

THE STOICHIOMETRY OF NUTRIENT AND ENERGY TRANSFER:
FROM ORGANELLES TO ORGANISMS

by

Kristopher Allen Hunt

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

of

Doctor of Philosophy

in

Chemical Engineering

MONTANA STATE UNIVERSITY
Bozeman, Montana

November 2016

©COPYRIGHT

by

Kristopher Allen Hunt

2016

All Rights Reserved

TABLE OF CONTENTS

1. INTRODUCTION	1
1-1 Motivation	1
1-2 Algorithm Developments and Theoretical Framework	3
1-3 Autotroph-heterotroph Interactions	6
1-4 Fungal Hydrocarbon Production	7
1-5 Summary.....	8
2. COMPLETE ENUMERATION OF ELEMENTARY FLUX MODES THROUGH SCALABLE, DEMAND-BASED SUBNETWORK DEFINITION	10
2-1 Contribution of Authors and Co-Authors.....	10
2-2 Manuscript Information Page	11
2-3 Abstract.....	12
2-4 Introduction	13
2-5 Methods	18
2-5-1 Experimental and Computational Systems	18
2-5-2 EFM Concatenation and Comparison	20
2-5-3 Optimizing EFMTTool Performance	20
2-5-4 k-1 EFM Correction	21
2-6 Results	21
2-6-1 Selecting Reactions for Defining Subnetworks	21
2-6-2 Efficacy of Splitting	26
2-6-3 Iterative Splitting.....	26
2-6-4 Optimizing the Number of Reactions for Splitting per Iteration	28
2-6-5 Minimizing Intractable Subnetwork Runtime.....	31
2-6-6 Application of the Iterative Splitting Algorithm to a Genome-Scale Diatom Model	31
2-7 Discussion.....	34
2-8 Symbols	37
2-9 Acknowledgements	37
3. INTERPRETING AND DESIGNING MICROBIAL COMMUNITIES FOR BIOPROCESS APPLICATIONS, FROM COMPONENTS TO INTERACTIONS TO EMERGENT PROPERTIES.....	38
3-1 Contribution of Authors and Co-Authors.....	38
3-2 Manuscript Information Page	39
3-3 Abstract	40
3-4 Introduction	41

TABLE OF CONTENTS - CONTINUED

3-5	Definitions	43
3-5-1	Community Components.....	44
3-5-2	Interaction Outcomes	45
3-5-3	Interaction Mechanisms	46
3-5-4	Emergent Properties	49
3-6	Ecological Theories for Interpreting and Designing Communities.....	50
3-6-1	Maximum Power Principle.....	52
3-6-2	Resource Ratio Theory.....	54
3-6-3	Resource Allocation Theory: Pareto Surfaces and Metabolic Tradeoff Analysis	56
3-7	Case Studies of Communities with Interpretation.....	60
3-7-1	Multilevel Interactions in a Naturally Occurring Phototrophic Community	60
3-7-2	Anaerobic Syntrophy in Methanogenic Communities	63
3-7-3	Cross-feeding Chemostat Communities.....	65
3-8	Conclusions	67
3-9	Acknowledgements	69
4.	STOICHIOMETRIC MODELING OF ASSIMILATORY AND DISSIMILATORY BIOMASS UTILIZATION IN A MICROBIAL COMMUNITY	70
4-1	Contribution of Authors and Co-Authors.....	70
4-2	Manuscript Information Page.....	71
4-3	Abstract.....	72
4-4	Introduction	73
4-5	Results	77
4-5-1	Construction of <i>in silico</i> Model and Application of Ecological Theory	77
4-5-2	Optimal Utilization of Macromolecules.....	81
4-5-3	Selective, Optimal Utilization of Monomers from Macromolecules.....	85
4-5-4	Simultaneous Versus Sequential Metabolism of Monomer Pools.....	91
4-5-5	Identification of Suboptimal Metabolic Strategies.....	91
4-6	Discussion.....	93
4-7	Methods	97
4-7-1	Metabolic Model Construction.....	97
4-7-2	Heterotroph Biomass Production	98
4-7-3	Stoichiometric Analyses.....	99
4-8	Acknowledgements	100

TABLE OF CONTENTS - CONTINUED

5. MULTISCALE ANALYSIS OF AUTOTROPH-HETEROTROPH INTERACTIONS IN A HIGH-TEMPERATURE MICROBIAL COMMUNITY	101
5-1 Contribution of Authors and Co-Authors	101
5-2 Manuscript Information Page	102
5-3 Abstract	103
5-4 Introduction	104
5-5 Methods	108
5-5-1 Genome Analysis and Model Construction	108
5-5-2 Computational Packages and Analyses	109
5-6 Results	111
5-6-1 Enzyme-level Analysis of Chemolithoautotroph Electron Donors and Acceptors	111
5-6-2 Cellular-level Analysis of Chemolithoautotroph Growth	112
5-6-3 Cellular-level Analysis of Organoheterotroph Growth	113
5-6-4 Cellular- and Community-level Analysis of Oxygen Requirements for Growth	115
5-6-5 Community-level Resource Competition as a Basis for Community Structure	116
5-6-6 System-level Sensitivity Analysis of Electron Donors Using Multiscale, Hybrid EFMA+FBA Models	121
5-7 Discussion	124
5-8 Acknowledgements	129
6. THEORETICAL AND PRACTICAL LIMITATIONS OF HYDROCARBON PRODUCTION FOR A CELLULOLYTIC, ENDOPHYTIC FILAMENTOUS FUNGUS	130
6-1 Contribution of Authors and Co-Authors	130
6-2 Manuscript Information Page	131
6-3 Abstract	132
6-4 Highlights	133
6-5 Introduction	133
6-6 Material and Methods	137
6-6-1 Culture Medium and Conditions	137
6-6-2 Construction of Fungal Endophyte Metabolic Model	138
6-6-3 Stoichiometric Analyses	139
6-7 Results	140
6-7-1 Modeling of Carbon Uptake and Product Formation	140
6-7-2 <i>In silico</i> Cellular Energy	142

TABLE OF CONTENTS – CONTINUED

6-7-3 Theoretical Product Yields Based on The Metabolic Potential of <i>A. sarcooides</i>	144
6-7-4 <i>In silico</i> Pathway Modification Strategies	148
6-7-5 <i>In vitro</i> Growth and Oxygen Utilization	152
6-7-6 <i>In vitro</i> Hydrocarbon Production	152
6-8 Discussion.....	154
6-9 Conclusion.....	157
6-10 Acknowledgements	157
7. CONCLUSION.....	158
7-1 Stoichiometric Analysis.....	158
7-2 Applicability of Heterotrophy Analysis	160
7-3 Community Analysis	161
7-4 Hydrocarbon Production.....	162
7-5 Closing Statements	163
REFERENCES CITED.....	165
APPENDICES	190
APPENDIX A: Additional Publications.....	191
APPENDIX B: Supplemental Material for Chapter 2	194
APPENDIX C: Supplemental Material for Chapter 4	201
APPENDIX D: Supplemental Material for Chapter 5	209
APPENDIX E: Supplemental Material for Chapter 6	219

LIST OF TABLES

Table	Page
2-1. Metabolic models used in the presented study.	16
2-2. Criteria for removal of reactions from enforce/suppress sets.	25
3-1. Recent community-relevant review articles categorized by focus.	42
3-2. Representative microbial communities organized by system components.....	44
3-3. Representative microbial communities organized by system interaction mechanisms.	48
3-4. Ecological theory studies.	50
3-5. Microbial community <i>in silico</i> analysis studies.	56
4-1. Summary of simulations for 29 forms of biomass-derived dissolved organic carbon (DOC).	78
4-2. Resource costs of heterotroph <i>Geoarchaeum</i> str. OSPB biomass production from the 29 analyzed biomass-derived dissolved organic carbon (DOC) pools.	89
4-3. Resource costs of heterotroph <i>Geoarchaeum</i> str. OSPB cellular energy production from the 29 analyzed biomass-derived dissolved organic carbon (DOC) pools.	90
5-1. Values and descriptions of parameters used.	120
5-2. Equations used to define mass balances, growth rates, and Fe oxidation rates in terms of other system variables (Table 1).	121
6-1. Commonly reported products of <i>A. sarcoides</i> NRRL50072 by hypothesized pathway.	141
6-2. Theoretical optimal yields across functional groups.	149
6-3. Observed hydrocarbons under varying oxygen limitations.	154

LIST OF FIGURES

Figure	Page
1-1. Nutrient and energy flow across trophic levels in an ecosystem.....	2
1-2. Summary of metabolic model building, stoichiometric analysis, and application of ecological theory.	5
2-1. Overview of EFM enumeration using EFMTTool and the splitting approach.....	17
2-2. Work flow of reaction identification and iterative splitting approaches.	19
2-3. Network compression effect on reactions.	25
2-4. Subnetwork EFM count reduction with increasing number of reactions for splitting (R).....	25
2-5. Comparison of two network splitting approaches.	28
2-6. Optimization of reactions for splitting applied per iteration (r).	33
3-B1. Maximum Power Principle	51
3-B2. Resource Ratio Theory	53
3-B3. Resource Allocation Theory: Pareto/Tradeoff Analysis.....	55
3-1. Resource investment and economies of scale for enzymatic flux.	58
3-2. Phototrophic community interactions.....	62
3-3. Sulfate-reducing bacteria (SRBs) – methanogen anaerobic syntrophy.	64
3-4. Naturally occurring <i>E. coli</i> crossing-feeding communities.	66
4-1. Conceptual representation of high-temperature acidic Fe(III)-oxide microbial mats.	74
4-2. Summary of ecological resource utilization theory used for quantifying competitive physiologies during utilization of autotroph biomass by the heterotroph, <i>Geoarchaeum</i> str. OSPB.	79

LIST OF FIGURES – CONTINUED

Figure	Page
4-3. Analysis of macromolecule utilization by the heterotroph <i>Geoarchaeum</i> str. OSPB.	84
4-4. Analysis of RNA macromolecule and monomer utilization by the heterotroph <i>Geoarchaeum</i> str. OSPB.	87
4-5. Analysis of protein macromolecule and amino acid utilization by the heterotroph <i>Geoarchaeum</i> str. OSPB.	88
5-1. Conceptual representation of a multiscale metabolic interaction model between a primary autotroph (<i>M. yellowstonensis</i>) and heterotroph (<i>Geoarchaeum</i>) present in high-temperature (70-80°C) acidic Fe-oxide mats.	110
5-2. Electron donor and acceptor resource requirements to produce biomass.	113
5-3. Total community growth rates (A), specific growth rates (B), and net community growth rates (C) predicted as functions of oxygen flux and the relative population abundance of autotroph and heterotroph.	119
5-4. Sensitivity analysis of system behavior as a function of geochemical variation.	123
6-1. Overview of acetyl-CoA metabolism.	137
6-2. Overview of hydrocarbon producing pathways.	140
6-3. Ecologically competitive strategies for cellular energy generation.	143
6-4. Observed product concentrations and yields from <i>A. sarcooides</i> NRRL 50072.	143
6-5. Hydrocarbon production strategies that minimize carbon and oxygen consumption.	146
6-6. Degree of reduction of hydrocarbons as a function of carbons per molecule.	147

LIST OF FIGURES – CONTINUED

Figure	Page
6-7. The effect of cellobiose phosphorylase (CP) and cytosolic pyruvate dehydrogenase (PDHcyt) activity on excreted products from <i>A. sarcoides</i> NRRL 50072.	151
7-1. Potential regulatory impacts on resource costs.	161

ABSTRACT

All life requires the acquisition and transformation of nutrients and energy, driving processes from cellular nutrient flow to planetary biogeochemical cycling. However, the organisms and communities responsible for these processes are often uncultivable and too complex to observe directly and understand. Stoichiometric modeling, a systems biology approach, analyzes the reactions in an organism and incorporates data from multiple sources to extract biologically meaningful parameters, such as theoretical limits of conversion and yields of a metabolism. These limits and yields quantify relationships between organisms to establish governing principles, from resource requirements to community productivity as a function of population composition.

The presented work expanded the stoichiometric modeling algorithm and identified fundamental principles that govern nutrient and energy transfer associated with heterotrophy, community composition, and intracellular compartmentalization. A scalable routine capable of analyzing complex metabolic networks by dividing them into tractable subnetworks was demonstrated for a eukaryotic diatom. The metabolic model contained approximately two billion routes through the network and established an international benchmark for elementary flux mode analysis. Additionally, a heterotrophic archaeon was examined for the resource requirements while consuming 29 different forms of biomass derived dissolved organic carbon. These resource requirements and limitations establish a basis to analyze heterotrophy with regard to the limiting nutrient in a variety of systems. The resulting resource requirements of heterotrophy were incorporated into a community where an iron oxidizing autotroph was hypothesized to be the primary source of carbon and energy. Analysis of the community model and *in situ* measurements of iron and oxygen utilization indicated additional electron donors were required to account for the observed acquisition of nutrients in some communities. Finally, limits and resource requirements for fungal production of hydrocarbons were identified as a function of carbon and energy partitioning using simulated genetic modifications, providing context regarding endophytic production of bioactive molecules for host resistance as well as endophyte capacity as a petroleum producing alternative.

CHAPTER 1

INTRODUCTION

1-1 Motivation

Alternatives for petroleum-based fuels and chemicals are a common goal, regardless of arguments about the current state of energy and future directions (e.g. depletion of fossil fuels, emissions, or national security (5, 194, 205)). Biologically derived fuel and chemicals are attractive, because they integrate into contemporary infrastructure and photosynthesis derived material (e.g. green plants) is atmospherically carbon neutral. First generation biofuels utilize easily saccharified feed stocks, such as starch and sugar to primarily produce ethanol, a modification of techniques humans have employed for millennia to produce fermented drinks. However, ethanol is not feasible as a complete liquid fuel replacement and first generation fuel and chemical production competes with food production and accelerates depletion of soil nutrients (51, 219). These limitations led to development of second generation biofuels using inedible material composed largely of cellulose. Cellulose is recalcitrant and requires harsh and expensive pretreatments, which often produce inhibitory by-products, to increase biological availability and convert at industrially relevant rates (196). An alternative to adding enzymes or separate organisms for hydrolysis is to use a single organism that consumes cellulose and produces chemicals, referred to as direct conversion. Direct conversion reduces the need for pretreatment and combines hydrolysis and conversion, alleviating processing costs and production of inhibitory compounds.

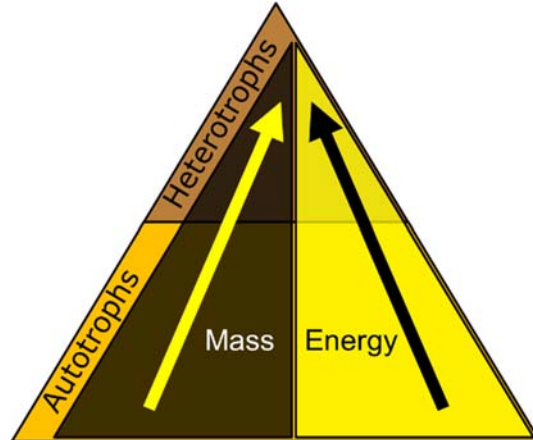


Figure 1-1. Nutrient and energy flow across trophic levels in an ecosystem. Ecosystems are governed by the supply of nutrients to each inhabiting population of organisms. In most ecosystems, this principle results in a governing community structure, where autotrophs, which fix carbon and energy from the environment, provide resources to heterotrophs through metabolite exchange or lysis.

Nutrient and energy transfer is fundamental to all scales of life, sustaining the hierarchy of trophic growth even in microbial communities (Figure 1-1). How nutrient recycling impacts ecosystem development has been a growing concern associated with both the use of bioprocesses to replace petroleum and the increasing demands of agriculture in general. Cellulosic feed stocks are expected to increase agricultural production and, therefore, rate of depletion of soil nutrients, leading to loss of soil fertility with no foreseeable replenishment on a human time scale (41, 69). Improper treatment of waste material from these processes leads to eutrophication by shedding large quantities of otherwise limiting nutrients into ecosystems and perturbing the balance of organisms that live there (20, 184). However, these ecosystems are often too complex to dissect interacting populations and track specific nutrients and energy flows. Generalized kinetic models have been used for decades to analyze processes like wastewater treatment (79), composting (155), and carbon flow in sediments (4).

However, these models rely on generalized kinetic parameters instead of using organism specific yields and rates, due to the complexity and difficulty of collecting such data. Examination of simpler ecosystems provides an understanding of ecological interactions and governing principles with which more complex and dynamic communities can be dissected.

1-2 Algorithm Developments and Theoretical Framework

Metabolic systems are complex and convoluted on a multitude of scales, from interacting pathways and metabolite pools to interacting organisms and trophic exchange. Stoichiometric analysis of metabolic systems simultaneously considers all modeled pathways in a system identifying optimization targets and ecologically relevant routes for nutrient utilization. There are two primary forms of stoichiometric analysis; both require an initial estimate of a metabolic network for the system of interest, commonly generated from genomic or other experimental data sets (121, 175, 195, 262). The metabolic network is represented by linear equations that describe cellular conversion of metabolites as either steady- or unsteady-state and reversible or irreversible (Figure 1-2). Flux balance analysis (FBA), the most computationally flexible form of stoichiometric analysis, uses this set of linear equations and bounds for reaction rates to obtain an optimal flux distribution that, for example, maximizes biomass production for a given substrate uptake rate (267). Elementary flux mode analysis (EFMA), another widely used form of stoichiometric analysis, uses the same set of linear equations to obtain the smallest set of genetically distinct, flux distributions known as elementary flux modes

(EFMs). The complete set of EFMs represents all possible steady-state phenotypes of the network through nonnegative linear combinations without cancellation (121, 140, 208).

EFMA has been the basis for both strain development and community analysis without knowledge of bounds to reaction rates (30, 242), making it more suitable for analysis of natural systems. However, as metabolic networks become increasing complex, they may contain exponentially more genetically distinct phenotypes, quickly making EFMA impractical without network simplification. Chapter 2 introduces an approach for dividing a metabolic network into tractable subnetworks and demonstrates an order of magnitude increase in the number of EFMs reported previously and sets the current internationally recognized record for EFMA.

The number of EFMs identified in a metabolic network can make manual analysis impractical without an approach to extract subsets with biological meaning. Application of ecological theory identifies subsets of EFMs expected to play a metabolic role, such as the EFMs that maximize the utility of a limiting resource. These EFMs are hypothesized to represent the majority of metabolisms in a population, providing a basis for predicting the environmental function of the population. Chapter 3 introduces several examples of ecological theory used to evaluate the effectiveness of an EFM under simulated carbon-, oxygen-, or enzyme-limitation, reducing the number of EFMs by orders of magnitude. An example of the application of stoichiometric modeling and ecological theory can be found in Figure 1-2. Additional review articles on stoichiometric modeling: (11, 65, 192, 262).

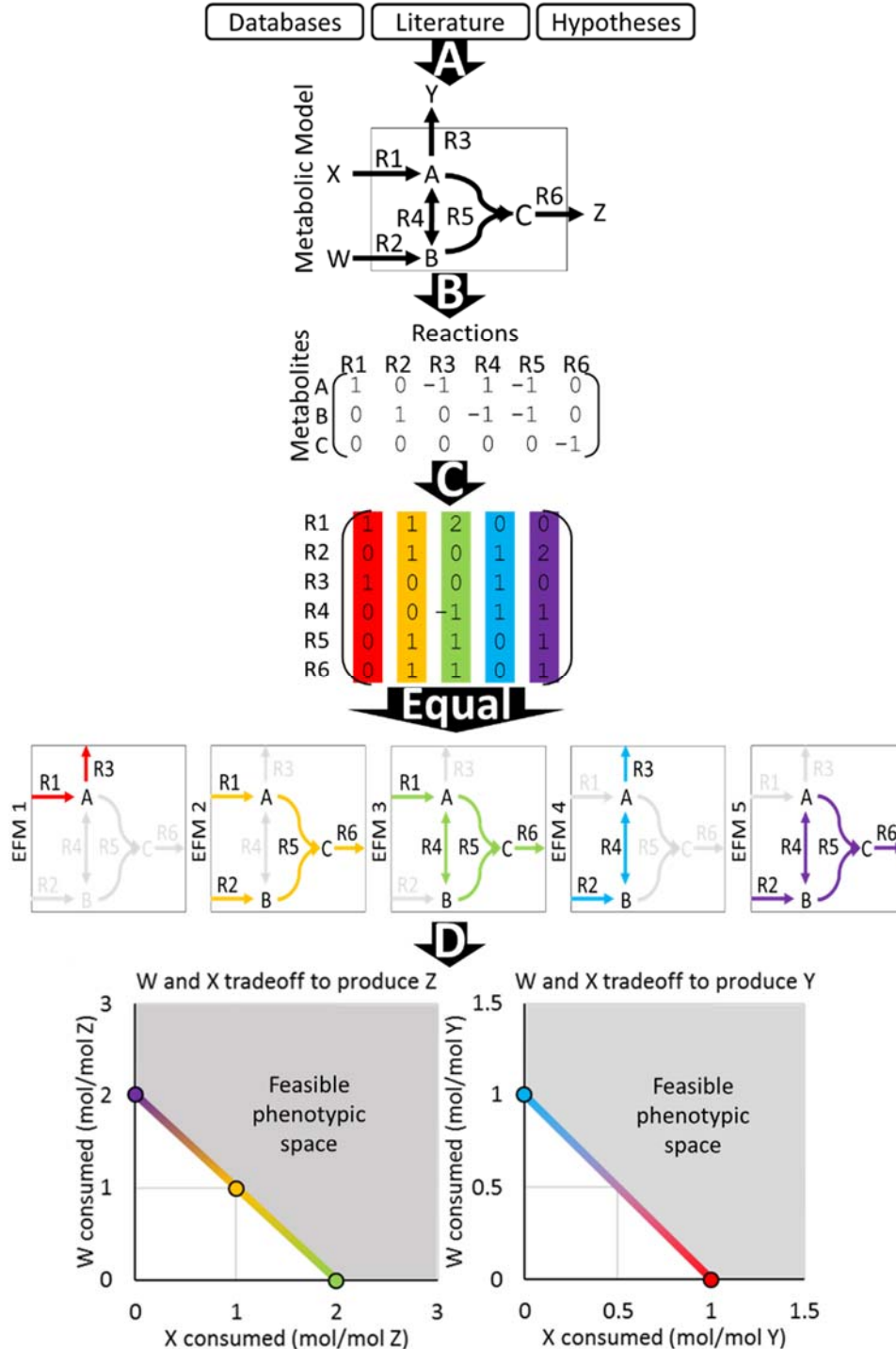


Figure 1-2. Summary of metabolic model building, stoichiometric analysis, and application of ecological theory. A) The metabolic model is a compilation of genomic annotations from databases and experimental yields reported in the literature. B) The model is represented as a series of reaction equations correlating the consumed and produced metabolites (i.e. R_i , where i corresponds to the reaction number and metabolites A, B and C; W, X, Y, and Z are sources and sinks). The reaction equations written as a

Figure 1-2 Continued

mathematical matrix, use positive coefficients for reaction products and negative coefficients for reaction substrates. C) Elementary flux mode (EFM) analysis finds all of the simplest, genetically distinct routes through the network (i.e. the five possible EFMs (color coded) for the sample metabolic model). D) Application of ecological theory to the analysis of these routes predicts competitive metabolic behavior. The theory assumes the reactions used to produce the desired product (e.g. biomass or cellular energy) will minimize the resource cost for the limiting nutrient (e.g. carbon, oxygen, or cellular volume). In the context of the depicted example, if W is limiting the production of Z, theory states the optimal strategy would be along the x-axis (left plot). Conversely, limitation of X to produce Z would be along the y-axis (left plot). Alternative products may favor alternative metabolic strategies (e.g. production of Y in the right plot). Modified from (96).

1-3 Autotroph-heterotroph Interactions

The iron oxidizing microbial mats in the hot springs of Yellowstone National Park provide an excellent model microbial community for the study of trophic interactions, specifically autotroph-heterotroph interactions. This community is primarily composed of 5 to 7 organisms, as determined from metagenomics and pyrosequencing analyses (101, 107, 128). The metagenomic analyses also provide a basis for reconstructing metabolic networks to represent the most abundant organisms present. Considering only populations that are expected to consume oxygen reduces the system even further to primarily two populations, an autotroph (e.g. *Metallosphaera yellowstonensis* MK1) and a heterotroph (e.g. *Geoarchaeum* str. OSPB), which represent the most abundant community members. The autotroph produces energy and biomass through chemolithoautotrophy, taking electrons from iron, respiring oxygen, and reducing carbon dioxide. The heterotroph produces energy and biomass through organoheterotrophy, consuming reduced carbon and respiring oxygen. Iron is provided by the source water

flowing over the mat and oxygen diffuses in from the air, as previously described (14). Carbon for the heterotroph in this system was shown to be 42-99% dissolved inorganic carbon in origin, suggesting major recycling of biomass to provide carbon to the non-autotrophic members of the community (107). This system is at a temperature that prohibits both photosynthesis and eukaryotic microbial predation (21, 23, 198), thereby removing biases associated with the production of other major macromolecules, such as chlorophyll. Therefore, viral lysis (23, 198) and senescence are two candidate mechanisms for carbon reallocation, providing an excellent basis for characterizing nutrient recycling. In chapter 4, this system is analyzed with respect to the heterotroph to ascertain the relative resource requirements to produce biomass and cellular energy from 29 different forms of biomass-derived dissolved organic carbon. In chapter 5, these resource costs are used in conjunction with observed iron and oxygen rates to characterize the metabolic activity of the community with respect to alternative electron and carbon sources.

1-4 Fungal Hydrocarbon Production

Ascocoryne sarcoides NRRL 50072, an endophytic fungus isolated from a Patagonian rainforest, produces a hydrocarbon mixture very similar to diesel and gasoline from a variety of substrates, including cellulose (150, 233). These hydrocarbons have been shown to be antibacterial and antifungal, and are hypothesized to confer protection to the host against infection (233). Hydrocarbon production by *A. sarcoides* has been primarily studied at the organism level (80, 150, 151, 233, 238), with limited pathway examination,

though the genome has been sequenced and annotated (76). Uncertainty of analytical methods and epigenetic variability presents a challenge for identifying all produced chemical species; however, a consensus of compounds reported across the literature provides a starting point for analysis. Analysis of the hydrocarbons produced establishes a basis for approximating pathways and understanding production limitations and resource costs. Chapter 6 examines these resource costs and metabolic conversion limitations using stoichiometric modeling to identify optimal hydrocarbons and the influence of culturing conditions on their production.

Production of energy rich biofuels requires cellular control over energy production and carbon storage. Eukaryotes maintain this control by partitioning acetyl-CoA between cellular compartments. Many hydrocarbons are produced through fatty acid synthesis, which occurs in the cytosol of fungi, though acetyl-CoA is generated in the mitochondria from pyruvate dehydrogenase. In chapter 6, the impact of this compartmentalization is examined by incorporating a cytosolic pyruvate dehydrogenase and cellobiose phosphorylase into the model. These genetic modifications have been shown to increase production efficiency and rate in other hydrocarbon producing organisms (18, 202). Building on such findings and identifying new control points will establish limits of hydrocarbon production potential.

1-5 Summary

The presented work developed tools expanding the applicability of stoichiometric modeling at the algorithm and data processing levels. Chapter 2 introduces an approach

for dividing a metabolic network into tractable subnetworks and demonstrates an order of magnitude increase in the number of EFMs reported previously. Chapter 3 introduces several examples of the application of ecological theory to evaluate the effectiveness of an EFM in a biologically meaningful context (e.g. carbon-, oxygen-, or enzyme-limitation) reducing the number of EFMs to consider for interpretation by orders of magnitude.

The application of stoichiometric analysis and ecological theory was used to establish theoretical metabolic conversion limitations and resource costs for ecologically and industrially relevant processes ranging from waste treatment to host defense and sustainable chemicals. In chapter 4, nutrient transfer across trophic levels through assimilatory and dissimilatory heterotrophy is analyzed to ascertain the relative resource costs and governing biological principles. In chapter 5, these resource costs and principles are used in conjunction with observed iron and oxygen rates to characterize the metabolic activity of the community with respect to alternative electron (e.g. arsenic and reduced sulfur species) and carbon (i.e. community fixed carbon versus landscape carbon) sources. In chapter 6, the effects of metabolite compartmentalization on hydrocarbon production are examined in a endophytic fungi. The partitioning is analyzed using a stoichiometric modeling approach to simulate alternative pathways resulting in varying levels of compartmentalization.

CHAPTER 2

COMPLETE ENUMERATION OF ELEMENTARY FLUX MODES THROUGH
SCALABLE, DEMAND-BASED SUBNETWORK DEFINITION

2-1 Contribution of Authors and Co-Authors

Author: Kristopher A. Hunt

Contributions: Assisted in concept development, helped with editing, generated figures, and wrote the paper.

Co-Author: James P. Folsom

Contributions: Assisted in concept development and helped with editing.

Co-Author: Reed L. Taffs

Contributions: Assisted in concept development and helped with editing.

Co-Author: Ross P. Carlson

Contributions: Assisted in concept development and helped with editing.

2-2 Manuscript Information Page

Kristopher A. Hunt, James P. Folsom, Reed L. Taffs, Ross P. Carlson
Bioinformatics

Status of Manuscript:

Prepared for submission to a peer-reviewed journal

Officially submitted to a peer-review journal

Accepted by a peer-reviewed journal

Published in a peer-reviewed journal

Published by Oxford University Press.

First published online February 3, 2014

COMPLETE ENUMERATION OF ELEMENTARY FLUX MODES THROUGH
SCALABLE, DEMAND-BASED SUBNETWORK DEFINITION

Kristopher A. Hunt^{1,2}, James P. Folsom^{1,2}, Reed L. Taffs^{1,2} and Ross P. Carlson^{1,2}

¹ Center for Biofilm Engineering, Montana State University, Bozeman, MT 59717-3980,
USA.

² Department of Chemical and Biological Engineering, Montana State University,
Bozeman, MT 59717-3920, USA.

2-3 Abstract

Motivation: Elementary flux mode analysis (EFMA) decomposes complex metabolic network models into tractable biochemical pathways, which have been used for rational design and analysis of metabolic and regulatory networks. However, application of EFMA has often been limited to targeted or simplified metabolic network representations due to computational demands of the method.

Results: Division of biological networks into subnetworks enables the complete enumeration of elementary flux modes (EFMs) for metabolic models of a broad range of complexities, including genome-scale. Here, subnetworks are defined using serial dichotomous suppression and enforcement of flux through model reactions. Rules for selecting appropriate reactions to generate subnetworks are proposed and tested; three test cases, including both prokaryotic and eukaryotic network models, verify the efficacy of these rules and demonstrate completeness and reproducibility of EFM enumeration.

Division of models into subnetworks is demand-based and automated; computationally intractable subnetworks are further divided until the entire solution space is enumerated.

To demonstrate the strategy's scalability, the splitting algorithm was implemented using an EFMA software package (EFMTool) and Windows PowerShell on a 50 node

Microsoft HPC cluster. Enumeration of the EFMs in a genome-scale metabolic model of a diatom, *Phaeodactylum tricornutum*, identified approximately two billion EFMs. The output represents an order of magnitude increase in EFMs computed compared to other published algorithms and demonstrates a scalable framework for EFMA of most systems.

2-4 Introduction

Extracting biologically meaningful information from the continuously expanding “omics” databases is often limited by bioinformatics tools. Stoichiometric modeling is one approach making advances in the quantification of metabolic phenotypes reflected by fluxomic data and inferred from metabolomic, proteomic, transcriptomic, and genomic data (195). Stoichiometric modeling assumes that during appropriate time scales, cellular biochemistry can be approximated as a steady state process with respect to intracellular enzyme and metabolite concentrations. This assumption simplifies the differential equations describing cellular mass balances, and, when coupled with reaction directionality and flux magnitude constraints, permits the calculation of physiologically reasonable flux distributions. There are two major stoichiometric modeling approaches with numerous sub-variations (121, 175, 195, 262). One stoichiometric modeling approach, elementary flux mode analysis (EFMA), enumerates all genetically distinct, indecomposable flux distributions in a metabolic network (216). These elementary flux modes (EFMs) are the minimal set of stoichiometrically balanced flux distributions representable as unique binary vectors based on participation of every model reaction. This set must also represent every steady state flux distribution in a given network

through nonnegative combinations without cancellation (121, 140, 208). Cancellation refers to removal of reaction participation through combination of fluxes with equal but opposite magnitudes (i.e. genetic distinction). Another stoichiometric modeling approach, flux balance analysis (FBA), utilizes linear programming to identify optimal flux distributions through a metabolic network. Various objective functions are possible, including the maximization of biomass yield from a substrate (267). These two stoichiometric modeling methods have been used for decades to direct metabolic engineering efforts and predict physiological and ecological behaviors (30, 33, 66, 138, 195, 242, 262, 268).

Both stoichiometric modeling approaches have strengths and limitations. A major strength of EFMA is enumeration of the entire solution space in an unbiased manner. FBA generates only a limited number of distinct solutions using a directed objective function (175, 195, 262). However, the directed FBA approach reduces computational costs relative to EFMA, permitting examination of complex metabolic reconstructions, commonly referred to as genome-scale models. Complex (e.g. genome-scale and microbial community) metabolic networks have often been intractable using EFMA (120, 252), despite algorithmic advances (117, 214, 216, 253, 264).

Approaches that circumvent the problem of enumerating all EFMs have been explored. One such approach, convex basis analysis (i.e. extreme pathway analysis (208)), identifies a subset of EFMs that still reproduce any feasible steady state flux distribution, but lack the biological interpretability of the complete EFM set (121, 140). Efforts have also been made to enumerate subsets or patterns of elementary flux vectors

that allow genome-scale investigations. These efforts include generating the conversion cone (a simpler projection of the solution space considering external fluxes only) (263, 265), enumerating only EFMs containing the largest number of zero fluxes (68), enumerating a set of EFMs that can explain a given flux distribution (104), applying regulation to minimize computation (112), and identifying elementary flux patterns, which are EFMs for subsystems that consider the constraints imposed by the parent network (113, 215). Finally, computationally intractable networks have been divided into manageable pieces based on biochemical knowledge (209, 270), metabolite connectivity (i.e. the number of reactions involving a particular metabolite) (217, 270), path length between reactions (145), examination of the nullspace of the stoichiometric matrix (190), and random sampling of the solution space in order to calculate eigenvectors which can be rotated to produce distinct reaction sets (7).

For many applications, the ideal approach is an unbiased investigation of the entire solution space, although this scenario was previously out of computational reach for complex networks (120, 140). It has been proposed that otherwise infeasible EFM enumerations can be performed through the dissection of metabolic networks into simpler subnetworks via suppression or enforcement of reaction fluxes, a technique referred to as splitting (117). Suppression of a reaction defines a subnetwork which does not contain the reaction while enforcement of a reaction defines a subnetwork that excludes EFMs that do not utilize the reaction, both of which reduce the computational burden (e.g. Figure 2-1). Complete coverage of a metabolic network's solution space without overlap is ensured at each split by paired suppression and enforcement of a single reaction. An

additional example of suppressing and enforcing can be found in (109) which used the EFMA algorithm, Elmo-Comp. Unfortunately, Elmo-Comp lacked enumeration completeness, as shown through a disagreement between the presented and published EFM counts (108, 109). The primary objective of the current study is to successfully implement a scalable, demand-based splitting technique for complete enumeration of all network EFMs. The objective is achieved through i) the development of rational rules for effectively splitting networks, ii) the successful use of a network splitting algorithm that continuously divides networks until they are computationally tractable, and iii) the use of a workstation-based computational cluster to distribute EFM enumeration of genome-scale models. The presented work demonstrates efficacy across network models containing EFM counts spanning four orders of magnitude. This algorithm is the first successful implementation of the splitting technique and establishes a scalable framework for EFMA of most metabolic models.

Table 2-1. Metabolic models used in the presented study

Organism	Model	Reactions /Metabolites	EFMs
<i>S. cerevisiae</i>	YEAST1 ^a	78/62	1,515,315 ^d
<i>S. cerevisiae</i>	YEAST2 ^a	83/63	68,868,602 ^d
<i>E. coli</i>	ECOLI ^b	95/94	226,269,020 ^d
<i>P. tricornutum</i>	DIATOM ^c	318/335	1,934,729,551 ^e

^a (109) ^b (112) ^c This Work

^d EFMs validated by comparing split and unsplit network results.

^e 14 reactions used for iterative splitting. Some subnetworks were recalculated varying the number of reactions for splitting to validate results.

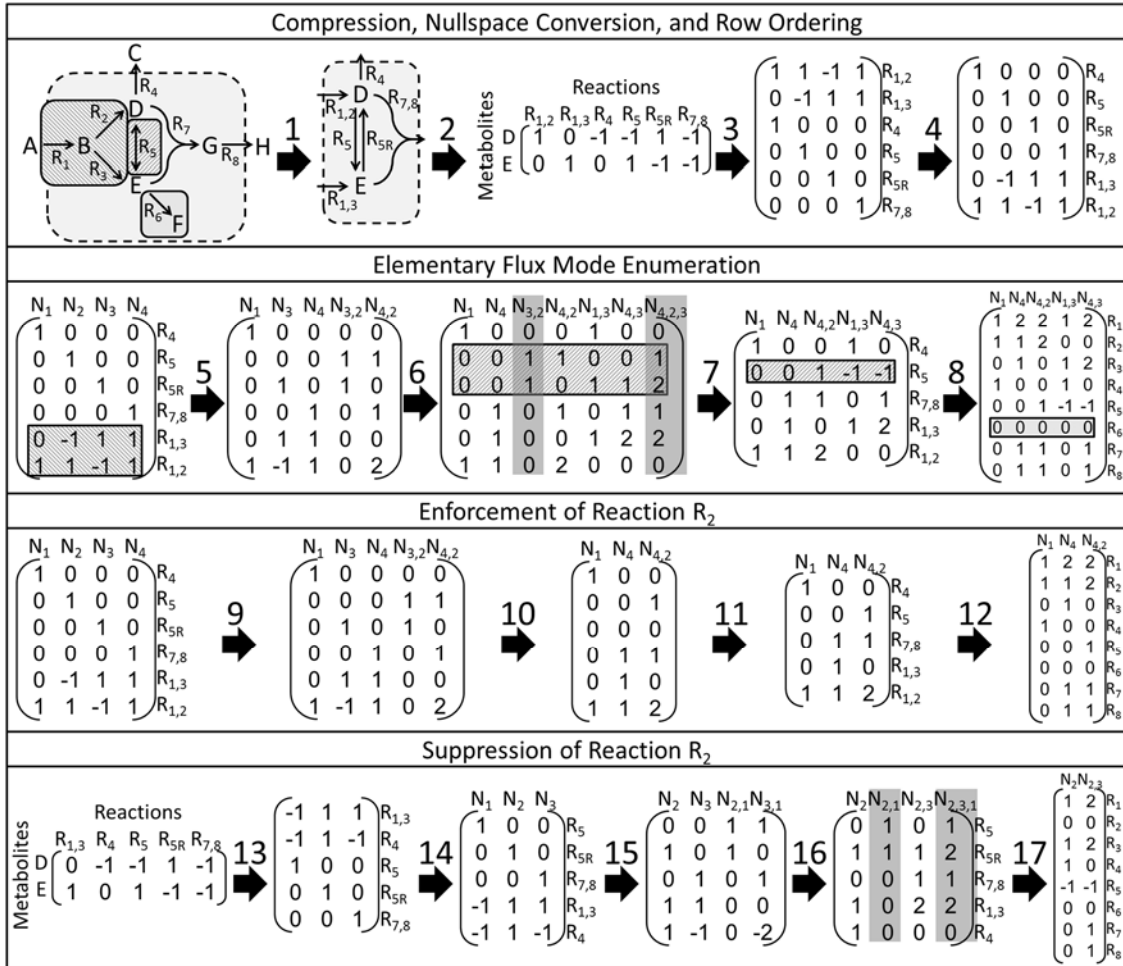





Figure 2-1. Overview of elementary flux mode enumeration using EFMTTool and the splitting approach. Step 1) The model is compressed removing dead-end reactions (R_6), separating reversible reactions into forward and reverse reactions (R_5 becomes R_5 and R_{5R}), and combining reaction sets which always occur together into a single reaction (R_7 and R_8 combined into $R_{7,8}$, removing metabolite G). Step 2) The compressed network, represented by a stoichiometric matrix with reactions in columns and metabolites in rows, is solved for the nullspace basis vectors (i.e. the nullspace kernel) (Step 3). External metabolites are not constrained by conservation relationships and are therefore not considered in the stoichiometric matrix. Step 4) The kernel rows are reordered to minimize memory usage and computational runtime. Steps 5-6) The EFMTTool algorithm then applies directionality constraints to each reaction sequentially; vectors containing negative fluxes in the reaction being constrained are replaced by all nonnegative linear combinations resulting in a zero flux through the reaction. The negative flux in reaction $R_{1,3}$ at N_2 is replaced by $N_{3,2}$ (N_3+N_2) and $N_{4,2}$ (N_4+N_2) while the negative flux in reaction $R_{1,2}$ at N_3 is replaced by $N_{1,3}$ (N_1+N_3), $N_{4,3}$ (N_4+N_3), and $N_{4,2,3}$ ($N_{4,2}+N_3$). EFMTTool bit-masks each row of values after the directionality constraint is applied, improving memory usage (not shown). Step 7) $N_{3,2}$ is removed during recompilation of reversible reactions and $N_{4,2,3}$ is removed because at least $q-m-1$ zeros per flux vector are

Figure 2-1 Continued

required by the degrees of freedom constraint, where q and m are the number of reactions and metabolites, respectively(75). Step 8) The remaining vectors are elementary flux modes and decompressed. This procedure can be modified to enforce (Steps 9-12) or suppress (Steps 13-17) reactions as needed. Enforced reactions are moved to the end of the algorithm and directionality constraints are applied as described above; however, when the enforced reaction is reached all positive flux vectors have already been enumerated (Step 9), therefore flux vectors which violate the directionality constraint or do not use the reaction are discarded (Step 10). Suppressed reactions result in a new stoichiometric matrix which is then processed as described above. The highlighted examples of a combined reaction in , reversible reaction in , and dead-end reaction in  are to be avoided for use in splitting.

2-5 Methods

2-5-1 Experimental and Computational Systems

Development and analysis of the presented splitting techniques was performed using three previously published metabolic models described in Tables 2-1 with additional detail available in Table 2-S1. The test models, chosen to demonstrate efficacy over a range of model complexity, included a prokaryotic (ECOLI), a simplified eukaryotic (YEAST1), and a multicompartment eukaryotic model (YEAST2) (Supplemental Material). These test models were calculated as both unsplit and split networks to validate the compiled subnetwork results. Unsplit networks were computed using the software package EFMTTool on a Windows 7 machine with a maximum configuration of 120 GB of RAM and 2 Intel Xeon processors (X5690). Implementation of EFMTTool (253) version 4.7.1 (www.csb.ethz.ch/tools/efmtool) used documented options described in Table 2-S2. The basic splitting and iterative splitting algorithms were written in Windows PowerShell v2.0 for implementation with Microsoft HPC, clustering software which supports workstation-based clusters. The computational cluster included 50 workstation

nodes, which ranged from 4 GB to 120 GB of RAM per node. Result compilations and analyses were performed using MATLAB on machines with 32 GB of RAM.

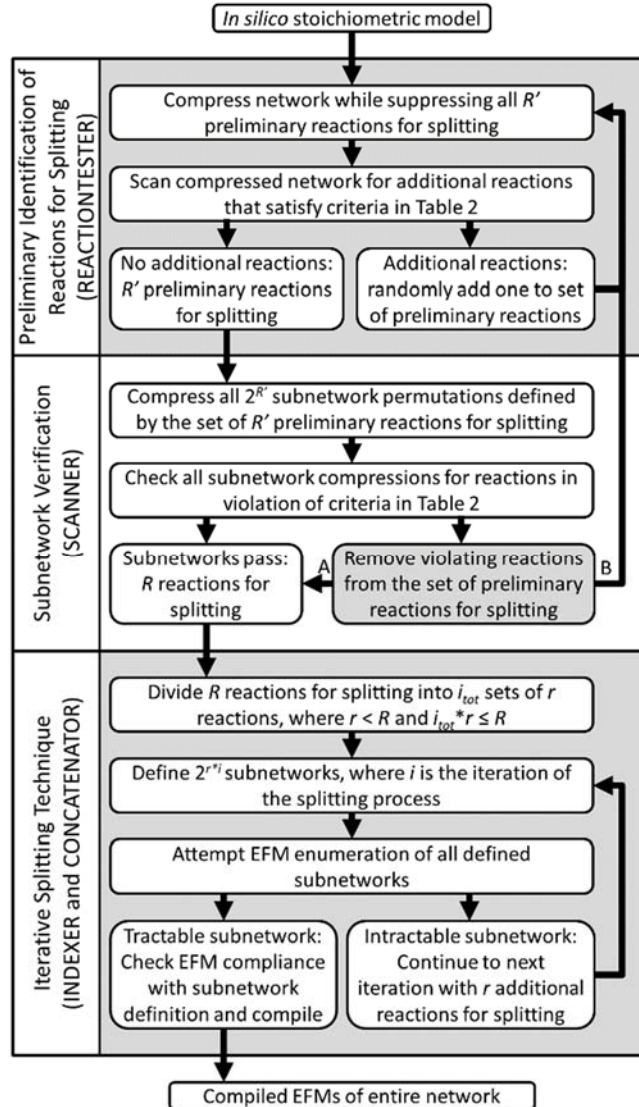


Figure 2-2. Work flow of reaction identification and iterative splitting approaches. Initially, the list of identified preliminary reactions for splitting is empty but increases by one reaction per loop through preliminary reaction identification. The final quantity of reactions for splitting was adjusted at the grey box by choosing route A if a predefined R was achieved or route B if more reactions for splitting were required. Depicted processes describe the pseudo-code or MATLAB functions listed in parentheses. See the Symbols section for additional variable definitions.

2-5-2 EFM Concatenation and Comparison

Subnetwork results produced by the splitting algorithm were screened for consistency with subnetwork definitions and concatenated into a single output set (as in Figure 2-2). EFMTool identifies futile cycles based on the number of nonzero fluxes in a flux vector, removes them from processing, and adds them back at the end of processing regardless of subnetwork definition. This necessitated screening all EFMs to verify that they did not violate the definition of their subnetwork. The screening process required minimal computational effort. For example, the CONCATENATOR script checked and compiled two billion EFMs in approximately one day; most of the elapsed time involved reading, decompressing, compressing, and writing the results, which occupy approximately one TB of hard drive space.

Network results were compared across different algorithms and splitting configurations by converting all EFMs to a binary representation based on reaction participation and identifying EFMs not shared by both result sets being compared. This comparison was done using standard MATLAB functions. Binary representation reduced rounding effects associated with normalization and allowed for comparison of all test models presented.

2-5-3 Optimizing EFMTool Performance

EFMTool's utilization of RAM was inspected and optimized using the management application, Jconsole, and Java options. Briefly, Java 1.6.x uses a generational memory construct that assigns new objects to one memory space while moving older objects to a separate predefined memory space. This is usually an efficient use of RAM, as most Java applications do not retain many old objects, and memory management is easier when

newer objects are kept together. However, EFMTTool stores each EFM as an object that stays in RAM for the duration of the calculation, eventually becoming an old object. EFMTTool filled the memory allocated for older objects which resulted in “out of RAM” errors, despite having only used 25-33% of the RAM assigned to Java. Thus, the default Java options limit EFMTTool execution to smaller networks. This practice was corrected and performance optimized using command line switches described in Table 2-S2.

2-5-4 $k-1$ EFM Correction

During early attempts at EFM enumeration using the EFMTTool functionality to enforce reactions, EFMs were absent from the resulting solution set. The missing EFMs contained $k-1$ enforced reactions, where k is the number of nonzero fluxes in the EFM. This issue was corrected through a modification in EFMTTool code to consider enforced reactions during the subroutine which removes futile cycles (Supplemental Material).

2-6 Results

2-6-1 Selecting Reactions for Defining Subnetworks

Appropriate reaction selection for subnetwork definition is essential for effective splitting. Selected reactions must permit complete enumeration of EFMs while distributing calculations over subnetworks. Initially, random reactions for splitting were screened using a brute force method to confirm functionality for splitting; analysis of the results identified four major categories of reactions to avoid. First, reactions for splitting should not produce or consume metabolites that are not consumed or produced by other

reactions (i.e. dead-end reactions, R_6 in Figure 2-1). Dead-end reactions result in empty subnetworks when enforced and do not reduce computations when suppressed.

Second, reactions for splitting should not include reversible reactions. EFMTTool divides reversible reactions into two separate irreversible reactions, a forward and reverse reaction, during EFM enumeration and then recombines the fluxes before reporting the final EFMs (R_5 in Figure 2-1). If used for splitting, the forward and reverse reactions are enforced or suppressed concurrently; EFMs with flux in only one of these reactions would be lost. This property makes reversible reactions unacceptable for splitting.

Network compression improves memory efficiency by combining reactions which must operate together into a single compressed reaction (e.g. R_1 and R_2 are combined into $R_{1,2}$ in Figure 2-1; (75)); however, compression can cause problems during splitting, resulting in a third category of reactions to avoid for splitting. Two adverse scenarios can result from compression. In a branching series, a single reaction can be combined separately with multiple reactions to form multiple compressed reactions. Using that single reaction as a reaction for splitting prevents certain subnetwork permutations because multiple independent compressed reactions are enforced or suppressed simultaneously (e.g. R_1 in Figures 2-1, 2-3, and 2-S1). In a linear series, a set of reactions which must occur together combine to form a single compressed reaction (e.g. R_7 and R_8 in Figure 2-1 and 2-S1). Using multiple reactions that define a compressed linear series is ineffective for the same reason as enforcing dead-end reactions (R_7 and R_8 in Figure 2-1 and S1). However, one reaction from a linear series or the branch reactions in a branched series may be effective for splitting (e.g. R_2 , R_3 , and R_7 or R_8 in Figure 2-1, 2-3, and 2-

S1). While more reactions may be available for subnetwork definition when compression is not used, compression greatly reduces computational burden during the enumeration process, vastly outweighing the benefit of additional potential reactions for splitting (data not shown; (117)).

A fourth type of reaction to avoid has poorly scaled coefficients that can lead to numerical instability. This often occurs with biomass synthesis associated reactions. For instance, accounting for the synthesis of a low abundance vitamin (e.g. B₁₂) and a common amino acid (e.g. glycine) can lead to a greater than 10³-fold difference between the reaction's largest and smallest coefficients. Such a model could not be analyzed with EFMTTool due to numerical instabilities in the stoichiometric matrix. This issue can be exacerbated during the network splitting process when potential EFMs with very small, but nonzero, fluxes in enforced reactions are removed based on the software's computational definition of zero. The removed EFMs are not present when the reaction is suppressed by definition, and are therefore missed in the results. One method explored to circumvent this problem used the higher accuracy arithmetic of fixed-point notation instead of the default floating-point notation (double to fractional arithmetic in EFMTTool nomenclature). The use of fixed-point notation permitted complete enumeration of EFMs for some test networks but introduced a substantial computational cost with a minimal increase in the number of usable reactions for splitting (data not shown). Increasing the precision of the zero definition for floating-point notation was also examined but not found to consistently improve EFM enumeration. While there are multiple methods to circumvent this problem, this study avoided using reactions containing poorly scaled

coefficients for splitting, defined as coefficients with a difference greater than 10 within a single reaction. This approach was sufficient for the three test models.

A list of effective reactions for splitting was identified through application of the rules summarized in Table 2-2 and confirmed through additional checks (i.e. REACTIONTESTER and SCANNER, respectively, in Figure 2-2). The rules were applied iteratively as follows: i) reactions were identified which did not violate the rules in Table 2-2 given the current network compression, ii) one such reaction was selected randomly and added to a list with R' preliminary reactions for splitting, and iii) the subnetwork was defined by suppressing all R' preliminary reactions for splitting, providing the basis for the next iteration of reaction selection. This was repeated until all remaining reactions violated the rules. Repeatedly adjusting the compression basis was necessary because as reactions are suppressed, they change the model compression through production of dead-ends and new sets of combined reactions. Then, to ensure effective splitting, all $2^{R'}$ subnetwork permutations defined using enforce/suppress combinations of the preliminary reactions for splitting were compressed and examined for rule violations. Any reactions which violated the rules under any enforce/suppress permutation were removed, resulting in a final list of R usable reactions for splitting. Reaction validation for all subnetworks also allowed for screening of reactions which create incompatibility issues with EFMTTool. These processes for preliminary reaction identification and subnetwork permutation checking, depicted in Figure 2-2, are described in the pseudo-code found in the Supplemental Material (REACTIONTESTER function and SCANNER subroutine, respectively).

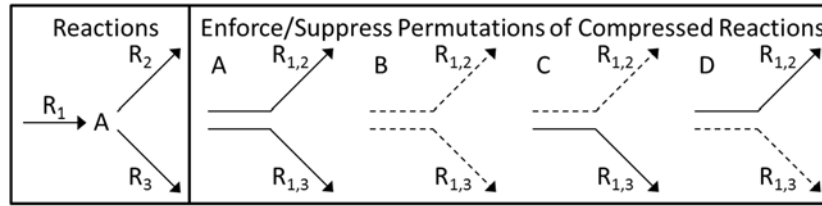


Figure 2-3. Network compression effect on reactions. Network compression results in two reactions: $R_{1,2}$ is $R_1 + R_2$, and $R_{1,3}$ is $R_1 + R_3$. Enforcing (solid lines)/suppressing (dotted lines) reaction R_1 generates only two (A and B) of the four reaction combinations needed to fully describe possible reaction participation (A, B, C, and D).

Table 2-2. Criteria for removal of reactions from enforce/suppress sets

Rule	Reaction Type	Issue
1	Dead-end	Subnetworks cannot evenly distribute EFMs
2	Reversible	Misses EFMs containing flux in either but not both directions
3	Combined ^a	May not allow for all reaction combinations
4	Poorly scaled coefficients	Poorly conditioned matrix and zero definition may miss EFMs (threshold difference of 10)

All reaction combinations must be tested to confirm that an EFM is not represented in the enforce/suppress set and that the subnetwork definition and compression does not rearrange reactions causing violation of the stated rules.

^a: See section 3.1 for detailed description of combined reactions to avoid

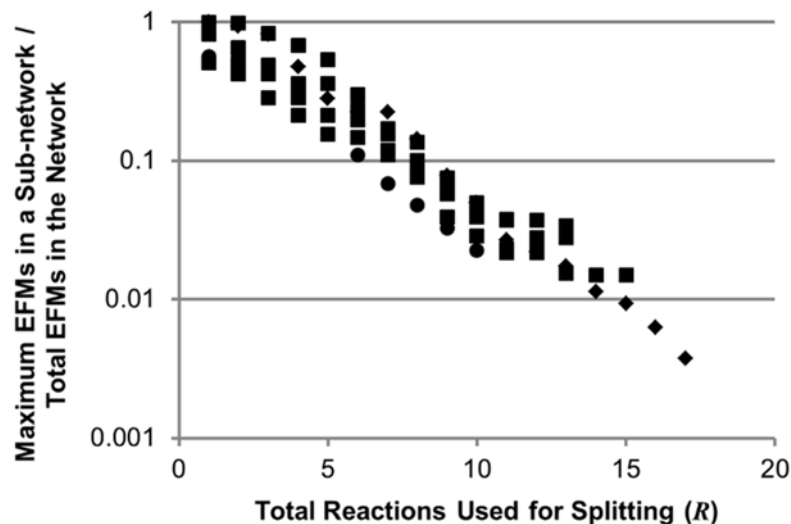


Figure 2-4. Subnetwork EFM count reduction with increasing number of reactions for splitting (R). The test models YEAST1 (squares), YEAST2 (diamonds), and ECOLI (circles) were analyzed in 77 separate configurations; every configuration resulted in complete enumeration of the solution space.

2-6-2 Efficacy of Splitting

The number and choice of reactions used to define subnetworks are important control points for splitting. To quantify splitting effectiveness based on these considerations, subnetworks were defined using a set of R reactions, EFMs for all subnetworks were enumerated, and the largest number of EFMs in any subnetwork was used as a basis to assess splitting efficacy. A single set of reactions for splitting was identified for each of the ECOLI and YEAST2 models while four distinct sets were identified for the YEAST1 model using the procedure detailed in the preceding section (pseudo-code presented in Supplemental Material). The four distinct sets of reactions for splitting exist due to the sequential random identification of reactions for splitting (e.g. most reactions could work in isolation, but only some sets of reactions work together). R was varied within each set of reactions for splitting by removing reactions without replacement to form smaller sets. The linear trend on a semi-log plot (Figure 2-4) suggests an exponential decay of maximum EFMs per subnetwork with increasing R . There were no deviations in the number or identity of the EFMs enumerated between the 77 splitting-based enumerations and the respective unsplit enumerations. This analysis also demonstrates that the reaction selection rules are functional over a broad range of R .

2-6-3 Iterative Splitting

Identifying the optimal number of reactions for splitting a model is a major theoretical and practical challenge. There are two competing effects with increasing magnitude of R which are i) the rate of new subnetwork definition and ii) the rate of reduction in subnetwork complexity. Splitting shifts the computational burden from memory

limitation to CPU limitation (Figure 2-5A, B). The number of subnetworks grows with 2^R when using basic splitting, resulting in a prohibitive number of files for larger values of R . Figure 2-5A demonstrates a case of basic splitting in grey where R is 4.

A demand-based network splitting algorithm can be implemented which minimizes the total number of analyzed subnetworks without sacrificing complete coverage of the solution. This strategy is referred to as iterative splitting (Figure 2-5B). The procedure initially uses a set of r reactions to define subnetworks, where r is less than the total number of reactions needed to complete the network (R_{min}). EFMs of calculable subnetworks are enumerated and subnetworks that are assessed to be intractable (e.g. EFM enumeration fails due to memory limitation or terminates via the algorithm discussed in section 2-6-5) form branches that are further divided using an additional set of r reactions. The process continues until each branch completes or the list of reactions for splitting is exhausted. Computational savings are realized when branches complete at intermediate iterations instead of the final iteration. Figure 2-5B illustrates a case where r is 1, R_{min} is 4 and the total number of iterations, i_{tot} , is 4. In this example, one branch completes per iteration resulting in a 2-fold decrease in total subnetworks attempted compared to basic splitting (Figure 2-5A, B). The YEAST2 model in Table 2-1 demonstrated a 76-fold decrease in number of subnetworks calculated using iterative splitting ($r=2$, $R_{min}=14$, and $i_{tot}=7$) as compared to basic splitting with the same 14 reactions (reactions for splitting: R4, R5, R9, R12, R14, R21, R27, R33, R37, R40, R46, R53, R58, and R61). R_{min} is a function of computational resources and was increased by

decreasing available memory for computation (e.g. 6718 MB heap size for the YEAST2 example).

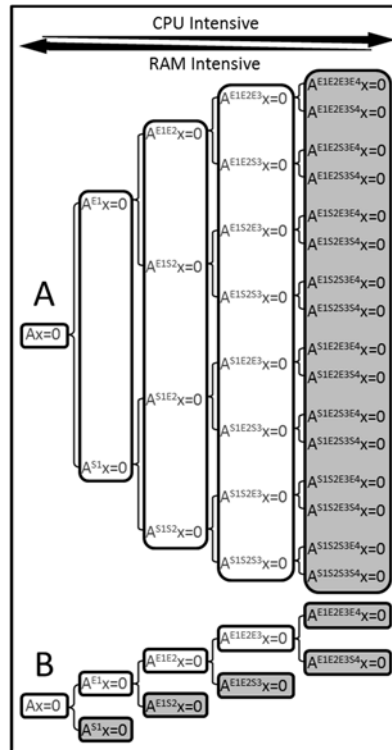


Figure 2-5. Comparison of two network splitting approaches. The basic splitting approach (A) uses combinations of enforced/suppressed reactions to create calculable subnetworks where the number of subnetworks increases exponentially with the number of reactions for splitting (117). The iterative splitting approach (B) only generates subnetworks when the previous ones are intractable (white) until all EFMs for all subnetworks in the branch can be enumerated (grey) (117). Both approaches shift the burden of computation from a RAM-limited problem to a CPU-limited problem as the subnetworks become smaller but greater in number. The example uses $R=4$ reactions for splitting with each iteration considering one additional reaction (i.e. $r=1$) for a total of $i=4$ iterations. Superscripts refer to the enforced (E) or suppressed (S) reactions in the labeled subnetworks.

2-6-4 Optimizing the Number of Reactions for Splitting per Iteration (r)

Iterative splitting has substantial computational benefits when branches complete before the final iteration; however, the computational costs of intractable subnetworks

may overwhelm this benefit. This tradeoff provides a basis for optimization, the number of reactions for splitting (r) applied per iteration, i . Increasing r defines a larger number of more constrained subnetworks per iteration which have an increased likelihood of completion compared to subnetworks defined using smaller r . This approach decreases failed enumeration attempts but increases the number of attempted subnetworks per iteration. Two scenarios bound the total possible number of subnetworks analyzed during iterative splitting. The upper limit occurs when all subnetworks are intractable until the last iteration at which point they all complete (Figure 2-5A, $r=1$). This scenario represents basic splitting when r equals R . The lower limit occurs when each step of iterative splitting produces only one intractable subnetwork until R reactions are used, at which point all subnetworks complete and the entire solution space for the whole network is covered (Figure 2-5B, $r=1$).

Optimization based on limiting the number of intractable subnetworks enables prediction of a generalized ideal r without *a priori* network specific information. The equation for the total number of subnetworks (N) is:

$$N = 2^r (f + 1) \quad (1)$$

where f is the number of intractable subnetworks. Equation 1 does not include the original unsplit network. As $r \rightarrow R_{min}$, $f \rightarrow 0$, therefore, basic splitting with at least the number of reactions required (R_{min}) eliminates time spent on intractable subnetworks. However, this neglects the benefit of early branch completion, reducing the total number of attempted subnetworks and therefore the total CPU time (Figure 2-5A, B).

Assigning a theoretical, relative CPU time for tractable and intractable subnetworks permits evaluation of effective values of r with respect to CPU time. For this analysis, a theoretical model is considered which requires one hour of CPU time to enumerate the EFMs from the unsplit network, $t_{unsplit}$. An approximate relation between $t_{unsplit}$ and the average CPU time to enumerate the EFMs of a subnetwork of the same network was determined empirically using the test models in Table 2-1. For the subnetwork defined by $r*i$ reactions, the CPU time for EFM enumeration ($t_{r,i}$) follows the relationship:

$$t_{r,i} = \frac{t_{unsplit}}{r * i} \quad (2)$$

(data not shown). For analysis purposes, the CPU time for subnetworks with failed EFM enumeration attempts, t_{fail} , was bracketed between 10-fold longer and 10-fold shorter than $t_{unsplit}$. 10-fold longer was used as an upper limit because Java garbage collection will run until it uses 90% of the total CPU time before crashing, inflating the run time ~10-fold. A 10-fold shorter run time was used for the lower limit because networks with an explosion in the number of potential EFMs may quickly fill available memory with objects which cannot be cleared (data not shown). Combining Equation 2 and the bounds for t_{fail} with the number of subnetworks defined by Equation 1, total CPU time boundaries were estimated as a function of r and normalized by $t_{unsplit}$ (Figure 2-6). Simulations considered three theoretical models with $R_{min} = 10, 20, \text{ and } 30$ and all combinations of i_{tot} and r producing a respective R_{min} . Increasing r reduces the maximum potential CPU time investment independent of R ; however, the time minimum increases significantly with r after approximately five reactions (Figure 2-6). The minimum time increases due to the large number of subnetworks and the associated redundant calculations during reaction

enforcement. Considering both the minimum relative CPU time for the lower t_{fail} limit scenario and the exponential decrease in relative CPU time for the upper t_{fail} limit scenario, an r in the range of 4-8 reactions appears optimal without *a priori* knowledge (Figure 2-6). The analysis considers total relative CPU time; implementation on computational clusters would reduce real-world time because of parallel computations.

2-6-5 Minimizing Intractable Subnetwork Runtime

Intractable subnetworks can have a high CPU time cost; therefore, large run time savings can be obtained by minimizing CPU time spent on intractable subnetworks. An empirical detection strategy was applied to identify EFM enumeration attempts that were not likely to complete; those attempts were then terminated before failure and the subnetwork was further divided. Applying the intractable subnetwork detection strategy resulted in substantial time savings for the models in this study (data not shown). Details of the prediction algorithm and utilized time thresholds can be found in the Supplemental Material.

2-6-6 Application of the Iterative Splitting Algorithm to a Genome-Scale Diatom Model

EFMA of a genome-scale metabolic model for the diatom *Phaeodactylum tricornerutum* Pt1 was conducted to demonstrate the iterative splitting algorithm. This model was reconstructed using the extensive knowledge base for *P. tricornerutum* Pt1 including: a finished genome (152), and literature-based, manual analyses (e.g. (60, 131)). The DIATOM model considers 680 genes that were manually compressed into 318 reactions with 335 metabolites. Reactions and metabolites were partitioned into five distinct

physical compartments based on analysis of organelle signal peptides (56) and biochemical studies (e.g. (241)). The metabolic model with genomic information, stoichiometries, biomass requirements, and a graphical representation can be found in the Supplemental Material.

The workflow for successful application of the iterative splitting algorithm (i.e. Figure 2-2) to the DIATOM model can be described as follows: a set of 14 reactions for splitting ($R=14$) was identified by calling the REACTIONTESTER function (Supplemental Material). The reactions were divided into two sets of seven reactions ($r=7$ and $i_{tot}=2$) based on the theoretical optimal values of r (Figure 2-6) and the complete use of the set of reactions for splitting. The INDEXER function submitted EFM enumeration jobs to the HPC computational cluster and, upon completion of the first iteration, scanned the results for intractable subnetworks to further divide with the next 7 reactions in the second iteration. The PowerShell script used is available in the Supplemental Material. EFMA of the DIATOM model resulted in 1,934,729,551 EFMs; this count was validated by reanalyzing random model subnetworks using different sets of reactions for splitting. The output exceeds the largest reported for EFMA to date by approximately one order of magnitude (112).

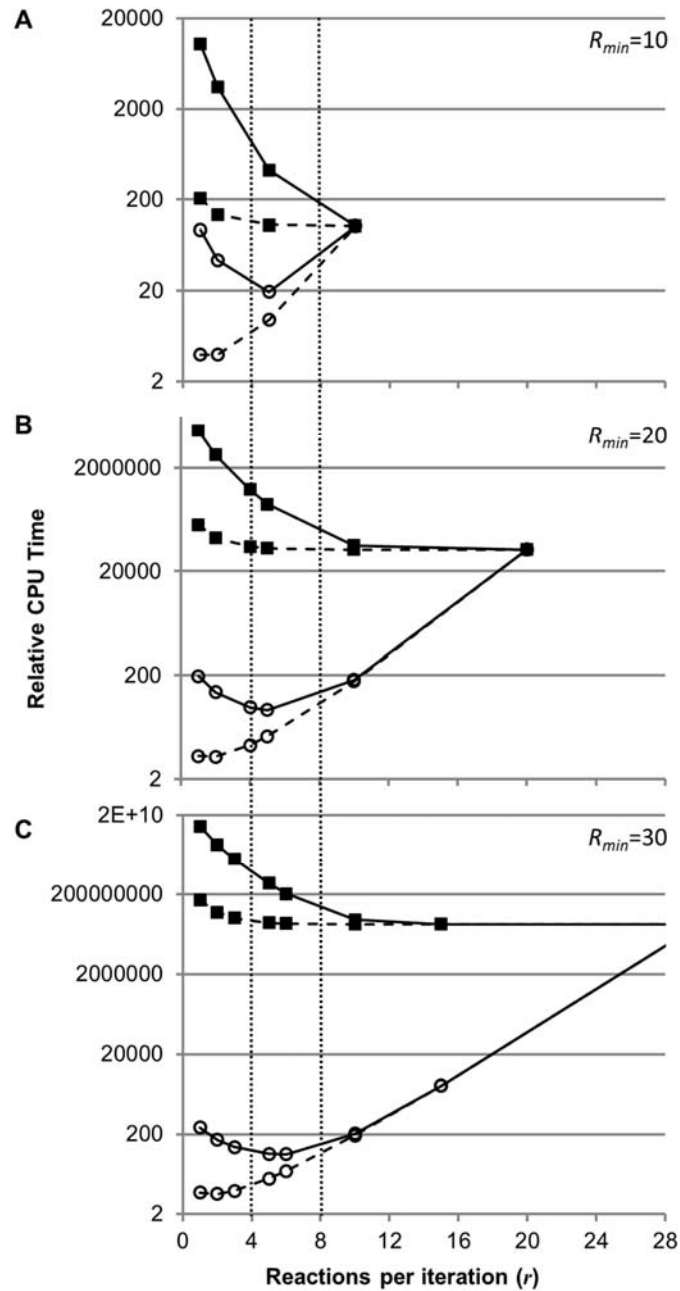


Figure 2-6. Optimization of reactions for splitting applied per iteration (r). Relative CPU time for theoretical networks requiring 10 (A), 20 (B), and 30 (C) total reactions for splitting to complete, assuming that one subnetwork is intractable per iteration (open circles) or all subnetworks are intractable until the last iteration (closed squares). Time for intractable subnetworks was varied between 10-fold increase (solid lines) and 10-fold decrease (dashed lines) relative to an unsplit model. Relative CPU time is the summation of time to complete all subnetworks normalized by time to complete the whole, unsplit network. Vertical dotted lines designate the optimal working range based on the minimum (open circles with solid lines) and the exponential decrease (closed squares), given no prior knowledge of network splitting behavior.

2-7 Discussion

The complete and reproducible EFMA of metabolic networks that were previously computationally intractable was demonstrated here through the definition of subnetworks based on serial enforcement and suppression of reaction fluxes. This is in contrast to other recent attempts to parallelize EFM enumeration by splitting which were found to be incomplete at times (108, 109). The presented basic and iterative splitting algorithms are expected to be fully compatible with that group's recent efforts to implement distributed EFMA computing with shared memory (110), offering additional computing power for recalcitrant subnetworks.

When using EFMTool, the reactions for splitting cannot contain a subset representing all nonzero flux reactions in an EFM; enforcing such a subset causes an error. The ability to identify the error by examining all subnetworks at the level of compression indicates the error is caused during preprocessing. Additionally, manual application of the nullspace algorithm did not yield an explanation for the error. The error was reproducibly avoided by identifying the subnetworks with the error and eliminating at least one of the reactions in the EFM from the set of reactions for splitting. This solution is algorithmically inconvenient since one of these reactions must be removed from the set of reactions for splitting and the new combination of reactions reanalyzed (Figure 2-2). However, subnetworks enforcing an EFM cannot contain more than the defined EFM since an EFM cannot be represented within another EFM (due to the decomposability constraint). The EFM in an enforced reaction set can be identified through EFMA of the problematic reaction set, a negligible computational investment. By confirming

enforcement of an EFM to be the problem, rechecking of reactions and empty subnetworks may be avoidable through early completion of the branch enforcing a complete EFM and addition of the identified EFM to the solution set.

Selection criteria were established for identifying effective reactions for splitting (Table 2-2). Some of these criteria could be incorporated into a new EFMA algorithm. Examples include i) changing the representation of reversible reactions and ii) changing when in the algorithm networks are compressed. Recording reversible reactions in the original network model as separate forward and reverse reactions would allow the algorithm to define enforce/suppress combinations involving currently unusable reactions thereby eliminating the need for the reversibility rule. Compressing networks before applying the enforce/suppress subroutine would prevent the network from reconfiguring with subnetwork definition, removing the need to for subnetwork verification. However, this operational order may reduce effectiveness of subnetwork definition since some suppressed reactions result in additional combinable reactions, as shown by the creation of a dead-end reaction when suppressing R_8 in Figure 2-1 and 2-S1 (117). While the likelihood of reaction rearrangement as a function of network complexity should be examined, the costs of accounting for network compression effects were not prohibitive to the analysis for presented models.

The presented splitting approaches are subject to the availability of reactions for splitting. For instance, a hypothetical model composed exclusively of reversible reactions or a model where every reaction had poorly scaled coefficients could not be analyzed with the presented approach. Fortunately, such models are not typically biologically

relevant. While all chemical reactions are theoretically reversible, many are effectively irreversible under physiologically relevant concentrations and temperatures. Additionally, the scaling of reaction coefficients for a metabolic model is dictated by product/substrate stoichiometries of the enzymes, which are often in a range that do not cause numerical instability. A more practical limit would be a model which does not have enough reactions satisfying the rules in Table 2-2 to sufficiently divide the network into tractable subnetworks. While such a model is theoretically possible, it is hypothesized that with computational resource improvements, most models will be tractable.

The basic splitting and iterative splitting algorithms allow for distributed computation of models through division into subnetworks. Microsoft HPC software permitted enumeration of the nearly two billion EFMs in the DIATOM model in less than one month using academic computer laboratories during idle hours. While the total CPU time is unavoidable, the real time required to run large models can be reduced substantially based on the number and computational power of nodes. For instance, the most complex subnetwork in the DIATOM model took approximately two days of run time; therefore, given appropriate resources, the EFMA could have been completed in two days if the other subnetworks were concurrently run on additional nodes. This highlights the lower limits of computational time for EFMA of genome-scale models given current computer capabilities. Using the splitting algorithm and sufficient computational resources, the presented study substantially expands the applicability of EFMA.

2-8 Symbols

f	Number of intractable subnetworks
i	Iteration in the iterative splitting process
i_{tot}	Number of iterations in the iterative splitting process
k	Number of reactions with nonzero coefficients within an EFM
m	Number of metabolites in the compressed model
N	Number of subnetworks attempted
q	Number of reactions in the compressed model
R	Number of reactions used for subnetwork definition during the splitting process
R_{min}	Minimum number of reactions needed for complete subnetwork enumeration during the splitting process
R'	Number of preliminary reactions for splitting (reactions that have not been verified at the subnetwork level)
$R_{\#}$	Reaction #, where # identifies a specific reaction from Figures 2-1 or 2-3
r	Number of reactions used for subnetwork definition at each iteration
t_{fail}	Average CPU time spent on intractable subnetworks
$t_{r,i}$	Average completion time for subnetwork defined by $r*i$ reactions
$t_{unsplit}$	Time for unsplit network to complete

2-9 Acknowledgements

The authors would like to thank Jeffrey J. Heys for helpful discussions.

Funding: This work was supported by the Centers of Biomedical Research Excellence (COBRE) Center for Analysis of Cellular Mechanisms and Systems Biology, NIH [P20RR024237 to J.P.F. and R.L.T.], Integrative Graduate Education and Research Traineeship (IGERT) in Geobiological Systems, NSF [DGE 0654336 to K.A.H. and R.L.T.], and the Emerging Frontiers in Research and Innovation (EFRI), NSF [0937613 to K.A.H.]. The diatom reconstruction was supported by the Air Force Office of Scientific Research (AFOSR grant FA9550-09-1-0243 to R.P.C.).

CHAPTER 3

INTERPRETING AND DESIGNING MICROBIAL COMMUNITIES FOR
BIOPROCESS APPLICATIONS, FROM COMPONENTS TO INTERACTIONS TO
EMERGENT PROPERTIES

3-1 Contribution of Authors and Co-Authors

Author: Ashley E. Beck

Contributions: Assisted in concept development, helped with editing, generated figures, and wrote the paper.

Author: Kristopher A. Hunt

Contributions: Assisted in concept development, helped with editing, generated figures, and wrote the paper.

Co-Author: Hans C. Bernstein

Contributions: Helped with editing.

Co-Author: Ross P. Carlson

Contributions: Assisted in concept development, helped with editing, generated figures, and wrote the paper.

3-2 Manuscript Information Page

Ashley E. Beck, Kristopher A. Hunt, Hans C. Bernstein, Ross P. Carlson
Biotechnology for Biofuel Production and Optimization

Status of Manuscript:

Prepared for submission to a peer-reviewed journal

Officially submitted to a peer-review journal

Accepted by a peer-reviewed journal

Published in a peer-reviewed journal

Published by Elsevier B.V.

First published online February 19, 2016

INTERPRETING AND DESIGNING MICROBIAL COMMUNITIES FOR
BIOPROCESS APPLICATIONS, FROM COMPONENTS TO INTERACTIONS TO
EMERGENT PROPERTIES

Ashley E. Beck^{1,2‡}, Kristopher A. Hunt^{1,3‡}, Hans C. Bernstein⁴, Ross P. Carlson^{1,3}

‡contributed equally

1. Center for Biofilm Engineering, Montana State University, Bozeman, MT 59717
2. Department of Microbiology and Immunology, Montana State University, Bozeman, MT 59717
3. Department of Chemical and Biological Engineering, Montana State University, Bozeman, MT 59717
4. Chemical and Biological Signature Science, Pacific Northwest National Laboratory, Richland, WA 99352

3-3 Abstract

Interest in microbial communities for bioprocessing has surged in recent years based on the potential to optimize multiple tasks simultaneously and to enhance process productivity and stability. The presence and magnitude of these desirable properties often result from interactions between community members. The importance of interactions has gained interest due to improving ‘omics techniques, polymicrobial culturing approaches, and computational methods, which have made the systems-level analysis of interacting components more tractable. This review defines and categorizes natural and engineered system components, interactions, and emergent properties, as well as presents three ecological theories relevant to microbial communities. Case studies are interpreted to illustrate components, interactions, emergent properties, and theoretical context. A general foundation is laid to facilitate interpretation of current systems and to aid the design of next generation bioprocesses.

3-4 Introduction

Most natural microbial ecosystems are the result of millions of years of natural selection in spatially and temporally dynamic landscapes. These ecosystems possess effective, highly evolved functions and are almost exclusively organized as polymicrobial communities. The study of natural and engineered microbial communities has benefited from new technologies such as increased resolution and throughput of 'omics measurements, development of new genetic systems for establishing model organisms, advanced cell isolation methods such as flow cytometry and cell sorting (167), comprehensive databases such as KBase (<http://kbase.us>), and ever-growing computational power for performing *in silico* experiments such as community-scale metabolic network modeling (148, 242). Interest in using polymicrobial systems for applied bioprocesses stems largely from an effort to mimic and ultimately control the beneficial emergent properties that are often observed in natural ecosystems. These attractive attributes of communities have a potential to enable superior catalytic function compared to traditional monocultures based on simultaneous optimization of multiple tasks, increased productivity, and greater stability.

While the study of applied microbial communities is growing in popularity, the appreciation and use of communities for societal purposes is not new. In fact, applications of microbial communities date back at least 5,000 years to early food preservation via lactic acid-producing bacterial communities used for yogurt production (244), and evidence suggests directed biogas production was practiced in Assyria and China going back at least 3,000 years (19). The significance of polymicrobial systems

was also observed and tested by the earliest pioneers of modern microbiology. In 1683, Antoine van Leuwenhoek recorded observations of morphologically distinct ‘animalcules’ collected from oral scrapings and, in 1877, Louis Pasteur tested antagonistic interactions between medically relevant bacteria (50, 72, 124, 183).

Numerous reviews reiterate, with overlapping content, the relevance and use of polymicrobial systems for human health, biological circuit synthesis, microbial computing, synthetic ecology, biomass degradation, and a myriad of other biological applications (Table 3-1). This review provides a generalized discussion of ecological foundations useful in understanding polymicrobial systems and highlights classical theories applicable to future microbial community engineering efforts. No attempt is made here to exhaustively outline every relevant study in this rapidly growing field; rather, case studies are selected to epitomize ecological themes and design motifs within the context of the current and future state of polymicrobial bioprocessing.

Table 3-1. Recent community-relevant review articles categorized by focus.

Focus	Reference	Focus	Reference
Artificial symbiosis	(163)	Food fermentations	(226)
Biodegradation	(160)	Industrial bioprocessing	(204)
Bioenergy, biomaterials	(176)	Microbiome	(256)
Bioengineering	(25)	Mining	(26)
Bioprocess	(6, 106)	Polymicrobial infections	(168)
Bioprocess, experimental and theoretical	(15)	Synthetic biology	(77)
Biotechnology, algal biofuels	(181)	Synthetic biology, bioprocess	(229)
Cellulose degradation, bioprocess	(287)	Synthetic communities	(82)
Cyanobacteria/microalgae and bacteria	(239)	Synthetic ecosystems	(201)
Ecological interactions, symbiosis	(98)	Viral interactions	(199)
Engineering, bioprocess	(222)		

3-5 Definitions

The study of microbial communities has expanded from traditional biological disciplines to include a wide cross-section of applied sciences. This broad expansion has resulted in the merging of concepts and terms from classical disciplines including biochemistry, computational biology, ecology, engineering, genetics, and microbiology. These fields use vernacular with varying connotations; hence, a list of terms is provided with the definitions used here to facilitate a unified discussion.

A microbiological community is a collection of populations that may be comprised of prokaryotic, eukaryotic, or viral components; definable interactions are not a requirement. A microbial community is a microbiological community limited to prokaryotic and eukaryotic populations with no explicit accounting for viruses and no constraints on interactions. However, the ubiquity of viruses suggests that the vast majority of microbial species hosts viruses, and the existence of virus-free microbial communities is therefore thought to be exceedingly rare. A microbial community exhibiting positive interactions is referred to as a consortium (plural consortia) (147); some uses of the term are more generic, referring to any interacting community. The origin of community populations can be used to further categorize the system. A natural community is defined as a collection of wild-type populations that have interacted in nature on evolutionary timescales. An artificial community is a collection of unmodified populations that have been assembled through manual intervention and are not thought to interact in natural habitats. A synthetic community is defined as a collection of genetically modified populations, whereas a semisynthetic community is a collection of

populations with at least one wild-type population and at least one genetically modified population. A review of recent case studies exemplifying these different organizations can be found in (15).

3-5-1 Community Components

Communities are ensembles of populations that serve as the system components.

Populations can be classified by phylogenetic and/or phenotypic distinction.

Phylogenetically distinct community populations are different species that range across the domains of microbial life and can include viruses. Table 3-2 provides examples of components comprising interacting communities ranging from virus-virus systems to archaeon-eukaryote systems. Virus-virus and host-virus interactions are often overlooked in reviews of applied microbial communities despite their common distribution across most natural systems and their potential to impact engineered systems. These community interactions are not limited to pathogenicity and can represent a spectrum of interaction outcomes, including mutually beneficial effects as reviewed in (199).

Table 3-2. Representative microbial communities organized by system components.

Community members	Description	Reference
Virus-Virus	Enhanced pathogenicity	(141)
Virus-Bacterium	Enhanced infectivity of human cells	(134)
Virus-Bacterium	Lysis of competitors	(206)
Virus-Eukaryote	Suppression of additional virus infection (HIV)	(59, 257)
Virus-Eukaryote	Prevention of diabetes	(174)
Virus-Eukaryote	Thermal stress tolerance	(153)
Bacterium-Bacterium	Cyanobacteria differentiating to fix nitrogen	(281)
Archaeon-Bacterium	Enhanced function via hydrogen exchange and methanogenesis	(158)
Bacterium-Eukaryote	Cellulose conversion to value-added biochemicals	(161)
Archaeon-Eukaryote	Enhanced cellulose degradation via hydrogen exchange and methanogenesis	(40)

Phenotypically distinct populations are not differentiable on the species level but exhibit separate expression patterns that often lead to niche differentiation. An ecological niche is the function or place of an organism within an ecosystem. The cyanobacterial strategy of cellular differentiation into specialized nitrogen-fixing heterocysts and vegetative cells is a well-studied example of a single-species population functioning as a community (133). Another example is synthetic or adapted communities of *Escherichia coli* strains with different substrate specificities which can lead to niche partitioning in biofilms (e.g., oxic or anoxic separation) (16, 54, 189, 200). Biofilms are polymer-encapsulated microbial populations attached to biological or abiotic surfaces (84). Mass transfer typically limits the availability of resources within a biofilm, resulting in spatial heterogeneity which, in turn, leads to phenotypic differentiation (84, 189, 236).

3-5-2 Interaction Outcomes

Community interactions can be classified as one of six outcomes. These interactions may be unidirectional, bidirectional, or higher order; they can also be obligatory or facultative. Each of the interaction outcome categories presented here is symbolically summarized for a two-population community using ‘+’ to indicate a population benefit, ‘-’ to indicate an adverse effect, and ‘0’ to indicate no effect. In practice, these interaction outcomes are often observed in communities comprised of more than two populations (93).

1) Mutualism (+/+): both populations benefit from the interaction(s). Syntrophy is a specific instance of mutualism associated with the cross-feeding of essential resources between populations (158). 2) Commensalism (+/0): one population benefits from the

interaction(s) while the other is unaffected. 3) Ammensalism (-/0): one population is adversely affected by the interaction(s) while the other is unaffected. 4)

Competition/antagonism (-/-): both populations are negatively affected by the interaction(s), which can be due to functional redundancy or antagonistic interactions. 5)

Parasitism (+/-): one population benefits and one population is adversely affected by the interaction(s). 6) Neutralism (0/0): neither population is affected by the interaction(s);

alternatively, there is no interaction between populations. Synthetic ecology, another scientific field witnessing rapid growth, often attempts to assemble tractable, albeit constrained, systems to test these interaction outcomes (201).

3-5-3 Interaction Mechanisms

The mechanisms that mediate known interactions can be divided into three generalized categories: metabolite exchange, physical interaction, and environmental modification.

These mechanisms are typically combinatorial and interrelated within natural and engineered microbial communities. Metabolite exchange is any transfer of material and/or chemical energy between community populations and requires some combination of active transport (e.g., ABC-type transporters) and/or passive transport (e.g., diffusion).

Exchanges of anabolic resources, resources used in biosynthetic processes (e.g., amino acids, nucleotides, vitamins, cofactors, and siderophores), are widely distributed among natural systems and represent targets for metabolically engineered polymicrobial systems (Table 3-3) (186, 277). Exchanged metabolites can also serve as catabolic resources, resources used to produce cellular energy. Examples include electron donors or acceptors

(e.g., hydrogen or oxygen, respectively), which are used in biologically mediated redox reactions to facilitate production of energetic molecules like NAD(P)H or ATP.

Exchanged material can serve functions other than anabolic or catabolic roles, such as modulating the behavior of community populations. Quorum sensing, a process by which organisms secrete and receive specific soluble metabolites that act as regulatory signals, is associated with a wide variety of natural multicellular functions, including coordinated biofilm formation, microbial pathogenicity, and culture bioluminescence (275). In addition, quorum sensing has been proposed as a means of interrogating local environments, permitting a feedback mechanism for regulating phenotype (e.g., resource-intensive strategies such as enzyme secretion) (46, 90). Modulating community population activity can also be realized through competitive strategies like the exchange of inhibitory or toxic metabolites (61, 91). Secretion of antibiotics and antimicrobial peptides is widely distributed and provides both offensive and defensive mechanisms within some communities.

Direct physical interactions between community populations can dictate community structure and function. These interactions are a hallmark mechanism of biofilms (179, 236); this physical association can be controlled or constrained by metabolite exchange (24). Evidence suggests that direct physical contact between community populations can also be used for the transfer of electrons through materials such as cytochrome-rich extracellular structures, which have been characterized as outer membrane and periplasmic extensions (188). Similarly, filamentous cable bacteria in microbial sediment

communities permit electron transfer across cm-length scales (185). Direct inter-population electron transfer remains an active area of investigation.

Environmental modification is a mechanism of interaction in which a community population influences the local environment and thereby alters the niche(s) that other populations can inhabit. Modulation often occurs when one population consumes a chemical species (e.g., organic acids, oxygen, hydrogen sulfide) that inhibits other populations (16, 36, 132). Table 3-3 provides examples of natural and engineered systems organized by mechanism of interaction.

Table 3-3. Representative microbial communities organized by interaction mechanisms.

Interaction mechanism	Community	Description	Reference
Anabolic metabolite exchange	Filamentous anoxygenic phototrophs / cyanobacteria	Vitamin	(197)
Anabolic metabolite exchange	Uncultured marine bacteria	Siderophore	(44)
Anabolic and catabolic metabolite exchange	Sulfate reducing bacteria / methanogens	Alanine and hydrogen / formate	(273)
Catabolic metabolite exchange	Sulfate reducing bacteria / methanogens	Hydrogen / formate	(225)
Catabolic metabolite exchange	Colon microbiota	Acetate cross-feeding	(63)
Quorum sensing exchange	Sludge community	Granulation and EPS production	(245)
Antibiotic / antimicrobial peptide exchange	<i>K. pneumoniae</i> / <i>E. coli</i>	Microcin-mediated ammensalism	(142)
Environment sensing exchange	<i>Vibrio</i> pathogen / crustacean hosts	Differential virulence impacts	(180)
Direct physical interaction	Cellulose-fed soil biofilm community	Filamentous structures for electron transfer	(105)
Environment modulation	Marine phytoplankton community	Hydrogen peroxide scavenging	(165)

3-5-4 Emergent Properties

A major advantage of polymicrobial systems over traditional monoculture biotechnology is the potential for emergence of higher order properties. Emergent properties are attributes that are either not present or increased in magnitude from those properties characteristic of the individual system components. Contrary to common connotations, emergent properties are not always beneficial to a community; however, the discussions here are limited to positive attributes. Stability is defined by a community's response to perturbations (203), although many other interpretations have been used for the concept of ecological stability and related properties such as robustness (81). Generally, a stable community returns to its initial state after a small perturbation, while an unstable community does not. Stable community behavior is attributed to the two quantifiable metrics resistance and resilience (187). Resistance is defined as the degree to which community behavior is insensitive to a perturbation, and resilience is the rate at which community behavior returns to its original condition (220). An increase in any combination of resistance and resilience is desirable for bioprocessing and often for ecological fitness, defined as the ability of an organism to survive and reproduce in an environment. Another set of properties that can emerge from community interactions is productivity and uptake. Productivity is defined by the rate at which material is produced; uptake is the material consumption rate. Many ecological studies have observed correlations between productivity and diversity (or species richness) (272), although a universally accepted mechanism has not been identified.

3-6 Ecological Theories for Interpreting and Designing Communities

Ecologists have historically studied multiscale material and energy flows between populations and their environments. Resource scarcity has influenced fundamental aspects of biological organization, including the elemental and macromolecular composition of microorganisms (47, 55, 284). Hence, theories focused on themes of resource acquisition, resource investment, energetic efficiencies, and tradeoffs have been developed to explain observations of natural phenomena. Fundamental theories help organize observations, describe community behaviors, and enable strategic engineering of community composition or manipulation of environmental factors to improve natural and biotechnological processes (25, 123, 193). The current section details three distinct but complementary ecological theories that provide explanations for the competitive basis of component organization and interaction outcomes found in many natural and engineered communities. These classical theories will remain relevant to ecologists and bioengineers as their respective fields further mature. Table 3-4 outlines some existing studies of microbial communities employing these theoretical concepts.

Table 3-4. Ecological theory studies.

Theory	Community	Description	Reference
Maximum power	Aquatic planktonic microcosms	Effect of pH-controlled light on power acquisition	(29)
Maximum power	Two closely-related picocyanobacteria	Stable coexistence due to light partitioning	(235)
Resource ratio	Competition studies with bacteria, phytoplankton, and zooplankton	Meta-analysis of prediction consistency	(276)
Resource ratio	Nine phytoplankton species	Competition for nitrogen, phosphorus, and light	(22)
Pareto / Tradeoff	Three-member hot springs mat community	Community productivity and inhibitory byproduct tradeoffs	(242)
Pareto / Tradeoff	Syntrophic <i>Geobacter</i> species	Interspecies electron transfer	(169)

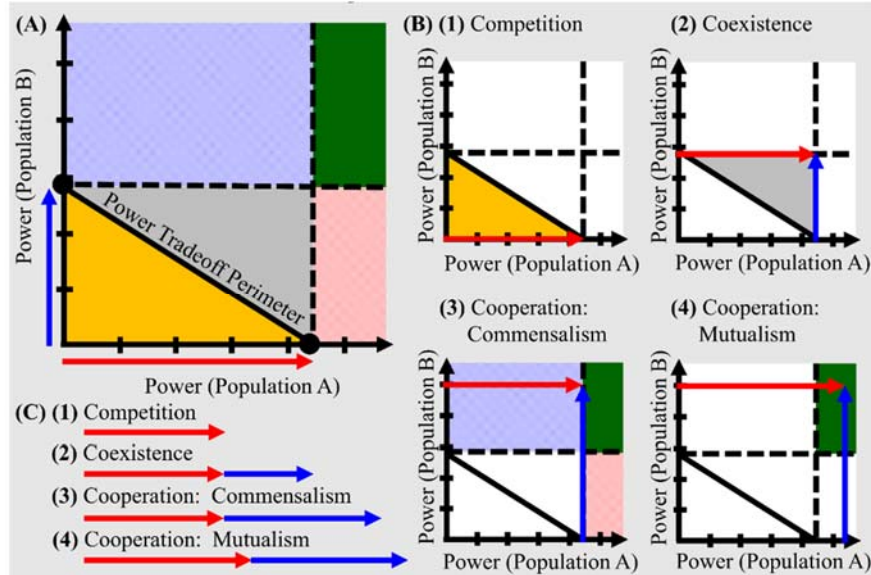


Figure 3-B1: Maximum Power Principle

(A) Graphical representation of key concepts from the maximum power principle.

Populations A and B are each characterized by a steady state metabolic power when growing in monoculture (black dots on axes; red and blue arrows represent magnitude of power for A and B, respectively). The line connecting these two points, the power tradeoff perimeter, represents the partitioning of metabolic power between the two populations. Any point within the plot represents a cumulative metabolic power of the community through summing the individual metabolic powers for populations A and B.

(B) Interaction outcomes interpreted with maximum power principle.

(B1) Populations A and B are competing for resources within the same ecological niche. The maximum power principle states that the fittest population has the largest metabolic power. Therefore, population A will outcompete population B. The total metabolic power of any mixture of two competing populations in the same niche is a non-negative linear combination of the individual metabolic powers and by definition cannot exceed the power tradeoff perimeter at steady state. For competition-based unsteady state scenarios, the community power will shift toward the population capable of the most metabolic power, which will eventually dominate (population A here).

(B2) Coexistence can occur when populations obtain energy from separate resources, allowing each population the potential to achieve its respective maximum power. Total community power may exceed the power tradeoff perimeter but is bounded by the magnitude of the individual population maximum powers. The gray region represents antagonism, as neither population is able to achieve its maximum power. A neutralistic community exists at the upper right corner (both populations achieve their maximum powers), while the edges bounding this region indicate an ammensal community in which one population realizes its maximum power and one does not.

(B3) Cooperating populations can exceed the maximum power of neutralistic populations if one population facilitates the acquisition of energy by the other population (e.g., commensal interaction) (boundary of the green region). Parasitic interactions may reduce such gains in community power because the increased power acquisition of one

Figure 3-B1 Continued

population occurs at the expense of the other population (red and blue patterned regions indicate parasitism favoring populations A and B, respectively).

(B4) The total community power can exceed the maximum power of neutralistic populations through mutualistic interactions. The two populations facilitate acquisition of metabolic power by each other, placing the community power in the green region. The final magnitude of an unsteady state trajectory is unknown without further information about the interaction.

(C) Comparing community metabolic powers. The example metabolic powers of populations A and B shown in (B1) – (B4) are summed and aligned to visualize the increase in metabolic power that may occur through community interactions.

Figure adapted from (144).

3-6-1 Maximum Power Principle

The maximum power principle describes community interactions based on acquisition of available energy. The maximum power principle asserts that the fitness of biological systems will increase with the rate of available energy harvest, resulting in a maximization of metabolic power (units of J/s) (143, 218). From an evolutionary viewpoint, the principle predicts selection of systems that capture previously unutilized energy sources (29). The ability to acquire available energy from the environment at faster rates leads to enhanced fitness by enabling more energy to be apportioned to survival and reproduction, while reducing available energy for competitors. Interactions between populations that increase the overall metabolic power of a system can lead to coexistence via cooperation (218) or niche differentiation such as the use of different substrates (29). Additional related postulates have been proposed over the years, including maximization or minimization of entropy (48, 154). Figure 3-B1 graphically demonstrates interaction outcomes for two populations with different metabolic powers.

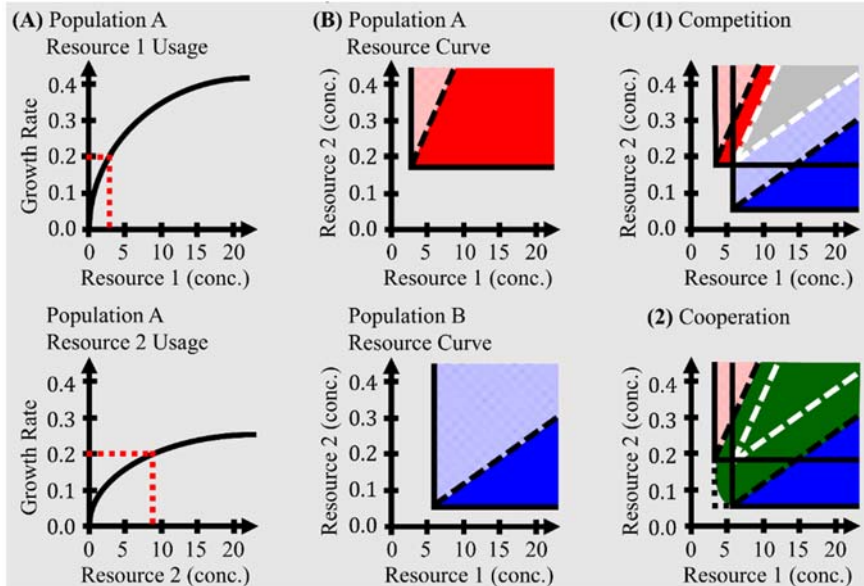


Figure 3-B2: Resource Ratio Theory

Graphical representation of key concepts from resource ratio theory.

(A) Determination of resource requirements. The resource concentrations required to support steady state growth of population A at different growth rates are tabulated using Monod expressions; an analogous set of data for population B is not shown. Resources 1 and 2 are essential for both species (e.g., nitrogen or phosphorus sources). The red dotted lines denote a particular growth rate selected for study and the corresponding resource concentrations required to support steady state growth.

(B) Population growth responses to limiting resources. The minimum requirements for resources 1 and 2 to maintain a steady state growth rate are indicated by the solid black edges for populations A (top) and B (bottom). Population A has a higher affinity for resource 1 and a lower affinity for resource 2 as compared to population B. Thus, populations A and B are more effective at depleting resources 1 and 2, respectively. The shaded areas represent conditions sustaining positive growth for each population. The slope of the dashed line separating the patterned region (resource 1-limited) from the solid region (resource 2-limited) represents the resource ratio requirement (resource 2 per resource 1 necessary for growth). If both resources are in excess, the population will maintain positive growth and consume resources as governed by its resource ratio requirement, until its consumption reaches the steady state resource level indicated by the solid lines.

(C) Interaction outcomes interpreted with resource ratio theory.

(C1) Resource ratio theory describes community interactions by combining the growth responses of the two populations and examining the resulting regions. If populations A and B are competing for resources 1 and 2, the population that can drive a resource concentration below the acquisition capability of the other will dominate the ecological niche. Population A will dominate the community in the red region due to its superior ability to deplete resource 1, and population B will dominate the community in the blue region due to its superior ability to deplete resource 2. These regions are bound by intersecting the resource ratio requirements of the two populations (white dashed lines).

Figure 3-B2 Continued

The gray region (coexistence), bound by the white dashed lines, represents an excess of resource 1 for population A and an excess of resource 2 for population B. The ratio of resources permits coexistence as both populations are limited by their non-competitive resources.

(C2) Resource ratio theory can describe cooperating populations utilizing mutualistic resource exchange. The region of coexistence (green region) is expanded (C1 gray region) due to resource trading. In the red patterned region, population A is resource 1-limited, leaving no excess resource 1 to trade with population B. Therefore, population A will dominate the community since it is more competitive not to trade with population B. In contrast, to the right of the red patterned region, population A is resource 2-limited and has excess resource 1, making it more competitive to exchange resource 1 with population B for resource 2. Analogous reasoning explains the solid blue region. Additionally, through resource trading, populations A and B are able to exist in a wider range of resource environments than is possible for either population individually (dotted box, lower left). In this region, both resources 1 and 2 are limiting for both populations. Thus, resource exchange is more competitive since neither population can maintain growth under these resource conditions without the other. The rectangular boundary around this extended resource availability region represents an ideal one-to-one trading scenario; however, the shape of this region may take on an ellipsoidal form depending on the costs and relative fluxes of resource exchange.

Figure adapted from (48).

3-6-2 Resource Ratio Theory

Resource ratio theory is an ecological theory describing resource consumption, competition, and niche partitioning in which inter-population interactions are defined with respect to shared resources (157, 260). These resources are often essential, but hemi-essential and substitutable resources have also been studied (258). Resource ratio theory postulates that the population best capable of depleting a limiting resource while maintaining a positive growth rate will be most competitive. The resource ratio theory has been used to assess outcomes (competitive exclusion or coexistence) between populations competing for shared limiting resources and can predict resource levels that will permit coexistence of multiple populations (182, 258–260). The theory can be

consistent with classic chemostat theory, which states that it is possible to sustain multiple stable populations when there are multiple limiting substrates, provided that each population is limited by a different resource (22, 28, 247).

Resource ratio theory can describe cooperative populations by accounting for mutualistic resource exchange. Cooperating populations that exchange limiting resources can exist in a wider range of resource environments than is possible for either population individually. This scenario describes what has been termed a super-competitor unit (157), a community with the emergent property of enhanced resource utilization that can deplete limiting resources more effectively than the individual monocultures. This ability to survive at reduced resource availabilities highlights an evolutionary advantage of cooperation that has been observed in natural ecosystems (157). Resource ratio theory extended to cooperation is analogous to the economic concept of comparative advantage, where specialization and resource exchange enable enhanced community function (58). Figure 3-B2 describes a generalized example of resource ratio theory applied to two different populations, each of which is more effective at depleting a different essential resource (157).

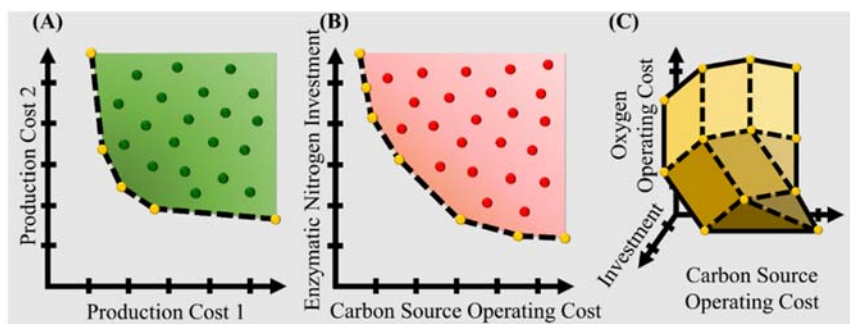


Figure 3-B3: Resource Allocation Theory: Pareto/Tradeoff Analysis

(A) A generalized Pareto tradeoff curve between two economic costs associated with the production of a good (for instance, the cost of fertilizer and pesticide required to produce

Figure 3-1 Continued

a bushel of corn). The plot quantifies the tradeoff between minimizing the two different costs. In the example, the yellow points and connecting line segments represent optimal cost-minimizing relationships between the variables, while green interior points represent higher-cost alternative strategies.

(B) A metabolic tradeoff curve analogous to a generalized Pareto tradeoff curve between two metabolic costs: (1) the carbon source operating cost, or the amount of carbon source required to synthesize a unit of bioproduct, and (2) enzymatic nitrogen investment, or the amount of nitrogen required to synthesize the enzymes in the utilized pathways. The yellow points represent the continuum of optimal pathways for producing the bioproduct given the objectives of minimizing the cost of carbon source or nitrogen investment; interior red points represent less efficient alternative pathways for producing the bioproduct.

(C) Three-dimensional metabolic tradeoff surface considering the requirements for three resources (carbon source, nitrogen, and oxygen). The resulting Pareto surface represents the optimal relationship between these three costs for synthesizing a bioproduct. For ease of visualization, the interior volume of the tradeoff surface is omitted.

3-6-3 Resource Allocation Theories: Pareto Surfaces and Metabolic Tradeoff Analysis

Table 3-5. Microbial community *in silico* analysis studies.

Focus	Description	Reference
Review of modeling	Application of genome-scale models to microbial communities	(148)
Cyanobacterial mat	Three EFMA modeling approaches	(242)
<i>Geobacter</i> electron transfer	Multi-omic modeling	(169)
Synthetic microbial ecosystems	Prediction of symbiotic communities via environmental constraints	(122)
SRB-methanogen interactions	FBA	(234)
Spatial community dynamics	Reaction-diffusion FBA	(88)
Three case studies of OptCom algorithm	Community FBA	(286)
Consortial studies	Dynamic community FBA	(86, 87, 285)

Abbreviations: EFMA, elementary flux mode analysis; FBA, flux balance analysis

Biological systems from enzymes to communities represent competitive resource allocation (232). Phenotypic plasticity, the ability to change phenotype with changing environment, can enable different relationships between cellular function and resource

allocation. However, phenotypic plasticity requires additional resources for the genes and regulatory systems to express different phenotypes under different environments. This investment cost for phenotypic plasticity may be ecologically justified for populations that persist in spatially and/or temporally dynamic environments (1).

Tradeoff theories analogous to those used in economics have been applied to metabolic systems to assess the competitive use of phenotypic plasticity (116). For many cellular behaviors, a benefit in one objective is realized only at the detriment of another objective. These tradeoffs are postulated to have tamed the ‘Darwinian demon,’ a super-species that can optimize all objectives simultaneously and which has never been observed in nature or the laboratory (137). The boundary that describes the tradeoff between resource uses is referred to as the Pareto front, efficiency frontier, or tradeoff surface (35, 116, 162, 223). Analysis of these resource allocation strategies is often paired with mathematical analysis such as stoichiometric modeling approaches (e.g., elementary flux mode and flux balance analyses) and/or kinetic models (Table 3-5) (162, 228). Phenotypic phase plane analysis and multi-objective optimization have been used in stoichiometric modeling approaches to examine the effects of resource allocation on competitive network function (31–33, 53, 173, 212). The tradeoff concept can be represented graphically via rates or efficiencies of resource utilization, a measure of cost. Figure 3-B3 demonstrates a generalized example of a Pareto optimization tradeoff curve for economic costs, with an analogous scenario translated into metabolic terms.

Single population examples of the predictions from resource allocation tradeoffs include the predicted use of the glyoxylate shunt and the Entner-Doudoroff glycolysis

pathway during *E. coli* nutrient-limited growth. These alternative pathways produce less ATP than the citric acid cycle and Embden-Meyerhof-Parnas glycolysis pathway, respectively, but also require fewer anabolic resources, such as nitrogen and carbon, to synthesize the pathway enzymes as compared to the higher ATP-yielding pathways (32, 70, 71).

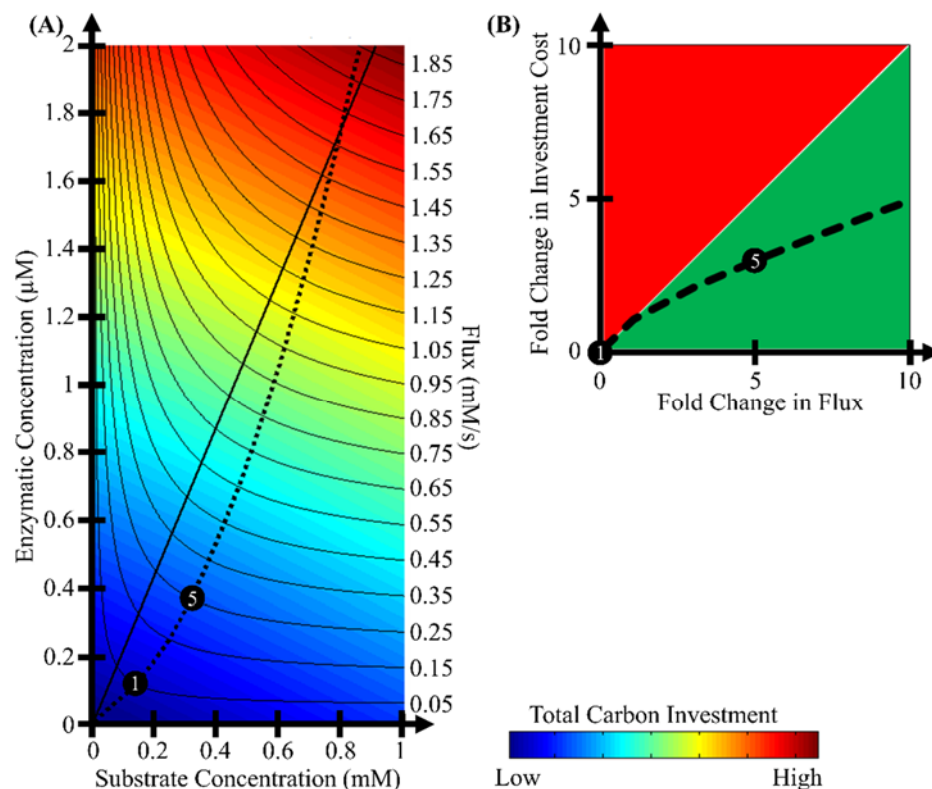


Figure 3-1. Resource investment and economies of scale for enzymatic flux. (A) A target enzymatic flux described by Michaelis-Menten kinetics can be achieved by a continuum of substrate or enzyme concentrations. Thin solid curves depict reaction isoflux lines that increase in $100 \mu\text{M/s}$ increments moving up and to the right. The substrate and enzyme concentrations each have associated resource costs related to the size and composition of the molecules. The total resource investment cost, depicted in color, sums the carbon investment from both the substrate and enzyme pools. The heavy solid line marks equal carbon investment contributions from substrate and enzyme pools. The dotted line represents the minimum total carbon investment for each isoflux line. Enzyme and substrate calculations model FumC from *E. coli* using $K_{\text{cat}}=1150 \text{ s}^{-1}$, $K_{\text{m}}=0.207 \text{ mM}$, assuming fumarate, and 8800 carbon atoms per functional enzyme. This principle can be extrapolated to entire pathways, posing a mechanism for improved

Figure 3-1 Continued

resource efficiency with increased specialization, as represented by the economies of scale. (B) The relationship between increasing flux and resource investment can be described based on the optimal minimization of investment. Total carbon investment per flux can be higher (red), lower (green), or the same (boundary) at larger fluxes as compared to smaller fluxes. In the example depicted, an increase in flux from 50 $\mu\text{M/s}$ (point 1) to 250 $\mu\text{M/s}$ (point 5) requires only three-fold more resources.

In addition to modulating metabolic pathways, resource investment transitions are predicted to occur based on allocation of resources to either enzyme or substrate pools. Classic Michaelis-Menten kinetics describe the driving of flux via enzyme ($v_{\text{max}} = k_{\text{cat}} \cdot [E]$) and substrate pools. These two pools create a continuum of drivers for a single flux, ranging from high relative substrate concentration to high relative enzyme concentrations (Figure 3-1A); these different flux-driving mechanisms are referred to as push- or pull-based mechanisms, respectively (34, 250, 266). The optimal distribution of resources between substrate and enzyme pools can vary depending on the required flux (Figure 3-1A, dotted line). Taking the FumC enzyme as an example, relatively small fluxes are best driven by a substrate-based push mechanism which minimizes overall resource requirements, while higher fluxes are best driven by a pull mechanism in which elevated enzyme concentrations represent a more efficient use of anabolic carbon.

Resource allocation between enzyme and substrate pools can also benefit microbial communities when populations exchange metabolites such as amino acids, vitamins, or nucleotides. Maximization of flux per investment of a valuable resource can be applied to a community exchanging metabolites such as cyanobacterial heterocyst-vegetative cell interactions (133). As a numerical illustration, five interacting populations can each individually support a flux of 50 $\mu\text{M/s}$ through enzyme X, which would require a total

substrate and enzyme carbon investment of 1.51 CmM (millimolar concentration of carbon). Alternatively, one population can specialize in that function and drive a flux of 250 $\mu\text{M/s}$ with a total carbon investment of 4.43 CmM and then exchange the product with the community. The division of labor strategy with product exchange reduces the total community carbon investment requirement by $\sim 41\%$, neglecting cost of metabolite exchange. Figure 3-1B demonstrates the general principle of flux as a function of total carbon investment into both enzyme and substrate pools. The net community investment savings continue to improve as the flux magnitude increases. Extrapolating to entire pathways will likely offset the involvement of transport proteins while maintaining resource investment savings. While this economy of scale holds from a purely reaction kinetics basis, additional factors such as diffusion rates and protein synthesis machinery could play substantial roles.

3-7 Case Studies of Communities with Interpretation

The maximum power principle, resource ratio theory, and resource allocation theory are useful for developing and interrogating the design principles of microbial communities. The current section provides an analysis of natural and engineered communities with dissection of relevant components, interactions, and theories.

3-7-1 Multilevel Interactions in a Naturally Occurring Phototrophic Community

Community Description: The naturally occurring phototrophic community “*Chlorochromatium aggregatum*” is comprised of the phototrophic green sulfur

bacterium *Chlorobium chlorochromatii* physically attached to the heterotrophic β -proteobacterium *Candidatus Symbiobacter mobilis* (Figure 3-2). This community is typically arranged as thirteen green sulfur bacteria surrounding a single central β -proteobacterium and is found in sulfide-containing, oxic-anoxic interfaces of stratified lakes (177, 178). The green sulfur bacteria are non-motile anoxygenic phototrophs which use light energy and electrons from sulfide to fix carbon and nitrogen. The heterotrophic β -proteobacterium is motile and possesses genetic evidence of phototactic and chemotactic capabilities but cannot harvest light energy and has limited respiratory capacities (139). The green sulfur bacterium, although culturable in the laboratory, is not found free-living in nature, and the β -proteobacterium is unculturable independent of the green sulfur bacteria (139, 271).

Interaction Mechanisms: “*C. aggregatum*” interacts via a number of mechanisms. First, the green sulfur bacteria exchange anabolic and catabolic metabolites (reduced carbon and nitrogen) with the β -proteobacterium. Metabolite exchange is hypothesized to be bidirectional; the β -proteobacterium may synthesize compounds such as acetate that could be assimilated by the green sulfur bacteria due to limited respiratory capabilities (139). Additionally, the β -proteobacterium and green sulfur bacteria are connected by periplasmic tubules. The direct physical interaction between the bacteria enables the β -proteobacterium to provide community motility toward light and sulfide (27, 166). These intercellular contacts could also permit exchange of electrons via soluble carriers (274), enabling community-wide energy transfer and redox balancing (139).

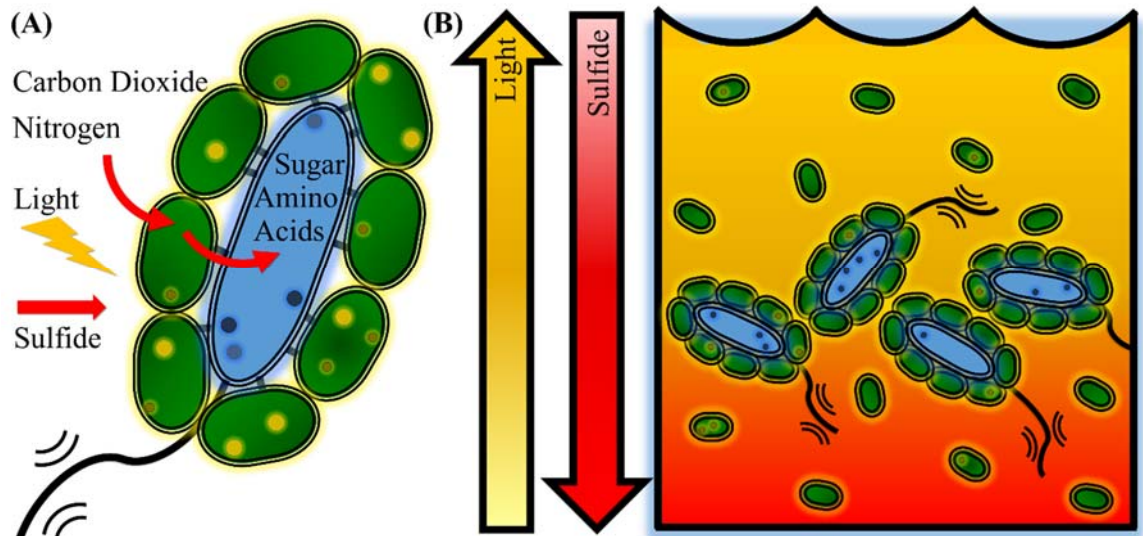


Figure 3-2. Phototrophic community interactions. (A) Green sulfur bacteria fix carbon dioxide and nitrogen from the environment and exchange reduced products with the heterotrophic β -proteobacteria. In return, the β -proteobacteria maintains a flagellum and provides taxis toward light and sulfide, which the green sulfur bacteria require for photosynthesis. (B) In a stratified lake environment, light is more intense at the top of the lake, and sulfide concentration increases toward the bottom. The motile community is able to take advantage of ideal concentrations of both resources. Free-living green sulfur bacteria cannot actively direct their movement toward resources, and the β -proteobacteria is not able to survive independently.

Theory Applications: Resource ratio and resource allocation theories provide an evolutionary rationale for the interactions of these two bacteria. In an environment with scarce resources (e.g., light, sulfide, reduced carbon, reduced nitrogen), this cooperative strategy confers a selective advantage (157). The physical interactions and metabolite exchange are consistent with the resource ratio theory concept of a super-competitor unit. The motility contributed by the β -proteobacterium enables a competitive advantage for growth at low sulfide concentrations and lower light intensities, directed movement toward essential resources (74). Additionally, the community configuration minimizing the associated resource investment costs. The direct physical connection between the bacteria permits fourteen cells to benefit from the ability to move actively toward light

and sulfide while only one cell needs to maintain the sensing and motility genes, synthesize flagellar proteins, and power flagellar operation (139).

3-7-2 Anaerobic Syntrophy in Methanogenic Communities

Community Description: Methane is a major focus of renewable energy efforts and plays an essential role in ecological food webs. Anaerobic communities produce methane by catabolizing organic feedstocks like biomass via a cascade of cross-feeding microorganisms. Sulfate-reducing bacteria (SRBs) and methanogenic archaea are key members of these anaerobic communities found in aquatic sediments and anaerobic digesters (172). The SRBs oxidize organic acids and produce hydrogen or formate in the absence of alternative electron acceptors. The methanogens catabolize hydrogen and formate for cellular energy, ultimately producing methane as a byproduct (Figure 3-3) (159, 279).

Interaction Mechanisms: Methanogenic communities have a strong requirement for metabolite exchange due to the thermodynamic constraints of hydrogen synthesis (210, 230, 255). Hydrogen partial pressures above a very small critical threshold (~ 10 Pa) shift the chemical reaction equilibrium from hydrogen synthesis to hydrogen consumption, stalling SRB metabolism in the absence of sulfate. The consumption of hydrogen by methanogens reduces local hydrogen partial pressures below the critical threshold, enabling further SRB catabolism (230). The mutualistic interactions can extend further. In addition to providing the methanogens with substrate, the SRBs can also modify the local environment by reducing sulfate to sulfide rather than forming hydrogen. Sulfide is

highly reactive with oxygen and reduces the environmental concentrations of oxygen, benefiting the oxygen-inhibited methanogens. In addition, research suggests that SRBs enhance methanogen consumption of hydrogen by exchanging anabolic resources like amino acids (273).

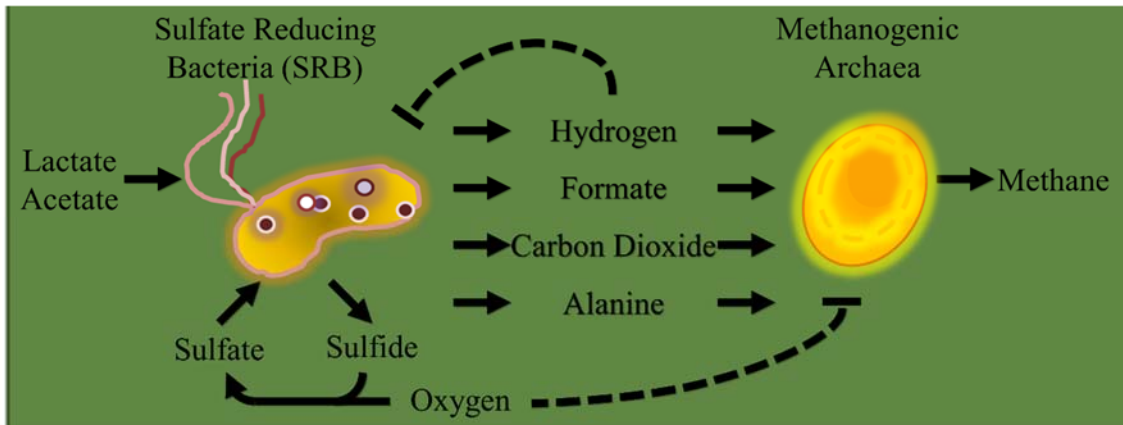


Figure 3-3. Sulfate-reducing bacteria (SRBs) – methanogen anaerobic syntrophy. SRBs create a reducing environment conducive to methanogenesis by reducing sulfate to sulfide, which can then abiotically react with oxygen. In the absence of sulfate, SRBs produce hydrogen; this metabolism becomes unfavorable without a hydrogen sink. Methanogens consume the hydrogen, facilitating SRB oxidation of carbon. SRBs cross-feed anabolic resources like amino acids to facilitate methanogen growth, resulting in a larger hydrogen sink.

Theory Applications: SRB-methanogen communities are a well-studied example of anaerobic syntrophy (158). The cross-feeding of hydrogen, which can also be viewed as an environmental modification interaction, represents a classic example of the maximum power principle by collectively increasing the potential for energy acquisition from the environment. Metabolite exchange allows the SRBs to consume substrate at higher rates, thus increasing the collective community metabolic power. The exchange of anabolic resources such as amino acids is proposed to facilitate methanogen growth by providing an energetically expensive metabolite, thus increasing methanogen consumption of

hydrogen (273). The amino acid cross-feeding strategy represents an example of improved function per anabolic resource investment according to resource allocation theory (Figure 3-1B).

The syntrophic interactions of methanogens are not strictly limited to hydrogen-producing SRBs. Hydrogen-producing fermenters, which provide the substrate for SRBs, also benefit from methanogens (17). Hydrogen transfer has also been observed in communities of hydrogen-producing anaerobic fungi and methanogens, which demonstrate the emergent property of more complete degradation of cellulosic biomass (40).

3-7-3 Cross-feeding Chemostat Communities

Community Description: Culturing a clonal population of *E. coli* in a glucose-limited chemostat for hundreds of generations results in the reproducible formation of multiple, phenotypically distinct *E. coli* populations which cross-feed secondary metabolites (Figure 3-4) (115, 200, 261). The interacting community consists of three distinct functional populations: a glucose-catabolizing glycerol- and acetate-producing specialist, an acetate-catabolizing specialist, and a glycerol-catabolizing specialist. Both the acetate- and glycerol-catabolizing specialists completely oxidize their respective substrates. The population functions are based primarily on mutations in gene regulatory sequences, resulting in altered gene expression (115). For instance, the glucose-catabolizing specialist expresses the high-affinity Mgl sugar transporter and has reduced expression of citric acid cycle enzymes, resulting in acetate and glycerol secretion. This strain, which

represents 80% of the community, grown in isolation has a lower specific growth rate and a lower biomass per glucose yield than the parent monoculture, yet the community demonstrates a 15% improvement in biomass per glucose yield. The reproducibility of the interacting consortium suggests the organization of components and interactions is more competitive and stable than an *E. coli* monoculture in a homogeneous environment like a chemostat.

Interaction Mechanisms: The community populations interact through the cross-feeding of catabolic resources. Additionally, consumption of acetate represents an environmental modification interaction that reduces the inhibition of the glucose-catabolizing strain.

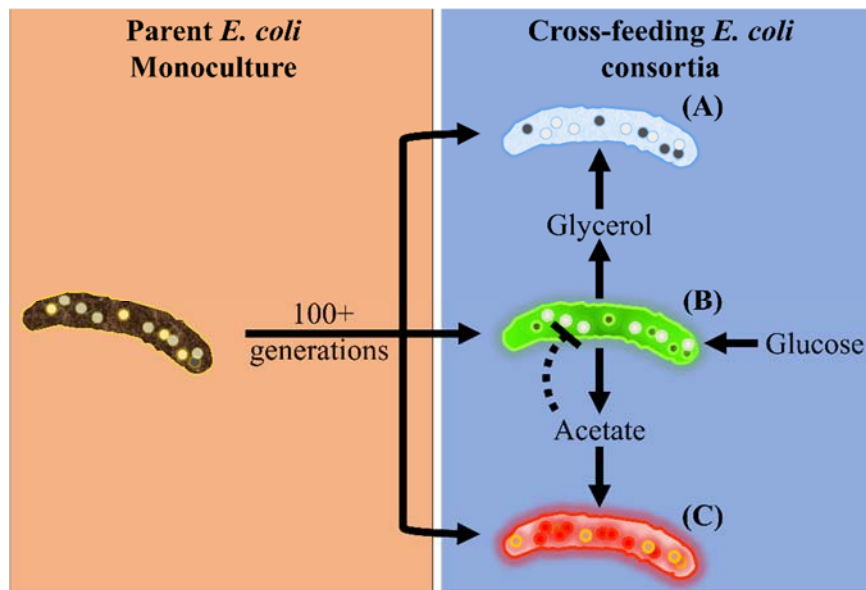


Figure 3-4. Naturally occurring *E. coli* crossing-feeding communities. *E. coli* monocultures grown long-term in glucose-limited chemostats evolve into cross-feeding communities. Community members include a glucose-catabolizing specialist that secretes acetate and glycerol (B), an acetate-catabolizing specialist (C), and a glycerol-catabolizing specialist (A). The cross-feeding community has a ~15% improvement on biomass per glucose yield as compared to the parent monoculture and depletes available substrates more completely than the parent monoculture.

Theory Applications: The cross-feeding *E. coli* community highlights key features of the maximum power principle, resource ratio theory, and resource allocation theory. The consortium has a higher metabolic power than the monoculture because inhibitory byproducts (e.g., acetate) are removed during growth, allowing for a higher community metabolic rate. The community, as a whole, demonstrates the emergent property of a super-competitor unit defined in resource ratio theory; through metabolite exchange, the community drives the concentrations of carbon-based resources to levels unattainable by the parent monoculture. The populations also demonstrate an enhanced resource allocation configuration through division of metabolic labor. Each respective population increases expression of enzymes associated with a portion of central metabolism. The glucose-consuming specialist has higher glycolytic expression but lower citric acid cycle fluxes, resulting in the secretion of acetate and glycerol (115, 200, 261). The strategy is representative of the tradeoff between substrate-driven and enzyme-driven fluxes; driving a reduced number of higher fluxes can be less resource intensive than driving a greater number of smaller fluxes.

A synthetic biology analog of the cross-feeding community has been created based on gene deletion rather than altered gene regulation (16). The synthetic system demonstrated improved biomass productivity under batch, chemostat, and biofilm culturing conditions.

3-8 Conclusions

The ebbs and flows of scientific efforts build largely on achievements of the predecessors of the modern scientific community. While gaining popularity in recent

years, the importance of microbial communities is not new. At this time of renewed interest in a systems-focused future, it is critical to keep a broad view. The early reductionist scientists are to be commended for laying a solid foundation that enables the current trajectory toward systems studies. In fact, reductionist approaches have a bright future because systems cannot be constructed solely on noble postulations but must be grounded in well-described components. It is not possible to define an emergent property until the basic components and interactions are appropriately catalogued.

A paramount goal and current challenge facing microbial community design is to mechanistically understand the design principles that govern higher order attributes of polymicrobial communities. The definitions presented here lay a foundation for a common language to bridge gaps across varying disciplines and enable greater collaboration between fields studying polymicrobial communities. The three ecological theories discussed – maximum power principle, resource ratio theory, and resource allocation theory – distill a broad range of literature regarding ecological theory into a concise synthesis and highlight the relevance of these theories to community interaction outcomes. The usefulness of these theories is illustrated through application to specific case studies in both natural and engineered environments. The tenets of resource acquisition, concentration, and allocation will remain useful for developing and examining the design principles of microbial communities and guide the future of polymicrobial bioprocess engineering.

3-9 Acknowledgments

The development of this work was supported by the National Institutes of Health awards (1U01EB019416-01, *P20GM103474*) and the Genomic Science Program (GSP), Office of Biological and Environmental Research (OBER), U.S. Department of Energy (DOE), and is a contribution of the PNNL Foundational Scientific Focus Area (Principles of Microbial Community Design) subcontracted to Montana State University. Ashley E. Beck was supported by the Office of the Provost at Montana State University through the Molecular Biosciences Program. Hans C. Bernstein was supported by the Linus Pauling Distinguished Postdoctoral Fellowship which is a Laboratory Directed Research and Development program at PNNL; operated for the DOE by Battelle Memorial Institute under Contract DE-AC05-76RLO 1830. *The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.*

CHAPTER 4

STOICHIOMETRIC MODELING OF ASSIMILATORY AND DISSIMILATORY
BIOMASS UTILIZATION IN A MICROBIAL COMMUNITY

4-1 Contribution of Authors and Co-Authors

Author: Kristopher A. Hunt

Contributions: Assisted in concept development, helped with editing, generated figures, and wrote the paper.

Co-Author: Ryan deM. Jennings

Contributions: Assisted in concept development and helped with editing.

Co-Author: William P. Inskeep

Contributions: Helped with editing.

Co-Author: Ross P. Carlson

Contributions: Assisted in concept development and helped with editing.

4-2 Manuscript Information Page

Kristopher A. Hunt, Ryan deM. Jennings, William P. Inskeep, Ross P. Carlson
Environmental Microbiology

Status of Manuscript:

Prepared for submission to a peer-reviewed journal

Officially submitted to a peer-review journal

Accepted by a peer-reviewed journal

Published in a peer-reviewed journal

Published online August 11, 2016

STOICHIOMETRIC MODELING OF ASSIMILATORY AND DISSIMILATORY BIOMASS UTILIZATION IN A MICROBIAL COMMUNITY

Kristopher A. Hunt^{1,2,3}, Ryan deM. Jennings^{1,3,4,5}, William P. Inskeep^{1,3,4}, Ross P. Carlson^{1,2,3}

¹Center for Biofilm Engineering, Montana State University, Bozeman, MT

²Department of Chemical and Biological Engineering, Montana State University, Bozeman, MT

³Thermal Biology Institute, Montana State University, Bozeman, MT

⁴Department of Land Resources and Environmental Sciences, Montana State University, Bozeman, MT

⁵Current address: Department of Biology, Mercer University, Macon, GA

4-3 Abstract

Assimilatory and dissimilatory utilization of autotroph biomass by heterotrophs is a fundamental mechanism for the transfer of nutrients and energy across trophic levels. Metagenome data from a tractable, thermoacidophilic microbial community in Yellowstone National Park was used to build an *in silico* model to study heterotrophic utilization of autotroph biomass using elementary flux mode analysis and flux balance analysis. Assimilatory and dissimilatory biomass utilization was investigated using 29 forms of biomass-derived dissolved organic carbon (DOC) including individual monomer pools, individual macromolecular pools, and aggregate biomass. The simulations identified ecologically competitive strategies for utilizing DOC under conditions of varying electron donor, electron acceptor, or enzyme limitation. Culturing environment affected which form of DOC was the most competitive use of nutrients; for instance, oxygen limitation favored utilization of less reduced and fermentable DOC while carbon-limited environments favored more reduced DOC. Additionally, metabolism was studied

considering two encompassing metabolic strategies: simultaneous versus sequential use of DOC. Results of this study bound the transfer of nutrients and energy through microbial food webs, providing a quantitative foundation relevant to most microbial ecosystems.

4-4 Introduction

Nutrient and energy transfers across trophic levels are essential in nearly all environmental, industrial, and medical microbial ecosystems. Primary producers obtain nutrients, including carbon, often via autotrophy using energy from chemolithotrophy or phototrophy. Heterotrophic utilization (e.g. predation, decomposition, and metabolite exchange) of resources from primary production requires nutrient assimilation and dissimilation, which contributes directly to biogeochemical cycling. Theoretical and experimental analysis of resource transfer between trophic levels is challenging due to the diversity of biomass constituents (i.e. macromolecules, monomers, cofactors, free metabolites, and minerals), the complexity of biochemical networks, and the number of intercellular metabolite exchanges that occur within microbial communities. Generalized kinetic models have been used for decades to analyze processes like wastewater treatment (79), composting (155), and carbon flow in sediments (4). However, these analyses typically use generic, instead of organism specific, yields or rates because direct measurements for each participating population of microorganisms and substrates are challenging.

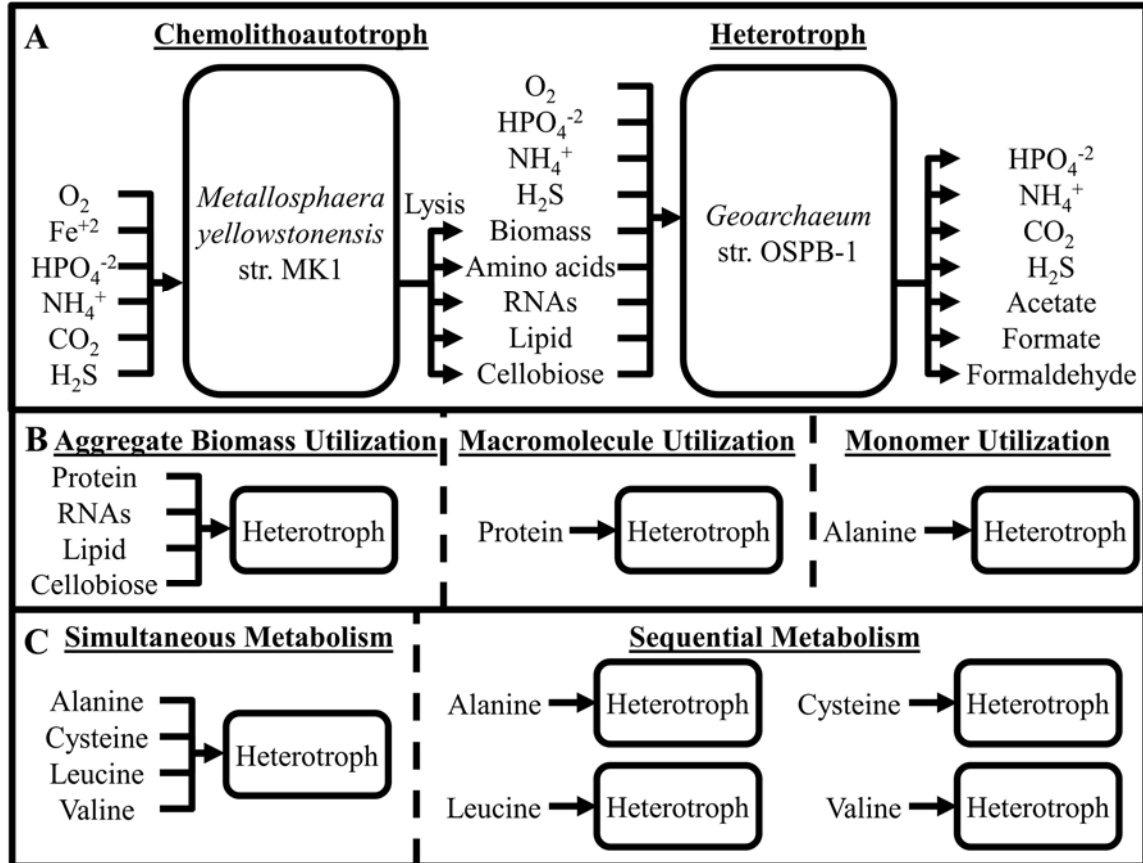


Figure 4-1 Conceptual representation of high-temperature acidic Fe(III)-oxide microbial mats. Primary productivity was represented by the chemolithoautotroph, *Metallosphaera yellowstonensis* str. MK1, which fixes carbon dioxide through iron oxidation and provides potential carbon and energy sources for community heterotrophs. Community heterotrophy was represented by *Geoarchaeum* str. OSPB, a numerically dominant heterotroph in the mats. Exchanges of nutrients and energy across trophic levels were analyzed considering 29 different biomass-derived dissolved organic carbon (B) ranging from individual monomers to macromolecules to aggregate autotroph biomass. Simulations also explored two distinct heterotroph strategies: simultaneous versus sequential metabolism of dissolved organic carbon pools (C).

Most naturally occurring, mesophilic ecosystems are complex assemblages of archaea, bacteria, eukaryotes, and a multitude of potential nutrients and energy sources. For instance, a gram of soil has an estimated 10^3 – 10^6 species representing all three domains of life and numerous trophic levels from primary producers to decomposers (45). These ecosystems are too complex to represent using current molecular-level modeling methods

without major simplifications. Relatively simple microbial mat communities, like those found in high-temperature environments of Yellowstone National Park (YNP), are tractable systems for examining assimilatory and dissimilatory biomass utilization. High-temperature acidic, iron-oxidizing microbial mats found in YNP have been studied extensively (101, 107, 128) and provide ideal model systems for examining nutrient and energy transfer across trophic levels (i.e. assimilatory and dissimilatory biomass utilization) (Figure 4-1A). High temperature (65–75 °C) and low pH (3–3.5) contribute to the simplicity of these microbial communities by inhibiting photosynthetic organisms and eukaryotes (21, 23, 198).

A well-studied system in One Hundred Springs Plain, Norris Geyser Basin YNP, contains 5–7 dominant microbial populations, of which 2-3 are metabolically active in the oxic region based on metagenomic analyses (101, 107, 128). *Metallosphaera yellowstonensis* str. MK1 is a dominant primary producer in these mats, and has been shown to fix inorganic carbon (DIC) during aerobic oxidation of ferrous iron (107). The numerically dominant heterotroph (30–50 % of the total aerobic community) in these mats is an archaeon from the newly described group of Geoarchaeota, referred to here as *Geoarchaeum* str. OSPB (128, 129). This organism functions as an aerobic chemoorganoheterotroph based on metabolic analysis of metagenome assemblies (129). Viral lysis and senescence are two candidate mechanisms for the release of resources derived from chemolithoautotroph biomass utilized by heterotrophic organisms including *Geoarchaeum* str. OSPB (23, 198). *M. yellowstonensis* str. MK1 and *Geoarchaeum* str. OSPB both require oxygen as a terminal electron acceptor for growth and cellular energy

production, and *in situ* analysis of oxygen as a function of mat depth indicated that oxygen diffusion, not reaction rate, is rate-limiting for aerobic growth (14). Therefore, the autotroph and heterotroph may compete for this limiting resource.

Metagenome-enabled stoichiometric modeling can examine molecular-level phenomena without requiring extensive kinetic parameters. Stoichiometric modeling of metabolic systems identifies all possible genome-encoded physiologies, including ecologically-relevant optimal and suboptimal phenotypes for nutrient utilization (121, 175, 195, 262). There are two widely applied methods of stoichiometric modeling. Flux balance analysis (FBA) is an optimization approach that uses a metabolic model to determine possible routes of nutrient flow that, for example, maximize biomass production assuming a substrate utilization rate (267). Elementary flux mode analysis (EFMA) is an unbiased stoichiometric modeling method that enumerates the smallest set of genotypes that describe nutrient flow through a metabolic model (elementary flux modes, EFMs) (see Figure 4-S1 for a detailed description). Nonnegative linear combinations of EFMs can represent all feasible phenotypes of the modeled organism (121, 140, 208). EFMA has enabled both strain development and community analysis without prior knowledge of the function of an organism (30, 242).

Stoichiometric modeling has been used to study microbial community ecology (39, 234, 242, 283); however, an assessment of assimilatory and dissimilatory uses of biomass-derived, dissolved organic carbon (DOC) as mechanisms of nutrient and energy transfer across trophic levels has not been conducted. This study examined biomass and cellular energy production in a tractable microbial mat community from YNP. Study

aims were to 1) identify the optimal biomass-derived DOC for biomass and cellular energy production by heterotroph *Geoarchaeum* str. OSPB as a function of different environmental scenarios (i.e. DOC-, oxygen-, and enzyme-limitation), 2) develop a relationship between microbial community member abundance based on inter-trophic nutrient transfer, 3) contrast simultaneous and sequential metabolism of biomass-derived DOC for heterotroph biomass and cellular energy production, and 4) predict distribution of byproducts during metabolism of biomass-derived DOC.

4-5 Results

4-5-1 Construction of *in silico* Model and Application of Ecological Theory

An *in silico* stoichiometric model was built to dissect potential autotroph-heterotroph interactions based on utilization of autotroph biomass components. Inorganic carbon fixed by *M. yellowstonensis* str. MK1 (hereafter termed autotroph) was modeled as the source of reduced carbon for the microbial community (Figure 4-1A). Community heterotrophy was represented by *Geoarchaeum* str. OSPB (hereafter termed heterotroph) utilization of autotroph biomass components (107). Heterotroph utilization of autotroph biomass was simulated using 29 forms of biomass-derived DOC ranging from individual monomer pools (e.g. alanine), macromolecular pools (e.g. protein), to aggregate biomass (Figure 4-1B and Table 4-1). The composition of monomers, macromolecules, and aggregate biomass were based on genetic attributes of the autotroph and assumptions detailed in the Materials and Methods. The geothermal source water was assumed to contain nonlimiting quantities of ammonium, phosphate, sulfide, and all micronutrients.

All other nutrient exchanges were modeled as depicted in Figure 4-1A. The metabolic model with metagenome justification and data verifying atom and electron conservation can be found in the Supplemental Material.

Table 4-1 Summary of simulations for 29 forms of biomass-derived dissolved organic carbon (DOC). Biomass production by heterotroph *Geoarchaeum* str. OSPB was evaluated using DOC ranging from monomers, to macromolecules, to aggregate biomass derived from the autotroph *M. yellowstonensis* str. MK1.

Carbon Source	Number of EFMs	Chemical Formula	Degree of Reduction ^a	Nutrients Consumed ^b	Byproducts
Biomass	14,994,531	CH _{1.6} N _{0.2} O _{0.4} P _{0.02} S _{0.005}	4.27 (4.92)		Acetate, Formate, NH ₄ ⁺ , HPO ₄ ⁻² , H ₂ S, H ₂ O ₂
Cellobiose	794,557	C ₁₂ H ₂₂ O ₁₁	4.00 (4.00)	HPO ₄ ⁻² , NH ₄ ⁺ , H ₂ S	Acetate, Formate, Formaldehyde, H ₂ O ₂
Archaeal Lipid	216,476	C ₈₆ H ₁₇₀ O ₁₂ P ₂	5.86 (5.86)	H ₂ S, NH ₄ ⁺	Acetate, Formate, Formaldehyde, HPO ₄ ⁻² , H ₂ O ₂
Nucleotides					
RNA	8,332,169	CHN _{0.4} O _{0.8} P _{0.1}	2.91 (4.07)		
AMP	438,345	C ₁₀ H ₁₂ N ₅ O ₇ P	3.00 (4.50)	H ₂ S	Acetate, Formate, NH ₄ ⁺ , HPO ₄ ⁻² , H ₂ O ₂
CMP	137,746	C ₉ H ₁₂ N ₃ O ₈ P	3.33 (4.33)		
GMP	433,861	C ₁₀ H ₁₂ N ₅ O ₈ P	2.80 (4.30)		
UMP	138,091	C ₉ H ₁₁ N ₂ O ₉ P	3.33 (4.00)		
Amino Acids					
Protein	303,907	CH _{1.6} N _{0.3} O _{0.3} S _{0.006}	4.21 (5.01)		
Ala	85,822	C ₃ H ₇ NO ₂	4.00 (5.00)		
Arg	8,721	C ₆ H ₁₅ N ₄ O ₂	3.67 (5.67)		
Asn	116,017	C ₄ H ₈ N ₂ O ₃	3.00 (4.50)		
Asp	92,836	C ₄ H ₆ NO ₄	3.00 (3.75)		
Cys	117,743	C ₃ H ₇ NO ₂ S	3.33 (4.33)		
Glu	12,408	C ₅ H ₈ NO ₄	3.60 (4.20)		
Gln	14,202	C ₅ H ₁₀ N ₂ O ₃	3.60 (4.80)		
Gly	100,664	C ₂ H ₅ NO ₂	3.00 (4.50)		
His	16,487	C ₉ H ₆ N ₃ O ₂	3.33 (4.83)	HPO ₄ ⁻²	Acetate, Formate, Formaldehyde, H ₂ O ₂ , NH ₄ ⁺ , H ₂ S
Ile	79,510	C ₆ H ₁₃ NO ₂	5.00 (5.50)		
Leu	88,657	C ₆ H ₁₃ NO ₂	5.00 (5.50)		
Lys	38,383	C ₆ H ₁₅ N ₂ O ₂	4.67 (5.67)		
Met	118,907	C ₅ H ₁₁ NO ₂	4.40 (5.00)		
Phe	55,714	C ₉ H ₁₁ NO ₂	4.44 (4.78)		
Pro	15,419	C ₅ H ₉ NO ₂	4.40 (5.00)		
Ser	104,376	C ₃ H ₇ NO ₃	3.33 (4.33)		
Thr	60,396	C ₄ H ₉ NO ₃	4.00 (4.75)		
Trp	165,116	C ₁₁ H ₁₂ N ₂ O ₂	4.18 (4.73)		
Tyr	56,065	C ₉ H ₁₁ NO ₃	4.22 (4.56)		
Val	30,653	C ₅ H ₁₁ NO ₂	4.80 (5.40)		

^a: Degree of reduction was calculated on NH₃(N₂) bases

^b: Nutrients consumed during biomass production

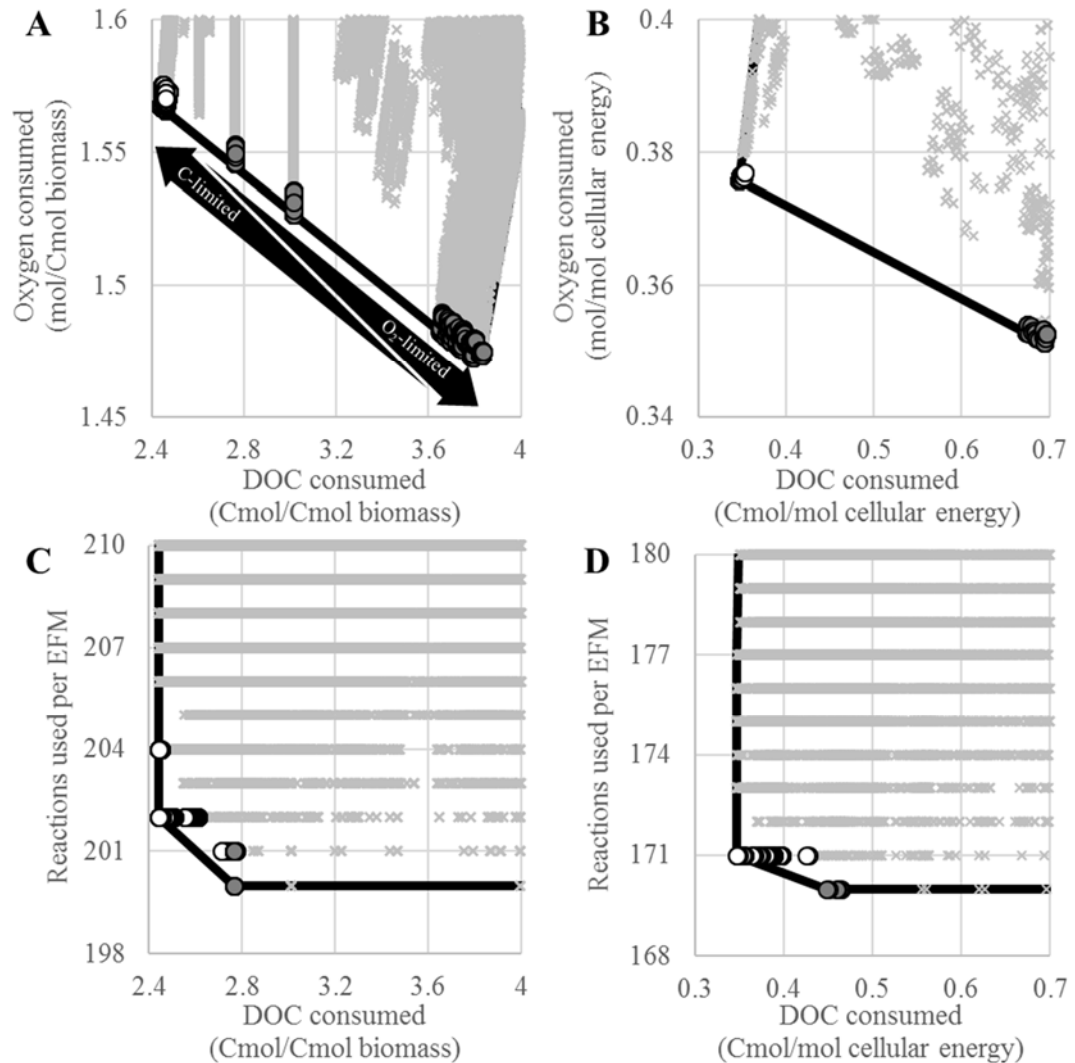


Figure 4-2 Summary of ecological resource utilization theory used for quantifying competitive physiologies during utilization of autotroph biomass by the heterotroph, *Geoarchaeum* str. OSPB. Resource costs (biomass-derived dissolved organic carbon (DOC), oxygen, and enzymes) were calculated for every elementary flux mode (EFM) that synthesized biomass or cellular energy. Biomass and cellular energy production costs are plotted for the carbon- versus oxygen-limited scenarios (A) and (B), respectively. Biomass and cellular energy production costs are plotted for the carbon- versus enzyme-limited scenarios (C) and (D), respectively. Linear combinations of the most efficient EFMs define an optimal tradeoff surface (line). EFMs plotted with circles represent suboptimal ecologically competitive strategies that are phenotypically indistinguishable from the optimal strategies. These strategies ranged from complete oxidation of DOC to acetate production (white and grey circles, respectively) in (A), (B), and (D), as well as energy production (grey circles) in (C). EFMs plotted with an x are suboptimal and phenotypically distinct from the optimal strategies. Reactions used per EFM are the number of reactions with a nonzero flux in each EFM, which is a proxy for the enzyme resource cost.

The metabolic model was analyzed using EFMA and FBA. Biologically-relevant flux distributions, represented as EFMs or FBA optimization solutions, were identified using ecological resource allocation theory, which assumes an organism will maximize the utility of a limiting resource (11, 242). The theory was applied to the complete set of EFMs by calculating the resource costs of each EFM as defined by the amount of reduced carbon source (biomass-derived DOC), electron acceptor (oxygen), or enzyme resource (e.g. nitrogen, sulfur, iron, or cell volume) required to produce either biomass or cellular energy (phosphodiester bond equivalents). Figure 4-2 provides a graphical representation of the analysis concepts using aggregate biomass as the example DOC. The enzyme resource cost for biomass and cellular energy production (Figures 4-2C and 4-2D, respectively) was modeled by summing the number of participating reactions in each EFM, which is proportional to the number of participating enzymes and therefore, the amount of enzyme resource required to synthesize these enzymes (73). This approach approximates all enzyme complexes to be the same size with the same amino acid distribution.

For every EFM, pairs of resource costs for biomass or cellular energy production were plotted to define tradeoff curves or ‘pareto fronts’ (Figure 4-2) (11), which identified the optimal phenotype for any position along the resource limitation gradient. The EFMs whose resource costs were closest to the plot axes and their nonnegative linear combinations (e.g. solid black lines in Figure 4-2) minimize resource costs, making them ecologically competitive (31). EFMs not on the curve represent suboptimal strategies for the simulated environment (x symbols in Figure 4-2). Abundance of suboptimal EFMs

illustrates the metabolic robustness of the reaction network, and EFMs near the tradeoff curve may have roles in metabolic resilience, such as responding to resource or genetic perturbations (circles in Figure 4-2). Biological implications of these specific suboptimal EFMs are discussed below.

4-5-2 Optimal Utilization of Macromolecules

Approximately 50 million EFMs were enumerated during the analysis of the 29 forms of DOC (Table 4-1). Every EFM was analyzed for the resource cost to produce biomass or cellular energy on gradients of DOC, oxygen, or enzyme resource limitation (Figures 4-3 through 4-5). Only the optimal tradeoff surfaces are presented for each form of DOC for clarity; the total number of EFMs for each form of DOC and common associated byproducts are listed in Table 4-1. Tradeoff curves generated using either EFMA or FBA were generally identical (Supplemental Material).

Lipid required the fewest Cmoles of DOC to produce a Cmole of heterotroph biomass or mole of cellular energy and was therefore the best form of DOC for carbon-limited environments relative to the other 28 forms of DOC (moving to left on x-axis in Figures 4-3A and 4-3B). The lower DOC resource cost for lipid was due primarily to the high degree of reduction (6 electrons per Cmole) (Table 4-1). After lipid, the most competitive macromolecular DOC to produce a Cmole of heterotroph biomass was aggregate biomass followed by protein, cellobiose, and lastly, RNA. The DOC produced biomass or cellular energy at varying resource costs due to 1) the degree of reduction of the DOC, which is a gross measure of energy content, 2) chemical structure of the DOC, which determines the entry point into central metabolism relative to substrate-level phosphorylation reactions,

3) the elemental stoichiometry of the DOC relative to heterotroph biomass, and/or 4) the biological accessibility of energy and nutrients in the DOC, which is related to the number of enzymes required to process the DOC. The ratio of Cmole DOC utilized per Cmole biomass produced quantifies the fraction of autotroph-derived carbon assimilated into heterotroph biomass with the balance dissimilated for cellular energy. This ratio establishes an important conceptual constraint between different trophic levels; at least 2.4 Cmoles of aggregate autotroph biomass are required to produce 1 Cmole of heterotroph biomass.

RNA required the fewest moles of oxygen per Cmole biomass or mole cellular energy produced among the macromolecular DOC for simulated oxygen-limited conditions (moving down y-axis in Figures 4-3A and 4-3B). The low requirement for oxygen was due to the low degree of reduction of RNA and the fermentable pentose sugar backbone. Oxygen resource costs to produce biomass or cellular energy increased for the other macromolecular DOC in the following order: cellobiose, aggregate biomass, protein, and finally, lipid. Reduced carbon byproducts were produced from most DOC as oxygen became more limiting with acetate and formate being common (a complete list of byproducts is available in Table 4-1). The metabolic model of the heterotroph did not permit biomass or cellular energy production under anoxic conditions for any considered forms of DOC, which agrees with experimental observations (9).

Heterotroph biomass produced from protein required the fewest enzymatic reactions, and by extension, the lowest enzyme resource cost of all macromolecular DOC considered. Firstly, the protein macromolecule supplied all 20 common amino acids,

which eliminated the need for amino acid synthesis reactions; secondly, the amino acid backbones were readily converted into central metabolism intermediates such as 3-phosphoglycerate and acetyl-CoA, which minimized the number of reactions involved in carbon backbone processing. After protein, the macromolecular DOC that minimized enzyme resource cost for biomass production were: cellobiose, RNA, aggregate biomass, and finally, lipid (Figure 4-3C). Heterotroph lipid (digeranylfarnesyl glycerol phospholipid (129)) was modeled as chemically distinct from autotroph lipid (cardarchaeol diphosphate (94)); therefore, utilization of autotroph lipid required enzymes for both degradation and reassembly to produce heterotroph biomass.

Cellular energy produced from cellobiose required the fewest enzymatic steps among the macromolecular DOC (Figure 4-3D); all other macromolecules were degraded to multiple central metabolism precursors, requiring more enzymatic reactions to produce cellular energy. Cellobiose was followed by lipid, RNA, protein, and then aggregate biomass as the most cost effective macromolecular DOC for cellular energy production with respect to enzyme resource cost. Nitrogen, phosphorous, and sulfur containing DOC produced noncarbon byproducts, including ammonium, phosphate, and sulfide when dissimilated to produce cellular energy (Table 4-1).

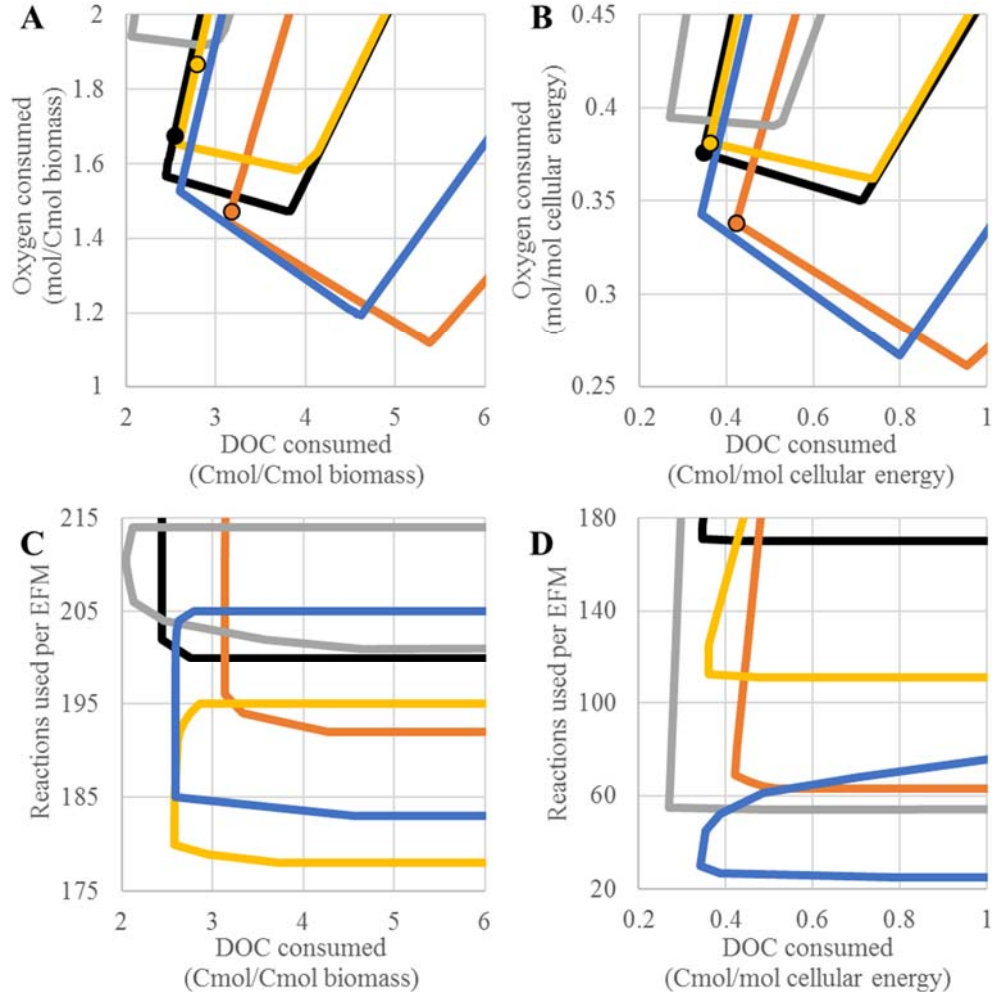


Figure 4-3 Analysis of macromolecule utilization by the heterotroph *Geoarchaeum* str. OSPB. Biomass-derived dissolved organic carbon (DOC) includes aggregate biomass (black), archaeal lipid (grey), cellobiose (blue), protein (yellow), and RNA (orange). Biomass and cellular energy production costs are plotted for the carbon- versus oxygen-limited scenarios (A) and (B), respectively. Biomass and cellular energy production costs are plotted for the carbon- versus enzyme-limited scenarios (C) and (D), respectively. Reactions used per elementary flux mode (EFM) are the number of reactions with a nonzero flux in each EFM, which is a proxy for the enzyme resource cost. Each tradeoff curve was calculated assuming simultaneous use of the macromolecule monomers, when applicable. Results from the sequential use of macromolecule monomers are highlighted with the circles in (A) and (B). For example, the solid yellow line in (A) outlines the tradeoff curve when the amino acids in the protein macromolecule were utilized simultaneously to produce biomass. The left most position on the yellow tradeoff curve represents the lowest DOC resource cost to produce biomass under carbon-limited conditions with its associated oxygen resource cost. The yellow circle (A) represents the DOC and oxygen resource costs to produce the same Cmole of heterotroph biomass when each amino acid is consumed sequentially; this simulation assumed all amino acids were metabolized without excretion.

4-5-3 Selective, Optimal Utilization of Monomers from Macromolecules

DOC simulations included both biomass-derived macromolecule and monomer pools enabling a comparison of different metabolic strategies. Biomass and cellular energy production from the RNA macromolecule and the four RNA monomers (AMP, CMP, GMP, and UMP) revealed differences in DOC energy content and the biological accessibility of the carbon backbones (Figure 4-4). Biomass and cellular energy production from the monomers CMP and UMP had lower DOC resource costs than either the aggregate RNA macromolecule, AMP, or GMP (Figures 4-4A and 4-4B); this was due to the higher degree of reduction and the connectivity of the carbon backbones to central metabolism. The enzyme resource cost was higher for the RNA macromolecule than any of the individual nucleotides because four separate monomer degradation pathways were necessary to generate central metabolism intermediates.

The 20 amino acids have a variety of metabolic properties including different elemental compositions and structures, which affected their potential use as substrates (Table 4-1). Isoleucine and leucine had the lowest DOC resource costs to produce both biomass and cellular energy compared to the protein macromolecule and other amino acids (Figure 4-5). This was due largely to the high degree of reduction of these amino acids and the entry points of the carbon backbones into central metabolism. Serine had the lowest oxygen resource cost to produce biomass of the amino acids, making it more competitive than the protein macromolecule or any other amino acid under oxygen-limited conditions. Serine and cysteine had the lowest oxygen resource requirement for cellular energy production of the amino acids. The serine backbone enters glycolysis at

pyruvate, which can be converted to acetate, carbon dioxide, and cellular energy using only a few enzymatic reactions including substrate-level phosphorylation. Moreover, acetate can be secreted, which removes reducing equivalents that would otherwise require oxygen. The enzyme resource costs to produce cellular energy from amino acids were a function of the carbon backbone (Figure 4-5C and 4-5D). Structurally simpler amino acids, such as glutamate and aspartate, could be integrated into central metabolism with fewer enzymatic reactions than cyclic or aromatic amino acids, such as phenylalanine, which required longer and more specialized enzymatic pathways. A similar trend was observed for biomass production, with the exception of histidine and arginine. Degradation of histidine and arginine required fewer enzymatic reactions than their synthesis (Supplemental Material); this differential made the DOC ideal for minimizing enzyme resource costs to produce biomass.

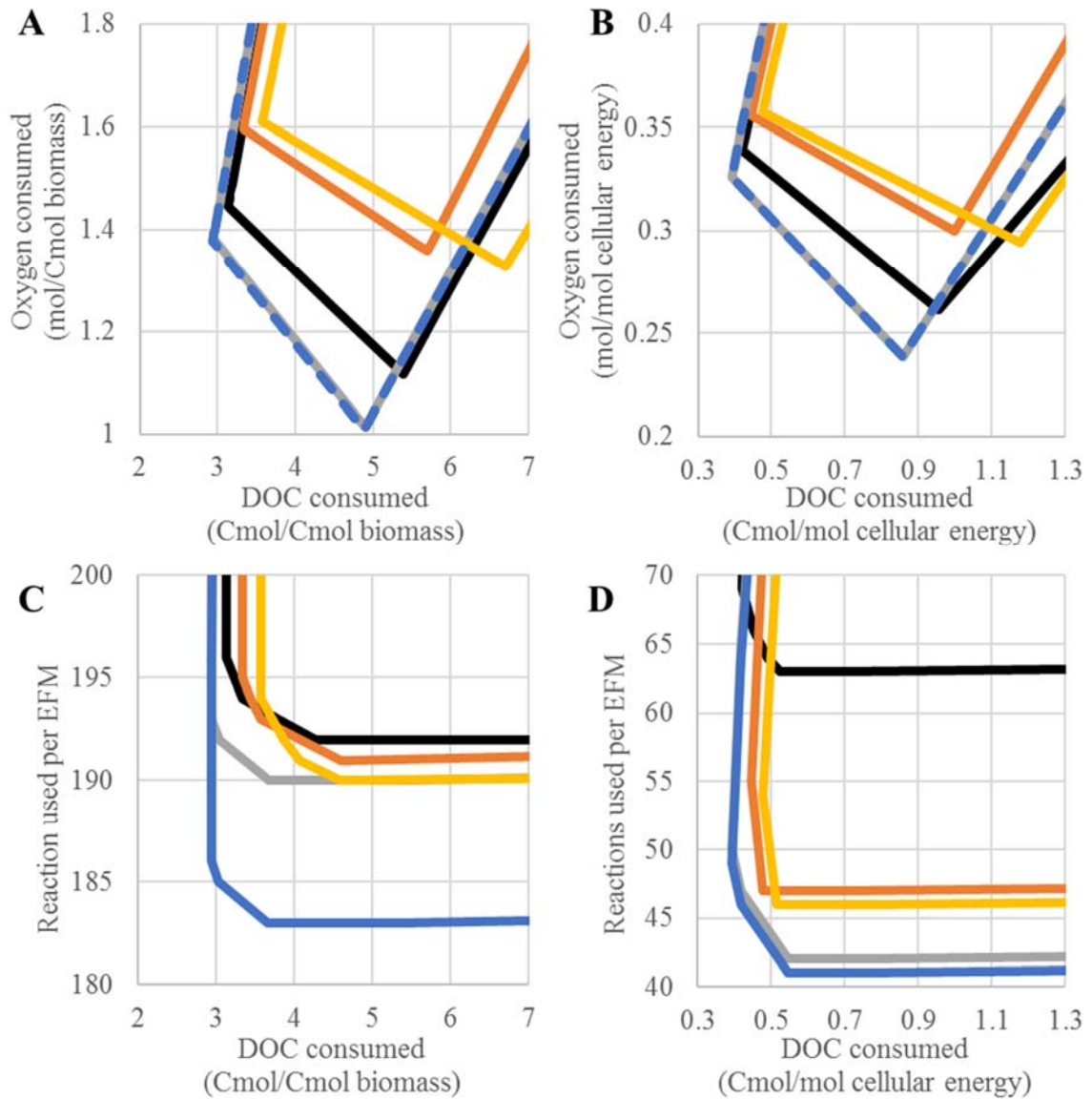


Figure 4-4 Analysis of RNA macromolecule and monomer utilization by the heterotroph *Geoarchaeum* str. OSPB. Biomass-derived dissolved organic carbon (DOC) includes RNA macromolecule (black), AMP (orange), CMP (grey), GMP (yellow), and UMP (blue). Biomass and cellular energy production costs are plotted for the carbon- versus oxygen-limited scenarios (A) and (B), respectively. Biomass and cellular energy production costs are plotted for the carbon- versus enzyme-limited scenarios (C) and (D), respectively. Reactions used per elementary flux mode (EFM) are the number of reactions with a nonzero flux in each EFM, which is a proxy for enzyme resource cost. UMP and CMP overlap in (A) and (B).

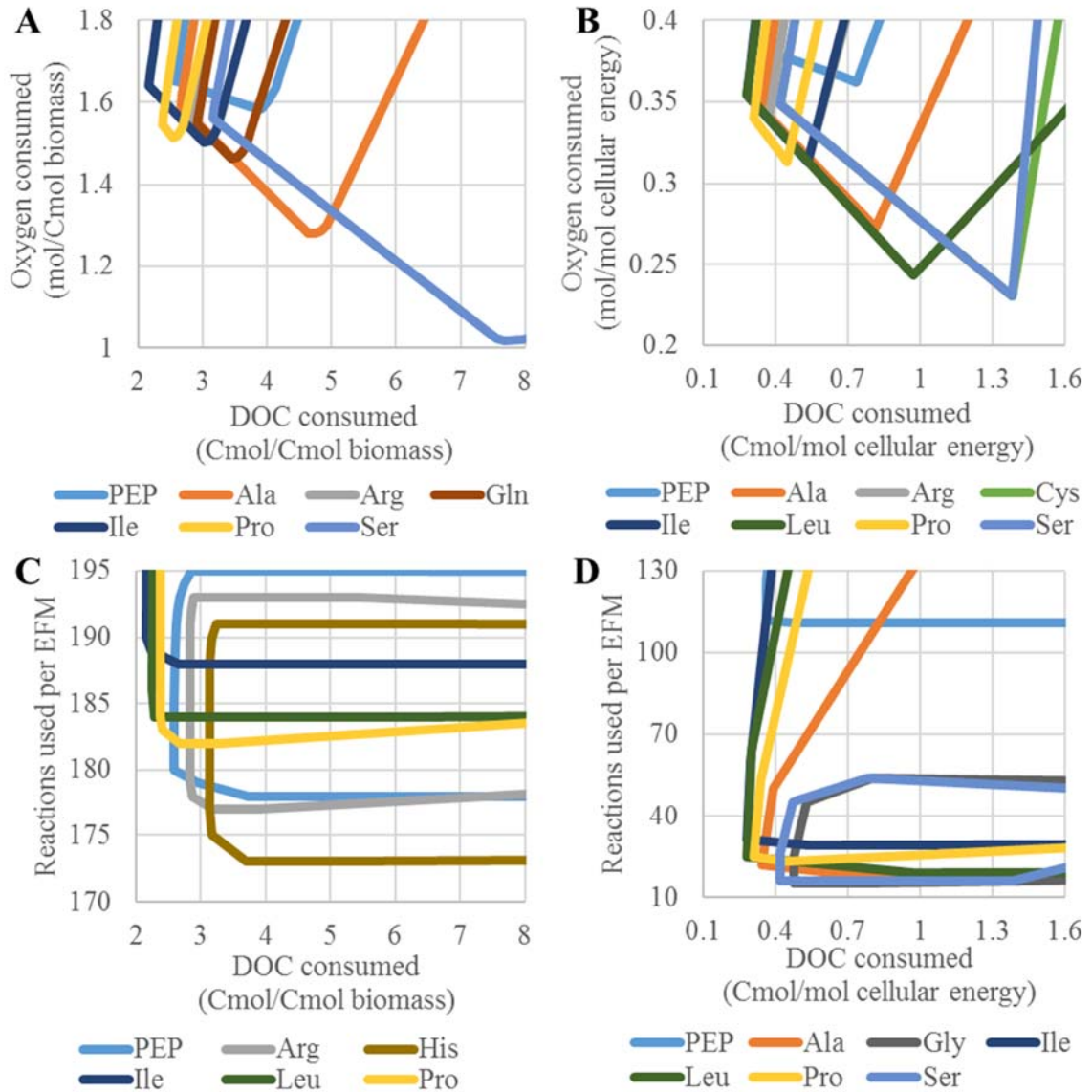


Figure 4-5 Analysis of protein macromolecule and amino acid utilization by the heterotroph *Geoarchaeum* str. OSPB. Biomass-derived dissolved organic carbon (DOC) includes protein macromolecule and individual amino acids. Biomass and cellular energy production costs are plotted for the carbon- versus oxygen-limited scenarios (A) and (B), respectively. Biomass and cellular energy production costs are plotted for the carbon- versus enzyme-limited scenarios (C) and (D), respectively. Reactions used per elementary flux mode (EFM) are the number of reactions with a nonzero flux in each EFM, which is a proxy for enzyme resource cost. Only amino acids with optimal properties are plotted in each subfigure; supplemental data contains the complete amino acid data set.

Table 4-2 Resource costs of heterotroph *Geoarchaeum* str. OSPB biomass production from the 29 analyzed biomass-derived dissolved organic carbon (DOC) pools. DOC is ordered from low to high biomass production costs for the three examined nutrient limitations.

	Carbon-limited Costs		Oxygen-limited Costs			Enzyme-limited Costs		
	DOC per Biomass	Oxygen per Biomass		Oxygen per Biomass	DOC per Biomass		Fewest Reactions Used per EFM	DOC per Biomass
LIP	2.06	1.94	UMP	1.01	4.89	His	173	3.70
Ile	2.17	1.64	CMP	1.01	4.90	Arg	177	3.20
Leu	2.25	1.74	Ser	1.02	7.68	PEP	178	3.73
Val	2.30	1.69	Cys	1.03	8.01	Gly	180	6.65
Pro	2.38	1.54	RNA	1.12	5.39	Ser	180	7.68
BIO	2.44	1.57	CEL	1.19	4.62	Cys	181	8.25
PEP	2.58	1.65	Asn	1.27	6.30	Pro	182	2.67
CEL	2.60	1.53	Asp	1.28	6.33	Gln	182	3.64
Ala	2.66	1.59	Ala	1.28	4.75	Glu	182	3.66
Met	2.75	2.23	GMP	1.33	6.71	Ala	182	4.88
Lys	2.84	2.24	Leu	1.34	4.86	Asn	182	6.48
Arg	2.84	1.54	AMP	1.36	5.70	Asp	182	6.51
Thr	2.86	1.79	Gly	1.40	6.54	UMP	183	3.67
Gln	2.91	1.55	Gln	1.46	3.45	CEL	183	4.57
Glu	2.92	1.56	His	1.47	4.16	Leu	184	2.29
UMP	2.94	1.38	Glu	1.47	3.47	Thr	184	4.17
CMP	2.94	1.38	BIO	1.47	3.79	Met	186	3.26
RNA	3.14	1.44	Ile	1.50	3.02	Lys	187	2.89
His	3.14	1.54	Arg	1.51	3.03	Ile	188	2.65
Ser	3.16	1.56	Pro	1.51	2.54	Trp	188	4.12
Cys	3.21	1.61	PEP	1.58	3.89	Val	190	2.35
AMP	3.33	1.59	Val	1.67	2.46	CMP	190	3.68
Trp	3.35	2.43	Thr	1.72	3.96	GMP	190	4.58
Asn	3.53	1.58	LIP	1.92	2.89	Tyr	190	6.51
Asp	3.55	1.59	Met	2.23	2.75	Phe	190	7.85
GMP	3.57	1.61	Lys	2.24	2.84	AMP	191	4.60
Gly	3.62	1.64	Trp	2.43	3.35	RNA	192	4.28
Tyr	3.76	2.90	Tyr	2.90	3.76	BIO	200	2.76
Phe	4.04	3.42	Phe	3.42	4.04	LIP	201	4.66

Biomass production costs utilizing biomass (BIO), archaeal lipid (LIP), cellobiose (CEL), a representative protein (PEP), and a representative distribution of nucleotides (RNA). Nucleotides and amino acids are labeled using standard biochemical conventions. Units are Cmoles DOC consumed, Cmoles biomass produced, and moles oxygen consumed.

Table 4-3 Resource costs of heterotroph *Geoarchaeum* str. OSPB cellular energy production from the 29 analyzed biomass-derived dissolved organic carbon (DOC) pools. DOC is ordered from low to high cellular energy production costs for the three examined nutrient limitations.

	Carbon-limited Costs		Oxygen-limited Costs			Enzyme-limited Costs		
	DOC per Energy	Oxygen per Energy		Oxygen per Energy	DOC per Energy		Fewest Reactions Used per EFM	DOC per Energy
LIP	0.270	0.395	Cys	0.230	1.382	Gly	15	0.480
Ile	0.283	0.354	Ser	0.230	1.382	Ser	16	0.418
Leu	0.283	0.354	CMP	0.238	0.857	Ala	17	0.817
Val	0.300	0.360	UMP	0.238	0.857	Cys	17	1.382
Pro	0.309	0.340	Leu	0.243	0.972	Leu	19	0.972
CEL	0.343	0.343	RNA	0.261	0.954	Asp	20	1.090
Ala	0.346	0.346	CEL	0.267	0.800	Glu	21	0.612
BIO	0.347	0.376	Ala	0.272	0.817	Gln	21	0.612
Lys	0.360	0.420	Asn	0.272	1.090	Pro	23	0.448
PEP	0.362	0.381	Asp	0.272	1.090	His	23	0.620
Met	0.366	0.439	GMP	0.294	1.176	Asn	23	1.090
Arg	0.371	0.340	Gly	0.299	1.198	Thr	24	0.705
Thr	0.375	0.375	AMP	0.300	1.000	Lys	24	0.900
Glu	0.380	0.342	Gln	0.306	0.612	CEL	25	0.800
Gln	0.380	0.342	His	0.306	0.734	Arg	26	0.537
UMP	0.391	0.326	Glu	0.306	0.612	Val	29	0.319
CMP	0.391	0.326	Arg	0.313	0.537	Ile	29	0.537
His	0.409	0.341	Pro	0.313	0.448	Met	30	0.882
Ser	0.418	0.349	Ile	0.313	0.537	Trp	31	1.082
Cys	0.418	0.349	Val	0.343	0.428	Tyr	31	1.102
RNA	0.421	0.339	BIO	0.351	0.707	Phe	34	1.349
Trp	0.437	0.457	Thr	0.353	0.705	UMP	41	0.545
AMP	0.444	0.356	PEP	0.362	0.734	CMP	42	0.545
Asp	0.461	0.346	LIP	0.391	0.509	GMP	46	0.513
Asn	0.461	0.346	Lys	0.420	0.360	AMP	47	0.476
GMP	0.476	0.357	Met	0.439	0.366	LIP	54	0.509
Gly	0.480	0.360	Trp	0.457	0.437	RNA	63	0.523
Tyr	0.495	0.523	Tyr	0.523	0.495	PEP	111	0.483
Phe	0.540	0.600	Phe	0.600	0.540	BIO	170	0.448

Energy production costs utilizing biomass (BIO), archaeal lipid (LIP), cellobiose (CEL), a representative protein (PEP), and a representative distribution of nucleotides (RNA).

Nucleotides and amino acids are labeled using standard biochemical conventions.

Units are Cmoles DOC consumed, moles energy produced, and moles oxygen consumed.

4-5-4 Simultaneous Versus Sequential Metabolism of Monomer Pools

Biomass utilization strategies were examined as two metabolic scenarios: simultaneous and sequential DOC metabolism (Figure 4-1C). Differences in the two metabolic strategies were exemplified by comparing the metabolism of the protein macromolecule to the sequential utilization of each amino acid (Figure 4-3). The sequential metabolism was quantified via nonnegative linear combinations of optimal EFMs for each amino acid (Tables 4-2 and 4-3). Sequential metabolism of individual amino acids had higher DOC and oxygen resource costs to produce biomass than the simultaneous metabolism of the amino acids (Figure 4-3A). Sequential consumption necessitated more metabolic processing; each amino acid had to be converted into the other 19 amino acids as well as the other biomass components. The added metabolic processing consumed more cellular energy and reducing equivalents and produced more carbon byproducts, which increased the DOC resource cost to produce biomass.

Simultaneous versus sequential DOC consumption had no effect on cellular energy production (Figure 4-3B). Cellular energy production exclusively harvests energy from the DOC, while biomass production uses DOC to produce both biomass components and cellular energy. This property resulted in non-linear relationships between different forms of DOC for biomass but not cellular energy production.

4-5-5 Identification of Suboptimal Metabolic Strategies

Resource allocation theory identified optimal strategies for biomass and cellular energy production along continuous gradients of resource limitations. However, this does

not capture the total robustness of the metabolic network. EFMA enumerated all mathematically possible biochemical pathways including optimal and suboptimal strategies. The suboptimal strategies can be biologically significant because they create metabolic redundancy and buffer against perturbations (149). The analysis of suboptimal strategies considered EFMs with biomass yields within 0.5% of the optimal tradeoff surface as phenotypically indistinguishable and were therefore part of the tradeoff surface. The 0.5% threshold value was set based on experimental observations (52). Analysis of suboptimal biomass production strategies expanded the relevant number of EFMs along the DOC versus oxygen tradeoff curve from 3 to 2073 for aggregate biomass utilization (Figure 4-2) while it had no effect on the 4 EFMs considered on the cellobiose utilization tradeoff curve.

The serine-glycine pathway is an example of an identified suboptimal metabolic strategy with competitive properties (Figure 4-2). The serine-glycine pathway uses serine and glycine synthesis reactions along with the glycine cleavage system to oxidize DOC (Figure 4-S2). This pathway permits complete oxidation of DOC, but requires fewer enzymatic steps than the TCA cycle, potentially making it competitive in enzyme-limited environments. However, fewer substrate-level phosphorylation reactions occurred in this pathway, which increased DOC resource cost. In the case of aggregate biomass utilization to produce biomass, the serine-glycine pathway is suboptimal on the enzyme resource plots, though it is optimal for some forms of DOC, such as cellobiose. Additionally, the serine-glycine pathway provides flexibility in response to oxygen limitation. In the presence of sufficient oxygen, the pathway can completely oxidize DOC

to maximize cellular energy production. Alternatively, under oxygen limitation, formate produced when regenerating tetrahydrofolate can be excreted to lower the requirement for oxygen. Consequently, results from EFMA analysis provide a theoretical foundation for understanding and/or predicting effects of different electron acceptor costs on microbial community response.

4-6 Discussion

Nutrient and energy transfer across trophic levels during assimilatory and dissimilatory biomass utilization is fundamental to most ecosystems. Heterotrophic utilization of reduced carbon constituents produced by autotrophs can occur as a result of metabolite exchange, predation (e.g. viral lysis), and/or decomposition. The presented work focused on decomposition of autotroph biomass and established quantitative relationships between community members. For example, a minimum of 2.4 Cmoles of autotroph biomass is required to synthesize 1 Cmole of heterotroph biomass. The exchange of metabolites is also common both within and across trophic levels (164, 273). Rational analysis of these synergistic exchanges requires a numerical basis. The value of exchanged DOC can be quantified in terms of the amount of biomass or cellular energy that can be produced from it (e.g. Tables 4-2 and 4-3). For instance, the amino acid isoleucine can be oxidized to produce 3.5 moles of cellular energy per Cmole, while phenylalanine can be oxidized to produce 1.9 moles of cellular energy per Cmole (Table 4-3). The presented study provides a quantitative foundation for analyzing metabolite exchange within microbial communities.

Stoichiometric modeling is well suited for analyzing complex systems such as metabolism and can quantify the extremes of cellular physiology without knowledge of substrate concentrations or specific enzyme parameters; this is a major strength of the modeling technique. Cellular phenotypes over the continuum of specific growth rates can be modeled using linear combinations of the EFMs at fast and slow growth rates, namely the EFMs that produce only biomass and only cellular energy, respectively. In this study, the modeled biomass production assumes a relatively high specific growth rate (0.1 hr^{-1}) and therefore, the majority of substrate is directed toward synthesizing biomass components and not toward cell maintenance (3.5 moles cellular energy per Cmole biomass). Conversely, at slow specific growth rates, most of the cellular activity is directed toward cell maintenance; these phenotypes are represented in the cellular energy plots (e.g. Figure 4-2B and 4-2C). Combinations of these two scenarios can be used to approximate any phenotype between the two extremes. Stoichiometric modeling does not automatically account for enzyme affinities and therefore accounting for low substrate concentrations and the expression of high affinity transporters would need to be implemented manually; the additional energetic requirements for a high affinity transporter vs. a low affinity transporter has been detailed previously (32). In theory, strategies that only produce cellular energy could describe the ultimate low nutrient environments found in the deep subsurface, where biomass doubling times on the order of 10^2 – 10^4 years have been estimated (92).

DOC utilization when multiple forms are present was bounded by two scenarios, simultaneous and sequential metabolism. These results can be used to study microbial

growth at either high or low DOC concentrations. Many experimental studies have examined sequential metabolism of simple carbohydrates and organic acids, often termed catabolite repression or diauxic growth (49, 78, 278). Sequential metabolism is competitive at high DOC concentrations where use of a single form of DOC can saturate the activity of a minimal enzyme investment and drive high growth rates. At low DOC concentrations, such as those found in most natural environments including the studied mat system, simultaneous metabolism of multiple forms of DOC is observed. This strategy requires more investment into degradation pathways, but is necessary to maximize growth rate (125).

The presented study establishes a foundation for the evaluation of biomass utilization by heterotrophs; additional complexity could be integrated during future work. Utilization of complex DOC mixtures is common in natural environments and is constrained by bioavailability (e.g. solubility and polymer hydrolysis), DOC transport into a cell, and the necessary degradation pathways (125). The solubility of DOC was not explicitly considered here, and could confound the predicted optimality of some forms of DOC, such as lipids which have lower solubilities ($\sim 5 \mu\text{g L}^{-1}$) than most other DOC constituents. A review of relevant DOC solubility can be found in the supplemental material (see Table 4-S1). Currently, all DOC is transported and hydrolyzed via a single macromolecule-specific mechanism (see Materials and Methods). While this permits equitable comparisons across simulations for simultaneous and sequential metabolism, it also highlights scientific knowledge gaps associated with enzyme specificity and enzyme resource costs of transport and hydrolysis reactions. Many microorganisms have multiple

transporters for the same molecule (e.g. low affinity versus high affinity transporters) or same class of molecules (e.g. glucose versus glucose oligomer transporters), which can influence phenotype (282). Another simplification made in the current study was the complete utilization of a metabolite pool (e.g. all amino acids in the protein macromolecule) as compared to the selective use of optimal monomers (e.g. metabolism of alanine combined with the secretion of phenylalanine), or the selective use of components from a complex molecule (e.g. metabolism of the ribose unit in an RNA nucleotide with the secretion of the recalcitrant aromatic base). Wastewater and sediment studies have commonly shown that recalcitrant molecules like aromatic amino acids are often utilized last (4, 42, 248). Finally, the broad applicability of the presented approach was demonstrated by applying the theory to an ecosystem with extensive DOC utilization data. The analysis compares theoretical and experimentally measured utilization of amino acids within a eutrophic marine ecosystem (presented in the supplemental material). The experimental measurements were predicted with remarkable accuracy using the oxygen limited acclimation strategies presented in this study highlighting how basic physiological concepts can be applied to many different ecosystems.

This *in silico* systems biology study quantified assimilatory and dissimilatory biomass utilization by an aerobic heterotrophic archaeon, *Geoarchaeum* str. OSPB. The study provided a metabolic basis for comparing optimal DOC utilization patterns under different nutrient-limiting conditions, evaluating simultaneous versus sequential DOC metabolism, and investigating byproducts secreted during simulated growth and cellular energy production. These relationships describe limits for community interactions

between an autotroph and a heterotroph, such as the amount and type of autotrophic biomass needed to sustain the community. The novel approaches and findings of this study are applicable to a wide range of microbial ecosystems.

4-7 Materials and Methods

4-7-1 Metabolic Model Construction

The *Geoarchaeum* str. OSPB (NCBI taxon ID 1448933) genome sequence (152), which has been assembled from metagenomes of iron oxide mats sampled from One Hundred Springs Plain (129), was used to build the *in silico* heterotroph metabolic network. Incorporation of enzyme-catalyzed reactions into the model was based on protein homology (greater than 30 % identity) and gene annotations. Due to the incomplete nature of metagenome assemblies, missing genes were modeled as present, when necessary, using genes found in other archaea (97). The metabolic model of *Geoarchaeum* str. OSPB exhibits typical heterotrophic archaeal central carbon metabolism (e.g. TCA cycle, archaeal glycolysis/gluconeogenesis, and electron transport) and common biosynthetic and degradation pathways. The model did not predict auxotrophy for any of the major biomass components (i.e. amino acids or nucleotides). Acetate, formate, and formaldehyde were the only reduced carbon byproducts considered in the model, based on gene annotation. No genes were identified to support the inclusion of other reduced carbon byproducts. Enzyme resource costs accounted for the number of enzymes in the manually compressed reactions, with the exception of transporters and degradation enzymes (Supplemental Material).

Macromolecules and monomers were not allowed to be excreted by the heterotroph due to EFMA computational explosion. Peptides were assumed to be transported actively using an ABC transporter for trimeric peptides and depolymerized independent of cellular energy. RNA was transported and depolymerized in a cellular energy neutral reaction. Cellobiose was assumed to be transported actively using an ABC transporter and depolymerized independent of cellular energy, but required phosphorylation to enter central metabolism. Archaeal lipid was assumed to be transported actively using an ABC transporter and required complete degradation to monomers before polymerization into heterotroph biomass. Monomers were transported and degraded as homopolymers to avoid biases associated with differences in transport costs of polymers and oligomers.

4-7-2 Heterotroph Biomass Production

The heterotroph biomass production reactions were developed using macromolecular ratios as previously described (170), tailoring amino acid and nucleotide distributions to organism-specific genome data (Supplemental Material). The amino acid distribution was determined using the average distribution of all open reading frames in the *de novo* assembly of *Geoarchaeum* str. OSPB from the metagenome. The nucleotide distribution of RNA was represented by the weighted average nucleotide distribution of major rRNA subunits (i.e. 16S and 23S); rRNA pools are the largest fraction of RNA in the cell (e.g. approximately 80% in *E. coli* (170)). Autotroph lipid was represented by ether-linked digeranylfarnesyl glycerol phospholipid (94). Polysaccharide was represented by cellobiose. The maintenance energy was set based on glucose utilization by *Alicyclobacillus acidocaldarius* DSM 446, a thermoacidophilic gram-positive bacterium,

grown at a pH and temperature of 4.3 and 51°C, respectively (64). The maintenance energy requirements were determined assuming the use of a cytochrome bd oxidase, which resulted in 1 proton translocated per electron, because a cytochrome c oxidase system, which resulted in 2 protons translocated per electron, resulted in uncharacteristically high maintenance energies (i.e. 3.5 versus 5 moles cellular energy per Cmole biomass or 150 versus 215 mmoles cellular energy per g cell dry weight for a cytochrome bd oxidase versus cytochrome c oxidase, respectively, at a growth rate of 0.1 hr⁻¹). One mole of cellular energy was modeled as one mole of phosphodiester bonds broken. EFMs using cytochrome c oxidase were subsequently separated from EFMs using cytochrome bd oxidase in the primary analysis for clarity as all trends are similar, but scaled by the difference in protons transferred per oxygen respired (Supplemental Material).

4-7-3 Stoichiometric Analyses

Model reactions and metabolites were assembled using Microsoft Excel, transferred to CellNetAnalyzer version 2014.1 (118, 119) for formatting, and exported to RegEFMTool version 2.0 (112) for EFM enumeration. Gene regulatory rules mutually excluding cyclic reactions were incorporated with RegEFMTool to minimize futile cycles and validated against network subsets analyzed with EFMTTool version 4.7.1 (251, 253) without regulatory rules (Supplemental Material). ‘Metabolic check valves’ (i.e. pseudo-metabolites and irreversible reactions) were used in addition to gene regulatory rules to minimize futile cycling of nucleotides, which are often involved in reactions as coenzymes (Supplemental Material). FBA-based resource costs were calculated from the

doubleRobustnessAnalysis function of COBRA Toolbox (207) on the model exported using the *CNA2Cobra* function of CellNetAnalyzer. The resulting output was converted to resource costs by normalizing to either the specific growth rate or the specific cellular energy production rate, and compared to resource cost tradeoff curves produced from EFMs (Supplemental Material). An example of model construction, EFMA, and resource costs is available in the supplemental material (Figure 4-S1). A general maintenance energy of 3.5 moles cellular energy per Cmole of biomass was held constant for all simulations, which included contributions from growth and non-growth associated maintenance energy at a growth rate of 0.1 hr^{-1} . Computations were performed on a machine with two X5690 Intel Xeon processors and 120 GB of RAM.

4-8 Acknowledgements

This study was supported by the National Science Foundation Integrative Graduate Education and Research Training (IGERT) Program (DGE 0654336) and subcontracts (112443, 254840) from Pacific Northwest National Laboratory (Richland, WA) to MSU as part of the DOE-BER Foundational Science Focus Area in Microbial Community Design Principles. K.A.H. was also supported by NIH (1U01EB019416) and the Department of Chemical and Biological Engineering at Montana State University.

CHAPTER 5

MULTISCALE ANALYSIS OF AUTOTROPH-HETEROTROPH INTERACTIONS IN
A HIGH-TEMPERATURE MICROBIAL COMMUNITY

5-1 Contribution of Authors and Co-Authors

Author: Kristopher A. Hunt

Contributions: Assisted in concept development, helped with editing, generated figures, and wrote the paper.

Co-Author: Ryan deM. Jennings

Contributions: Assisted in concept development, helped with editing, generated figures, and wrote the paper.

Co-Author: William P. Inskeep

Contributions: Assisted in concept development and helped with editing.

Co-Author: Ross P. Carlson

Contributions: Assisted in concept development and helped with editing.

5-2 Manuscript Information Page

Kristopher A. Hunt, Ryan deM. Jennings, William P. Inskeep, Ross P. Carlson
The International Society for Microbial Ecology Journal

Status of Manuscript:

- Prepared for submission to a peer-reviewed journal
- Officially submitted to a peer-review journal
- Accepted by a peer-reviewed journal
- Published in a peer-reviewed journal

Multiscale Analysis of Autotroph-Heterotroph Interactions in a High-Temperature
Microbial Community

Kristopher A. Hunt^{1,2,3}, Ryan deM. Jennings^{1,3,4,5}, William P. Inskeep^{1,3,4}, Ross P.
Carlson^{1,2,3}

¹Center for Biofilm Engineering, Montana State University, Bozeman, MT

²Department of Chemical and Biological Engineering, Montana State University,
Bozeman, MT

³Thermal Biology Institute, Montana State University, Bozeman, MT

⁴Department of Land Resources and Environmental Sciences, Montana State University,
Bozeman, MT

⁵Current address: Department of Biology, Mercer University, Macon, GA

5-3 Abstract

Microorganisms constitute a major portion of the biosphere, driving planetary biogeochemical cycles through their metabolic activity. Fe(III)-oxide mats in the acidic, high-temperature springs of Yellowstone National Park are inhabited by relatively few microbial populations and geochemically well characterized. These mats contain excellent model communities for elucidating metabolic activity, trophic interactions and community structure-function relationships. A multiscale analysis integrated *in situ* and *in silico* data from enzymatic activity, cellular metabolism, community interactions, and ecosystem geochemistry to predict and quantify the functional limits of a relatively simple autotroph-heterotroph community. Metagenomic, transcriptomic, and metaproteomic data were used to reconstruct carbon and energy metabolisms of the autotroph and heterotroph. Standard and hybrid elementary flux mode and flux balance analyses of metabolic models predicted cellular- and community-level metabolic acclimations to simulated environmental stresses. *In situ* geochemical analysis, including

spatial oxygen profiles, total Fe(III)-oxide deposition rates, isotopic carbon fractionation, and mat biomass concentrations, were combined with the cellular models to predict community-level metabolic activity. Integration of these data, including the relative population abundances of autotrophs to heterotrophs (0.3-0.5), demonstrates the two studied mat communities function as predicted by the maximum power principle, maximizing total community growth rate as opposed to net community growth rate. Integration of multiscale data with practical ecological theory provide a basis for predicting autotroph-heterotroph interactions and community-level cellular organization.

5-4 Introduction

Microorganisms are the largest component of the biosphere and drive biogeochemical cycles through metabolic activity (62). Microorganisms commonly exist as biofilms or mats that contain numerous microenvironments due to the interplay between convection-, diffusion-, and microbial activity-induced chemical concentration gradients (8). Microbial communities within these mats are complex and temporally dynamic with numerous interacting microbial populations, nutrients, and energy sources. Microbial interactions are explicitly linked to key metabolic activities within a community. For example, competition is linked to nutrient and energy transfer. Most natural microbial communities, including soil and marine systems, are exceedingly complex due to the phylogenetic diversity and the array of nutrient and energy sources present, which precludes detailed functional analyses. Natural microbial communities comprised of

tractable nutrient inputs and populations are excellent systems to elucidate the concepts that organize microbial metabolism and interaction.

The phylogenetic diversity of microbial communities present in Fe(III)-oxide mats of acid-sulfate-chloride springs in Yellowstone National Park (YNP) is limited due to high temperature (65 – 75°C) and low pH (~ 3) (102, 103, 129). These biomineralizing communities are formed and inhabited by 5 to 7 distinct phylotypes, of which only a few are aerobes, including archaea from the order Sulfolobales (e.g. *Metallosphaera yellowstonensis* str. MK1) and novel archaeal groups (e.g. *Geoarchaeum* str. OSPB) (9, 10, 101, 103, 128–130). The aqueous and solid-phase geochemistry of two such environments in Beowulf and One Hundred Springs Plain (OSP) hot springs have been studied in detail (9, 14, 100, 102, 136) and provide bounds and context for modeling microbial community interactions. The primary electron donors that drive chemolithoautotrophy in Fe(III)-oxide microbial mats include Fe(II) (25 – 40 μM) and possibly reduced forms of sulfur (dissolved sulfide < 10 μM) and As(III) (25 – 30 μM) (128). The oxidation of Fe(II) coupled with the reduction of oxygen provides energy necessary for the reduction of carbon dioxide by *M. yellowstonensis*, a major autotroph in these mats (14, 103, 107, 127, 130). Oxygen consumption is diffusion-limited in Fe(III)-oxide microbial mats resulting in steep oxygen gradients from 60 μM to below 1 μM dissolved oxygen over a distance of 0.5 to 1 mm (14) and establishing a probable microbial competition for this electron acceptor. Genomic and mRNA data indicate both the major autotroph (e.g. *M. yellowstonensis*) and major heterotroph (e.g. *Geoarchaeum* str. OSPB) populations have heme Cu oxidases and utilize oxygen as an electron acceptor

(127, 129, 243). The observed localization of these microbial populations with respect to mat depth corresponds to measured oxygen concentrations (9).

Carbon dioxide fixation contributes 42 to 99 % of the total microbial biomass carbon in the Beowulf and OSP Fe(III)-oxide mats (107). Carbon dioxide fixation has been demonstrated by the major oxic zone autotroph populations (e.g. *M. yellowstonensis*) during *in vitro* isolate characterization and *ex situ* Fe(III)-oxide mat incubations (107). Metagenome analysis has established *Geoarchaeum* str. OSPB as one of the primary aerobic heterotrophs, often comprising 30 to 50 % of the total microbial community in the oxic zones of high temperature Fe(III)-oxide mats (9, 129). Chemolithoautotrophy by autotrophs and subsequent transfer of nutrients and energy to heterotrophs (e.g. *Geoarchaeum* str. OSPB) is hypothesized to drive major autotroph-heterotroph interactions along with competition for the primary terminal electron acceptor, oxygen. Sulfolobales viruses are highly represented in the metagenome sequence of these Fe(III)-oxide mats (101), which suggests that viral predation of *M. yellowstonensis* contributes to the turnover of autotrophic biomass and creates reduced carbon sources for heterotrophs.

Genome-enabled stoichiometric modeling is a powerful systems biology approach for examining metabolic acclimations to environmental stressors on size scales from individual cells to communities of interacting populations (30, 67, 242). Briefly, these approaches construct *in silico* representations of cellular metabolisms inferred from genome sequence analysis (262). The metabolic models define possible routes of cellular energy production and central carbon metabolism, including details of electron transport, the acquisition of carbon, and anabolic processes necessary to synthesize biomass. There

are two major classes of stoichiometric modeling, elementary flux mode analysis (EFMA) and flux balance analysis (FBA). EFMA identifies all enzymatically distinct and indecomposable routes through a metabolic network; these routes are termed elementary flux modes (EFMs) (216). EFMs, and non-negative linear combinations thereof, describe all possible physiologies independent of kinetic parameters, which makes EFMA well suited for evaluating energetic efficiencies of different electron donors, acceptors, and nutrient sources involved in biomass production. FBA identifies optimal solutions that, for example, maximize growth rate for a given substrate uptake rate. Stoichiometric modeling has been used to predict optimal genotypes in engineered systems (30, 267), interpret physiological behavior (33, 67), and evaluate the transfer of mass and energy between distinct populations in a natural phototrophic microbial community (242).

A primary goal in environmental microbiology is to understand and predict microbial behavior in communities, where interactions among different populations lead to emergent properties, such as enhanced productivity, stability, and robustness (123). Consequently, the objectives of this study were to 1) construct metagenome-enabled representations of enzymatic and abiotic reactions for the dominant microbial populations in the oxic zone of the Fe(III)-oxide mat with a focus on relevant electron donors and acceptors; 2) analyze cellular-level efficiency of electron donors and acceptors with respect to the production of biomass and cellular energy as a function of simulated environmental stress; 3) integrate individual cellular models into a community representation of autotroph-heterotroph interactions that considers competition for oxygen; and 4) perform a sensitivity analysis based on *in situ* parameters to determine

model limitations and identify future priorities for field measurements. This is the first multiscale analysis of a tractable, naturally occurring, archaeal mat community using metagenome-enabled stoichiometric interaction models and community-level competition models, which integrates data from the nanoscale of electron transport to microscale spatial oxygen profiles to mesoscale measurements of biotic Fe(III)-oxide deposition.

5-5 Materials and Methods

5-5-1 Genome Analysis and Model Construction

In silico stoichiometric models were constructed for autotroph *M. yellowstonensis* str. MK1 (NCBI Taxon ID 671065, GOLD ID Gi04920) and heterotroph *Geoarchaeum* str. OSPB (NCBI Taxon ID 1448933, GOLD ID Gi0000638). Elementally and electronically balanced reactions that represent carbon and energy metabolism were compiled based on genome sequence (Joint Genome Institute - Integrated Microbial Genomes (JGI-IMG) (152)). Prior genome analyses (101, 103, 107, 127, 129, 130), literature surveys, and the MetaCyc (37, 38) and KEGG databases (114) were cross-referenced during metabolic reconstruction and reactions were assumed to close pathway gaps that would have resulted in auxotrophy. The modeled macromolecular compositions of biomass for each population were based on previous reports (170). The publically available genomes were mined for monomer distributions of DNA, RNA, and protein, based on the GC content, the ribosomal subunits, and the average amino acid distribution of all protein encoding genes, respectively. Maintenance energy requirements were adjusted to calibrate the *M. yellowstonensis* and *Geoarchaeum* str. OSPB models to observed yields for

Acidithiobacillus ferrooxidans, a representative thermoacidophilic autotrophic Fe(II) oxidizing bacterium (224) and *Alicyclobacillus acidocaldarius*, a representative thermoacidophilic heterotrophic bacterium (64), respectively. Nongrowth associated maintenance energy (3.44 and 3 mmol ATP g biomass⁻¹ h⁻¹ for *M. yellowstonensis* and *Geoarchaeum* str. OSPB, respectively) was assumed to be 2.5 % of the growth associated maintenance energy (137.6 and 120 mmol ATP g biomass⁻¹ for *M. yellowstonensis* and *Geoarchaeum* str. OSPB, respectively) (67). Yield prediction using the above maintenance energies for other electron and carbon sources were within reported experimental error.

5-5-2 Computational Packages and Analyses

In silico stoichiometric models were constructed using CellNetAnalyzer version 2014.1 (118), and exported to RegEFMTool version 2.0 (112) to enumerate EFMs for each metabolic model. The macronutrients available to both populations were modeled as illustrated in Figure 5-1. Metabolic reconstructions for both populations, including reactions, genomic evidence, specific literature references, metabolites, stoichiometric balance, and gene regulatory rules can be found in the supplemental material. Community FBA was performed using functions from the COBRA Toolbox (207) and overall reactions obtained from EFMs deemed ecologically competitive based on resource utilization. Functions used for the community analysis are available in the supplemental material. All computations were processed on a machine with at most two Intel Xeon X5690 and 120 GB RAM.

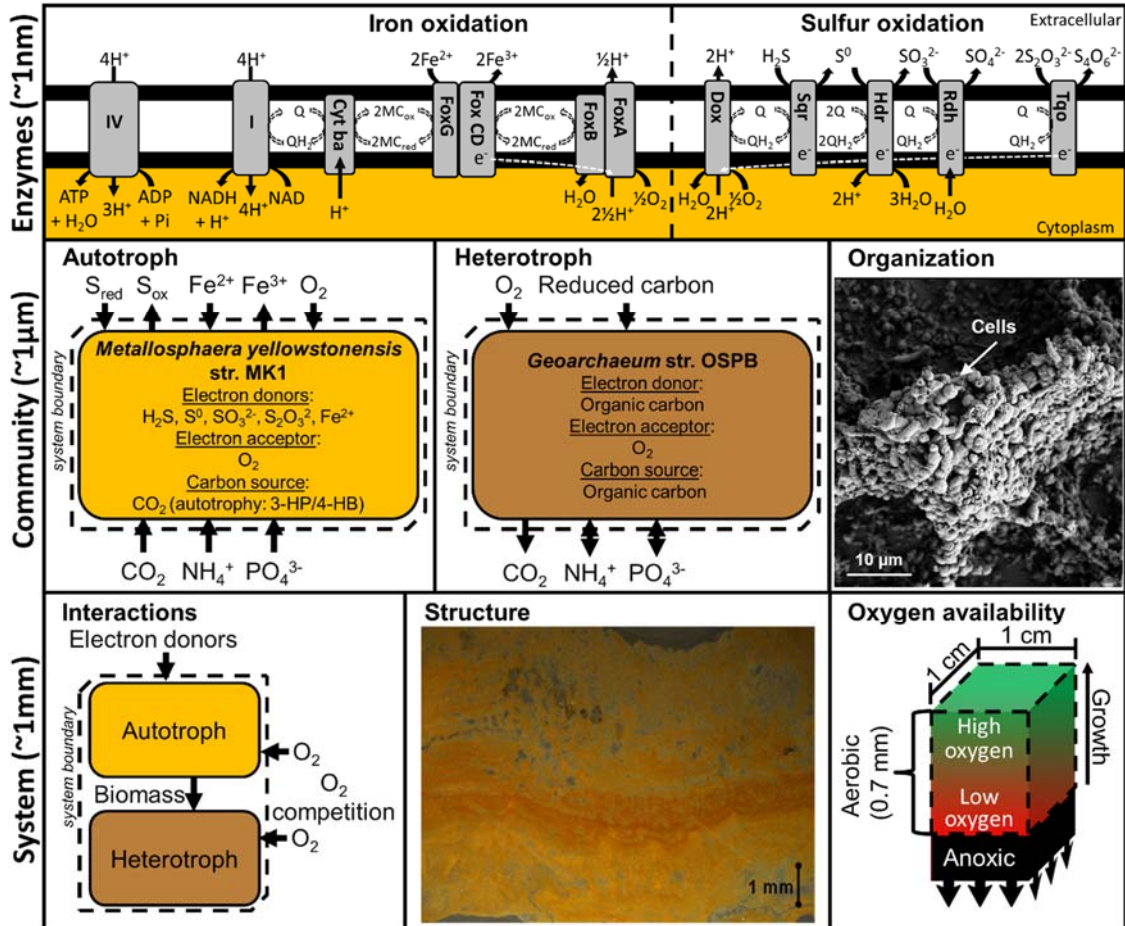


Figure 5-1. Conceptual representation of a multiscale metabolic interaction model between a primary autotroph (*M. yellowstonensis*) and heterotroph (*Geoarchaeum* str. OSPB) present in high-temperature (70-80°C) acidic Fe-oxide mats. Metagenomic and genomic annotations were used to identify enzymes in a cellular-level model to quantify electron donor and acceptor utilization for the primary autotroph and heterotroph to produce biomass and cellular energy using elementary flux mode analysis. Electron flow is indicated by white dashed lines. The nutrient requirements of each population provided a basis for comparison of the community-level modeling of autotroph-heterotroph interactions to *in situ* measurements including relative population abundances, isotope fractionation analyses, Fe(III) deposition rates, and oxygen flux into the mat. The system volume considered in these analyses is defined as the aerobic region of the mat (0.07 cm by 1 cm by 1 cm) moving vertically up as the mat grows. The effect of high and low oxygen associated with the top and bottom of the aerobic region provided bounds for electron acceptor stresses. Abbreviations: I - energy conserving NADH:ubiquinone oxidoreductase; Cyt ba - cytochrome/quinol oxidase; Fox AB and FoxCDG - Fe(II) oxidation complexes; MCO - Multicopper protein; Q - quinone; QH₂ - quinol; IV - ATP synthase; Sqr - Sulfur-quinone reductase; Hdr - Heterodisulfide reductase; Rdh - Rhodanese-related sulfur transferase; Tqo - Thiosulfate-quinone oxidoreductase; and Dox - Sulfur associated terminal oxidase.

5-6 Results

5-6-1 Enzyme-level Analysis of Chemolithoautotroph Electron Donors and Acceptors

M. yellowstonensis str. MK1 is a major primary producer in the Beowulf and OSP Fe(III)-oxide mats, which converts inorganic chemical species, such as carbon dioxide and ammonia, into the nutrients that sustain microbial community activity (107, 130). The interface between environmental electron donors and acceptors and cellular metabolism is mediated by the enzymes of electron transport. Metagenomic and genomic data were used to construct *in silico* representations of the electron transport network responsible for the oxidation of Fe(II) and sulfur species and the reduction of terminal electron acceptor, oxygen (Figure 5-1). Genomic and culturing evidence indicated that the chemolithoautotroph, *M. yellowstonensis*, can oxidize Fe(II) using the proteins encoded by the *fox* operon and can oxidize a variety of sulfur species, including sulfide, elemental sulfur, sulfite, and thiosulfate, to reduce the quinone pool to quinol (127). Electrons in the quinol pool can drive cellular energy production via oxidative phosphorylation through the reduction of oxygen using a high affinity heme copper oxidase. Alternatively, electrons in the quinol pool can reduce NAD⁺, enter central carbon metabolism, and be used to reduce inorganic carbon via the 3-hydroxypropionate / 4-hydroxybutyrate (3-HP / 4-HB) pathway (107). These electron transport pathways are hypothesized to provide the majority of mass and energy to the mat microbial community. The modeled central carbon metabolism of *M. yellowstonensis* included the

tricarboxylic acid cycle, gluconeogenesis, the pentose phosphate pathway via ribulose monophosphate (227), and the mevalonate pathway (156).

5-6-2 Cellular-level Analysis of Chemolithoautotroph Growth

Enzyme-level analysis of *M. yellowstonensis* electron donor and acceptor pathways was integrated into a cellular-level metabolism model to quantify the relationship between consumption of environmental resources and the production of biomass (Figure 5-1, supplemental material). A total of 6,337 elementary flux modes (EFMs) were identified that produce *M. yellowstonensis* biomass from the inorganic electron donors, Fe(II), sulfide, elemental sulfur, sulfite, or thiosulfate. Each EFM was plotted as a function of moles of electron donor and moles of electron acceptor required to produce one carbon mole (Cmole) of *M. yellowstonensis* biomass (Figure 5-2A). Biological systems often minimize use of limiting resources to produce biomass, providing an ecologically relevant basis for the *in silico* prediction of metabolic phenotypes (73). The moles of electron donor required to form a Cmole of biomass was lowest for sulfide and highest for Fe(II) (Figure 5-2A), which follows the thermodynamic favorability expected for the oxidation of these electron donors coupled to the reduction of oxygen (3). The oxidation of Fe(II) also required the most moles of oxygen consumed per Cmole of biomass produced of the considered electron donors. These relationships were consistent on both a moles of electron donor and moles of oxidized electrons basis (Figure 5-S1, Table 5-S1 for relevant degrees of reduction).

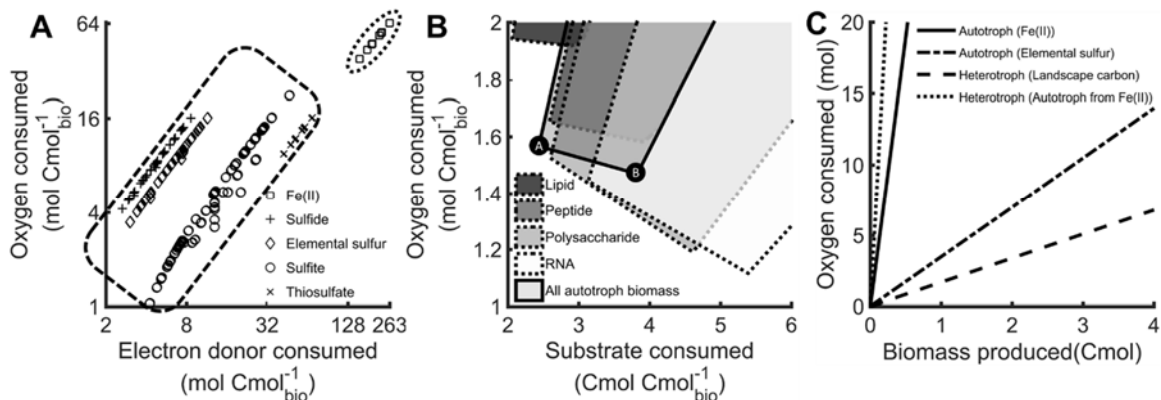


Figure 5-2. Electron donor and acceptor resource requirements to produce biomass. A) Moles of oxygen and electron donor required to produce one Cmole of *M. yellowstonensis* biomass for all biomass producing elementary flux modes. Clusters of electron donors, which include Fe(II), sulfide, sulfur, sulfite, and thiosulfate, are circled by dashed lines. B) Moles of oxygen and carbon substrate required to produce one Cmole of *Geoarchaeum* str. OSPB biomass for all biomass producing elementary flux modes. Specific carbon substrates include lipids, peptides, polysaccharides, RNA, and complete autotroph biomass (modified from prior work) (96). Optimizations for carbon- and oxygen-limited scenarios are marked by points A and B, respectively. C) Relative oxygen consumption plotted as a function of biomass production for autotroph, *M. yellowstonensis*, utilizing Fe(II) or elemental sulfur, and for heterotroph, *Geoarchaeum* str. OSPB, utilizing autotroph biomass generated from Fe(II) oxidation or exogenous sources of reduced organic carbon (i.e. landscape carbon).

5-6-3 Cellular-level Analysis of Organoheterotroph Growth

Genomic analysis of *Geoarchaeum* str. OSPB indicated this organism is an organoheterotroph with the metabolic potential to utilize a wide variety of reduced carbon species, including biomass macromolecules (i.e. lipids, peptides, polysaccharides, and nucleic acids). Community simulations assumed autotroph biomass was the major source of organic carbon for heterotrophs. Autotrophic biomass was either represented in simulations as individual macromolecular pools or as an aggregate substrate comprised of the autotroph in entirety. More than 24 million EFMs were identified for the production of *Geoarchaeum* str. OSPB biomass from *M. yellowstonensis* biomass and

macromolecules. A more detailed analysis of these relationships were elucidated in a prior report, a summary of which is presented here (96). Briefly, the efficiency of heterotroph biomass synthesis from individual macromolecular carbon sources was evaluated with respect to the Cmoles of carbon substrate or moles of oxygen required to produce a Cmole of biomass (x-axis of Figure 5-2B, Table 5-S2). Metabolism of the different carbon substrates demonstrated trade-offs between the Cmoles of carbon substrate and moles of oxygen required to produce *Geoarchaeum* str. OSPB biomass. The differences in substrate requirements were based on the different carbon backbones, entry points into central metabolism (i.e. substrate-level phosphorylation reactions) and the degrees of reduction (i.e. the number of oxidizable electrons in a molecule per either carbon in the molecule or one for inorganic molecules). For example, the high degree of reduction for lipids resulted in a greater oxygen requirement than for utilization of either peptides or polysaccharides. Due to modest levels of reduced carbon species available *in situ*, all subsequent results assumed the heterotroph metabolized autotroph biomass in its entirety.

The aggregate biomass substrate simulations established an important environment dependent mass and energy relationship between trophic levels. The synthesis of a Cmole of *Geoarchaeum* str. OSPB biomass required 2.4 Cmoles of autotrophic biomass to supply the required cellular energy and structural components for growth under carbon-limited conditions. This relationship was dependent on simulated environmental stress. Synthesis of a Cmole of *Geoarchaeum* str. OSPB biomass under oxygen-limited conditions requires 3.9 Cmoles of autotroph biomass (Figure 5-2B).

5-6-4 Cellular- and Community-level Analysis of Oxygen Requirements for Growth

Analysis of *in situ* oxygen profiles indicated that the productivity of the Beowulf and OSP mats is limited by electron acceptor availability (14). Autotrophic production of biomass required substantial oxygen and was governed largely by the degree of reduction of the electron donor and by the energy available between the electron donor and acceptor pair (Figure 5-2C, Table 5-S1). Autotroph biomass production via elemental sulfur oxidation required 3.5 moles of oxygen per Cmole of biomass produced while Fe(II) oxidation required an order of magnitude more oxygen, 38 moles oxygen per Cmole of biomass produced.

Heterotroph growth was analyzed on two bases: 1) a monoculture that utilized allochthonous organic carbon (i.e. organic carbon from the surrounding landscape), hereafter landscape carbon, and 2) a community member that utilized autotroph biomass. Heterotrophic growth on landscape carbon required 1.6 moles oxygen per Cmole of biomass produced; substantially less oxygen per Cmole of biomass produced than autotrophic growth from either elemental sulfur or Fe(II), a 54 and 96 % reduction in required oxygen, respectively (Figure 5-2C, Table 5-S2). The lower oxygen requirement for organoheterotrophic growth was due to utilization of reduced carbon substrates as both electron donors and anabolic precursors. By comparison, autotrophic metabolism required substantial energy to reduce carbon dioxide leading to high oxygen requirements.

In situ growth of heterotrophs is hypothesized to be dependent on the turnover of autotroph biomass. Oxygen requirements to produce heterotroph biomass increased

sharply when community-level interactions were included in the oxygen consumption calculations. *Geoarchaeum* str. OSPB requires 2.4 Cmoles of autotroph biomass to produce a Cmole of heterotroph biomass under carbon-limited conditions (Table 5-S2). The total requirement for oxygen to produce heterotroph biomass from Fe(II) growth autotrophic biomass also included the oxygen requirements of the autotroph consumed, which resulted in 93 moles of oxygen per Cmole heterotroph biomass (Figure 5-2C). Appropriate system boundaries that adequately represent population dependencies were critical for interpreting resource requirements for trophic level activities.

5-6-5 Community-level Resource Competition as a Basis for Community Structure

Oxygen is a required resource for both the autotroph and heterotroph to produce biomass and cellular energy, which results in resource competition. Additionally, the heterotroph consumes autotroph biomass as an electron donor creating a competitive interdependence. Relative population abundance and growth rates were used to quantify community structure as a function of measured environmental parameters. The system was analyzed using the oxic zone as a control volume, which was defined by the *in situ* depth of oxygen penetration, 0.7 mm (14) (Figure 5-1). The control volume, which was considered to be homogeneous to simplify the analysis, remained in the oxic zone by moving vertically as Fe(III)-oxide and biomass were deposited. Mass balances on the autotroph and heterotroph in the control volume were described as functions of growth rates, resource requirements, and biomass concentrations (Table 5-1 and 5-2). The system was assumed to be at steady state for the analyzed time scale of days to weeks, which

resulted in a relationship between the steady state specific growth rate of the autotroph and heterotroph. This relationship and the flux of oxygen into the mat were used to solve for steady state net and total community growth rates (Table 5-2). The rate of Fe(II) oxidation was then solved using the autotroph growth rate. These relationships (Table 5-2) provided a mechanism to examine the effects of oxygen flux into the mat and the relative population abundance of autotroph to heterotroph on individual and community growth rates (Figure 5-3). Three oxygen fluxes into the mat were analyzed, 50, 100, and 200 % of the average *in situ* flux ($10 \mu\text{mol cm}^{-2} \text{ day}^{-1}$) measured with microelectrodes (9). The relative population abundance of autotroph to heterotroph (0.3 – 0.5) was determined based on community composition analyses from Fe(III)-oxide mat samples collected from Beowulf and OSP springs (Table 5-S4).

The total community growth rate was constrained by competition for oxygen and the autotroph biomass requirement of the heterotroph. Total community growth rate increased with oxygen flux but decreased with relative population abundance of autotroph to heterotroph (Figure 5-3A). This increase in total community growth rate with increased heterotroph abundance was the result of an increase in autotroph biomass production, which was then consumed by the heterotroph, which indicated that a higher metabolic activity is obtained for a community with more heterotrophy. Additionally, two scenarios were examined for heterotrophic growth: 1) a carbon-limited scenario, which was expected to occur at the top of the mat where oxygen is plentiful, and 2) an oxygen-limited scenario, which was expected toward the bottom of the oxic zone (Figure 5-3). Optimized carbon usage by the heterotroph, hypothesized to occur under carbon

limitation, increased community growth rates by minimizing the autotroph biomass required to support the heterotroph. Conversely, optimized oxygen usage by the heterotroph caused an increase in heterotrophic carbon requirements that decreased net community growth rates by consuming more autotroph biomass (Figure 5-3B and C).

Two constraints bounded the feasible total biomass concentrations in the simulated community: 1) maximum specific growth rate and 2) oxygen flux into the mat. The relative specific growth rates of the autotroph and heterotroph were constrained by competition for oxygen and the heterotroph requirement for autotroph biomass as a reduced carbon source (Figure 5-3B). An increase in oxygen flux into the mat increased the specific growth rates of both populations for both carbon- and oxygen-limited simulations. Autotrophic growth consumed the majority of oxygen entering the mat. A decrease in the ratio of autotroph to heterotroph increased the specific growth rate for autotroph and decreased the specific growth rate for heterotroph, highlighting the increased need for autotrophic biomass to sustain an increase in heterotroph abundance. However, a finite maximum specific growth rate of 0.1 h^{-1} has been reported based on *in vitro* studies of *M. yellowstonensis* (95, 130). Therefore, an increase in autotroph specific growth rate established a minimum feasible total biomass concentration of $\sim 0.2 \text{ mg cm}^{-3}$, below which the maximum specific growth rate for the autotroph was exceeded, based on the *in situ* relative population abundances (autotroph to heterotroph) of 0.3 to 0.5 for the oxygen fluxes examined (Figure 5-3B). The net community growth rate was constrained by competition for oxygen and requirement for autotroph biomass by the heterotroph. Net community growth rate increased with oxygen flux and relative population abundance of

autotroph to heterotroph (Figure 5-3C). The observed relative population abundances established a maximum total biomass concentration of $\sim 2.0 \text{ mg cm}^{-3}$, above which the net community growth rate was negative. A negative net community growth rate indicated that the total biomass concentration in the mat was not sustainable and would decrease for the oxygen fluxes examined (Figure 5-3C).

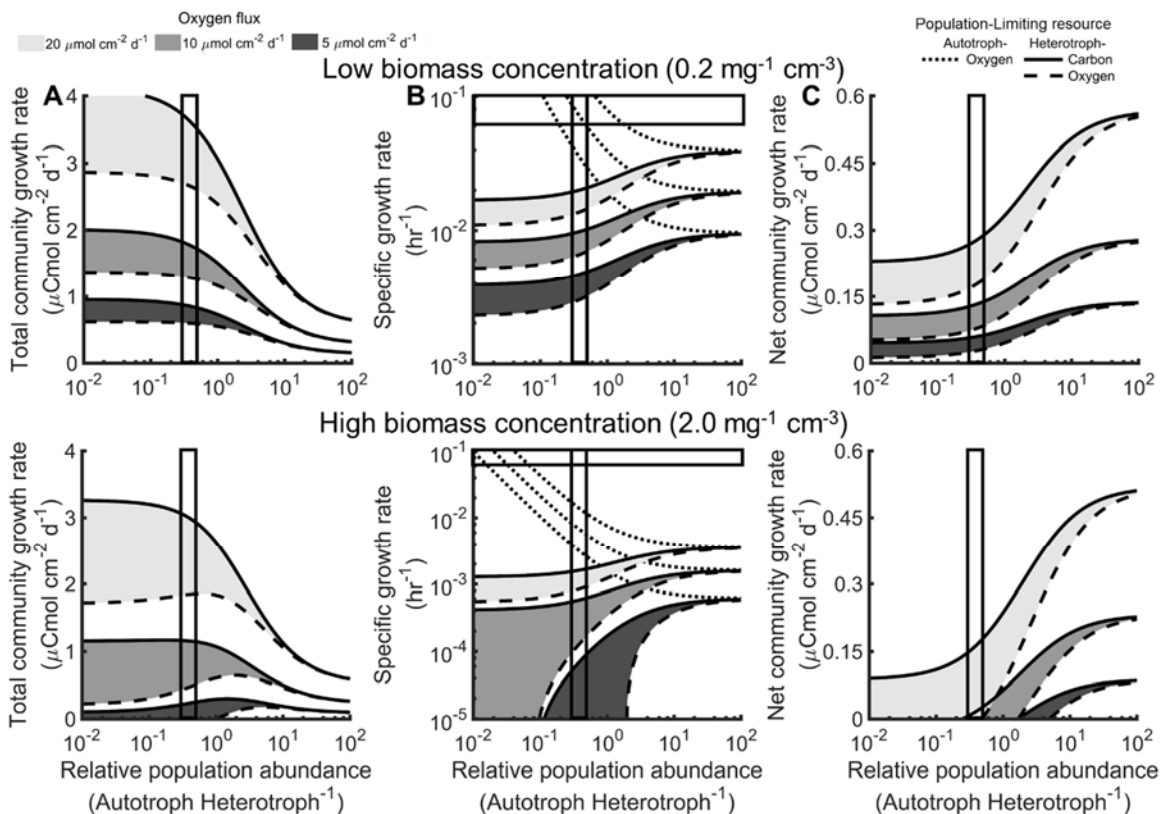


Figure 5-3. Total community growth rates (A), specific growth rates (B), and net community growth rates (C) predicted as functions of oxygen flux and the relative population abundance of autotroph and heterotroph. Three oxygen fluxes, 5, 10, and $20 \mu\text{mol}$ dissolved oxygen $\text{cm}^{-2} \text{ day}^{-1}$, represent 50, 100, and 200 %, respectively, of the average observed oxygen flux (9, 14). Each modeled oxygen flux was analyzed using the predicted most efficient metabolic strategy for oxygen (solid line) and carbon (dashed line). Active biomass concentrations were 0.2 (top) and 2.0 (bottom) mg cm^{-3} and an aerobic volume of $0.07 \text{ cm}^3 \text{ cm}^{-2}$ allowed for the prediction of specific growth rates under the examined conditions assuming a maximum growth rate of 0.1 h^{-1} . The observed relative population abundance was 0.3 and 0.5 (vertical dashed lines) based on sequence analyses.

Table 5-1. Values and descriptions of parameters used.

Variable	Description	Units	Value
μ_i	Specific growth rate of population “i”	$\text{g g}^{-1} \text{h}^{-1}$	Calculated
μ_{\max}	Maximum growth rate of any population	$\text{g g}^{-1} \text{h}^{-1}$	0.1
$Y_{\text{Fe/BA}} / Y_{\text{O}_2/\text{BA}}$	Iron / oxygen to produce autotroph biomass ^a	mol g^{-1}	6.7 / 1.63
$Y_{\text{Fe/EA}} / Y_{\text{O}_2/\text{EA}}$	Iron / oxygen to produce autotroph cellular energy	mol mol^{-1}	16 / 4
$Y_{\text{S/BA}} / Y_{\text{O}_2/\text{BA}}$	Sulfide / oxygen to produce autotroph biomass ^a	mol g^{-1}	0.112 / 0.181
$Y_{\text{S/EA}} / Y_{\text{O}_2/\text{EA}}$	Sulfide / oxygen to produce autotroph cellular energy	mol mol^{-1}	0.25 / 0.5
$Y_{\text{A/BH}} / Y_{\text{O}_2/\text{BH}}$	Carbon / oxygen to produce heterotroph biomass ^a	$\text{g g}^{-1} / \text{mol g}^{-1}$	2.29 (3.45) ^b / 0.057 (0.054) ^b
$Y_{\text{A/EH}} / Y_{\text{O}_2/\text{EH}}$	Carbon / oxygen to produce heterotroph cellular energy	$\text{g mol}^{-1} / \text{mol mol}^{-1}$	16.8 (8.2) ^b / 0.351 (0.376) ^b
X_i	Biomass concentration of population “i”	$\mu\text{g cm}^{-3}$	Calculated
X_{Tot}	Biomass concentration of all populations	$\mu\text{g cm}^{-3}$	0.2 – 2.0
GAM_A	Cellular energy to sustain autotroph growth	mmol g^{-1}	137.6
GAM_H	Cellular energy to sustain heterotroph growth	mmol g^{-1}	120
M_A	Cellular energy to sustain nongrowing autotroph	$\text{mmol g}^{-1} \text{h}^{-1}$	3.44
M_H	Cellular energy to sustain nongrowing heterotroph	$\text{mmol g}^{-1} \text{h}^{-1}$	3
	Depth of oxic zone	cm	0.07
V	Volume of aerobic region	cm^3	0.07
F	Volumetric flow rate	$\text{cm}^3 \text{d}^{-1}$	–
j_{O_2}	Oxygen flux into the mat	$\mu\text{mol cm}^{-2} \text{d}^{-1}$	Calculated
	Iron oxidation rate	$\mu\text{mol cm}^{-2} \text{d}^{-1}$	Calculated
RPA	Relative population abundance	g g^{-1}	0.3 – 0.5
RCF	Lowest fraction of biomass carbon from DIC	g g^{-1}	0.42 (0.67) ^d ^c
	Net community growth rate	$\mu\text{mol cm}^{-2} \text{d}^{-1}$	Calculated
	Total community growth rate	$\mu\text{mol cm}^{-2} \text{d}^{-1}$	Calculated
	Cmoles per gram of autotroph	Cmol g^{-1}	0.0430
	Cmoles per gram of heterotroph	Cmol g^{-1}	0.0438

^a: Growth associated resource requirements excluding nongrowth resource requirements

^b: Lowest carbon (oxygen) resource requirements

^c: Values for Beowulf (OSP) hot springs

^d: Constrained by mass balance

Table 5-2. Equations used to define mass balances, growth rates, and Fe oxidation rates in terms of other system variables (Table 1).

Parameter	Equation
Autotroph accumulation (dX_A/dt)	$\mu_A X_A V - (\mu_H Y_{A/BH} + M_H Y_{A/EH}) X_H V - F X_A$
Heterotroph accumulation (dX_H/dt)	$\mu_H X_H V - F X_H$
Oxygen flux into the mat (j_{O_2})	$(\mu_A Y_{O_2/BAF} + M_A Y_{O_2/EAF}) X_A V + (\mu_H Y_{O_2/BH} + M_H Y_{O_2/EH}) X_H V$
Total Community Growth Rate	$\mu_A X_A V + \mu_H X_H V$
Net Community Growth Rate	$\mu_A X_A V + \mu_H X_H V - (Y_{A/BH} \mu_H + M_H Y_{A/EH}) X_H V$
Fe oxidation rate	$(\mu_A Y_{Fe/BA} + M_A Y_{Fe/EA}) X_A V$

5-6-6 System-level Sensitivity Analysis of Electron Donors Using Multiscale, Hybrid EFMA+FBA Models

In situ measurements in natural ecosystems are complicated by numerous variables, including seasonal variation in weather, wind, precipitation, temperature, and inherent heterogeneity. A sensitivity analysis was performed on alternative electron donors, sulfide and landscape carbon, for the autotroph and heterotroph, respectively, using a multiscale, hybrid EFMA and FBA approach. The hybrid method first identified cellular-level, metabolic acclimation strategies to simulated environments, such as carbon and oxygen limitation, using EFMA (Figure 5-1). These cellular-level activities were summarized by overall resource transformation reactions in a model of community-level function using FBA. The system control volume was identical to the previous section, a 0.7 mm deep oxic mat. The system was assumed to be at steady state with the control volume moving vertically as Fe(III)-oxide and biomass were deposited. Three permutations of electron donors were considered (Scenario A-C in Figure 5-4).

Scenario A was the base scenario with Fe(II) and autotroph biomass as the only electron donors for the autotroph and heterotroph, respectively. Scenario A examined the capacity of the community to oxidize Fe(II) based on varying oxygen flux into the mat (Scenario A of Figure 5-4 and Figure 5-S3). The resource requirements for the heterotroph to produce biomass and cellular energy were bounded by the carbon- and oxygen-limited EFMA simulation results to approximate the high and low oxygen regions in the mat, respectively. The growth rate constraints used for FBA are described in Figure 5-4 and included site specific relative population abundance and total biomass concentration as well as the maximum specific growth rates of autotroph and heterotroph. The relative population abundance of autotroph to heterotroph was set to the observed values for each mat community as described in the previous section. The maximum specific growth rate of both populations were set to 0.1 h^{-1} based on *in vitro* studies of *M. yellowstonensis* (95, 130).

Scenario B built on scenario A by considering both autotroph biomass and landscape carbon as carbon sources for the heterotroph. The scenario examined the effect of 42 to 99 % of the biomass carbon in the system having been fixed by the Fe(II) oxidizing autotroph and the balance from exogenous, landscape origins (Scenario B of Figure 5-4 and Figure 5-S4) as indicated by isotopic analyses (107). The relative final biomass carbon fractions were simulated by fixing the relative population abundance of autotroph and autotroph consuming heterotroph relative to the landscape carbon consuming heterotroph.

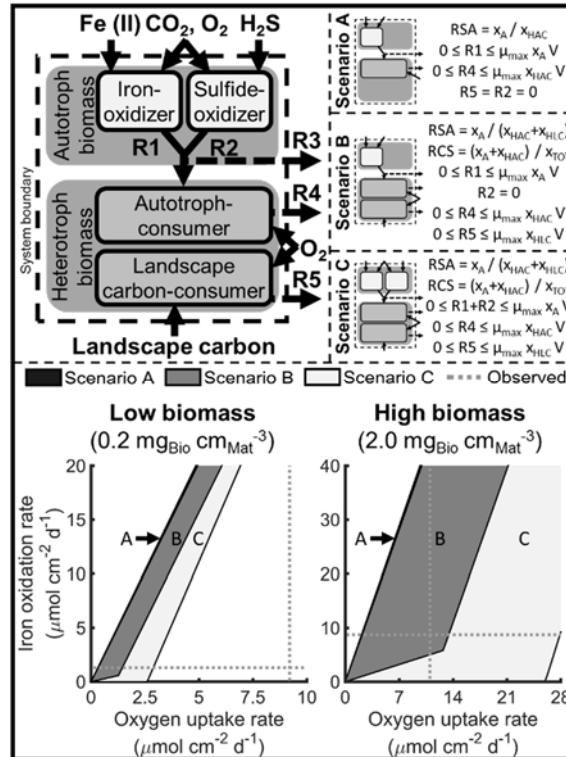


Figure 5-4. Sensitivity analysis of system behavior as a function of geochemical variation. Three scenarios predicted the metabolic space feasible for autotroph-heterotroph interactions. Scenario A analyzes a community composed solely of an Fe(II) oxidizing autotroph (R1) and autotroph-consuming heterotroph (R4), which are present at relative population abundances (RPA) of 0.3 and 0.5 autotroph to heterotroph. Scenario B analyzes the impact of additional carbon sources for the heterotroph (R5), such as the consumption of landscape carbon, which results in a relative carbon signature (RCS) of at least 42 % autotroph-originating biomass. Scenario C determines the impact of an additional electron donor (sulfide) for the autotroph (R2) in addition to a landscape carbon source for the heterotroph. The combined unconsumed autotroph is denoted by R3. Maximum specific growth rate (μ_{max}) and control volume (V) were set to 0.1 h^{-1} and 0.07 cm^3 , respectively. Yields were determined by elementary flux mode analysis, which were then used as inputs to flux balance analysis to predict rates. Averages of observed Fe(III)-oxide deposition (1.3 and $8.7 \mu\text{mol cm}^{-2} \text{ day}^{-1}$ for Beowulf and OSP, respectively) and oxygen uptake rates (9 and $11 \mu\text{mol cm}^{-2} \text{ day}^{-1}$ for Beowulf and OSP, respectively) ($9, 14$) (dotted lines) are shown for comparison.

Scenario C built on Scenario B by considering the alternative electron donor, sulfide, for the autotroph. Elemental sulfur and sulfide can be present at low levels in Fe(II) oxidizing mats. The presented analysis only considered sulfide as it was predicted to have

the largest impact (Scenario C in Figure 5-4 and Figure 5-S5). Autotrophy was allowed to vary between oxidation of sulfide exclusively and Fe(II) exclusively, but total autotroph maximum specific growth rate was limited to 0.1 h^{-1} . This treatment resulted in a tapering effect as a larger population of Fe(III) oxidizing autotroph decreased the possible sulfide oxidizing population (Scenario C in Figure 5-4 and Figure 5-S5).

The three scenarios were compared to the *in situ* measurements of Fe(III)-oxide deposition and oxygen flux into the mat. In a spring operating with a high biomass concentration, $\sim 2.0 \text{ mg cm}^{-3}$ as determined above, the average *in situ* measurements for OSP Spring were within the predicted metabolic space for both scenarios B and C. This overlap suggests that oxidation of landscape carbon or combinations of landscape carbon and reduced sulfur species could account for the additional aerobic activity in these microbial communities. However, in a spring with a low biomass concentration, $\sim 0.2 \text{ mg cm}^{-3}$ as determined above, the average *in situ* measurements for Beowulf Spring were outside the model predictions, which indicated additional activity or phenomena were occurring, such as oxidation of additional electron donors or incomplete iron deposition (see Discussion).

5-7 Discussion

The organization of microbial life spans orders of magnitude in scale, from enzymes and individual cells to populations and communities to macroscopic mats that can ultimately impact planetary biogeochemical cycling (Figure 5-1). Metagenomic and genomic data from thermal Fe(III)-oxide mats in YNP were used to identify enzymes

responsible for electron donor/acceptor functionality of the autotroph and heterotroph community members within the oxic zone. The enzyme-level data were integrated into stoichiometric models to quantify cellular-level resource requirements to produce biomass and cellular energy (Figure 5-2). Metagenomic data and geochemical analyses established *in situ* relative population abundances and total biomass concentrations, which provided context for the community-level modeling. *In situ* measurements of oxygen flux into the mat using microelectrodes (9, 14), system carbon sources bounded by isotope analysis (107), and Fe(III)-oxide accretion rates from long-term temporal studies (9) (Figure 5-4) were integrated with the *in silico* models and used to quantify the total contribution of these populations to the biogeochemical activity of the natural Fe(III)-oxide mats. Autotroph-heterotroph interaction modeling indicated that the majority of the measured oxygen flux *in situ* is consumed by the autotrophic populations (e.g. *M. yellowstonensis*). The autotrophic dominance of oxygen consumption was due to the relatively low energy content of the electron donor (i.e. Fe(II)) and the high energy demands of carbon fixation. Conversely, heterotrophic activity had little effect on predicted Fe(II) oxidation rates as a function of oxygen consumed (Scenario A and B of Figure 5-4). The effect of the heterotroph was to provide a sink for autotroph biomass, which increased total community growth rate yet decreased net community growth (Figure 5-3).

The turnover of autotroph biomass is important to nutrient and energy flux through the mat community. It is hypothesized that viral predation and subsequent cell-lysis represents a major mechanism of carbon cycling in microbial communities (198, 240).

Numerous Sulfolobales viruses have been identified (191, 280), and more specifically, the genome of *M. yellowstonensis* contains an extremely high abundance of transposases and integrases (126), which suggests that viruses are very important in the life-cycle of these organisms. Assembled sequence of Sulfolobales viruses has been obtained from Fe(III)-oxide microbial mats (101), and the lysis of archaeal cells by viruses has been observed *in situ* using scanning electron microscopy (Inskeep, unpubl). Virus induced turnover of autotroph biomass enables production of heterotroph biomass leading to increased total community activity (Figure 5-3A). The observed *in situ* relative population abundances suggest that a community naturally maximizes its metabolic activity in agreement with the maximum power principle (11). This theory postulates that communities that maximize their metabolic activity (energy acquisition from the environment) are more competitive than communities that exhibit lower rates of energy acquisition. This theory, when combined with metagenomics data, provides an *a priori* predictive tool for community structure-function relationships.

The principle of ecosystem carrying capacity governs the minimum and maximum biomass concentrations in the studied, oxic autotroph-heterotroph communities. Carrying capacity, defined here as the maximum sustainable population supported by an ecosystem, can predict or describe the maximum population of both autotrophs and heterotrophs. Community analysis of the carrying capacity of Fe(III)-oxide mats indicated that a minimum and maximum total biomass concentration exist where the production of autotroph biomass and oxygen flux into the mat become limiting, respectively. The minimum predicted biomass carrying capacity of 0.2 mg cm^{-3} (Figure

5-3B) corresponds remarkably well with the calculated biomass concentration in Beowulf Spring (0.3 mg cm^{-3}) (Table 5-S5), which suggests that this mat system is governed by the maximum specific growth rate of autotroph. In this scenario of autotroph growth limitation, an increase in heterotroph population would require an autotroph specific growth rate beyond the value observed experimentally *in vitro* and is therefore not feasible. Conversely, the maximum predicted biomass carrying capacity of 2.0 mg cm^{-3} (Figure 5-3C) corresponds well with the calculated biomass concentration of OSP Spring (2.2 mg cm^{-3}) (Table 5-S5), which suggests that this mat system is governed by the flux of oxygen. Oxygen limitation is similar to autotroph limitation; however, instead of requiring the autotroph to grow faster than biologically possible, there is a lack of oxygen to support growth. The presented modeling framework combined with the classic ecological theory provide a very powerful tool for predicting community structure and function and can be applied to more complicated or less characterized systems than the Fe(III)-oxide mats of YNP.

A variety of additional oxygen consuming processes in the Fe(III)-oxide mats could explain the higher amounts of measured oxygen flux relative to the measured Fe(III)-oxide deposition rates (Figure 5-4). For example, at least three major oxygen consuming processes (i.e., additional electron donors) could account for the higher measured oxygen flux than predicted for Fe(II) oxidation alone; these include the oxidation of reduced sulfur species, As(III), and reduced carbon from landscape carbon sources. Small amounts of sulfide ($< 5 \text{ } \mu\text{M}$) and occasional flocs of elemental sulfur can be present in Fe(III)-oxide depositing zones of acid-sulfate-chloride springs, such as Beowulf Spring,

where high sulfide at discharge results in the deposition of elemental sulfur upstream of Fe(II) oxidation. Additionally, *Geoarchaeum* str. OSPB was modeled as the sole heterotroph in this system; however, additional aerobic heterotrophs have been observed (e.g. NAG2). In the high biomass concentration systems, such as OSP Spring based on Fe(III)-oxide density (Table 5-S5), reduced sulfur and carbon may account for the oxygen consumed. However, in the low biomass concentration mats, such as Beowulf Spring, these alternatives do not adequately describe oxygen consumed per Fe(III)-oxide deposited. The populations of NAG2 and *Geoarchaeum* str. OSPB have been shown to replace each other as functions of geochemical parameters, often resulting in a constant relative population abundance of autotroph to heterotroph. Preliminary metabolic potential analyses suggest that these organisms may serve same role in their respective communities. Additionally, 25 to 30 μM As(III) is present in many acid-sulfate-chloride springs, and has been shown to be oxidized to As(V) by the other major autotroph in these mats (*Hydrogenobaculum* spp.) (85); elemental analysis of the poorly-crystalline Fe(III)-oxide phases indicates that significant amounts of As(V) is incorporated into the solid phase (102). Any of these reduced species would result in consumption of additional oxygen if oxidized. It has been shown that reduced sulfur species are important to *Hydrogenobaculum* spp. up-gradient of Fe(III)-oxide deposition zones (43), and that *de novo* sequence assemblies of these populations contain necessary sulfur and As(III) oxidation pathways that utilize oxygen as an electron acceptor (243). *Hydrogenobaculum* spp. account for a minority of the total community in mature (0.5 – 1 cm) Fe(III)-oxide mats (101, 103, 107) but are a dominant population in early stages (0 – 14 d) of Fe(III)-

oxide mat assembly (9, 146). The effects of As(III) oxidation by this population on the total oxygen consumption in Fe(III)-oxide mats warrants further study.

The integration of *in situ* measurements and *in silico* analyses provide unprecedented approaches to understand microbial interactions and community function. The analyses applied in the presented study provide context for the relative population abundance of autotrophs and heterotrophs, the minimum and maximum biomass concentration of microbial mats, and the possible effects of additional electron donors. Multiscale analyses, such as those presented here, provide context for the emergent properties of complex organization, such as increased stability and robustness, highlighting shared principles across microbial and mesoscale ecology.

5-8 Acknowledgements

This study was supported by the National Science Foundation Integrative Graduate Education and Research Training (IGERT) Program (DGE 0654336) and subcontracts (112443, 254840) from Pacific Northwest National Laboratory (Richland, WA) to MSU as part of the DOE-BER Foundational Science Focus Area in Microbial Community Design Principles. K.A.H. was also supported by NIH (1U01EB019416) and the Department of Chemical and Biological Engineering at Montana State University, and W.P.I. appreciates support from the Montana Agricultural Experiment Station (Project 911300).

CHAPTER 6

THEORETICAL AND PRACTICAL LIMITATIONS OF HYDROCARBON
PRODUCTION FOR A CELLULOLYTIC, ENDOPHYTIC FILAMENTOUS FUNGUS

6-1 Contribution of Authors and Co-Authors

Author: Kristopher A. Hunt

Contributions: Assisted in concept development, helped with editing, generated figures, and wrote the paper.

Co-Author: Natasha D. Mallette

Contributions: Assisted in concept development, helped with editing, generated figures, and wrote the paper.

Co-Author: Brent M. Peyton

Contributions: Helped with editing.

Co-Author: Ross P. Carlson

Contributions: Assisted in concept development and helped with editing.

6-2 Manuscript Information Page

Kristopher A. Hunt, Natasha D. Mallette, Brent M. Peyton, Ross P. Carlson
Metabolic Engineering

Status of Manuscript:

- Prepared for submission to a peer-reviewed journal
- Officially submitted to a peer-review journal
- Accepted by a peer-reviewed journal
- Published in a peer-reviewed journal

THEORETICAL AND PRACTICAL LIMITATIONS OF HYDROCARBON PRODUCTION FOR A CELLULOLYTIC, ENDOPHYTIC FILAMENTOUS FUNGUS

Kristopher A. Hunt^{1,2}, Natasha D. Mallette^{1,2,†}, Brent M. Peyton^{1,2}, Ross P. Carlson^{*1,2}¹Center for Biofilm Engineering, Montana State University, Bozeman, MT²Department of Chemical and Biological Engineering, Montana State University, Bozeman, MT[†] Current address: Department of Chemical, Biological, and Environmental Engineering, Oregon State University, Corvallis, OR6-3 Abstract

Hydrocarbons have a variety of ecological and industrial uses from chemical signals and antifungal/antibacterial agents to fuels and specialty chemicals. The hydrocarbon production potential of the cellulolytic, endophytic fungus, *Ascocoryne sarcoides*, was determined using genome-enabled stoichiometric modeling. In silico analysis identified available hydrocarbon production routes including both anabolic- and catabolic-based strategies and determined correlations between the type and size of molecules and culturing parameters, such as oxygen and carbon limitation. The analysis quantified the limits of wild-type hydrocarbon production from cellulose-based substrates and identified influential metabolic engineering targets, including cellobiose phosphorylase (CP) and cytosolic pyruvate dehydrogenase complex (PDHcyt). CP and PDHcyt activity increased the theoretical hydrocarbon production limits most substantially under anoxic conditions where the metabolism was less efficient at extracting energy from substrate. Incorporation of both modifications resulted in near complete recovery of electrons. The in silico framework was integrated with in vitro fungal batch growth experiments to support predictions of electron acceptor limitation and hydrocarbon production. This is

the first reported metabolic reconstruction of an endophytic filamentous fungus and describes pathways for both specific and general hydrocarbon production strategies applicable to many eukaryotic hosts.

6-4 Highlights

Oxygen limitation decreased theoretical hydrocarbon yield due to energy limitation.

Theoretical hydrocarbon yield increased with chain length.

Theoretical hydrocarbon yield decreased with metabolic compartmentalization.

6-5 Introduction

Ascocoryne sarcooides is an endophytic fungus isolated from the *Eucryphia cordifolia* tree in Chilean Northern Patagonia and has been shown to produce hydrocarbons (233). Endophytes are ubiquitous components of plant microbiomes that may contribute advantageous and essential functions to their plant host. Examples of such functions are tolerance to nutrient limitation, water, temperature, and salinity stresses as well as protective strategies against animals and insects (98). *A. sarcooides* produces a variety of hydrocarbons hypothesized to play a role in a symbiotic relationship with its host (237, 269). The hydrocarbons *A. sarcooides* produces could also be used as renewable chemicals, such as biofuels and flavor compounds, made from low cost substrates, including cellulose and other agricultural wastes (2, 76, 80, 150, 151, 238). This consolidation of hydrocarbon production and diverse range of feedstocks makes *A. sarcooides* a candidate for bioprocessing. Hydrocarbon production by *A. sarcooides* has

been studied *in vitro* at the organism-level, but, to the best of the authors knowledge, there have been minimal molecular-level pathway examinations and no *in silico* reconstructions of its metabolism (76).

The production of hydrocarbons, defined here to include oxygen or sulfur containing organic compounds, by eukaryotes has been shown to occur via four primary metabolic routes: traditional fermentation, fatty acid production, fatty acid degradation, and Ehrlich pathways. Traditional fermentation pathways produce simple hydrocarbons as central metabolism byproducts, such as ethanol. Fatty acid biosynthetic machinery can produce a plethora of hydrocarbons as a function of the initiating molecule, number of elongation cycles, cycle completeness, and molecules incorporated, such as octane, octanol, or hexanoic acid (213). Linoleic acid degradation has been shown to produce a variety of C₈ to C₁₀ hydrocarbons through the oxidation of double bonds and subsequent redox reactions, such as 1-octen-3-ol (76). Ehrlich pathways have been shown to produce a variety of C₄ to C₁₀ hydrocarbons through the deamination and decarboxylation of amino acids or their derivatives, such as 1-methyl-3-butanol (89). Ehrlich pathways have been linked to nitrogen cycling in organisms. Additionally, alcohols and acids from the described routes can be used in the production of esters, further expanding the range of possible hydrocarbons (213). However, a thorough analysis of the yields and energy production during growth with these hydrocarbon producing strategies has not been done in an endophytic fungus.

Acetyl-CoA is a fundamental building block for many hydrocarbons and their biomass precursors, such as lipids and amino acids. Production and degradation of fatty acids in

eukaryotes is chemically partitioned through the use of acyl carrier protein (ACP) and coenzyme A (CoA), respectively, to activate intermediates. These functions are also spatially partitioned through localization of related enzymes between the cytosol, mitochondria, and peroxisome (249). The pathways used to generate acetyl-CoA by various organisms correlate with their metabolic strategy. In oleaginous yeast, acetyl-CoA is generated from the pyruvate dehydrogenase complex in mitochondria where the acetyl-CoA can be oxidized in the TCA cycle and electron transport chain (Figure 6-1). In non-oleaginous yeast, acetyl-CoA is generated in the cytosol through pyruvate decarboxylase, acetaldehyde oxidase, and acetyl-CoA synthetase (221). This compartmentalization creates a natural control point for balancing mitochondrial generated cellular energy and cytosolic production of fatty acids from carbon and energy. Acetyl-CoA is transported between the mitochondria and the cytosol using primarily three routes in fungi with each having different cellular energy requirements (Figure 6-1). The ATP-citrate lyase (Acl) route cleaves citrate produced in the mitochondria into oxaloacetate and acetyl-CoA in the cytosol (12, 18, 99). The oxaloacetate can also be oxidized to malate providing reducing equivalents. The acetate route converts acetyl-CoA to acetate in the mitochondria, transports the acetate to the cytosol, and reactivates it via acetyl-CoA synthetase (Acs). The acyl-carnitine route uses acyl-carnitine transport to move acyl groups between the cellular compartments for β -oxidation, but has been shown to be insufficient as a major route for transporting acetyl-CoA from mitochondria to the cytosol (99).

Stoichiometric modeling identifies targets for optimization and ecologically relevant phenotypes in a metabolic network. There are two primary forms of stoichiometric analysis; both require an initial estimate of the metabolic network, commonly generated from genomic or other experimental data (121, 175, 195, 262). The metabolic network is represented as linear equations that define metabolite conversions. Flux balance analysis (FBA) uses these linear equations, bounds on reaction rates, and an objective function (i.e. maximizing biomass production) to obtain an optimal flux distribution (267). Elementary flux mode analysis (EFMA) uses the same set of linear equations to obtain the smallest set of genetically distinct flux distributions known as elementary flux modes (EFMs). The complete set of EFMs represents all possible steady-state metabolisms of the network through nonnegative linear combinations allowing for predictions of maximum yield and the effect of genetic modifications (121, 140, 208).

The presented work is the first metabolic reconstruction of an endophytic fungus, an important natural component of plant microbiomes and potential consolidated bioprocess catalyst, and maps both specific and general hydrocarbon production strategies applicable to many eukaryotes. The results define the stoichiometric limitations of hydrocarbon production with respect to i) the theoretically optimal chain length of products from cellulose, ii) the theoretical limits of cellulosic feedstocks considering genetic modifications, and iii) relative effect of different cytosolic acetyl-CoA generating pathways and iv) provide context for the impact of culturing conditions on hydrocarbon profiles. Comparison of theoretical limits with observed production trends provide the

foundation for hydrocarbon production with direct applicability to biotechnology, microbiotics, and biochemical and genetic regulatory theory.

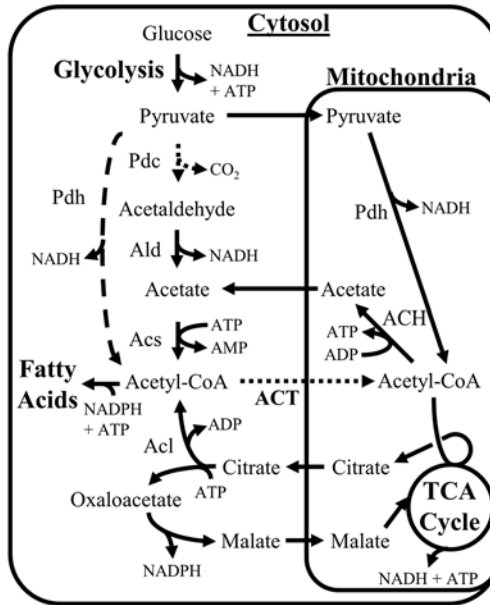


Figure 6-1. Overview of acetyl-CoA metabolism. Common pathways in fungi are shown in solid arrows using either a mitochondrial pyruvate dehydrogenase and transport or a cytosolic pyruvate decarboxylase (not found in *Ascooryne sarcooides* NRRL 50072). A cytosolic pyruvate dehydrogenase (dashed line) may increase hydrocarbon production, but has not been found in fungi. Enzyme abbreviations: Ack – acetate kinase, Acl – acetyl-CoA lyase, Acs – acetyl-CoA synthetase, ACT – acylcarnitine transport, Ald – aldehyde dehydrogenase, Pdh – pyruvate dehydrogenase, Pdc – pyruvate decarboxylase.

6-6 Material and Methods

6-6-1 Culture Medium and Conditions

Ascooryne sarcooides (NRRL 50072) was grown in 20 mL sealed test tubes on a minimal cellulose medium consisting of (per liter): KH_2PO_4 (13.6 g), sodium carboxyl methylcellulose (CMC) (10 g), $\text{NaH}_2\text{PO}_4 \cdot 2\text{H}_2\text{O}$ (2.75 g), NH_4Cl (1.76 g), $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ (0.86 g), $\text{Ca}(\text{NO}_3)_2 \cdot 4\text{H}_2\text{O}$ (0.28 g), yeast extract (0.05 g), and trace salts: KCl (60 mg), KNO_3 (80 mg), FeCl_3 (2 mg), MnCl_2 (5 mg), ZnSO_4 (2.5 mg), H_3BO_3 (1.4 mg), and KI

(0.7 mg). The initial medium pH was 4.9 and was inoculated with a 7-day culture prepared from Microbank™ Bead frozen stocks (-80°C). Three initial headspace oxygen concentrations were examined, 21%, 10%, and 7%. 10% and 7% oxygen concentrations were achieved by purging with N₂ and initial concentrations were confirmed by gas chromatography. Culture tubes were incubated at 21°C at a fixed 45° angle on rotary shakers at 150 rpm. In addition, each condition was repeated in duplicate as 10 mL volumes in 20 mL solid phase microextraction (SPME) vials.

Biomass was determined in triplicate by cell dry weight (CDW) as described previously (151) and by optical density at 600 nm (OD₆₀₀). Ammonium was assayed with Hach Method 10031 for High Range Nitrogen (Hach Company, Loveland, CO). CMC, glucose, acetate, and ethanol were measured by high performance liquid chromatography (HPLC) on an Agilent 1200 with an Aminex HPX-87H ion exclusion column at 45°C with 0.005 M H₂SO₄ eluent. Concentrations of oxygen were measured with gas chromatography (GC) by 1 mL direct injection of culture tube headspace into a SRI8610C (SRI Instruments, Torrance, CA) with a 6 ft molecular sieve 13x column with helium carrier gas. Volatile organic carbons (VOCs) were measured in the headspace of 20 mL SPME compatible vials (HS-SPME). HS-SPME was performed as described in Mallette et al. 2014 (150).

6-6-2 Construction of Fungal Endophyte Metabolic Model

An *A. sarcooides* NRRL 50072 metabolic model was constructed from the genome annotations and transcriptome data available at <http://asco.gersteinlab.org> (76) and

incomplete pathways were filled using assumed reactions based on completeness of a pathway and reference organisms (supplementary material). Enzyme localization was modeled based on predictions from TargetP 1.1 (57) and considered two compartments, the mitochondria and cytosol. The macromolecular composition of biomass was modeled after *Thielavia terrestris* as a representative filamentous fungus (83). Protein and RNA monomer composition was determined by the average distribution of all open reading frames in the genome and the ribosomal subunits, respectively. DNA monomer composition was modeled to represent the guanine-cytosine content of the genome. Micronutrients were assumed non-limiting and nitrogen supplied by ammonium, sulfur by sulfate, and phosphorous by phosphate in accordance with published *A. sarcooides* NRRL 50072 hydrocarbon studies (2, 76, 80, 151, 238). Oxygen was required for the production of essential fatty acids and therefore biomass.

6-6-3 Stoichiometric Analyses

The stoichiometric model was compiled as described in the following section using Microsoft EXCEL, converted to appropriate formats using CellNetAnalyzer version 2014.1 (118, 119), and flux distributions were enumerated using either RegEFMTool version 2.0 (111, 112) or the constraints based reconstruction and analysis (COBRA) toolbox (207). Gene regulatory rules excluding the simultaneous use of mutually exclusive reactions were applied to minimize the number of physiologically irrelevant possibilities (supplemental material). All computations were performed on a machine with a maximum of 2 Intel Xeon X5690s and 120GB of memory.

6-7 Results

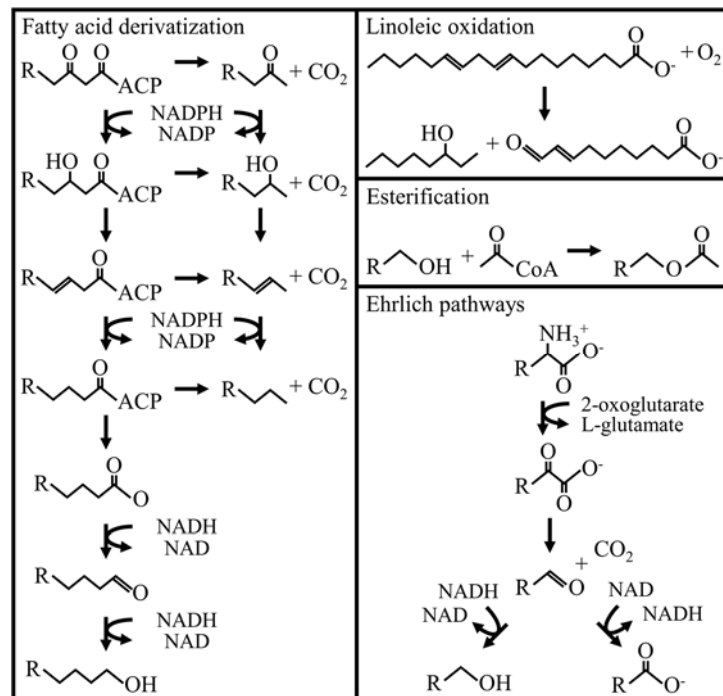
6-7-1 Modeling of Carbon
Uptake and Product Formation

Figure 6-2. Overview of hydrocarbon producing pathways. Four pathways were considered: fatty acid derived products, linoleic oxidation products, esters of acetate, and Ehrlich pathway products. ACP refers to the acyl carrier protein associated with fatty acid synthesis.

A. sarcoides has been shown to utilize cellulose as a carbon substrate. Fungi have been shown to degrade insoluble cellulose to soluble oligosaccharides (i.e. 1–5 glucose units) extracellularly. The energetic impact of oligosaccharide chain length was bounded by modeling the carbon and electron sources cellopentaose and glucose, which enter the cell through ABC transporters, depolymerize independent of phosphate, and are finally activated by a hexose kinase. The generation and transportation of cytosolic acetyl-CoA was based on the annotated genes for the Acl and Acetate routes. A Pdc was not found in

the genome and the acyl-carnitine reaction is not a major route for transporting acetyl-CoA from mitochondria to the cytosol (99). Reactions that led to the transport of acetyl-CoA were modeled as irreversible under normal physiological conditions.

Table 6-1. Commonly reported products of *A. sarcooides* NRRL50072^a by hypothesized pathway.

Fatty acid derived	Ehrlich products	Linoleic acid derived
1-butanol	2-methyl-1-propyl acid, alcohol, and esters	3-octanone
Hexanoate / 1-hexanol		1-octen-3-ol
1-heptanol	3-methyl-1-butyl acid, alcohol, and esters	Miscellaneous
Octane		
2-nonanone	2-phenylethyl alcohol and esters	Propyl-cyclopropane ^b
C ₂ , C ₅ -C ₁₀ alkyl acetates	Phenyl methanol	2-pentyl-furan ^b

^a All hydrocarbons were reported from at least three separate sources (6–10, this study)

^b Excluded from model

Hydrocarbon production reactions were built using genomic and product profile data. A consensus list of hydrocarbon compounds was assembled by collating any chemical species reported by at least three published *A. sarcooides* sources (Table 6-1). These hydrocarbons were modeled using fatty acid production, linoleic acid degradation, and Ehrlich pathways (Figure 6-2). Fatty acid metabolism can produce a wide variety of hydrocarbons, subject to the initiating molecule (e.g. acetyl-CoA or propanoyl-CoA), monomers incorporated during elongation (e.g. C₂ and C₃ monomers), and final functional group of the product upon excretion. These variations in components and pathways can have a combinatorial effect (135, 211, 213) and explain most hydrocarbons reported for *A. sarcooides* (Table 6-1). Only acetyl- and propanoyl-CoA were considered here to initiate the hydrocarbon chains with only acetyl-CoA as an elongation monomer to avoid excessive computational complexity, which resulted in either even- or odd-chained hydrocarbons depending on the number of decarboxylations. Esterification of

hydrocarbons resulting from the aforementioned pathways accounts for reported esters (Table 6-1). In total, 161 different hydrocarbons ranging from C₃ to C₂₀ representing eight different functional groups were modeled.

6-7-2 *In silico* Cellular Energy

Cellular energy production is essential for all life. The most competitive phenotypes for *A. sarcooides* that produce cellular energy were examined for metabolite secretion using ecological resource allocation theory, which postulates that a cell would maximize the value of a limiting resource by minimizing resource cost per product (31). The carbon and oxygen resource costs to produce cellular energy were determined for each EFM and were defined as the Cmoles of carbon source and moles of oxygen, respectively, consumed to produce a mole of cellular energy (ATP phosphodiester bonds broken). EFMs that minimized carbon and oxygen resource costs form a trade-off surface (black line in Figure 6-3). EFMs that minimize the carbon resource cost to produce cellular energy are to the left in the plot; the most efficient of these metabolic strategies completely oxidize cellulose to carbon dioxide and water at the depicted carbon (0.2 Cmole cellulose per mole cellular energy) and oxygen costs (0.2 mole oxygen per mole cellular energy). As oxygen becomes more limited, metabolic strategies that produce increasingly reduced carbon products become competitive, producing acetate, acetaldehyde, and ultimately, ethanol under oxygen limiting simulations (i.e. moving from the top left to the bottom right in Figure 6-3). Increasing ethanol and acetate yield were observed *in vitro* with decreasing headspace oxygen (Figure 6-4). The model predicted anoxic cellular energy production (1.3 Cmole cellulose per mole of cellular

energy), but not anoxic biomass production due to an oxygen requirement to produce sterols and lipids (data not shown). Cellopentaose was found to have lower resource costs than glucose for all hydrocarbons.

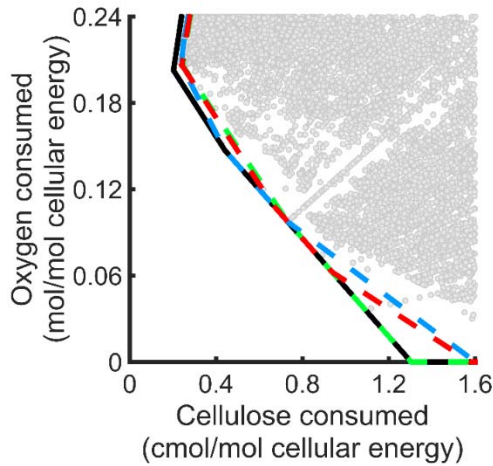


Figure 6-3. Ecologically competitive strategies for cellular energy generation. Optimal oxygen and substrate utilization strategies that produce cellular energy were identified that minimize resource costs (black boundary, all possible metabolic strategies are grey points). Strategies which produce acetate (blue dashed line), acetaldehyde (red dashed line), and ethanol (green dashed line) are highlighted to distinguish metabolic strategies. Cellopentaose was the optimal carbon source.

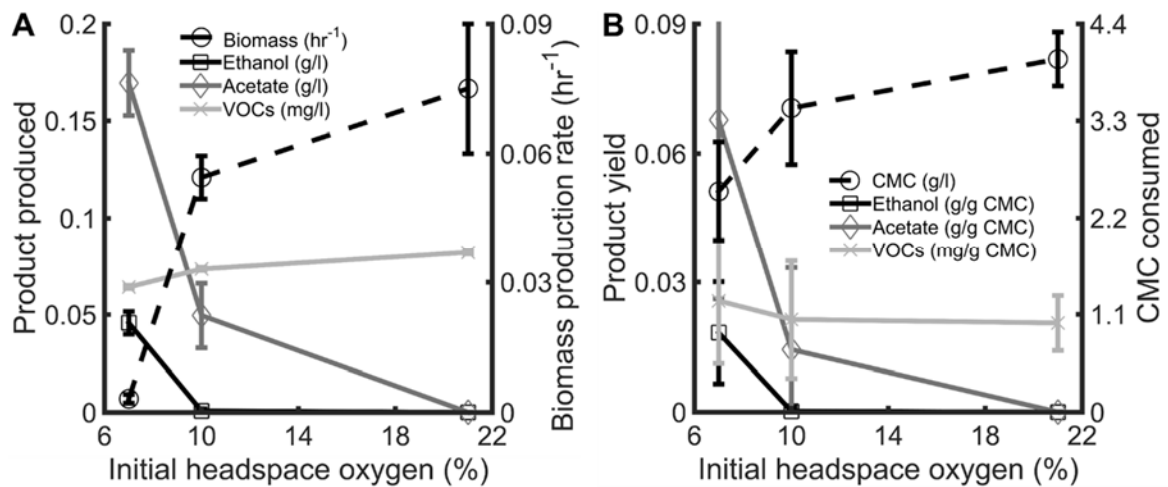


Figure 6-4. Observed product concentrations and yields from *A. sarcoides* NRRL 50072. Product and biomass concentrations (A) and yields with respect to grams of carboxymethyl cellulose (B). Error bars represent standard error of measurement for n=6, except biomass production rate with n=3.

6-7-3 Theoretical Product Yields Based on The Metabolic Potential of *A. sarcooides*

Hydrocarbon production is a balance of cellular energy production, electron trafficking, and resource utilization, which are all sensitive to culturing conditions. Gradients of carbon source and oxygen limitation were examined for competitive hydrocarbon producing phenotypes based on both carbon and oxygen resource costs. The carbon and oxygen resource costs of hydrocarbon production were defined as the Cmoles of carbon source or moles of oxygen, respectively, consumed to make a Cmole of hydrocarbons. The eight functional groups characterized were examined separately for production of C₃ to C₂₀ hydrocarbons with respect to chain length and pathway with the lowest carbon and oxygen resource costs to the left and bottom, respectively, for a given functional group (Figure 6-5). The analysis of odd- and even-chained hydrocarbons, which represent acetyl- and propanoyl-CoA as an initiator for fatty acids, was performed on the same set of EFMs but plotted separately for clarity. Analysis of general hydrocarbon production strategies revealed all hydrocarbons, except aromatic hydrocarbons, could be produced under anoxic conditions. The deamination of aromatic amino acids was modeled as requiring oxygen because phenylpyruvate decarboxylase was not annotated. The lowest predicted carbon resource cost to produce the hydrocarbons occurred under oxic conditions, where every Cmole of hydrocarbon produced consumed at least 1.16 Cmoles of carbon source (the grey lines in Figure 6-5). Strategies that minimized resource costs to exclusively produce isolated hydrocarbons of a given chain length are depicted by colored lines. For example, under anoxic conditions (along x-axis in the acids panels of Figure 6-5) the lowest carbon resource cost to

produce acids was achieved with propanoate (red line) followed by 2-methyl-butanoate, progressively longer odd chain lengths of fatty acid derived hydrocarbons, then progressively shorter even chain lengths. This general trend of Ehrlich product followed by fatty acids was observed for all primary functional groups, whereas Ehrlich pathways did not result in secondary functional groups.

Ehrlich Pathway Products: Cellopentaose-derived Ehrlich products were predicted to have the lowest carbon resource cost under oxic conditions due to their degree of reduction. For example, the lowest carbon cost under anoxic conditions was achieved by propanoate production because all electrons can be conserved on the carbon available. Larger Ehrlich products, such as indole-3-acetate and 4-hydroxyphenylacetate, require a much higher oxygen resource cost (green lines in the even chain length plot of Figure 6-5), and retain a large fraction of the carbon from the carbon source (i.e. carbon costs approximately 1.5), but have lower degrees of reduction clustering around 4.5 (Figure 6-6). Alternatively, the fatty acid derived hydrocarbons all lose at least one carbon dioxide per monomeric acetyl-CoA, but retain most if not all the electrons from the carbon source as evidenced by the degrees of reduction approaching 6 (Figure 6-6). Exceptions to this trend amongst Ehrlich products include isobutyl and isoamyl derivatives, which have degrees of reduction from 6 to nearly 8, but have increased carbon resource costs. The primary fermentative metabolism products formate, acetate, acetaldehyde, and ethanol were excluded from the analysis as they were coproduced by most longer carbon chained hydrocarbons under anoxic conditions, but range from 2 to 6 in degree of reduction.

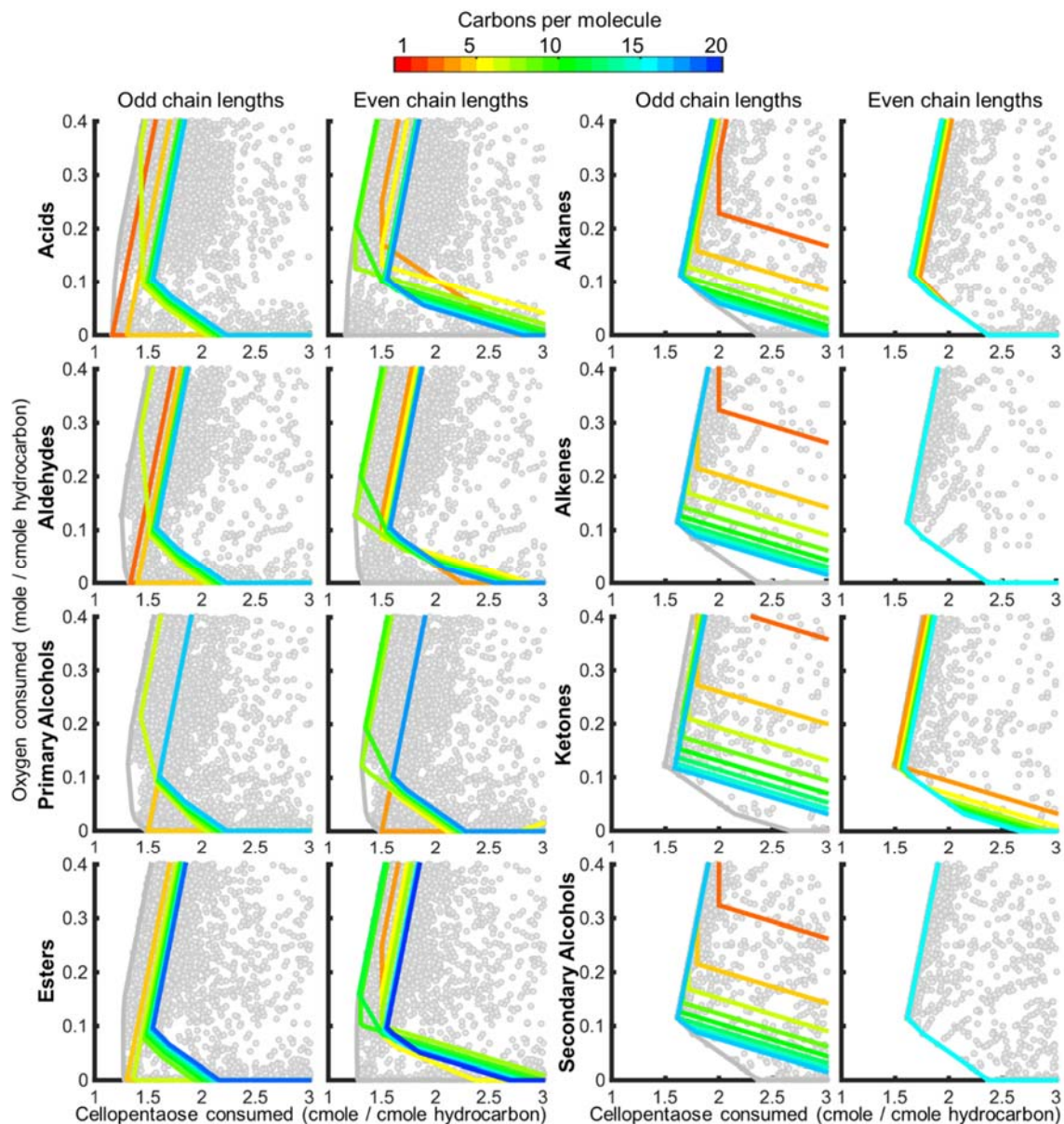


Figure 6-5. Hydrocarbon production strategies that minimize carbon and oxygen consumption. Optimal oxygen and carbon source utilization strategies that produce hydrocarbons were identified that minimize resource costs (grey boundary, all possible metabolic strategies are grey points). Metabolic strategies that produce hydrocarbons with a given number of carbons per molecule are denoted by the colored lines. Formate, acetate, acetaldehyde, and ethanol were only counted as hydrocarbons for their respective functional groups. Odd and even chained hydrocarbons were initiated by acetyl-CoA and propanoyl-CoA as appropriate. Resource profiles for even chained alkanes, even chained secondary alcohols appear as lines due to overlap.

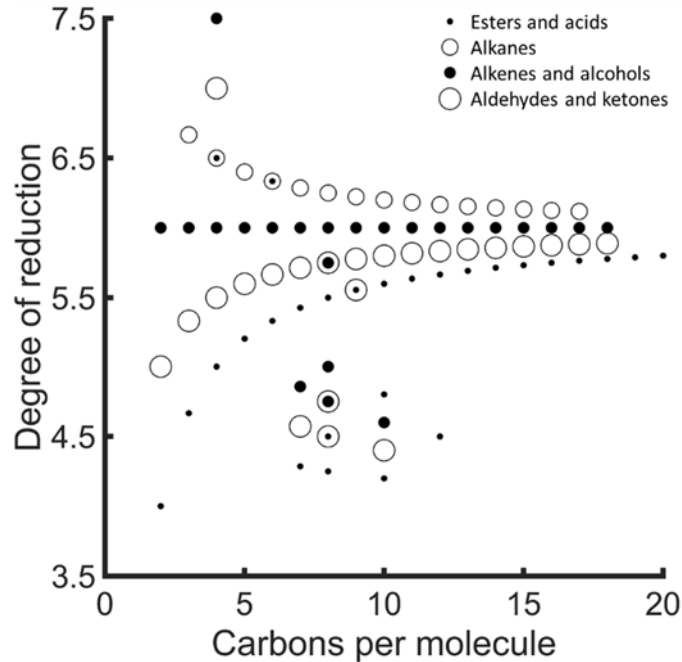


Figure 6-6. Degree of reduction of hydrocarbons as a function of carbons per molecule. Degree of reduction for each functional group was plotted considering esters and acids (•), alkanes (○), alkenes and alcohols (●), and aldehydes and ketones (◐). Outliers represent Ehrlich and linoleic pathway products.

Acetyl-CoA-derived Products: The carbon resource cost of fatty acid chain lengths is affected by the energy production requirements for a given fatty acid. Longer even chain fatty acid derivatives with primary functional groups reduce the carbon resource cost associated with formate. Formate is used to expel extra electrons since the initiating molecule does not go through reduction cycles associated with monomer additions. Odd chain length fatty acid derivatives with primary functional groups do not have extra electrons as they use propanoyl-CoA for an initiator. Propanoate has a degree of reduction of 4 with 3 carbons, thereby matching the electron input of saccharide with output of acid. In contrast, fatty acid derivatives with secondary functional groups reverse these trends as odd and even chain length fatty acids are initiated by acetyl- and

propanoyl-CoA, respectively. A similar trend was observed for oxic production of fatty acid derivatives and across functional groups. Cellopentaose consistently had lower or equal costs for all hydrocarbons (Figure 6-5 and 6-S1).

6-7-4 *In silico* Pathway Modification Strategies

In silico models provide a framework for predicting network modifications that permit enhanced catalytic function. Two modifications were tested using the presented *A. sarcooides* model for their capacity to decrease transport costs: 1) the incorporation of a cellobiose phosphorylase (CP) and 2) a cytoplasmic pyruvate dehydrogenase (PDH_{cyt}).

Cellobiose Phosphorylase: CP decreased transport costs by conserving the energy in a $\beta(1,4)$ glucose bond through phosphorylation with inorganic phosphate. This additional energy conservation alleviates the cost of activating n-1 glucose molecules using ATP, where n is the number of monomers in the oligosaccharide being transported.

Incorporation of CP activity decreased the predicted carbon and oxygen resource costs of all products with the exception of 2-methyl-propanoate and 2-methyl-butanoate, which did not require a net input of energy to produce (Figure 6-7). This effect was exemplified under anoxic conditions where energy is more resource intensive (Table 6-2). The carbon resource cost to produce hydrocarbons decreased by as much as 17 % when cellopentaose was the carbon source (Figure 6-S2). No change was predicted for cultures grown with glucose as the carbon source.

Table 6-2. Theoretical optimal yields across functional groups.

Functional group	Wildtype (WT)		WT + CP		WT + PDH _{cyt}		WT + CP + PDH _{cyt}	
	Carbon cost	Oxygen cost	Carbon cost	Oxygen cost	Carbon cost	Oxygen cost	Carbon cost	Oxygen cost
Acids	1.17 (1.2)	0.00	1.15 (1.2)	0.03	1.16 (1.2)	0.003	1.15 (1.2)	0.03
Aldehydes	1.25 (1.3)	0.13	1.23 (1.3)	0.08	1.25 (1.3)	0.13	1.23 (1.3)	0.08
Esters	1.29	0	1.29	0	1.29	0	1.29	0
Primary alcohols	1.31 (1.5)	0.12	1.28 (1.5)	0.09	1.31 (1.5)	0.12	1.28 (1.5)	0.09
Ketones	1.5 (2.7)	0.12	1.5 (2.2)	0.08	1.5 (2.0)	0.08	1.5 (1.7)	0.08
Alkenes	1.62 (2.4)	0.12	1.57 (2.0)	0.07	1.52 (1.7)	0.02	1.5	0
Secondary alcohols	1.62 (2.4)	0.12	1.57 (2.0)	0.07	1.52 (1.7)	0.02	1.5	0
Alkanes	1.64 (2.4)	0.11	1.60 (2.0)	0.07	1.54 (1.6)	0.01	1.53	0

Carbon and oxygen cost are the cmoles of cellopentaose and moles of oxygen, respectively, consumed per cmole of hydrocarbons produced under carbon- (oxygen-) limited conditions.

Each functional group was simulated in isolation.

Cytosolic Pyruvate Dehydrogenase: The transport cost of acetyl-CoA from the mitochondria to the cytosol is a limiting factor in cytosolic production of fatty acids. All acetyl-CoA transport routes described above require the consumption of energy to transfer mitochondria-derived acetyl-CoA to the cytosol where it can be incorporated into fatty acids. A PDH_{cyt} activity was hypothesized to uncouple cytosolic acetyl-CoA production from the energetic requirements of transport and provide cytosolic reducing equivalents, leading to an increase in yield of fatty acid-derived products (dashed green line in Figure 6-1). This effect was exemplified under anoxic simulations where the

carbon resource costs decreased by as much as 30 % and for products predominately derived from acetyl-CoA, such as longer chained fatty acids. The modeled PDH_{cyt} was less significant for oxic conditions (up to 6 % decrease in substrate cost, Table 6-2 and 6-S3) due the higher cellular energy yields per sugar under aerobic conditions, or shorter fatty acids that were initiated by something besides acetyl-CoA (Figure 6-7). PDH_{cyt} activity did not impact Ehrlich pathway products with the exception of 3-methylbutanoate; Ehrlich pathway products were primarily derived from central carbon metabolism intermediates and therefore insensitive to the effects of PDH_{cyt} activity. Additionally, some amino acids have been shown to originate in the mitochondria, such as leucine, thereby further removing the impact of PDH_{cyt} activity on Ehrlich pathway products. A similar effect predicted for utilization of glucose as a carbon source (Figure 6-S4).

Simultaneous Modification: The combined effect of both modifications decreased the carbon resource cost to produce hydrocarbons by up to 7 and 38 % under oxic and anoxic conditions, respectively (Table 6-2 and Figure 6-S5). The largest reductions were observed for fatty acid derived hydrocarbons as they are the most energy intensive and dependent on cytosolic acetyl-CoA. Additionally, fatty acid derived hydrocarbons were primarily limited to a minimum theoretical carbon resource cost of 1.5 due to the loss of a carbon dioxide during the formation of acetyl-CoA. Exceptions to this are fatty acids derived from alternative initiating molecules, such as propanyl-CoA, that start with a lower carbon resource cost and thereby increase in carbon cost with chain length.

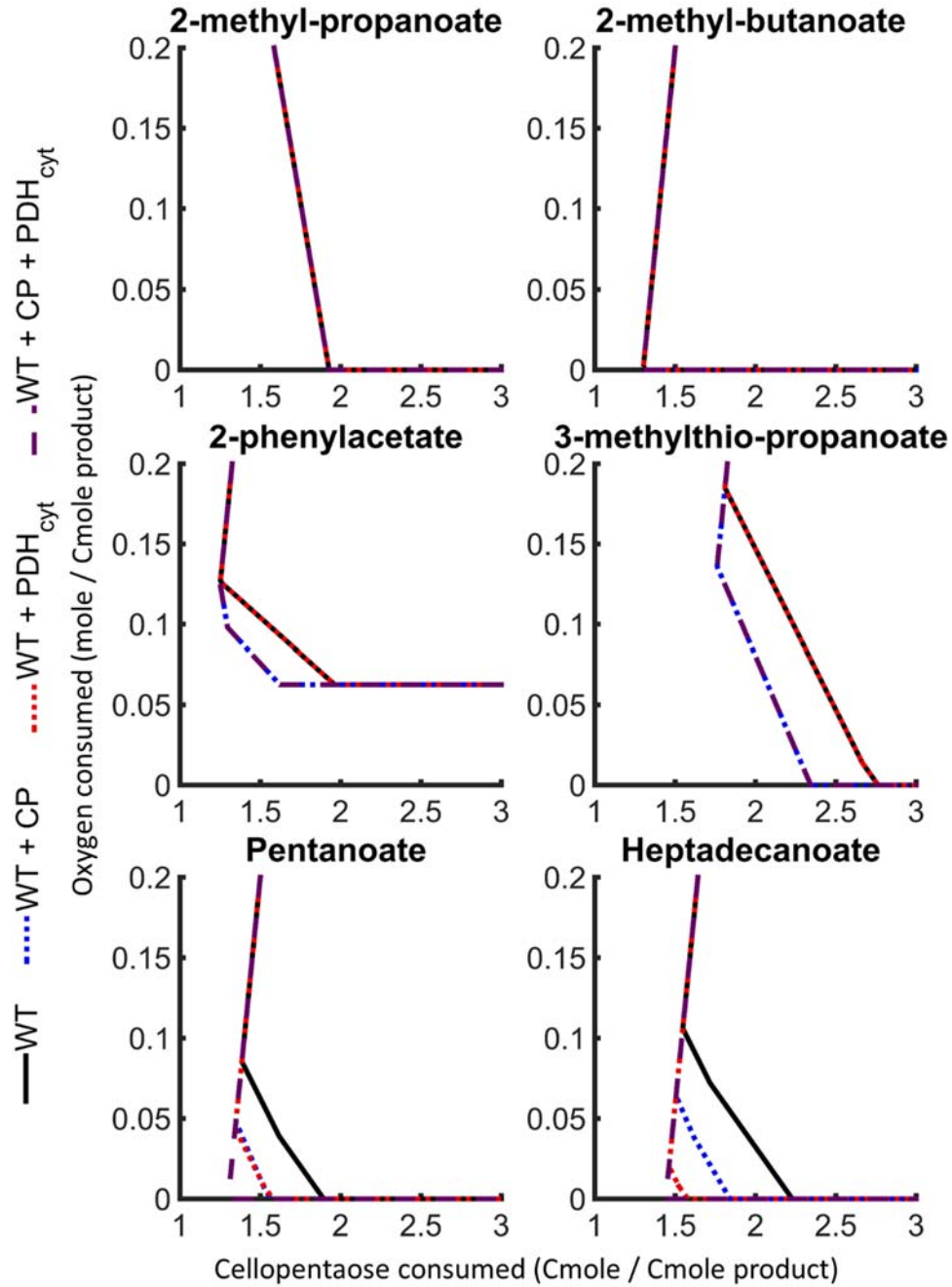


Figure 6-7. The effect of cellobiose phosphorylase (CP) and cytosolic pyruvate dehydrogenase (PDH_{cyt}) activity on excreted products from *A. sarcooides* NRRL 50072. Products that have a net yield or are independent of cellular energy were unaffected by the incorporation of these activities. Products independent of the localization of acetyl-CoA increased solely with the incorporation of CP activity, such as most aromatic Ehrlich products. Finally, products derived from acetyl-CoA were subject to PDH_{cyt} activity with change in yield proportional to the percent derived from acetyl-CoA. For example, the short fatty acid, pentanoate, had a comparable effect from both CP and PDH_{cyt} while PDH_{cyt} was the dominate activity for production of heptadecanoate.

6-7-5 In vitro Growth and Oxygen Utilization

The sodium salt of carboxymethyl cellulose (CMC), a soluble cellulose analog, was used to facilitate measurement of biomass production. *A. sarcooides* grew rapidly after approximately 24 hours of lag phase and reached stationary phase within 4 days (Figure 6-S6). Growth was very similar for 10 and 21% initial headspace oxygen concentrations but significantly reduced for 7%, with a final biomass concentration of less than 6% compared to the other oxygen concentrations (Figure 6-4A). Oxygen consumption based on measurement of the headspace (liquid accounted for less than 4% of the total vessel oxygen) was 15 ± 11 , 124 ± 10 , and 188 ± 19 mM O₂ (averages \pm standard deviations) for initial headspace oxygen concentrations of 7% (270 mM), 10% (380 mM), and 21% (740 mM), respectively. Final oxygen concentration of 221 ± 11 , 191 ± 10 , and 540 ± 19 mM O₂ were measured for 7, 10, and 21 %, respectively. The difference in growth and oxygen utilization between 7 and 10% indicated a minimum oxygen concentration for *A. sarcooides* growth was reached. CO₂ production mimicked oxygen consumption across all conditions (i.e. 23 ± 3 , 129 ± 8 , and 162 ± 12 mM for 7, 10, and 21%, respectively), indicating most of the oxygen consumed was respired.

6-7-6 In vitro Hydrocarbon Production

A. sarcooides produced the most acetate and ethanol at the 7% starting oxygen concentration (Figure 6-4A). Ethanol and acetate were not detected in the 21% starting oxygen concentration cultures. Acetate concentrations at 7% starting oxygen concentration were 2.0 mM or greater, and 0.7 mM at the 10% starting oxygen concentration, despite much higher biomass at the 10% starting oxygen concentration.

There was no glucose detected in any sample, which indicated immediate utilization of any released glucose from CMC or depolymerization of CMC as short oligomers. Increased production of reduced carbon byproduct indicated a metabolic shift to fermentation under reduced oxygen conditions that corresponded to 200 mM O₂ or 7% starting oxygen concentration. The greatest amount of fermentation byproducts, ethanol and acetate, were seen for the 7% followed by the 10%, and 21% starting oxygen concentration did not produce detectable amounts (Figure 6-4A). A final O₂ concentration of ~540 mM indicates that the 21% initial oxygen concentration condition did not become oxygen limited, which prevented a switch from complete oxidation to fermentation.

The greatest diversity of hydrocarbon species produced occurred below an initial headspace oxygen concentration of 21% (Figure 6-4). The number and total amount of hydrocarbons produced at 7% was similar to 10%, despite the observed growth impairment at 7%. The compounds identified included alcohols, ketones, aromatics, alkanes, an alkene, an ether, acids and their esters (Table 6-3) and ranged from C₂ to C₁₅. Esters represented the greatest number of identified compounds for all conditions. The 7% oxygen condition had the most identified compounds, while 21% returned the lowest numbers of compounds. Quantification of individual compound titers were limited by the HS-SPME technique due to the varied fiber affinity of the compounds in the headspace mixture (151, 231) and therefore not reported (Table 6-3). However, an integration of the peak areas of the entire mass spectra may approximate the total amount of material adsorbed by the fiber without speciation. The amount of VOC material desorbed from the

SPME fiber ranged from 123 to 163% higher than the controls, which correlated to 64 to 82 µg per ml hydrocarbon based on internal controls. The highest titers were at the lowest oxygen condition of 7%.

Table 6-3. Observed products under varying oxygen limitations.

Products detected	Initial headspace oxygen		
	7%	10%	21%
1,3-octadiene; (Z)-3-octen-1-yl-acetate			
4-hydroxy-butanoate; (Z)-4-Hexen-1-yl-acetate; (E)-3-decen-2-ol; Furfurylmethylamphetamine; 2-(1-Cyclopent-1-enyl-1-methylethyl) cyclopentanone			
2-heptanyl-acetate; 3-methyl-butanyl-acetate* ; Dimethyl ether; C ₂ , C ₅ , and C ₁₀ alkyl acetate esters*			
2-methyl-propanol* ; 1,1-dimethyl-cyclohexane; 2-ethyl-hexanol; 2-phenylethanol* ; 3-methyl-butyl-2-methyl-butanoate; C ₈ alkyl acetate esters*			
Acetate* ; 3-methyl-butanol* ; Propyl-cyclopropane* ; Octane* ; 2-nonanone* ; C ₇ and C ₉ alkyl acetate esters*			
3-methyl-hexane			

*: previously reported (6–10)

6-8 Discussion

The presented analysis of products from the endophytic fungus *A. sarcooides* quantified theoretical yields of hydrocarbon production. These theoretical yields provide context for products in both ecology and biotechnology. For example, a host would need to supply at least 1.6 Cmoles of cellopentaose to receive a Cmole of C₁₈ alkane (Table 6-2).

Alternatively put, 3.1 mg per liter of cellopentaose would result in 1 mg per liter of C₁₈ alkane. However, these yields were not observed *in vitro* indicating the possibility of a signaling molecule produced by the host to control the production of these molecules *in vivo*.

In vitro changes in *A. sarcooides* growth and excreted products (Figure 6-4) suggests that the transition to anoxic metabolism occurs above 200 mM O₂. The increase in types of hydrocarbons produced below this oxygen threshold (Table 6-3) is expected to be due to an increasingly reduced internal environment or regulatory effects favoring the excretion of reduced byproducts. The cost to produce alkane was predicted to increase with oxygen limitation (2.37 Cmoles) nearly doubling substrate costs. The *in vitro* product yield, as determined from HS-SPME, was relatively independent of oxygen concentration (Figure 6-4), therefore while the internal environment may be more reducing the net excretion of hydrocarbons was generally unchanged. The shift from common signaling molecules during hyphal growth, such as the C₈ linoleic acid degradation products, to more reduced biomass derivatives, such as Ehrlich pathway products, supports this (Table 6-3).

The theoretical yield of all examined hydrocarbons decrease slightly when derived for shorter oligosaccharides in wild type *A. sarcooides*. The ATP required to transport the carbon source is inversely proportional to the number of monomers in the oligosaccharide (i.e. n^{-1}). This may provide a feedback mechanism for endophytes that provide host defense. For example, a pathogen that results in increased oligomer concentrations and free amino acids, as expected during cellulose degradation and cell lysis, would increase the production of antibiotics by the endophyte. The effect of metabolizing free fatty acids was not analyzed here, but warrants further study.

Fatty acid production can account for most commonly reported hydrocarbons from *A. sarcooides* (Table 6-1). The presented stoichiometric model examined the effect of

different starting material, including common central carbon metabolism (i.e. acetyl-CoA) and amino acid derived intermediates (i.e. propanoyl-CoA). These two initiating molecules impacted the production of hydrocarbons through electron and energy balances. For example, propanoyl-CoA requires more cellular energy and electrons to produce than acetyl-CoA resulting in increased initiation costs for each resulting hydrocarbon. Ehrlich pathway products are also feasible initiating molecules as seen in alicyclic fatty acid production (213); hydrocarbons derived from such initiators would follow the trends observed for propanoyl-CoA because all Ehrlich products require more carbon, cellular energy, and/or electrons than acetyl-CoA.

Introduction of additional metabolic activity (i.e. CP and PDH_{cyt}) reduced the theoretical cost of hydrocarbon production to the theoretical limit. Stoichiometric analysis predicts addition of CP and PDH_{cyt} activities to hydrocarbon producing cells should result in up to a 38% increase in hydrocarbon yields, Table 6-2. CP activity reduced the activation cost of the carbon source and PDH_{cyt} activity reduced the energetic requirement for cytosolic acetyl-CoA, increasing the yield of hydrocarbons and reducing cellular control over their production. The combined impact of these modifications reduced the carbon resource cost to produce fatty acid derived hydrocarbons to 1.5 Cmoles carbon source per Cmole of hydrocarbon. These modifications and others, such as the reversal of some of these acetyl-CoA partitioning control mechanisms, have increased hydrocarbon production yields and rates in other genetically modified organisms (18, 202).

6-9 Conclusion

The study presented here represents the most extensive stoichiometric analysis of hydrocarbons produced by either an endophytic or filamentous fungus. Comparing across functional groups provides context for biological production of bioactive specialty chemicals, as well as limits of production with which to gauge metabolic engineering efforts. The examination of enzymatic localization provides insight into possible targets to improve yields and their effect on the resulting suite of excreted products.

6-10 Acknowledgements

The authors would like to thank the National Science Foundation, Emerging Frontiers in Research and Innovation [Grant Number 093761] for funding this research. Our appreciation goes to Dr. Albert Parker for help with statistical analyses.

CHAPTER 7

CONCLUSIONS

7-1 Stoichiometric Analysis

Stoichiometric analysis has been used for decades to facilitate metabolic engineering and understand ecological processes, but computational restrictions have limited its application to simplified systems. Different approaches to stoichiometric analysis, such as flux balance analysis (FBA), alleviate this restriction by only finding a subset of the routes through the network. Trouble shooting a model with FBA is difficult because some routes through the network may never satisfy a common objective function and therefore never be observed. EFMA has expanded by orders of magnitude in the last decade. One of the early software packages, MetaTool, could analyze networks containing up to a few million EFMs, though it became unstable with larger networks. EfmTool increased this threshold with the nullspace algorithm and bit-pattern trees to track solutions using binary representation, analyzing networks containing hundreds of millions of EFMs. Combining EfmTool with the subdivision approach presented in chapter 2, a network containing billions of EFMs was analyzed on multiple machines in a month. These advancements and others have since been incorporated into flux mode calculator (FMC), which reproduced the billions of EFMs on a single machine in approximately 24 hours.

Two recent approaches to EFMA, RegEfmTool and tEfmTool, remove EFMs during the analysis process that are not expected to be biologically or thermodynamically relevant, respectively. These approaches reduce the total computational requirements of a

given network, increasing data incorporation in the form of gene regulation and metabolite concentrations, while simplifying downstream processing. Genetic regulatory relationships can be identified by combining stoichiometric analyses that use preprocessing restrictions, such as RegEfmTool and tEfmTool, with unrestricted analyses. In chapter 4, the use of gene regulatory rules was examined for a heterotrophic metabolic network using RegEfmTool and EfmTool. The gene regulatory rules applied made reactions at branch points between degradation and biosynthesis mutually exclusive (i.e. specific metabolites in the network could be degraded, synthesized, or neither, but not both). Indeed, some metabolic pathways excluded by the rules resulted in lower resource costs to produce biomass and cellular energy and would therefore be expected to operate, such as valine and arginine (Figure 7-1). However, these pathways have been shown to differentially regulate under the two scenarios. Additional analyses using tEfmTool are expected to provide similar insight, and as networks become more complex, computationally advantageous approaches, such as FBA, may become increasingly important to examine these assumptions.

Stoichiometric analysis should be a routine modeling tool for under-characterized systems. The increasingly available wealth of sequencing data and annotation tools provide a convenient first pass to understanding the functionality of an organism. The organization of data, such as gene annotations and metabolites, required to construct a functional model provides a basis for the storage and passing of information within and across research projects. Creating a unified framework for data organization should provide a basis for examining data structure, improving integration with other techniques

and data sets while allowing for comparison across phylogeny and geography. For example, large scale projects working to characterize the pathways and interactions of a community and its inhabitants, can work on their individual models testing hypotheses and validating pathways, providing a convenient form to share their individual components and create a community level analysis. These community analyses could also be examined for governing principles of community organization.

7-2 Applicability of Heterotrophy Analysis

Assimilatory and dissimilatory utilization of biomass-derived dissolved organic carbon (DOC) are prevalent metabolic aspects of all biological systems. Just as biological information is perpetuated and stored, so too must antiquated or corrupted information be removed and its energy repurposed. In chapters 4 and 5, comparison of serial and parallel utilization of 29 different forms of DOC highlight the effect of nutrient transformation in growth and cellular energy production. The relative resource costs of amino acid utilization by an archaeon from Yellowstone National Park under oxygen limitation are directly proportional to volumetric consumption rates in a eutrophic estuary in North Carolina. Indeed, oxygen limitation is common in eutrophic environments, such as the dead zone in the Gulf of Mexico and Lake Erie. This relationship suggests a constant volumetric growth rate amongst populations in a community competing for the limiting resource, highlighting a stoichiometric governing principles in community organization and supporting a convenient assumption when incorporating relative abundance data into modeling.

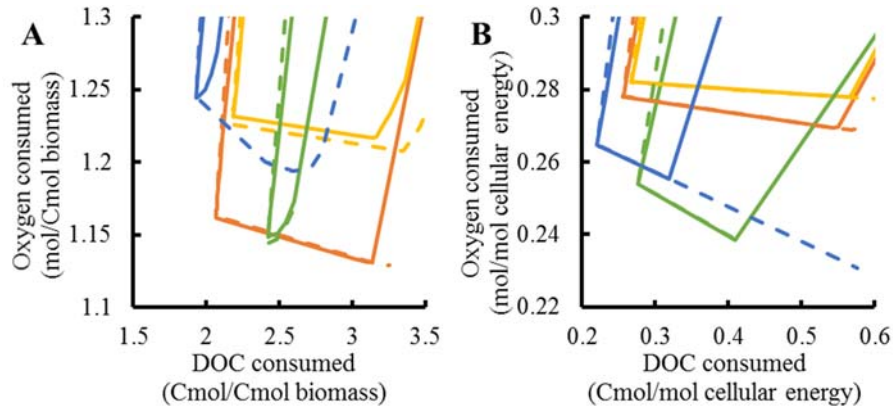


Figure 7-1 Potential regulatory impacts on resource costs. DOC and oxygen resource costs were quantified with (solid lines) and without (dashed lines) gene regulatory rules excluding degradation and biosynthesis for arginine (green), valine (blue), protein (yellow), and aggregate biomass (orange). The oxygen resource cost to produce biomass from valine and arginine is lower without reaction rules indicating alternative optimal EFMs (A), however the reactions involved in these EFMs have been shown to be differently regulated and therefore not active under the same conditions. The difference in costs translated up through the macromolecular forms to lesser degrees due to the other monomers which comprise the macromolecules. This effect is similar when producing cellular energy (B). Figure modified from (96).

7-3 Community Analysis

Continued extraction of meaningful and quantitative relationships between community populations requires advancement in algorithm performance and innovative approaches. The EFM-FBA hybrid approach described in chapter 5, used organism-level EFMA to determine specific parameters with respect to nutrient limitation, such as biomass yields under oxygen limitation, for each of the model organisms. The resulting yields were used with measurements of the natural system to determine community productivity as a function of electron acceptor, relative population abundance, and specific growth rate. Major ecological inferences about the operation of this natural community, such as the requirement of additional electron acceptor, were made by simply using a steady state mathematical treatment. These inferences provide a basis for further community

exploration, such as transcriptomic or utilization rate studies, as well as bounding parameters for finer resolution modeling approaches, which incorporate structure and transport limitations.

7-4 Hydrocarbon Production

Metabolic compartmentalization by organelles create control points for the partitioning of cellular functions, such as excretion or assimilation of energy rich cellular components. The higher resource cost to produce hydrocarbons when monomer production and assembly is partitioned, as opposed to together, suggests a large fraction of cellular energy is directed toward this control mechanism. In prokaryotes, cellular energy and fatty acid production occur in the same compartment and effectively compete for acetyl-CoA, while in eukaryotes the mitochondria are the primary source of cellular energy. This means unlike prokaryotes, eukaryotic lipid synthesis machinery has no way to assess the energy state of the cell relative to the acetyl-CoA pool. Therefore, a eukaryote with a cytosolic pyruvate dehydrogenase could over produce lipid draining the cell of its cellular energy. Fatty acid derived products are a simple example of this since they specifically require as much or more energy to produce than present in the original carbon substrate. However, this principle can be extrapolated to other cellular products such as proteins or other small molecules.

Biological hydrocarbon production is subject to the same limitations of mass and energy conservation as all other industrial processes. Products that are more reduced than the substrates used to produce them, such as alkanes from glucose, result in waste carbon

since the carbon source is also the only electron source in traditional fermentation. As evident by the theoretical limits for hydrocarbon production described in detail in chapter 6, if electrons are the sole limiting reagent (i.e. the product does not require the concentration of energy as well as electrons) a simple mass balance reveals that Cmoles of hydrocarbon produced per Cmole of carbon dioxide produced equals the degree of reduction (DOR) of the substrate over the difference of DOR for the hydrocarbon produced (Eq. 1).

$$\frac{DOR_{Substrate}}{DOR_{Hydrocarbon} - DOR_{Substrate}} = \frac{Cmol_{Hydrocarbon}}{Cmol_{CO_2}} \quad \text{Eq. 1}$$

For example, if glucose (DOR=4) is used to produce alkanes (DOR=6) the max carbon selectivity is 2 or 67% of the carbon in glucose. However, in biological systems additional energy may be needed to drive the reaction (i.e. transport or overcoming thermodynamic limits) as examined in chapter 6 thereby reducing the yield even further. This is both an advantage and disadvantage to biological systems, which require energy from the process to propagate their biological information, but can also move energy around (i.e. bifurcating enzymes).

7-5 Closing Statements

Nutrient and energy transfers are fundamental aspects of all life manifesting at various scales from pathogenesis to planetary biogeochemical cycling. Complex modeling and analytical methods have been developed to understand the function and structure of these populations and communities, but can be prohibitively expensive or complex for routine application. Stoichiometric modeling at the organism and community level simply

required genomic or metagenomics data, which is becoming increasingly available, to construct the relevant metabolic networks. Analysis and interpretation using simple kinetic assumptions, such as limits for growth rate, was able to illustrate and quantify principles that govern the metabolic nature of these interactions, such as intra- and inter-cellular compartmentalization and maximum biomass production as a function of relative population abundance. Developments in stoichiometric analysis and data processing will continue to expand the applicability of this approach and ease of use.

The presented community analysis has considered biomass as the primary means of nutrient and energy transfer; however, other means of exchange may be more effective when growth becomes prohibitively slow. In microbial communities where some populations inhibit others, this decrease in growth rate may result in leaking of metabolites that are more amenable to the energy transfer needs of specific community members. For example, a low energy community, such as those found in the deep subsurface or hydrogen based exchanges to produce methane, the death of populations would be very costly to the community. Therefore, in these communities it would be advantageous to exchange metabolites as opposed to biomass. These communities could be modeled by setting uptake rates and allowing for the variability of specific growth rate to quantify the impact of exchanging different metabolites between live populations as opposed to the consumption of one populations biomass to sustain another. Quantification of these governing principles of nutrient and energy transfer through readily available datasets, such as metagenomes, could set the stage for understanding and prediction of more complex systems, such as multicellular organisms.

REFERENCES CITED

REFERENCES CITED

1. Agrawal AA. Phenotypic plasticity in the interactions and evolution of species. *Science* 294: 321–326, 2001.
2. Ahamed A, Ahring BK. Production of hydrocarbon compounds by endophytic fungi *Gliocladium* species grown on cellulose. *Bioresour Technol* 102: 9718–22, 2011.
3. Amend JP, Shock EL. Energetics of overall metabolic reactions of thermophilic and hyperthermophilic Archaea and Bacteria. *FEMS Microbiol. Rev.* 25: 175–243, 2001.
4. Arndt S, Jørgensen BB, LaRowe DE, Middelburg JJ, Pancost RD, Regnier P. Quantifying the degradation of organic matter in marine sediments: A review and synthesis. *Earth-Science Rev* 123: 53–86, 2013.
5. Asif M, Muneer T. Energy supply, its demand and security issues for developed and emerging economies. *Renew Sustain Energy Rev* 11: 1388–1413, 2007.
6. Bader J, Mast-Gerlach E, Popovic MK, Bajpai R, Stahl U. Relevance of microbial coculture fermentations in biotechnology. *J Appl Microbiol* 109: 371–387, 2010.
7. Barrett CL, Herrgård MJ, Palsson BØ. Decomposing complex reaction networks using random sampling, principal component analysis and basis rotation. *BMC Syst Biol* 3: 30, 2009.
8. Battin TJ, Sloan WT, Kjelleberg S, Daims H, Head IM, Curtis TP, Eberl L. Microbial landscapes: new paths to biofilm research. *Nat Rev Microbiol* 5: 76–81, 2007.
9. Beam JP, Bernstein HC, Jay ZJ, Kozubal MA, Jennings R deM., Tringe SG, Inskeep WP. Assembly and Succession of Iron Oxide Microbial Mat Communities in Acidic Geothermal Springs. *Front Microbiol* 7, 2016.
10. Beam JP, Jay ZJ, Kozubal MA, Inskeep WP. Niche specialization of novel Thaumarchaeota to oxic and hypoxic acidic geothermal springs of Yellowstone National Park. *ISME J* 8: 938–951, 2014.
11. Beck AE, Hunt KA, Bernstein HC, Carlson RP. Interpreting and designing microbial communities for bioprocess applications, from components to interactions to emergent properties. In: *Biotechnology for Biofuel Production and Optimization*, edited by Eckert C, Trinh CT. Elsevier, p. 407–432.
12. Beopoulos A, Cescut J, Haddouche R, Uribe Larrea J-L, Molina-Jouve C, Nicaud J-M. *Yarrowia lipolytica* as a model for bio-oil production. *Prog Lipid Res* 48: 375–

87, 2009.

13. Berlin H, Westerberg J. Über die Unterscheidung von Muskel- und Hefeadenylsäure. *Hoppe-Seyler's Zeitschrift für Physiol Chemie* 281: 98–101, 1944.
14. Bernstein HC, Beam JP, Kozubal MA, Carlson RP, Inskeep WP. *In situ* analysis of oxygen consumption and diffusive transport in high-temperature acidic iron-oxide microbial mats. *Environ Microbiol* 15: 2360–70, 2013.
15. Bernstein HC, Carlson RP. Microbial consortia engineering for cellular factories: *in vitro* to *in silico* systems. *Comput Struct Biotechnol J* 3: e201210017, 2012.
16. Bernstein HC, Paulson SD, Carlson RP. Synthetic *Escherichia coli* consortia engineered for syntrophy demonstrate enhanced biomass productivity. *J Biotechnol* 157: 159–66, 2012.
17. Bizukojc M, Dietz D, Sun J, Zeng A-P. Metabolic modelling of syntrophic-like growth of a 1,3-propanediol producer, *Clostridium butyricum*, and a methanogenic archeon, *Methanosarcina mazei*, under anaerobic conditions. *Bioprocess Biosyst Eng* 33: 507–523, 2010.
18. Blazeck J, Hill A, Liu L, Knight R, Miller J, Pan A, Otoupal P, Alper HS. Harnessing *Yarrowia lipolytica* lipogenesis to create a platform for lipid and biofuel production. *Nat Commun* 5: 3131, 2014.
19. Bond T, Templeton MR. History and future of domestic biogas plants in the developing world. *Energy Sustain Dev* 15: 347–354, 2011.
20. Boruff CS. Recovery of fermentation residues as feeds. *Ind Eng Chem* 39: 602–607, 1947.
21. Boyd ES, Leavitt WD, Geesey GG. CO₂ uptake and fixation by a thermoacidophilic microbial community attached to precipitated sulfur in a geothermal spring. *Appl Environ Microbiol* 75: 4289–96, 2009.
22. Brauer VS, Stomp M, Huisman J. The nutrient-load hypothesis: patterns of resource limitation and community structure driven by competition for nutrients and light. *Am Nat* 179: 721–740, 2012.
23. Breitbart M, Wegley L, Leeds S, Schoenfeld T, Rohwer F. Phage community dynamics in hot springs. *Appl Environ Microbiol* 70: 1633–40, 2004.
24. Brenner K, Arnold FH. Self-organization, layered structure, and aggregation enhance persistence of a synthetic biofilm consortium. *PLoS One* 6: e16791, 2011.

25. Brenner K, You L, Arnold FH. Engineering microbial consortia: a new frontier in synthetic biology. *Trends Biotechnol* 26: 483–489, 2008.
26. Brune KD, Bayer TS. Engineering microbial consortia to enhance biomining and bioremediation. *Front Microbiol* 3: 203, 2012.
27. Bryant DA, Liu ZF, Li T, Zhao FQ, Costas AMG, Klatt CG, Ward DM, Frigaard NU, Overmann J. Comparative and functional genomics of anoxygenic green bacteria from the taxa Chlorobi, Chloroflexi, and Acidobacteria. *Funct Genomics Evol Photosynth Syst* 33: 47–102, 2012.
28. Bull AT. The renaissance of continuous culture in the post-genomics age. *J Ind Microbiol Biotechnol* 37: 993–1021, 2010.
29. Cai TT, Montague CL, Davis JS. The maximum power principle: An empirical investigation. *Ecol Modell* 190: 317–335, 2006.
30. Carlson R, Fell D, Sreenc F. Metabolic pathway analysis of a recombinant yeast for rational strain development. *Biotechnol Bioeng* 79: 121–34, 2002.
31. Carlson R, Sreenc F. Fundamental Escherichia coli biochemical pathways for biomass and energy production: creation of overall flux states. *Biotechnol Bioeng* 86: 149–62, 2004.
32. Carlson RP. Metabolic systems cost-benefit analysis for interpreting network structure and regulation. *Bioinformatics* 23: 1258–1264, 2007.
33. Carlson RP. Decomposition of complex microbial behaviors into resource-based stress responses. *Bioinformatics* 25: 90–7, 2009.
34. Carlson RP, Oshota OJ, Taffs RL. Systems analysis of microbial adaptations to simultaneous stresses. *Subcell Biochem* 64: 139–157, 2012.
35. Carlson RP, Taffs RL. Molecular-level tradeoffs and metabolic adaptation to simultaneous stressors. *Curr Opin Biotechnol* 21: 670–676, 2010.
36. Carlton RG, Richardson LL. Oxygen and sulfide dynamics in a horizontally migrating cyanobacterial mat: Black band disease of corals. *FEMS Microbiol Ecol* 18: 155–162, 1995.
37. Caspi R, Altman T, Dale JM, Dreher K, Fulcher CA, Gilham F, Kaipa P, Karthikeyan AS, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Paley S, Popescu L, Pujar A, Shearer AG, Zhang P, Karp PD. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 38: D473-9, 2010.

38. Caspi R, Altman T, Dreher K, Fulcher CA, Subhraveti P, Keseler IM, Kothari A, Krummenacker M, Latendresse M, Mueller LA, Ong Q, Paley S, Pujar A, Shearer AG, Travers M, Weerasinghe D, Zhang P, Karp PD. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res* 40: D742-53, 2012.
39. Cerqueda-García D, Martínez-Castilla LP, Falcón LI, Delaye L. Metabolic analysis of *Chlorobium chlorochromatii* CaD3 reveals clues of the symbiosis in “Chlorochromatium aggregatum”. *ISME J* 8: 991–8, 2014.
40. Cheng YF, Edwards JE, Allison GG, Zhu WY, Theodorou MK. Diversity and activity of enriched ruminal cultures of anaerobic fungi and methanogens grown together on lignocellulose in consecutive batch culture. *Bioresour Technol* 100: 4821–4828, 2009.
41. Childers DL, Corman J, Edwards M, Elser JJ. Sustainability challenges of phosphorus and food: Solutions from closing the human phosphorus cycle. *Bioscience* 61: 117–124, 2011.
42. Crawford CC, Hobbie JE, Webb KL. The utilization of dissolved free amino acids by estuarine microorganisms. *Ecology* 55: 551–563, 1974.
43. D’Imperio S, Lehr CR, Oduro H, Druschel G, Kuhl M, McDermott TR. Relative Importance of H₂ and H₂S as Energy Sources for Primary Production in Geothermal Springs. *Appl Environ Microbiol* 74: 5802–5808, 2008.
44. D’Onofrio A, Crawford JM, Stewart EJ, Witt K, Gavrish E, Epstein S, Clardy J, Lewis K. Siderophores from neighboring organisms promote the growth of uncultured bacteria. *Chem Biol* 17: 254–264, 2010.
45. Dance A. Soil ecology: What lies beneath. *Nature* 455: 724–725, 2008.
46. Decho AW, Norman RS, Visscher PT. Quorum sensing in natural environments: emerging views from microbial mats. *Trends Microbiol* 18: 73–80, 2010.
47. Dekel E, Alon U. Optimality and evolutionary tuning of the expression level of a protein. *Nature* 436: 588–592, 2005.
48. DeLong JP. The maximum power principle predicts the outcomes of two-species competition experiments. *Oikos* 117: 1329–1336, 2008.
49. Deutscher J. The mechanisms of carbon catabolite repression in bacteria. *Curr Opin Microbiol* 11: 87–93, 2008.
50. Dobell C, Leeuwenhoek A van. Antony van Leeuwenhoek and his “Little animals”; being some account of the father of protozoology and bacteriology and

his multifarious discoveries in these disciplines; Harcourt, Brace and company,.

51. Donner SD, Kucharik CJ. Corn-based ethanol production compromises goal of reducing nitrogen export by the Mississippi River. *Proc Natl Acad Sci U S A* 105: 4513–8, 2008.
52. Dykhuizen D, Hartl D. Selective neutrality of 6PGD allozymes in *E. coli* and the effects of genetic background [Online]. *Genetics* 96: 801–17, 1980. <http://www.genetics.org/content/96/4/801.short> [28 Mar. 2013].
53. Edwards JS, Ramakrishna R, Palsson BØ. Characterizing the metabolic phenotype : A phenotype phase plane analysis. *Biotechnol Bioeng* 77: 27–36, 2002.
54. Eiteman MA, Lee SA, Altman E. A co-fermentation strategy to consume sugar mixtures effectively. *J Biol Eng* 2: 3, 2008.
55. Elser JJ, Acquisti C, Kumar S. Stoichiogenomics: the evolutionary ecology of macromolecular elemental composition. *Trends Ecol Evol* 26: 38–44, 2011.
56. Emanuelsson O, Brunak S, von Heijne G, Nielsen H. Locating proteins in the cell using TargetP, SignalP and related tools. *Nat Protoc* 2: 953–71, 2007.
57. Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *J Mol Biol* 300: 1005–1016, 2000.
58. Enyeart PJ, Simpson ZB, Ellington AD. A microbial model of economic trading and comparative advantage. *J Theor Biol* (2014). doi: 10.1016/j.jtbi.2014.09.030.
59. Ernst D, Greer M, Akmatova R, Pischke S, Wedemeyer H, Heiken H, Tillmann HL, Schmidt RE, Stoll M. Impact of GB virus C viraemia on clinical outcome in HIV-1-infected patients: a 20-year follow-up study. *HIV Med* 15: 245–250, 2014.
60. Fabris M, Matthijs M, Rombauts S, Vyverman W, Goossens A, Baart GJE. The metabolic blueprint of *Phaeodactylum tricornutum* reveals a eukaryotic Entner-Doudoroff glycolytic pathway. *Plant J* 70: 1004–14, 2012.
61. Fajardo A, Martinez JL. Antibiotics as signals that trigger specific bacterial responses. *Curr Opin Microbiol* 11: 161–167, 2008.
62. Falkowski PG, Fenchel T, Delong EF. The microbial engines that drive Earth's biogeochemical cycles. *Science* 320: 1034–1039, 2008.
63. Falony G, De Vuyst L. Ecological interactions of bacteria in the human gut. (2009). doi: 10.1007/978-0-387-79058-9_16.

64. Farrand SG, Jones CW, Linton JD, Stephenson RJ. The effect of temperature and pH on the growth efficiency of the thermoacidophilic bacterium *Bacillus acidocaldarius* in continuous culture. *Arch Microbiol* 135: 276–283, 1983.
65. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 7: 129–43, 2009.
66. Feist AM, Scholten JCM, Palsson BØ, Brockman FJ, Ideker T. Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol Syst Biol* 2: 2006.0004, 2006.
67. Feist AM, Scholten JCM, Palsson BØ, Brockman FJ, Ideker T. Modeling methanogenesis with a genome-scale metabolic reconstruction of *Methanosarcina barkeri*. *Mol Syst Biol* 2: 2006.0004, 2006.
68. de Figueiredo LF, Podhorski A, Rubio A, Kaleta C, Beasley JE, Schuster S, Planes FJ. Computing the shortest elementary flux modes in genome-scale metabolic networks. *Bioinformatics* 25: 3158–65, 2009.
69. Filippelli GM. The global phosphorus cycle: Past, present, and future. *Elements* 4: 89–95, 2008.
70. Fischer E, Sauer U. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *J Biol Chem* 278: 46446–46451, 2003.
71. Flamholz A, Noor E, Bar-Even A, Liebermeister W, Milo R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc Natl Acad Sci U S A* 110: 10039–10044, 2013.
72. Florey HW. The use of micro-organisms for therapeutic purposes. *Yale J Biol Med* 19: 101-, 1946.
73. Folsom JP, Carlson RP. Physiological, biomass elemental composition and proteomic analyses of *Escherichia coli* ammonium-limited chemostat growth, and comparison with iron- and glucose-limited chemostat growth. *Microbiology* 161: 1659–70, 2015.
74. Frostl JM, Overmann J. Physiology and tactic response of the phototrophic consortium “*Chlorochromatium aggregatum*.” *Arch microbiol* 169: 129–135, 1998.
75. Gagneur J, Klamt S. Computation of elementary modes: a unifying framework and the new binary approach. *BMC Bioinformatics* 5: 175, 2004.
76. Gianoulis TA, Griffin MA, Spakowicz DJ, Dunican BF, Alpha CJ, Sboner A, Michael Sismour A, Kodira C, Egholm M, Church GM, Gerstein MB, Strobel SA. Genomic analysis of the hydrocarbon-producing, cellulolytic, endophytic fungus

Ascocoryne sarcoides. *PLoS Genet* 8: e1002558, 2012.

77. Goers L, Freemont P, Polizzi KM. Co-culture systems and technologies: taking synthetic biology to the next level. *J R Soc Interface* 11: 20140065–20140065, 2014.
78. Görke B, Stülke J. Carbon catabolite repression in bacteria: many ways to make the most out of nutrients. *Nat Rev Microbiol* 6: 613–24, 2008.
79. Grady CPLJ, Daigger GT, Love NG, Filipe CDM. *Biological Wastewater Treatment*. 3rd ed. Taylor & Francis, 2011.
80. Griffin MA, Spakowicz DJ, Gianoulis TA, Strobel SA. Volatile organic compound production by organisms in the genus *Ascocoryne* and a re-evaluation of myco-diesel production by NRRL 50072. *Microbiology* 156: 3814–29, 2010.
81. Grimm V, Wissel C. Babel, or the ecological stability discussions: An inventory and analysis of terminology and a guide for avoiding confusion. *Oecologia* 109: 323–334, 1997.
82. Grosskopf T, Soyer OS. Synthetic microbial communities. *Curr Opin Microbiol* 18: 72–77, 2014.
83. Grube M, Zagreba E, Gromozova E, Fomina M. Comparative investigation of the macromolecular composition of mycelia forms *Thielavia terrestris* by infrared spectroscopy. *Vib Spectrosc* 19: 301–306, 1999.
84. Hall-Stoodley L, Costerton JW, Stoodley P. Bacterial biofilms: from the natural environment to infectious diseases. *Nat Rev Microbiol* 2: 95–108, 2004.
85. Hamamura N, Macur RE, Korf S, Ackerman G, Taylor WP, Kozubal M, Reysenbach AL, Inskeep WP. Linking microbial oxidation of arsenic with detection and phylogenetic analysis of arsenite oxidase genes in diverse geothermal environments. *Environ Microbiol* 11: 421–431, 2009.
86. Hanly TJ, Henson MA. Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. *Biotechnol Bioeng* 108: 376–385, 2011.
87. Hanly TJ, Urello M, Henson M a. Dynamic flux balance modeling of *S. cerevisiae* and *E. coli* co-cultures for efficient consumption of glucose/xylose mixtures. *Appl Microbiol Biotechnol* 93: 2529–2541, 2012.
88. Harcombe WR, Riehl WJ, Dukovski I, Granger BR, Betts A, Lang AH, Bonilla G, Kar A, Leiby N, Mehta P, Marx CJ, Segre D, Segre D. Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial

- dynamics. *Cell Rep* 7: 1104–1115, 2014.
89. Hazelwood LA, Daran J-M, van Maris AJA, Pronk JT, Dickinson JR. The Ehrlich pathway for fusel alcohol production: a century of research on *Saccharomyces cerevisiae* metabolism. *Appl Environ Microbiol* 74: 2259–66, 2008.
 90. Hense BA, Kuttler C, Muller J, Rothballer M, Hartmann A, Kreft JU. Does efficiency sensing unify diffusion and quorum sensing? *Nat Rev Microbiol* 5: 230–239, 2007.
 91. Hibbing ME, Fuqua C, Parsek MR, Peterson SB. Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol* 8: 15–25, 2010.
 92. Hoehler TM, Jørgensen BB. Microbial life under extreme energy limitation. *Nat Rev Microbiol* 11: 83–94, 2013.
 93. Holland JN, DeAngelis DL. Consumer-resource theory predicts dynamic transitions between outcomes of interspecific interactions. *Ecol Lett* 12: 1357–1366, 2009.
 94. Hopmans EC, Schouten S, Pancost RD, Van Der Meer MTJ, Sinninghe Damsté JS. Analysis of intact tetraether lipids in archaeal cell material and sediments by high performance liquid chromatography/atmospheric pressure chemical ionization mass spectrometry. *Rapid Commun Mass Spectrom* 14: 585–9, 2000.
 95. Huber G, Spinnler C, Gambacorta A, Stetter KO. *Metallosphaera sedula* gen. nov. represents a new genus of aerobic, metal-mobilizing, thermoacidophilic archaeobacteria. *Syst Appl Microbiol* 12: 38–47, 1989.
 96. Hunt KA, Jennings R deM., Inskeep WP, Carlson RP. Stoichiometric modelling of assimilatory and dissimilatory biomass utilisation in a microbial community. *Environ. Microbiol.* (August 2016). doi: 10.1111/1462-2920.13444.
 97. Huson DH, Auch AF, Qi J, Schuster SC. MEGAN analysis of metagenomic data. *Genome Res* 17: 377–86, 2007.
 98. Hussa E a, Goodrich-Blair H. It takes a village: ecological and fitness impacts of multipartite mutualism. *Annu Rev Microbiol* 67: 161–178, 2013.
 99. Hynes MJ, Murray SL. ATP-citrate lyase is required for production of cytosolic acetyl coenzyme A and development in *Aspergillus nidulans*. *Eukaryot Cell* 9: 1039–48, 2010.
 100. INSKEEP WP, ACKERMAN GG, TAYLOR WP, KOZUBAL M, KORF S, MACUR RE. On the energetics of chemolithotrophy in nonequilibrium systems: case studies of geothermal springs in Yellowstone National Park. *Geobiology* 3:

297–317, 2005.

101. Inskeep WP, Jay ZJ, Herrgard MJ, Kozubal MA, Rusch DB, Tringe SG, Macur RE, Jennings R de M, Boyd ES, Spear JR, Roberto FF. Phylogenetic and functional analysis of metagenome sequence from high-temperature archaeal habitats demonstrate linkages between metabolic potential and geochemistry. *Front Microbiol* 4: 95, 2013.
102. Inskeep WP, Macur RE, Harrison G, Bostick BC, Fendorf S. Biomineralization of As(V)-hydrous ferric oxyhydroxide in microbial mats of an acid-sulfate-chloride geothermal spring, Yellowstone National Park. *Geochim Cosmochim Acta* 68: 3141–3155, 2004.
103. Inskeep WP, Rusch DB, Jay ZJ, Herrgard MJ, Kozubal MA, Richardson TH, Macur RE, Hamamura N, Jennings R, Fouke BW, Reysenbach AL, Roberto F, Young M, Schwartz A, Boyd ES, Badger JH, Mathur EJ, Ortmann AC, Bateson M, Geesey G, Frazier M. Metagenomes from high-temperature chemotrophic systems reveal geochemical controls on microbial community structure and function. *PLoS One* 5: e9773, 2010.
104. Ip K, Colijn C, Lun DS. Analysis of complex metabolic behavior through pathway decomposition. *BMC Syst Biol* 5: 91, 2011.
105. Ishii S, Shimoyama T, Hotta Y, Watanabe K. Characterization of a filamentous biofilm community established in a cellulose-fed microbial fuel cell. *BMC Microbiol* 8: 6, 2008.
106. Jagmann N, Philipp B. Design of synthetic microbial communities for biotechnological production processes. *J Biotechnol* 184: 209–218, 2014.
107. Jennings RM, Whitmore LM, Moran JJ, Kreuzer HW, Inskeep WP. Carbon dioxide fixation by *Metallosphaera yellowstonensis* and acidothermophilic iron-oxidizing microbial communities from Yellowstone National Park. *Appl Environ Microbiol* 80: 2665–71, 2014.
108. Jevremović D, Boley D. Parallel computation of elementary flux modes in metabolic networks using global arrays. In: *The 6th Conference on Partitioned Global Address Space Programming Models*. Santa Barbara, CA: 2012.
109. Jevremović D, Boley D, Sosa CP. Divide-and-conquer approach to the parallel computation of elementary flux modes in metabolic networks. In: *2011 IEEE International Symposium on Parallel and Distributed Processing Workshops and Phd Forum*. IEEE, p. 502–511.
110. Jevremović D, Trinh CT, Srien F, Sosa CP, Boley D. Parallelization of nullspace algorithm for the computation of metabolic pathways. *Parallel Comput* 37: 261–

78, 2011.

111. Jungreuthmayer C, Ruckerbauer DE, Zanghellini J. Utilizing gene regulatory information to speed up the calculation of elementary flux modes [Online]. <http://arxiv.org/abs/1208.1853> [19 Jun. 2014].
112. Jungreuthmayer C, Ruckerbauer DE, Zanghellini J. regEfmttool: Speeding up elementary flux mode calculation using transcriptional regulatory rules in the form of three-state logic. *Biosystems* 113: 37–9, 2013.
113. Kaleta C, de Figueiredo LF, Schuster S, Figueiredo L de. Can the whole be less than the sum of its parts? Pathway analysis in genome-scale metabolic networks using elementary flux patterns. *Genome Res* 19: 1872–83, 2009.
114. Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. [Online]. *Nucleic Acids Res* 28: 27–30, 2000. <http://www.ncbi.nlm.nih.gov/pubmed/10592173> [1 Nov. 2012].
115. Kinnersley M, Wenger J, Kroll E, Adams J, Sherlock G, Rosenzweig F. Ex uno plures: clonal reinforcement drives evolution of a simple microbial community. *PLoS Genet* 10: e1004430, 2014.
116. Kitano H. Violations of robustness trade-offs. *Mol Syst Biol* 6: 384, 2010.
117. Klamt S, Gagneur J, von Kamp A. Algorithmic approaches for computing elementary modes in large biochemical reaction networks. *IEE Proc - Syst Biol* 152: 249–55, 2005.
118. Klamt S, von Kamp A. An application programming interface for CellNetAnalyzer. *BioSystems* 105: 162–8, 2011.
119. Klamt S, Saez-Rodriguez J, Gilles ED. Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst Biol* 129: 329–51, 2007.
120. Klamt S, Stelling J. Combinatorial complexity of pathway analysis in metabolic networks. [Online]. *Mol Biol Rep* 29: 233–236, 2002. <http://www.ncbi.nlm.nih.gov/pubmed/12241063> [28 Mar. 2013].
121. Klamt S, Stelling J. Two approaches for metabolic pathway analysis? *Trends Biotechnol* 21: 64–9, 2003.
122. Klitgord N, Segrè D, Segre D. Environments that induce synthetic microbial ecosystems. *PLoS Comput Biol* 6: e1001002, 2010.
123. Konopka A. What is microbial community ecology? *ISME J* 3: 1223–1230, 2009.

124. Korgaonkar A, Trivedi U, Rumbaugh KP, Whiteley M. Community surveillance enhances *Pseudomonas aeruginosa* virulence during polymicrobial infection. *Proc Natl Acad Sci U S A* 110: 1059–1064, 2013.
125. Kovarova-Kovar K, Egli T. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics [Online]. *Microbiol Mol Biol Rev* 62: 646–66, 1998.
<http://mibr.asm.org/cgi/content/abstract/62/3/646>
<http://mibr.asm.org/cgi/content/full/62/3/646>.
126. Kozubal MA. Geomicrobiology of iron oxyhydroxide mats in acidic geothermal springs of Yellowstone National Park, Wyoming, United States of America. .
127. Kozubal MA, Dlakic M, Macur RE, Inskeep WP. Terminal oxidase diversity and function in “*Metallosphaera yellowstonensis*”: gene expression and protein modeling suggest mechanisms of Fe(II) oxidation in the *Sulfolobales*. *Appl Environ Microbiol* 77: 1844–53, 2011.
128. Kozubal MA, Macur RE, Jay ZJ, Beam JP, Malfatti SA, Tringe SG, Kocar BD, Borch T, Inskeep WP. Microbial iron cycling in acidic geothermal springs of Yellowstone National Park: integrating molecular surveys, geochemical processes, and isolation of novel Fe-active microorganisms. *Front Microbiol* 3: 109, 2012.
129. Kozubal MA, Romine M, Jennings R deM, Jay ZJ, Tringe SG, Rusch DB, Beam JP, McCue LA, Inskeep WP. Geoarchaeota: a new candidate phylum in the Archaea from high-temperature acidic iron mats in Yellowstone National Park. *ISME J* 7: 622–34, 2013.
130. Kozubal M, Macur RE, Korf S, Taylor WP, Ackerman GG, Nagy A, Inskeep WP. Isolation and distribution of a novel iron-oxidizing crenarchaeon from acidic geothermal springs in Yellowstone National Park. *Appl Environ Microbiol* 74: 942–949, 2008.
131. Kroth PG, Chiovitti A, Gruber A, Martin-Jezequel V, Mock T, Parker MS, Stanley MS, Kaplan A, Caron L, Weber T, Maheswari U, Armbrust EV, Bowler C. A model for carbohydrate metabolism in the diatom *Phaeodactylum tricorutum* deduced from comparative whole genome analysis [Online]. *PLoS One* 3: e1426, 2008. <http://dx.plos.org/10.1371/journal.pone.0001426>.
132. Kuhl M, Jorgensen BB. Microsensor measurements of sulfate reduction and sulfide oxidation in compact microbial communities of aerobic biofilms [Online]. *Appl Env Microbiol* 58: 1164–1174, 1992.
<http://www.ncbi.nlm.nih.gov/pubmed/16348687>.
133. Kumar K, Mella-Herrera RA, Golden JW. Cyanobacterial heterocysts. *Cold Spring Harb Perspect Biol* 2: a000315, 2010.

134. Kuss SK, Best GT, Etheredge CA, Pruijssers AJ, Frierson JM, Hooper L V, Dermody TS, Pfeiffer JK. Intestinal microbiota promote enteric virus replication and systemic pathogenesis. *Science* 334: 249–252, 2011.
135. Ladygina N, Dedyukhina EG, Vainshtein MB. A review on microbial synthesis of hydrocarbons. *Process Biochem* 41: 1001–1014, 2006.
136. Langner HW, Jackson CR, McDermott TR, Inskeep WP. Rapid oxidation of arsenite in a hot spring ecosystem, Yellowstone National Park [Online]. *Env Sci Technol* 35: 3302–3309, 2001. <http://www.ncbi.nlm.nih.gov/pubmed/11529568>.
137. Law R. Optimal life histories under age-specific predation. *Am Nat* 114: 399–417, 1979.
138. Liao JC, Hou S-Y, Chao Y-P. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnol Bioeng* 52: 129–40, 1996.
139. Liu Z, Muller J, Li T, Alvey RM, Vogl K, Frigaard NU, Rockwell NC, Boyd ES, Tomsho LP, Schuster SC, Henke P, Rohde M, Overmann J, Bryant DA. Genomic analysis reveals key aspects of prokaryotic symbiosis in the phototrophic consortium “*Chlorochromatium aggregatum*.” *Genome Biol* 14: R127, 2013.
140. Llaneras F, Picó J. Which metabolic pathways generate and characterize the flux space? A comparison among elementary modes, extreme pathways and minimal generators. *J Biomed Biotechnol* 2010: 753904, 2010.
141. Lopez-Ferber M, Simon O, Williams T, Caballero P. Defective or effective? Mutualistic interactions between virus genotypes. *Proc Biol Sci* 270: 2249–2255, 2003.
142. De Lorenzo V, Martinez JL, Asensio C. Microcin-mediated Interactions Between *Klebsiella pneumoniae* and *Escherichia coli* Strains. *Microbiology* 130: 391–400, 1984.
143. Lotka AJ. Contribution to the energetics of evolution [Online]. *Proc Natl Acad Sci U S A* 8: 147–151, 1922. <http://www.ncbi.nlm.nih.gov/pubmed/16576642>.
144. Lotka AJ. Natural selection as a physical principle. *Proc Natl Acad Sci U S A* 8: 151–154, 1922.
145. Ma H-W, Zhao X-M, Yuan Y-J, Zeng A-P. Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. *Bioinformatics* 20: 1870–6, 2004.
146. Macur RE, Langner HW, Kocar BD, Inskeep WP. Linking geochemical processes

with microbial community analysis: successional dynamics in an arsenic-rich, acid-sulphate-chloride geothermal spring. *Geobiology* 2: 163–177, 2004.

147. Madigan MT, Clark DP, Stahl DA, Martinko JM. *Brock Biology of Microorganisms*. 13th ed. Benjamin Cummings, 2010.
148. Mahadevan R, Henson MA. Genome-based modeling and design of metabolic interactions in microbial communities. *Comput Struct Biotechnol J* 3: e201210008, 2012.
149. Mahadevan R, Lovley DR. The degree of redundancy in metabolic genes is linked to mode of metabolism. *Biophys J* 94: 1216–20, 2008.
150. Mallette N, Pankratz EM, Parker AE, Strobel GA, Busse SC, Carlson RP, Peyton BM. Evaluation of cellulose as a substrate for hydrocarbon fuel production by *Ascocoryne sarcoides* (NRRL 50072). *J Sustain Bioenergy Syst* 4: 33–49, 2014.
151. Mallette ND, Knighton B, Strobel GA, Carlson RP, Peyton BM. Resolution of volatile fuel compound profiles from *Ascocoryne sarcoides*: a comparison by proton transfer reaction-mass spectrometry and solid phase microextraction gas chromatography-mass spectrometry. *AMB Express* 2: 23, 2012.
152. Markowitz VM, Chen I-MMA, Palaniappan K, Chu K, Szeto E, Grechkin Y, Ratner A, Jacob B, Huang J, Williams P, Huntemann M, Anderson I, Mavromatis K, Ivanova NN, Kyrpides NC. IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res* 40: D115-22, 2012.
153. Marquez LM, Redman RS, Rodriguez RJ, Roossinck MJ. A virus in a fungus in a plant: three-way symbiosis required for thermal tolerance. *Science* 315: 513–515, 2007.
154. Martyushev L. Entropy and entropy production: Old misconceptions and new breakthroughs. *Entropy* 15: 1152–1170, 2013.
155. Mason IG. Mathematical modelling of the composting process: A review. *Waste Manag* 26: 3–21, 2006.
156. Matsumi R, Atomi H, Driessen AJM, van der Oost J. Isoprenoid biosynthesis in Archaea - Biochemical and evolutionary implications. *Res Microbiol* 162: 39–52, 2011.
157. de Mazancourt C, Schwartz MW. A resource ratio theory of cooperation. *Ecol Lett* 13: 349–359, 2010.
158. McInerney MJ, Sieber JR, Gunsalus RP. Syntrophy in anaerobic global carbon cycles. *Curr Opin Biotechnol* 20: 623–632, 2009.

159. Meyer B, Kuehl J, Deutschbauer AM, Price MN, Arkin AP, Stahl DA. Variation among *Desulfovibrio* species in electron transfer systems used for syntrophic growth. *J Bacteriol* 195: 990–1004, 2013.
160. Mikeskova H, Novotny C, Svobodova K. Interspecific interactions in mixed microbial cultures in a biodegradation perspective. *Appl Microbiol Biotechnol* 95: 861–870, 2012.
161. Minty JJ, Singer ME, Scholz S a, Bae C-HH, Ahn J-HH, Foster CE, Liao JC, Lin XN. Design and characterization of synthetic fungal-bacterial consortia for direct production of isobutanol from cellulosic biomass. *Proc Natl Acad Sci U S A* 110: 14592–14597, 2013.
162. Molenaar D, van Berlo R, de Ridder D, Teusink B. Shifts in growth strategies reflect tradeoffs in cellular economics. *Mol Syst Biol* 5: 323, 2009.
163. Momeni B, Brileya KA, Fields MW, Shou W. Correction: Strong inter-population cooperation leads to partner intermixing in microbial communities. *Elife* 3: e02945, 2014.
164. Morris BEL, Henneberger R, Huber H, Moissl-Eichinger C. Microbial syntrophy: interaction for the common good. *FEMS Microbiol Rev* 37: 384–406, 2013.
165. Morris JJ, Johnson ZI, Szul MJ, Keller M, Zinser ER. Dependence of the cyanobacterium *Prochlorococcus* on hydrogen peroxide scavenging microbes for growth at the ocean's surface. *PLoS One* 6: e16805, 2011.
166. Muller J, Overmann J. Close interspecies interactions between prokaryotes from sulfurous environments. *Front Microbiol* 2: 146, 2011.
167. Muller S, Nebe-von-Caron G. Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities. *FEMS Microbiol Rev* 34: 554–587, 2010.
168. Murray JL, Connell JL, Stacy A, Turner KH, Whiteley M. Mechanisms of synergy in polymicrobial infections. *J Microbiol* 52: 188–199, 2014.
169. Nagarajan H, Embree M, Rotaru A-EE, Shrestha PM, Feist AM, Palsson BØ, Lovley DR, Zengler K. Characterization and modelling of interspecies electron transfer mechanisms and microbial community dynamics of a syntrophic association. *Nat Commun* 4: 2809, 2013.
170. Neidhardt F, Ingraham JL, Schaechter S, Neidhardt FC, Ingraham JL, Schaechter M. Physiology of the bacterial cell: a molecular approach [Online]. 1st ed. Sinauer Associates Inc, Sunderland, MA. <http://www.abebooks.co.uk/Physiology-Bacterial-Cell-Molecular-Approach-Frederick/9026572972/bd>.

171. Nozaki Y, Tanford C. The solubility of amino acids and related compounds in aqueous urea solutions. *J Biol Chem* 238: 4074–81, 1963.
172. Offre P, Spang A, Schleper C. Archaea in biogeochemical cycles. *Annu Rev Microbiol* 67: 437–57, 2013.
173. Oh YG, Lee DY, Lee SY, Park S. Multiobjective flux balancing using the NISE method for metabolic network analysis. *Biotechnol Prog* 25: 999–1008, 2009.
174. Oldstone M. Prevention of type I diabetes in nonobese diabetic mice by virus infection. *Science* 239: 500–502, 1988.
175. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol* 28: 245–8, 2010.
176. Ortiz-Marquez JCF, Do Nascimento M, Zehr JP, Curatti L. Genetic engineering of multispecies microbial cell factories as an alternative for bioenergy production. *Trends Biotechnol* 31: 521–9, 2013.
177. Overmann J. The phototrophic consortium “Chlorochromatium aggregatum” - a model for bacterial heterologous multicellularity. *Adv Exp Med Biol* 675: 15–29, 2010.
178. Overmann J, Tuschak C, Sass H, Frostl J. The ecological niche of the consortium “Pelochromatium roseum.” *Arch microbiol* 169: 120–128, 1998.
179. Paerl HW, Pinckney JL. A mini-review of microbial consortia: Their roles in aquatic production and biogeochemical cycling. *Microb Ecol* 31, 1996.
180. Pande GS, Natrah FM, Sorgeloos P, Bossier P, Defoirdt T. The *Vibrio campbellii* quorum sensing signals have a different impact on virulence of the bacterium towards different crustacean hosts. *Vet Microbiol* 167: 540–545, 2013.
181. Pandhal J, Noirel J. Synthetic microbial ecosystems for biotechnology. *Biotechnol Lett* 36: 1141–1151, 2014.
182. Passarge J, Hol S, Escher M, Huisman J. Competition for nutrients and light: Stable coexistence, alternative stable states, or competitive exclusion? *Ecol Monogr* 76: 57–72, 2006.
183. Pasteur Joubert, J. L. Charbon et septicemie. *Compt Rend Acad Sci* 85: 101–105, 1877.
184. Perrin RK, Fretes NF, Sesmero JP. Efficiency in Midwest US corn ethanol plants: A plant survey. *Energy Policy* 37: 1309–1316, 2009.

185. Pfeffer C, Larsen S, Song J, Dong M, Besenbacher F, Meyer RL, Kjeldsen KU, Schreiber L, Gorby YA, El-Naggar MY, Leung KM, Schramm A, Risgaard-Petersen N, Nielsen LP. Filamentous bacteria transport electrons over centimetre distances. *Nature* 491: 218–221, 2012.
186. Phelan V V, Liu WT, Pogliano K, Dorrestein PC. Microbial metabolic exchange--the chemotype-to-phenotype link. *Nat Chem Biol* 8: 26–35, 2012.
187. Pimm SL. The complexity and stability of ecosystems. *Nature* 307: 321–326, 1984.
188. Pirbadian S, Barchinger SE, Leung KM, Byun HS, Jangir Y, Bouhenni RA, Reed SB, Romine MF, Saffarini DA, Shi L, Gorby YA, Golbeck JH, El-Naggar MY. *Shewanella oneidensis* MR-1 nanowires are outer membrane and periplasmic extensions of the extracellular electron transport components. *Proc. Natl. Acad. Sci.* (August 20, 2014). doi: 10.1073/pnas.1410551111.
189. Poltak SR, Cooper VS. Ecological succession in long-term experimentally evolved biofilms produces synergistic communities. *ISME J* 5: 369–378, 2011.
190. Poolman MG, Sebu C, Pidcock MK, Fell DA. Modular decomposition of metabolic systems via null-space analysis. *J Theor Biol* 249: 691–705, 2007.
191. Prangishvili D. The wonderful world of archaeal viruses. *Annu Rev Microbiol* 67: 565–585, 2013.
192. Price ND, Reed JL, Palsson BØ. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol* 2: 886–97, 2004.
193. Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, Green JL, Green LE, Killham K, Lennon JJ, Osborn AM, Solan M, van der Gast CJ, Young JPW. The role of ecological theory in microbial ecology. *Nat Rev Microbiol* 5: 384–392, 2007.
194. Rahman MM, Mostafiz SB, Paatero J V., Lahdelma R. Extension of energy crops on surplus agricultural lands: A potentially viable option in developing countries while fossil fuel reserves are diminishing. *Renew Sustain Energy Rev* 29: 108–119, 2014.
195. Reed JL, Palsson BO. Thirteen years of building constraint-based *in silico* models of *Escherichia coli*. *J Bacteriol* 185: 2692–9, 2003.
196. Regalbuto JR. Cellulosic biofuels--got gasoline? *Science* 325: 822–4, 2009.
197. Rodionova IA, Li X, Plymale AE, Motamedchaboki K, Konopka AE, Romine MF, Fredrickson JK, Osterman AL, Rodionov DA. Genomic distribution of B-vitamin

- auxotrophy and uptake transporters in environmental bacteria from the *Chloroflexi* phylum. *Environ Microbiol Rep* 7: 204–210, 2015.
198. Rohwer F, Prangishvili D, Lindell D. Roles of viruses in the environment. *Environ Microbiol* 11: 2771–4, 2009.
 199. Roossinck MJ. The good viruses: viral mutualistic symbioses. *Nat Rev Microbiol* 9: 99–108, 2011.
 200. Rosenzweig RF, Sharp RR, Treves DS, Adams J. Microbial evolution in a simple unstructured environment - genetic differentiation in *Escherichia Coli*. *Genetics* 137: 903–917, 1994.
 201. De Roy K, Marzorati M, Van den Abbeele P, Van de Wiele T, Boon N. Synthetic microbial ecosystems: an exciting tool to understand and apply microbial communities. *Env Microbiol* 16: 1472–1481, 2014.
 202. Runguphan W, Keasling JD. Metabolic engineering of *Saccharomyces cerevisiae* for production of fatty acid-derived biofuels and chemicals. *Metab Eng* 21: 103–13, 2014.
 203. Rykiel EJ. Towards a definition of ecological disturbance. *Austral Ecol* 10: 361–365, 1985.
 204. Sabra W, Dietz D, Tjahjajari D, Zeng A-P. Biosystems analysis and engineering of microbial consortia for industrial biotechnology. *Eng Life Sci* 10: 407–421, 2010.
 205. Sadeghinezhad E, Kazi SN, Sadeghinejad F, Badarudin A, Mehrali M, Sadri R, Reza Safaei M. A comprehensive literature review of bio-fuel performance in internal combustion engine and relevant costs involvement. *Renew Sustain Energy Rev* 30: 29–44, 2014.
 206. Sandaa RA, Gomez-Consarnau L, Pinhassi J, Riemann L, Malits A, Weinbauer MG, Gasol JM, Thingstad TF. Viral control of bacterial biodiversity--evidence from a nutrient-enriched marine mesocosm experiment. *Env Microbiol* 11: 2585–2597, 2009.
 207. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S, Kang J, Hyduke DR, Palsson BØ. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* 6: 1290–1307, 2011.
 208. Schilling CH, Letscher D, Palsson BØ. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *J Theor Biol* 203: 229–48, 2000.

209. Schilling CH, Palsson BØ. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *J Theor Biol* 203: 249–83, 2000.
210. Schink B. Energetics of syntrophic cooperation in methanogenic degradation [Online]. *Microbiol Mol Biol Rev* 61: 262–280, 1997. <http://mmbbr.asm.org/content/61/2/262.short> [27 Oct. 2014].
211. Schirmer A, Rude MA, Li X, Popova E, del Cardayre SB. Microbial biosynthesis of alkanes. *Science* 329: 559–62, 2010.
212. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U. Multidimensional optimality of microbial metabolism. *Science* 336: 601–604, 2012.
213. Schulz S, Dickschat JS. Bacterial volatiles: the smell of small organisms. *Nat Prod Rep* 24: 814–42, 2007.
214. Schuster S, Fell D a, Dandekar T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat Biotechnol* 18: 326–32, 2000.
215. Schuster S, de Figueiredo LF, Kaleta C. Predicting novel pathways in genome-scale metabolic networks. *Biochem Soc Trans* 38: 1202–5, 2010.
216. Schuster S, Hilgetag C. On elementary flux modes in biochemical reaction systems at steady state. *J Biol Syst* 2: 165–82, 1994.
217. Schuster S, Pfeiffer T, Moldenhauer F, Koch I, Dandekar T. Exploring the pathway structure of metabolism: decomposition into subnetworks and application to *Mycoplasma pneumoniae*. [Online]. *Bioinformatics* 18: 351–61, 2002. <http://www.ncbi.nlm.nih.gov/pubmed/11847093>.
218. Sciubba E. What did Lotka really say? A critical reassessment of the “maximum power principle.” *Ecol Modell* 222: 1347–1353, 2011.
219. Searchinger T, Heimlich R, Houghton RA, Dong F, Elobeid A, Fabiosa J, Tokgoz S, Hayes D, Yu T-H. Use of U.S. croplands for biofuels increases greenhouse gases through emissions from land-use change. *Science* 319: 1238–40, 2008.
220. Shade A, Peter H, Allison SD, Baho DL, Berga M, Burgmann H, Huber DH, Langenheder S, Lennon JT, Martiny JB, Matulich KL, Schmidt TM, Handelsman J. Fundamentals of microbial community resistance and resilience. *Front Microbiol* 3: 417, 2012.
221. Shiba Y, Paradise EM, Kirby J, Ro D-K, Keasling JD. Engineering of the pyruvate dehydrogenase bypass in *Saccharomyces cerevisiae* for high-level production of

- isoprenoids. *Metab Eng* 9: 160–8, 2007.
222. Shong J, Jimenez Diaz MR, Collins CH. Towards synthetic microbial consortia for bioprocessing. *Curr Opin Biotechnol* 23: 798–802, 2012.
223. Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, Dekel E, Kavanagh K, Alon U. Evolutionary trade-offs, Pareto optimality, and the geometry of phenotype space. *Science* 336: 1157–1160, 2012.
224. Shrihari, Kumar R, Gandhi KS. Modeling of Fe-2+ Oxidation by *Thiobacillus ferrooxidans*. *Appl Microbiol Biotechnol* 33: 524–528, 1990.
225. Sieber JR, McInerney MJ, Gunsalus RP. Genomic insights into syntrophy: the paradigm for anaerobic metabolic cooperation. *Annu Rev Microbiol* 66: 429–452, 2012.
226. Smid EJ, Lacroix C. Microbe-microbe interactions in mixed culture food fermentations. *Curr Opin Biotechnol* 24: 148–154, 2013.
227. Soderberg T. Biosynthesis of ribose-5-phosphate and erythrose-4-phosphate in archaea: a phylogenetic analysis of archaeal genomes [Online]. *Archaea* 1: 347–352, 2005.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2685555&tool=pmcentrez&rendertype=abstract>.
228. Song H-S, Cannon W, Beliaev AS, Konopka A. Mathematical modeling of microbial community dynamics: A methodological review. *Processes* 2: 711–752, 2014.
229. Song H, Ding M-ZZ, Jia X-QQ, Ma Q, Yuan Y-JJ. Synthetic microbial consortia: from systematic analysis to construction and applications. *Chem Soc Rev* (July 14, 2014). doi: 10.1039/c4cs00114a.
230. Stams AJM, Plugge CM. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. *Nat Rev Microbiol* 7: 568–77, 2009.
231. Stashenko EE, Martínez JR. Sampling volatile compounds from natural products with headspace/solid-phase micro-extraction. *J Biochem Biophys Methods* 70: 235–242, 2007.
232. Sterner RW, Elser JJ. Ecological stoichiometry: The biology of elements from molecules to the biosphere [Online]. Princeton University Press.
<http://books.google.com/books?id=53NTDvppdYUC>.
233. Stinson M, Ezra D, Hess WM, Sears J, Strobel G. An endophytic *Gliocladium sp.* of *Eucryphia cordifolia* producing selective volatile antimicrobial compounds.

- Plant Sci* 165: 913–922, 2003.
234. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, Stahl DA. Metabolic modeling of a mutualistic microbial community. *Mol Syst Biol* 3: 92, 2007.
 235. Stomp M, Huisman J, De Jongh F, Veraart AJ, Gerla D, Rijkeboer M, Ibelings BW, Wollenzien UI, Stal LJ. Adaptive divergence in pigment composition promotes phytoplankton biodiversity. *Nature* 432: 104–107, 2004.
 236. Stoodley P, Sauer K, Davies DG, Costerton JW. Biofilms as complex differentiated communities. *Annu Rev Microbiol* 56: 187–209, 2002.
 237. Strobel GA, Daisy B, Castillo UF, Harper J. Natural products from endophytic microorganisms. *J Nat Prod* 67: 257–68, 2004.
 238. Strobel GA, Knighton B, Kluck K, Ren Y, Livinghouse T, Griffin MA, Spakowicz DJ, Sears J. The production of myco-diesel hydrocarbons and their derivatives by the endophytic fungus *Gliocladium roseum* (NRRL 50072). *Microbiology* 154: 3319–28, 2008.
 239. Subashchandrabose SR, Ramakrishnan B, Megharaj M, Venkateswarlu K, Naidu R. Consortia of cyanobacteria/microalgae and bacteria: biotechnological potential. *Biotechnol Adv* 29: 896–907, 2011.
 240. Suttle CA. Marine viruses - major players in the global ecosystem. *Nat Rev Microbiol* 5: 801–812, 2007.
 241. Tachibana M, Allen AE, Kikutani S, Endo Y, Bowler C, Matsuda Y. Localization of putative carbonic anhydrases in two marine diatoms, *Phaeodactylum tricornutum* and *Thalassiosira pseudonana*. *Photosynth Res* 109: 205–21, 2011.
 242. Taffs R, Aston JE, Brileya K, Jay Z, Klatt CG, McGlynn S, Mallette N, Montross S, Gerlach R, Inskeep WP, Ward DM, Carlson RP. *In silico* approaches to study mass and energy flows in microbial consortia: a syntrophic case study. *BMC Syst Biol* 3: 114, 2009.
 243. Takacs-Vesbach C, Inskeep WP, Jay ZJ, Herrgard MJ, Rusch DB, Tringe SG, Kozubal MA, Hamamura N, Macur RE, Fouke BW, Reysenbach AL, McDermott TR, Jennings R, Hengartner NW, Xie G. Metagenome sequence analysis of filamentous microbial communities obtained from geochemically distinct geothermal channels reveals specialization of three aquificales lineages. *Front Microbiol* 4: 84, 2013.
 244. Tamime AY, Robinson RK. *Yoghurt: Science and technology*. Woodhead Publishing, 1999.

245. Tan CH, Koh KS, Xie C, Tay M, Zhou Y, Williams R, Ng WJ, Rice SA, Kjelleberg S. The role of quorum sensing signalling in EPS production and the assembly of a sludge community into aerobic granules. *ISME J* 8: 1186–1197, 2014.
246. Taylor JB. The water solubilities and heats of solution of short chain cellulosic oligosaccharides. *Trans Faraday Soc* 53: 1198, 1957.
247. Taylor PA, le BWPJ. Theoretical studies on the coexistence of competing species under continuous-flow conditions [Online]. *Can J Microbiol* 21: 90–98, 1975. <http://www.ncbi.nlm.nih.gov/pubmed/1116041>.
248. Tegelaar EW, de Leeuw JW, Derenne S, Largeau C. A reappraisal of kerogen formation. *Geochim Cosmochim Acta* 53: 3103–6, 1989.
249. Tehlivets O, Scheuringer K, Kohlwein SD. Fatty acid synthesis and elongation in yeast. *Biochim. Biophys. Acta - Mol. Cell Biol. Lipids* 1771: 255–270, 2007.
250. Tepper N, Noor E, Amador-Noguez D, Haraldsdottir HS, Milo R, Rabinowitz J, Liebermeister W, Shlomi T. Steady-state metabolite concentrations reflect a balance between maximizing enzyme efficiency and minimizing total metabolite load. *PLoS One* 8: e75370, 2013.
251. Terzer M. Accelerating the computation of elementary modes using pattern trees. *Algorithms Bioinforma* 2006: 333–43, 2006.
252. Terzer M, Maynard ND, Covert MW, Stelling J. Genome-scale metabolic networks. *Wiley Interdiscip Rev Syst Biol Med* 1: 285–97, 2009.
253. Terzer M, Stelling J. Large-scale computation of elementary flux modes with bit pattern trees. *Bioinformatics* 24: 2229–35, 2008.
254. Tetko I V., Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin VA, Radchenko E V., Zefirov NS, Makarenko AS, Tanchuk VY, Prokopenko V V. Virtual Computational Chemistry Laboratory – Design and Description. *J Comput Aided Mol Des* 19: 453–463, 2005.
255. Thauer RK, Kaster A-KK, Seedorf H, Buckel W, Hedderich R. Methanogenic archaea: ecologically relevant differences in energy conservation. *Nat Rev Microbiol* 6: 579–91, 2008.
256. Thiele I, Heinken A, Fleming RMT. A systems biology approach to studying the role of microbes in human health. *Curr Opin Biotechnol* 24: 4–12, 2013.
257. Tillmann HL, Heiken H, Knapik-Botor A, Heringlake S, Ockenga J, Wilber JC, Goergen B, Detmer J, McMorrow M, Stoll M, Schmidt RE, Manns MP. Infection

- with GB virus C and reduced mortality among HIV-infected patients. *N Engl J Med* 345: 715–724, 2001.
258. Tilman D. Resources - a graphical-mechanistic approach to competition and predation. *Am Nat* 116: 362–393, 1980.
259. Tilman D. Tests of resource competition theory using four species of Lake Michigan algae. *Ecology* 62: 802, 1981.
260. Tilman D. *Resource competition and community structure*. Princeton: Princeton University Press, 1982.
261. Treves DS, Manning S, Adams J. Repeated evolution of an acetate-crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Mol Biol Evol* 15: 789–797, 1998.
262. Trinh CT, Wlaschin A, Srieenc F. Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. *Appl Microbiol Biotechnol* 81: 813–26, 2009.
263. Urbanczik R. Enumerating constrained elementary flux vectors of metabolic networks. *IET Syst Biol* 1: 274–279, 2007.
264. Urbanczik R, Wagner C. An improved algorithm for stoichiometric network analysis: theory and applications. *Bioinformatics* 21: 1203–10, 2005.
265. Urbanczik R, Wagner C. Functional stoichiometric analysis of metabolic networks. *Bioinformatics* 21: 4176–80, 2005.
266. Valenzuela J, Mazurie A, Carlson RP, Gerlach R, Cooksey KE, Peyton BM, Fields MW. Potential role of multiple carbon fixation pathways during lipid accumulation in *Phaeodactylum tricorutum*. *Biotechnol Biofuels* 5: 40, 2012.
267. Varma A, Boesch BW, Palsson BØ. Stoichiometric interpretation of *Escherichia coli* glucose catabolism under various oxygenation rates. [Online]. *Appl Environ Microbiol* 59: 2465–73, 1993.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=182307&tool=pmcentrez&rendertype=abstract>.
268. Varma A, Palsson BØ. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type *Escherichia coli* W3110. [Online]. *Appl Environ Microbiol* 60: 3724–31, 1994.
<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=201879&tool=pmcentrez&rendertype=abstract>.
269. Verma VC, Kharwar RN, Strobel GA. Chemical and functional diversity of natural

- products from plant associated endophytic fungi. [Online]. *Nat Prod Commun* 4: 1511–1532, 2009. <http://www.cabdirect.org/abstracts/20093357817.html> [19 Oct. 2013].
270. Verwoerd WS. A new computational method to split large biochemical networks into coherent subnets. *BMC Syst Biol* 5: 25, 2011.
271. Vogl K, Glaeser J, Pfannes KR, Wanner G, Overmann J. *Chlorobium chlorochromatii* sp. nov., a symbiotic green sulfur bacterium isolated from the phototrophic consortium “Chlorochromatium aggregatum.” *Arch microbiol* 185: 363–372, 2006.
272. Waide RB, Willig MR, Steiner CF, Mittelbach G, Gough L, Dodson SI, Juday GP, Parmenter R. The relationship between productivity and species richness. *Annu Rev Ecol Syst* 30: 257–300, 1999.
273. Walker CB, Redding-Johanson AM, Baidoo EE, Rajeev L, He Z, Hendrickson EL, Joachimiak MP, Stolyar S, Arkin AP, Leigh JA, Zhou J, Keasling JD, Mukhopadhyay A, Stahl DA. Functional responses of methanogenic archaea to syntrophic growth. *ISME J* 6: 2045–2055, 2012.
274. Wanner G, Vogl K, Overmann J. Ultrastructural characterization of the prokaryotic symbiosis in “Chlorochromatium aggregatum.” *J Bacteriol* 190: 3721–3730, 2008.
275. Waters CM, Bassler BL. Quorum sensing: cell-to-cell communication in bacteria. *Annu Rev Cell Dev Biol* 21: 319–346, 2005.
276. Wilson JB, Spijkerman E, Huisman J. Is there really insufficient support for Tilman’s R* concept? A comment on Miller et al. *Am Nat* 169: 700–706, 2007.
277. Wintermute EH, Silver PA. Emergent cooperation in microbial metabolism. *Mol Syst Biol* 6: 407, 2010.
278. Wolfe AJ. The acetate switch. *Microbiol Mol Biol Rev* 69: 12–50, 2005.
279. Xiangzhen L, McInerney MJ, Stahl DA, Krumholz LR. Metabolism of H₂ by *Desulfovibrio alaskensis* G20 during syntrophic growth on lactate. *Microbiology* 157: 2912–2921, 2011.
280. Young M, Wiedenheft B, Snyder J, Spuhler J, Roberto F, Douglas T. Archaeal viruses from Yellowstone’s high-temperature environments. *Geotherm. Biol. Geochemistry Yellowstone Natl. Park.* .
281. Zhang CC, Laurent S, Sakr S, Peng L, Bedu S. Heterocyst differentiation and pattern formation in cyanobacteria: a chorus of signals. *Mol Microbiol* 59: 367–

- 375, 2006.
282. Zhang Y-HP, Lynd LR. Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. *Proc Natl Acad Sci* 102: 7321–5, 2005.
283. Zhuang K, Izallalen M, Mouser P, Richter H, Risso C, Mahadevan R, Lovley DR. Genome-scale dynamic modeling of the competition between *Rhodofex* and *Geobacter* in anoxic subsurface environments. *ISME J* 5: 305–16, 2011.
284. Zinn M, Witholt B, Egli T. Dual nutrient limited growth: models, experimental observations, and applications. *J Biotechnol* 113: 263–279, 2004.
285. Zomorodi AR, Islam MM, Maranas CD. d-OptCom: Dynamic multi-level and multi-objective metabolic modeling of microbial communities. *ACS Synth Biol* 3: 247–257, 2014.
286. Zomorodi AR, Maranas CD. OptCom: a multi-level optimization framework for the metabolic modeling and analysis of microbial communities. *PLoS Comput Biol* 8: e1002363, 2012.
287. Zuroff TR, Curtis WR. Developing symbiotic consortia for lignocellulosic biofuel production. *Appl Microbiol Biotechnol* 93: 1423–1435, 2012.

APPENDICES

APPENDIX A

ADDITIONAL PUBLICATIONS

The included publications account for extraneous work not directly linked to this dissertation, but performed during its duration.

Citation

Barnhart, E.P., McClure, M.A., Johnson, K., Cleveland, S., Hunt, K.A., Fields, M.W. (2015) Potential Role of Acetyl-CoA Synthetase (*acs*) and Malate Dehydrogenase (*mae*) in the Evolution of the Acetate Switch in Bacteria and Archaea. *Sci. Rep.* 5, 12498; doi: 10.1038/srep12498.

Abstract

Although many Archaea have AMP-Acs (acetyl-coenzyme A synthetase) and ADP-Acs, the extant methanogenic genus *Methanosarcina* is the only identified Archaeal genus that can utilize acetate via acetate kinase (*Ack*) and phosphotransacetylase (*Pta*). Despite the importance of *ack* as the potential urkinase in the ASKHA phosphotransferase superfamily, an origin hypothesis does not exist for the acetate kinase in Bacteria, Archaea, or Eukarya. Here we demonstrate that Archaeal AMP-Acs and ADP-Acs contain paralogous ATPase motifs previously identified in *Ack*, which demonstrate a novel relation between these proteins in Archaea. The identification of ATPase motif conservation and resulting structural features in AMP- and ADP-acetyl-CoA synthetase proteins in this study expand the ASKHA superfamily to include acetyl-CoA synthetase. Additional phylogenetic analysis showed that *Pta* and *MaeB* sequences had a common ancestor, and that the *Pta* lineage within the halophilic archaea was an ancestral lineage. These results suggested that divergence of a duplicated *maeB* within an ancient halophilic, archaeal lineage formed a putative *pta* ancestor. These results provide a

potential scenario for the establishment of the Ack/Pta pathway and provide novel insight into the evolution of acetate metabolism for all three domains of life.

Citation

Schoen, H.R., Hunt, K.A., Strobel, G., Peyton, B., Carlson, R.P. (2016) Carbon chain length of biofuel- and flavor-relevant hydrocarbons produced by lignocellulolytic fungal endophytes vary with culturing temperature. *Mycoscience*, In review

Abstract

Objective: To determine hydrocarbon synthesis profiles, substrate utilization and growth properties of three consolidated bioprocessing-relevant, fungal endophyte isolates from the genus *Nodulisporium*.

Results: The fungi synthesized a number of long and short-chain hydrocarbons, including eucalyptol; 1-butanol, 3-methyl; 1-octen-3-ol; and benzaldehyde, with potential applications as biofuel or flavor compounds. All three fungi grew on a wide range of carbon substrates ranging from simple sugars to waste biomass sources, including wood chips, corn stover, and sugar beet pulp, highlighting potential consolidated bioprocessing applications. The fungi grew faster, with doubling times less than a day, and produced more concentrated biomass, up to 19 g cell dry weight per liter, than the previously reported hydrocarbon-producing fungal endophytes.

Conclusion: Three *Nodulisporium* isolates produced hydrocarbons with potential biofuel or flavor compound applications; the results expand the number of described hydrocarbon-producing fungi and provide prerequisite data for choosing strains for optimization and genetic modification efforts.

APPENDIX B

SUPPLEMENTAL MATERIAL FOR CHAPTER 2

Treatment of Numerical Error

Numerical errors are a common computational problem. The use of bit-masking in EFMTTool element representation could expand the numerical precision of the algorithm and relax the coefficient rule (Table 2-2). When numerical errors occur in EFMTTool, they result in lost EFMs. This can occur when the digit definition is insufficient to distinguish between true zero and coefficients with very small, but nonzero magnitudes. Algorithm definitions account for these errors predictably and change nonzero coefficients below a threshold to zeros. These changes justify the coefficient rule since maintaining skewed coefficient ratios throughout the calculation, as done by moving enforced reactions to the bottom of the matrix (Figure 2-1, Steps 9-12), increases the likelihood of small numbers. However, active rescaling of coefficients within a potential EFM after a step is bit-masked (75) may increase the acceptable coefficient range within a reaction. As the EFMTTool algorithm bit-masks values, differences within column values may be hidden, and since each EFM is only unique to a constant and bit-masked values do not have magnitudes, the non-bit-masked portions of each column could be rescaled such that values are more numerically compatible. The increased quantity of nonzero values may then enumerate additional EFMs since a large component of the algorithm depends on whether a value is positive, negative, or zero. This may also relax the restrictions of the coefficient rule (Table 2-2) by decreasing sensitivity to row arrangement, permitting a larger range of potential reactions for splitting. The permitted coefficient range applied during this study was empirically derived and was expected to be conservative.

Minimizing Failed Subnetwork Run time

Failed subnetworks can have a high CPU time cost; therefore, large run time savings can be obtained by minimizing CPU time spent on intractable subnetworks. A detection strategy was applied to identify subnetworks that were not likely to complete and terminate those attempts before failure. The detection strategy used a predicted CPU time, D_{model} , to process iteration $n+1$ of the nullspace algorithm (used by EFMTTool to enumerate EFMs) as an estimate of model difficulty, defined as:

$$D_{model} = \frac{adj_{n+1} \left[\left(n \sum_1^n adj_j t_j \right) - \left(\sum_1^n adj_j \sum_1^n t_j \right) \right]}{\left(n \sum_1^n adj_j^2 \right) - \left(\sum_1^n adj_j \right)^2} \quad (1)$$

where t_j is the time in seconds to process adj_j potential EFMs (i.e. adjacencies) produced by iteration j , n is the total number of iterations completed, and adj_{n+1} is the number of adjacencies in the current incomplete iteration. This definition of model difficulty permitted robust model difficulty assessment for initial iterations where processing time per adjacency is unstable or late iterations where processing time per adjacency is nonlinear (data not shown). Model difficulty was calculated every iteration after an initial two seconds of run time. To continue the enumeration, the model difficulty was required to be less than the empirical threshold:

$$D_{threshold} = 3240 * RAM + 6794280 \quad (2)$$

where RAM is the megabytes of RAM available for calculations. This threshold could be raised or lowered depending on the effective cost of failed subnetworks. Additionally, the threshold may not be linearly related to available RAM over the whole range of possible computer configurations, but was appropriate for the presented work. Applying

the difficulty detection strategy resulted in substantial time savings for the models and computational cluster used in this study.

Additional Model Descriptions

The models described in Table 2-1 of the main text are represented here with additional network information. The Reactions / Reversible and the Metabolites / External categories include the calculated values for the models after network compression. Compression combines sequences of reactions that always occur together into aggregate reaction representations reducing the computational requirements. The convex basis vectors (CBVs) for each model were also enumerated to approximate the model structures. The number of EFMs and CBVs are sensitive to network structure, as well as the number of reactions and external metabolites.

Table 2-S1. Metabolic models used in the presented study.

Organism	Model	Reactions / Reversible ^a	Metabolites / External ^a	CBVs ^b	EFMs
<i>S. cerevisiae</i>	YEAST1 ^c	78 / 31 (54 / 17)	62 / 11 (34 / 0)	11603	1,515,315 ^f
<i>S. cerevisiae</i>	YEAST2 ^c	83 / 34 (61 / 23)	63 / 11 (39 / 0)	13567	68,868,602 ^f
<i>E. coli</i>	ECOLI ^d	95 / 59 (59 / 26)	94 / 23 (37 / 0)	174	226,269,020 ^f
<i>P. tricornutum</i>	DIATOM ^e	318 / 103 (74 / 31)	335 / 17 (64 / 0)	145958	1,934,729,551 ^g

^a Original network (compressed network) descriptors calculated using EFMTTool compression

^b Convex basis vectors (CBVs) calculated using CellNetAnalyzer (119)

^c (109) ^d (112) ^e This Work

^f EFMs validated by comparing split and unsplit network results.

^g 14 reactions used for iterative splitting. Some subnetworks were recalculated varying the number of reactions for splitting to validate results.

Java and EFMTTool Options Used

The command line switches used to execute EFMTTool and improve its performance are described in Table 2-S2. The EFMTTool options used are shown in gray.

Table 2-S2. Description of Java and EFMTool options used

Command	Description
-Xmx7g	Specifies RAM available for EFM calculation (i.e. 7 GB)
-XX:+UseParallelOldGC	For multi-core machines
-XX:+UseAdaptiveSizePolicy	Permits memory reallocation during calculation
-XX:+UseTLAB -XX:+ResizeTLAB	For multi-core machines
-XX:MaxPermSize=15m	Optional: EFMTool was not found to use more than 15 MB

Example command line with Java (bold) and EFMTool options (gray):

```
JAVA -Xmx7g -XX:+UseParallelOldGC -XX:+UseAdaptiveSizePolicy -
XX:+UseTLAB -XX:+ResizeTLAB -XX:MaxPermSize=15m -cp lib\metabolic-efm-
all.jar;lib\dom4j-1.6.1.jar;lib\junit-3.8.1.jar
ch.javasoft.metabolic.efm.main.CalculateFluxModes -kind stoichiometry -stoich
tmp\stoich.txt -rev tmp\revs.txt -meta tmp\mnames.txt -reac tmp\rnames.txt -arithmetic
double -zero 1e-010 -out matlab tmp\efms.mat -compression default -log console -level
INFO -tmpdir \tmp -maxthreads -1 -normalize min -adjacency-method pattern-tree-
minzero -rowordering MostZerosOrAbsLexMin
```

k-1 EFM Correction

It should be noted that, to date, EFMTool removes EFMs which have k-1 reactions enforced, where k is the number of reactions with nonzero fluxes in a given EFM. This occurs when the futile-cycle removing subroutine only considers the bit-masked part of EFMs, missing fluxes being enforced which are not bit-masked at that stage in the algorithm. This was repaired by allowing the subroutine to take into account all components of the EFM. In the supplemental file

“FutileCycleColumnFilterchanged.java” line 65, the code should read

```
if (model.getFinalNumericSize() != 0) {
    not
    if (model.getFinalNumericSize() == 0) {
```

as found in the original code obtained for EFMTool version 4.7.1. Upon replacement, EFMTool should be recompiled. The change is shown below. A java file with the fix,

FutileCycleColumnFilterchanged.java, has been provided as part of these supplemental materials, which is ready for replacing and recompiling of EFMTool.

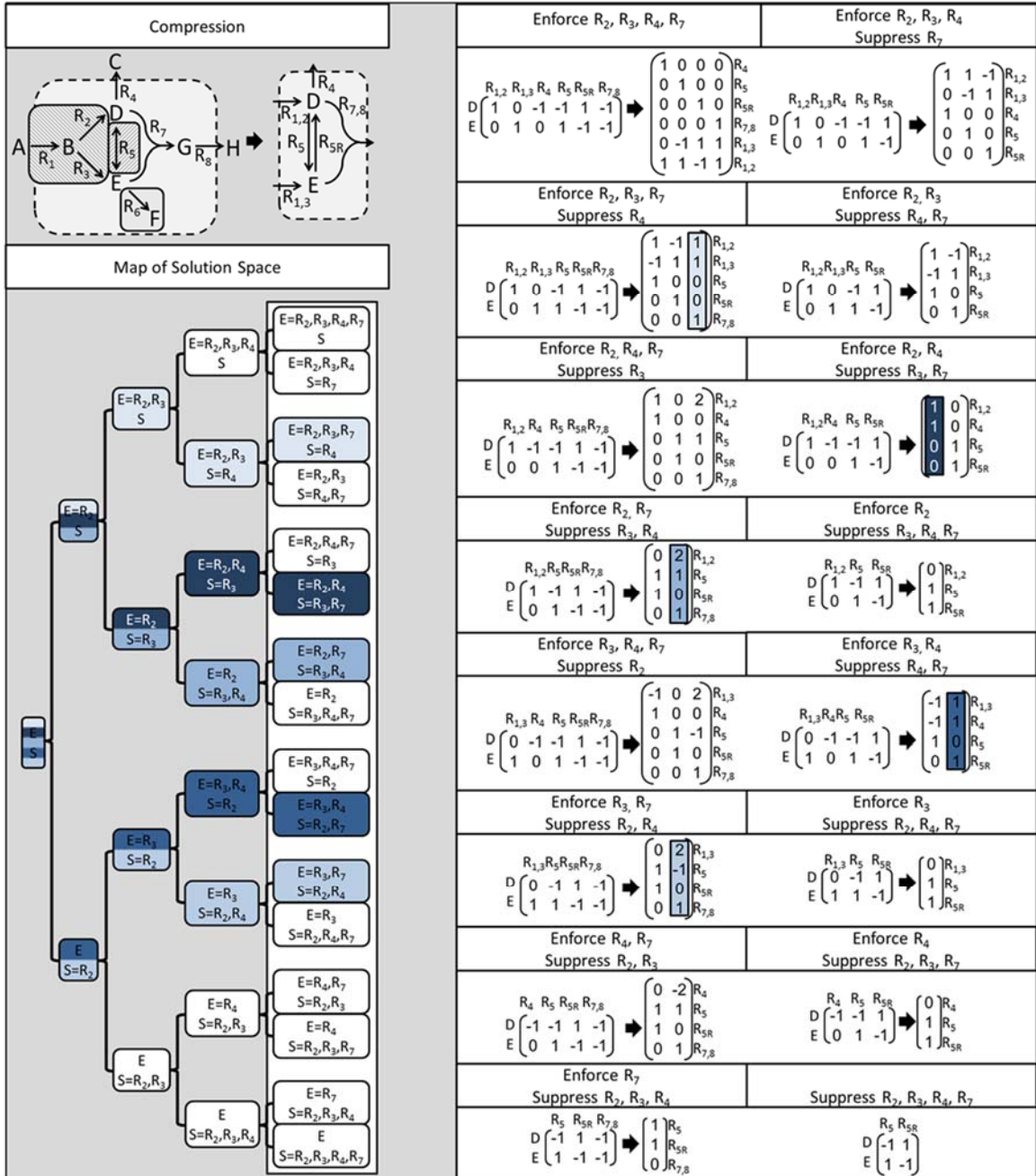





Figure 2-S1. Application of the splitting algorithm. An example model in original (left) and compressed (right) format is shown in the upper left panel titled Compression. The depicted model was split using the maximum number of reactions for splitting that satisfy the criteria in Table 2-3 ($R=4$); the subnetwork definitions are shown in the Map of

Figure 2-S1 Continued

Solution Space (lower left panel). As demonstrated, all possible combinations of reactions were examined, covering the entire solution space. The right panel presents every subnetwork defined in the last column of the Map of the Solution Space in the same order as they appear in the lower left panel. The subnetworks are represented by stoichiometric matrices (left) and nullspace basis vectors (right). Shaded nullspace vectors in the right panel represent EFMs in the subnetwork of the same shade; nullspace vectors that are not shaded violate the subnetwork definition and would be removed. A similar result was obtainable using fewer reactions to define subnetworks as shown by the multishade boxes. While R7 and R8 individually satisfy the criteria in Table 2-3; as described in section 2-6-1, using both reactions for splitting results in dead-end reactions, violating rules 1 and 3. The highlighted examples of a combined reaction in , reversible reaction in , and dead-end reaction in  are to be avoided for use in splitting.

Example of Splitting Algorithm

An additional example of the splitting approach is provided in Figure 2-S1.

Additional Supplementary Files

Supplemental materials are available at Bioinformatics online.

APPENDIX C

SUPPLEMENTAL MATERIAL FOR CHAPTER 4

Basic Stoichiometric Modeling Concepts

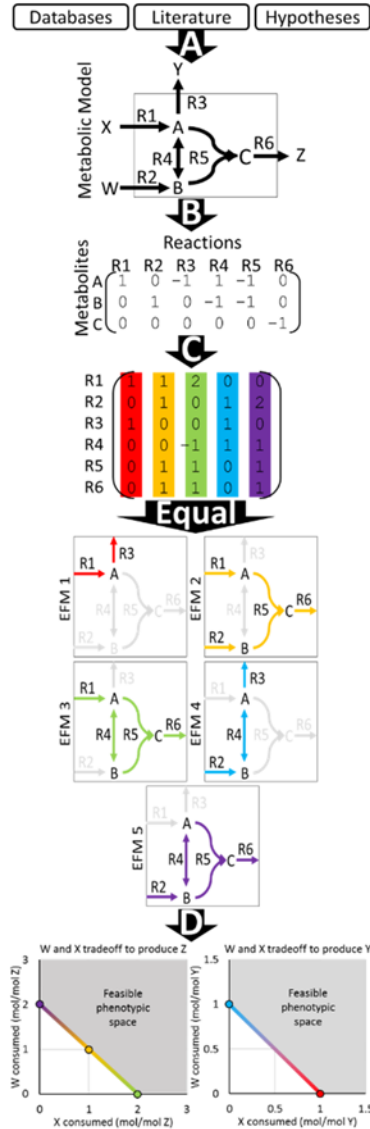


Figure 4-S1. Summary of metabolic model building, stoichiometric analysis, and application of ecological theory. A) The metabolic model is a compilation of genomic annotations from databases, experimental yields and efficiencies reported in the literature, and hypothesized reactions. Hypothesized reactions include those missing from annotations, but expected due to literature data (*i.e.* an organism that grows on a mineral medium has complete biosynthetic pathways). B) The model is represented as a series of reaction equations correlating the consumed metabolites and the produced metabolites (*i.e.* R_i , where i corresponds to the reaction number and is written in terms of metabolites A, B and C; W, X, Y, and Z are sources and sinks). The reaction equations can be written concisely as a mathematical matrix, positive coefficients are reaction products and negative coefficients are reaction substrates. C) These reaction equations are analyzed to

Figure 4-S1 Continued

identify metabolic routes through the network that are stoichiometrically balanced (i.e. all metabolites produced must be consumed and *vice versa*) and all reactions are limited to directions permitted by chemical thermodynamics. In flux balance analysis (FBA), these routes must also satisfy an objective function (e.g. maximize flux through R6 while constraining the magnitude of flux through reactions R1 and R2). Elementary flux mode (EFM) analysis finds all of the simplest, genetically distinct routes through the network (i.e. the five possible EFMs (color coded) for the sample metabolic model). D) Application of ecological theory to the analysis of these routes predicts competitive metabolic behavior. The theory assumes the reactions used to produce the desired product (e.g. biomass or cellular energy) will minimize the resource cost for the limiting nutrient (e.g. carbon, oxygen, or cellular volume). In the context of the depicted example, if W is limiting and a population must produce Z, theory states the optimal strategy would be along the x-axis (plot to the left). Conversely, limitation of X to produce Z would be along the y-axis (plot to the left). In addition, alternative products may favor alternative metabolic strategies (e.g. production of Y in the plot to the right). Additional review articles on stoichiometric modeling: (11, 65, 192, 262).

Serine-glycine Pathway and Tetrahydrofolate Regeneration

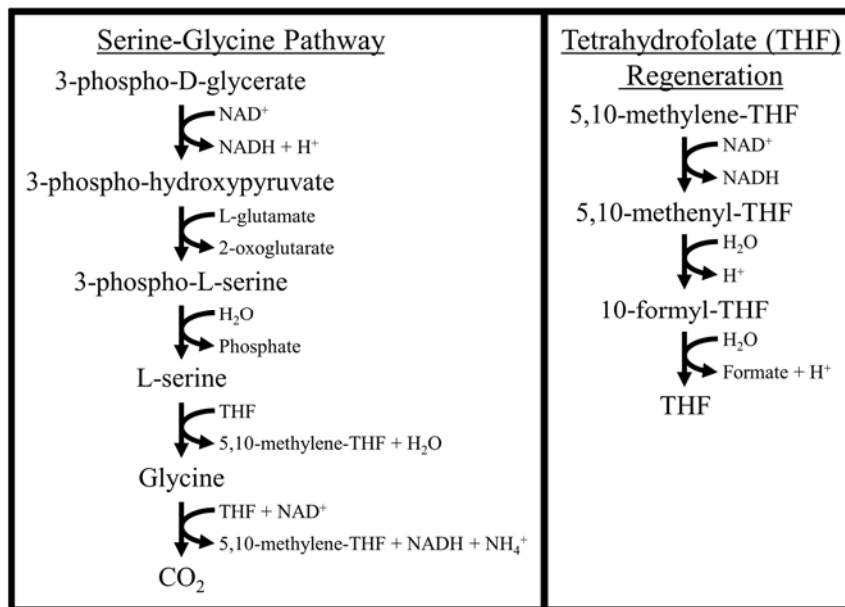


Figure 4-S2. Graphical representation of the serine-glycine pathway and tetrahydrofolate regeneration. The net pathway of serine-glycine and tetrahydrofolate regeneration results in formate and reducing equivalents. If the formate produced is then oxidized, the net result is electronically identical to complete oxidation through the tricarboxylic acid cycle but uses fewer enzymes.

Examination of Gene Regulation Assumptions.

Stoichiometric analysis using RegEFMTool and EFMTool provided redundant checks, and information difficult to obtain by any individual method. RegEFMTool allowed for the incorporation of gene regulatory descriptions, which reduced the computational burden while providing the entire set of physiologically relevant EFMs. Comparison of results incorporating regulation with subsets of unregulated EFMs assessed the appropriateness of the rules applied. Examples of deviations found in this study were valine catabolism through the modeled biosynthetic route for isoleucine. This pathway was expected to down regulate during valine catabolism and therefore not function; however, its use decreased the biomass and cellular energy production costs for valine (Supplemental Material). Arginine catabolism also showed a similar result, where a bifurcated urea cycle resulted in additional substrate-level phosphorylation not expected due to regulation. These findings suggest that with advancements in data incorporation, such as transcriptomics, complementary approaches will become increasingly necessary to interpret biological phenomena and produce viable hypotheses.

4-9-4 Effect of Terminal Oxidases and Substrate-level Phosphorylation

Accounting for differences between archaeal metabolism and other forms of life expand the applicability of the presented findings. One variable across metabolism is the amount of energy conserved in substrate-level phosphorylation reactions relative to oxidative phosphorylation (data not shown). For example, some archaea lack phosphorylation reactions in Embden-Meyerhof-Parnas glycolysis, possibly eliminating

substrate-level phosphorylation entirely from a pathway that, in bacteria, commonly results in a net of two phosphorylations per glucose consumed. The cytochrome bd oxidase and cytochrome c oxidase shift the relative importance of substrate-level and oxidative phosphorylation by changing the number of protons conserved as proton motive force per oxygen (i.e. 12 and 16 protons per oxygen, respectively). Proton motive force is the driving force for oxidative phosphorylation, often resulting in 10 fold more phosphorylations per glucose in aerobic organisms. Similar trends across the two oxidases for both cost of biomass and cellular energy under the three limited environments examined, suggest that differences in substrate-level phosphorylation steps may ultimately be negligible relative to the whole of metabolism for aerobic assimilatory and dissimilatory biomass utilization (Supplemental Material). This difference warrants further study, as does anaerobic metabolism.

Interpretation of Physiological Stress from Measured DOC Utilization Rates

Identifying the limiting nutrient in an environment can be instrumental in understanding the physiological processes of inhabiting microbial populations. The theory developed in this study was applied to a data set from a eutrophic estuary (42). Comparing the predicted resource costs (amount of a substrate to produce biomass or ATP) to the measured utilization rates in an environment can be used to interpret physiological stress (e.g. carbon-, oxygen-, or enzyme-limitation). The specific utilization rate for a given form of DOC (q_s) is the product of the specific growth rate (μ) and the

DOC resource cost to produce biomass ($Y_{S/B}$, mol DOC consumed /mol biomass produced) (Eq. 1).

$$q_S = \mu Y_{S/B} \quad (\text{Eq. 1})$$

Additionally, the volumetric utilization rate for a given form of DOC (r_S) is the product of the specific utilization rate (q_S) and the concentration of biomass (X_B) (Eq. 2).

$$r_S = q_S X_B = \mu Y_{S/B} X_B \quad (\text{Eq. 2})$$

Therefore, the DOC resource costs ($Y_{S/B}$) that best fit experimental utilization rates (i.e. have the highest R^2) could provide physiological context for the environment. The experimental amino acid utilization rates reported for a eutrophic estuary were compared to the amino acid resource costs ($Y_{S/B}$) for carbon-, oxygen-, or enzyme-limitation. There was a remarkably strong correlation with the oxygen-limited data, suggesting that the experimental system was oxygen limited (Figure 4-S3B). Indeed, oxygen limitation is common in eutrophic environments (e.g. the dead zone in the Gulf of Mexico and Lake Erie). This analysis should be applicable to many environmental systems.

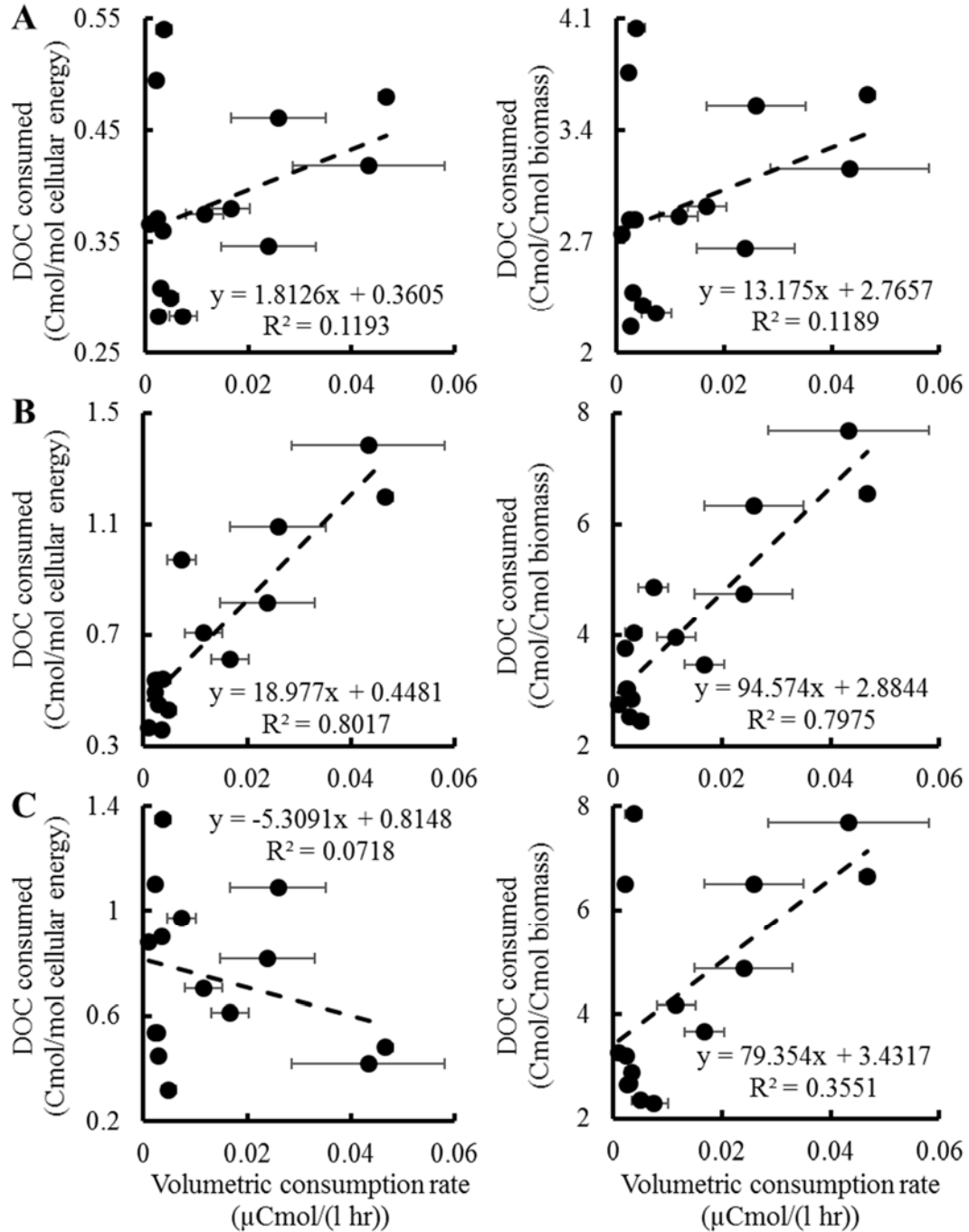


Figure 4-S3. Correlation of theoretical amino acid carbon resource cost and experimental volumetric amino acid consumption rate. Carbon resource costs to produce cellular energy (Cmol amino acid / mol cellular energy) (left) and biomass (Cmol amino acid / Cmol biomass) (right) under carbon- (A), oxygen- (B), or enzyme-limited conditions (C) are shown for all amino acids except asparagine, cysteine, glutamine, histidine, and tryptophan, which were not measured. Volumetric consumption rates are the average consumption rates calculated from (42). Error bars are standard error of measurements for up to 12 measurements.

DOC Solubility

Table 4-S1. Solubilities of common forms of dissolved organic carbon at 25 °C.

Carbon Source	Solubility (g l ⁻¹)
Cellobiose	100 ^a
Cellopentaose	6 ^a
Caldarchaeol	0.000005 ^b
C18 diacylglyceride	0.00005 ^b
Geranylarnesyl pyrophosphate	0.001 ^b
Palmitic acid	0.001 ^b
RNA	NA ^c
AMP	10 ^d
CMP	16 ^b
GMP	4 ^b
UMP	12 ^b
Protein	40 (bovine serum albumin) ^e
Ala	17 ^f
Arg	15 ^f
Asn	3.5 ^f
Asp	0.78 ^f
Cys	280 ^g
Glu	0.86 ^f
Gln	2.5 ^f
Gly	25 ^f
His	4.2 ^f
Ile	4.1 ^f
Leu	2.4 ^f
Lys	1000 ^h
Met	3.4 ^f
Phe	3.0 ^f
Pro	160 ^f
Ser	5.0 ^f
Thr	9.8 ⁱ
Trp	1.1 ^f
Tyr	0.045 ^f
Val	8.9 ^f

^a: (246) ^b: (254) ^c: RNA is not expected to be stable in natural environments due to free RNase activity. ^d: (13)

^e: http://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/a4919pis.pdf

^f: https://www.anaspec.com/html/amino_acids_properties.html

^g: http://www.chemicalbook.com/ChemicalProductProperty_EN_CB7388480.htm

^h: https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Product_Information_Sheet/2/15501pis.pdf ⁱ: (171)

Additional Supplementary Files

Supplemental materials are available at Environmental Microbiology.

APPENDIX D

SUPPLEMENTAL MATERIAL FOR CHAPTER 5

EFMA of Individual Population Models

The RegEFMTool software package employs user-defined gene regulatory rules governing the co-occurrence of reactions within an EFM to simplify the model, thus decreasing the computational burden. The incorporation of gene regulatory rules removed futile cycles that both synthesized and degraded potential carbon or energy sources (*Geoarchaeum*) or used multiple electron donors (*M. yellowstonensis*). Moreover, each carbon source was simulated independent of other carbon sources for the *Geoarchaeum* model, while the *M. yellowstonensis* model included a single carbon source (carbon dioxide) and EFMs were enumerated in full through a single simulation.

Heterotroph Turnover and Electron Acceptor Recycling

Little is known regarding viruses specific to *Geoarchaeum*. It is highly likely that *M. yellowstonensis* is subject to significant viral predation as well. Cyclical catabolism of heterotrophic biomass by heterotrophic organisms is highly probable but for every new Cmole of heterotroph, at least 2.4 Cmoles of old heterotroph would be consumed. In anoxic environments, a Cmole of anaerobe biomass is expected to require at least 10 Cmoles of carbon source. This indicates the recycling of heterotroph biomass would decrease by 2 to 10 fold with each cycle. Additionally, assuming all autotroph biomass is consumed by anaerobic heterotrophs, for every mole of Fe(II) or sulfide oxidized by the autotroph at most 0.03 or 0.21 moles, respectively, would be reduced by the anaerobe. However, the carbon consumed by anaerobes is expected to be closer to 10 % of the total produced, further reducing the impact of anoxic electron cycling in these mats.

Approximation of Population Independent
Oxidation of Reduced Sulfur Species

Populations of *M. yellowstonensis* have not been found in sulfidic systems of acid-sulfate-chloride geothermal springs of NGB, despite the fact that these Sulfolobales contain the genetic potential for the oxidation of elemental sulfur and have been shown to do so in culture in the absence of significant sulfide (130). Fe(II) oxidation terminal oxidase genes (i.e. *fox* complex) are more highly expressed under field conditions than sulfur or organic carbon-associated terminal oxidases (127), which is consistent with the realized niche for *M. yellowstonensis* in zones with higher oxygen concentrations (e.g., DO > 20 μ M). Elemental sulfur generally co-occurs with high sulfide concentrations (e.g., DS > 10 μ M) in acid-sulfate and/or acid-sulfate-chloride geothermal springs of YNP, which precludes the necessary oxygen flux to support the generation of *M. yellowstonensis* biomass. Consequently, EFMA and FBA have assisted in understanding why *M. yellowstonensis* occupies an Fe(II) oxidizing niche *in situ*, and has provided possible explanations for autotroph-heterotroph interactions occurring in high-temperature Fe-mat communities.

Resource Requirements for
Autotroph and Heterotroph Growth

Table 5-S1. Electron donors and acceptor required to produce biomass for the most efficient metabolisms identified through elementary flux mode analysis for the autotroph, *M. yellowstonensis*.

Electron Donor	Electron Donor per Biomass ^a	Oxygen per Biomass ^b	Degree of Reduction
Fe(II)	156	38.0	1 ^c
Sulfide	2.6	4.2	8
Elemental sulfur	3.0	3.5	6
Sulfite	4.2	1.0	2
Thiosulfate	42.3	9.5	^d

^a: Moles of electron donor consumed per Cmole of biomass produced

^b: Mole of oxygen consumed per Cmole of biomass produced

^c: Assuming Fe(III) has a degree of reduction of 0

^d: One electron is released per thiosulfate oxidized

Table 5-S2. Carbon sources and electron acceptor required to produce biomass for the most efficient metabolisms identified through elementary flux mode analysis for the heterotroph, *Geoarchaeum*.

Carbon Source	Carbon-limitation		Oxygen-limitation		Degree of Reduction ^c
	Carbon per Biomass ^a	Oxygen per Biomass ^b	Carbon per Biomass ^a	Oxygen per Biomass ^b	
Aggregate Biomass	2.4	1.6	3.8	1.5	4.3
Lipid	2.1	1.9	2.9	1.9	5.9
Peptide	2.6	1.7	3.9	1.6	5.0
Polysaccharide	2.6	1.5	4.6	1.2	4.0
RNA	3.1	1.4	5.4	1.1	4.1

^a: Cmoles of carbon substrate consumed per Cmole of biomass produced

^b: Mole of oxygen consumed per Cmole of biomass produced

^c: Degree of reductions were calculated on NH₃ bases

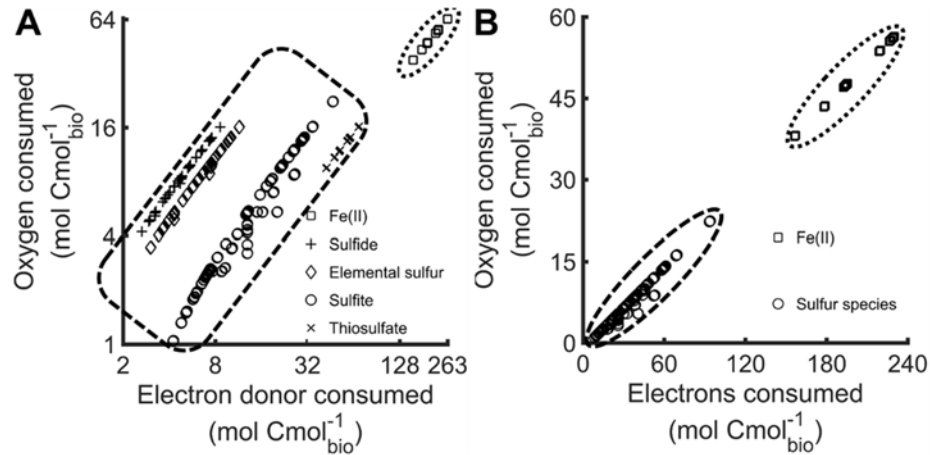


Figure 5-S1. Electron donor required to produce autotroph biomass. The oxygen consumption for all biomass producing elementary flux modes of *M. yellowstonensis* plotted as a function of mole electron donor (A) or mole electrons (B) consumed to produce a Cmole of biomass. Specific electron donors include Fe(II), sulfide, elemental sulfur, sulfite, and thiosulfate.

Determination of Relative Population Abundance of Autotroph to Heterotroph

The relative abundance of autotroph to heterotroph was approximated by previously published sequence analyses (Table 5-S3) (9). Various representations of autotroph, including *M. yellowstonensis* and *Hydrogenobaculum*, and heterotroph, including *Geoarchaeum* str. OSPB and Novel Archaeal Group 2, were examined to determine a representative distribution of governing populations (Table 5-S4). A relative population abundance of autotroph to heterotroph of 0.3 to 0.5 was used to include the range of both examined springs.

Table 5-S3. Distribution of populations in One Hundred Springs Plain (OSP) and Beowulf hot springs from 454 and Illumina sequence analyses.

	OSP			Beowulf		
	454	Illumina	Averaged Percent	454	Illumina	Averaged Percent
Autotrophs						
<i>M. yellowstonensis</i>	10.3	16	21%	6.6	12	14%
<i>Hydrogenobaculum</i>	3.4	3	5%	1.2	4.3	4%
Heterotrophs						
<i>Geoarchaeum</i> str. OSPB	43.4	33	60%	3	6	7%
Novel Archaeal Group 2	4	6	8%	38	27	48%
Others						
<i>Thaumarchaeota</i>	0.4	0.4	1%	4.4	2	5%
Novel Archaeal Group 3	0.1	0.12	0%	3	5	6%
Other <i>Sulfolobales</i>	2.5	4	5%	10	11	16%
<i>Nanoarchaeota</i>	0.2	0.6	1%	0.1	0.6	1%
Total	64.3	63.12	100%	66.3	67.9	100%

Modified from (9).

Table 5-S4. Relative population abundance of autotroph to heterotroph considering different representative populations based on Table 5-S3.

Calculation of Relative Population Abundance	OSP		Beowulf	
	454	Illumina	454	Illumina
MK1/Geo	0.24	0.48	2.20	2.00
MK1/(Geo+NAG2)	0.22	0.41	0.16	0.36
(MK1+Hydro)/(Geo + NAG2)	0.29	0.49	0.19	0.49

Abbreviations are as follows: MK1 - *M. yellowstonensis*; Hydro - *Hydrogenobaculum*; Geo - *Geoarchaeum* str. OSPB; NAG2 - Novel Archaeal Group 2;

Determination of Total Biomass Concentration from *in situ* Fe Measurements

The total biomass concentrations of OSP and Beowulf springs were approximated from the Fe deposition rate using the elemental composition of Fe(III)-oxide mats. Fe-oxide deposition rates of 8.73 and 1.28 $\mu\text{mol}_{\text{Fe}} \text{cm}^{-2} \text{d}^{-1}$ at OSP and Beowulf springs were determined from previous characterization studies (9) assuming a constant deposition rate in a mature Fe(III)-oxide mat (i.e. older than 40 days) (Figure 5-S2) (9). Total biomass

concentrations were extrapolated using the elemental composition of the mats and density as derived from the reported vertical growth rate of the Fe(III)-oxide mat (Table 5-S5).

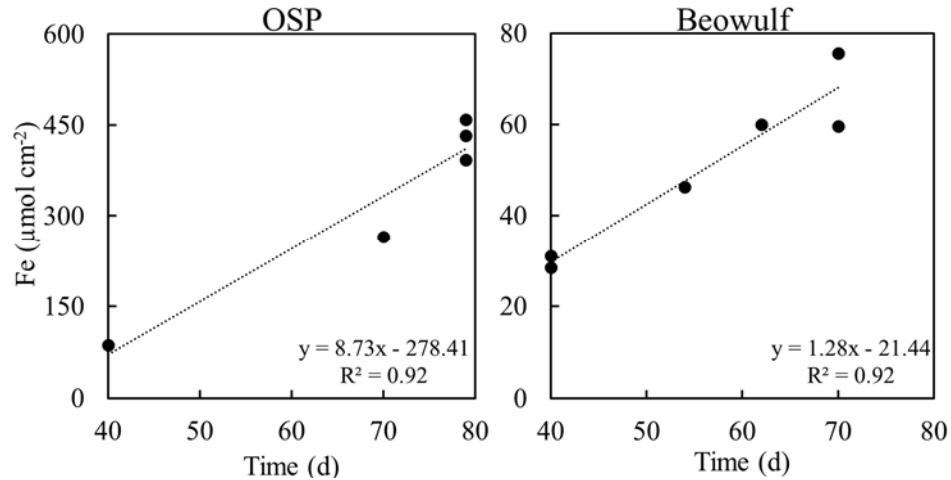


Figure 5-S2. Fe-oxide deposition rate for One Hundred Springs Plain (OSP) and Beowulf hot springs. A linear approximation was used to determine the deposition rate assuming that a developed mat would have constant microbial activity in the oxic zone. Data from (9).

Table 5-S5. Biomass concentration calculations based on Fe-oxide deposition rate.

Description	OSP	Beowulf	Units
Fe deposition rate	8.73	1.28	$\mu\text{mol}_{\text{Fe}} \text{cm}^{-2} \text{d}^{-1}$
Fe concentration across mat growth rates ^a	0.24	0.036	$\text{g}_{\text{Fe}} \text{cm}^{-3}$
Mat carbon per mat iron ^b	0.046	0.046	$\text{g}_{\text{CTot}} \text{g}_{\text{Fe}}^{-1}$
Biomass carbon per total mat carbon ^c	0.1	0.1	$\text{g}_{\text{CBio}} \text{g}_{\text{CTot}}^{-1}$
Carbon per gram of biomass from the model ^d	0.043	0.043	$\text{Cmol g}_{\text{Bio}}^{-1}$
Biomass concentration	2.15	0.32	$\text{mg}_{\text{Bio}} \text{cm}^{-3}$

^a: An average vertical growth rates of $22.5 \mu\text{m cm}^{-2} \text{d}^{-1}$ was based on (9). ^b: (102). ^c: (107). ^d: (96).

Graphical Representations of Metabolic
Space Simulated Using Hybrid EFM-FBA

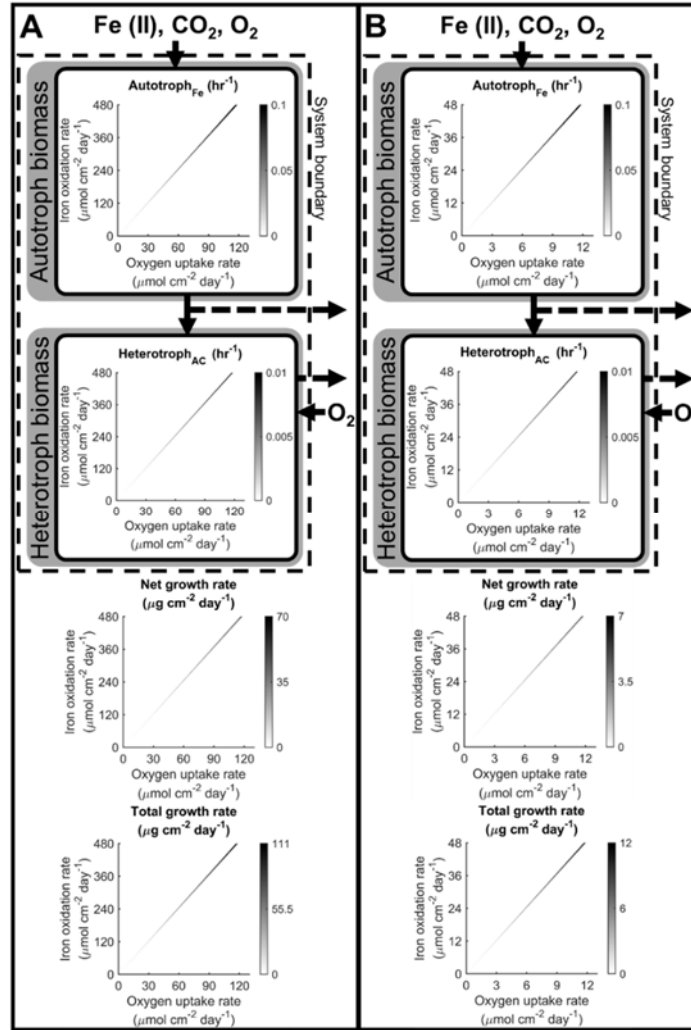


Figure 5-S3. Metabolic space feasible for a community composed of an iron oxidizing autotroph and autotroph consuming heterotroph. Active biomass concentrations were 0.2 (A) and 2.0 (B) mg cm⁻³ and an aerobic volume of 0.07 cm³ cm⁻² allowed for the prediction of specific growth rates under the examined conditions bounding the possible physiological space assuming a maximum growth rate of 0.1 hr⁻¹. The observed relative population abundances were 0.3 to 0.5 based on sequence analyses.

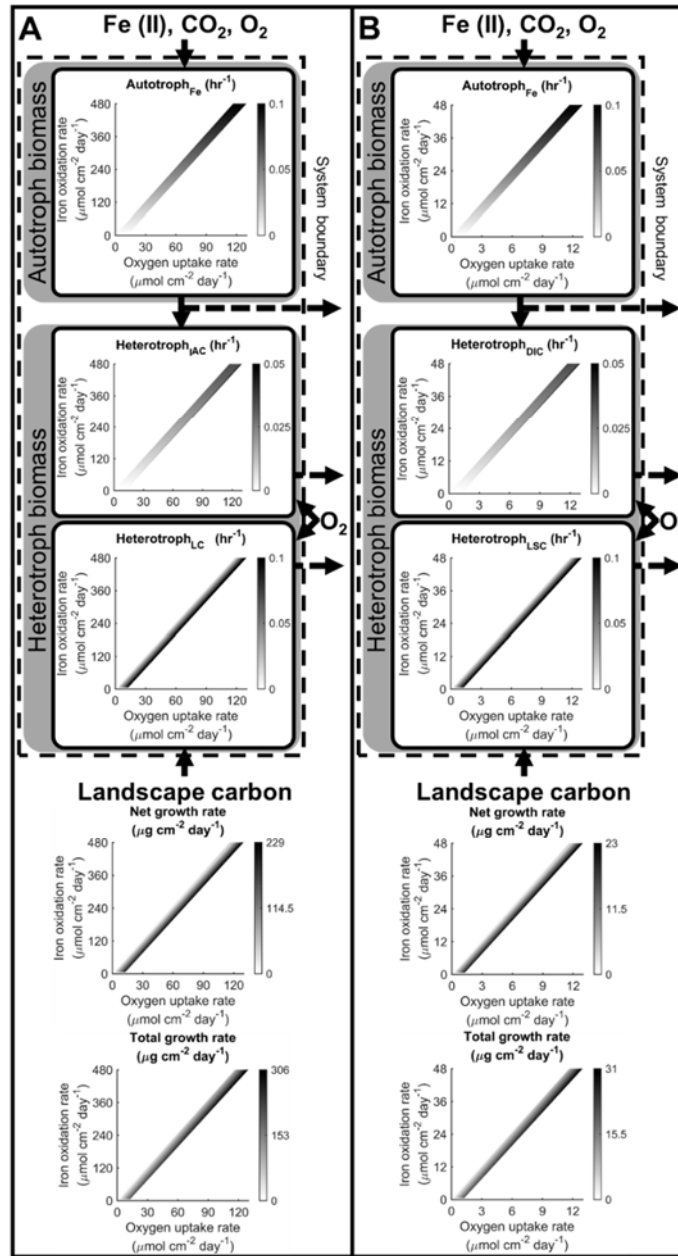


Figure 5-S4. Metabolic space feasible for a community composed of an iron oxidizing autotroph, an autotroph consuming heterotroph and a landscape carbon consuming heterotroph. Active biomass concentrations were 0.2 (A) and 2.0 (B) mg cm^{-3} and an aerobic volume of $0.07 \text{ cm}^3 \text{ cm}^{-2}$ allowed for the prediction of specific growth rates under the examined conditions bounding the possible physiological space assuming a maximum growth rate of 0.1 hr^{-1} . The observed relative population abundances were 0.3 to 0.5 based on sequence analyses. A maximum of 33 and 58 % of mat biomass carbon was allowed to be of landscape carbon origins for Beowulf and OSP, respectively.

APPENDIX E

SUPPLEMENTAL MATERIAL FOR CHAPTER 6

Impact of Genetic Modifications by
Chain Length and Functional Group

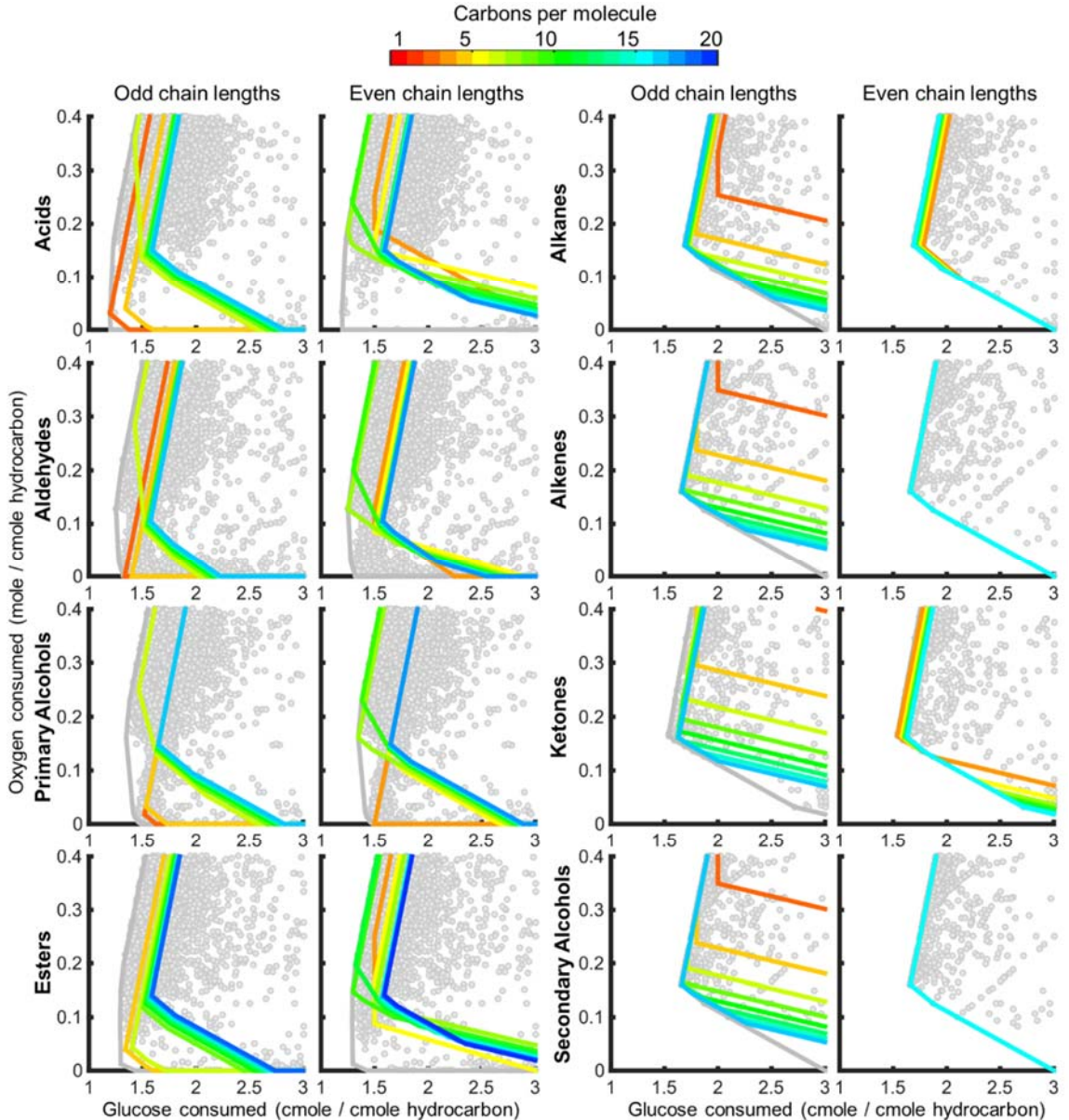


Figure 6-S1. Hydrocarbon production strategies that minimize glucose and oxygen consumption. Optimal oxygen and carbon source utilization strategies that produce hydrocarbons were identified that minimize resource costs (grey boundary, all possible metabolic strategies are grey points). Metabolic strategies that produce hydrocarbons with a given number of carbons per molecule are denoted by the colored lines. Formate, acetate, acetaldehyde, and ethanol were only counted as hydrocarbons for their respective functional groups. Odd and even chained hydrocarbons were initiated by acetyl-CoA and propanoyl-CoA as appropriate.

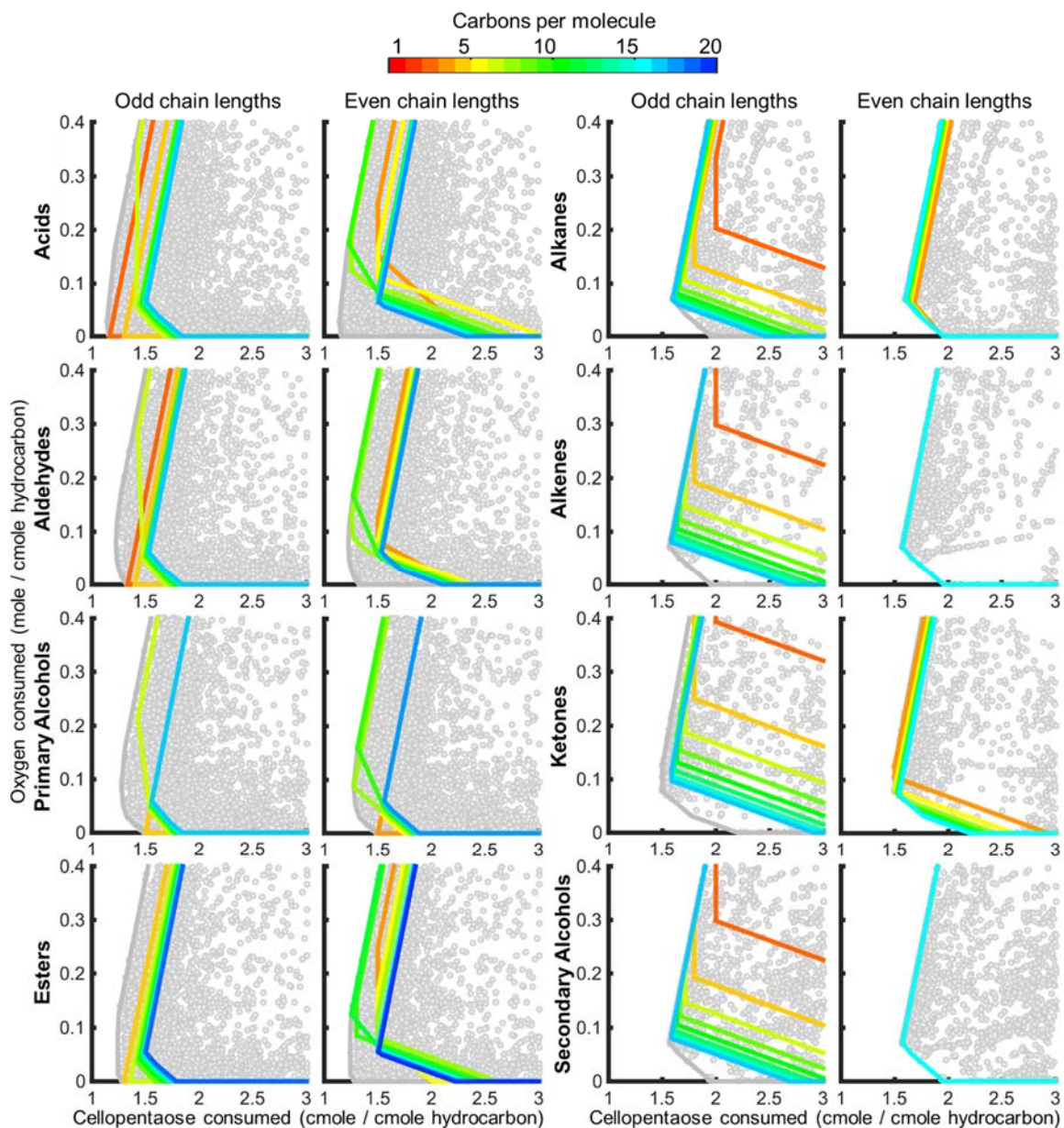


Figure 6-S2. Hydrocarbon production strategies that minimize cellopentaose and oxygen consumption when cellobiose phosphorylase is present. Optimal oxygen and carbon source utilization strategies that produce hydrocarbons were identified that minimize resource costs (grey boundary, all possible metabolic strategies are grey points). Metabolic strategies that produce hydrocarbons with a given number of carbons per molecule are denoted by the colored lines. Formate, acetate, acetaldehyde, and ethanol were only counted as hydrocarbons for their respective functional groups. Odd and even chained hydrocarbons were initiated by acetyl-CoA and propanoyl-CoA as appropriate.

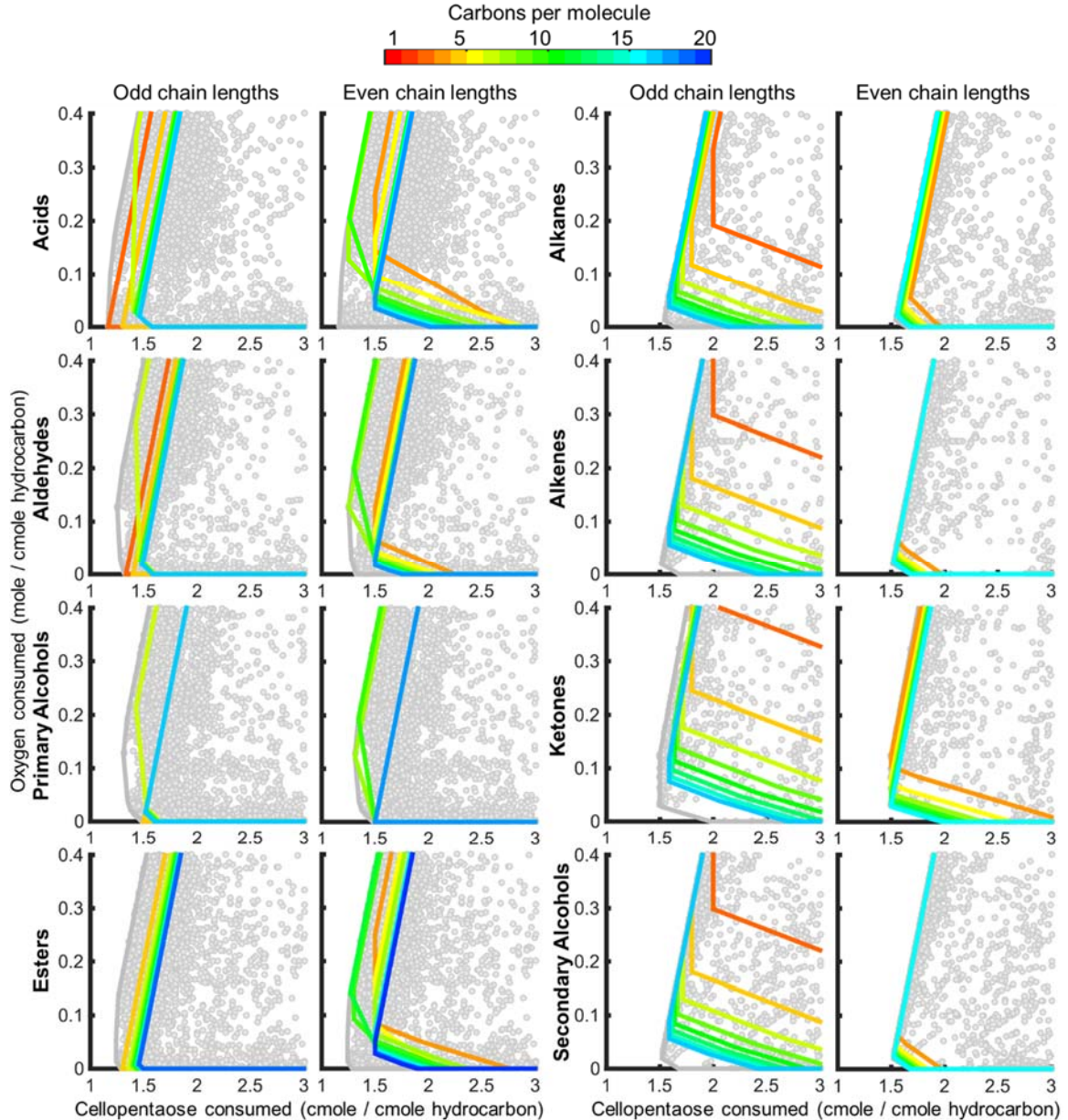


Figure 6-S3. Hydrocarbon production strategies that minimize cellopentaose and oxygen consumption when a cytosolic pyruvate dehydrogenase is present. Optimal oxygen and carbon source utilization strategies that produce hydrocarbons were identified that minimize resource costs (grey boundary, all possible metabolic strategies are grey points). Metabolic strategies that produce hydrocarbons with a given number of carbons per molecule are denoted by the colored lines. Formate, acetate, acetaldehyde, and ethanol were only counted as hydrocarbons for their respective functional groups. Odd and even chained hydrocarbons were initiated by acetyl-CoA and propanoyl-CoA as appropriate.

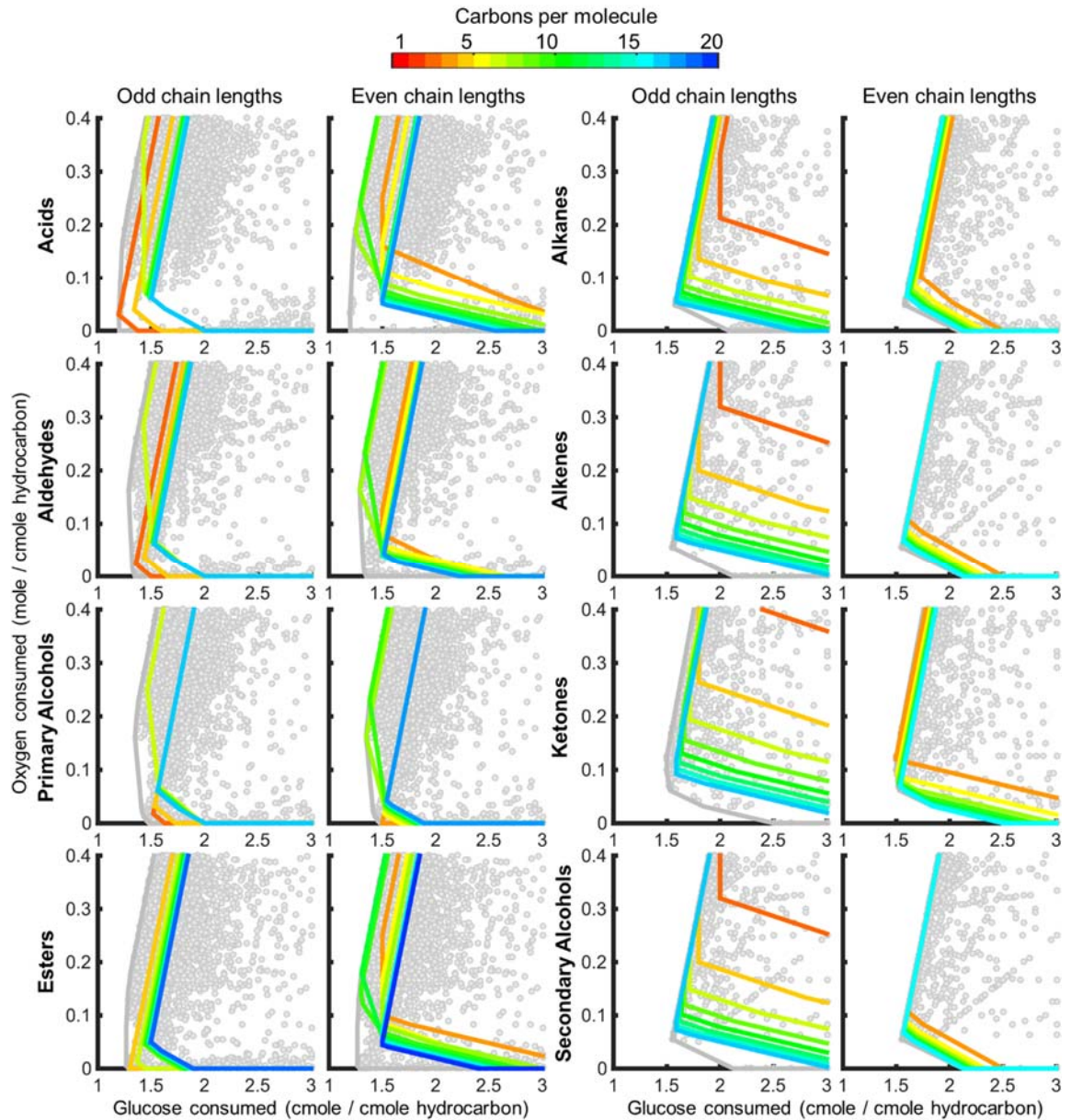


Figure 6-S4. Hydrocarbon production strategies that minimize glucose and oxygen consumption when cytosolic pyruvate dehydrogenase is present. Optimal oxygen and carbon source utilization strategies that produce hydrocarbons were identified that minimize resource costs (grey boundary, all possible metabolic strategies are grey points). Metabolic strategies that produce hydrocarbons with a given number of carbons per molecule are denoted by the colored lines. Formate, acetate, acetaldehyde, and ethanol were only counted as hydrocarbons for their respective functional groups. Odd and even chained hydrocarbons were initiated by acetyl-CoA and propanoyl-CoA as appropriate.

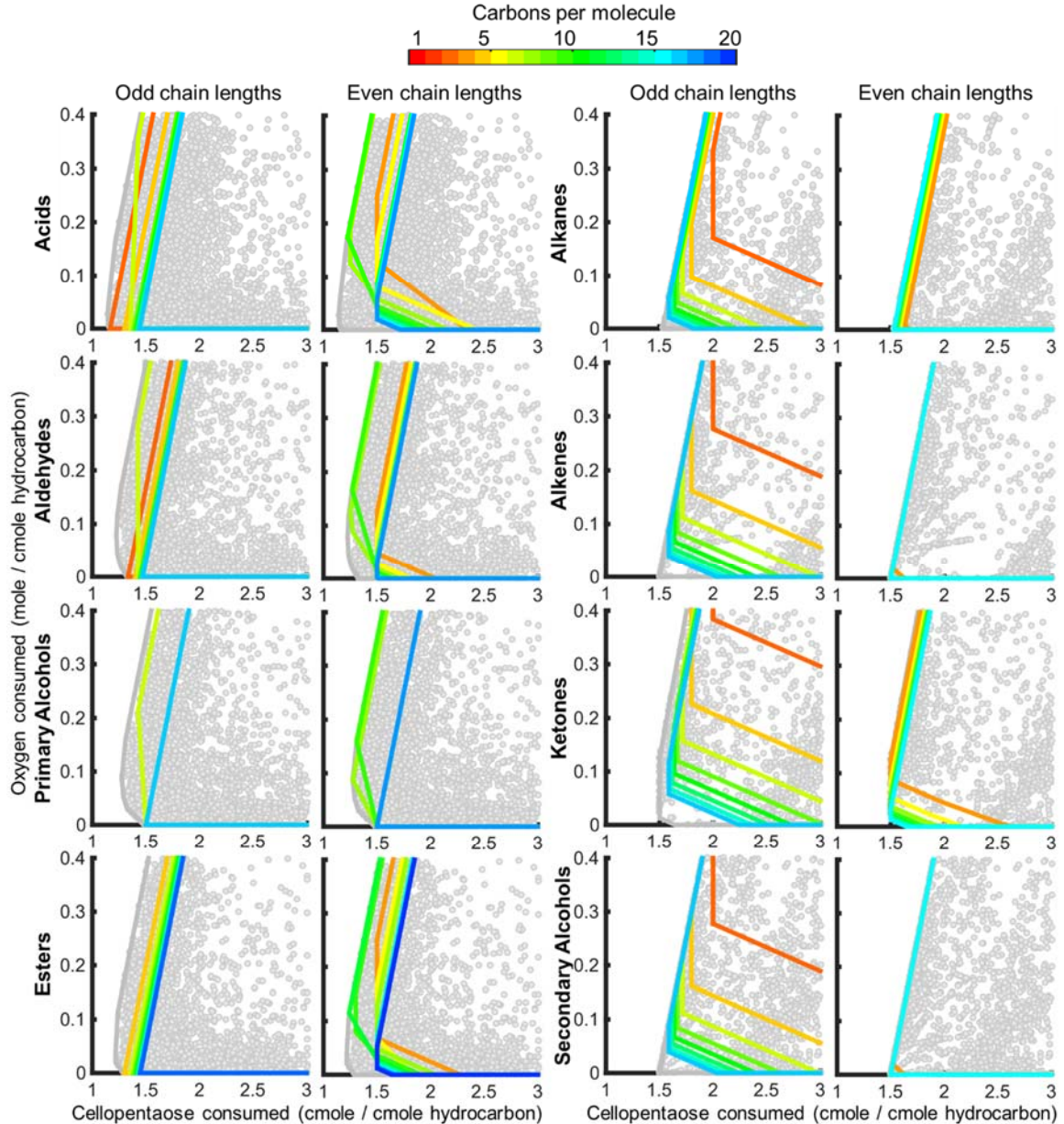


Figure 6-S5. Hydrocarbon production strategies that minimize cellopentaose and oxygen consumption when both cellobiose phosphylase and cytosolic pyruvate dehydrogenase are present. Optimal oxygen and carbon source utilization strategies that produce hydrocarbons were identified that minimize resource costs (grey boundary, all possible metabolic strategies are grey points). Metabolic strategies that produce hydrocarbons with a given number of carbons per molecule are denoted by the colored lines. Formate, acetate, acetaldehyde, and ethanol were only counted as hydrocarbons for their respective functional groups. Odd and even chained hydrocarbons were initiated by acetyl-CoA and propanoyl-CoA as appropriate.

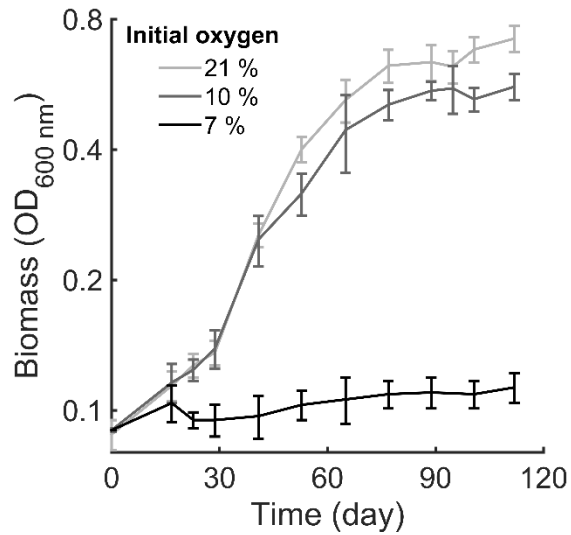


Figure 6-S6. Biomass production under varying oxygen limitation. Three initial headspace oxygen concentrations were examined, 7, 10, and 21%, during biomass production from carboxymethyl cellulose. Values shown are the average and standard error of measurements of biological triplicates.

Additional Supplementary Files

Supplemental materials are available through the MSU library