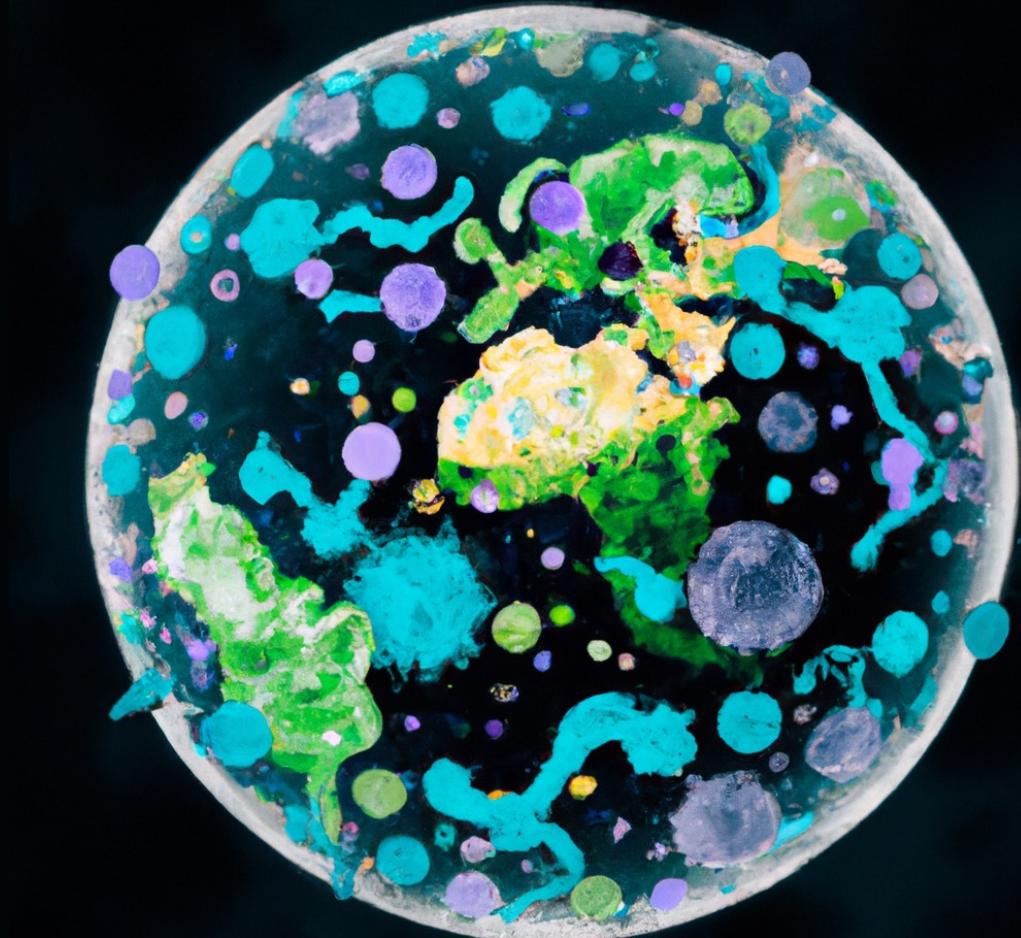


# CARBON ACROSS SPACE & TIME

*Multi-Scale Modeling of Microbial  
Metabolism in the Marine Carbon Cycle*



Helen Scott

Qualifying Exam | December 2022

# Table of Contents

page

**3**

---

*Introduction*

---

page

**29**

---

*Challenge 4: Multi-Scale Modeling*

---

page

**11**

---

*Challenge 1: Diversity of Organisms*

---

page

**34**

---

*Challenge 5: Community Education*

---

page

**16**

---

*Challenge 2: Availability of Well-Curated Genome-Scale Metabolic Models*

---

page

**40**

---

*Project Map & Timeline*

---

page

**22**

---

*Challenge 3: Predicting Carbon Use Efficiency from Flux Balance Analysis Results*

---

page

**44**

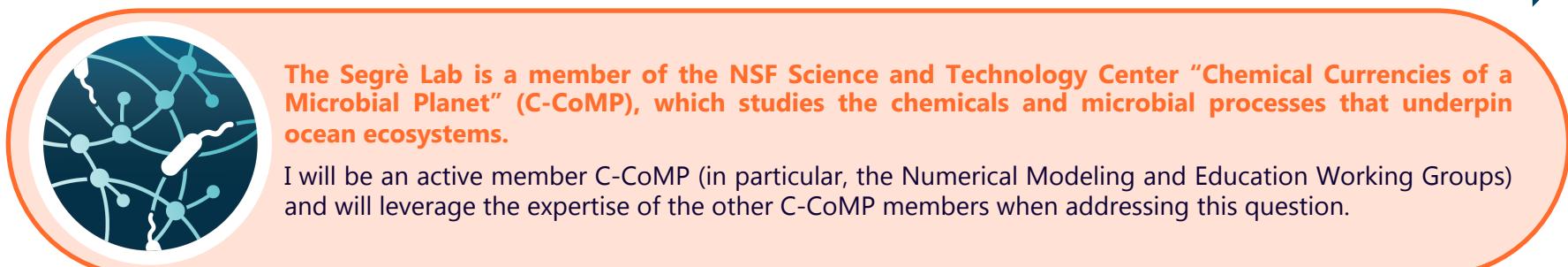
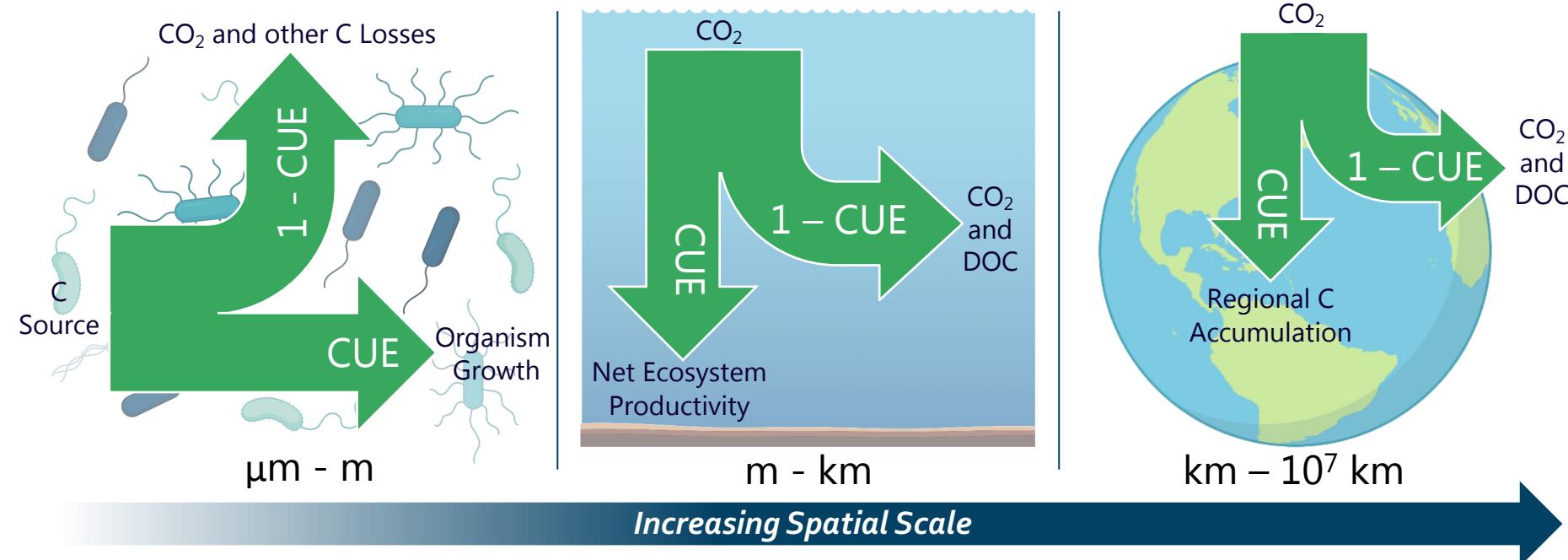
---

*References and Acknowledgments*

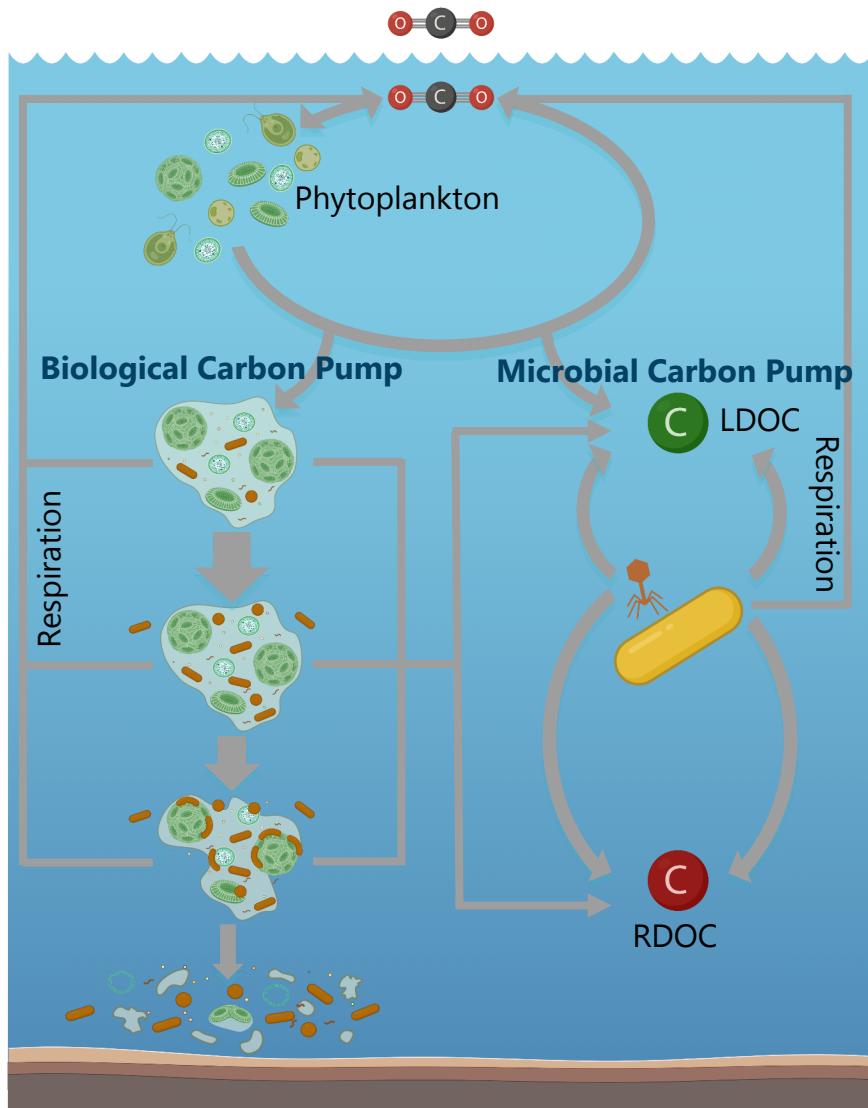
---

# The Big Question: How does Bacterial Carbon Use Efficiency Vary in Space and Time?

*It is important to understand how bacterial carbon use efficiency varies across scales to evaluate the role of heterotrophic bacteria as a photosynthetic carbon sink in the marine ecosystem (Manzoni, 2018).*



# Microbial Metabolism is Critical to the Marine Carbon Cycle



*The global cycling of carbon supports life on Earth and affects the state of the biosphere within which humans reside. The flux and fate of carbon in the ocean is mediated largely by microbial activity (Moran, 2022).*

Phytoplankton are responsible for most of the transfer of carbon dioxide ( $\text{CO}_2$ ) from the atmosphere to the ocean via photosynthesis. Some of these phytoplankton are bacteria, such as *Prochlorococcus*, or are eukaryotic, such as coccolithophores.

After fixation, carbon may take the form of Particulate Organic Carbon (POC) or Dissolved Organic Carbon (DOC). These distinct forms of carbon support different carbon sequestration processes.

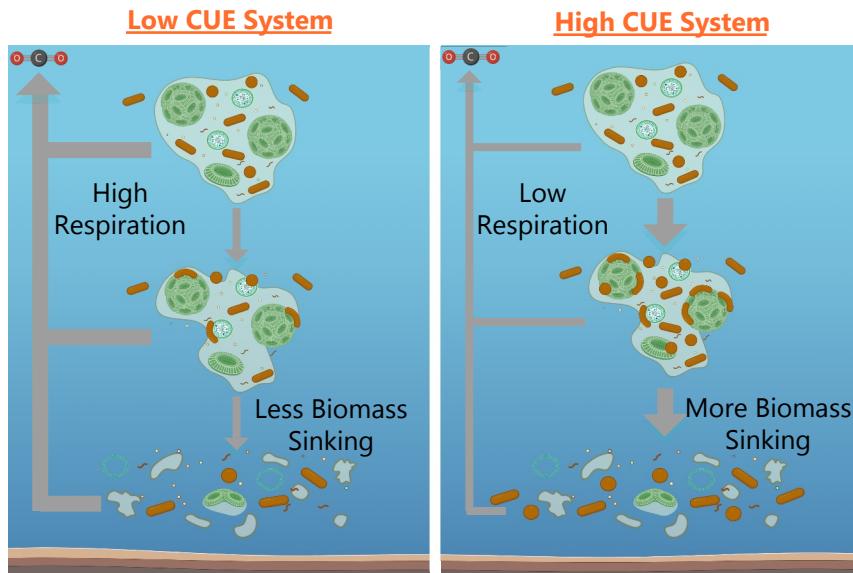
**In the Biological Carbon Pump (BCP),** POC sinks deeper into the ocean, where it can be stored for geologic timescales. However, very little organic carbon survives to be buried in deep-sea sediments. The vast majority of the carbon in POC is taken up by heterotrophic bacteria and is remineralized into  $\text{CO}_2$  or decomposed and released as DOC (Jiao, 2010).

**In the Microbial Carbon Pump (MCP),** heterotrophs transform labile DOC (LDOC) into recalcitrant DOC (RDOC). RDOC is resistant to further biological degradation and is thus maintained in the ocean for decades to millennia (Jiao, 2011).

*Microbial communities regulate the sequestration efficiency of both the BCP and the MCP.*

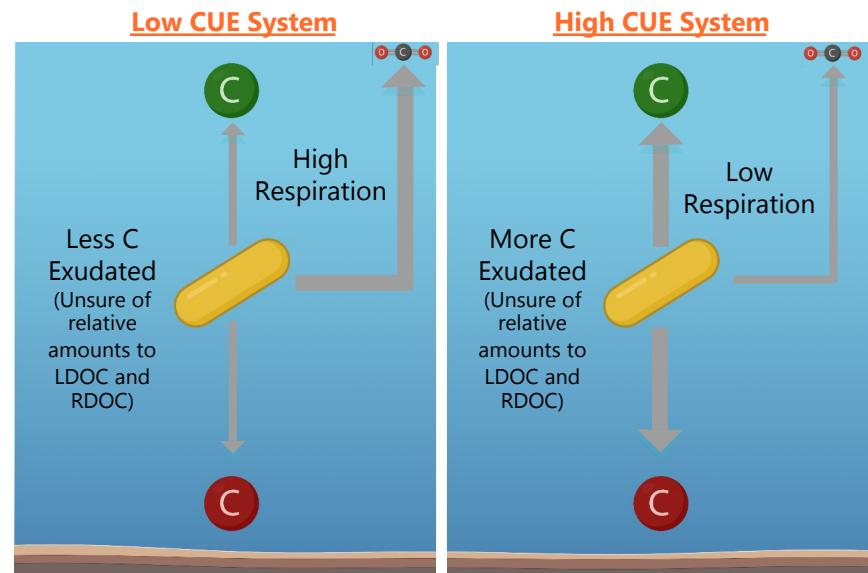
# Carbon Use Efficiency Determines the Fate of Carbon in the Biological and Microbial Carbon Pumps

*Carbon Use Efficiency (CUE) is defined as the ratio of carbon sequestered by a system to carbon that enters that system. In a living system, biological CUE is the fraction of carbon taken up from the environment that is allocated to biosynthesis (Sinsabaugh, 2013).*



In the BCP, CUE conveys how quickly POC is remineralized. The timing of particle degradation is very important as the depth of the CO<sub>2</sub> is directly related to the duration of carbon sequestration in the ocean.

An organism with a high CUE, which respires very little, could potentially allow for more organic carbon (i.e. its own biomass) to sink to deeper depths before being remineralized or recycled into DOC.



In the MCP, CUE conveys how much DOC is being remineralized, but it would not necessarily confirm if the DOC has been transformed into a recalcitrant state.

Because of its ecological importance, CUE is a critical parameter to include in ecosystem models for the better prediction of the carbon flux between the surface ocean and the atmosphere (Moran, 2022).

# Ocean Biogeochemical Models have Rudimentary Treatment of Heterotrophic Bacteria

Ocean biogeochemical models (OBMs) describe the ocean's circulation, physical properties, biogeochemical properties, and their transformations using coupled differential equations (Fennel, 2022).

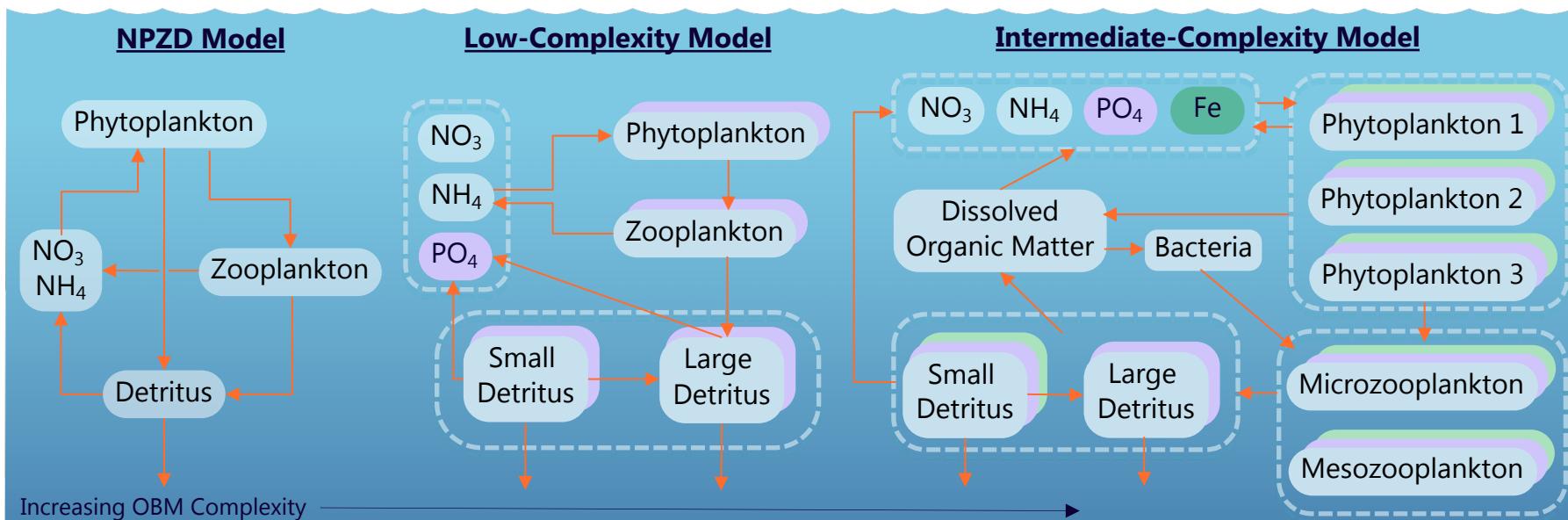
OBMs can be simple or can become more complex by increasing the number of state variables and biogeochemical transformations represented. The figure below shows OBMs increasing in complexity from left to right (adapted from Fennel, 2022).

The simplest, the Nutrient-Phytoplankton-Zooplankton-Detritus (NPZD) model, includes four state variables and one nutrient currency, often nitrogen.

A typical low-complexity model includes several nutrient currencies, which are stored as multiple state variables for each functional group or particulate pool. With greater complexity, there are more granular functional groups and nutrient pools, and new state variables like dissolved organic matter.

While OBMs may include hundreds of phytoplankton groups and multiple size classes of zooplankton, very few models directly account for the activity of heterotrophic bacteria.

For example, of the suite of ocean biogeochemistry models that the IPCC uses for climate modeling, only one includes active heterotrophic bacteria (Lovato, 2022).



# One Ocean Biogeochemical Model that Includes Heterotrophic Bacteria is the BATS-1D-VAR Model

BATS-1D-VAR is a one-dimensional (1D) variational data assimilation model specific to the Bermuda Atlantic Time-series Study site (BATS) which simulates biotic and abiotic processes in a 1D vertical water column (Heather Kim, unpublished).

The BATS-1D-VAR model simulates stocks and flows of C, N, and P through multiple state variables. The schematic to the right shows these state variables and the transformations that convert one state to another (see key below).

## State Variables

- SP: Small Phytoplankton
- TR: Trichodesmium
- UN: Unicellular, N<sub>2</sub> fixing cyanobacteria
- FL BAC: Free-living Bacteria
- PA BAC: Particle-attached bacteria
- LDOM: Labile dissolved organic matter (DOM)
- SDOM: Semi-labile DOM
- RDOM: Refractory DOM
- Higher Level: Higher-level zooplankton

## Transformations

- 1: Primary production
- 2: Respiration
- 3: Viral Mortality
- 4: Grazing
- 5: DOM Uptake
- 6: Excretion
- 7: Nutrient Uptake
- 8: Regeneration
- 9: Particle Uptake
- 10: Particle Sinking Flux

In this OBM, heterotrophic bacteria are represented by the state variables FL BAC and PA BAC. Their activities in the model have been parameterized using BATS data, however, this does not capture the mechanisms behind their activities.

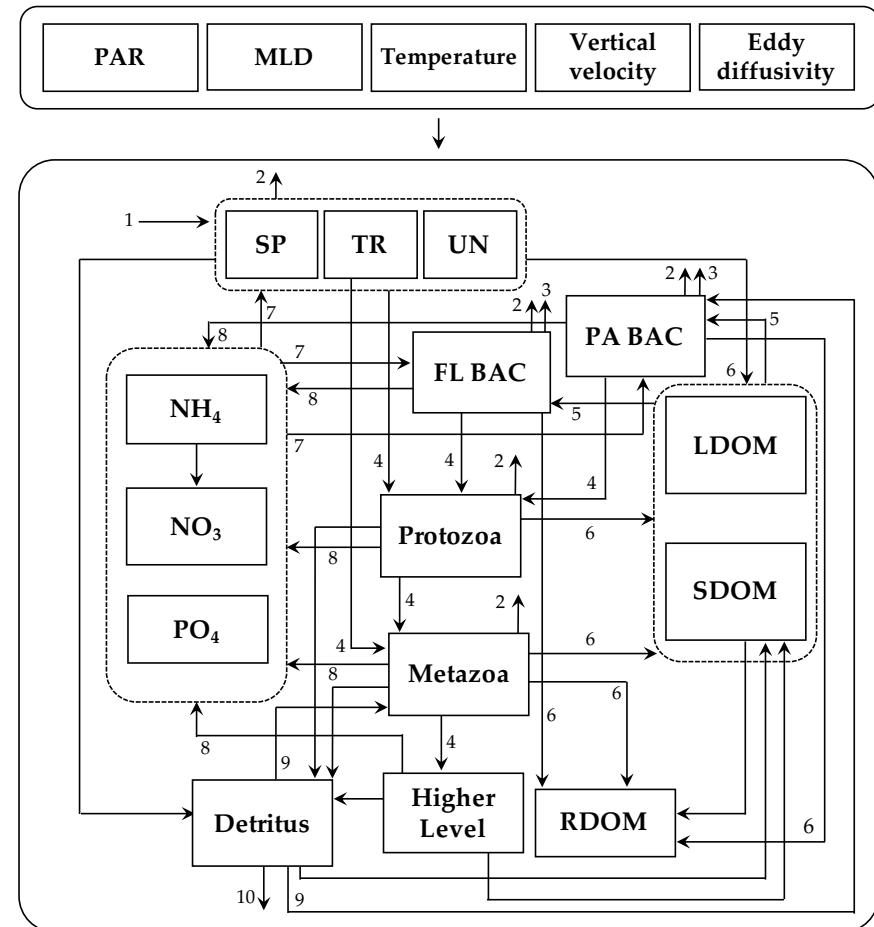


Figure from Heather Kim, unpublished

# My Project Will Address Microbial Roles in Carbon Biogeochemistry Using Multi-Scale Modeling

The successful modeling of relevant microbial metabolisms at regional and global scales will close one of the largest knowledge gaps in the global carbon cycle.

I envision a regional OBM that explicitly includes heterotrophic CUE as a function of microbes' metabolic networks and available nutrient sources.

The challenge for modeling is to bridge all relevant spatial and temporal scales. Spatially, the model must span thirteen orders of magnitude, from cell metabolism at the scale of  $10^{-6}$  m to ocean flux at the scale of  $10^7$  m. Temporally, the model must capture dynamics from a single generation of bacteria to years or even decades.

Component models that focus on specific portions of these scales already exist or are being developed.

*I will focus on developing models at the microbial community scale and on connecting the Flux Balance Analysis modeling framework across scales.*

To fully achieve this vision, there are multiple challenges that need to be addressed at the microbial community scale, in connecting component models, and in the microbiome research community at large.



# Challenges in Achieving this Vision

## Challenge 1: Diversity of Organisms

With the immense diversity of marine microbes, I must first select strains of photo-autotrophs and heterotrophs to use in my studies. Ideally, these model organisms will be culturable in the lab, have sequenced and annotated genomes, have a body of literature around them, and be representative of clades or larger groups of microbes.

## Challenge 5: Community Education

Many microbiome scientists do not use metabolic modeling because of a lack of computational expertise and concern about the accuracy of predictions. There are not many resources available, to teach metabolic modeling to researchers of diverse backgrounds.

## Challenge 2: Availability of Well-Curated GEMs

For metabolic modeling, I will need genome-scale metabolic models (GEMs) for each organism. To become truly representative of the organism of interest, each model will need to be curated and potentially include experimental data.



## Challenge 3: Predicting CUE from FBA Results

Using a consistent methodology to define CUE across a broad range of microbial taxa is necessary to determine how physiological variation in resource use between taxa impacts CUE.

## Challenge 4: Multi-Scale Modeling

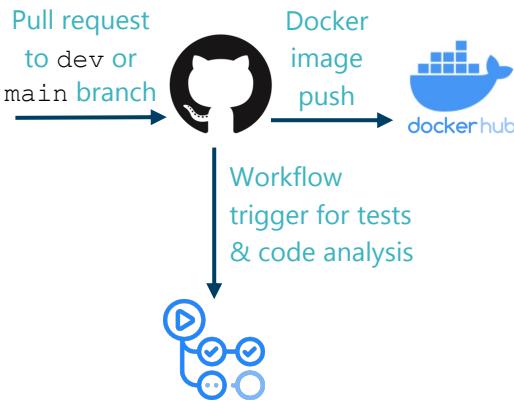
To integrate models at each scale, there needs to be standardization of the inputs and outputs of each so that the "hand-off" between models can occur.

# My Open Science Commitment

## Goal: I want any software I write to be usable after I graduate.

To ensure that all code I write is usable, I will use Continuous Integration and Continuous Testing. This means that all code will be version-controlled and publicly available on GitHub. I will write tests for all my code to ensure that I find bugs sooner and to improve the quality of the code that I publish.

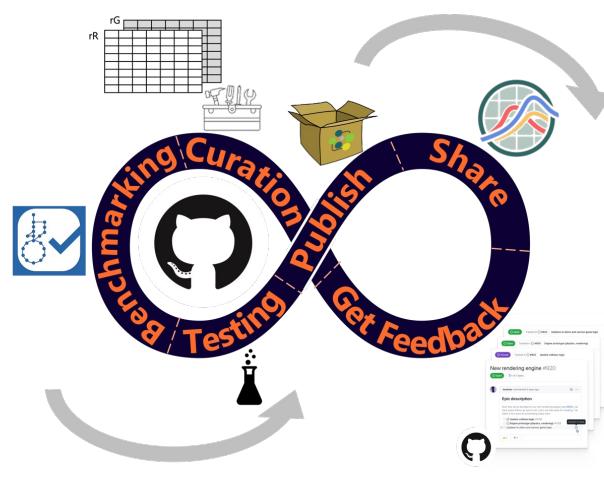
I will package my software with all its dependencies in Docker images. Using Docker will ensure that users do not have to do any further configuration to install or run my tools. I will use GitHub actions to create automated workflows that build a Docker image and run automated tests.



## Goal: I want GEMs I make to be re-used and iterated upon.

The field is moving towards community-driven GEM development on online open-source platforms. To ensure that my work is part of this community effort, I will:

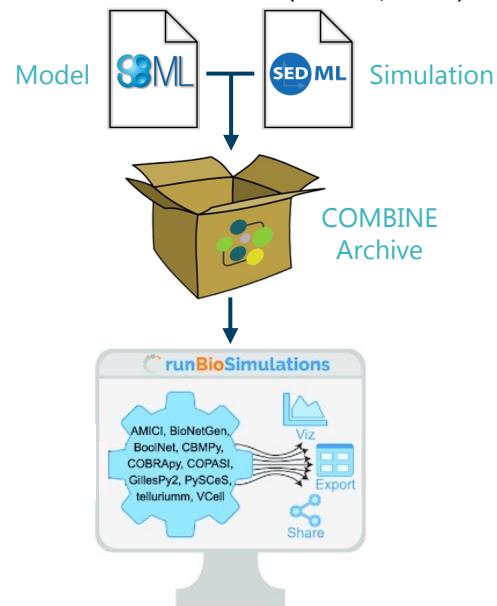
- Use version control to document automated and manual curation.
- Use MEMOTE for standardized model benchmarking and testing (Lieven, 2020).
- Create and share reconstruction metadata in a COMBINE archive (Bergmann, 2014).
- Create recall matrices of removed content (Seif, 2021).
- Share the model on at least one publicly available database (e.g. BioModels).



## Goal: I want simulations I run to be reproducible.

In order to share my results with collaborators and ensure that anyone can reproduce my results easily I will use COMBINE archives. A COMBINE archive is a single file that contains all the information necessary for a modeling and simulation experiment (Bergmann, 2014).

Where possible, I will use RunBioSimulations to execute these COMBINE archives to run and publish entire simulation studies (Shaikh, 2021).



**Challenge 1:**

# Diversity of Organisms

# *Prochlorococcus* and *Alteromonas* form a Model Organism System

To gain insights into marine microbial communities without trying to work with all possible microbes, I will work with a model organism system of a photoautotrophic and a heterotrophic microbe.

The phototrophic partner of my model system will be the *Prochlorococcus* strain MED4 (PRO). The *Prochlorococcus* cyanobacterial group is the numerically dominant phototroph in the oceans and is a well-known model organism in the field of microbial ecology (Biller, 2015b).

PRO co-evolved with heterotrophic bacteria and can truly only be understood within the context of its greater microbial community. PRO supplies the heterotrophs with photosynthetically fixed carbon, and in turn, heterotrophic bacteria both aid and inhibit the growth of PRO (Biller, 2015b).

One such heterotroph is *Alteromonas macleodii* MIT1002 (ALT), a well-studied  $\gamma$ -proteobacterium that was initially isolated from a *Prochlorococcus* culture (Biller, 2015a). In addition to its close relationship to PRO, ALT is a copiotroph which is a desirable trait in a model organism as it means the bacteria grows well under laboratory conditions and is amenable to genetic manipulation.

I will use this model organism system for my initial work and simulations. However, a system of just one heterotroph and one phototroph does not begin to capture the diversity of microbial life in the ocean, and so throughout my project, I will gradually need to expand the scope to work with a larger set of diverse microbes.

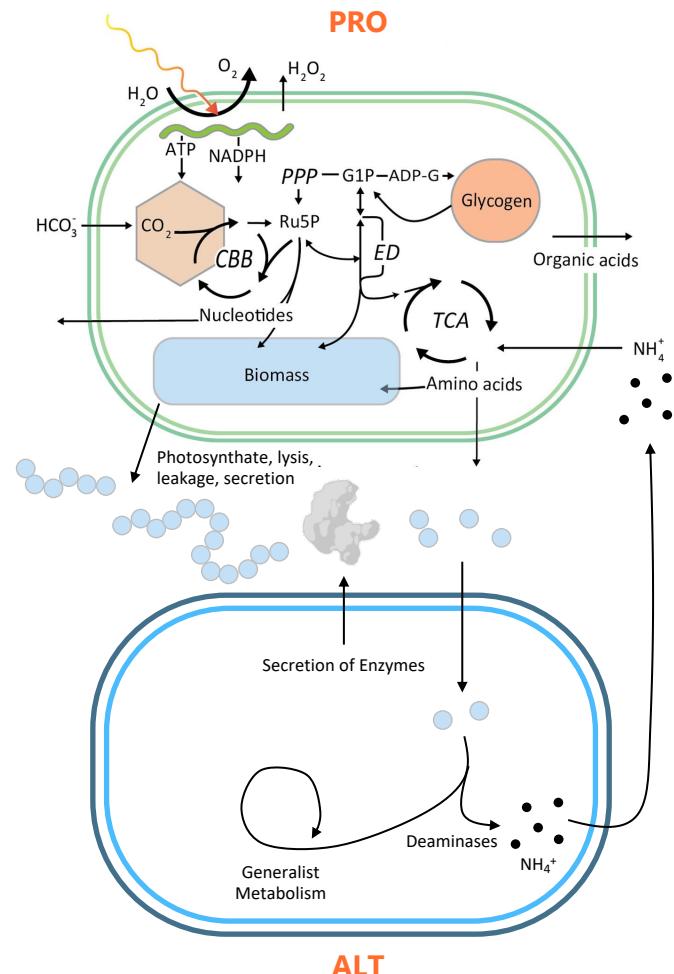


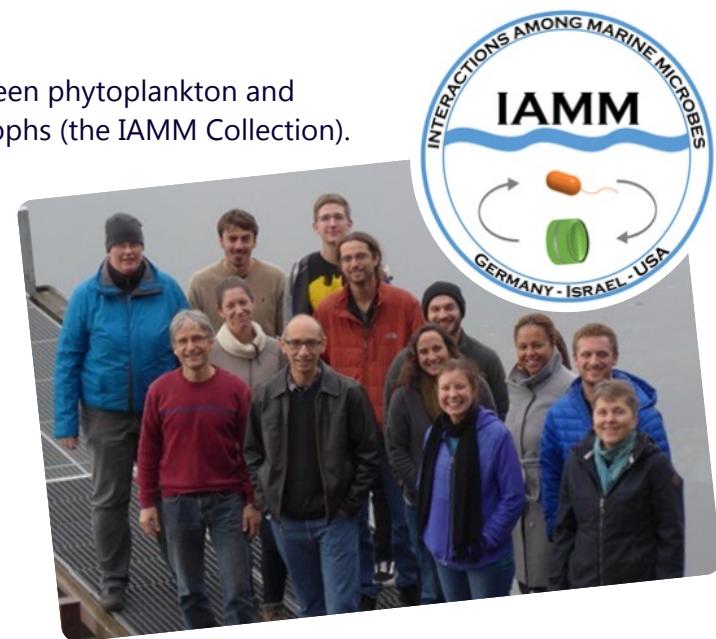
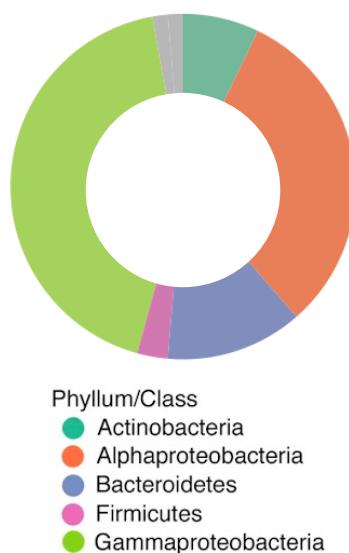
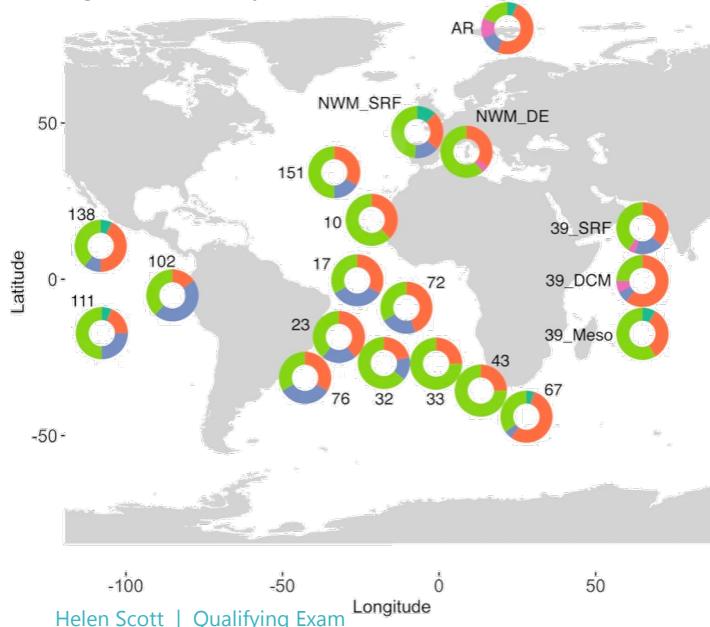
Figure adapted from (Ofaim, 2021)

# The IAMM Collection of Marine Heterotrophs Captures Marine Diversity

The Interactions Among Marine Microbes (IAMM) project studied the coupling between phytoplankton and heterotrophic microbes. In the project, the team created a library of marine heterotrophs (the IAMM Collection).

To form the collection, culturable strains were selected from a larger library of hundreds of assembled genomes (Zoccarato, 2022). The final 65 strains represent a total of 51 marine genome functional clusters (GFCs) and outliers, including human-associated and soil bacteria as controls.

While the GFCs were defined to create groups of bacteria that behave similarly in terms of their interactions with other microbes, the collection also shows a wide diversity representative of the ocean microbiome. Below, a map of pie charts indicating the proportion of isolates retrieved affiliating with the different phyla, or classes in the case of Proteobacteria from different sampling stations is shown along with the phylum/class pie chart for the IAMM collection (Sanz-Sáez, 2020).



## The IAMM Team PIs:

- Daniel Segrè (Boston University)
- Daniel Sher (University of Haifa)
- Hans-Peter Grossart (Leibniz Institute of Freshwater Ecology and Inland Fisheries)
- Maren Voss (Leibniz Institute for Baltic Sea Research)

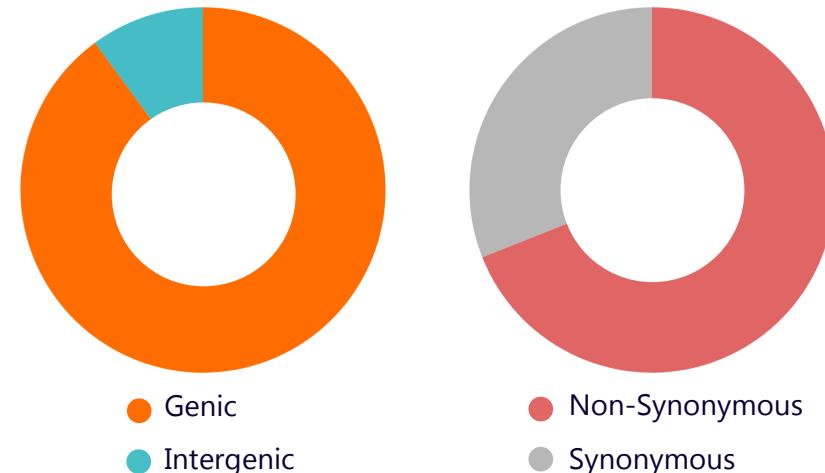
# A Group Project Began Verifying that the IAMM Strains Have not Diverged Due to Laboratory Domestication

The strains of the IAMM collection were assembled in the Segrè lab from sources including collaborator labs and large strain repositories such as the American Type Culture Collection (ATCC). After the collection was assembled in the lab, all the strains were grown overnight, and their genomes were sequenced.

Since being isolated from the environment and the initial publication of the genomes, it is not always known how the strains have been stored and handled. Strains that have been propagated for many generations under laboratory conditions could have developed into a cultivated strain distinct from the wild ancestor. In this case, the large change in conditions between the bacterium's natural environment and the laboratory could have created new selective pressures that caused changes to the bacterium's genome.

*Along with Bioinformatics students Natasha Gurevich and Josseline Velasquez-Reyes, I began characterizing how the genomes of laboratory stocks of the IAMM collection compare to published reference genomes.*

As part of our first-year "Challenge Project", we created a variant-calling pipeline to identify SNPs in the re-sequenced genomes when compared to the previously published references. Across all strains, we found an average of ~24,000 SNPs per strain, most of which were non-synonymous mutations in known gene-coding regions. We also began identifying which genes had the highest number of mutations.



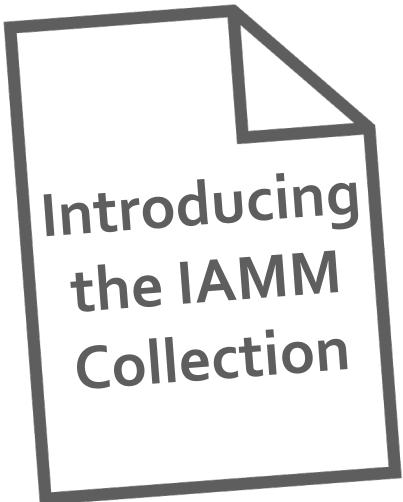
Most Mutated Genes Across 30 Samples

Gene	Number of Samples
rpoB	12
rpoC	7
mexB	6
rcsC_1	3
tycC_2	3

I will continue this work to identify the specific genes and pathways enriched for mutations across the strains.

I hope to use these findings to ensure that future experimental work with the laboratory stocks and models developed from the reference genome do not diverge and produce different results.

# Potential Publications



**Target journal:** TBD

**My role:** Contributor (one of many authors)

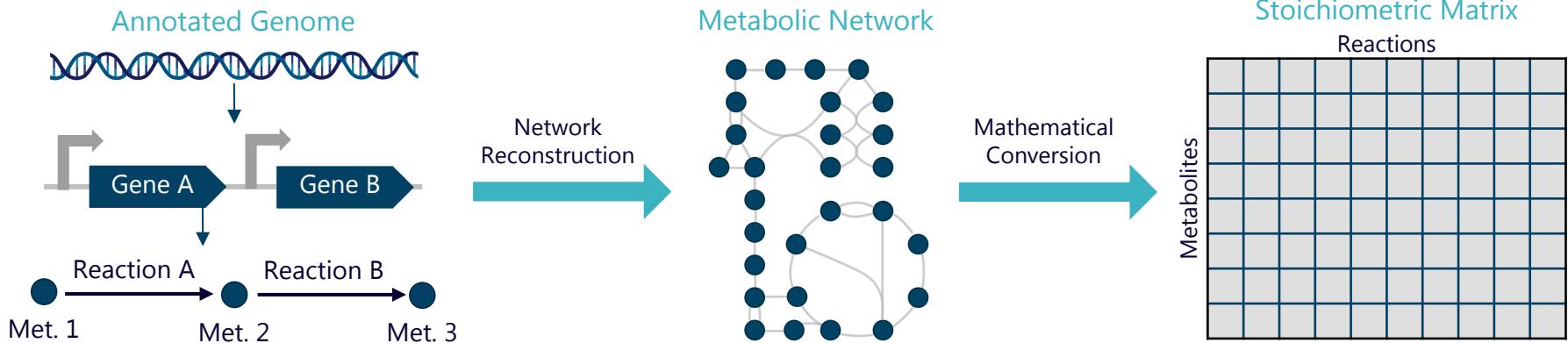
**Target submission date:** TBD by lead author

I will contribute results on the resequencing of the IAMM strains to a larger publication about the collection. My contribution will include a small written section and one figure summarizing the functional impact of mutations found in the resequencing. My goal for the figure is to summarize which genes and which functional pathways have been impacted by SNPs across the collection.

Challenge 2:

# Availability of Well-Curated Genome-Scale Metabolic Models

# GEMs are Key Computational Tools for the Systems-Level Study of Metabolic Networks



Genome-scale metabolic models (GEMs) are organized and systematized knowledge bases that have multiple uses, including conversion into computational models that interpret and predict phenotypic states and the consequences of environmental and genetic perturbations (Fang, 2020).

The development of a GEM requires a curated metabolic knowledge base and an annotated genome sequence of the organism of interest. By mapping the annotated genome sequence to the knowledge base, one can reconstruct a metabolic network composed of all known metabolic reactions. This metabolic network can be converted into a mathematical format, called the stoichiometric matrix (S matrix). In this matrix the columns represent reactions, and rows represent metabolites. Each entry is the corresponding coefficient for that particular metabolite in that reaction.

As the manual reconstruction of a GEM is laborious and time-consuming, many automated network reconstruction tools have been developed to accelerate the reconstruction process, including: ModelSeed (Seaver, 2021), CarveMe (Machado, 2018), and RAVEN (Wang, 2018).

However, questions regarding the quality of the GEMs resulting from these automated workflows have been raised (Faria, 2018; Seif, 2021).

The products of these automated workflows are often called "draft models" and they may result in an inaccurate description of the organism and unreliable predictions.

One way to avoid these pitfalls is by manually refining a draft model.

# Manual Curation is Needed to Refine GEMs

The curation process produces GEMs that are increasingly more organism specific, have larger metabolic coverage, contain increasingly comprehensive manually curated pathways, and show improved predictive power (Seif, 2021).

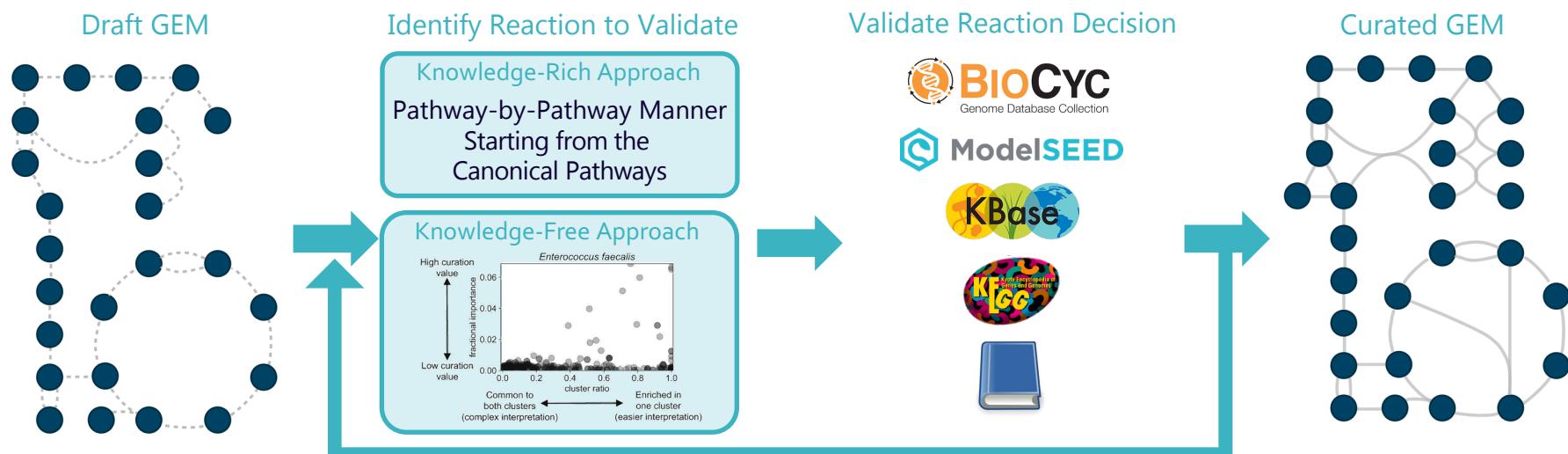
The quality of the starting draft reconstruction heavily affects the amount of time subsequently spent on curating the network, with a strong trade-off between model size, model quality, and human hours spent on model curation.

In the curation process, each gene, reaction, metabolite, and gene-reaction mapping needs to be checked against literature evidence and assigned confidence scores.

Traditionally, the reactions/metabolites/genes are curated in a knowledge-rich manner by going pathway by pathway, starting

from the canonical pathways. This approach has the advantage that reactions are evaluated within their metabolic context and gaps in pathways are easily identified (Seif, 2021). More recently, machine learning has been used to prioritize curation efforts in “knowledge-free” approaches. Tools such as AMMEDEUS use ensemble modeling to generate curation metrics to rank which parameters (i.e., presence/absence of a reaction) have the largest effect on the uncertainty of simulations (Medlock, 2020).

Once a reaction is chosen, there are multiple approaches to validate its presence or absence, including wet lab experiments, targeted bioinformatic analyses, and database or literature searches. One helpful tool for this process is COBRAMod, which integrates data from various metabolic pathway databases as well as user-curated information (Camborda, 2022).



# Status of GEMs for Model Organism Systems

The Segrè and Sher labs have previously collaborated on curating a GEM for *Prochlorococcus MED4* (Ofaim, 2021).

The latest version of the model from this collaboration is called iSO595. iSO595 was an advancement from previous PRO GEMs because it features a complete Entner–Doudoroff pathway, rewired glycogen metabolism and increased coverage of the genome. The figure below shows the most relevant metabolic features of the model.

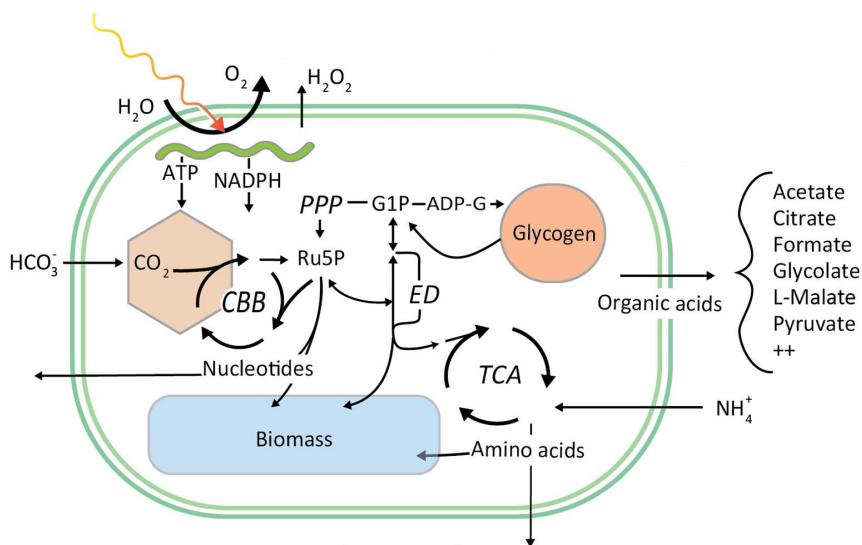


Figure adapted from (Ofaim, 2021)

I will use iSO595 and improve upon it as needed, such as by changing the nomenclature used for the metabolites to make the model more compatible with GEMs generated with the ModelSEED database.

I have generated a draft model of ALT and have begun the manual curation process.

As a starting point for curation, I have generated a draft model using CarveMe (Machado, 2018). CarveMe is an automated tool that aims to increase the quality of draft models by taking a "top-down" approach. In this "top-down" system, CarveMe uses a curated universal model that includes all the reactions and metabolites in the BiGG database. CarveMe takes this universal model and makes an organism-specific model using the organism's annotated genome through a process called "carving".

From CarveMe, I have generated a GEM for ALT that includes import/export reactions, a generic gram-negative biomass reaction, and contains no block or unbalanced reactions. However, this is still a draft model, and will require additional manual curation, especially to capture secondary metabolism.

I have used this draft model as the starting point for iterative and continuous curation, as described in "My Open Science Commitment".

I have generated draft models for the IAMM Collection strains, but manual curation for all strains is not feasible.

I have generated draft models for the IAMM collection strains using CarveMe, but I do not plan to manually curate all strains. In order to curate GEMs for the entire IAMM collection, an entirely different process will need to be developed that decreases the amount of manual curation needed.

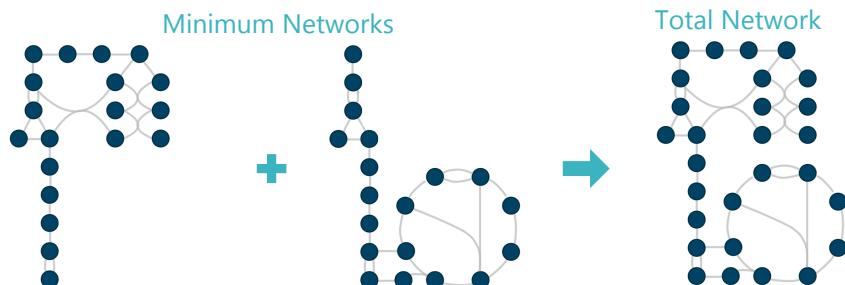
# Development of Novel Network Reconstruction Methods

*Because I am interested in a large number of diverse strains, one potential avenue for future work is to develop methods of network reconstruction that will allow for faster and less manual curation.*

## Idea #1: Mass Phenotyping

Experimental data is often used in network reconstruction and model refinement. One class of data often used is phenotypic data, such as growth on various carbon sources. High-throughput systems can evaluate a microbe's growth under many conditions at once, such as on individual carbon or nitrogen sources.

Where this data is often incorporated into a network reconstruction during manual curation, it could also, in principle, be used to construct the network *ab initio*. In this new framework, a universal metabolic model would be used to identify the minimum metabolic network to grow on each nutrient source. The resulting union of all minimal networks would represent the total metabolic network.

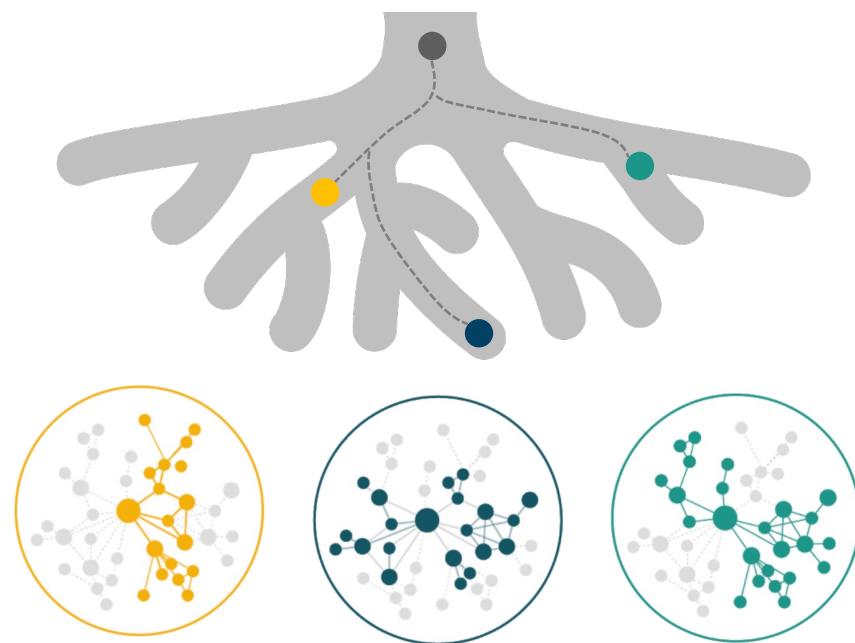


Growth data could also be coupled with high-performance chromatography to determine which metabolites are excreted on a specific nutrient source. This approach may allow us to circumvent issues with poorly annotated genomes and with gap-filling.

## Idea #2: Evolutionary Algorithm

One other new paradigm for constructing metabolic networks could leverage the most curated model, *E. coli*, to deduce any other organism's network.

In this system, it is assumed that the metabolic networks for all organisms exist in a continuous space determined by the course of evolution. So, in theory, by taking small steps through evolutionary space, any organism will be related to all others.



# Potential Publications



**Target journal:** TBD

**My role:** First Author

**Target submission date:** Summer 2023

This publication will serve as the formal release of the manually curated ALT model and will discuss the features of the curated ALT metabolic network and defend the model's accuracy by comparing *in silico* results and experimental data reported in the literature.

Because of the close association between ALT and PRO, we will also demonstrate the value of this model as a tool to probe the interspecies interactions mediated by metabolites. To this aim, I will extend currently existing COMETS simulations of day-night cycles of PRO growth to include ALT.



**Target journal:** TBD

**My role:** First Author

**Target submission date:** Summer/Fall 2023

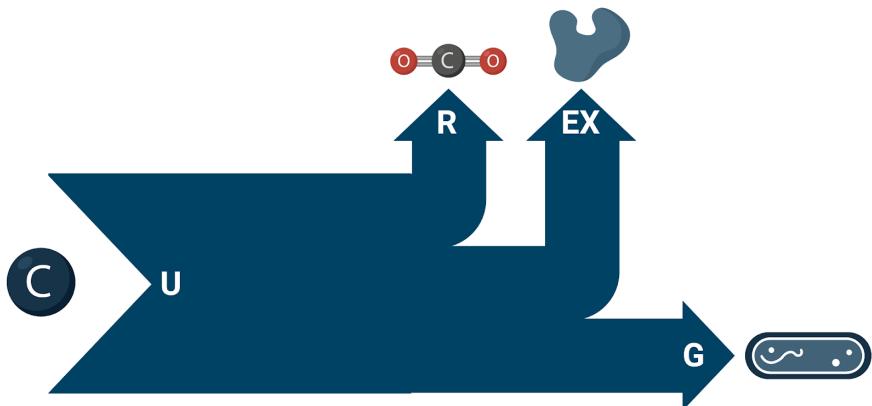
Current literature on GEM curation often discusses the importance of iterative development and community engagement in the curation process. However, these publications have not put forward a formal framework for those practices, leaving groups to develop their own systems in an *ad hoc* manner. In this review, I will detail a protocol for the manual curation of a model that is analogous to continuous development in computer science, which I call continuous curation. Continuous curation emphasizes automated testing, and editing only small sections of the model at a time.

My ideal goal would be to create a template GitHub repository to include in the publication for others to use. This template would include all the automated testing steps as workflows run by GitHub actions. By using this template, scientists could create a shared repository for their model without needing to set up the automated benchmarking and tests recommended by best-practices.

Challenge 3:

# Predicting Carbon Use Efficiency from Flux Balance Analysis Results

# Defining the CUE



For biological systems (from cells to communities), CUE is defined as the ratio between the amount of carbon allocated to biosynthesis and the amount of carbon taken up (Manzoni, 2018).

To calculate this ratio, we define four fluxes of carbon. Carbon uptake (U) is the flux of all organic carbon taken up by the cell, which can be used for growth or energy generation. Two fluxes represent secretion- R is the flux of  $\text{CO}_2$  being released from the cell through respiration, and EX is the exudation rate of organic carbon. Finally, the growth flux, G, is the flux of carbon used to generate biomass.

Despite the deceptively simple definition, different definitions of CUE have proliferated across many disciplines in biology, ecology, and Earth sciences. The main difficulty is to unambiguously define what represents growth and release of extracellular compounds or carbon storage. In addition to multiple definitions, the concept of CUE has many other names, including yield, apparent or growth yield, or growth efficiency (Manzoni, 2018).

The term CUE often arises in the terrestrial biogeochemical literature, where researchers are primarily concerned about the rate of  $\text{CO}_2$  lost from the ecosystem through microbial respiration.

I will use the two following definitions in my work (Manzoni, 2018):

$$\text{CUE} = 1 - \frac{R}{U}$$

This definition of CUE is consistent with previous work on plant carbon budgets. It considers respiration (R) as the only pathway for carbon to leave the system. All organic carbon, regardless if it is exuded from the organism, is still considered to have been used to promote growth and is still a part of the system.

However, in the current literature on soil microbes, the definition for CUE differs. There, typically only biomass synthesis is considered to be carbon that remained in the system, and thus any organic carbon that has been exuded has been lost. This definition of CUE is equivalent to gross growth efficiency (GGE). To avoid confusion I will use the term GGE for this definition.

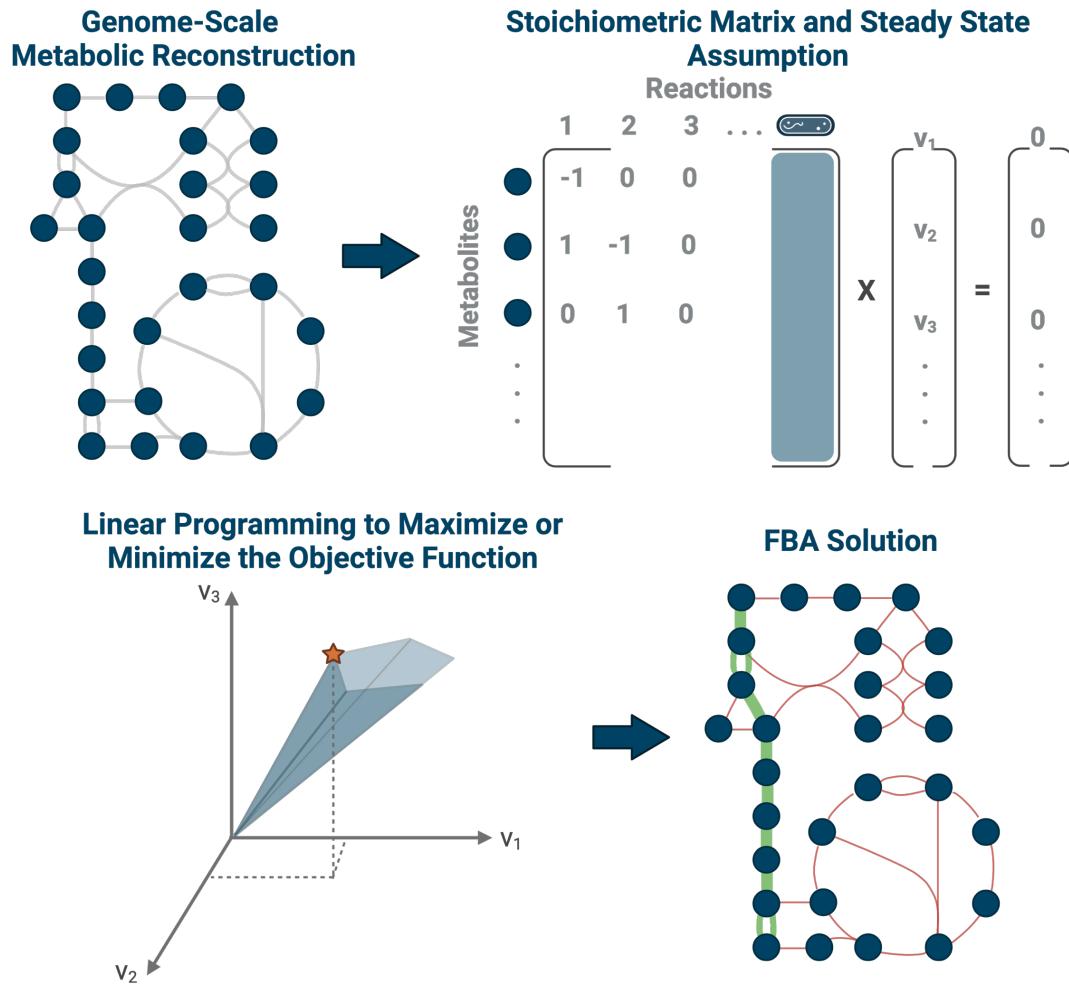
$$\text{GGE} = \frac{G}{U}$$

A value of CUE is specific to the conditions the organism is experiencing. Variables such as growth rate, or the synthesis and release of products, and the availability of carbon sources all vary in time, so CUE is also expected to change. To calculate CUE dynamically, we will track changes in biomass and carbon substrate over time.

# Flux Balance Analysis is a Mathematical Approach for Predicting the Flow of Metabolites through a Metabolic Network

Once curated, a GEM can be used to perform flux balance analysis (FBA). To perform FBA, the reactions in the GEM must be stoichiometrically balanced and then can be converted into a mathematical model by forming a matrix (Orth, 2010). The rows of the matrix represent the metabolites, and the columns reactions, the entry in the matrix is the stoichiometric coefficient for that metabolite in that reaction. The stoichiometric matrix will be sparse, as many entries will be zero. Biomass production is represented by a pseudoreaction in the stoichiometric matrix, called the biomass reaction (shaded blue column in the figure to the right).

The steady-state assumption is used to define a system of linear equations. In complex systems, there are many possible solutions that will fit this constraint. Ideally, a single solution can be found using an objective function. Most commonly, the objective function is to maximize the biomass reaction. With the linear objective function defined, linear programming can then be used to identify the optimal flux distribution (Orth, 2010).



# Calculating CUE from Reaction Fluxes

The four fluxes needed to calculate CUE (organic carbon uptake, respiration, organic carbon exudation, and biomass production) are all easily parsed from the reaction fluxes resulting from FBA.

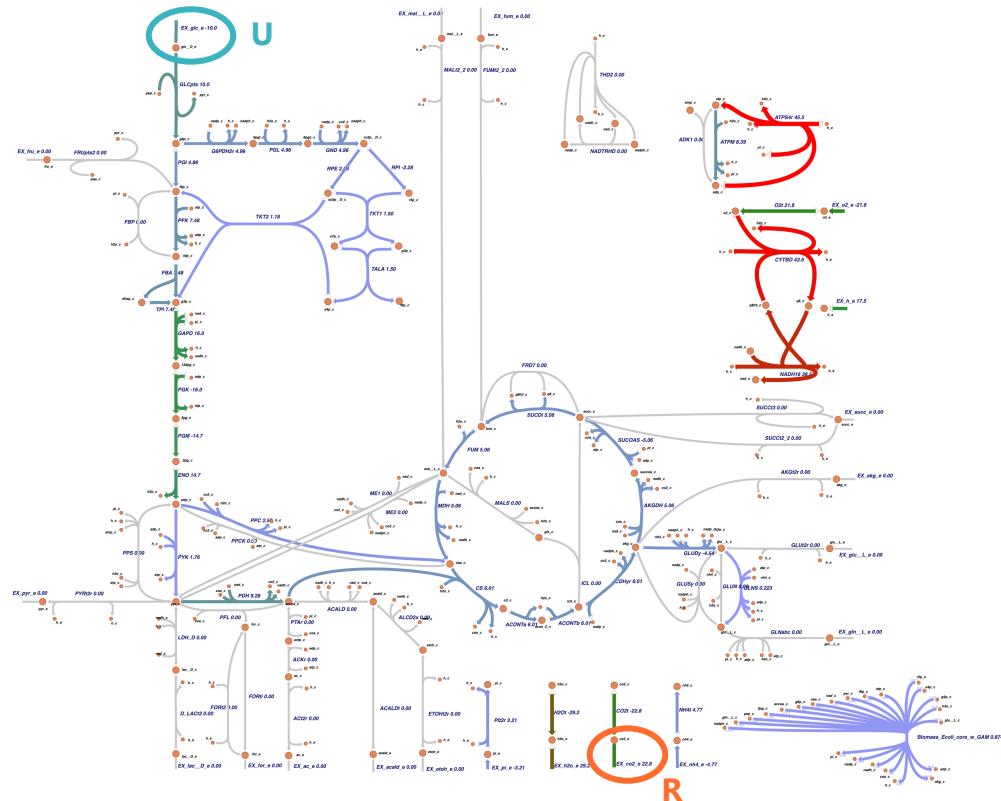
(Saifuddin, 2019) used FBA to estimate the CUE for hundreds of soil bacteria. A simplified version of this process is explained below.

A drawing of the metabolic network for the *E. coli* core model (a small-scale model that can be used for educational purposes) is shown to the right. The colors and widths of the edges of the network indicate the flux of that reaction (with gray being zero, and dark red being the highest).

On this network, the uptake flux (reaction EX\_glc\_e) is circled in blue, and the respiration flux (reaction EX\_co2\_e) is circled in orange. The fluxes for these two reactions are then used to calculate CUE, as shown below the figure.

Programmatically, this is done in python by assembling a list of possible carbon exchange reactions for the model and their elemental stoichiometries. The results of FBA are then searched for each exchange reaction which can be totaled to calculate the total uptake and total loss of carbon.

Because this code functions only on the model and the reaction fluxes, it is compatible with any FBA tool, such as COBRA or COMETS.



$$\text{CUE} = 1 - \frac{R}{U}$$

$$= 1 - \frac{22.81}{(10 \times 6)}$$

$$= 0.62$$

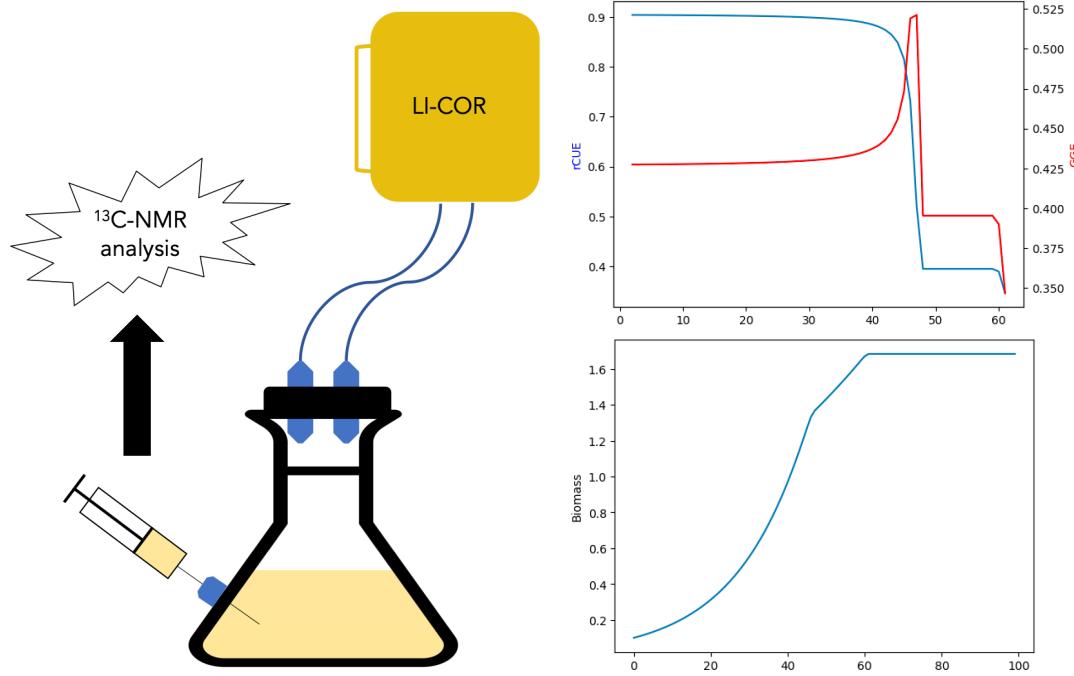
# Comparisons Between Predicted and Experimentally Measured CUE

I will work with Dr. Zac Cooper (Moran Lab, University of Georgia) to compare CUE predictions from FBA to experimental observations for *Alteromonas macleodii* and *Rugeria pomeroyi*.

The Moran Lab (UGA) has planned experiments to measure CUE along a time course for these two heterotrophs. In these experiments, axenic cultures of the heterotrophs will be grown on a known amount of <sup>13</sup>C labeled substrate (acetate or glucose). A gas flux analyzer (LI-COR) will be used to continuously measure CO<sub>2</sub> production, and the liquid media will be sampled periodically to quantify the substrate left in the media using NMR spectroscopy. These experiments are planned for winter 2022-2023.

I will then compare the experimentally calculated CUE with the values predicted from FBA. Because FBA can offer snapshots of the CUE at a finer resolution than the experiments can, we expect that the two values may differ.

For example, FBA may be able to capture the changes in CUE along a diauxic shift as an organism depletes glucose and begins to grow on previously secreted acetate. See plots to the right for predictions of CUE along this diauxic shift for the *E. coli* core model (a small-scale model that can be used for educational purposes).



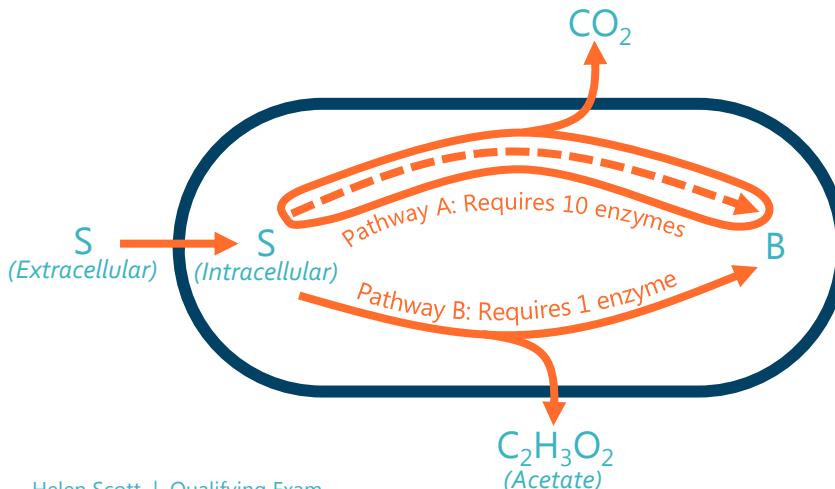
While this initial work will focus entirely on glucose and acetate as the sole carbon source, we would expect CUE to be different on ecologically relevant carbon sources, such as phytoplankton exudates or biomass components. Future work could focus on predicting/measuring the CUE of these strains on a broader collection of carbon sources, including peptides, amino acids, lipids, and oligo-saccharides (Forchielli, 2022).

# Protein Allocation May Cause Real CUE to Differ from Predictions based on FBA

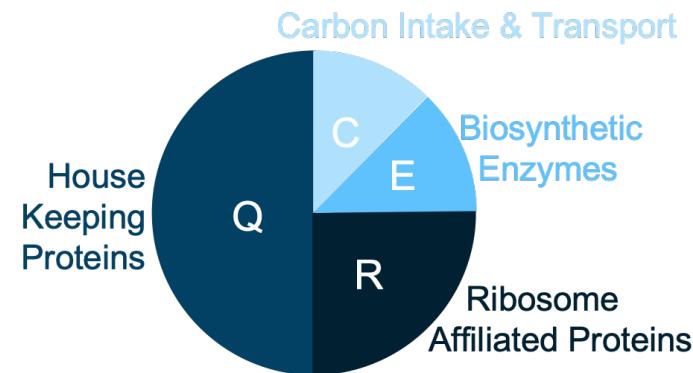
Bacterial "growth laws" directly relate the protein, DNA, and RNA content of a cell to the growth rate. However, these laws are not directly reflected in constraint-based models like FBA.

In order to capture these growth laws, one needs to go beyond the basic elements of FBA and incorporate the costs associated with gene expression and protein synthesis into models of cellular metabolism.

By incorporating these costs, predicted reaction fluxes and therefore the CUE may change. You can imagine a system where there are two pathways to synthesize some biomass precursor (B) from the substrate (S). One pathway may cause the loss of one carbon atom but requires many enzymatic steps (and therefore high protein investment). The other pathway may release more carbon and be more inefficient, but if it requires substantially less protein investment it may still be favored by the cell.



To account for these effects of biomass synthesis on CUE, I will use Constrained Allocation Flux Balance Analysis (caFBA).



caFBA adds parameters to FBA that represent the tug-of-war in the allocation of cellular resources across ribosomal, transport, and biosynthetic proteins that has been observed in experiments. By imposing that the ribosomal share of the proteome behaves in accordance with empirically established growth laws, caFBA is able to better predict cell behavior than typical FBA.

caFBA is typically parameterized based on established growth laws from *E. coli*. We will collaborate with Dr. Mak Saito (Woods Hole Oceanographic Institute) to obtain proteomic data for our strains of interest. This data will then be used to parameterize caFBA for the relative adjustment of proteome sectors at different growth rates.

# Potential Publications



**Target journal:** TBD

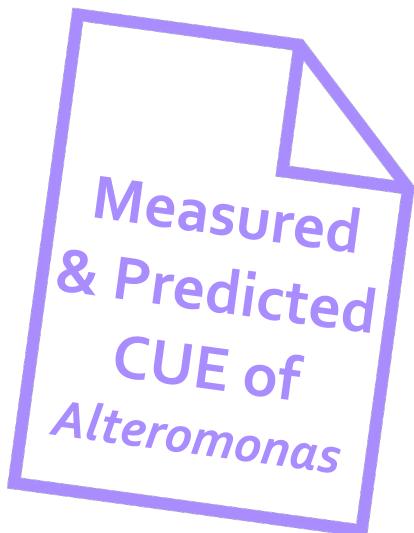
**My role:** First Author

**Target submission date:** Spring 2023

My main publication for this goal will be a simulation-based paper exploring the CUE of toy communities of *E. coli*. By focusing on a model organism, I will bypass the need to manually curate a GEM prior to simulation.

To my knowledge, this would be the first spatially explicit simulation of carbon use efficiency for a microbial community, and I believe that this could be a high-impact publication.

I will aim to address this paper to a broader scientific audience by discussing the importance of CUE to microbial ecology, rather than only focusing on the biogeochemical aspects of ecosystem C sequestration.



**Target journal:** TBD

**My role:** Co-Author

**Target submission date:** Summer 2023

This publication will be a collaboration with the Moran lab and other members of C-CoMP, notably Dr. Zac Cooper, who will be collecting measurements of CUE for ALT in the lab.

In this publication, I will use the same simulation techniques used with *E. coli* in the previous publication and compare the results to what was measured experimentally.

This comparison between FBA-based simulation and experimental measurement will be novel and will be of interest to a scientific audience beyond just those interested in this ALT strain. Because dFBA can simulate fine-grained levels of detail, we hope to be able to define rules of how community CUE emerges from community dynamics.

**Challenge 4:**

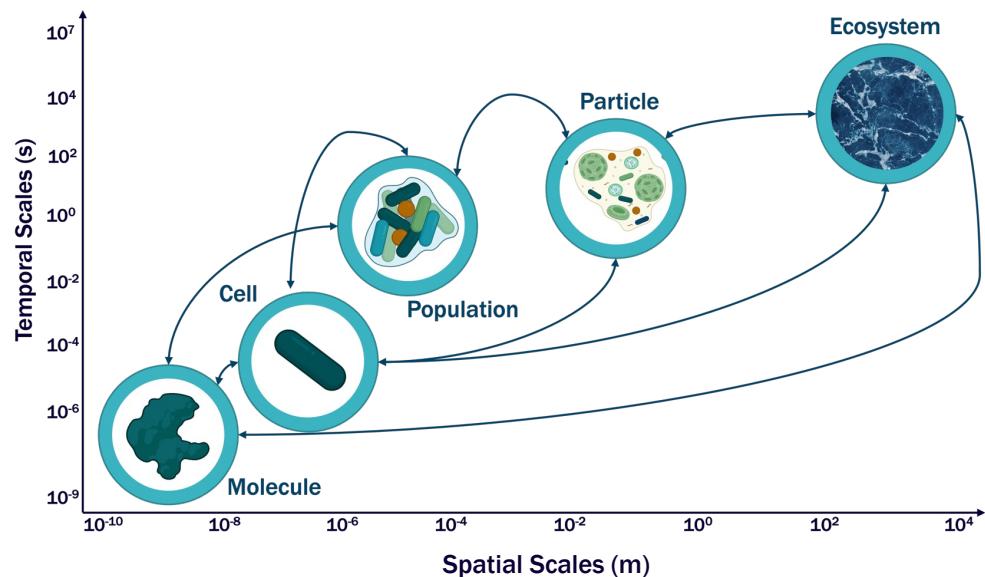
# **Multi-Scale Modeling**

# FBA can be Integrated with Intracellular Mechanistic Models and Environmental Mechanistic Models

FBA models steady-state metabolism for a single organism, averaged over many cells. Because of the steady-state assumption in FBA, the flux predictions are instantaneous values that may not be representative of a cell several hours, or even a few minutes, later in time. GEMs used in FBA represent the average cell's intracellular reactions and interactions with its immediate surroundings (i.e. transport reactions) and so are limited to representing small regions of space at approximately the size of a single cell ( $10^{-6}$  m).

However, when modeling the marine carbon cycle, many other spatial and temporal scales are relevant. For example, genetic regulation, through processes like quorum sensing, can have a large effect on microbial phenotypes, however they are not captured by FBA at the metabolic network scale. And on the other side of the spectrum, OBMs may include biotic processes at higher trophic levels (i.e. grazing) or abiotic factors that occur only at larger spatial and temporal scales.

*In order to model other spatial and temporal scales, FBA can be integrated with models that constrain the flux solution or that use the flux solution as an input.*



**By using another model to impose flux constraints on the FBA problem, I can incorporate more biological information into FBA.**

This may include a chemical reaction network (CRN) model of transcription, translation, regulation, and the enzymatic activity of genes of interest.

**The results of FBA can also be used within greater environmental models that include biomass and extracellular metabolite concentrations.**

This may include a micro or macro-scale environment, including other organisms.

# Using Vivarium as a Platform to Interface FBA with Other Models

Vivarium is a Pythonic software for building integrative multiscale models. It provides an interface that makes individual models into modules that can be wired together in large composite models (Agmon, 2022).

**Vivarium's basic elements are processes and stores (Agmon, 2022).**

**Processes** can implement any kind of dynamical model, such as FBA, or systems of ODEs.

**Stores** are databases for the state variables. Processes can read the stores, and then the stores can apply each Processes' updates to its value.

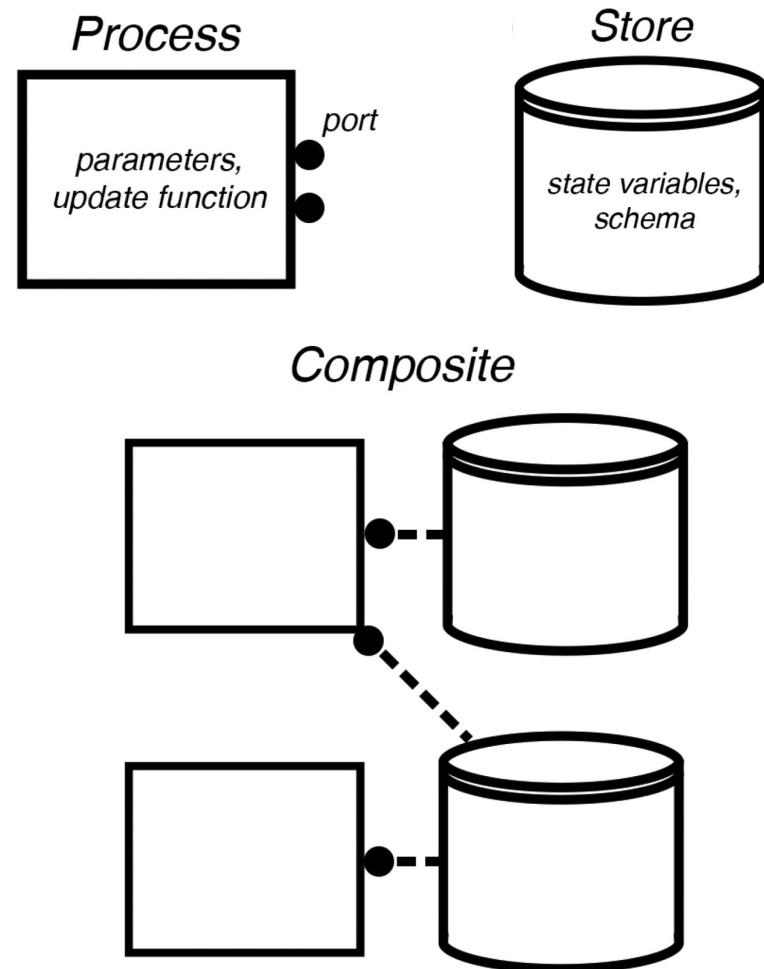
Processes and stores can be thought of as software implementations of the update functions and state variables of dynamical systems. For example, for the difference equation  $\Delta x = f(r, x) \cdot \Delta t$ , the Vivarium store would hold the state variable  $x$ , whereas the process would contain the update function  $f$ .

For example, the dynamic system of transcription can be modeled using a transcription process and stores for the DNA and mRNA concentrations.

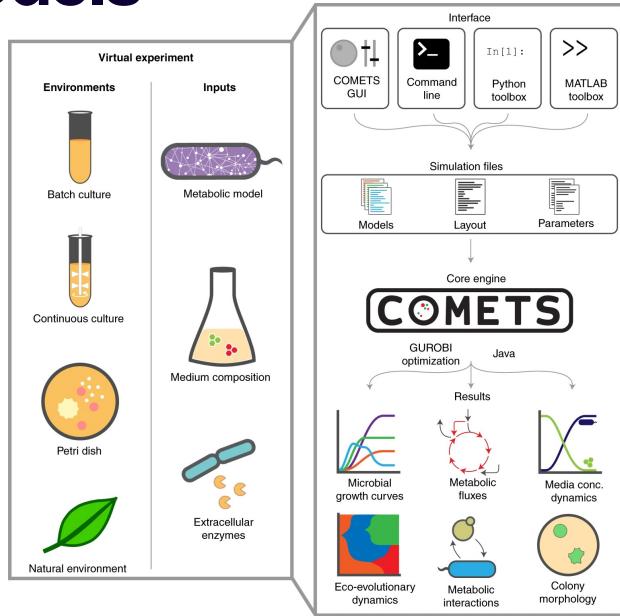
**Vivarium can be used to build composite models that combine several modeling frameworks (Agmon, 2022).**

Multiple processes can be coupled to create a composite model. Processes are connected through their ports using a topology to declare the connections.

The modularity of Vivarium lends itself well to the reuse of models developed by domain experts. This will ensure that composite models can build on deep knowledge bases for each individual module.



# A Vivarium Wrapper will Connect COMETS with other Models



**COMETS (Computation of Microbial Ecosystems in Time and Space) is an extension of dynamic FBA to generate simulations of multiple microbial species in molecularly complex and spatially structured environments (Dukovski, 2021).**

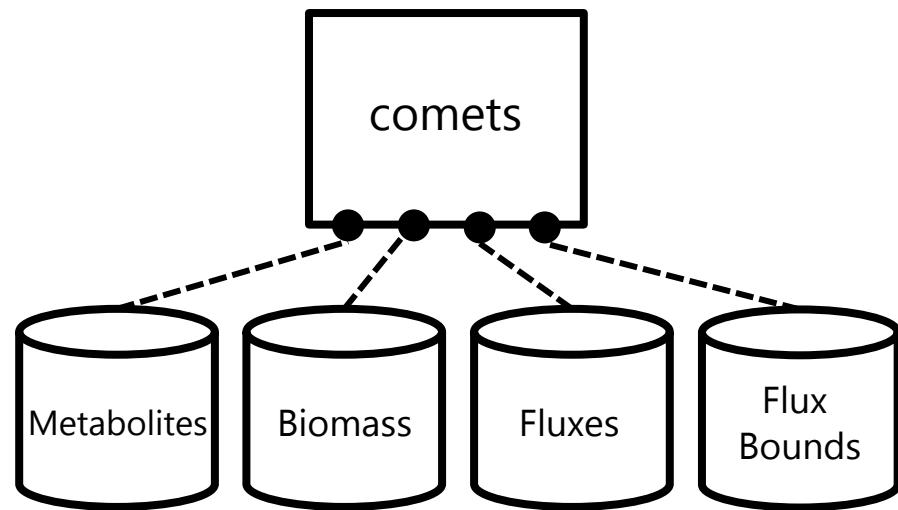
However, like all tools, COMETS has limitations, and by using it exclusively, one may miss out on insights that could be gained from other tools. Heterogenous model integration using Vivarium would allow us to extend COMETS with additional modules (Agmon, 2022).

*The first step for COMETS to join the Vivarium ecosystem of models is to write a vivarium process wrapper for the COMETSpy library.*

This wrapper will include a Process called "Comets" that has ports to connect to Stores of the 2D dimensional grids for biomass, metabolites, flux bounds, and resulting fluxes.

Notably, other processes can constrain individual fluxes in COMETS through the "Flux Bounds" port, allowing for integrative models.

In other Vivarium composites, adaptor processes play critical roles in integrating disparate processes (Agmon, 2022). These adaptors handle conversions between units, reference frames, namespaces, and data formats. While currently, no adaptor processes are planned, they will likely be necessary once we begin connecting the COMETS process with other processes.



# Connecting the Results of FBA to ODE Models

*With the ultimate goal of embedding the mechanistic understanding of FBA into the BATS-1D-VAR model, I will use calls to FBA to replace parameterized rates in the ecosystem scale ODEs.*

A schematic of the source and sink terms for heterotrophic bacteria in BATS-1D-VAR is shown to the right, and the corresponding equations are shown below.

Currently, BATS-1D-VAR is parameterized from 30 years of observed data at BATS. By replacing these parameters with results from an FBA-based simulation, we hope to better capture biological reality.

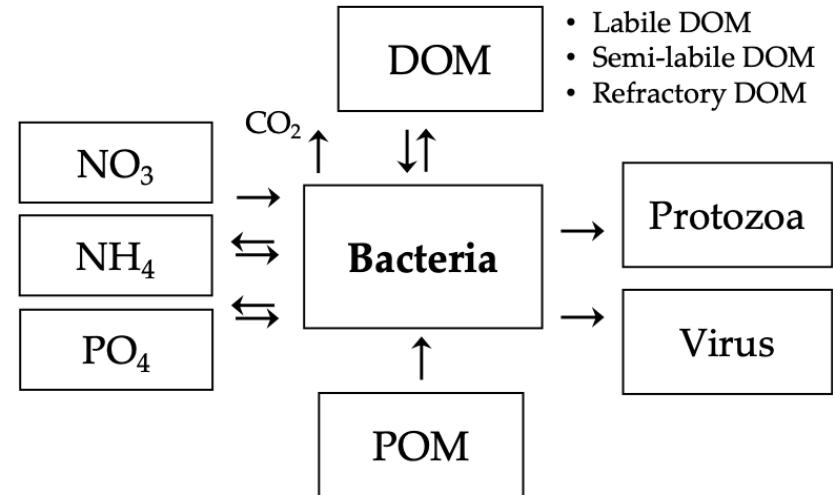
For example, FBA will inform dynamic organic carbon fate parameters. These new FBA-based values can help address the question of how the presence and abundance metabolite classes affect ocean biogeochemistry.

This work is preliminary, and any future work and publications will be highly collaborative and most likely led by partners who work at the ecosystem scale, including Dr. Scott Doney (University of Virginia) and Dr. Heather Kim (Woods Hole Oceanographic Institute).

$$\frac{dB^C}{dt} = \text{Uptake}^C_{DOM} + \text{Uptake}^C_{POM} - \text{Grazing}^C - \text{Virus}^C - \text{Respiration}^C - \text{Excretion}^C_{DOM}$$

$$\frac{dB^N}{dt} = \text{Uptake}^N_{DOM} + \text{Uptake}^N_{POM} + \text{Uptake}^N_{NH_4} + \text{Uptake}^N_{NO_3} - \text{Grazing}^N - \text{Virus}^N - \text{Excretion}^N_{DOM} - \text{Remineralization}^N$$

$$\frac{dB^P}{dt} = \text{Uptake}^P_{DOM} + \text{Uptake}^P_{POM} + \text{Uptake}^P_{PO_4} - \text{Grazing}^P - \text{Virus}^P - \text{Excretion}^P_{DOM} - \text{Remineralization}^P$$



## Bacterial variables

- Bacterial C biomass ( $= B^C$ )
- Bacterial respiration ( $= R^C$ )
- Bacterial C demand ( $= G^C_{DOM} + G^C_{POM}$ )
- Bacterial production ( $= G^C_{DOM} + G^C_{POM} - R^C$ )

**Challenge 5:**

# Community Education

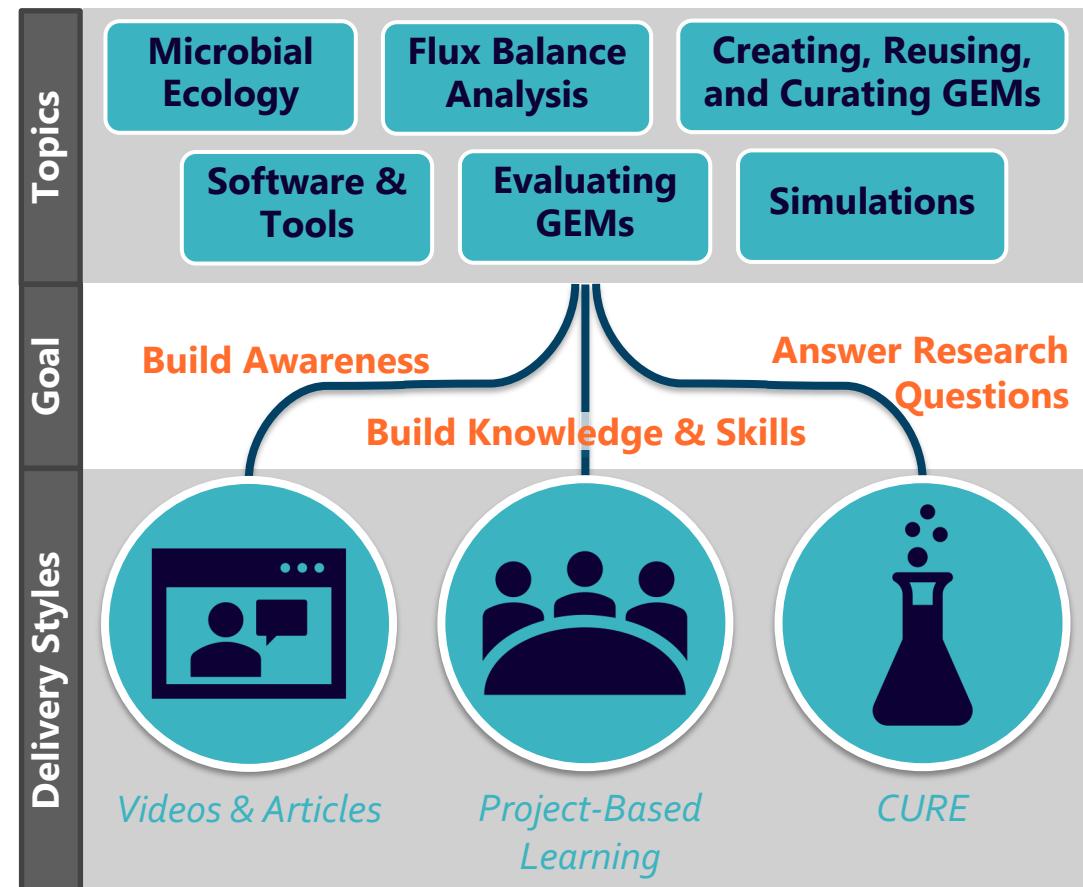
# Opportunities to Establish Training and Curricula to Prepare Scientists to Apply Genome-Scale Metabolic Modeling in Microbiome Studies

*Metabolic modeling has the potential to advance microbiome science. Yet, these modeling techniques have not been fully integrated into the microbiome research community (Ankrah, 2021).*

This is especially true in empirically focused research groups. A survey showed that over 70% of empirical researchers were interested in using metabolic modeling in their work, but a lack of computational expertise and concern about the accuracy of predictions has prevented these researchers from integrating models into their research (Ankrah, 2021).

To more fully integrate metabolic modeling into the community, modeling techniques need to be more accessible, and the communication and interpretation of simulations need to be more transparent.

To this end, I will curate and create educational materials that address these boundaries. I aim to create these materials using student-centered, evidence-based pedagogy.



All material will be written with Jupyter Book- an open-source tool for building publication-quality books and documents from computational material.

# CUREs: An Organized and Evidence-Based Revolution in STEM Education

A Course-based Undergraduate Research Experience (CURE) is a classroom-based course where students engage in original research (Dolan, 2021).

In a CURE, students will:

## 1. Engage in scientific practices.

This includes collecting and analyzing data, creating and revising models, generating and evaluating arguments, and communicating their findings.

## 2. Make discoveries that are relevant to stakeholders outside the classroom.

The scientific outcomes of the CURE should be unknown to the students, with the possibility of true discovery. The student's findings should be of interest to the broader scientific community or other stakeholders outside of the classroom.

## 3. Engage in iterative work.

Troubleshooting and problem-solving in a CURE increase student ownership of the research (Dolan, 2021). Repeating aspects of their work also increases confidence in the scientific results.

## 4. Collaborate with peers and more expert scientists.

This may include sharing progress, getting feedback, and compiling results to draw conclusions.

*I hope to use a CURE to help develop a student's identity as a Science Researcher.*

Pfiefer et. al have identified three levels of science researcher identity (unpublished). Each level of identity involves a different purpose of work, degree of intellectual responsibility and amount of operational autonomy. The three levels are:

**Science Student**  
Purpose of Work:  
Learn how to be a researcher

**Science Researcher**  
Purpose of Work:  
Answer research questions

**Professional Researcher**  
Purpose of Work: Ask research questions

I believe that a CURE will foster a student's Science Researcher identity. The CURE itself will serve as a research environment in which the student can work. In the CURE project, I aim to give students a high degree of intellectual responsibility for their work, and operational autonomy in selecting how to achieve their research goal.

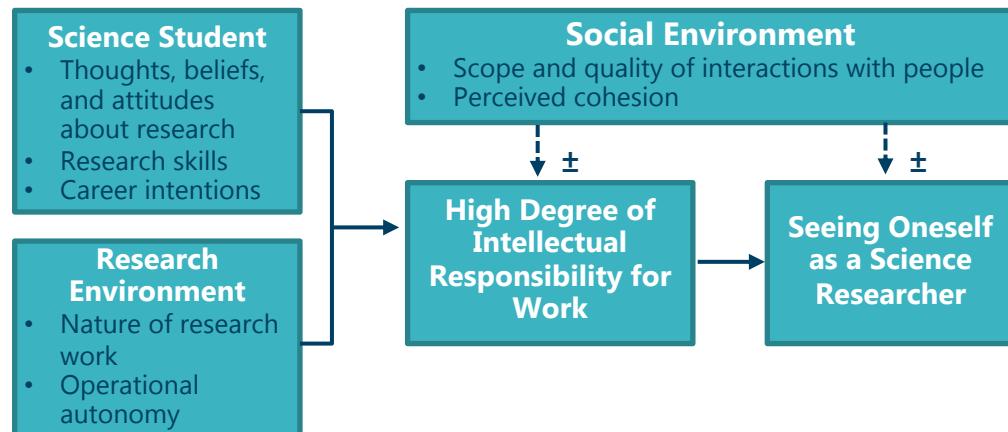


Figure Adapted from Mariel Pfiefer, unpublished | 36

# Developing a CURE to teach Metabolic Modeling

In metabolic engineering (ME), bioengineers manipulate the metabolic network of an organism to optimize the production of a compound of interest. Metabolic modeling with FBA is the basis of rational ME, as it allows engineers to test designs *in silico* (Sarkar, 2019).

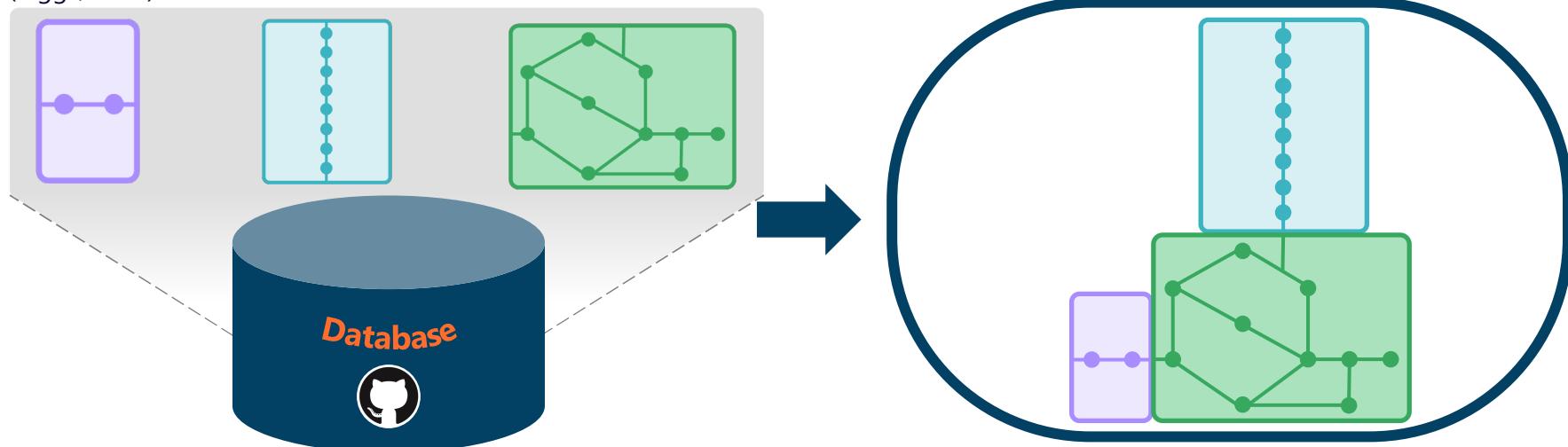
While ME describes the goal of a project, it does not necessarily describe the means. Often, ME projects lack the standardization and modularization developed in Synthetic Biology. Without this modularization, successes in the field are isolated and can not be transitioned to different chassis organisms or target molecules (Stephanopoulos, 2012).

Enzymes in metabolic pathways can be grouped into discrete modules. These metabolic modules can then be mixed and matched, and regulated independently to create new ME designs (Biggs, 2014).

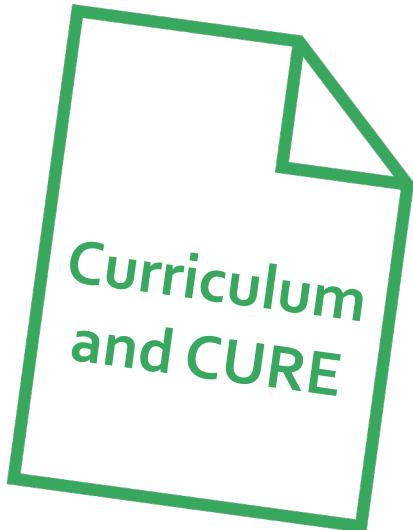
*I will develop a CURE where students will perform novel research to enable the modular engineering of metabolism.*

In this CURE, students will be guided through the process of generating a synthetic metabolic network to solve a problem they are interested in. In order to create these networks, students will use pre-defined metabolic modules (i.e. from a starter set I will generate) but will likely also need to curate new modules. At the end of the project, students will add these new modules to a publicly available collection, similar to the Registry of Standard Biological Parts.

This collection of metabolic modules will be of interest to the greater scientific community, including iGEM competition teams.



# Potential Publications



**Curriculum  
and CURE**

**Target journal:** CourseSource

**My role:** First Author

**Target submission date:** Fall 2023

I aim to publish all undergraduate-level material in CourseSource, an open-access journal of peer-reviewed teaching resources for undergraduate biology and physics.

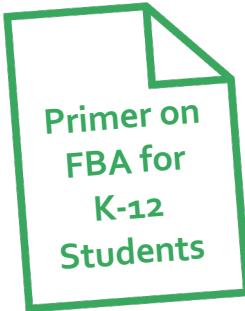
The CourseSource Learning Framework for Bioinformatics lays out learning goals and objectives relevant to undergraduate biological sciences majors. These objectives cover a variety of topics, including Systems Biology and Metabolomics. Within each topic, articles should address a specific learning goal.

To date, there have been no CourseSource articles published that address the learning goal "How can bioinformatics tools be employed to examine flow of molecules within pathways?". I believe my work, developing both the CURE and a project-based learning curriculum, will directly address this learning goal.

After pilot testing the materials, I will organize and format all materials so that it will be easy for other instructors to reuse them. For computational projects, this may include developing docker containers or GitHub Classrooms.

I will have support in developing this publication from collaborator Erin Dolan (UGA and member of C-CoMP) who has worked extensively with CourseSource.

# Potential Publications Continued



**Target journal:** Frontiers for Young Minds

**My role:** First Author

**Target submission date:** January 2023

This publication will be a Frontiers for Young Minds "Core Concepts" article that aims to explain linear programming and Flux Balance Analysis to a middle school audience.

The article walks through two simple optimization problems, one of a bicycle factory and one of a cell. In the examples, the solution to each problem is found graphically by plotting linear inequalities (see figures below).

**Ex: Optimize profit in a bicycle factory.**

$$2 \text{ (wheel)} + 6 \text{ (tire)} = \text{Bicycle}$$

$$3 \text{ (wheel)} + 3 \text{ (tire)} = \text{Tricycle}$$

$$\text{Max } 2 \text{ (wheel)} = 25 \quad 1 \text{ (tire)} = \$7$$

$$\text{Max } 3 \text{ (wheel)} = 45 \quad 1 \text{ (tire)} = \$8$$

	Bicycle Line	Tricycle Line
Wheel	-2	-3
Tire	-6	-3
Bicycle	1	0
Tricycle	0	1

**Ex: Optimize biomass production in a cell.**

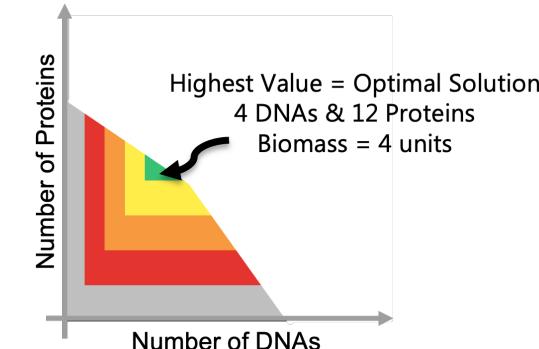
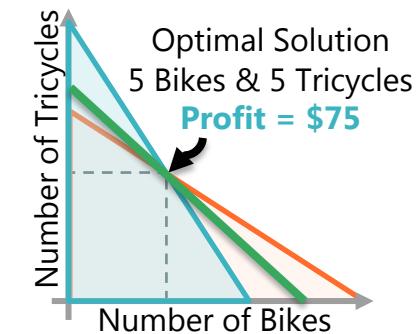
$$5 \text{ (C)} + 2 \text{ (N)} = \text{DNA}$$

$$4 \text{ (C)} + 1 \text{ (N)} = \text{Protein}$$

$$1 \text{ (DNA)} + 3 \text{ (Protein)} = \text{Biomass unit}$$

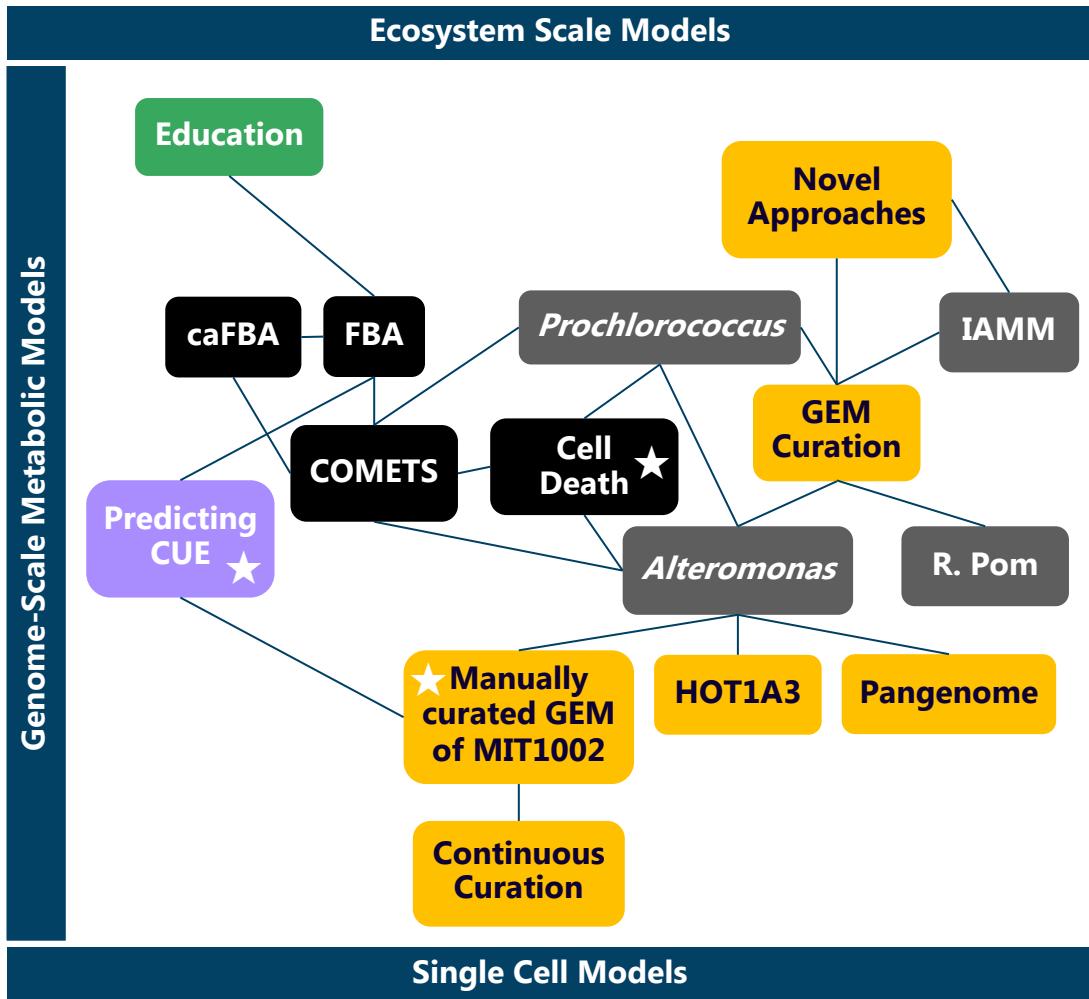
$$\text{Max C} = 75 \quad \text{Max N} = 24$$

	DNA Synthesis	Protein Synthesis	Biomass Reaction
C	-5	-4	0
N	-2	-1	0
DNA	1	0	-1
Protein	0	1	-3
Biomass Unit	0	0	1



# Project Map & Timeline

# Mind Map of all Projects

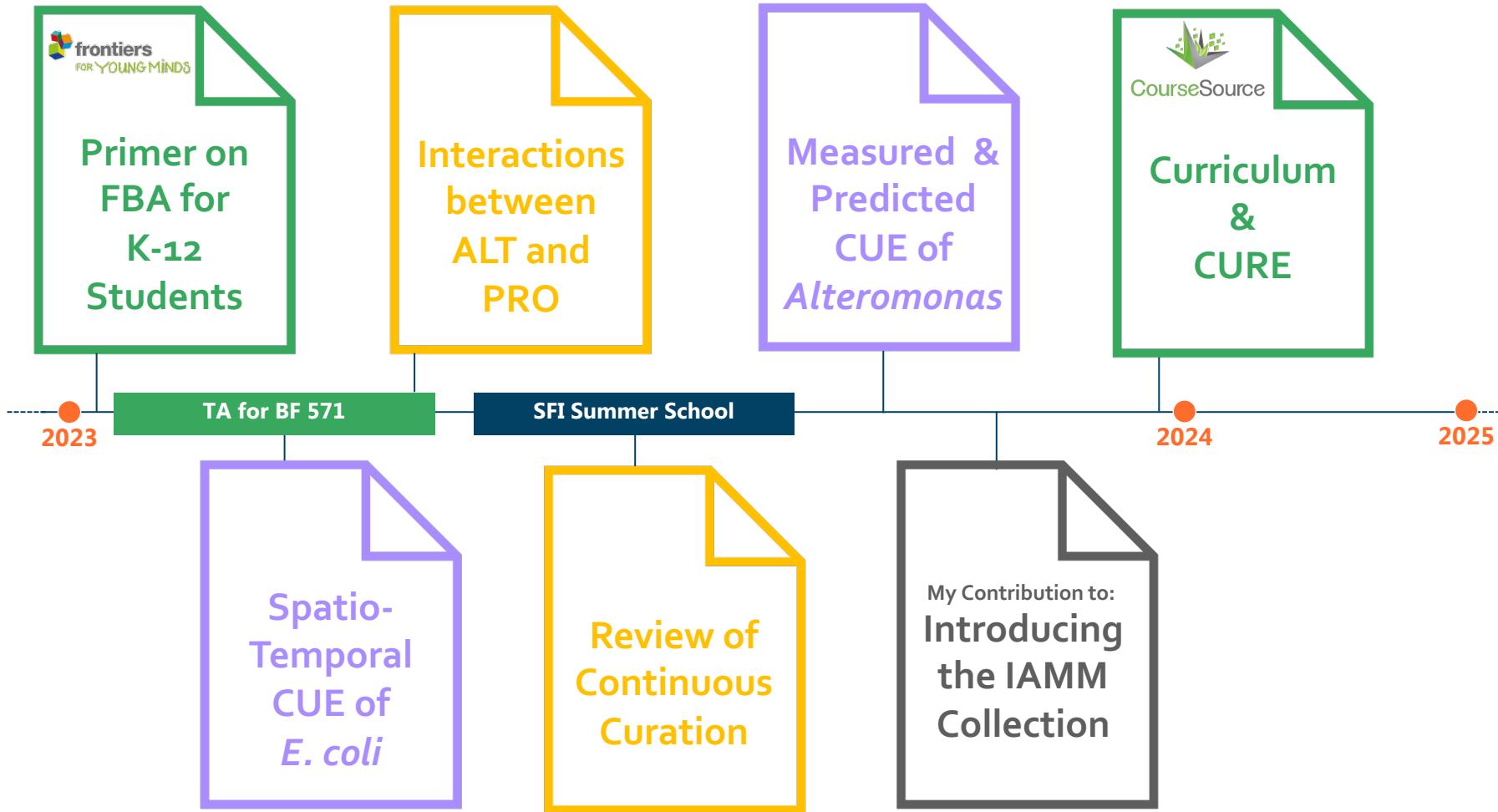


While these challenges may seem disparate, they are highly connected and offer many intermediate milestones.

Here, I highlight the sub-projects associated with each challenge and how the different sub-projects are connected.

Projects marked with a star are research projects (i.e. not reviews) where I will be the lead researcher.

# Project Milestones



# THANK YOU

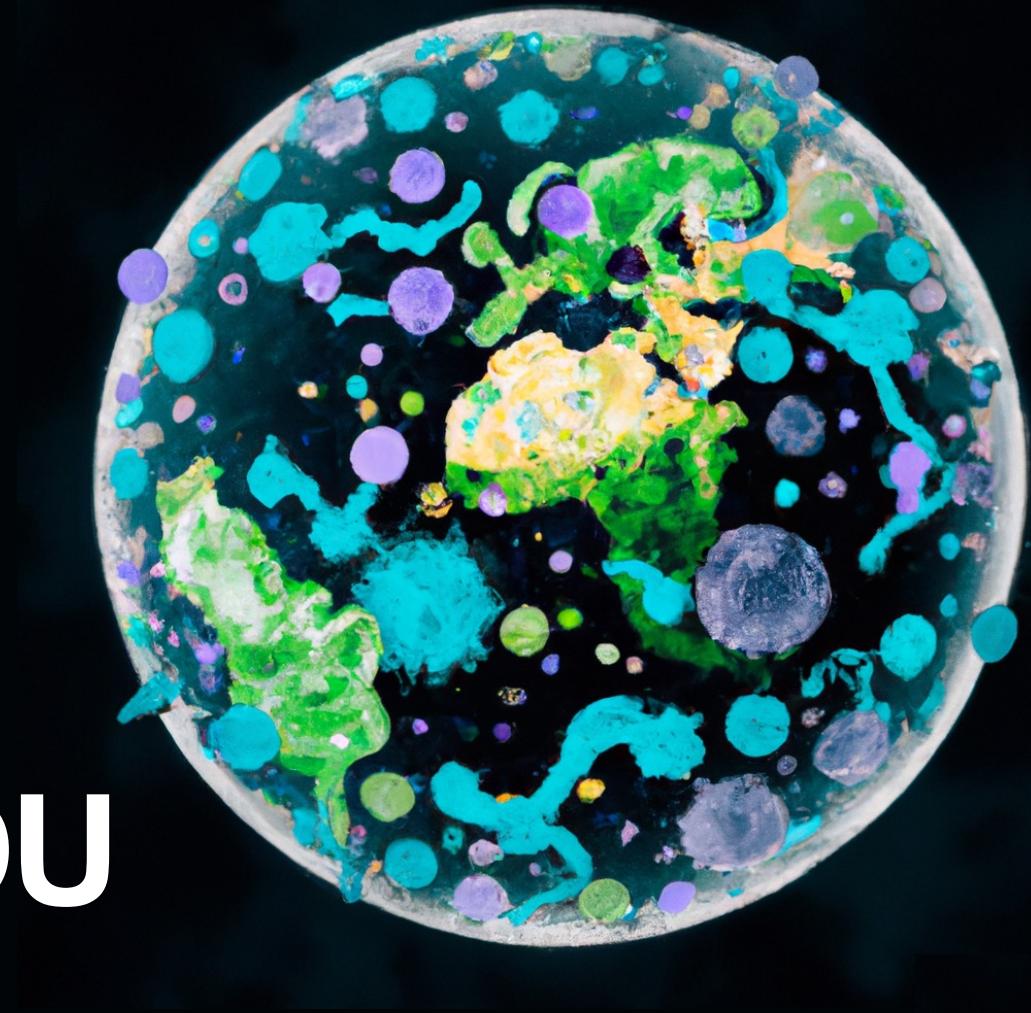
---

## *Contact Information*

*For more information, please contact me at:*

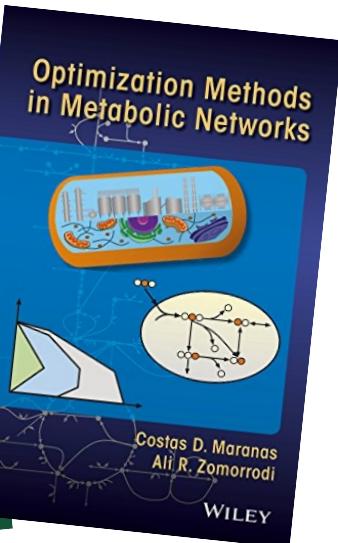
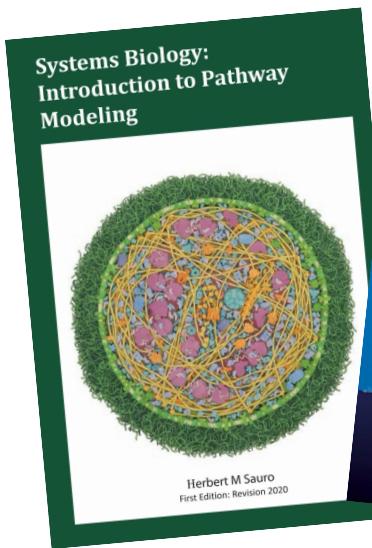
*[hscott@bu.edu](mailto:hscott@bu.edu)*

---

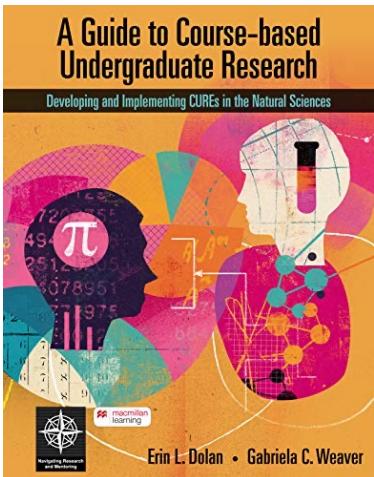


# Core Resources for My Project

## *Textbooks on Systems Biology and FBA*



## *Guide on Developing a CURE*



## *Review on Microbial CUE*

# ECOLOGY LETTERS

Ecology Letters, 2010, 13(9): 930–939

doi:10.1111/j.1462-9923.2010.02513.x

## REVIEW AND SYNTHESIS

### Carbon use efficiency of microbial communities: stoichiometry, methodology and modelling

Robert L. Shmida,<sup>a</sup> Stefanie Mansor,<sup>a</sup> Daryl L. Moorhead<sup>b</sup> and Andrew Bode<sup>b\*</sup>

**Abstract** Carbon use efficiency (CUE) is a fundamental parameter for ecological models and is central to the planning of environmental management. CUE determines energy and material flows to higher trophic levels, conversion of plant biomass to animal biomass, and the fate of energy and material inputs to ecosystems. In this review we show that the calculation of CUE requires a minimum CUE value of ~0.04 (CUE<sub>min</sub>). Kinetic and stoichiometric constraints on microbial growth support this CUE in a multi-source limited nutrient system, while stoichiometric approaches predict CUE values ranging from 0.04 to 0.55 because the methods used to estimate CUE in aquatic and terrestrial systems differ greatly. The range of CUE values is wider than the range of CUE<sub>min</sub>, which reflects the difficulty of measuring CUE in natural environments. We also show that CUE is often used as a measure of ecosystem productivity, which can also lead to overestimation of CUE. We recommend that the dual-scale model is a CUE value of 0.04, unless there is evidence for lower values as a result of precise nutrient limitations. Ecological models should include both CUE<sub>min</sub> and CUE values to predict the fate of energy and material inputs to ecosystems, as well as environmental drivers, to predict the CUE of microbial communities.

#### Keywords

biochemical stoichiometry, ecosystem ecology, ecological stoichiometry, microbial production, nutrient limitation, threshold element ratio.

*Ecol. Lett.* (2010), 13, 930–939

## INTRODUCTION

Most of the net primary production of the biosphere is transferred through decomposer flow (Cubitt & Lawlor 2006). The major role of decomposers in the cycling of energy and material mass from the metabolism of dead organic matter. The efficiency of this conversion, often termed carbon use efficiency (CUE), controls the fate of energy and material inputs to ecosystems, rates of energy conversion, energy and material fluxes through ecosystems and ecosystem productivity.

The same growth yields, growth efficiency, metabolic efficiency and CUE are often used interchangeably (Shmida & Ellner 1979).

Adding to the confusion, there are terms used interchangeably that differ in their capacity to represent the metabolism of ecosystems. These include the terms 'ecosystem respiration' between systems and the development of productive models (Moorhead & Shmida 2009). The term 'ecosystem respiration' is the thermodynamic, physiological and ecological consequences on ecosystems of the metabolism of ecosystems. This contrasts with the growth of microbial communities; the limitations of existing methodology for assessing microbial community growth; and the unique properties of microbial communities that distinguish them from other organisms. All these give rise to clearly apparent inconsistencies in the literature, particularly the terminology of microbial communities.

<sup>a</sup>Biology Department, University of New Mexico, Albuquerque, NM, 87131, USA

<sup>b</sup>Centre for Environmental Risk Assessment and Remediation, Durham University, Durham, UK, DH1 3LE

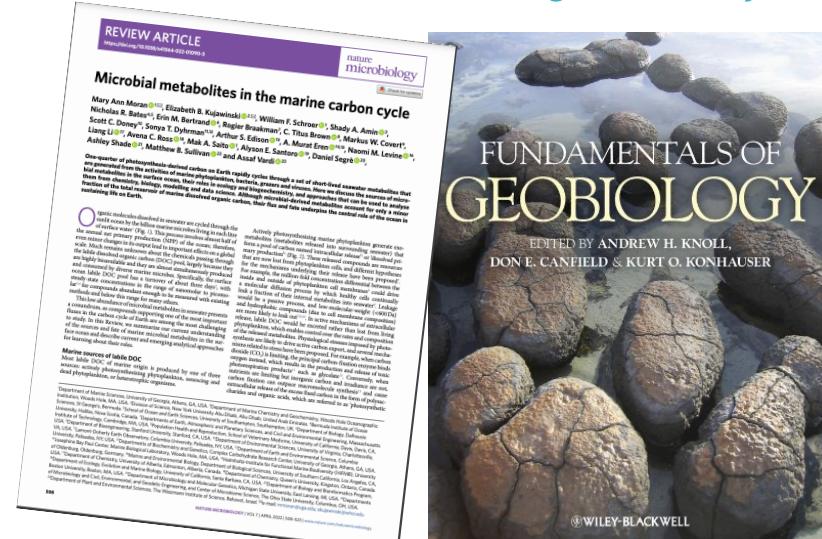
obtain across systems and recommended strategies for representing the CUE of microbial communities in ecological process models.

## MICROBIAL CUE: DEFINITIONS AND CONTROLS

The stoichiometry of microbial growth can be measured in many ways. The most common method is to culture growth curves in media with a range of elemental ratios (e.g. C:N:P) and measure the CUE<sub>min</sub> (ratio of substrate elements) required for separation of growth from death (Shmida & Ellner 1979). CUE<sub>min</sub> is the ratio of net energy consumed from the medium by the assimilated substrate (Shmida & Ellner 1979; Shmida 1981; Shmida & Ellner 1984; Pavia & Shmida 1985; Pavia & Shmida 1987; Van Steenbergh & Merviel 1993; Levine & Kajikawa 1994).

Microbial growth can also be measured by calculating the rate of increase in mass, rather than energy. In the case of pure cultures, the CUE<sub>max</sub> (ratio of substrate elements to the rate of growth) is equal to the CUE<sub>min</sub>. CUE<sub>max</sub> is generally defined as the ratio of growth rate to the rate of energy assimilation (Shmida & Ellner 1979; Ellner & Shmida 1982; Ellner & Shmida 1984; Ellner & Shmida 1985; Ellner & Shmida 1986; Ellner & Shmida 1987; Ellner & Shmida 1988; Ellner & Shmida 1989; Ellner & Shmida 1990; Ellner & Shmida 1991; Ellner & Shmida 1992; Ellner & Shmida 1993; Ellner & Shmida 1994; Ellner & Shmida 1995; Ellner & Shmida 1996; Ellner & Shmida 1997; Ellner & Shmida 1998; Ellner & Shmida 1999; Ellner & Shmida 2000; Ellner & Shmida 2001; Ellner & Shmida 2002; Ellner & Shmida 2003; Ellner & Shmida 2004; Ellner & Shmida 2005; Ellner & Shmida 2006; Ellner & Shmida 2007; Ellner & Shmida 2008; Ellner & Shmida 2009; Ellner & Shmida 2010; Ellner & Shmida 2011; Ellner & Shmida 2012; Ellner & Shmida 2013; Ellner & Shmida 2014; Ellner & Shmida 2015; Ellner & Shmida 2016; Ellner & Shmida 2017; Ellner & Shmida 2018; Ellner & Shmida 2019; Ellner & Shmida 2020; Ellner & Shmida 2021; Ellner & Shmida 2022; Ellner & Shmida 2023; Ellner & Shmida 2024; Ellner & Shmida 2025; Ellner & Shmida 2026; Ellner & Shmida 2027; Ellner & Shmida 2028; Ellner & Shmida 2029; Ellner & Shmida 2030; Ellner & Shmida 2031; Ellner & Shmida 2032; Ellner & Shmida 2033; Ellner & Shmida 2034; Ellner & Shmida 2035; Ellner & Shmida 2036; Ellner & Shmida 2037; Ellner & Shmida 2038; Ellner & Shmida 2039; Ellner & Shmida 2040; Ellner & Shmida 2041; Ellner & Shmida 2042; Ellner & Shmida 2043; Ellner & Shmida 2044; Ellner & Shmida 2045; Ellner & Shmida 2046; Ellner & Shmida 2047; Ellner & Shmida 2048; Ellner & Shmida 2049; Ellner & Shmida 2050; Ellner & Shmida 2051; Ellner & Shmida 2052; Ellner & Shmida 2053; Ellner & Shmida 2054; Ellner & Shmida 2055; Ellner & Shmida 2056; Ellner & Shmida 2057; Ellner & Shmida 2058; Ellner & Shmida 2059; Ellner & Shmida 2060; Ellner & Shmida 2061; Ellner & Shmida 2062; Ellner & Shmida 2063; Ellner & Shmida 2064; Ellner & Shmida 2065; Ellner & Shmida 2066; Ellner & Shmida 2067; Ellner & Shmida 2068; Ellner & Shmida 2069; Ellner & Shmida 2070; Ellner & Shmida 2071; Ellner & Shmida 2072; Ellner & Shmida 2073; Ellner & Shmida 2074; Ellner & Shmida 2075; Ellner & Shmida 2076; Ellner & Shmida 2077; Ellner & Shmida 2078; Ellner & Shmida 2079; Ellner & Shmida 2080; Ellner & Shmida 2081; Ellner & Shmida 2082; Ellner & Shmida 2083; Ellner & Shmida 2084; Ellner & Shmida 2085; Ellner & Shmida 2086; Ellner & Shmida 2087; Ellner & Shmida 2088; Ellner & Shmida 2089; Ellner & Shmida 2090; Ellner & Shmida 2091; Ellner & Shmida 2092; Ellner & Shmida 2093; Ellner & Shmida 2094; Ellner & Shmida 2095; Ellner & Shmida 2096; Ellner & Shmida 2097; Ellner & Shmida 2098; Ellner & Shmida 2099; Ellner & Shmida 20100; Ellner & Shmida 20101; Ellner & Shmida 20102; Ellner & Shmida 20103; Ellner & Shmida 20104; Ellner & Shmida 20105; Ellner & Shmida 20106; Ellner & Shmida 20107; Ellner & Shmida 20108; Ellner & Shmida 20109; Ellner & Shmida 20110; Ellner & Shmida 20111; Ellner & Shmida 20112; Ellner & Shmida 20113; Ellner & Shmida 20114; Ellner & Shmida 20115; Ellner & Shmida 20116; Ellner & Shmida 20117; Ellner & Shmida 20118; Ellner & Shmida 20119; Ellner & Shmida 20120; Ellner & Shmida 20121; Ellner & Shmida 20122; Ellner & Shmida 20123; Ellner & Shmida 20124; Ellner & Shmida 20125; Ellner & Shmida 20126; Ellner & Shmida 20127; Ellner & Shmida 20128; Ellner & Shmida 20129; Ellner & Shmida 20130; Ellner & Shmida 20131; Ellner & Shmida 20132; Ellner & Shmida 20133; Ellner & Shmida 20134; Ellner & Shmida 20135; Ellner & Shmida 20136; Ellner & Shmida 20137; Ellner & Shmida 20138; Ellner & Shmida 20139; Ellner & Shmida 20140; Ellner & Shmida 20141; Ellner & Shmida 20142; Ellner & Shmida 20143; Ellner & Shmida 20144; Ellner & Shmida 20145; Ellner & Shmida 20146; Ellner & Shmida 20147; Ellner & Shmida 20148; Ellner & Shmida 20149; Ellner & Shmida 20150; Ellner & Shmida 20151; Ellner & Shmida 20152; Ellner & Shmida 20153; Ellner & Shmida 20154; Ellner & Shmida 20155; Ellner & Shmida 20156; Ellner & Shmida 20157; Ellner & Shmida 20158; Ellner & Shmida 20159; Ellner & Shmida 20160; Ellner & Shmida 20161; Ellner & Shmida 20162; Ellner & Shmida 20163; Ellner & Shmida 20164; Ellner & Shmida 20165; Ellner & Shmida 20166; Ellner & Shmida 20167; Ellner & Shmida 20168; Ellner & Shmida 20169; Ellner & Shmida 20170; Ellner & Shmida 20171; Ellner & Shmida 20172; Ellner & Shmida 20173; Ellner & Shmida 20174; Ellner & Shmida 20175; Ellner & Shmida 20176; Ellner & Shmida 20177; Ellner & Shmida 20178; Ellner & Shmida 20179; Ellner & Shmida 20180; Ellner & Shmida 20181; Ellner & Shmida 20182; Ellner & Shmida 20183; Ellner & Shmida 20184; Ellner & Shmida 20185; Ellner & Shmida 20186; Ellner & Shmida 20187; Ellner & Shmida 20188; Ellner & Shmida 20189; Ellner & Shmida 20190; Ellner & Shmida 20191; Ellner & Shmida 20192; Ellner & Shmida 20193; Ellner & Shmida 20194; Ellner & Shmida 20195; Ellner & Shmida 20196; Ellner & Shmida 20197; Ellner & Shmida 20198; Ellner & Shmida 20199; Ellner & Shmida 20200; Ellner & Shmida 20201; Ellner & Shmida 20202; Ellner & Shmida 20203; Ellner & Shmida 20204; Ellner & Shmida 20205; Ellner & Shmida 20206; Ellner & Shmida 20207; Ellner & Shmida 20208; Ellner & Shmida 20209; Ellner & Shmida 20210; Ellner & Shmida 20211; Ellner & Shmida 20212; Ellner & Shmida 20213; Ellner & Shmida 20214; Ellner & Shmida 20215; Ellner & Shmida 20216; Ellner & Shmida 20217; Ellner & Shmida 20218; Ellner & Shmida 20219; Ellner & Shmida 20220; Ellner & Shmida 20221; Ellner & Shmida 20222; Ellner & Shmida 20223; Ellner & Shmida 20224; Ellner & Shmida 20225; Ellner & Shmida 20226; Ellner & Shmida 20227; Ellner & Shmida 20228; Ellner & Shmida 20229; Ellner & Shmida 20230; Ellner & Shmida 20231; Ellner & Shmida 20232; Ellner & Shmida 20233; Ellner & Shmida 20234; Ellner & Shmida 20235; Ellner & Shmida 20236; Ellner & Shmida 20237; Ellner & Shmida 20238; Ellner & Shmida 20239; Ellner & Shmida 20240; Ellner & Shmida 20241; Ellner & Shmida 20242; Ellner & Shmida 20243; Ellner & Shmida 20244; Ellner & Shmida 20245; Ellner & Shmida 20246; Ellner & Shmida 20247; Ellner & Shmida 20248; Ellner & Shmida 20249; Ellner & Shmida 20250; Ellner & Shmida 20251; Ellner & Shmida 20252; Ellner & Shmida 20253; Ellner & Shmida 20254; Ellner & Shmida 20255; Ellner & Shmida 20256; Ellner & Shmida 20257; Ellner & Shmida 20258; Ellner & Shmida 20259; Ellner & Shmida 20260; Ellner & Shmida 20261; Ellner & Shmida 20262; Ellner & Shmida 20263; Ellner & Shmida 20264; Ellner & Shmida 20265; Ellner & Shmida 20266; Ellner & Shmida 20267; Ellner & Shmida 20268; Ellner & Shmida 20269; Ellner & Shmida 20270; Ellner & Shmida 20271; Ellner & Shmida 20272; Ellner & Shmida 20273; Ellner & Shmida 20274; Ellner & Shmida 20275; Ellner & Shmida 20276; Ellner & Shmida 20277; Ellner & Shmida 20278; Ellner & Shmida 20279; Ellner & Shmida 20280; Ellner & Shmida 20281; Ellner & Shmida 20282; Ellner & Shmida 20283; Ellner & Shmida 20284; Ellner & Shmida 20285; Ellner & Shmida 20286; Ellner & Shmida 20287; Ellner & Shmida 20288; Ellner & Shmida 20289; Ellner & Shmida 20290; Ellner & Shmida 20291; Ellner & Shmida 20292; Ellner & Shmida 20293; Ellner & Shmida 20294; Ellner & Shmida 20295; Ellner & Shmida 20296; Ellner & Shmida 20297; Ellner & Shmida 20298; Ellner & Shmida 20299; Ellner & Shmida 20300; Ellner & Shmida 20301; Ellner & Shmida 20302; Ellner & Shmida 20303; Ellner & Shmida 20304; Ellner & Shmida 20305; Ellner & Shmida 20306; Ellner & Shmida 20307; Ellner & Shmida 20308; Ellner & Shmida 20309; Ellner & Shmida 20310; Ellner & Shmida 20311; Ellner & Shmida 20312; Ellner & Shmida 20313; Ellner & Shmida 20314; Ellner & Shmida 20315; Ellner & Shmida 20316; Ellner & Shmida 20317; Ellner & Shmida 20318; Ellner & Shmida 20319; Ellner & Shmida 20320; Ellner & Shmida 20321; Ellner & Shmida 20322; Ellner & Shmida 20323; Ellner & Shmida 20324; Ellner & Shmida 20325; Ellner & Shmida 20326; Ellner & Shmida 20327; Ellner & Shmida 20328; Ellner & Shmida 20329; Ellner & Shmida 20330; Ellner & Shmida 20331; Ellner & Shmida 20332; Ellner & Shmida 20333; Ellner & Shmida 20334; Ellner & Shmida 20335; Ellner & Shmida 20336; Ellner & Shmida 20337; Ellner & Shmida 20338; Ellner & Shmida 20339; Ellner & Shmida 20340; Ellner & Shmida 20341; Ellner & Shmida 20342; Ellner & Shmida 20343; Ellner & Shmida 20344; Ellner & Shmida 20345; Ellner & Shmida 20346; Ellner & Shmida 20347; Ellner & Shmida 20348; Ellner & Shmida 20349; Ellner & Shmida 20350; Ellner & Shmida 20351; Ellner & Shmida 20352; Ellner & Shmida 20353; Ellner & Shmida 20354; Ellner & Shmida 20355; Ellner & Shmida 20356; Ellner & Shmida 20357; Ellner & Shmida 20358; Ellner & Shmida 20359; Ellner & Shmida 20360; Ellner & Shmida 20361; Ellner & Shmida 20362; Ellner & Shmida 20363; Ellner & Shmida 20364; Ellner & Shmida 20365; Ellner & Shmida 20366; Ellner & Shmida 20367; Ellner & Shmida 20368; Ellner & Shmida 20369; Ellner & Shmida 20370; Ellner & Shmida 20371; Ellner & Shmida 20372; Ellner & Shmida 20373; Ellner & Shmida 20374; Ellner & Shmida 20375; Ellner & Shmida 20376; Ellner & Shmida 20377; Ellner & Shmida 20378; Ellner & Shmida 20379; Ellner & Shmida 20380; Ellner & Shmida 20381; Ellner & Shmida 20382; Ellner & Shmida 20383; Ellner & Shmida 20384; Ellner & Shmida 20385; Ellner & Shmida 20386; Ellner & Shmida 20387; Ellner & Shmida 20388; Ellner & Shmida 20389; Ellner & Shmida 20390; Ellner & Shmida 20391; Ellner & Shmida 20392; Ellner & Shmida 20393; Ellner & Shmida 20394; Ellner & Shmida 20395; Ellner & Shmida 20396; Ellner & Shmida 20397; Ellner & Shmida 20398; Ellner & Shmida 20399; Ellner & Shmida 20400; Ellner & Shmida 20401; Ellner & Shmida 20402; Ellner & Shmida 20403; Ellner & Shmida 20404; Ellner & Shmida 20405; Ellner & Shmida 20406; Ellner & Shmida 20407; Ellner & Shmida 20408; Ellner & Shmida 20409; Ellner & Shmida 20410; Ellner & Shmida 20411; Ellner & Shmida 20412; Ellner & Shmida 20413; Ellner & Shmida 20414; Ellner & Shmida 20415; Ellner & Shmida 20416; Ellner & Shmida 20417; Ellner & Shmida 20418; Ellner & Shmida 20419; Ellner & Shmida 20420; Ellner & Shmida 20421; Ellner & Shmida 20422; Ellner & Shmida 20423; Ellner & Shmida 20424; Ellner & Shmida 20425; Ellner & Shmida 20426; Ellner & Shmida 20427; Ellner & Shmida 20428; Ellner & Shmida 20429; Ellner & Shmida 20430; Ellner & Shmida 20431; Ellner & Shmida 20432; Ellner & Shmida 20433; Ellner & Shmida 20434; Ellner & Shmida 20435; Ellner & Shmida 20436; Ellner & Shmida 20437; Ellner & Shmida 20438; Ellner & Shmida 20439; Ellner & Shmida 20440; Ellner & Shmida 20441; Ellner & Shmida 20442; Ellner & Shmida 20443; Ellner & Shmida 20444; Ellner & Shmida 20445; Ellner & Shmida 20446; Ellner & Shmida 20447; Ellner & Shmida 20448; Ellner & Shmida 20449; Ellner & Shmida 20450; Ellner & Shmida 20451; Ellner & Shmida 20452; Ellner & Shmida 20453; Ellner & Shmida 20454; Ellner & Shmida 20455; Ellner & Shmida 20456; Ellner & Shmida 20457; Ellner & Shmida 20458; Ellner & Shmida 20459; Ellner & Shmida 20460; Ellner & Shmida 20461; Ellner & Shmida 20462; Ellner & Shmida 20463; Ellner & Shmida 20464; Ellner & Shmida 20465; Ellner & Shmida 20466; Ellner & Shmida 20467; Ellner & Shmida 20468; Ellner & Shmida 20469; Ellner & Shmida 20470; Ellner & Shmida 20471; Ellner & Shmida 20472; Ellner & Shmida 20473; Ellner & Shmida 20474; Ellner & Shmida 20475; Ellner & Shmida 20476; Ellner & Shmida 20477; Ellner & Shmida 20478; Ellner & Shmida 20479; Ellner & Shmida 20480; Ellner & Shmida 20481; Ellner & Shmida 20482; Ellner & Shmida 20483; Ellner & Shmida 20484; Ellner & Shmida 20485; Ellner & Shmida 20486; Ellner & Shmida 20487; Ellner & Shmida 20488; Ellner & Shmida 20489; Ellner & Shmida 20490; Ellner & Shmida 20491; Ellner & Shmida 20492; Ellner & Shmida 20493; Ellner & Shmida 20494; Ellner & Shmida 20495; Ellner & Shmida 20496; Ellner & Shmida 20497; Ellner & Shmida 20498; Ellner & Shmida 20499; Ellner & Shmida 20500; Ellner & Shmida 20501; Ellner & Shmida 20502; Ellner & Shmida 20503; Ellner & Shmida 20504; Ellner & Shmida 20505; Ellner & Shmida 20506; Ellner & Shmida 20507; Ellner & Shmida 20508; Ellner & Shmida 20509; Ellner & Shmida 20510; Ellner & Shmida 20511; Ellner & Shmida 20512; Ellner & Shmida 20513; Ellner & Shmida 20514; Ellner & Shmida 20515; Ellner & Shmida 20516; Ellner & Shmida 20517; Ellner & Shmida 20518; Ellner & Shmida 20519; Ellner & Shmida 20520; Ellner & Shmida 20521; Ellner & Shmida 20522; Ellner & Shmida 20523; Ellner & Shmida 20524; Ellner & Shmida 20525; Ellner & Shmida 20526; Ellner & Shmida 20527; Ellner & Shmida 20528; Ellner & Shmida 20529; Ellner & Shmida 20530; Ellner & Shmida 20531; Ellner & Shmida 20532; Ellner & Shmida 20533; Ellner & Shmida 20534; Ellner & Shmida 20535; Ellner & Shmida 20536; Ellner & Shmida 20537; Ellner & Shmida 20538; Ellner & Shmida 20539; Ellner & Shmida 20540; Ellner & Shmida 20541; Ellner & Shmida 20542; Ellner & Shmida 20543; Ellner & Shmida 20544; Ellner & Shmida 20545; Ellner & Shmida 20546; Ellner & Shmida 20547; Ellner & Shmida 20548; Ellner & Shmida 20549; Ellner & Shmida 20550; Ellner & Shmida 20551; Ellner & Shmida 20552; Ellner & Shmida 20553; Ellner & Shmida 20554; Ellner & Shmida 20555; Ellner & Shmida 20556; Ellner & Shmida 20557; Ellner & Shmida 20558; Ellner & Shmida 20559; Ellner & Shmida 20560; Ellner & Shmida 20561; Ellner & Shmida 20562; Ellner & Shmida 20563; Ellner & Shmida 20564; Ellner & Shmida 20565; Ellner & Shmida 20566; Ellner & Shmida 20567; Ellner & Shmida 20568; Ellner & Shmida 20569; Ellner & Shmida 20570; Ellner & Shmida 20571; Ellner & Shmida 20572; Ellner & Shmida 20573; Ellner & Shmida 20574; Ellner & Shmida 20575; Ellner & Shmida 20576; Ellner & Shmida 20577; Ellner & Shmida 20578; Ellner & Shmida 20579; Ellner & Shmida 20580; Ellner & Shmida 20581; Ellner & Shmida 20582; Ellner & Shmida 20583; Ellner & Shmida 20584; Ellner & Shmida 20585; Ellner & Shmida 20586; Ellner & Shmida 20587; Ellner & Shmida 20588; Ellner & Shmida 20589; Ellner & Shmida 20590; Ellner & Shmida 20591; Ellner & Shmida 20592; Ellner & Shmida 20593; Ellner & Shmida 20594; Ellner & Shmida 20595; Ellner & Shmida 20596; Ellner & Shmida 20597; Ellner & Shmida 20598; Ellner & Shmida 20599; Ellner & Shmida 20600; Ellner & Shmida 20601; Ellner & Shmida 20602; Ellner & Shmida 20603; Ellner & Shmida 20604; Ellner & Shmida 20605; Ellner & Shmida 20606; Ellner & Shmida 20607; Ellner & Shmida 20608; Ellner & Shmida 20609; Ellner & Shmida 20610; Ellner & Shmida 20611; Ellner & Shmida 20612; Ellner & Shmida 20613; Ellner & Shmida 20614; Ellner & Shmida 20615; Ellner & Shmida 20616; Ellner & Shmida 20617; Ellner & Shmida 20618; Ellner & Shmida 20619; Ellner & Shmida 20620; Ellner & Shmida 20621; Ellner & Shmida 20622; Ellner & Shmida 20623; Ellner & Shmida 20624; Ellner & Shmida 20625; Ellner & Shmida 20626; Ellner & Shmida 20627; Ellner & Shmida 20628; Ellner & Shmida 20629; Ellner & Shmida 20630; Ellner & Shmida 20631; Ellner & Shmida 20632; Ellner & Shmida 20633; Ellner & Shmida 20634; Ellner & Shmida 20635; Ellner & Shmida 20636; Ellner & Shmida 20637; Ellner & Shmida 20638; Ellner & Shmida 20639; Ellner & Shmida 20640; Ellner & Shmida 20641; Ellner & Shmida 20642; Ellner & Shmida 20643; Ellner & Shmida 20644; Ellner & Shmida 20645; Ellner & Shmida 20646; Ellner & Shmida 20647; Ellner & Shmida 20648; Ellner & Shmida 20649; Ellner & Shmida 20650; Ellner & Shmida 20651; Ellner & Shmida 20652; Ellner & Shmida 20653; Ellner & Shmida 20654; Ellner & Shmida 20655; Ellner & Shmida 20656; Ellner & Shmida 20657; Ellner & Shmida 20658; Ellner & Shmida 20659; Ellner & Shmida 20660; Ellner & Shmida 20661; Ellner & Shmida 20662; Ellner & Shmida 20663; Ellner & Shmida 20664; Ellner & Shmida 20665; Ellner & Shmida 20666; Ellner & Shmida 20667; Ellner & Shmida 20668; Ellner & Shmida 20669; Ellner & Shmida 20670; Ellner & Shmida 20671; Ellner & Shmida 20672; Ellner & Shmida 20673; Ellner & Shmida 20674; Ellner & Shmida 20675; Ellner & Shmida 20676; Ellner & Shmida 20677; Ellner & Shmida 20678; Ellner & Shmida 20679; Ellner & Shmida 20680; Ellner & Shmida 20681; Ellner & Shmida 20682; Ellner & Shmida 20683; Ellner & Shmida 20684; Ellner & Shmida 20685; Ellner & Shmida 20686; Ellner & Shmida 20687; Ellner & Shmida 20688; Ellner & Shmida 20689; Ellner & Shmida 20690; Ellner & Shmida 20691; Ellner & Shmida 20692; Ellner & Shmida 20693; Ellner & Shmida 20694; Ellner & Shmida 20695; Ellner & Shmida 20696; Ellner & Sh

*Textbook & Review on Biogeochemistry*



# Acknowledgments

*I am very grateful for all the people who are making this project possible, from Boston University, C-CoMP, and other external collaborations, including:*

## The Segrè Lab:

- Kathryn Atherton
- Sarah Bald
- Ryan Cornelius
- Dr. Ilija Dukovski
- Dr. Konrad Herbst
- Dr. Dileep Kishore
- Devlin Moyer
- Dr. Melisa Osbone
- Dr. Daniel Segrè
- Hui Shi
- Michael Silverstein
- Kaitlin Valli
- Jing Zhang

## My Thesis Advisory Committee:

- Dr. Brian Cleary
- Dr. Wally Fulweiler

## C-CoMP:

- Dr. Eran Agmon
- Dr. Rogier Braakman
- Dr. Victoria Centurino
- Dr. Zac Cooper
- Dr. Markus Covert
- Michelle DeMers
- Dr. Erin Dolan
- Dr. Scott Doney
- Dr. Laura Gray
- Helen Harding
- Dr. Sofia Ibarrarán-Viniegra
- Dr. Heather Kim
- Dr. Elizabeth Kujawinski
- Dr. Sam Miller
- Dr. Mary Ann Moran
- Dr. Mak Saito
- Michael Yang

## The Sher Lab:

- Dr. Daniel Sher
- Osnat Weissberg

## iGEM Engineering Committee:

- Dr. Jake Beal

## The Bioinformatics Program:

- Dr. Gary Benson
- Mary Ellen Gibson-FitzPatrick
- Dave King
- Caroline Lyman
- Dr. Thomas Tullius
- Johanna Vasquez

# References

- Agmon, Eran, Ryan K. Spangler, Christopher J. Skalnik, William Poole, Shayn M. Peirce, Jerry H. Morrison, and Markus W. Covert. 2022. "Vivarium: An Interface and Engine for Integrative Multiscale Modeling in Computational Biology." *Bioinformatics*, February. <https://doi.org/10.1093/bioinformatics/btac049>.
- Ankrah, Nana Y. D., David B. Bernstein, Matthew Biggs, Maureen Carey, Melinda Engevik, Beatriz García-Jiménez, Meiyappan Lakshmanan, Alan R. Pacheco, Snorre Sulheim, and Gregory L. Medlock. 2021. "Enhancing Microbiome Research through Genome-Scale Metabolic Modeling." *mSystems*, December, e0059921.
- Bergmann, Frank T., Richard Adams, Stuart Moodie, Jonathan Cooper, Mihai Glont, Martin Golebiewski, Michael Hucka, et al. 2014. "COMBINE Archive and OMEX Format: One File to Share All Information to Reproduce a Modeling Project." *BMC Bioinformatics* 15 (December): 369.
- Biggs, Bradley Walters, Brecht De Paepe, Christine Nicole S. Santos, Marjan De Mey, and Parayil Kumaran Ajikumar. 2014. "Multivariate Modular Metabolic Engineering for Pathway and Strain Optimization." *Current Opinion in Biotechnology* 29 (October): 156–62.
- Biller, Steven J., Allison Coe, Ana-Belen Martin-Cuadrado, and Sallie W. Chisholm. 2015. "Draft Genome Sequence of Alteromonas Macleodii Strain MIT1002, Isolated from an Enrichment Culture of the Marine Cyanobacterium Prochlorococcus." *Genome Announcements* 3 (4). <https://doi.org/10.1128/genomeA.00967-15>.
- Biller, Steven J., Paul M. Berube, Debbie Lindell, and Sallie W. Chisholm. 2015. "Prochlorococcus: The Structure and Function of Collective Diversity." *Nature Reviews. Microbiology* 13 (1): 13–27.
- Camborda, Stefano, Jan-Niklas Weder, and Nadine Töpfer. 2022. "CobraMod: A Pathway-Centric Curation Tool for Constraint-Based Metabolic Models." *Bioinformatics*, February. <https://doi.org/10.1093/bioinformatics/btac119>.
- Dukovski, Ilija, Djordje Bajić, Jeremy M. Chacón, Michael Quintin, Jean C. C. Vila, Snorre Sulheim, Alan R. Pacheco, et al. 2021. "A Metabolic Modeling Platform for the Computation of Microbial Ecosystems in Time and Space (COMETS)." *Nature Protocols*, October. <https://doi.org/10.1038/s41596-021-00593-3>.
- Fang, Xin, Colton J. Lloyd, and Bernhard O. Palsson. 2020. "Reconstructing Organisms in Silico: Genome-Scale Models and Their Emerging Applications." *Nature Reviews. Microbiology* 18 (12): 731–43.
- Faria, José P., Miguel Rocha, Isabel Rocha, and Christopher S. Henry. 2018. "Methods for Automated Genome-Scale Metabolic Model Reconstruction." *Biochemical Society Transactions* 46 (4): 931–36.
- Fennel, Katja, Jann Paul Mattern, Scott C. Doney, Laurent Bopp, Andrew M. Moore, Bin Wang, and Liuqian Yu. 2022. "Ocean Biogeochemical Modelling." *Nature Reviews Methods Primers* 2 (1): 1–21.

# References Continued

- Forchielli, Elena, Daniel Sher, and Daniel Segrè. 2022. "Metabolic Phenotyping of Marine Heterotrophs on Refactored Media Reveals Diverse Metabolic Adaptations and Lifestyle Strategies," January. <https://doi.org/10.1101/2022.01.07.475430>.
- Jiao, Nianzhi, Gerhard J. Herndl, Dennis A. Hansell, Ronald Benner, Gerhard Kattner, Steven W. Wilhelm, David L. Kirchman, et al. 2010. "Microbial Production of Recalcitrant Dissolved Organic Matter: Long-Term Carbon Storage in the Global Ocean." *Nature Reviews Microbiology* 8 (8): 593–99.
- Jiao, Nianzhi, and Qiang Zheng. 2011. "The Microbial Carbon Pump: From Genes to Ecosystems." *Applied and Environmental Microbiology* 77 (21): 7439–44.
- Lieven, Christian, Moritz E. Beber, Brett G. Olivier, Frank T. Bergmann, Meric Ataman, Parizad Babaei, Jennifer A. Bartell, et al. 2020. "MEMOTE for Standardized Genome-Scale Metabolic Model Testing." *Nature Biotechnology* 38 (3): 272–76.
- Lovato, T., D. Peano, M. Butenschön, S. Materia, D. Iovino, E. Scoccimarro, P. G. Fogli, et al. 2022. "CMIP6 Simulations with the CMCC Earth System Model (CMCC-ESM2)." *Journal of Advances in Modeling Earth Systems* 14 (3). <https://doi.org/10.1029/2021ms002814>.
- Machado, Daniel, Sergej Andrejev, Melanie Tramontano, and Kiran Raosaheb Patil. 2018. "Fast Automated Reconstruction of Genome-Scale Metabolic Models for Microbial Species and Communities." *Nucleic Acids Research* 46 (15): 7542–53.
- Manzoni, Stefano, Petr Čapek, Philipp Porada, Martin Thurner, Mattias Winterdahl, Christian Beer, Volker Brüchert, et al. 2018. "Reviews and Syntheses: Carbon Use Efficiency from Organisms to Ecosystems – Definitions, Theories, and Empirical Evidence." *Biogeosciences* 15 (19): 5929–49.
- Medlock, Gregory L., and Jason A. Papin. 2020. "Guiding the Refinement of Biochemical Knowledgebases with Ensembles of Metabolic Networks and Machine Learning." *Cell Systems* 10 (1): 109–19.e3.
- Moran, Mary Ann, Elizabeth B. Kujawinski, William F. Schroer, Shady A. Amin, Nicholas R. Bates, Erin M. Bertrand, Rogier Braakman, et al. 2022. "Microbial Metabolites in the Marine Carbon Cycle." *Nature Microbiology* 7 (4): 508–23.
- Ofaim, Shany, Snorre Sulheim, Eivind Almaas, Daniel Sher, and Daniel Segrè. 2021. "Dynamic Allocation of Carbon Storage and Nutrient-Dependent Exudation in a Revised Genome-Scale Model of Prochlorococcus." *Frontiers in Genetics* 12 (February): 586293.
- Orth, Jeffrey D., Ines Thiele, and Bernhard Ø. Palsson. 2010. "What Is Flux Balance Analysis?" *Nature Biotechnology* 28 (3): 245–48.
- Saifuddin, Mustafa, Jennifer M. Bhatnagar, Daniel Segrè, and Adrien C. Finzi. 2019. "Microbial Carbon Use Efficiency Predicted from Genome-Scale Metabolic Models." *Nature Communications* 10 (1): 3568.
- Sanz-Sáez, Isabel, Guillem Salazar, Pablo Sánchez, Elena Lara, Marta Royo-Llonch, Elisabet L. Sà, Teresa Lucena, et al. 2020. "Diversity and Distribution of Marine Heterotrophic Bacteria from a Large Culture Collection." *BMC Microbiology* 20 (1): 207.

# References Continued

- Sarkar, D., and C. D. Maranas. 2019. "Engineering Microbial Chemical Factories Using Metabolic Models." *BMC Chemical Engineering*. <https://link.springer.com/article/10.1186/s42480-019-0021-9>.
- Seaver, Samuel M. D., Filipe Liu, Qizhi Zhang, James Jeffryes, José P. Faria, Janaka N. Edirisinghe, Michael Mundy, et al. 2021. "The ModelSEED Biochemistry Database for the Integration of Metabolic Annotations and the Reconstruction, Comparison and Analysis of Metabolic Models for Plants, Fungi and Microbes." *Nucleic Acids Research* 49 (D1): D575–88.
- Seif, Yara, and Bernhard Ørn Palsson. 2021. "Path to Improving the Life Cycle and Quality of Genome-Scale Models of Metabolism." *Cell Systems* 12 (9): 842–59.
- Shaikh, Bilal, Gnaneswara Marupilla, Mike Wilson, Michael L. Blinov, Ion I. Moraru, and Jonathan R. Karr. 2021. "RunBioSimulations: An Extensible Web Application That Simulates a Wide Range of Computational Modeling Frameworks, Algorithms, and Formats." *Nucleic Acids Research* 49 (W1): W597–602.
- Sinsabaugh, Robert L., Stefano Manzoni, Daryl L. Moorhead, and Andreas Richter. 2013. "Carbon Use Efficiency of Microbial Communities: Stoichiometry, Methodology and Modelling." *Ecology Letters* 16 (7): 930–39.
- Stephanopoulos, Gregory. 2012. "Synthetic Biology and Metabolic Engineering." *ACS Synthetic Biology* 1 (11): 514–25.
- Wang, Hao, Simonas Marcišauskas, Benjamín J. Sánchez, Iván Domenzain, Daniel Hermansson, Rasmus Agren, Jens Nielsen, and Eduard J. Kerkhoven. 2018. "RAVEN 2.0: A Versatile Toolbox for Metabolic Network Reconstruction and a Case Study on Streptomyces Coelicolor." *PLoS Computational Biology* 14 (10): e1006541.
- Zoccarato, Luca, Daniel Sher, Takeshi Miki, Daniel Segrè, and Hans-Peter Grossart. 2022. "A Comparative Whole-Genome Approach Identifies Bacterial Traits for Marine Microbial Interactions." *Communications Biology* 5 (1): 1–13.