PROJECT SUMMARY

Overview:

Recent advances in massively parallel DNA sequencing techniques have allowed an unprecedented assessment of capability and expression of metabolic function in marine systems via metagenomic and metatranscriptomic analyses, respectively. Over the next decade omics-based metabolic observations will become common place, but the ability to synthesize these data and translate them to properly constrained ocean-scale biogeochemistry is lacking. Regression-based analyses of metabolic function will allow some assessment of marine ecosystem function, but what is needed are mechanistic-based ocean biogeochemistry models that can integrate these new observations to further both our understanding of marine biogeochemistry as well as increase our prognostic capabilities.

Intellectual Merit :

This project builds upon the Darwin Project, a trait and selection based modeling approach for describing marine plankton communities and biogeochemical cycles. The approach relies on local competition to select from a diverse population and determines the functional characteristics of microorganisms that mediate biogeochemical cycles. Key challenges are defining and constraining the costs and benefits of key traits and functionalities. Here we proposed to combine this selection-based modeling approach with a distributed metabolic network perspective previously develop to facilitate calculating reaction thermodynamics. This will provide mechanistic and quantitative description of key metabolic functions and allow the new model to be directly mappable to omics-based observations. The project will utilize new modeling design criteria based on the maximum entropy production (MEP) conjecture to determine allocation of metabolic machinery and its expression, such as metabolic switching between nitrogen fixation and ammonium uptake. Model testing will rely on existing oceanographic surveys and observations. Once validated, the coupled model will be used to investigate losses of functional biodiversity, generalist versus specialists, temporal planktonic strategies as well as losses in community complementarity on ecosystem biogeochemistry. A significant output from the project will be a predicted, global functional biogeography, mapping metabolic function and expression (such as nitrogen fixation and ammonium oxidation), that can be tested with, and used to interpret, directed omics observations.

Broader Impacts :

Predicting how marine biogeochemistry will respond to global change is a pressing issue for society. This project will directly advance modeling skills for predicting such changes using the MEP paradigm that should prove to be more robust to extrapolation. In addition, research efforts will allow observations from the rapidly developing omics disciplines to be related to model predictions and allow measurements at the genome scale to be quantitatively extrapolated to global-scale biogeochemical processes. This project will support one postdoctoral scholar in this new interface between ocean biogeochemistry modeling, thermodynamics and molecular observations. The project will also support summer internships as part of the Woods Hole Partnership Education Program (PEP), which is a consortium of institutions committed to increasing diversity in Woods Hole, as well as support two undergraduate independent research projects per year as part of the Semester in Environmental Science Program at MBL. To broaden exposure of MEP concepts in marine biogeochemistry and explore its place in the broader context of recent advances in metabolic modeling and theory, we plan to propose a workshop to NSF?s Ocean Carbon & Biogeochemistry Program to be held in year 2. Ocean model code developed during the project will be open source and disseminated from mitgcm.org.

i. Response to Previous Review

This is the third submission of the proposal to Bio OCE, and we greatly appreciate the constructive input from the previous reviews. The proposal ranked highly in the last panel (E, E, E/VG, E/VG), but there were some concerns regarding integration and complementation of the Maximum Entropy Production (MEP) model to established approaches. In Section 3 we now show that the MEP model represents more of a natural extension to the existing Darwin model rather than a completely *de novo* synthesis. MEP is simply a means to implement Lotka's (1922) ideas within a quantitative framework that can be readily implemented within the existing Darwin model. We have also presented some proof-of-concept results in Sections 3 (**Fig. 1**) and Section 4.4.3 (**Fig. 4**). Finally, we have made our Broader Impacts more project specific and plan to hold a workshop in year 2 on thermodynamic approaches to understanding marine biogeochemistry to broaden appreciation of the MEP approach.

1. Introduction

Current ocean ecosystem and biogeochemistry models aim to represent the cycling of carbon and other elements, and the relevant structure of the microbial community, in order to provide diagnostic and predictive tools for ecological, biogeochemical and climate studies. Building on the early models of Riley (1946), Steele (1954), Fasham et al. (1990) and others, a paradigm shift occurred in the JGOFS era with the development of such models which resolve "functional types" of primary producers (e.g., Chai et al. 2002, Moore et al. 2002, Le Quere et al. 2005). The resolution of functional types of phytoplankton allows an explicit representation between contrasting "microbial loop" and "bloom" communities and the implication for new and recycled production, export efficiency and so on. More recently, models have explored the role of finer-scale diversity within the phototrophs (e.g. Bruggeman & Kooijman 2007, Follows et al. 2007, Dutkiewicz et al. 2009) which extends the application of such models to finer-grained views of biogeochemical function (e.g., Monteiro et al. 2010) and a link to a broader set of ecological questions (e.g. patterns of biodiversity, Barton et al. 2010) and the driving forces behind the "self-organization" of complex and diverse communities (e.g., Dutkiewicz et al. 2012).

This more recent ecological perspective emphasizes the role of traits and trade-offs in the organization of microbial communities. Locally, a population of diverse phenotypes is "sorted" by their relative fitness in the given environment. Fitter combinations of traits (physiological and biophysical characteristics of the organism which mediate its interaction with the environment) will be selected for, where less fit trait combinations will be excluded. However, less fit combinations may be fittest in some other region or season. Local bottom-up selection does not lead to complete competitive exclusion because of dispersal (e.g., Barton et al. 2010, d'Ovidio et al. 2010), top-down pressure from predators or viral activity ("kill the winner" effects; e.g. Thingstad & Lignell 1997, Prowe et al. 2012), phenotypic plasticity and (on appropriate timescales) genetic adaptation, though the latter processes are not yet resolved in ocean models. This selection-based approach is the foundation of the Darwin model (Follows et al. 2007, Dutkiewicz et al. 2009, Ward et al. 2012, Ward et al. 2014), which we seek to build upon in this proposal with an energy rather than organismal emphasis.

The focus of most, if not all, biogeochemical models has been on elucidating and quantifying the mechanisms by which planktonic communities interact and function. While this has been and will continue to be an extremely useful approach, placing emphasis on the organisms has two undesirable features: 1) the emergent properties of a system can be difficult to control because the model parameters operate at the individual rather than system scale and 2) the details necessary to quantify the planktonic community results in models with a relatively large number of biological parameters that are often poorly known and available observations are insufficient to resolve (Ward et al. 2010). Recent models have reduced parameter uncertainty by using competition as a means to select parameter values, but the focus still remains on the interactions of individuals or guilds and no selection is placed at the systems level.

However, organisms do not grow in isolation, but rather are tightly coupled to adaptation and evolution of other community members, which supply elemental resources, free energy, or both by their actions; the fitness landscape is not fixed but is as dynamic as the organisms themselves (Chave 2013). Cooperative

interactions (Wilson 1997, Nowak 2006) are particularly important for microbially-based ecosystems, such as marine planktonic systems, because it allows them to function more like coordinated, but distributed, metabolic networks (Vallino 2003) that organize to effectively extract nearly all free energy available to an ecosystem. Systems level assessments are not new and date back to the pioneering work of Lotka (1922) and H.T. Odum (1955, 1983), who were the first to speculate on the governing principles that may organize ecosystems. However, this systems-level view has not yet been explored or exploited in current ocean ecosystem and biogeochemistry models. Over the last decade, there has been a growing appreciation for the mechanism that drives system organization; namely, the dissipation of available energy. This perspective does not consider the individuals, but rather the properties exhibited by their collective actions. This is similar to, and derives from, thermodynamics that describes an ensemble of particles, not the particle interactions themselves. Because the free energy extracted by the community is ultimately dissipated as heat (i.e., biomass accumulation does not increase indefinitely but reaches a pseudo-steady state) organization of ecosystem function can be described as a type of maximum entropy production (MEP) process (Paltridge 1975, Dewar 2003, Dewar 2005). When applied to microbial systems, MEP provides a powerful organizing principle that provides a direction to the complex interactions of a community that ultimately derives from the actions of its individual constituents (Vallino 2010, Vallino 2011). Perhaps Lineweaver and Egan (2008) summarize this change in perspective best with "*This represents a paradigm shift from 'we eat food' to 'food has produced us to eat it'."* This new paradigm will serve as foundation for model development.

In this proposal we will focus our attention on the dissipation of potential energy via the synthesis of catalytic agents or, as Falkowski (2008) describes them, metabolic machines. The Darwin modeling approach will be used to determine, via competition, the allocation and expression of molecular machinery to metabolic pathways constrained by the availability of resources (i.e.,C, N, P, Si, Fe, etc) and information contained in the planktonic community's metagenome. Self-selected catalytic machines will then be assessed using entropy production as the metric, which we can use to falsify the premises that systems organize to maximize entropy production. This new thermodynamic addition to the Darwin model has several advantages, which we highlight below. We view this new branch as a complement, not replacement, to the existing Darwin models.

2. Problem Statement and Proposal Objectives

Marine biogeochemistry models that use dozens of biological parameters have far more degrees of freedom than can be constrained by available observations, even under base case scenarios (Vallino 2000, Ward et al. 2010, Ward et al. 2013). Consequently, there exist many different parameter configurations (technically an infinite number) that can reproduce the available observations, yet each parameterization will produce different dynamics when operating outside of the data envelope used for parameter calibration. That is, models interpolate observations well, but often extrapolate beyond them poorly. Yet, it is often the need for extrapolation that drives model development; because we are often interested in predicting how marine biogeochemistry may change under conditions that have not yet occurred, such as increased temperature or pCO_2, higher nutrient concentrations, decreases in pH, or losses in biodiversity. This proposal seeks to develop a modeling approach that is expected to extrapolate more gracefully, because it is based on the systems level property of maximum entropy production. The MEP approach will be added to the Darwin model and is based on a metabolic framework which exploits redox and energy constraints that has been developed and explored in idealized marine settings (Vallino 2010, Vallino 2011, Algar & Vallino 2014, Vallino et al. 2014) but has not yet been brought to bear in "realistic", basin or global scale simulations. The proposal objectives and associated advantages are as follows:

1. The model will be developed with entropy production as the design criterion as opposed to organismal growth, which will provide the following benefits:
   a. Enhance prognostic capabilities in changing environments.
   b. Provide a thermodynamic underpinning for key metabolic processes such as nitrogen fixation

and nutrient cycling.

2. A metabolic representation will be used to capture the collective actions of the planktonic community, the results from which can be more readily mapped to metagenomic and metatranscriptomic observations. *A major product of the proposed work is to present hypothesized/predicted global maps of metabolic function in the surface ocean*.
3. By describing microbes as a collection of "metabolic machines" (Falkowski et al. 2008), the costs and benefits of packaging multiple metabolic machines together, or separately, can be evaluated quantitatively. For example:
   a. What are the trade-offs between generalism (supporting numerous metabolic strategies) and specialism ("streamlining") or different extents of luxury uptake (e.g., Tozzi et al. 2004)?
   b. Under what environmental circumstances do these different strategies emerge as useful?
   c. Are systems dominated by specialists more likely to suffer loss of function?
4. The proposed model framework, in which selection occurs from an initialized population which resolves a broad and diverse set of metabolic functionality, holds promise for solving MEP-based problems. We may ask, from a diverse initialized community, is the subset that emerges as "fittest" that which maximizes entropy production of the whole system?
   a. Is the MEP solution more realistic than other solutions in any key regard?
   b. *We will explicitly test the hypothesis that the MEP principle operates to regulate the organization of complex and diverse ecosystems by comparting MEP solutions to oceanographic observations*.

In Section 3 we discuss the Darwin ocean model that forms the model foundation. In section 4 we develop the MEP-based addition. In Section 5 we outline the simulations that will be run and the hypotheses that will be tested. Section 6 describes the project's broader impacts.

3. Darwin Ocean Model

The ocean modeling framework to be developed and applied is the MITgcm including the fluid dynamics (Marshall et al. 1997) and biogeochemical components as well as some aspects of the existing ecosystems components (e.g., Follows et al. 2007, Ward et al. 2012), which we describe below.

3.1 MITgcm: Physical model, model configurations and biogeochemical tracers.

MITgcm is an ocean modeling platform based on an efficient primitive equation solver (Marshall et al. 1997) with an open-source code-base (http://mitgcm.org) which includes many extensions. The level-based model can be configured for any bathymetry (from 1-dimensional column to realistic eddy-resolving global simulations). Parameterizations of sub-gridscale processes include buoyancy and shear driven vertical mixing (e.g., Large et al. 1994) and the rectified effect of mesoscale eddies on tracer transport (Gent & McWilliams 1990). Biogeochemical and ecological modules (e.g., Parekh et al. 2005, Follows et al. 2007, Ward et al. 2012) are deeply integrated and modifications of these components – as proposed in Section 4 – are easy to implement in any physical configuration. Here we aim to develop, test and apply new biogeochemical and ecological model components in the context of three dimensional ocean circulations. For development we will employ a single basin (double gyre) configuration at low-resolution (Lévy et al. 2014) which extends beyond the box-model of Vallino (2011) but remains highly idealized. We will ultimately apply the new parameterizations in more realistic global settings where the physical state is constrained to be consistent with remote sensing and in situ hydrographic measurements (Wunsch & Heimbach 2007) (e.g. **Fig. 1**) (e.g., Follows et al. 2007, Dutkiewicz et al. 2009, Ward et al. 2012). The biogeochemical model resolves dissolved inorganic carbon (see, e.g. Follows et al. 2006) alkalinity, phosphate, nitrate, silicic acid, particulate (detrital) and dissolved organic forms of carbon, nitrogen, phosphorus and iron. Iron chemistry includes explicit complexation with an organic ligand of specified concentration, scavenging by particles, and a representation of aeolian and sedimentary sources (Dutkiewicz et al. 2012).

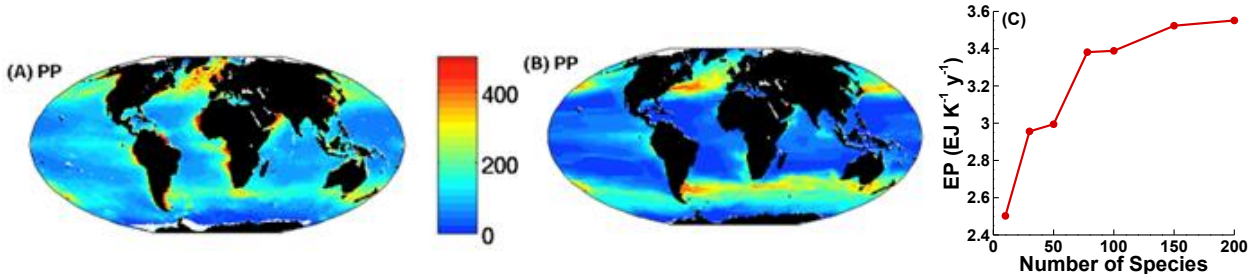

***Fig. 1**. (A) Global annual primary production derived from remote sensing products (g C m^{-2} yr^{-1}). (B) Annual primary production (g C m^{-2} yr^{-1}) as simulated by the Darwin model (reproduced from Follows et al, 2007; supplementary material). (C) Simulated global entropy production as a function of the number of phytoplankton phenotypes ("species") seeded in the simulation. Increasing the coverage of trait-space in the species leads to enhanced resource utilization, higher global primary productivity, and higher entropy production (EP), which is consistent with MEP theory. The simulation is based on that described in Follows et al (2007; the "Darwin Project model") where the traits of each individual phytoplankton species are drawn stochastically from specified distributions and EP is estimated from GPP (Unpublished study: J. Bragg, S. Dutkiewicz, M. Follows, J. Vallino).*

3.2 Current Ecosystem Model.

In the Darwin model of Follows et al. (2007) and Dutkiewicz et al. (2009), the ecological sub-model resolves order 100 phenotypes of potentially viable, virtual phytoplankton and (micro) zooplankton. All are initialized and interactions of the community and environment organize both through a natural selection process. Biogeochemical and biological tracers interact through the formation, transformation, and remineralization of organic matter. Inorganic nutrients are taken up by phytoplankton, and these are grazed by zooplankton. Mortality, sloppy feeding, and egestion transfer living organic material into sinking particulate and dissolved organic detritus, which is currently returned to inorganic form through a simple parameterization of bacterial remineralization. The time-dependent change in the biomass of each of the modeled plankton types is described in terms of growth, sinking, grazing, and other mortality, alongside transport and mixing by the fluid flow. Growth rate is a function of temperature, light and available nutrient resources. Up to 6 taxonomic "functional groups" are resolved; analogs of *Prochlorococcus*; picoeukaryotes, *Synechococcus*, coccolithophores, dinoflagellates and diatoms (Dutkiewicz et al. 2015).

Following many earlier examples, the Darwin model (as described by Follows, 2007; Dutkiewicz et al, 2009) employed a governing equation for the biomass of phytoplankton type *j*, i.e. *Pj* (moles R m^{-3}), similar to this simplified form:

$$\frac{\partial P_j}{\partial t} = \mu_{o,i}(T,I)\frac{R}{R+K_{R,j}}P_j - G_{j,k}(P_j Z_k) - m_j P_j - \nabla\cdot(uP_j) + \nabla\cdot(\kappa\nabla P_j). \quad (1)$$

Here the first term on the right parameterizes population growth using Monod kinetics where R is the limiting resource, $\mu_{o,i}(T,I)$ (s^{-1}) is a temperature and light dependent maximum growth rate and $K_{R,j}$ (mol m^{-3}) is a half-saturation for R limited growth. The second term on the right indicates grazing by predator Z_k, the third term on the right represents losses including respiration and sinking, and the last two terms represent advection and diffusion by the ocean currents and turbulence (see, e.g. Dutkiewicz et al. 2009 for more complete forms). The parameterizations and parameter values associated with growth and loss processes encapsulate the traits of the organisms. A key challenge in "trait-based" modeling, (such as the Darwin model) is to appropriately capture the costs and benefits of particular traits such that the virtual organisms self-assemble into appropriate populations with appropriate biogeochemical functionality.

Focusing on growth, $\mu_{o,i}$ and $K_{R,j}$ are key traits which differentiate key ecological strategies: opportunistic "bloomers" such as diatoms are characterized by high maximum growth rates and the ability to grow fast

in resource replete conditions. In contrast, gleaners such as *Prochlorococcus* eke out a living in very oligotrophic environments where their high nutrient affinities (related to low $K_{R,j}$) make them the fittest competitors for scarce resources. These key traits depend, in part, on body size and this has been demonstrated empirically and with robust mechanistic theory for volume e.g. (Irwin et al. 2006, Litchman et al. 2007). For example, maximum growth rates (within taxonomic groups) scale with cell size as a power law. To reflect these constraints, we have resolved multiple size classes and employed allometric constraints to key structure traits in the Darwin model (e.g. Ward et al., 2012; 2014)[1].

Though maximum growth rates scale with size within taxonomic groups, there are significant differences between those groups. For example, picocyanobacteria have relatively slow maximum growth rates despite being the smallest cells, and size-for-size diatoms typically have higher maximum growth rates. These inter-guild differences reflect diversity of physiological and metabolic organization that is not fully understood. They represent an additional dimension in trait space beyond body size. Similarly, such metabolic organization determines diverse heterotrophic and chemotrophic lifestyles which are currently absent or poorly represented in the Darwin model. The MEP approach proposed here (described below) provides a strategy for encapsulating such unresolved metabolic organization and trade-offs by hypothesizing that, though the details are not explicit, they conform to a broader governing principle. Significantly, since the MEP-based approach described in Section 4 is couched in terms of analogues of $\mu_{o,i}$ and $K_{R,j}$ and Monod-like kinetics, it will be technically straightforward to reframe the Darwin model to an MEP-based structure. In turn, the fitness and selection-based self-organization of the Darwin model framework will provide an efficient alternative to the optimization of parameters previously employed in the MEP approach (details below).

4. MEP-Based Biogeochemistry Model

Our objective is to develop a new branch to the existing Darwin model described above that replaces the food web representation with a metabolic reaction network that is more amendable for implementing MEP. While we considered modifying the current Darwin ecosystem model, its control variables (i.e., adjustable parameters) are design to influence competition nuances between organisms for resources or prey, and not on the direct dissipation of free energy. Because the MEP approach is thermodynamically based, the structure of the planktonic community is not explicitly constrained. For instance, bacteria grazed by protists or lysed by viruses are indistinguishable processes. In fact, the MEP conjecture indicates that there should be many different community configurations that give rise to the same entropy production, which has been observed in microbial systems (Fernandez et al. 1999, Wittebolle et al. 2008, Vallino et al. 2014). Because of this organismal interchangeability, the ecosystem is represented by a distributed metabolic network (Vallino 2003) that is capable of expressing known metabolic functions, such ammonium oxidation or N_2 fixation. Each metabolic function is catalyzed by an associated "biological structure", $\mathbb{S}_j$, that is synthesized from resources and free energy available in the environment. While this approach does not provide details of food web structure, as does the current Darwin model, the MEP-based model should be more robust for extrapolated environmental conditions, such as occurred in the distant past or may occur in the future. Using an analogy, the current Darwin model is similar to a weather model that can predict detailed state information, while the MEP model functions more like a climate model that is expected to have more robust long-term forecasting at the expense of details. As discussed below, the MEP model will use the approach of populating a system with a large number of competitors and allowing communities to self-assemble, as pioneered in the

[1] In parallel, we have extended the model to resolve flexible internal stores of nitrogen, iron and other elements (Ward et al. 2012, 2014) relative to a cellular carbon pool (e.g., Droop 1968). A dynamic pigment pool regulates carbon fixation in accord with the model of Geider et al. (1998). Recent developments allow simple and flexible configuration of the Darwin model such that the user defines whether Monod or Droop kinetics are used on organism-by-organism and nutrient-by-nutrient basis.

Darwin model. This will allow us to solve an otherwise difficult optimization problem (see Section 4.4.3).

4.1 MEP Background

Numerous theories describing ecosystem organization and function have been proposed in theoretical ecology dating back to at least Lotka (1922), who proposed ecosystems organize to maximize power. Here, however, we are particularly interested in the maximum entropy production (MEP) conjecture (Paltridge 1975, Dewar 2003, Dewar 2005, Niven 2009), which asserts that steady state, non-equilibrium systems with many degrees of freedom will likely be found in a macrostate that maximizes internal entropy production, where entropy here refers to the classic thermodynamic definition of Clausius, Gibbs and Boltzmann (i.e., the dispersal of energy; Clausius 1867). If internal self-organization, such as vortices and macroscopic structures, facilitates internal entropy production, then those structures will likely develop (Lorenz 2003), but the theory makes no distinction between biotic or abiotic systems. Similar to equilibrium thermodynamics that requires isolated systems to be found in a state of maximum entropy, MEP indicates that non-equilibrium systems will head towards equilibrium along the fastest possible pathway, which may be facilitated by internal system organization. That is, they will dissipate free energy as fast as possible within the constraints imposed on the system (Makela & Annila 2010, Vallino 2010). Several phenomena appear consistent with MEP, including planetary-scale heat transport (Lorenz et al. 2001, Kleidon et al. 2003), laminar to turbulent flow transition (Martyushev 2007), plant evapotranspiration (Wang & Bras 2011), atmospheric and ocean circulation (Kleidon et al. 2003, Shimokawa & Ozawa 2007), processes in the critical zone (Quijano & Lin 2014) and many others (see Dewar et al. 2014). MEP provides directionality to evolution of the biosphere, in that it should progress towards states of higher entropy production. The global succession of anoxygenic phototrophs by oxygenic phototrophs is one example of this progression.

While actively discussed by the community, uncertainty remains regarding the scale over which MEP applies (Lucia 2012, Martyushev 2013). Our preliminary work indicates that MEP applies at a systems level as defined by the extent of matter and information connectivity (Vallino 2011). Hence, a single bacterium does not maximize entropy production because it does not dissipate all chemical potential (some free energy is used to produce more bacteria instead of CO_2 and water). However, a microbial community does achieve a MEP state because the growth of each organism is more or less consumed by a predator, and the predator by its predator, and so on. Total biomass of the community will increase until either all incoming energy is consumed (energy limited), or resources limit the amount of biomass that can be supported (resource limited). For instance, the surface ocean is resource limited (typically by N, P or Fe), while the deep ocean is energy limited. Regardless of the limitation, the ecosystem operates near a pseudo steady state where the assimilated free energy is simply dissipated as heat, since it is not stored in other potentials (**Fig. 2**), which is the definition of thermodynamic entropy. Microbes and microbial consortium so effectively organize that nearly all chemical potentials found in the environment are readily dissipated; it has been estimated that reaction free energy potentials as low as 9 kJ mol^{-1} can be exploited by microbes (Hoehler 2004). Hence, microbial communities are exquisitely evolved and organized to extract, and subsequently dissipate, chemical potentials and electromagnetic radiation.

We have developed a theoretical framework for describing microbial biogeochemistry as a type of dissipative system governed by maximum entropy production (Vallino 2010, Vallino 2011) that can be used to direct how communities may organize at the metabolic level. The MEP model is founded on the hypothesis that microbial communities evolve, adapt and organize to extract as much free energy from the environment as available resources (N, P, S, etc.) and information (i.e., metabolic capability) allow (**Fig. 2**). Biodiversity, which we view as a reservoir of genomic information, is critical, as it

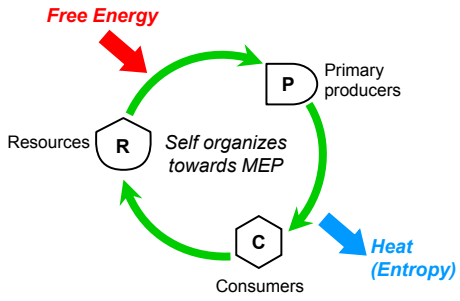

***Fig. 2**. How an ecosystem functions as a purely dissipative system. Either free energy input or resources will limit the cycling of the system once organized.*

ultimately determines the set of metabolic machines (Falkowski et al. 2008)—catalysts in particular—and metabolic functions that can be constructed from the available environmental resources. This information also includes designs for the metabolic machines that turnover machinery, namely protists, predatory bacteria and viruses, that allow the system to be dynamic and adaptive to changes in environmental drivers via reallocation of biological structure by predation.

4.2 Temporal strategies and MEP

An intriguing hypothesis to derive from application of the MEP principle to microbial systems is a proposed distinction between biotic and abiotic systems. When entropy production (EP) is maximized instantaneously, then no bacteria are produced and any biomass initially present slowly consumes itself. The results are similar to how fire behaves. If instead, EP is maximized over a period of time, then the model performs in a manner similar to real microbial communities. In fact, the model can simulate experimental methanotrophic microcosms accurately using only two adjustable parameters (Vallino et al. 2014). The approach also works well for describing metabolic switching between denitrification, anammox and dissimilatory nitrate reduction to ammonium (DNRA) pathways during anaerobic nitrate reduction (Algar & Vallino 2014). Our hypothesis is that abiotic systems, such as fire or a rock rolling down a hill, maximize instantaneous EP by following a steepest descent trajectory along the potential energy surface, but such Markovian trajectories can lead to metastable states (the flame gets extinguished or the rock gets trapped in a ditch) that limit total EP over time. Living systems choose alternate pathways down the potential energy surface that avoid metastable traps by forecasting future events. While this leads to lower instantaneous EP, averaged EP is increased (combustion persists–for 3.8 Gyr now–or the rock rolls all the way down the hill). Living systems discover alternate pathways via evolution and natural selection and store this information in the metagenome. For example, storage of internal energy reserves (e.g., starches) permit metabolic function to persist when external energy sources become temporarily unavailable, such as in animal hibernation and plant dormancy over winter (persistent combustion). Circadian rhythm is another example that allows phytoplankton to "predict" the sun will return and can orchestrate metabolic machinery appropriately before sunrise (Dvornyk et al. 2003). Recently, similar temporal strategies have been discovered for entire microbial communities (Ottesen et al. 2014). While we have implemented ideas from optimal control theory to specifically address biological anticipatory control (Vallino et al. 2014), temporal strategies within the Darwin model will be implemented using internal storage pools.

4.3 Structure of MEP model

The structure of the MEP model is based on the following simple logic. All non-equilibrium systems attempt to race down free energy surfaces towards equilibrium (maximum entropy), but many get trapped in meta-stable states due to activation energy barriers. For instance, at room temperature methane and oxygen remain relatively stable. Introduction of the proper catalyst, however, can free systems from meta-stable states allowing them to reach equilibrium quickly (i.e., MEP). Consequently, microbial systems can be viewed simply as complex catalysts that hasten the destruction of chemical potential or electromagnetic radiation. Because biological catalysts are largely protein, in oxidizing environments they contain considerable chemical potential; they are far from equilibrium. The optimization objective and model design criterion then is to produce enough catalyst to maximize the dissipation of available free energy from resources available in the environment (C, N, P, S, Fe, etc), while keeping the amount of catalyst at a minimum, *which is the first fundamental design principal of the MEP model*. Over synthesis of catalyst (i.e., biomass) would not lead to an MEP state because free energy contained in the excess catalyst could have been dissipated producing more entropy.

To illustrate model design concepts, consider the following simple two-reaction "network" for glucose oxidation that consists of two types of catalyst, $\mathbb{S}_1$ ("producer") and $\mathbb{S}_2$ ("consumer"):

$$r_{1,1}\colon\ \mathrm{CH_2O} + \varepsilon_1\gamma\mathrm{NH_3} + (1-\varepsilon_1)\mathrm{O_2} \xrightarrow{\mathbb{S}_1} \varepsilon_1\mathbb{S}_1 + (1-\varepsilon_1)\mathrm{H_2CO_3} \tag{2}$$

$$r_{1,2}\colon\ \pi_1 \mathbb{S}_1 + \pi_2 \mathbb{S}_2 + (1-\varepsilon_2)\mathrm{O}_2 \xrightarrow{\mathbb{S}_2} \varepsilon_2 \mathbb{S}_2 + (1-\varepsilon_2)(\mathrm{H_2CO_3} + \gamma \mathrm{NH_3}) \tag{3}$$

For simplicity we have assumed the chemical composition of the biological structures, or catalysts $\mathbb{S}_j$, are given by $\mathrm{CH_2O(NH_3)_\gamma}$, and π_i represent "feeding preference", if any, by $\mathbb{S}_2$. This simple two-reaction network has several important features critical to the MEP approach. 1) Each reaction is catalyzed by its respective biological structure, $\mathbb{S}_j$, so reaction rates depend on the concentration of the biological structure, $c_{\mathbb{S}_j}$, just as they do for any bacterial growth model (i.e., they are autocatalytic reactions that can grow exponentially). 2) The coupled reactions can operate in a futile cycle perpetually turning over biological structure fueled by glucose oxidation (e.g., **Fig. 2**). 3) The thermodynamic efficiency parameter, ε_j, selects the degree to which reduced organic carbon is either converted to more catalyst (pure anabolic reaction as $\varepsilon_j \to 1$) or oxidized to CO_2 and H_2O (pure catabolic reaction as $\varepsilon_j \to 0$), where the latter reaction produces large amounts of entropy provided sufficient catalyst is present. Note, the free energy of reaction $r_{1,1}$ remains negative (i.e., can occur spontaneously) even when $\varepsilon_j = 1$, because, contrary to popular belief, living organisms are ***not*** low entropy structures (Morrison 1964, Martyushev 2013).

To describe the growth kinetics of reactions $r_{1,1}$ and $r_{1,2}$, and all others in a reaction network, we use the following general expression (Vallino 2011, Vallino et al. 2014),

$$r_{i,j} = \nu_j^* \varepsilon_j^2 F_T\left(\Delta_{r_{i,j}} G\right) c_{\mathbb{S}_j} \prod_{k=1}^{n_c} \left(\frac{c_k}{c_k + \kappa_j^* \varepsilon_j^4}\right)^{\Lambda_{i,j,k}} \omega_{i-1,j} \prod_{l=i}^{n_{\mathbb{S}_j}-1} (1-\omega_{l,j}). \tag{4}$$

where j corresponds to a particular catalyst and i a sub-reaction catalyzed by $\mathbb{S}_j$. This equation is similar to the classic multi-substrate Monod growth model where c_k are substrate concentrations, (c_{CH_2O} and c_{NH_3}) or ($c_{\mathbb{S}_1} + c_{\mathbb{S}_2}$) in the above example, but the maximum specific uptake rate is replace by $\nu_j^* \varepsilon_j^2 F_T\left(\Delta_{r_{i,j}} G\right)$ and the half saturation, or Monod, constant is given by $\kappa_j^* \varepsilon_j^4$. The term $F_T\left(\Delta_{r_{i,j}} G\right)$ is the thermodynamic driving force (Jin & Bethke 2003, Jin et al. 2013) that limits reaction kinetics as Gibbs free energy of reaction, $\Delta_{r_{i,j}} G$, approaches 0 and provides the tradeoff between reaction speed and efficiency. As anabolic-catabolic coupled reactions approach 100% thermodynamic efficiency ($\varepsilon_j \to 1, \Delta_r G \approx 0$), they must proceed reversibly, so infinitely slowly, which explains why evolution has not favored growth efficiencies at or near 100% (Pfeiffer & Bonhoeffer 2002). *This represents the second fundamental design principle of the model; efficient reactions proceed slowly, while fast reactions dissipate large amounts of free energy* (*low efficiency*).

The parameters ν_j^* and κ_j^* are chosen to capture bacterial growth kinetics observed in nutrient deplete (i.e., oligotrophic, $\mu \ll 1\ \mathrm{d}^{-1}$) to nutrient abundant (i.e., eutrophic, $\mu > 50\ \mathrm{d}^{-1}$) conditions. That is, ν_j^* and κ_j^* are independent of community composition and are typically fixed at 350 d^{-1} and 5000 mmol C m^{-3}, respectively, for all reactions. The exponent $\Lambda_{i,j,k}$ is set to either 0 or 1 depending on reaction stoichiometry for the n_c state concentration variables, c_k, and $\omega_{l,j}$ determines how $\mathbb{S}_j$ is partitioned to its associated $n_{\mathbb{S}_j}$ sub-reactions, where $\omega_{0,j} = 1$ for all reactions (see Section 4.4 below). The advantage of Eq. (4) is that it only depends on the value of ε_j, where values of ε_j between 0 and 1 produce a family of curves that describes kinetics over substrate concentrations from nM to mM (**Fig. 3**).

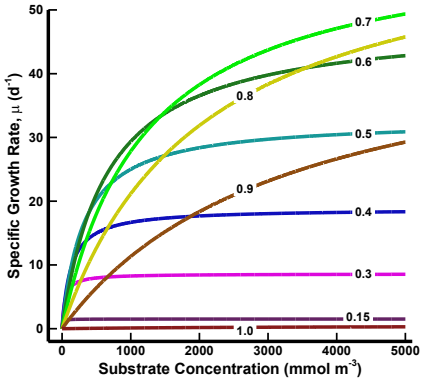

Fig. 3. *Growth kinetics parameterized by* $\boldsymbol{\varepsilon_j}$ *as a function of substrate concentration.*

Internal entropy production rate, $\dot{\sigma}_r$ (J K^{-1} d^{-1}), associated with a reaction network of $n_{\mathbb{S}}$ catalysts is readily calculated from the reaction rates, $r_{i,j}$ (mmol m^{-3} d^{-1}), Gibbs free energy of reaction, $\Delta_{r_{i,j}} G$ (J mmol^{-1}), element volume, V (m^3), and temperature, T (K), as given by (Vallino 2010),

$$\dot{\sigma}_r = -\frac{V}{T}\sum_{j=1}^{n_{\mathbb{S}}}\sum_{i=1}^{n_{\mathbb{S}_j}} r_{i,j}(\varepsilon_j, \omega_{i,j}, \mathbf{c})\Delta_{r_{i,j}}G(\varepsilon_j, \mathbf{c}). \quad (5)$$

where $n_{\mathbb{S}_j}$ is the number of subreactions catalyzed by $\mathbb{S}_j$. We account for concentration of reactants and products as well as for activity coefficients in $\Delta_{r_{i,j}}G$ calculations (Alberty 2003). We also explicitly account for proton dissociation equilibria between chemical species via pH, so "H_2CO_3" and "NH_3" in the above reactions represent $H_2CO_3 + HCO_3^- + CO_3^{2-}$ and $NH_3(aq) + NH_4^+$, respectively (all weak acids and bases are accounted for similarly) (Alberty 2003). As evident in Eqs. (2-4), both reaction rates and associated Gibbs free energy of reaction depend on the thermodynamic efficiency and partitioning control variables $(\varepsilon_j, \omega_{i,j})$ and the concentration of state variables, **c**. Though not shown here, we also calculate entropy production from mixing, $\dot{\sigma}_m$, but these terms are only a small fraction of the entropy of reaction, $\dot{\sigma}_r$ (see Vallino 2011 for details).

To reiterate, our MEP model design places all degrees of freedom that normally reside in adjustable parameters, such as half saturation constants, maximum growth rates, etc., within the thermodynamic reaction efficiency parameters, ε_j. We note that the growth term in equation (4) is analogous to that of the Monod kinetics version of the Darwin model in equation (1): biological structure $\mathbb{S}_j$ replaces biomass P_j etc. Hence it is technically simple to implement in the ocean model framework.

4.4 MEP integration with the Darwin Model

The MEP-based biogeochemistry model will be incorporated into the MITgcm taking advantage of the existing ecosystem modeling framework (e.g. structures to deal with multiple tracers transported by the physical flow, carbonate chemistry, interaction of state variables, etc.) and will initially consist of biological structures catalyzing aerobic phototrophy and heterotrophy, including N_2 fixation and nitrification. Each biological structure, $\mathbb{S}_j$, includes one or more catabolic reactions coupled with one or more anabolic reaction, except for predation (**Table 1**). For instance, canonical photosynthesis consists of catabolic reaction c_1 combined with anabolic reaction a_5, assuming the phytoplankton specializes only in NH_3 uptake for nitrogen requirements (**Table 1**). This framework also allows us to include generalists; for instance, combining reactions c_1 with $a_5 + a_6 + a_7 + a_8$, would allow the phytoplankton to use all dissolved inorganic nitrogen sources as well as fix N_2. Similarly, combing reactions c_1 with c_2 would permit mixotrophic growth. When a biological structure contains multiple catabolic and/or anabolic reactions, the partitioning of catalyst to each sub-reaction is governed by the $\omega_{i-1,j}$ control variable in Eq. (4); consequently, there are some costs associated with being a generalist. For instance, the NH_3 uptake specialist ($\omega_{0,j} = 1$) allocates

Table 1*. Catabolic, c_i, anabolic, a_i, and first order, k_i, reactions used to represent the metabolic capability of an aerobic planktonic community, where $\mathbb{S}$ has the C-normalized elemental composition of* $CH_\alpha O_\beta N_\gamma P_\delta Si_\zeta Fe_\xi$*, and* $[\![NH_3, HNO_2, HNO_3, N_2]\!]_i$ *represents the choice of the* N *source. We have excluded details on* P, Si *and* Fe *for clarity.*

Rxn	Reaction Type
	Free energy dissipating reactions (catabolic reactions)
c_1	$h\nu_H \xrightarrow{\mathbb{S}_{c1}} h\nu_L$
c_2	$CH_2O + O_2 \xrightarrow{\mathbb{S}_{c2}} H_2CO_3$
c_3	$NH_3 + \frac{3}{2}O_2 \xrightarrow{\mathbb{S}_{c3}} HNO_2 + H_2O$
c_4	$HNO_2 + \frac{1}{2}O_2 \xrightarrow{\mathbb{S}_{c4}} HNO_3$
	Biological structure synthesis (anabolic reactions)
$a_{1\ldots4}$	$CH_2O + \gamma[\![NH_3, HNO_2, HNO_3, N_2]\!]_i + \delta H_3PO_4 \xrightarrow{\mathbb{S}_{a1\ldots4}} \mathbb{S}_{a1\ldots4}$
$a_{5\ldots8}$	$H_2CO_3 + \gamma[\![NH_3, HNO_2, HNO_3, N_2]\!]_{i-4} + \delta H_3PO_4 \xrightarrow{\mathbb{S}_{a5\ldots8}} \mathbb{S}_{a5\ldots8}$
	Biological structure turnover (predation)
p_1	$\mathbb{S}_i + O_2 \xrightarrow{\mathbb{S}_{p1}} \varepsilon_{p1}\mathbb{S}_{p1} + \varepsilon H_2CO_3 + (1-\varepsilon)C_D + \gamma(\varepsilon NH_3 + (1-\varepsilon)N_D) + \delta(\varepsilon H_3PO_4 + (1-\varepsilon)P_D) + \xi(\varepsilon Fe + (1-\varepsilon)Fe_D) + \cdots$
	First order decomposition reactions
k_1	$C_D \longrightarrow CH_2O$
k_2	$N_D \longrightarrow NH_3$
k_3	$P_D \longrightarrow H_3PO_4$
k_4	$Fe_D \longrightarrow Fe$

100% of its biomass, $c_{\mathbb{S}_j}$, to NH_3 uptake, while a generalist that can consume both NH_3 and HNO_3 may only allocate 50% of $c_{\mathbb{S}_j}$ to NH_3 uptake if $\omega_{1,j} = 0.5$. Consequently, the specialist will be able to grow twice as fast as the generalist, but of course the generalist can still grow on HNO_3 when NH_3 concentration is very low. We realize that this is not a perfect representation of generalist versus specialist, but our general design criterion is to incorporate known metabolic constraints while introducing no additional parameters. This design criterion is also evident in the reaction associated with biological structure turnover (i.e., predation; p_1 in **Table 1**). In this reaction all resources needed by the predator are met by the prey, and the control parameter that governs thermodynamic efficiency, ε_j, also determines how much of the prey gets converted into labile C, N, P and Fe (as CH_2O, NH_3, H_3PO_4 and Fe) versus detrital (i.e., refractory) pools: C_D, N_D, P_D and Fe_D. The rational is that the more efficient the reaction is, the slower the reaction proceeds, so more C, N, P and Fe of the prey will be extracted, while faster reaction rates with lower ε_j will result in more refractory material produced. Again, this is not a perfect representation, but it does capture expectation and does not introduce any new parameters. Finally, the most difficult material to consider is the refractor, or detrital pools: C_D, N_D, P_D and Fe_D. Because detritus is often a poorly defined biopolymer, we represent its decomposition into labile pools via uncatalyzed first order decay reactions (k_i, **Table 1**). These type of poorly define biological parameters are in general what we try to avoid including, but it is impossible to remove them all.

In this formulation, the reaction control variables ε_j and $\omega_{i,j}$ are analogous to trait parameters in the current Darwin model. Importantly, the control parameters are all bounded between 0 and 1, which facilitates implementation.

4.4.1 Primary Production, Photoautotrophy To date, we have successfully used MEP to model microbial systems consisting of chemoorganoheterotrophs, chemolithoautotrophs and chemolithoheterotrophs. This will provide an excellent framework for incorporating a broader set of microbial lifestyles into the MITgcm/Darwin framework. However, we have not yet incorporated phototrophs into the MEP framework. Consequently, we describe here some of the details associated with combining catabolic and anabolic reactions of **Table 1** for photoautotrophs that use only NH_3, but anoxygenic photoautotrophs based on S cycling can also be readily included for studies investigating biogeochemistry in oxygen minimum zones. The combined oxygenic photoautotrophy reaction ($c_1 + a_5$ reactions, **Table 1**) in abbreviated form is,

$$\varepsilon_j H_2CO_3 + \varepsilon_j \gamma NH_3 + n_{\gamma_{1,j}} h\nu_H \xrightarrow{\mathbb{S}_j} \varepsilon_j \mathbb{S}_j + \varepsilon_j O_2 + n_{\gamma_{1,j}} h\nu_L. \quad (6)$$

For clarity, we have not shown the other elements necessary for $\mathbb{S}_j$ synthesis, and other combinations of reactions in **Table 1** could also be used. In this reaction, high frequency light, ν_H, is converted to infrared light, ν_L, as a function of intercepted photons captured by phototrophs (see below), where h is Planck's constant. The parameter $n_{\gamma_{1,j}}$ (mmol-photon mmol-rxn^{-1}) is determined such that as ε_j approaches 1, 100% of the usable light energy is transferred to chemical potential, so that the overall reaction free energy equals 0. Of course, this means the reaction proceeds infinitely slowly. At the other extreme, when ε_j approaches 0, all light energy is dissipated as infrared radiation without any biosynthesis.

A radiative transfer framework, already present in the MITgcm/Darwin model system, will be used to determine light interception by both water and particles. In the MEP formulation, any interception of light not reflected leads to entropy production if the light is not converted to another energy potential, so silt laden water dissipates light energy as effectively as phytoplankton. *Consequently, increasing entropy production can be achieved via the formation of particles when none are present, which is effectively what phytoplankton achieve in the MEP context*. To formulate a MEP model for phototrophs, we calculate the light intercepted by each biological structure in a layer of water Δz_l thick at a depth of z_l, which we define as $I_{\mathbb{S}_j}(\Delta z_l)$. By increasing the concentration of $c_{\mathbb{S}_j}$ in a layer, the amount of light intercepted and potentially dissipated as heat increases, but the function does saturate at high $\mathbb{S}_j$ concentrations, as expected. The reaction rate equation (4) is modified for phototrophs because the maximum rate is set by the photon flux, so the rate parameter, ν^*, is replaced by,

$$v^*(\text{phototrophs}) = \frac{I_{\mathbb{S}_j}(\Delta z_l)}{n_{\gamma_{i,j}} \Delta z_l} \frac{1}{c_{\mathbb{S}_j}}. \quad (7)$$

Free energy for the phototrophic reaction (6) and associated entropy production can be readily calculated for a given value of ε_j, concentrations of substrates and products and accounting for the thermodynamic efficiency for converting electromagnetic radiation to chemical potential (Candau 2003).

4.4.2 Internal Storage, Temporal Strategy and Variable Stoichiometry We assume that the elemental composition for biological structure, $\mathbb{S}_j$, that produces catalytic activity is fixed, and its elemental composition is given by, $CH_{\alpha_j}O_{\beta_j}N_{\gamma_j}P_{\delta_j}Fe_{\xi_j}$; however, we allow C, N, P and Fe (and any other elements) to be internally stored and concentrated as CH_2O ($C_{\mathbb{S}_j}$), NH_3 ($N_{\mathbb{S}_j}$), H_3PO_4 ($P_{\mathbb{S}_j}$) and Fe ($Fe_{\mathbb{S}_j}$) in a manner similar to Droop's (1973) approach and used in the Darwin model (Ward et al. 2012). The combined core composition with variable internal pools allows elemental composition of $\mathbb{S}_j$ to vary over time and space as well as a means to implement temporal strategies as discussed in Section 4.2. Additional metabolic reactions to those in Table 1 are used to account for resource uptake into internal pools that catalyst is synthesized from. This introduces new state variables for the internal pools for each biological structure, but these additions will allow us to directly assess the importance of K- versus R-selection (Pianka 1970, Salmaso et al. 2015) on entropy production over global scales. We will use internal pool volume to set biological structure size and consequently its sinking velocity as used in the Darwin models (Ward et al. 2012, 2014).

4.4.3 Monte Carlo with competition replaces optimization In applications to date, we have determined how thermodynamic efficiencies, ε_j, and partitioning of $\mathbb{S}_j$ to sub-reactions given by $\omega_{i,j}$ vary over time by solving a receding horizon optimal control problem (Vallino 2010, Algar & Vallino 2014, Vallino et al. 2014) or a standard optimization problem for steady state systems (Vallino 2011). While this approach works well for zero-dimensional problems and has been extended to 1D problems with some success, the optimization procedure becomes computationally demanding for 2 or more dimensions. We believe the approach developed for the Darwin models can be used to solve the MEP problem for higher dimensions.

In a simple two-box MEP ocean model (Vallino 2011), a Monte Carlo approach showed that multiple solutions produce entropy at very near the maximum rate; however, in the steady state analysis we did not examine competition between multiple ε_j and $\omega_{i,j}$ parameterizations for each $\mathbb{S}_j$ at the same time. The Darwin model uses a combination of Monte Carlo-like simulation coupled with Darwinian-like competition that allows the model to find the optimum solution for a given subset of traits or phenotypes from the set of all possible traits. This approach can also be used to solve the MEP problem in higher spatial dimensions in the MITgcm. In the simplest version of the MEP ocean model (specialists only), we would have five catalysts types (**Table 1**, $c_1 \ldots c_4, p_1$), each having approximately 6 control variables (one ε_j and 5 ω_{ij}). As in previous implementations of the Darwin model (Follows et al. 2007), we will populate the model with approximately 100 (or more) $\mathbb{S}_j$ each with a different parameterization selected at random for ε_j and ω_{ij}. Because some parameterizations will not produce viable catalysts, we can cull them from available parameter space to speed subsequent searches.

As a first proof of concept, encouraged by previous reviewers, we have tested the Darwin approach for solving the MEP problem using the two-box model (Vallino 2011). Preliminary results show that food web structure is important, but even simple viral-like predation structures (unique prey-predator pairs) can find MEP solutions purely by competition (**Fig. 4a**). Interestingly, these models do not exhibit competitive exclusion and retain a "rare biosphere" (Sogin et al. 2006) (**Fig. 4b**). Furthermore, we demonstrate that increasing the number of degrees of freedom in the current Darwin model (i.e., number of species) results in greater entropy production (**Fig. 1c**), which is consistent with MEP theory regarding sufficient degrees of freedom (Ozawa et al. 2003).

5. Model simulations and Hypotheses Testing

The MEP metabolic network addition to the Darwin model will be used to test the following hypotheses:

Hypothesis I: The metabolic network with associated biological structures that self-assembles in a given environment from a diverse set of control variables will maximizes entropy production.
Approach: We will populate a global ocean model with a functionally diverse population of virtual plankton to test whether the emergent community structure reflects the MEP state.

Hypothesis II: We expect that dynamic environments will host a greater percentage of generalist versus specialists, and the opposite distribution to be the case for stable environments.
Approach: We will examine the communities in dynamic environments, such as in temperate zones, with those associated with stable environments, such as in the tropics. Specialists are defined as those biological structures with only one anabolic and one catabolic reaction, while generalists are defined with more than one anabolic or catabolic reaction (**Table 1**). Environmental dynamics will be based on the fluctuation of nutrients in a localized area.

Hypothesis III: Implementing a MEP-based representation of the ecosystem and biogeochemical cycles will produce a more robust model that can be extrapolated beyond data used for its calibration. This will improve forecasts of how future global changes may alter functional biogeography of the ocean.
Approach: We will develop, implement and test the MEP-based framework and evaluate the plausibility of solutions for key biogeochemical distributions and fluxes under forcing conditions beyond those used for calibration, such as high CO_2, high temperature and/or high nutrients, or hindcast to conditions that existed in the distant past (Norris et al. 2013) for comparison to available paleoceanographic data.

5.1 Entropy production in Darwin Model (Hypothesis I)

To test the hypothesis that natural communities organize to maximize entropy production and that the MEP conjecture is a useful metric to guide model development, we will generate an ensemble of solutions for global ocean biogeochemistry from different initial concentrations and parameterizations of the biological structures, including a random mixing of generalists and specialists. Total internal entropy production for each ensemble member will be calculated and each solution ranked by entropy production. Because we will not be able to span the entire parameter space, we expect that some solutions in the ensemble will produce more entropy than others. For each solution member, we will also calculate a goodness of fit (GOF) metric between model output and oceanographic observations, such as NO_3^- concentration and primary production (see below). If the MEP conjecture is correct, we should find a linear correlation between total entropy production and the GOF metric. On the contrary, if we find that there exists ensemble members with relatively low entropy production, but high GOF metric, then we can conclude that MEP is not a useful tool for describing marine biogeochemistry; that is, **the main hypothesis can be falsified**.

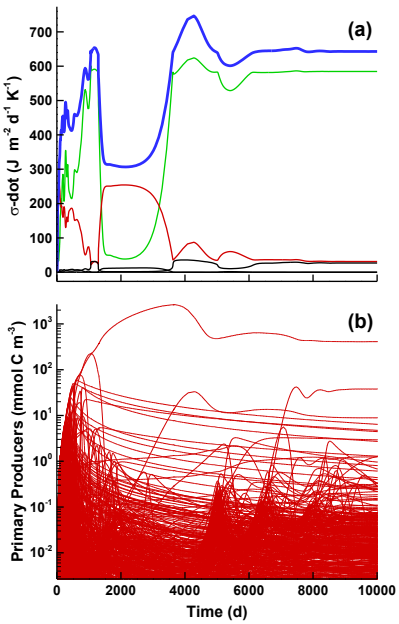

Fig. 4 *(a) Entropy production by primary produces (green), consumers (red), diffusion (black) and system (blue). (b) Population dynamics of 1000 produces in surface layer in Darwin version of two-box model.*

5.2 Generalists versus Specialists (Hypothesis II)

To examine the hypothesis (and its complement) that generalist will perform better in variable environments, we will construct a rate of change (ROC) metric based on the time derivative of several modeled nutrients (see below) averaged over an appropriate spatial scale for each solution in the simulated ensemble. Within each parcel, we will assess the proportion of generalist (i.e., have more than one anabolic or catabolic reaction) to specialists. If our hypothesis is correct, we should find a positive correlation between concentration of generalist relative to specialists and the ROC metric. On a related hypothesis, we also expect to find that in areas with a high ROC metric, a higher number of biological structures with large storage pools for C, N, P and/or Fe will be found compared to areas with low ROC metric (See Section 4.4.2). We

will also explore a corollary that loss of metabolic function, such as N_2 fixation, might be more likely if only specialists exist, since a transient event of sufficient duration can result in local extinction. To explore this idea we will conduct simulations where generalists are not allowed and examine the distribution of metabolic functions compared to the nominal simulations.

We will also explore a third corollary. Organisms that integrate entropy production over longer time scales can exploit resources separated over space more effectively than organisms that maximize entropy production over short time scales (Vallino 2011). For example, diel vertical migration (DVM) allows phytoplankton to access nutrients in deep water at night and energy (light) near the surface during the day (Inoue & Iseri 2012), and zooplankton effectively transport nutrients from deep to surface waters by DVM as well (Steinberg et al. 2002, Haupt et al. 2010). To test this corollary, we will produce an ensemble of solutions where internal storage of C, N, P and Fe is not permitted. Bringing an MEP perspective to the hypothesis of Tozzi et al. (2004), if resources are being collected at one location/time to facilitate energy dissipation in another region/time, we expect entropy production to decrease as well as become more homogenously distributed over the model domain compared to the nominal simulations.

Nutrient cycling is of course very important in the surface ocean because it allows for greater free energy dissipation when resources for catalyst synthesis are limited (consider **Fig. 1**). In oligotrophic areas, such as subtropical gyres, we expect a greater proportion of biomass allocated to structure turnover (rxn. p_1, **Table 1**) because predation can enhance primary productivity under nutrient limitation (Selph et al. 2003, Schmitz et al. 2010, Trommer et al. 2012). We also expect cells with lower C, N, P and Fe pools because smaller cells sink slower in the model formulation. Because of the importance of the marine N cycle (Zehr & Kudela 2010), we will explore solutions regarding the energetics tradeoffs associated with N_2 fixation versus N cycling and compare our MEP results to Darwin model results (Monteiro et al. 2011).

5.3 Robust Extrapolation (Hypothesis III)

If the MEP conjecture proves to be a useful design criterion (Hypothesis I), then we expect models based on MEP should be more robust to predicting biogeochemistry under conditions were calibration data are unavailable because MEP would apply under those conditions as well. However, testing this hypothesis is challenging because the paleoceanographic data are sparse. Consequently, this hypothesis cannot be rigorously tested, but we will explore how the MEP model performs under paleo conditions based on reconstructed expectations (Canfield 2006, Norris et al. 2013), or in the oxygen minimum zones (Canfield et al. 2010, Reed et al. 2014).

5.4 Global maps of metabolic function (Project Output)

With the rapid increase in the use of metagenomics, metatranscriptomics and proteomics (Saito et al. 2014) in oceanographic surveys, there is a growing need to incorporate such observations in global ocean models. Currently, microbial ecologist rely on linear correlation models to extrapolate sparsely observed genomic information to global scale maps based on independent variables, such as season, time of day, latitude, month, temperature and salinity (Ladau et al. 2013). While regression analysis does provide some useful interpretation, regression does not provide any information regarding processes that underpin the relationships, nor are regressions very useful as prognostic tools. Consequently, we will present model outputs from the MEP model in a functional context (**Table 1**) that can be related to metagenomic, metatranscriptomic and proteomic data, i.e. functionally, rather than taxonomically, based mappings. Our model will also have information on how functional genes may be packaged in the context of specialist (single function) versus generalist (multiple functions). Global maps of functional distributions can then be used to assess omics data and provide visualizations and hypotheses which could help shape future omics surveys. In particular, the concentration of biological structure, $\mathbb{S}_i$ allocated to a particular function is proportional to metagenomic data, while reaction rate (governed by ε_i) is proportional to gene expression, or metatranscriptomic/proteomic data.

5.5 Metrics from model solutions.

5.5.1 Goodness of Fit Metric Our GOF metric will be evaluated using a standard root mean squared

(RMS) error to assess both location and magnitude of error between model solution and observations. We will evaluate cumulative RMS model-data differences for key biogeochemical fluxes including global primary production (see **Fig. 1a**,**b** for example comparison) and export production (or the ef-ratio), the pattern of air-sea flux of CO_2 and patterns of nitrogen fixation. In addition, we will quantitatively evaluate the simulations of the global scale distributions of macro-nutrients (specifically inorganic nitrogen species, phosphate, Si), micro-nutrients (here Fe) as well as derived indicators of biogeochemical function, e.g. $P^*=PO_4^{3-} - NO_3^-/16$. We will evaluate the modeled distribution of DOC, DON, DOP and the profile of sinking particle fluxes. We will use numerous data sets to evaluate the RMS error for each global simulation. These include global, seasonal primary productivity derived from remote sensing products using the VGPM (Behrenfeld & Falkowski 1997); global distributions of biogeochemically significant species: nitrate, nitrite, phosphate, dissolved inorganic carbon, dissolved oxygen (GLODAP; Key et al. (2004) and/or World Ocean Atlas, 2009); regional variations in export efficiency, e-f ratio (JGOFS and other process study data; Buesseler and Boyd (2009)).

5.5.2 Rate of Change Metric The ROC metric will be determined at each model grid point based on the time derivative for each modeled nutrient (NH_4^+, NO_3^-, NO_2^-, PO_4^{3-}, etc.). The maximum rate at each grid point will be selcted, then the average determined over an appropriate spacial scale (i.e., patch, ~10^6 km^2). This will give us an index for how quickly resouces are changing in a given area.

6. Broader Impacts

With the projected increases in global temperature, atmospheric CO_2 (and associated decrease in ocean pH), and nitrogen loading to coastal oceans from terrestrial ecosystems, there is a growing societal need to understand how biogeochemical processes will respond to these changes. Results from our project will bring a thermodynamic perspective to marine biogeochemistry models that should result in their improved performance. Our MEP-based model should improve forecast accuracy when extrapolate beyond available calibration data and our thermodynamic extension to the Darwin model should improve understanding and prediction of evolutionary strategies. Global maps of microbial metabolic function that will be produced by the project will allow comparison to the increasingly used omics-based surveys and should provide a mechanistic context to understand those observations.

The MIT ocean model is open source and freely available at *http://www.mitgcm.org*. All algorithm and code developments from this project will be published and freely available. We regularly host training visits and provide user support. Proposed developments will be released as part of the source code. The MIT group regularly participates in outreach events, communicating fundamental understanding and visualizations of marine phytoplankton communities and their role in the global carbon cycle. Continuing activities include (i) a collaboration with Dr. Jen Frazier on an interactive exhibit in the San Francisco Exploratorium about the diversity of marine phytoplankton (Living Liquids). (ii) A collaboration with Dr. Isabelle Klauk and colleagues at the City of Science and Industry's planetarium in Paris, where visualizations of our ocean models are featured in narrated shows discussing the Earth as viewed from space. The Follows group regularly participate in local science events using live microscope projections of plankton tows from the Charles River and local coast to illustrate to the general public the richness of microbial life in natural waters and to discuss their role in the carbon cycle. Forthcoming events include participation in the annual Cambridge Science Festival (April 2016; MIT Museum) and at MIT's Carlson Lecture at the New England Aquarium (October 2015).

We will work with the undergraduate programs at MBL and MIT to train new students at the interface of biogeochemical modeling, molecular microbiology, and microbial biochemistry. We will do this through both undergraduate classroom teaching and research projects and internships. PI Vallino is a faculty member in the Semester in Environmental Science (SES) program at MBL (http://courses.mbl.edu/SES), which annually draws up to 24 juniors and seniors from over 60 colleges and universities around the country and draws many students in underrepresented groups in science (classes average 84% women and several minority colleges and universities participate in the SES program). Students give a public presentation to the Woods Hole community and write "journal ready" manuscripts regarding their

research that are published on the SES web site. In each of the three project years, Vallino will mentor two SES independent student research projects associated with thermodynamic constraints on microbial processes involving laboratory or field work. Follows co-teaches an undergraduate Ecology class (1.018) at MIT. Concepts of thermodynamics, metabolism and marine microbial ecology related to this project are key elements of the class. Results from this research will provide materials with which to discuss if and how ecosystems reflect organization related to overarching thermodynamic and energetic principles. We will link classical ideas from Odum and Lotka to current and emerging perspectives. We will also seek undergraduate participation through the Woods Hole Partnership Educational Program (PEP) (http://www.woodsholediversity.org/pep/). PEP specifically targets underrepresented groups in science and exposes them to multiple disciplines and approaches to research-based science by pairing students with mentors at research labs in Woods Hole. PEP students present their research results to the Woods Hole community at a one day PEP symposium held in mid-August. Project PIs will train summer students in the use of models for understanding microbial processes and ocean biogeochemistry.

To broaden exposure of MEP concepts in marine biogeochemistry and explore its place in the broader context of recent advances in metabolic modeling and theory, we plan to propose a workshop to NSF's Ocean Carbon & Biogeochemistry (OCB) Program to be held in year 2. The workshop, provisionally entitled "Thermodynamic constraints on microbial metabolism and biogeochemical cycles" would bring together experts in MEP along with experts in thermodynamic and redox cascades from cells, to systems biology through to global biogeochemical cycles. We feel the time is right to ask if and how these new, or as yet under-exploited tools, can help advance marine ecology and biogeochemistry in the forthcoming years.

Finally, this project will support one postdoctoral scholar in this new interface between ocean biogeochemistry modeling, thermodynamic modeling and their integration with molecular observations. The postdoctoral scholar will have the opportunity to conduct research at both MBL and MIT.

7. Results from prior NSF grants

Theory: Biological systems organize to maximize entropy production subject to information and biophysicochemical constraints. EF-0928742, 9/2009-8/2013: $750,000. PIs: **Vallino** and Huber. *Intellectual Merit:* This project examined the hypothesis that biological systems evolve and organize in a manner that results in MEP. One of the project's main hypotheses is that living systems differ from abiotic systems, such as fire, by integrating entropy production over time using information stored in the organismal metagenome. A MEP-based model developed during the project has been able to simulate observations using only two adjustable parameters. Model results indicate the communities are inherently well adapted to handling cyclic energy inputs up to periods of at least 20 days. *Broader Impacts*: Experimental and modeling results to date have been presented at 8 international conferences and numerous departmental seminars, five papers have been published (Vallino 2010, Vallino 2011, Algar & Vallino 2014, Vallino et al. 2014, Vallino & Algar 2016), one submitted (Chapman et al. *submitted*) and one soon-to-be submitted (Fernandez Gonzalez et al. *to be submitted*). The project has supported 9 undergraduate research projects and one postdoc.

M. Follows *and C. Hill,* OCE-1029900 DATES, 09/01/2010-08/31/2013, $969,769, **The Biogeography of Primary Producers in the subpolar North Atlantic**. *Intellectual Merit:* Through analysis of the Continuous Plankton Recorder Survey data, idealized modeling and numerical simulations of regional circulation and ecosystem, we sought to understand the organization of diatom and dinoflagellate populations in the North Atlantic. Twelve refereed publications included Ward et al. (2011) examining the costs and benefits of mixotrophy in plankton; Barton et al (2013) a trait-based characterization of diatom and dinoflagellate variations in space and time in the CPR data set, and a pair of manuscripts exploring controls on the size-structure of plankton populations (Ward et al. 2012, Ward et al. 2014). *Broader Impacts:* This project brought significant international links through connections to the Euro-BASIN project: a funded EU program, for example creating connections between the Follows lab and the Center for Ocean Life (DTU, Copenhagen).

References

Alberty, R. A. 2003. *Thermodynamics of biochemical reactions*. Wiley & Sons, Hoboken, NJ.

Algar, C. K., and J. J. Vallino. 2014. Predicting microbial nitrate reduction pathways in coastal sediments. *Aquat.Microb.Ecol.* **71**:223-238, doi:10.3354/ame01678.

Barton, A. D., S. Dutkiewicz, G. Flierl, J. Bragg, and M. J. Follows. 2010. Patterns of Diversity in Marine Phytoplankton. *Science* **327**:1509-1511, doi:10.1126/science.1184961.

Behrenfeld, M. J., and P. G. Falkowski. 1997. Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography* **42**:1-20.

Bruggeman, J., and S. Kooijman. 2007. A biodiversity-inspired approach to aquatic ecosystem modeling. *Limnology and Oceanography* **52**:1533-1544, doi:10.4319/lo.2007.52.4.1533.

Buesseler, K. O., and P. W. Boyd. 2009. Shedding light on processes that control particle export and flux attenuation in the twilight zone of the open ocean. *Limnology and Oceanography* **54**:1210-1232, doi:10.4319/lo.2009.54.4.1210.

Candau, Y. 2003. On the exergy of radiation. *Solar Energy* **75**:241-247, doi:10.1016/j.solener.2003.07.012.

Canfield, D. E. 2006. Models of oxic respiration, denitrification and sulfate reduction in zones of coastal upwelling. *Geochimica et Cosmochimica Acta* **70**:5753-5765, doi:10.1016/j.gca.2006.07.023.

Canfield, D. E., F. J. Stewart, B. Thamdrup, L. De Brabandere, T. Dalsgaard, E. F. DeLong, N. P. Revsbech, and O. Ulloa. 2010. A Cryptic Sulfur Cycle in Oxygen-Minimum Zone Waters off the Chilean Coast. *Science* **330**:1375-1378, doi:10.1126/science.1196889.

Chai, F., R. C. Dugdale, T. H. Peng, F. P. Wilkerson, and R. T. Barber. 2002. One-dimensional ecosystem model of the equatorial Pacific upwelling system. Part I: model development and silicon and nitrogen cycle. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **49**:2713-2745, doi:10.1016/s0967-0645(02)00055-3.

Chapman, E. J., D. L. Childers, and J. J. Vallino. *submitted*. A review of how The Second Law of Thermodynamics has informed ecosystem ecology through its history. *BioScience*.

Chave, J. 2013. The problem of pattern and scale in ecology: what have we learned in 20 years? *Ecology Letters* **16**:4-16, doi:10.1111/ele.12048.

Clausius, R. 1867. *The mechanical theory of heat: with its applications to the steam-engine and to the physical properties of bodies*. J. van Voorst.

d'Ovidio, F., S. De Monte, S. Alvain, Y. Dandonneau, and M. Lévy. 2010. Fluid dynamical niches of phytoplankton types. *Proceedings of the National Academy of Sciences* **107**:18366-18370, doi:10.1073/pnas.1004620107.

Dewar, R. 2003. Information theory explanation of the fluctuation theorem, maximum entropy production and self-organized criticality in non-equilibrium stationary states. *Journal of Physics A: Mathematical and General* **36**:631-641.

Dewar, R. C. 2005. Maximum entropy production and the fluctuation theorem. *Journal of Physics A: Mathematical and General* **38**:L371-L381, doi:10.1088/0305-4470/38/21/L01.

Dewar, R. C., C. H. Lineweaver, R. K. Niven, and K. Regenauer-Lieb, editors. 2014. Beyond the second law: Entropy production and non-equilibrium systems. Springer-Verlag, Berlin, doi:10.1007/978-3-642-40154-1.

Droop, M. R. 1968. Vitamin b12 and marine ecology. 4. Kinetics of uptake growth and inhibition in *monochrysis lutheri*. *Journal of the Marine Biological Association of the United Kingdom* **48**:689-&.

Droop, M. R. 1973. Some thoughts on nutrient limitation in algae. *Journal of Phycology* **9**:264-272.

Dutkiewicz, S., M. J. Follows, and J. G. Bragg. 2009. Modeling the coupling of ocean ecology and biogeochemistry. *Global Biogeochemical Cycles* **23**, doi:10.1029/2008gb003405.

Dutkiewicz, S., J. J. Morris, M. J. Follows, J. Scott, O. Levitan, S. T. Dyhrman, and I. Berman-Frank. 2015. Impact of ocean acidification on the structure of future phytoplankton communities. *Nature Clim. Change* **advance online publication**:1-8, doi:10.1038/nclimate2722.

Dutkiewicz, S., B. A. Ward, F. Monteiro, and M. J. Follows. 2012. Interconnection of nitrogen fixers and iron in the Pacific Ocean: Theory and numerical simulations. *Global Biogeochemical Cycles* **26**:GB1012, doi:10.1029/2011gb004039.

Dvornyk, V., O. Vinogradova, and E. Nevo. 2003. Origin and evolution of circadian clock genes in prokaryotes. *Proceedings of the National Academy of Sciences* **100**:2495-2500, doi:10.1073/pnas.0130099100.

Falkowski, P. G., T. Fenchel, and E. F. DeLong. 2008. The Microbial Engines That Drive Earth's Biogeochemical Cycles. *Science* **320**:1034-1039, doi:10.1126/science.1153213.

Fasham, M. J. R., H. W. Ducklow, and S. M. McKelvie. 1990. A nitrogen-based model of plankton dynamics in the ocean mixed layer. *J.Mar.Res.* **48**:591-639.

Fernandez, A., S. Huang, S. Seston, J. Xing, R. Hickey, C. Criddle, and J. Tiedje. 1999. How Stable Is Stable? Function versus Community Composition. *Applied and Environmental Microbiology* **65**:3697-3704.

Fernandez Gonzalez, N., J. A. Huber, and J. J. Vallino. *to be submitted*. Microbial community dynamics under periodic perturbations. *ISME J*.

Follows, M. J., S. Dutkiewicz, S. Grant, and S. W. Chisholm. 2007. Emergent Biogeography of Microbial Communities in a Model Ocean. *Science* **315**:1843-1846, doi:10.1126/science.1138544.

Follows, M. J., T. Ito, and S. Dutkiewicz. 2006. On the solution of the carbonate chemistry system in ocean biogeochemistry models. *Ocean Modelling* **12**:290-301, doi:10.1016/j.ocemod.2005.05.004.

Geider, R. J., H. L. MacIntyre, and T. M. Kana. 1998. A dynamic regulatory model of phytoplanktonic acclimation to light, nutrients, and temperature. *L&O* **43**:679-694.

Gent, P. R., and J. C. McWilliams. 1990. ISOPYCNAL MIXING IN OCEAN CIRCULATION MODELS. *Journal of Physical Oceanography* **20**:150-155, doi:10.1175/1520-0485(1990)020<0150:imiocm>2.0.co;2.

Haupt, F., M. Stockenreiter, E. S. Reichwaldt, M. Baumgartner, W. Lampert, M. Boersma, and H. Stibor. 2010. Upward phosphorus transport by Daphnia diel vertical migration. *Limnology and Oceanography* **55**:529-534, doi:10.4319/lo.2009.55.2.0529.

Hoehler, T. M. 2004. Biological energy requirements as quantitative boundary conditions for life in the subsurface. *Geobiology* **2**:205-215, doi:10.1111/j.1472-4677.2004.00033.x.

Inoue, T., and Y. Iseri. 2012. Diel vertical migration and nutrient transport of the dinoflagellate Peridinium bipes f. occultatum in a thermally stratified reservoir. *Water Science and Technology* **66**:1212-1219, doi:10.2166/wst.2012.302.

Irwin, A. J., Z. V. Finkel, O. M. E. Schofield, and P. G. Falkowski. 2006. Scaling-up from nutrient physiology to the size-structure of phytoplankton communities. *Journal of Plankton Research* **28**:459-471, doi:10.1093/plankt/fbi148.

Jin, Q., and C. M. Bethke. 2003. A New Rate Law Describing Microbial Respiration. *Applied and Environmental Microbiology* **69**:2340-2348.

Jin, Q. S., E. E. Roden, and J. R. Giska. 2013. Geomicrobial Kinetics: Extrapolating Laboratory Studies to Natural Environments. *Geomicrobiology Journal* **30**:173-185, doi:10.1080/01490451.2011.653084.

Key, R. M., A. Kozyr, C. L. Sabine, K. Lee, R. Wanninkhof, J. L. Bullister, R. A. Feely, F. J. Millero, C. Mordy, and T. H. Peng. 2004. A global ocean carbon climatology: Results from Global Data Analysis Project (GLODAP). *Global Biogeochemical Cycles* **18**:1-23.

Kleidon, A., K. Fraedrich, T. Kunz, and F. Lunkeit. 2003. The atmospheric circulation and states of maximum entropy production. *Geophys.Res.Lett.* **30**:1-4, doi:10.1029/2003GL018363.

Ladau, J., T. J. Sharpton, M. M. Finucane, G. Jospin, S. W. Kembel, J. O'Dwyer, A. F. Koeppel, J. L. Green, and K. S. Pollard. 2013. Global marine bacterial diversity peaks at high latitudes in winter. *ISME J* **21 March**:1-9, doi:10.1038/ismej.2013.37.

Large, W. G., J. C. McWilliams, and S. C. Doney. 1994. Oceanic vertical mixing - a review and a model with a nonlocal boundary-layer parameterization. *Reviews of Geophysics* **32**:363-403, doi:10.1029/94rg01872.

Le Quere, C., S. P. Harrison, I. Colin Prentice, E. T. Buitenhuis, O. Aumont, L. Bopp, H. Claustre, L. Cotrim Da Cunha, R. Geider, X. Giraud, C. Klaas, K. E. Kohfeld, L. Legendre, M. Manizza, T. Platt, R. B. Rivkin, S. Sathyendranath, J. Uitz, A. J. Watson, and D. Wolf-Gladrow. 2005. Ecosystem dynamics based on plankton functional types for global ocean biogeochemistry models. *Global Change Biology* **11**:2016-2040, doi:10.1111/j.1365-2486.2005.1004.x.

Lévy, M., O. Jahn, S. Dutkiewicz, and M. J. Follows. 2014. Phytoplankton diversity and community structure affected by oceanic dispersal and mesoscale turbulence. *Limnology and Oceanography: Fluids and Environments* **4**:67-84, doi:10.1215/21573689-2768549.

Lineweaver, C. H., and C. A. Egan. 2008. Life, gravity and the second law of thermodynamics. *Physics of Life Reviews* **5**:225-242, doi:10.1016/j.plrev.2008.08.002.

Litchman, E., C. A. Klausmeier, O. M. Schofield, and P. G. Falkowski. 2007. The role of functional traits and trade-offs in structuring phytoplankton communities: scaling from cellular to ecosystem level. *Ecology Letters* **10**:1170-1181, doi:10.1111/j.1461-0248.2007.01117.x.

Lorenz, R. 2003. Full Steam Ahead-Probably. *Science* **299**:837-838.

Lorenz, R. D., J. I. Lunine, and P. G. Withers. 2001. Titan, Mars and Earth : Entropy production by latitudinal heat transport. *Geophysical Research Letters* **28**:415-418.

Lotka, A. J. 1922. Contribution to the Energetics of Evolution. *Proceedings of the National Academy of Sciences* **8**:147-151.

Lucia, U. 2012. Maximum or minimum entropy generation for open systems? *Physica A: Statistical Mechanics and its Applications* **391**:3392-3398, doi:10.1016/j.physa.2012.01.055.

Makela, T., and A. Annila. 2010. Natural patterns of energy dispersal. *Physics of Life Reviews* **7**:477-498, doi:10.1016/j.plrev.2010.10.001.

Marshall, J., A. Adcroft, C. Hill, L. Perelman, and C. Heisey. 1997. A finite-volume, incompressible Navier Stokes model for studies of the ocean on parallel computers. *Journal of Geophysical Research-Oceans* **102**:5753-5766, doi:10.1029/96jc02775.

Martyushev, L. 2007. Some interesting consequences of the maximum entropy production principle. *Journal of Experimental and Theoretical Physics* **104**:651-654, doi:10.1134/S1063776107040152.

Martyushev, L. 2013. Entropy and Entropy Production: Old Misconceptions and New Breakthroughs. *Entropy* **15**:1152-1170, doi:10.3390/e15041152.

Monteiro, F. M., S. Dutkiewicz, and M. J. Follows. 2011. Biogeographical controls on the marine nitrogen fixers. *Global Biogeochem.Cycles* **25**:1-8, doi:10.1029/2010GB003902.

Monteiro, F. M., M. J. Follows, and S. Dutkiewicz. 2010. Distribution of diverse nitrogen fixers in the global ocean. *Global Biogeochem.Cycles* **24**:1-16.

Moore, J. K., S. C. Doney, J. A. Kleypas, D. M. Glover, and I. Y. Fung. 2002. An intermediate complexity marine ecosystem model for the global domain. *Deep-Sea Research Part Ii-Topical Studies in Oceanography* **49**:403-462, doi:10.1016/s0967-0645(01)00108-4.

Morrison, P. 1964. A Thermodynamic Characterization of Self-Reproduction. *Reviews of Modern Physics* **36**:517-524.

Niven, R. K. 2009. Steady state of a dissipative flow-controlled system and the maximum entropy production principle. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)* **80**:021113-021115, doi:10.1103/PhysRevE.80.021113.

Norris, R. D., S. K. Turner, P. M. Hull, and A. Ridgwell. 2013. Marine Ecosystem Responses to Cenozoic Global Change. *Science* **341**:492-498, doi:10.1126/science.1240543.

Nowak, M. A. 2006. Five Rules for the Evolution of Cooperation. *Science* **314**:1560-1563, doi:10.1126/science.1133755.
Odum, H. T. 1983. *Systems Ecology*. John Wiley, Toronto.
Odum, H. T., and R. C. Pinkerton. 1955. Time's speed regulator: the optimum efficiency for maximum power output in physical and biological systems. *Am.Sci.* **43**:321-343.
Ottesen, E. A., C. R. Young, S. M. Gifford, J. M. Eppley, R. Marin, S. C. Schuster, C. A. Scholin, and E. F. DeLong. 2014. Multispecies diel transcriptional oscillations in open ocean heterotrophic bacterial assemblages. *Science* **345**:207-212, doi:10.1126/science.1252476.
Ozawa, H., A. Ohmura, R. D. Lorenz, and T. Pujol. 2003. The second law of thermodynamics and the global climate system: A review of the maximum entropy production principle. *Reviews of Geophysics* **41**:1-24, doi:10.1029/2002RG000113.
Paltridge, G. W. 1975. Global dynamics and climate-a system of minimum entropy exchange. *Q.J.Roy.Met.Soc.* **104**:927-945.
Parekh, P., M. J. Follows, and E. A. Boyle. 2005. Decoupling of iron and phosphate in the global ocean. *Global Biogeochemical Cycles* **19**, doi:10.1029/2004gb002280.
Pfeiffer, T., and S. Bonhoeffer. 2002. Evolutionary consequences of tradeoffs between yield and rate of ATP production. *Zeitschrift für Physikalische Chemie* **216**:51-63, doi:10.1524/zpch.2002.216.1.051.
Pianka, E. R. 1970. R-Selection and K-Selection. *American Naturalist* **104**:592-597.
Prowe, A. E. F., M. Pahlow, S. Dutkiewicz, M. Follows, and A. Oschlies. 2012. Top-down control of marine phytoplankton diversity in a global ecosystem model. *Progress in Oceanography* **101**:1-13, doi:10.1016/j.pocean.2011.11.016.
Quijano, J., and H. Lin. 2014. Entropy in the Critical Zone: A Comprehensive Review. *Entropy* **16**:3482-3536, doi:10.3390/e16063482.
Reed, D. C., C. K. Algar, J. A. Huber, and G. J. Dick. 2014. Gene-centric approach to integrating environmental genomics and biogeochemical models. *Proceedings of the National Academy of Sciences* **111**:1879-1884, doi:10.1073/pnas.1313713111.
Riley, G. A. 1946. Factors controlling phytoplankton population on Georges Bank. *J.Mar.Res.* **6**:54-73.
Saito, M. A., M. R. McIlvin, D. M. Moran, T. J. Goepfert, G. R. DiTullio, A. F. Post, and C. H. Lamborg. 2014. Multiple nutrient stresses at intersecting Pacific Ocean biomes detected by protein biomarkers. *Science* **345**:1173-1177, doi:10.1126/science.1256450.
Salmaso, N., L. Naselli-Flores, and J. Padisák. 2015. Functional classifications and their application in phytoplankton ecology. *Freshwater Biology* **60**:603-619, doi:10.1111/fwb.12520.
Schmitz, O. J., D. Hawlena, and G. C. Trussell. 2010. Predator control of ecosystem nutrient dynamics. *Ecology Letters* **13**:1199-1209, doi:10.1111/j.1461-0248.2010.01511.x.
Selph, K. E., M. R. Landry, and E. A. Laws. 2003. Heterotrophic nanoflagellate enhancement of bacterial growth through nutrient remineralization in chemostat culture. *Aquatic Microbial Ecology* **32**:23-37, doi:10.3354/ame032023.
Shimokawa, S., and H. Ozawa. 2007. Thermodynamics of irreversible transitions in the oceanic general circulation. *Geophysical Research Letters* **34**:1-5, doi:10.1029/2007GL030208.
Sogin, M. L., H. G. Morrison, J. A. Huber, D. M. Welch, S. M. Huse, P. R. Neal, J. M. Arrieta, and G. J. Herndl. 2006. Microbial diversity in the deep sea and the underexplored "rare biosphere". *Proceedings of the National Academy of Sciences*:0605127103, doi:10.1073/pnas.0605127103.
Steele, J. H. 1954. Plant production in the northern North Sea. *Mar.Res.Ser.Scot.Home Dep.* **7**:1-36.
Steinberg, D. K., S. A. Goldthwait, and D. A. Hansell. 2002. Zooplankton vertical migration and the active transport of dissolved organic and inorganic nitrogen in the Sargasso Sea. *Deep-Sea Research Part I-Oceanographic Research Papers* **49**:1445-1461, doi:10.1016/s0967-0637(02)00037-7.
Thingstad, T. F., and R. Lignell. 1997. Theoretical models for the control of bacterial growth rate, abundance, diversity and carbon demand. *Aquatic Microbial Ecology* **13**:19-27, doi:10.3354/ame013019.

Tozzi, S., O. Schofield, and P. Falkowski. 2004. Historical climate change and ocean turbulence as selective agents for two key phytoplankton functional groups. *Marine Ecology Progress Series* **274**:123-132, doi:10.3354/meps274123.

Trommer, G., P. Pondaven, M. Siccha, and H. Stibor. 2012. Zooplankton-mediated nutrient limitation patterns in marine phytoplankton: an experimental approach with natural communities. *Marine Ecology Progress Series* **449**:83-94, doi:10.3354/meps09508.

Vallino, J. J. 2000. Improving marine ecosystem models: use of data assimilation and mesocosm experiments. *Journal of Marine Research* **58**:117-164.

Vallino, J. J. 2003. Modeling Microbial Consortiums as Distributed Metabolic Networks. *The Biological Bulletin* **204**:174-179.

Vallino, J. J. 2010. Ecosystem biogeochemistry considered as a distributed metabolic network ordered by maximum entropy production. *Philosophical Transactions of the Royal Society B: Biological Sciences* **365**:1417-1427, doi:10.1098/rstb.2009.0272.

Vallino, J. J. 2011. Differences and implications in biogeochemistry from maximizing entropy production locally versus globally. *Earth Syst.Dynam.* **2**:69-85, doi:10.5194/esd-2-69-2011.

Vallino, J. J., and C. K. Algar. 2016. The Thermodynamics of Marine Biogeochemical Cycles: Lotka Revisited. *Annual Review of Marine Science* **8**:(in press), doi:10.1146/annurev-marine-010814-015843.

Vallino, J. J., C. K. Algar, N. F. González, and J. A. Huber. 2014. Use of Receding Horizon Optimal Control to Solve MaxEP-Based Biogeochemistry Problems. Pages 337-359 *in* R. C. Dewar, C. H. Lineweaver, R. K. Niven, and K. Regenauer-Lieb, editors. Beyond the Second Law - Entropy production and non-equilibrium systems. Springer Berlin Heidelberg, doi:10.1007/978-3-642-40154-1_18.

Wang, J., and R. L. Bras. 2011. A model of evapotranspiration based on the theory of maximum entropy production. *Water Resour.Res.* **47**:1-10, doi:10.1029/2010WR009392.

Ward, B., S. Dutkiewicz, A. D. Barton, and M. J. Follows. 2011. Biophysical Aspects of Resource Acquisition and Competition in Algal Mixotrophs. *American Naturalist* **178**:98-112, doi:10.1086/660284.

Ward, B. A., S. Dutkiewicz, and M. J. Follows. 2014. Modelling spatial and temporal patterns in size-structured marine plankton communities: top–down and bottom–up controls. *Journal of Plankton Research* **36**:31-47, doi:10.1093/plankt/fbt097.

Ward, B. A., S. Dutkiewicz, O. Jahn, and M. J. Follows. 2012. A size-structured food-web model for the global ocean. *Limnology and Oceanography* **57**:1877-1891, doi:10.4319/lo.2012.57.6.1877.

Ward, B. A., M. A. M. Friedrichs, T. R. Anderson, and A. Oschlies. 2010. Parameter optimisation techniques and the problem of underdetermination in marine biogeochemical models. *Journal of Marine Systems* **81**:34-43, doi:10.1016/j.jmarsys.2009.12.005.

Ward, B. A., M. Schartau, A. Oschlies, A. P. Martin, M. J. Follows, and T. R. Anderson. 2013. When is a biogeochemical model too complex? Objective model reduction and selection for North Atlantic time-series sites. *Progress in Oceanography* **116**:49-65, doi:10.1016/j.pocean.2013.06.002.

Wilson, D. S. 1997. Biological communities as functionally organized units. *Ecology* **78**:2018-2024.

Wittebolle, L., H. Vervaeren, W. Verstraete, and N. Boon. 2008. Quantifying Community Dynamics of Nitrifiers in Functionally Stable Reactors. *Applied and Environmental Microbiology* **74**:286-293, doi:10.1128/AEM.01006-07.

Wunsch, C., and P. Heimbach. 2007. Practical global oceanic state estimation. *Physica D-Nonlinear Phenomena* **230**:197-208, doi:10.1016/j.physd.2006.09.040.

Zehr, J. P., and R. M. Kudela. 2010. Nitrogen Cycle of the Open Ocean: From Genes to Ecosystems. *Annual Review of Marine Science* **3**:197-225, doi:10.1146/annurev-marine-120709-142819.