A NEW METHOD OF GENOME-SCALE METABOLIC MODEL VALIDATION FOR BIOGEOCHEMICAL APPLICATION

by

BENJAMIN MORRIS SHAPIRO

A THESIS

Presented to the Department of Earth Sciences
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Master of Science

June 2017

THESIS APPROVAL PAGE

Student: Benjamin Morris Shapiro

Title: A New Method of Genome-Scale Metabolic Model Validation for Biogeochemical Application

This thesis has been accepted and approved in partial fulfillment of the requirements for the Master of Science degree in the Department of Earth Sciences by:

Jim Watkins	Chairperson
Qusheng Jin	Advisor
Edward Davis	Member
Scott Bridgham	Member

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2017

©2017 Benjamin Morris Shapiro

THESIS ABSTRACT

Benjamin Morris Shapiro

Master of Science

Department of Earth Sciences

June 2017

Title: A New Method of Genome-Scale Metabolic Model Validation for Biogeochemical Application

We propose a new method to integrate genome-scale metabolic models into biogeochemical reaction modeling. This method predicts rates of microbial metabolisms by combining flux balance analysis (FBA) with microbial rate laws. We applied this new hybrid method to methanogenesis by *Methanosarcina barkeri*.

Our results show that the new method predicts well the progress of acetoclastic, methanol, and diauxic metabolism by M. barkeri. The hybrid method represents an improvement over dynamic FBA. We validated genome-scale metabolic models of *Methanosarcina barkeri*, *Methanosarcina acetivorans*, *Geobacter metallireducens*, *Shewanella oneidensis*, *Shewanella putrefaciens* and *Shewanella sp.* MR4 for application to biogeochemical modeling. FBA was used to predict the response of cell metabolism, and ATP and biomass yield. Our analysis provides improvements to these models for the purpose of applications to natural environments.

CURRICULUM VITAE

NAME OF AUTHOR: Benjamin Morris Shapiro

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon
University of Puget Sound, Tacoma, Washington

DEGREES AWARDED:

Master of Science, Geological Sciences, 2017, University of Oregon
Bachelor of Science, Geology, 2010, University of Puget Sound

AREAS OF SPECIAL INTEREST:

Geobiology
Geochemistry
Hydrogeology
Biochemistry

PROFESSIONAL EXPERIENCE:

Graduate Employee, Department of Earth Sciences, University of Oregon, Eugene, 2016-2017

Graduate Teaching Fellow, Department of Earth Sciences, University of Oregon, Eugene, 2013-2016

Research Assistant, Department of Earth Sciences, University of Oregon, Eugene, 2011-2012

Water Resources Intern, U.S. Geological Survey, Menlo Park, 2009-2010

Teaching Assistant, Department of Geology, University of Puget Sound, Tacoma, 2008-2009

GIS Lab Intern, U.S. Geological Survey, Menlo Park, 2006-2007

GRANTS, AWARDS, AND HONORS:

Outstanding GTF Award, University of Oregon, 2014

Condon Scholarship, University of Oregon, 2013

Johnston Scholarship, University of Oregon, 2012

Thayer Scholarship, University of Oregon, 2012

Cordilleran Section Student Presentation Award, Geological Society of America, 2009

Al Eggers Research Award, University of Puget Sound, 2009

ACKNOWLEDGMENTS

I would like to thank my advisor, Qusheng Jin, for moving in a new research direction with me; for your academic, emotional, and financial support. I thank my peers and teachers in both the Department of Earth Sciences at the University of Oregon and the Department of Geology at the University of Puget Sound for years of instruction and input. I express a special thanks to the office staff within the Department of Earth Sciences: Sandy Thoms, Marla Trox, and Dave Stemple. I thank my thesis and guidance committee members, Edward Davis, Jim Watkins, and Scott Bridgham, for assisting me on my academic journey. To my friends and family that have guided me and stood by my side, I thank you deeply.

in memory of Rod Park; he was a great scientist, role model, and my grandfather.

TABLE OF CONTENTS

Chapter Page

I. A NEW METHOD OF GENOME-INFORMED BIOGEOCHEMCIAL REACTION MODELING 1

Introduction 1

Methods 4

Biogeochemical Reaction Modeling 4

Flux Balance Analysis 5

Metabolic Control Analysis 6

Hybrid Method 7

dFBA 7

Application 8

Results and Discussion 10

Constraints of Exchange Fluxes 10

Metabolic Control Analysis 11

De Trop Reaction 12

Acetoclastic Methanogenesis 13

Methanol Methanogenesis 16

Diauxic Growth 19

Discussion 21

II. GENOME-SCALE METABOLIC VALIATION FOR THE NATURAL ENVIRONEMNT 28

Introduction 28

Methods 29

Chapter Page

Catabolic and Biosynthesis Networks 29

Flux Balance Analysis 30

Validation and Revision 31

Application 32

Results 32

Dependence of Biosynthesis on Nutrient Uptake 32

ATP Yield 33

Growth Yield per ATP 34

Revision 34

Methanosarcina barkeri 35

Methanosarcina acetivorans 35

Geobacter metallireducens 36

Shewanella 37

Assembly ATP Cost 39

Futile Loops 44

Prediction 45

Sweeping Results 45

ATP Yield 46

Growth Yield per ATP 46

Biomass Composition 47

Discussion 48

REFERENCES CITED 78

LIST OF FIGURES

Figure Page

1. Variations in exchange fluxes of acetate and CO_2 with rates of methanogenesis by *M. barkeri*.......... 61

2. Variations in the flux control coefficients by *M. barkeri* 61

3. Results of hybrid method and dFBA simulations of acetoclastic methanogenesis by *M. barkeri*.......... 62

4. Results of hybrid method and dFBA simulations of methanol methanogenesis by *M. barkeri*.......... 63

5. Results of hybrid method simulation of diauxic growth on a mixture of acetate and methanol by *M. barkeri*.......... 64

6. Predicted biosynthesis rates from a range of substrate consumption rates (H_2/CO_2, acetate, lactate, methanol) using published models 65

7. Variations in the flux control coefficients of biosynthesis rate with substrate consumption rates (H_2/CO_2, acetate, lactate, methanol).......... 66

8. Predicted ATP formation rates from consumption of electron donors (H_2, acetate, lactate, methanol) from the catabolic network.......... 67

9. Variations in the flux control coefficients of ATP formation flux with electron donor uptake flux (H_2, acetate, lactate, methanol).......... 68

10. Predicted ATP formation rates from carbon source consumption rates in the biosynthesis network on different carbon sources (CO_2, acetate, lactate, methanol).......... 69

11. Variations in the flux control coefficients of biosynthesis with ATP flux on different carbon sources (CO_2, acetate, lactate, methanol).......... 70

12. Amino acid abundance in biomass objective function in various models including E. coli. 71

13. Predicted biosynthesis rates from a range of substrate consumption rates (H_2/CO_2, acetate, lactate, methanol) with modifications to the combined catabolic and biosynthesis network (genome-scale network).......... 72

Figure Page

14. Variations in the flux control coefficients of biosynthesis rate with substrate consumption rates (H_2/CO_2, acetate, lactate, methanol) with corrections to the genome-scale network.. 73

15. Predicted ATP formation rates from a range of electron donor consumption rates (H_2, acetate, lactate, methanol) from the modified catabolic network.......... 74

16. Variations in the flux control coefficients of ATP flux with consumption of electron donors (H_2, acetate, lactate, methanol) in the corrected catabolic network.. 75

17. Predicted biosynthesis rates from a range of ATP flux rates from the corrected biosynthesis network on different carbon sources (CO_2, acetate, lactate, methanol).. 76

18. Variations in the flux control coefficients of biosynthesis with ATP flux on different carbon sources (CO_2, acetate, lactate, methanol) with the corrected biosynthesis network.. 77

LIST OF TABLES

Table Page

1. Kinetic parameters of methanogenesis and enzymes of *M. barkeri* 53
2. Genome-scale metabolic models, numbers of metabolites and reactions, number of reactions in catabolic and biosynthesis models, and electron donors and accepters tested using FBA. 54
3. Carbon sources for biomass yield predictions using FBA.................................... 54
4. ATP yield of original catabolic network models and observations 55
5. Growth yield of original biosynthesis network models 56
6. Survey of GAM estimated values use in metabolic models 56
7. ATP required for monomer biosynthesis and the cost of biomass assembly 57
8. Comparison of ATP required for synthesis of amino acids in iMB745, iMG746, and iAF1260. 58
9. Maximum flux of loop reactions constrained by flux direction 59
10. ATP yield of validated catabolic network models and observations.................... 59
11. Growth yield (with 50% efficiency) of validated biosynthesis network models with BAC estimations from this study.. 59
12. Growth yield of validated biosynthesis network models...................................... 60
13. Elemental biomass composition from biosynthesis network models 60

CHAPTER I

A NEW METHOD OF GENOME-INFORMED BIOGEOCHEMCIAL REACTION MODELING

1. Introduction

Biogeochemical reaction modeling has become an essential tool for investigating the microbial processes of natural environments (Bethke; Jin and Roden). This method builds on both geochemical and microbial reaction models. Geochemical models describe abiotic chemical reactions, such as chemical speciation, redox reaction, and mineral precipitation and dissolution. Microbial models capture metabolic reactions of microorganisms. By coupling the simulations of abiotic and biological processes, biogeochemical modeling offers a quantitative assessment of microbial processes in the context of concurrent geochemical reactions. It has been applied to both theoretical and practical problems, from biomineralization, to element cycling, and to the remediation of contaminants.

Current approaches of biogeochemical modeling describe microbial metabolisms using rate laws, such as the Monod equation and its modifications. These rate laws are developed for microbes in laboratory and industrial reactors (Monod; Simkins and Alexander), and thus may not be effective in simulating microbial metabolisms in nature. For example, most laboratory experiments offer optimal growth conditions, and microbial metabolism proceeds at maximum rates. However, in the natural environment, where nutrients are often limited and of fluctuating concentrations, microbes must regulate biochemical pathways and fine-tune metabolic rates to effectively utilize available nutrients and gain competitive advantage. As a result, natural microbes may behave differently from the predictions of the rate laws developed from laboratory experiments.

Recently, a genome-based method, dynamic flux balance analysis (dFBA), has been applied as an alternative to microbial rate laws in predicting microbial rates. This method takes advantage of genome-scale metabolic network models, the stoichiometric reaction equations of enzymes within entire organisms (Orth, Thiele, and Palsson). It applies linear optimization to estimate metabolic rates and other phenotypes from the rates of microbial nutrient uptake. FBA brings genome-scale biochemical pathways into the simulation of microbial metabolism, which make it possible to incorporate biochemical pathways and metabolic regulation into biogeochemical modeling.

The method of dFBA has been applied to *Geobacter* and *Methanosarcina* species, as well as to other prokaryotes of biogeochemical significance. However, its application has not always been successful. Previous studies have shown that dFBA may overestimate the rates of microbial metabolism, sometimes by a factor of 10. This overestimation has been attributed to the under-constrained nature of genome-scale metabolic models (Scheibe et al.; Tartakovsky et al.).

In addition, the overestimation may result from the way dFBA computes the uptake fluxes of nutrients. Specifically, dFBA computes the uptake fluxes by applying the Michaelis-Menten equation to enzymes that consumes nutrients. This equation assumes that enzyme reactions depend only on the concentrations of substrates, but enzyme activities may also be limited by thermodynamic drives. Thus, the Michaelis-Menten equation may overestimate the uptake fluxes of nutrients.

The overestimation of dFBA may also arise from *de trop* reactions in genome-scale metabolic models. In constructing these models, hypothetical reactions, without gene association, are often added, to account for documented phenotypes (Thiele and Palsson). But non-gene-associated reactions may become *de trop* – that is, in addition to producing the desired phenotype, they may interact with other gene-associated enzyme reactions to carry out erroneous functions that are beyond the capability of the microbes of interest.

Here we propose a new method for applying genome-scale metabolic models to biogeochemical modeling. We first separate genome-scale metabolic models into catabolic and biosynthesis networks. These network models are validated to ensure their applicability to the environment using metabolic control analysis. We predict the ATP fluxes of the catabolic network by combining the modified Monod equation and FBA. We then predict microbial growth rates by applying the ATP fluxes and FBA to the biosynthesis network.

Specifically, network models are validated using metabolic control analysis and nutrient fluxes typical of natural environments. Metabolic control analysis quantifies the response of individual enzyme reactions of metabolic networks to the variations in nutrient uptake fluxes. The results of the analysis are applied to identify *de trop* reactions and to ensure that model predictions are consistent with the physiology of the microbes of interest.

The new method then constrains genome-scale metabolic networks using the fluxes of chemical compounds produced or consumed by cell catabolism only. Examples of catabolism-specific compounds include electron acceptors A and their reduced forms

A^{-}. According to the principle of chemical kinetics, at steady state, rates of enzyme reactions (including uptake fluxes of A or release fluxes of A^{-}), are stoichiometrically equivalent to the rates of catabolism. Thus we can constrain the fluxes of A or A^{-} using the rates of respiration, and then compute respiration rates according to a modified Monod equation that accounts for both kinetics and thermodynamics of microbial respiration (Jin and Bethke).

We illustrated the new method by applying it to methanogenesis. Methanogenesis is a key biogeochemical process in carbon cycling and global climate change. We simulate the metabolism of *Methanosarcina barkeri*. *M. barkeri* is a representative methanogen that has been extensively studied. Specifically, we are using the genome-scale metabolic network model for *M. barkeri* strain Fusaro (Gonnerman et al.). *M. barkeri* is metabolically diverse, capable of making methane from H_2/CO_2, acetate, methanol, and other methyl-containing compounds. We simulated acetoclastic and methanol methanogenesis by *M. barkeri* in laboratory bioreactors using both the hybrid method and dFBA. The simulation results show that the new method well predicts experimental observations, and represents an improvement over the current method of dFBA.

2. Methods

2.1. Biogeochemical Reaction Modeling

Microbes derive energy for growth and maintenance by catalyzing redox reactions. Biogeochemical reaction models describe microbial metabolisms using rates of respiration, growth, and maintenance. Current methods calculate respiration rate ($\mathrm{mol \cdot g^{-1} \cdot s^{-1}}$) according to the modified Monod equation,

$$r'_{\mathrm{Ck}} = k \cdot \frac{m_{\mathrm{D}}}{m_{\mathrm{D}} + K_{\mathrm{D}}} \cdot \frac{m_{\mathrm{A}}}{m_{\mathrm{A}} + K_{\mathrm{A}}} \cdot \left[1 - \exp\left(\frac{\Delta G + \nu_{\mathrm{P}} \cdot \Delta G_{\mathrm{P}}}{\chi RT} \right) \right] \tag{1}$$

where k is the rate constant ($\mathrm{mol \cdot g^{-1} \cdot s^{-1}}$), m_{D} and m_{A} are molal concentrations of electron donors and acceptors, respectively, ΔG is the Gibbs free energy change of redox reactions, χ is the average stoichiometric number, ν_{P} is the ATP yield, ΔG_{P} is the phosphorylation energy, R is the gas constant, and T is the absolute temperature. Microbial growth rate is computed according to

$$\frac{d[\mathrm{X_k}]}{dt} = \left(r'_{\mathrm{X}} - r'_{\mathrm{M}} \right) \cdot [\mathrm{X_k}] \tag{2}$$

where [X] is the biomass concentration, the dry weight per unit volume ($\mathrm{g \cdot L^{-1}}$), r'_{Xk} is the specific growth rate, and r'_{Mk} is the specific maintenance rate ($\mathrm{s^{-1}}$).

Microbial metabolism consumes chemical species in the environment. The rate at which chemical compound i is consumed is calculated according to

$$\frac{dm_{\mathrm{i}}}{dt} = \sum_{\mathrm{k}} \left(\nu_{\mathrm{i}}^{\mathrm{Ck}} \cdot r'_{\mathrm{Ck}} + \nu_{\mathrm{i}}^{\mathrm{Xk}} \cdot r'_{\mathrm{Xk}} + \nu_{\mathrm{i}}^{\mathrm{Mk}} \cdot r'_{\mathrm{Mk}} \right) \cdot [\mathrm{X_k}] \tag{3}$$

where m_{i} is the molal concentration; $\nu_{\mathrm{i}}^{\mathrm{Ck}}$ and $\nu_{\mathrm{i}}^{\mathrm{Xk}}$ are stoichiometric coefficients of compound *i* in catabolism and biosynthesis, respectively. Biogeochemical reaction models simulate abiotic reactions using the standard protocol of geochemical reaction modeling. For example, chemical speciation is assumed at equilibrium. Redox reaction and mineral precipitation and dissolution can be simulated using either thermodynamic equilibrium or kinetic rate laws.

2.2. Flux Balance Analysis

Flux balance analysis (FBA) is a method for analyzing genome-scale metabolic models. These models describe genome-scale metabolic networks using a stoichiometric matrix S of size $m \times n$. Here, n is the number of enzymes within the entire organism, and m is the number of metabolites consumed or produced by the enzymes. Element S_{ij} is the stoichiometric coefficient of metabolite i in the reaction equation of enzyme j. Applying the principle of mass balance to these models leads to J = S·R. Here J and R are vectors; element Ji is the net rate of metabolite A_i production ($mmol \cdot g^{-1} \cdot s^{-1}$), and element Rj is the flux (or net reaction rate) of enzyme j ($mmol \cdot g^{-1} \cdot s^{-1}$).

The goal of FBA is to estimate enzyme fluxes at steady state. It takes the fluxes of nutrient uptake as input and assumes that the objective of metabolic networks is to maximize rates of microbial growth. It solves for enzyme fluxes using linear optimization.

2.3. Metabolic Control Analysis

Metabolic Control Analysis quantifies the relative response of enzyme flux R_j to the variations in the uptake flux R_i of metabolite i using flux control coefficient C_i^j,

$$C_i^j = \frac{F_i \cdot dF_j}{F_j \cdot dF_i} \tag{4}$$

Flux control coefficients quantifies the response of enzyme fluxes to external perturbations, and thus can be applied to validate genome-scale metabolic networks. Specifically, microbes respond to nutrient availability by regulating biochemical pathways and modulating the fluxes of enzymes. In the cases of no response, enzyme fluxes vary linearly with the fluxes of nutrient uptake, and the flux control coefficients remain at 1. In the case of metabolic responses, the relative increase in enzyme fluxes can be either larger or smaller than those of nutrient uptake fluxes, and the flux control coefficients become either larger or smaller than 1.

2.4. Hybrid Method

The hybrid method simulates microbial metabolism by combining FBA of genome-scale metabolic models and rate laws of microbial respiration. This method first separates the genome-scale metabolic models into a catabolic network model and a network model of biosynthesis, and then applies FBA separately to the respiration and biosynthesis networks. The catabolic network consumes electron donors and acceptors from the environment, and produces ATP. The biosynthesis network takes up nutrients, and uses the ATP to make new cells. It computes respiration rate according to the modified Monod equation.

The hybrid method applies FBA to the catabolic network using uptake rates of electron donors or acceptors as a constraint. The results include the yield and rate of ATP production. It then applies FBA to the biosynthesis network, using the ATP production rate as a constraint. The output includes the rates of growth and nutrient consumption.

2.5. dFBA

dFBA computes microbial growth rates by applying FBA directly to genome-scale metabolic models. It first applies the Michaelis-Menten equation to enzymes that directly consume electron donors or acceptors, and computes the rates at which electron donors and acceptors are consumed,

$$F_{\mathrm{N}} = -\sum_{\mathrm{j}} V'_{\max,\mathrm{j}} \cdot \prod_{\mathrm{S}} \frac{m_{\mathrm{S}}}{m_{\mathrm{S}} + K_{\mathrm{S,j}}} \tag{5}$$

where $V'_{\max,\mathrm{j}}$ is the maximum activity of enzyme j per cell dry weight (mmol·g^{-1} dw·hr^{-1}, or mmol·g^{-1}·hr^{-1}), m_{S} is the concentration of enzyme substrate, and $K_{\mathrm{S,j}}$ is the affinity constant. The negative sign appears because current genome-scale models describe the exchange of chemical compound between cell and the environment as the movement of nutrients from the extracellular space to the environment, or the release of nutrients from the cell. It then applies FBA to the genome-scale metabolic models, using the consumption rates of electron donors (or acceptors) as a constraint.

2.6. Application

We applied the hybrid method to methanogenesis by *M. barkeri.* Specifically, we first constructed a network model of methanogenesis and a network model of biosynthesis on the basis of the biochemical reactions within the genome-scale metabolic network model, iMG746 (Gonnerman et al.). The new network models included the following modifications: Na^+/H^+ antiporter that translocate one proton per Na^+, acetate transporter moves one acetate, together with one proton, into the cytoplasm, and the maintenance reaction is excluded from the biosynthesis network.

We validate the network models using the flux control coefficients of nutrient uptake. The flux control coefficients are calculated by applying FBA to the networks of methanogenesis and biosynthesis across a wide range of fluxes of substrate uptake or ATP production. The objective for the catabolic network is the maximization of ATP production fluxes. The objective for the biosynthesis network is the maximization of growth rate.

We simulate the metabolisms of *M. barkeri* using both the hybrid method and dFBA. Applying the hybrid method requires the kinetic parameters for methanogenesis and maintenance. Applying dFBA requires the kinetic parameters of enzymes that directly consume substrates, including methanol methyltransferase (MTA) and acetate transporters (ACT). The parameters for methanogenesis and for MTA are taken directly from previous experimental studies. The parameters for acetate transporters remain to be determined, and thus are assigned on the basis of the measurements for *Cycloclasticus oligotrophus* (table 1) (Button).

Applying both the hybrid method and dFBA also requires the rate of cellular maintenance. Gonnerman et al. estimated that *M. barkeri* had a maintenance rate of 1.75 mmol ATP$\cdot$g$^{-1}\cdot$hr^{-1}. Taking the growth yield of *M. barkeri* as 5.0 g$\cdot$mol-ATP^{-1}, *M. barkeri* has a specific maintenance rate of 2.4×10^{-6} s^{-1}.

We implemented the hybrid method by linking COBRA and PHREEQC software packages and by using the Microsoft Component Object Model (COM) Server as a control and data management source. COBRA and PHREEQC specialize in FBA and biogeochemical reaction modeling, respectively. COM is a Microsoft foundations technology for exchanging information among software packages of different platforms.

At each time step, the hybrid method first computes the specific rate of methanogenesis using the modified Monod equation (eq. 8) and chemical composition of the medium. The specific rate is then applied as a constraint in the analysis of the methanogenesis network using FBA. The results include the rates of ATP and methane production. Next, the ATP production rate is fed to the application of FBA to the biosynthesis network. The application assumes that the efficiency of biosynthesis is 0.5, accounting for the observation that actual growth yields are 50% of theoretical predictions. The analysis of the biosynthesis network predicts the rates of growth, nutrient consumption, and metabolite production, which are then utilized to compute the chemical composition and biomass composition in the medium at next time step.

3. Results and Discussion

We simulated the methanogenesis and growth of *M. barkeri* using the genome-scale metabolic model, iMG746. Metabolic network models include (1) stoichiometric equations and directions of enzyme reactions, and (2) exchange fluxes of energy sources and growth nutrients. We validated model iMG746 by checking the consistency in the constraints for exchange fluxes, and by carrying out metabolic control analysis. We then applied the hybrid method to simulate acetoclastic and methanol methanogenesis of *M. barkeri* in laboratory bioreactors. For comparison, we also simulated the metabolism of *M. barkeri* using dFBA.

3.1. Constraints of Exchange Fluxes

Like other genome-scale metabolic models, iMG746 describes nutrient uptake from the environment and the release of metabolites into the environment using 74 non-gene-associated exchange reactions. iMG746 considers a total of 15 nutrients. Out of these nutrients, acetate is constrained by a maximum uptake rate, and sulfite is limited by a narrow range – from a maximum uptake rate of 10^{-2} $\mathrm{mmol{\cdot}g^{-1}{\cdot}hr^{-1}}$ to a maximum production rate of $+10^{-4}$ $\mathrm{mmol{\cdot}g^{-1}{\cdot}hr^{-1}}$. For other nutrients, there is no limitation (see supplementary table S1).

There is no explanation on the special treatment on the constraints of sulfite exchange in model iMG746. If a typical constraint for element sources, -10^4 to $+10^4$ $\mathrm{mmol{\cdot}g^{-1}{\cdot}hr^{-1}}$, is applied to sulfite, FBA predicts that regardless of methanogenesis rates, acetate or methanol is produced at 10^3 $\mathrm{mmol{\cdot}g^{-1}{\cdot}hr^{-1}}$, the maximum rate allowed by COBRA (see figure 1).

3.2. Metabolic Control Analysis

Our metabolic control analysis first focused on the network of methanogenesis using acetate as a substrate. To compute flux control coefficients, we carried out a series of FBA by varying acetate exchange fluxes from 10^{-5} to 10 $\mathrm{mmol \cdot g^{-1} \cdot hr^{-1}}$. Note that we can perform the same analysis by varying the exchange flux of methane, and arrive at the same results (data not shown).

Figure 2 shows how the control coefficient of acetate on methane production varies with the uptake flux of acetate. The flux control coefficient remains constant at 1 over the entire range of analysis. A constant control coefficient suggests that there is little response from the enzymes with the methanogenesis network to the changes in acetate uptake fluxes and, as a result, the rate of methane production varies linearly with acetate uptake flux. We repeated the analysis for methanol methanogenesis, and the flux control coefficient of methanol on methane production also remains at unity, regardless of the uptake fluxes of methanol.

We then carried out the metabolic control analysis on the biosynthesis network by varying the flux of ATP supply from 10^{-5} to 10 $\mathrm{mmol \cdot g^{-1} \cdot hr^{-1}}$. Figure 2 shows how the flux control coefficient of biosynthesis rate varies with ATP formation rate of *M. barkeri*. The flux control coefficient for always stays above 1, indicating that in addition to the ATP flux from methanogenesis, there is an additional energy source that powers the synthesis of biomass. Furthermore, the negative slope of the control coefficient suggests that where the flux of ATP from methanogenesis becomes smaller, the contribution of this unknown energy to biosynthesis becomes more important.

Variations of control coefficients are not unexpected. As methanogens adapt to the changes in environmental conditions, they may regulate the activities of individual enzymes and re-route the pathways of metabolic functions, thereby modulating the significance of enzymes and the stoichiometric coefficients of nutrients and metabolic products in overall metabolism. However, the ranges of the flux control coefficient in the biosynthesis network is too large to be physiologically feasible. Specifically, the variations in the flux control coefficient suggests that in addition to methanogenesis, there is an additional source of ATP in the metabolism *M. barkeri* – a prediction that is inconsistent with the current state of knowledge.

3.3. *De Trop* Reaction

The metabolic control analysis pointed out that the biosynthesis network may harbor *de trop* reactions of ATP production. We computed the flux control coefficients for 200 non-gene-associated reactions in model iMG746 (Gonnerman et al.) (see supplementary table S2). Based on the results, we identified a *de trop* reaction of sulfite reductase (SULR2).

$$3\ \mathrm{F420\text{-}2H2} + \mathrm{SO3} \rightleftarrows 3\ \mathrm{F420\text{-}2} + 3\ \mathrm{H2O} + \mathrm{H2S} + \mathrm{H+}$$

This reaction was first identified in *Methanococcus maripaludis*, and its function was assumed as detoxifying or assimilating sulfite (Johnson and Mukhopadhyay). But FBA results show that this reaction proceeds backwards, oxidizing sulfide.

The occurrence of sulfide oxidation arises from the consumption of cysteine. Cysteine is an essential nutrient for the growth of *M. barkeri*. FBA predicted that, in model iMG746, cysteine desulfhydrase (CYSDS) utilize cysteine to produce pyruvate and sulfide,

$$\text{cysteine-L} + H_2O \rightarrow H_2S + NH_4^+ + \text{pyruvate}$$

SULR2 then proceeds backwards, oxidizing sulfide by reducing cofactor F420. Pyruvate is then converted to acetyl-CoA, and finally to acetate. The conversion of acetyl-CoA to acetate is carried out by PTAs and acetate kinase, producing ATP. In other words, FBA of iMG746 predicts that *M. barkeri* utilizes cysteine as an energy source.

To prevent the erroneous prediction, we applied two modifications to model iMG746. First, we set exchange fluxes for sulfite as –inf and +inf. Second, we limit the function of SULR2 to the assimilation of sulfite, by setting the reaction only in the forward direction. These modifications eliminated the erroneous predictions, and produced a constant flux control coefficient of ATP on growth over the range of ATP fluxes tested here.

3.4. Acetoclastic Methanogenesis

M. barkeri can make methane by dismutating acetate,

$$\text{Acetate} + H_2O \qquad CH_4 + HCO_3^-$$

Fukuzaki et al. examined the progress of acetate consumption by *M. barkeri*. In their experiments, cells grew at 37 ^{o}C in batch reactors of 125 mL that contained 50 mL of complex growth medium. The medium had pH 7.1 and contained 19.7 mM acetate. Figure 3 shows, according to the experimental observations, how acetate concentrations varied with time.

We simulated the experiments using the hybrid method. We evaluated the rates of acetate consumption by methanogenesis using the modified Monod equation and the kinetic parameters determined for *M. barkeri* (table 1). We set an initial biomass concentration of 50 $mg{\cdot}L^{-1}$ with a specific maintenance rate of $8.92{\times}10^{-3}$ h^{-1}.

As shown in Figure 3, the simulation results match with the observations of Fukuzaki et al. At the beginning of the experiments, acetate concentration decreased almost linearly with time. After two days into the experiments, the decrease slowed down. At day 4, acetate reached a constant concentration. On the other hand, methane and biomass concentrations increased linearly during the first two days of the experiments. At day 4, methane reached its maximum concentration of 0.3 mM. Note that because of the limited solubility, most methane built up in the headspace. At about day 2, biomass concentration also reached its maximum. Afterwards, biomass concentration decreased steadily with time.

Figure 3C shows the results of flux balance analysis on the networks of methanogenesis and biosynthesis. The ATP yield of the methanogenesis network remained constant during the experiments, and the value was 0.75. Acetate consumption by biosynthesis network also varied linearly with the ATP flux from the methanogenesis network. These linear variations gave a constant fraction of acetate utilized by the biosynthesis network. The acetate fraction to biosynthesis is 14.3%.

According to the rate law (eq. 1), methanogenesis rate depends on acetate concentrations and the energy available to *M. barkeri*. As shown in figure 3D, the kinetic factor of acetate decreased with time, due to the decrease in acetate concentration. The thermodynamic factor also decreased with time, because of the decrease in the available energy (fig 3E). As a result, in the first two days, the rate decreased linearly with time. At day 4, the available energy decreased close to the saved energy, which decreased the thermodynamic drive and thus the rate of methanogenesis to 0.

Microbial growth rate depends on biomass concentration and the specific growth rate. At the beginning, the growth rate increased with time, due to the increase in biomass concentrations (fig 3F). After reaching a maximum value at day 2, the growth rate started to decrease. This is because of the significant decrease in the specific growth rate. After day 4, the specific growth rate decreased near 0.

We also simulated the experimental progress by using dynamic FBA (dFBA). dFBA constrained the whole-cell metabolic network, including both catabolic (methanogenesis) and biosynthesis networks, using the rate of acetate uptake. *M. barkeri* uses acetate transporters to move acetate from the environment to the cytoplasm. Because the kinetic parameters of this enzyme are to be determined for methanogens, we evaluated the rate of acetate uptake using the Michaelis-Menten equation, and the parameters for acetate transporters (table 1). The simulation set an initial biomass concentration of 120 mg·L^{-1}.

As shown in figure 3, the dFBA predicted well the progress of acetate consumption at the beginning of the experiments. However, after day 2.5, dFBA returned an infeasible solution. This occurred because dFBA estimates growth rates using the method of FBA. FBA builds cell maintenance directly into genome-scale metabolic networks. Specifically, FBA represents cell maintenance using the reaction of ATP hydrolysis, and quantifies maintenance rate using the rate of ATP consumption. Under laboratory conditions, *M. barkeri* has a maintenance rate of 1.76 mmol·g^{-1}·hr^{-1}.

According to the results of dFBA, the rate of ATP production decreased with time, because of the decrease in acetate concentrations. At day 2.5, where the rate of ATP production by methanogenesis decreases to the rate of ATP consumption by maintenance, FBA returns infeasible solution, and the simulation stopped.

The difference between the two simulation results arises from the control of thermodynamics on methanogenesis rate. Methanogenesis is subject to significant thermodynamic control (Bethke et al.). The biogeochemical reaction modeling accounts for the energy available in the environment using the thermodynamic factor F_T. As shown in figure 3D and E, the thermodynamic factor was near unity at the beginning of the experiment, and decreased with time, due to the decrease in acetate concentrations and the accumulation of bicarbonate and methane. After four days into the experiment, the value decreased near 0, limiting significantly the rate of methanogenesis. In contrast, dFBA describes the rate of acetate uptake using the Michaleis-Menten equation, which calculates the uptake rate as a function of acetate concentration. As a result, methanogenesis continues until acetate was too low to support cell maintenance.

3.5. Methanol Methanogenesis

M. barkeri can also make methane using methanol,

$$\text{Methanol} = \frac{1}{4}\text{H}^{+} + \frac{1}{4}\text{H}_2\text{O} + \frac{1}{4}\text{HCO}_3^{-} + \frac{3}{4}\text{CH}_4 .$$

Smith and Mah analyzed methane production and growth of *M. barkeri* that utilized methanol. In their experiments, cells grew between 35-37 ^{o}C in 300 mL flasks that contained 100 mL complex medium. The medium had pH 6.5 and 25 mM methanol. Figure 4 shows, in accordance with the experiments, how dissolved methane and biomass concentrations varied with time.

We first simulated the experimental progress using the hybrid method. Specifically, in order to apply model iMG746 to methanol methanogenesis, the exchange flux of acetate is set to 0. We computed the specific rates of methanol consumption using the modified Monod equation (eq. 1) and the kinetic parameters determined for *M. barkeri* (table 1). We set an initial biomass concentration of 1.5 mg·L^{-1} with a specific maintenance rate of 8.92·10^{-3} h^{-1}.

The simulation results match the experimental observations of Smith and Mah (fig 4). According to the simulation results, during the first two days of the experiments, methanol concentration decreases gradually with time. Afterwards, methanol concentration decreased sharply, and decreased near 0 at day 4. Dissolved methane and biomass concentrations increased slowly at the beginning and after day 2, the concentrations increased relatively fast. At day 4, methane and biomass concentrations reached their maximum values, 0.33 mM and 85 mg·L^{-1}, respectively.

Figure 4C shows the results of flux balance analysis on the networks of methanogenesis and biosynthesis. For methanogenesis, the ATP yield remained at 0.75 during the experiments. Methanol was consumed by both methanogenesis and biosynthesis networks. The fraction of methanol consumption by biosynthesis in the total methanol consumption was constant, and the value was 18%.

According to the modified Monod equation, methanogenesis rate depends on the kinetic factor of methanol and the thermodynamic factor of the available energy. As shown in figure 4D, both the kinetic and thermodynamic factor were near unity at the beginning of the experiment (see table 1). The kinetic factor decreased slowly with time. After 3.5 days into the experiments, the kinetic factor decreased sharply. After the kinetic factor decreased near 0 at day 4, the thermodynamic factor also decreased to 0. As a result, the methanogenesis rate is controlled mainly by the kinetic factor, and its variation with time followed the same pattern as that of the kinetic factor.

Microbial growth rate depends on biomass concentration and the specific growth rate. According to the FBA results of the biosynthesis network, the specific growth rate is proportional to the specific rate of methanogenesis (fig 4F,G). At the beginning, the growth rate increased with time, due to the increase in biomass concentration (fig 4H). After reaching a maximum value at day 4, the growth rate started to decrease and decrease below 0, because of the significant decrease in the specific rate of growth. As a result, the population sizes started to decline.

We simulated the experiments using dFBA. We calculated the rate of methanol uptake by applying the Michaelis-Menten equation to methanol methyltransferase (MTA) of *M. barkeri* (see table 1). The simulation set an initial biomass concentration of 10^{-8} $\mathrm{mg \cdot L^{-1}}$ or 1 cell per reactor – the lowest possible concentration in the experiments.

As shown in figure 4, the dFBA did not predict well the methane production or microbial growth. After 11 hours, dFBA returned an infeasible solution. This occurred because dFBA estimates growth rates using the method of FBA. FBA builds cell maintenance directly into genome-scale metabolic networks. Specifically, FBA represents cell maintenance using the reaction of ATP hydrolysis, and quantifies maintenance rate using the rate of ATP consumption. Under laboratory conditions, *M. barkeri* has a maintenance rate of 1.75 $\mathrm{mmol \cdot g^{-1} \cdot hr^{-1}}$.

According to the results of dFBA, the rate of ATP production decreased with time, because of the decrease in methanol concentrations (fig 4A). After hour 11, where the rate of ATP production by methanogenesis decreases to the rate of ATP consumption by maintenance, FBA returns an infeasible solution, and the simulation stopped.

3.6. Diauxic Growth

Where both acetate and methanol are present, *M. barkeri* demonstrates a diauxic pattern in its growth. *M. barkeri* first grows by utilizing methanol as a substrate. After methanol is depleted, it starts to use acetate as a substrate (Smith and Mah). Diauxic growth is widespread in prokaryotes, and is attributed to cell's preferential usage of substrates that support fast growth. Genome-informed biogeochemical modeling explicitly simulates microbial growth rates, and thus provides a framework for simulating diauxic growth of microbes.

Smith and Mah grew *M. barkeri* using both acetate and methanol. In their experiments, cells grew at 35-37 ^{o}C in a medium that had pH 6.5. Figure 5 shows, according to the experimental observations, how methane and biomass concentrations vary with time. The isotope labeling experiments demonstrated that methane was first produced from methanol and, after methanol was consumed, then derived from acetate (Smith and Mah).

We simulated the diauxic growth experiments of *M. barkeri* using genome-informed biogeochemical modeling. We set the initial methanol concentration at 24.4 mM, and the initial acetate concentration at 53.0 mM. Following the above examples, we computed the uptake rates of acetate and methanol using the modified Monod equation (eq. 1, table 1). We assume that, where acetate and methanol are both available, *M. barkeri* use the one that supports higher growth rate. We also assumed that after one substrate is depleted, there was a lag time before the utilization of the other substrate. We found that a lag phase of 11.87 days (285 hours) describes well the experimental observations. We set an initial biomass concentration of 2 mg·L^{-1} with a specific maintenance rate of 8.92×10^{-3} h^{-1}.

The simulation results describe well the experimental observations of Smith and Mah. As shown in figure 5, *M. barkeri* first grew by utilizing methanol as a substrate, decreasing methanol concentrations. After 4.5 days into the experiments, methanol was depleted. Methanol methanogenesis accumulated inorganic carbon and methane in the media. At the end of methanol methanogenesis, methane and TIC reached a concentration of 0.23 mM and 4.4 mM respectively.

After methanol was depleted, we assumed that *M. barkeri* entered a lag phase till day 12. During this lag phase, the biomass concentration decreased because of cellular maintenance. After day 12, *M. barkeri* started to grow by using acetate as substrate, accumulating more TIC and methane in the reactors. After day 19, methane and TIC reached their maximum concentrations of 0.72 mM and 40 mM respectively.

FBA predicted that, at the beginning of the experiments (fig 5C), *M. barkeri* acquired a growth rate of 1.05 day^{-1} by utilizing methanol, larger than the rate of 0.65 day^{-1} by using acetate. As a result, methanol was utilized to support the growth. After day 4 no methanogenesis or growth occurred as the result of a lag phase imposed on acetoclastic methanogenesis.

At about day 12, the methanogenesis and growth of *M. barkeri* resumed, consuming acetate as a substrate. The decreases in acetate concentration decreased the available energy (fig 5E). After day 20, the available energy was close to the value of the saved energy, which decreased the thermodynamic drive and thus the rate of methanogenesis and growth to 0.

3.7. Discussion

We applied the hybrid method of FBA and Monod equation to simulating the metabolisms of *M. barkeri* utilizing acetate and methanol. The hybrid method predicted well the laboratory observations, including the diauxic metabolism of acetate and methanol. The hybrid method represents an improvement over both dFBA and the traditional approach of rate laws for simulating the kinetics of microbial metabolisms.

The hybrid method brings out unprecedented details of genome-scale metabolic dynamics to biogeochemical reaction modeling. The traditional approach separates microbial metabolism into respiration and growth. It simulates respiration reactions using rate laws, such as the Monod equation and its modified forms. For microbial growth, because of the uncertainties in the reaction stoichiometry, the simulated rates of nutrient consumptions are not up to the accuracy in the simulations of catabolic reactions.

The traditional approach is also limited in that it takes microbial cells as a black box and does not account for biochemical pathways or metabolic regulation. It assumes that reaction stoichiometry and microbial parameters remain constant, regardless of environmental conditions.

The hybrid method builds on dFBA, and includes significant modifications for improving the quality of the predictions. The method of dFBA combines FBA with the Michaleis-Menten equation. FBA predicts metabolic fluxes and growth rates from the uptake rates of substrates, not the concentrations of substrates. The prediction is arrived at using linear optimization. In essence, FBA predicts the maximum yields of microbial metabolism, the production of metabolites and biomass per substrate.

Substrate uptake fluxes are computed by applying the Michaleis-Menten equation to the enzymes that directly consume substrates. But applying the Michaelis-Menten equation to natural microbes is challenging. First, evaluating the Michaelis-Menten equation requires both Michaelis constant for substrates and the enzyme activity per cell. But both parameters are not available for most natural microbes. Microbes are known for their capability of adapting to environmental conditions. Such adaptation may result in modifications in the whole-cell activity at different environmental conditions, such as different substrate concentrations.

The Michaleis-Menten equation oversimplifies the kinetics of enzymes by only accounting for substrate concentrations. But both reaction products and reaction thermodynamics may place a significant control on fluxes of enzymes, especially for those that proceeds close to thermodynamic equilibrium. Furthermore, for most cells, the concentrations of metabolites are not known, and so is the thermodynamic drives.

Furthermore, by applying activities of individual enzymes as constraints for the entire metabolic networks, dFBA assumes that microbial metabolism is limited locally by these nutrient-consuming enzymes. However, this assumption may not be applicable to most metabolisms. Previous studies show that microbial metabolism is often limited by enzymes within cells. Previous sensitivity analyese show that dFBA predictions were highly dependent on the kinetic parameters of substrate uptake enzymes (Klier). Previous applications also show that dFBA may overestimate microbial activities by one order of magnitude (Scheibe et al.).

The hybrid method differs from dFBA by using rate laws to account for environmental conditions. In the environment, electron donors and acceptors are often limited, and so is the energy available to microbes. In the modified Monod equation (eq. 1), the kinetic factors account for the availability of electron donors and acceptors, and the thermodynamic factor accounts for the energy available in the environment. The hybrid method applies the respiration rates as a constraint for steady-state enzyme fluxes within the entire network of catabolism, and thus can be viewed as a global constraint on microbial metabolic networks.

The hybrid method takes full advantage of the predictive power of FBA by separating genome-scale metabolic networks into the networks of respiration and biosynthesis. First, it applies FBA to catabolic networks to predict the maximum yield of ATP production, which feeds to microbial rate laws to compute the rates of respiration. Second, it applies FBA to biosynthesis networks to predict the maximum rates of microbial growth. In this way, the hybrid method does not set the yields of ATP and microbial growth as constants, but computes in real time these parameters by accounting for the dynamics of environmental conditions. In other words, the hybrid method brings unprecedented details of metabolic dynamics into the simulation of microbial metabolisms and how microbes may respond to changes in environmental conditions.

The hybrid method improves predictions by eliminating infeasible solutions of FBA. Previous studies show that at small substrate concentrations, FBA returns infeasible solutions and suggests that microbial metabolism fails to proceed. Such predictions underestimated the extent of substrate consumption by microbial metabolism (Scheibe et al.; Tartakovsky et al.). Previous studies attributed the infeasible solution and the under-prediction of reaction extents to the under-constrained nature of genome-scale metabolic models – large numbers of unknown enzyme fluxes compared to few defined relationships among the fluxes (Scheibe et al.; Tartakovsky et al.).

In the case of iMG746, the infeasible solutions arise from the consideration of cellular maintenance by FBA. Specifically, FBA includes cell maintenance into genome-scale metabolic networks using a hypothetical reaction of ATP hydrolysis. At small concentrations of substrates, as in most natural environments, enzyme fluxes, including those of ATP synthase, are also small. Where the ATP fluxes decrease below the flux of ATP consumption by maintenance, there is no solution of enzyme fluxes that can support microbial growth.

The hybrid method keeps the classical framework of biogeochemical modeling, by accounting for maintenance rate in predicting biomass concentrations. It removes the hypothetical reaction of cellular maintenance from the genome-scale metabolic model. In this way, it relaxes the constraint of cellular maintenance on both catabolism and biosynthesis, and makes possible for FBA to locate feasible solutions at small rates of substrate uptake.

The accuracy of the predictions by the hybrid method and dFBA depends on both uptake fluxes and genome-scale metabolic models. In the natural environment, both energy sources and nutrients can vary from limited in oligotrophic environments to abundant in eutrophic settings. Where the concentrations of energy sources and nutrients change, microbes have to regulate their metabolism by modulating fluxes through individual enzymes, in order to survive and grow.

Most genome-scale models are developed primarily for cell metabolism under laboratory conditions. These methods are validated based on laboratory observations. Thus, before applying these models to natural environments, it is import to ensure that these models are consistent with the physiology of the targeted microbes, and applicable to the environment of interest.

We propose to validate genome-scale metabolic models using metabolic control analysis. Genome-scale metabolic models are primarily developed and validated for laboratory and industrial applications. In laboratory and industrial reactors, energy sources and nutrients are abundant, driving microbial metabolisms at relatively large rates. But biogeochemical modeling mainly target natural environments, where nutrients are often limited and microbial metabolisms are sluggish. Thus a genome-scale metabolic model validated for abundant nutrient conditions may not be directly applicable to natural environments.

Metabolic control analysis is a sensitivity analysis of enzyme fluxes in biochemical pathways (Kacser and Burns; Heinrich and Rapoport). It quantifies the impact of individual enzymes on overall pathways using flux control coefficients (Flint et al.; Middleton and Kacser; Torres et al.). In addition, the metabolic control analysis has also been applied to investigate the regulation and design of biochemical pathways (Groen et al.; Salter, Knowles, and Pogson; Fell and Snell).

Metabolic control analysis quantifies the responses of enzyme fluxes within metabolic networks to external perturbations. Thus it represents a powerful tool for validating genome-scale metabolic models. Here, we designed a framework for validating genome-scale metabolic networks for the purpose of biogeochemical application. This method samples across a wide range of nutrient uptake fluxes, from those in oligotrophic environments to those typical of laboratory reactors. At each nutrient uptake flux, we applied FBA to the genome-scale model of interest, using the objective of maximizing the rates of ATP production or growth. The predicted variations in the fluxes of individual enzymes were then applied to compute the flux control coefficients.

We validated the current genome-scale metabolic model of *M. barkeri*, iMG746. The metabolic fluxes over a wide spectrum of the fluxes of substrate uptake and ATP synthesis were examined. FBA expresses the flux of an enzyme as a linear function of the fluxes of other enzymes in the networks. We expect that enzymes fluxes vary linearly with uptake fluxes, and the flux control coefficients remains constant. Any deviation from this linear trend either results from metabolic switch of different pathways or potential misbehavior of network models. This exercise on the model iMG746 revealed that the misbehavior of genome-scale models can arise from un-intended role of enzyme reactions in cell metabolism, and can be avoided by modifying reaction directionality. The common use of dFBA may be expected to work well for describing microbes in laboratory reactors. Metabolic models are currently validated for laboratory reactors, where available energy is not limed. However, application of this method to the natural environment presents a challenge, because energy availability is small. Integrating microbial rate laws improves FBA predictions at small respiration rates observed in the natural environment. Genome-scale metabolic models add an unprecedented amount of information to biogeochemical modeling. However, more of these metabolic models need to be validated for the natural environment. The hybrid method pairs the thermodynamic strengths of microbial rate laws with FBA yield predictions, to improve the simulation of microbes in the natural environment.

CHAPTER II

GENOME-SCALE METABOLIC VALIATION FOR THE NATURAL ENVIRONEMNT

1. Introduction

Genome-scale models are an increasingly important tool in many systems biology approaches. Growing knowledge and the creation of genetic databases have paved the way for the systematic reconstruction of metabolic models based on microbial genomes. Of 130 genome-scale metabolic models in the web-based resource SEED Model, only 22 of them validated with laboratory observations (Henry et al.). With the advent of high-throughput genome sequencing and annotation, as well as the use of systems biology approaches, (Thiele and Palsson) have already enabled the construction of manually-curated, genome-scale models describing the complete metabolism of several microorganisms (e.g., *Geobacter sulfurreducens*, *Geobacter metallireducens*, *Shewanella oneidensis*, *Methanosarcia barkeri,* and *Methanosarcina acetivorans*), which play an important role in the bioremediation of contaminated groundwater (Mahadevan et al.; Sun et al.; Pinchuk et al.; Ong et al.; Gonnerman et al.; Benedict et al.).

The representation of microbial processes in these reactive transport models is often based on the assumption of simple reaction rates and Monod kinetic equations that do not always account for microbial processes observed in the environment. Quantitative prediction techniques are needed to describe the individual metabolic activities of microorganisms to directly parametrize metabolic microbial processes.

Compared to kinetic models, which require many kinetic parameters that are difficult to measure, constraint-based modeling does not require extensive model parametrization (Orth, Thiele, and Palsson; King et al.). In genome-scale modeling, the specific uptake flux rates of substrates are usually determined by Monod equations. However, in some cases Michaelis−Menten equations, which have the same mathematical relationship are used. These two relationships differ in scale, one representing the response of the entire microorganism, the other describing enzyme response. The flux of substrates is an important capacity constraint in constraint-based modeling and any uncertainties in reaction rates or reaction directions have a significant influence on predicted microbial growth rates.

Genome scale metabolic models have been applied to predict the metabolic activities of natural microorganisms, including *G. sulfurreducens* in a uranium-contaminated aquifer (Scheibe et al.; Fang et al.; Tartakovsky et al.). However, the direct applications of constraint-based modeling tend to overproduce the activity of microorganisms. Growth yield prediction from genome-scale models are scaled by a control coefficient to better fit field observations (Scheibe et al.).

This study investigated whether use of the current genome-scale models of *Geobacter metallireducens*, *Mehtanosarcina barkeri*, *Methanosarcina acetivorans*, and *Shewanella* stains MR1/4/W318 are applicable for natural environments, as described by their original authors. We introduce a new method for validating genome-scale metabolic models for application to the natural environment.

2. Method

2.1. Catabolic and Biosynthesis Networks

Genome-scale metabolic models summarize entire biochemical reactions of cells using stoichiometric reaction equations. The genome-scale reaction equations can be summarized using a stoichiometric matrix S of size $m \times n$. Here, n is the number of biochemical reactions of the entire organism, and m is the number of metabolites consumed or produced by the enzymes. Element S_{ij} is the stoichiometric coefficient of metabolite i in the reaction equation of enzyme j. These models may also specify reaction directions, and include nutrient fluxes in or out of the cell.

To validate these models, we group the biochemical reactions in genome-scale metabolic models into catabolic and biosynthesis models. The catabolic models includes biochemical reactions that take up electron donors and acceptors from the environment, oxidize electron donors, and transfer electrons to the acceptors, and produces ATP.

The biosynthesis models consider biochemical reactions that transport nutrients into the cytoplasm, and uses ATP to synthesize amino acids, nucleic acids, and other monomers. Following the common practice in flux balance analysis, we represent biosynthesis using an unbalanced reaction equation that converts ATP and monomers to ADP, phosphate, and other waste products. This biosynthesis reaction has a rate unit of per hr.

2.2. Flux Balance Analysis

Flux balance analysis (FBA) is a standard method for analyzing genome-scale metabolic models. These method assumes that cell metabolism is at steady state, and applies the principle of mass balance to metabolites, $J = S \cdot R$. Here J and R are vectors; element Ji is the net rate or flux of metabolite A_i production ($mmol \cdot g^{-1} \cdot s^{-1}$), and element Rj is the rate of enzyme j ($mmol \cdot g^{-1} \cdot s^{-1}$). For internal metabolites, the fluxes are 0.

The goal of FBA is to estimate the rates of biochemical reactions that achieve specific metabolic objectives. A common objective is to maximize rates of microbial growth. FBA solves for enzyme rates using linear optimization. We implemented FBA by using the COBRA toolbox for MatLab and Gruobi 6.5. Gruobi 6.5 is a linear optimization solver that provides acceptable accuracy at small nutrient fluxes.

2.3. Validation and Revision

We validated genome-scale metabolic models by analyzing the response of cell metabolism to uptake nutrients. According to metabolic control analysis (MCA), these responses can be quantified using flux control coefficient $C_{\mathrm{i}}^{\mathrm{j}}$,

$$C_{\mathrm{i}}^{\mathrm{j}} = \frac{F_{\mathrm{i}} \cdot dF_{\mathrm{j}}}{F_{\mathrm{j}} \cdot dF_{\mathrm{i}}}$$

To compute flux control coefficients, we apply FBA across a wide range of nutrient fluxes, from 10^{-5} mmol·g^{-1}·hr^{-1}, as reported in natural environments, to typical values in bioreactors, 10^{+2} mmol·g^{-1}·hr^{-1}. In applying FBA to the catabolic and biosynthesis models, we assume that cells maximize the rates of ATP and biomass synthesis, respectively.

In theory, FBA predictions, both enzyme and biosynthesis rates, increase linearly with nutrient fluxes. As a result, the flux control coefficients remain constant, regardless of nutrient uptake fluxes. The slopes of the increases give stoichiometric ratios of external substrates, internal metabolites, and final products in both catabolism and biosynthesis, including the yields of ATP other metabolites, and biomass.

We validate genome-scale models on the basis of flux control confidents and the ATP and biomass yields. Where the flux control coefficients vary with uptake nutrients, or where the yield predictions differ significantly from the values reported previously, we consider that genome-scale metabolic models require refinement.

To improve metabolic models, we first ensure that biochemical reactions are consistent with genome annotation and current knowledge. We set reaction directions and rates to allow cells to carry out essential functions.

To reflect biogeochemical conditions of the environments, we only allow the uptake of growth factors, but disabled the uptake fluxes of other complex organic compounds. We also limit the production and release of metabolites into the environment on the basis of cell physiology.

2.4. Application

We validated the published genome-scale metabolic models for *Methanosarcina barkeri*, *Methanosarcina acetivorans*, *Shewanella oneidensis, Shewanella putrefaciens, Shewanella sp. MR4*, *Geobacter metallireducens* (table 2).

3. Results

3.1. Dependence of Biosynthesis on Nutrient Uptake

We applied FBA to genome-scale metabolic models to predict biosynthesis rates across a wide ranges of substrate uptake fluxes (table 3). We varied substrate exchange fluxes from 10^{-5} to 100 $\text{mmol}\cdot\text{g}^{-1}\cdot\text{hr}^{-1}$. In theory, FBA predictions vary linearly with substrate uptake fluxes. Linear optimization problems maximize or minimize an objective function that depends linearly on variables (Sloan). Specifically, we constrained the substrate uptake (Bordbar et al.). But metabolic shift, utilization of new nutrients and pathways, may move the predictions away from the linear dependence.

Figures 6 and 7 show the biosynthesis rates predicted at different substrate uptake fluxes and the flux control coefficients for biosynthesis. For *Geobacter metallireducens* and three *Shewanella* strains, we predicted their biosynthesis rates by providing ferric mineral as an electron acceptor. For *Geobacter metallireducens* (model iAF987) oxidizing acetate and the three *Shewanella* strains oxidizing lactate, their biosynthesis rates vary linearly with the uptake fluxes. By providing H_2 as an electron donor and CO_2 as a carbon source, the biosynthesis rate of strain MR1 also vary linearly with H_2 uptake flux. But the biosynthesis rate of strain MR1 becomes invariant with H_2 uptake fluxes by replacing CO_2 with acetate as a carbon source. No biosynthesis is predicted for strain W318 and MR4 growing on H_2. The biosynthesis rates of *M. acetivorans* (model iMB745) and *M. barkeri* (iMG746) do not follow a linear trend with the uptake fluxes of their substrates.

3.2. ATP Yield

We evaluated the flux control coefficient of substrate uptake on ATP synthesis by applying FBA to the catabolic reaction models. We varied the substrate uptake fluxes from 10^{-5} to 100 $\text{mmol}\cdot\text{g}^{-1}\cdot\text{hr}^{-1}$, and assumed that the catabolic models maximize the rates of ATP production (figure 8). The results in show that the biosynthesis flux control coefficients remained constant for *M. acetivorans* and *G. metallireducens* (figure 9). The biosynthesis coefficients of *M. barkeri* are 1 at substrate uptake rate larger than 0.1 $\text{mmol}\cdot\text{g}^{-1}\cdot\text{hr}^{-1}$. Below this threshold, the value of the control coefficient decreases to 0. Based on the FBA predictions, we estimated the ATP yield per substrate. As shown in Table 4, the predicted yield for *Geobacter* is close to the previous estimation. The predicted ATP yields of methanogenesis for both methanogens are smaller than, but within 70% of, the current estimations. The predictions for lactate oxidation by *Shewanella* are half of the current estimations. According to the FBA result, *Shewanella* species do not make any ATP by oxidizing H2. The FBA prediction was 3 ATP per lactate, about 20% higher.

3.3. Growth Yield per ATP

We computed the flux control coefficients for the biosynthesis network models. We enabled the transport of essential nutrients (electron donors and acceptors, nitrogen, sulfur, and vitamins) while providing ATP, and maximizing the biosynthesis rate (figure 10). The FBA failed for provide a solution for *M. acetivorans* utilizing acetate or methanol as a carbon source. For other microbes, their biosynthesis rates vary linearly with the flux of ATP supply (figure 11).

We computed the biomass yield as the ratios of the predicted biosynthesis rates and the flux of ATP supply (table 5). The predicted growth yields of *M. barkeri* are larger, but within a factor of two, than previous laboratory observations. But the predicted values for other microbes are much larger than the observations, up to three orders of magnitude.

4. Revision

The above results suggest that the genome-scale metabolic models of iMB745, iMG746, iAF987, iMR1_799, iW318_789, and iMR4_812 require improvements before applying to natural environments. When energy yields of genome-scale metabolic models do not agree with experimental observations there are three factors that contribute to that erroneous results. The stoichiometric number of protons translocated outside the membrane determine the flux of ATP synthase. It is also possible that some reaction equations in the catabolic network do not conform to the mass balance law assumption. The last possibility is that there could be unconstrained, de trop, reactions causing misbehavior in the network. We first evaluated the reaction stoichiometry of proton translocation reactions to ensure they were mass balanced and reflected current knowledge (see supplementary table S3).

4.1. *Methanosarcina barkeri*

In *M. barkeri*, iMG746, we changed the stoichiometric coefficients of sodium pump (NAt3_1) to translocate one proton per sodium cation. iMG746 described acetate transporter (ACt3r) as a bicarbonate antiporter. Based on gene annotation and laboratory observations of *Methanosarcina mazei* (Welte, Kröninger, and Deppenmeier), we changed to a proton symporter:

ac[e] + h[e] <=> ac[c] + h[c]

iMG746 includes a sulfite transporter (SO3t) to transport sulfite across the membrane,

h[e] + so3[e] <= h[c] + so3[c]

Because no gene homology is identified from the genome, we removed this reaction. Additionally, the model sets a maximum rate of alcohol dehydrogenase (ALCD1y) at 1 $\text{mmol}\cdot\text{g}^{-1}\cdot\text{hr}^{-1}$. When degrading methanol as the carbon source and at ATP fluxes above 5 $\text{mmol}\cdot\text{g}^{-1}\cdot\text{hr}^{-1}$, the model is NADP limited with this constraint. We set the maximum rate for ALCD1y at 1000 $\text{mmol}\cdot\text{g}^{-1}\cdot\text{hr}^{-1}$.

4.2. *Methanosarcina acetivorans*

In *M. acetivorans*, iMB745 assumed that methanophenazine reductase (RNF) translocates one sodium cation per ferredoxin. We changed to two sodium cations per ferredoxin:

2 fdred[c] + 2 h[c] + mphen[c] + 2 na1[c] -> 2 fdox[c] + mphenh2[c] + 2 na1[e]

iMB745 described acetate transporter (ACt3r) as a bicarbonate antiporter. Based on *Methanosarcina mazei*, we changed to a proton symporter:

ac[e] + h[e] <=> ac[c] + h[c]

iMB745 sets the minimum and maximum rate of carbon monoxide dehydrogenase (CODHr) at 0. We set the minimum and maximum rate at -1000 and 1000 mmol·g^{-1}·hr^{-1}, respectively. The also constrains the maximum rate of pyruvate synthase (POR2) at 1 mmol·g^{-1}·hr^{-1}. We set the maximum rate of POR2 at 1000 mmol·g^{-1}·hr^{-1}. Without this modification the model becomes pyruvate limited when acetate is the carbon source and at ATP fluxes greater than 10 mmol·g^{-1}·hr^{-1}.

4.3. *Geobacter metallireducens*

For *G. metallireducens*, iAF987 assumed that the ATP synthase XXX makes three ATP by translocating X protons. We changed the H$^+$/ATP ratio to 4,

adp[c] + pi[c] + 4 h[p] <=> atp[c] + 3 h[c] + h2o[c].

this stoichiometric ratio is the standard convention for most metabolic networks.

In model iAF987, three cytochrome oxidoreductase reactions (CYTMQORpp, CYTBMQORpp, and PPCCYTCpp) move protons across membrane. The reaction CYTMQOR3pp denotes the transfer of electrons from the menaquinone pool to the cytochromes at the inner membrane,

2 ficytC[c] + mql8[c] -> 2 focytC[c] + mqn8[c] + 3 h[p]

where ficytC and focytC are ferro cytochromes in oxidized and reduced forms, mql8 and mqn8 are menaquinone-8 in oxidized and reduced forms. We changed this enzyme translocates one proton per electron,

2 ficytC[c] + mql8[c] -> 2 focytC[c] + mqn8[c] + 2 h[p]

PPCCYTCpp transfers electrons from the periplasmic cytochromes to the outer-membrane cytochromes of model iAF987,

focytC[c] + h[p] + ppco[p] <=> ficytC[c] + h[c] + ppcr[p]

where focytC and ficytC are ferro cytochromes of the cytoplasm and ppco and ppcr are cytochromes of the periplasm. Because no mechanism is available to account for the link between proton translocation and electron-transfer outside of the cytoplasmic membrane, we removed proton translocation from the reaction.

G. metallireducens genome codes only one gene of NADH dehydrogenase (Butler, Young, and Lovley). Model iAF987 has two NADH dehydrogenase – NADH17pp translocate protons and NADH10 does not. The model assumes that NADH17pp translocates 3 protons per NADH,

3 h[c] + mqn8[c] + nadh[c] -> mql8[c] + nad[c] + 2 h[p]

We assume that NADH17pp translocates two protons per NADH.

4.4. *Shewanella*

Pitchuck et al. built a genome-scale metabolic model iSO783 for *Shewanella onedensis* MR-1. Ong et al. revised iSO783, and renamed iMR1_799. Ong et al. (2014) also built genome-scale models for *Shewanella putrifaciens* (iW318_789) and *Shewanella* strain MR4 (iMR4_812).

We modified the following reactions in iMR1_799, iW318_789, and iMR4_812. Hydrogenase HYDi5 and HYDi6 oxidize hydrogen gas, but do not translocate proton. But hydrogenases are capable of proton translocation (Cao and Hall). We changed the reaction to pump out one proton per election,

h2[e] + 2 h[c] + mqn7[c] -> mql7[c] + 2 h[e] h2[e] + 2 h[c] + mmqn7[c] -> mmql7[c] + 2 h[e]

Lactate dehydrogenase (LDH5) oxidizes lactate to pyruvate by reducing methylmenaquinone to methylmenaquinol, without proton translocation. We modified the reaction to pump 2 protons per lactate,

d-lactate[c] + 3 h[c] + mmqn7[c] -> mmql7[c] + 2 h[e] + pyr[c]

In the three *Shewanella* models, lipid synthesis requires phosphatidylglycerophosphate (pglyp). This compound can be produced from dihydroxyacetone phosphate (dhap) by glycerol-3-phosphate dehydrogenase (G3PD2) and glycerol-3-phosphate 3-phosphatidyltransferase (PGSA),

glyc3p [c] + nadp[c] <=> dhap[c] + h[c] + nadph[c]

cdpdag[c] + glyc3p[c] <=> cmp[c] + h[c] + pglyp[c]

consuming NADPH.

The *Shewanella* genome sequence contains three NADH dehydrogenases (NADH12, NADH14, NADH4). Under anaerobic conditions, two NADH dehydrogenases can make NADH using electrons from the reduced forms of menaquinone (NADH12) and methylmenaquinone (NADH14). Ubiquinone is used by NADH14, which is only during aerobic respiration, so we disabled this reaction (Søballe and Poole).

The three *Shewanella* models assume that the NADH dehydrogenases can only oxidize NADH. We allow the enzymes to either oxidize or produce NADH. These models also assume that NAD transhydrogenase (THD5) transfer electrons only from NADPH to NAD, and NADP transhydrogenase (THD2) transfers electrons only from NADH to NADP:

nad[c] + nadph[c] => nadh[c] + nadp[c]

2 h[e] + nadh[c] + nadp[c] => 2 h[c] + nad[c] + nadph[c]

We assume that the enzymes can drive the reactions both forward and backward (Søballe and Poole).

4.5. Assembly ATP Cost

Genome-scale metabolic models describe cell reproduction in two steps. First, biosynthesis pathways convert sources of C, N, P, and S to amino acids, nucleic acids, peptidoglycan, and other monomers. Second, the monomers are then converted to new biomass. The first step is represented using a series of enzyme-driven biochemical reactions, but the second step is simplified as a single step, mass-unbalanced, hypothetical reaction. Both steps consume ATPs, and the stoichiometric number of ATP in the two steps determine the biomass yield per ATP. The ATP consumed by the assembly reaction is called GAM in genome-scale modeling (Thiele and Palsson). In microbiology and geobiology, maintenance refers to metabolic processes that consume ATP but do not contribute to the production of new biomass. Thus, we rename GAM as biomass-assembly ATP cost (BAC).

A standard protocol estimates GAM and NGAM by fitting FBA predictions to nutrient uptake fluxes at different growth rates observed in chemostat experiments (Thiele and Palsson). This method has been applied to the estimation of GAM for *M. barkeri* and *Shewallena oneidensis*. Another method is based on the theoretical estimations of biomass synthesis (Stouthamer). This method accounts for the amino acid compositions and ATP cost for protein maturation and mRNA turnover. This method has been applied to *M. acetivorans* and *G. metallireducens*. Table 6 shows the values of

GAM and NGAM estimated for various genome-scale metabolic models. The values span over one order of magnitude, from 18 to 220 mmol ATP·g^{-1}.

We estimated the biomass assembly cost by accounting for protein maturation, mRNA turnover, and phosphate and nitrogen transport. Protein maturation consumes 5 ATP per amino acid (Arnold and Nikoloski). The total protein maturation cost is calculated from the total amount of amino acid per cell, multiplied by the amount of ATP necessary to synthesize an amino acid. Although the genome-scale metabolic models include the first step of the maturation process, t-RNA charging, these reactions never participate in biosynthesis, because in these models, biosynthesis reactions take amino acids as reactants, and hence excluding t-RNA charging. The ATP costs for mRNA turnover and nutrient transport are taken as 1.39 mmol ATP·g^{-1}, and 5.2 mmol ATP·g^{-1}, respectively, the values estimated for E. coli (Stouthamer).

The direct application of Stouthamer's method gives a relatively small GAM values, which in turn overestimates the yield of biomass. To circumvent this setback, previous studies have included an unknown factor in order to raise GAM. Alternatively, we suggest that the efficiency of biomass synthesis is 50%, which reduces the biomass yield by half.

Table 7 shows the GAM estimated for different microbes. The values fall into a narrow range, 30.9 to 35.5 mmol ATP·g^{-1}·hr^{-1}. These values fall within the lower range of the values used in current models.

The monomer synthesis cost is the sum of ATP consumption by all the enzyme reactions of biosynthesis. We estimated the monomer synthesis cost by applying FBA to the biosynthesis models, and by removing the ATP cost from the biosynthesis reaction.

Table 7 compares the estimated monomer synthesis costs. The GAM values for many of the models are close (45.1-62.7 mmol ATP·g^{-1}), except for *E. coli* K-12, which was substantially larger. The values estimated for *M. acetivorans* was lower than those for rest of the field.

The differences in the monomer synthesis cost arise from the differences in biomass composition. The synthesis of different macromolecules requires different ATPs. Furthermore, different microbes have different amino acid compositions and the ATP costs are different for the synthesis of different amino acids (supplementary table S4). We estimated the ATP cost for individual amino acids by applying FBA to the biosynthesis network, removing the biosynthesis reaction, and maximizing the production flux of specific amino acids.

The relative significances of amino acids are derived either from mass of protein from experiments (*E. coli* K-12 model iAF1260, and *Shewanella* models). In *M. barkeri* model iMG746 and *M. acetivorans* model iMB745, amino acid significance is directly applied from *E. coli* K-12, model iAF1260 (Feist et al.). Another approach is to count codons in genome sequence (*G. mellireducenes* model iAF987). The number of codons for each amino acid is converted to molar fraction by dividing the total codons reads from the genome. The molar fraction is multiplied by the molecular weight of each amino to give units of mmol AA·g protein^{-1} for the abundance. Because gene frequency does not always translate directly to protein frequencies, this method is less desirable.

Between iMG746 and iMB745, most AA significance are similar however, there are a few abundances that stand out between the two species (figure 12). For *M. barkeri*:

Alanine, glutamate, glycine, are higher than in *M. acetivorans*. Glutamine, serine, and tyrosine are higher in *M. acetivorians*.

In *G. mellireducenes*, proteins account less for the biomass than *M. barkeri* and *M. acetivorans*, but more than *Shewanella*. For the three *Shewanella* species, their amino acid compositions are the same. Note their serine and aspartic acid are comparable to *M. acetivorans*.

The revised models were compared to *Escherichia coli K-12,* model iAF1260. This model needs modification to be comparable to the other models. Model iAF1260 has a different mechanism of nutrient transport from the other models. This model uses active transport of nutrients from the periplasm compartment to the cytoplasmic compartment. While active transport reactions are included in the other models, these reactions are supplemented with proton symporters/antiporters, and diffusion reactions. We replaced transport reactions in model iAF1260 with facilitated transport mechanisms.

We used FBA to compare the predicted energy cost of producing each amino acid between models iMB745 and iMG746 (table 8). The amino acid composition measured for iAF1260 was used to make iMB745 and iMG746 models comparable. Both models were constrained to uptake the same essential nutrients. The largest difference in energy cost are reflected in Arginine, Serine, Leucine, and Glutamate (0.7-0.95 ATP·mmol^{-1} amino acid). The difference in formation energy could be explained by different biosynthesis reactions or different active pathways between the two models.

In model iMG746, serine is derived from acetate. Once acetate is converted to pyruvate through the acetoclastic pathway. Now in the form of acetyl-CoA, pyruvate synthase (POR2) converts it to pyruvate, diverted to the glycolysis pathway. PPDK consumes pyruvate and produces phosphoenolpyruvate (PEP) and this compound exits the pathway as 3-Phosphohydroxypyruvate. Phosphor-L-serine phosphatase (PSP_L) cleaves off the phosphorus group producing serine. Some of the acetate consumed is diverted to carbon monoxide dehydrogenase (CODH) to drive the reduction of ferrodoxin which is required by POR2 to proceed in the reverse direction.

Model iMB745 did not predict the same pathway for serine synthesis of serine as model iMG746. Instead serine is produced directly from cysteine. Cysteine synthase (CYSS) proceeds in reverse, consuming cysteine and acetate to produce acetyl-serine and hydrogen sulfide. Hydrogen sulfide is converted and expelled through methylsulfide synthase (MSS) which generates methyl sulfide. This reaction is non-gene associated in iMB745 and is not present in iMG746. The presence of this reaction allows CYSS to run in the forward direction, unconstrained. The production of methylsulfide has been reported in laboratory experiments, but the predicted methylsulfide production is larger than laboratory observations. The occurrence of the predicted pathways for serine synthesis remains to be tested.

We also compared the formation energy of amino acids to *E. coli* model, iAF1260. The amount of energy requires ranges from 1.13 – 2.29 times more than model iMG746. Aspartic acid, glutamine, serine and prolines show the largest increase when compared to *M. barkeri* and *M. acetivorans*. Model efficiency is probably cause of the greater energy requirements. Model iAF1260 has more biosynthesis reactions in each amino acid pathway compared to the methanogens. The synthesis of amino acids involves more enzymes and consumes more ATP. For example, to synthesis aspartic acid, iAF1260 involves 35 reactions and 2.1 ATPs, but in model iMB745, only 27 enzymes are required and, as a result, 0.52 ATPs are consumed per amono acids.

4.6. Futile loops

Metabolic network reactions with shared mechanisms and unconstrained reaction directions create futile loops that approach infinite flux. According to the FBA results, model iW318_789 contains three futile cycles. These cycles are driven by two or more enzymes of reaction rates at or near 1000 mmol·g^{-1}·hr^{-1} (table 9). The first cycle is driven by transaminases and dehydrogenases of three amino acids, leucine, isoleucine, and valanine, entered futile loops, with reaction rates at or near 1000 mmol·g·hr^{-1}. Transaminases transfer amine-group (NH_2-) between amino acids and α-keto acid,

akg[c] + amino_acid[c] <=> x-oxopentanoate [c] + glu_l[c]

Dehydrogenases oxidize amino acids,

h2o[c] + nad[c] + amino_acid[c] <=> x-oxopentanoate + h[c] + nadh[c] + nh4[c]

We removed the futile loops by imposing reaction directions on the transaminases and dehydrogenases. For isoleucine, both transaminase and dehydrogenase must run in backward direction; for isoleucine, transaminase runs in backward, and the dehydrogenase can run in both directions; for valanine, the transaminase can proceed in both directions, but the dehydrogenase must run backward.

The second cycle is from threonine transport. According to model iW318_789, threonine is transported into the cytoplasm together with proton (THRHT) or sodium cation (THRT4),

h[e] + thr_l[e] <=> h[c] + thr_l[c]

na1[e] + thr_l[e] -> na1[c] + thr_l[c]

Together with proton/sodium antiporter that exchanges sodium and proton across the membrane, the transporters carry out a futile cycle. To remove the loop, we assume that the threonine/proton symporter runs forward only.

The third cycle is at kinases. AMP kinases, adenylate kinase (ADK1) and guanylate kinase (ADK3), transfer phosphate group from atp or gtp to amp,

$$\text{amp[c]} + \text{atp[c]} \iff 2\ \text{adp[c]}$$

$$\text{amp[c]} + \text{gtp[c]} \iff \text{adp[c]} + \text{gdp[c]}$$

nucleoside-diphosphate kinase, NDPK1, exchanges the phosphate group between atp and gdp,

$$\text{atp[c]} + \text{gdp[c]} \iff \text{adp[c]} + \text{gtp[c]}$$

To remove the loop, we set ADKs and NDPK1 to proceed forward. Futile cycles can be more complex than two or three reactions, making it hard to determine reaction direction from hypothesis testing. These cycles exist because the reaction constraints on two or more enzymes are not known. This is probably the result of limited information about reaction mechanism and thermodynamics.

Removing futile loops through reaction direction hypothesis testing is time consuming and does not influence the overall result of the FBA solution. We manually reduced the futile loops for the remaining models. Any reaction with a flux approaching 1000 $\text{mmol}\cdot\text{g}^{-1}\cdot\text{hr}^{-1}$ we removed from the reaction network.

5. Prediction

5.1. Sweeping Results

We repeated the flux control analyses for the catabolic network, biosynthesis network, and the combined catabolic and biosynthesis network (or genome-scale metabolic network).

Figures 13 and 14 show how biosynthesis rates vary with nutrients and the results of MCA. Figures 15 and 16 show how the ATP formation rate of models iMB745, iMG746, iAF987, and *Shewanella* (iMR1_799, iW813, iMR4_812) vary with the consumption substrates and the flux control coefficient of ATP. Figures 17 and 18 show how the biosynthesis rate of models iMB745, iMG746, iAF987, and *Shewanella* (iMR1_799, iW813, iMR4_812) vary with production of ATP and how biosynthesis rate responds to changes in ATP flux. The modifications corrected the misbehavior of original models, and predicted linear dependence of ATP formation rates on substrate consumption rates, and the biosynthesis rates on ATP supply fluxes.

5.2. ATP Yield

The slopes of ATP fluxes over electron donor fluxes give the ATP yield per electron donor (table 10). The energy yield values all fell within the range of reported values. For model iMB745, the predicted yield was 0.5 molATP·mol acetate^{-1} and 0.625 mol ATP·mol mehtanol^{-1} which was in the range of 0-1 mol ATP·mol substrate^{-1} reported in literature. Model iMG746 also fell within this reported range for use of methanol with a predicted yield of 0.75 mol ATP·mol acetate^{-1}. This model predicted the reported value for hydrogenotrophic and acetoclastic methanogenesis. Model iAF987 and *Shewanella* models match the reported values for energy yield of those organisms.

5.3. Growth Yield per ATP

We used the modification to the metabolic network models mentioned above to evaluate the growth yield of each organism on different substrates (table 11). We expected to find growth yield that were comparably the same. The predicted growth yields for methanogens where within 10% of observed values, except for *M. acetivorans* degrading methanol which was 5% higher. *Shewanella* models (iMR1_799, iW318_789, iMR4_812) predicted growth yield within 13% of observed values when using lactate or acetate as carbon sources. *S. oneidensis* and *S. putrefaciens* predicted yield values within 1% of observations when using lactate as the carbon source. The predicted yield for *S. sp.* MR4 was father from observered values during lactate utilization. These models predicted a slightly lower yield for acetate as the carbon source. *S. oneidensis, S. putrefaciens, and S. sp.* MR4 under predicted the growth yield by 51% when using CO_2 as a carbon source. The estimated ATP assembly cost has the largest effect on growth yield (table 12).

5.4. Biomass Composition

The biomass equation in metabolic network models defines the average macromolecular composition of the cell. However, macromolecular cell composition is growth dependent. For example, as growth rate increases the cellular content of RNA increases, while the amount of protein, DNA, and content of cell wall polymers decreases (Novak et al.). FBA takes the rates of cellular metabolism at steady-state so an average macromolecular composition is used. The composition of protein, RNA, DNA, lipids, cell wall, and polysaccharides are set based on data available from literature (supplementary table S4). However, many reconstructions only have some of this data available. Another organism is usually used to fill in the gaps for biomass composition and *Escherichia coli* is normally used.

By using the nutrient and product exchange fluxes we determined the biomass composition be taking the elemental mass consumed by each organism. The biomass formula is shown in table 13. Elemental composition was normalized to the composition of nitrogen. There is not much variation in biomass composition between organisms. The most variation is observed in hydrogen composition, iMR4_812 and iW318_789 having the most and iMG746 having the least hydrogen. The similarity between the model may be due to their use of E. coli DNA, RNA, and lipid composition.

6. Discussion

We applied our validation method for de-coupled metabolic network models to methanogens and iron-reducing organisms. The validation method predicted energy and growth yield closer to reported in the literature, than the original models. The validation method represents a way to ensure metabolic models developed for laboratory environment can be applied to biogeochemical reaction modeling in the natural environment.

This new method for validation of genome-scale metabolic models ensures that they can be applied reliably to biogeochemical reaction modeling. The current method for model validation only evaluates that model under a limited range of substrate uptake rates and observed physiological behavior of microorganisms.

Our new approach tests the sensitivity of the model over a wide range of constraining rates. It decouples microbial metabolism; simulating respiration reaction separate from biosynthesis reactions.

The traditional approach to genome-scale metabolic models takes the cell as a system of reaction equations, simulating both respiration and biosynthesis pathways in the same linear programing problem. It assumes that reaction stoichiometry and microbial yield remain constant, regardless of conditions in the environment. The simulation of both catabolism and anabolism can lead to these pathways assimilating or stagnation reactions outside of their normal pathways.

In laboratory and industrial reactors, energy sources and nutrients are abundant, driving microbial metabolisms at relatively large rates. But biogeochemical modeling mainly target natural environments, where nutrients are often limited and microbial metabolisms are sluggish. Thus, a genome-scale metabolic model validated for abundant nutrient conditions may not be directly applicable to natural environments.

The new method of validation builds on laboratory validated and curated genome-scale metabolic models for improving the quality of microbial predictions when applied to the natural environment. The method use a sequence of FBA predictions of energy yield and biomass yield to identify issues to be revised in the model. FBA predicts metabolic fluxes and growth rates from the uptake rates of substrates, not the concentrations of substrates. The prediction is arrived at using linear optimization. FBA predicts the maximum yields of microbial metabolism, the production of metabolites and biomass per substrate.

We propose to validate genome-scale metabolic models using metabolic control analysis. Metabolic control analysis is a sensitivity analysis of enzyme fluxes in biochemical pathways (Kacser and Burns; Heinrich and Rapoport; Heinrich and Rapoport). It quantifies the impact of individual enzymes on overall pathways using flux control coefficients (Flint et al.; Middleton and Kacser; Torres et al.). In addition, the metabolic control analysis has also been applied to investigate the regulation and design of biochemical pathways (Groen et al.; Salter, Knowles, and Pogson; Fell and Snell).

Metabolic control analysis quantifies the responses of enzymes fluxes within metabolic networks to external perturbations. Thus, it represents a powerful tool for validating genome-scale metabolic models. Here, we designed a framework for validating genome-scale metabolic networks for biogeochemical application. This method samples across a wide range of nutrient uptake fluxes, from those in oligotrophic environments to those typical of laboratory reactors. At each nutrient uptake flux, we apply FBA to the genome-scale model of interest, using the objective of maximizing the rates of ATP production or growth. The objective function that was compared to the constraining rate. If these quantities deviated form a linear relationship, revision of the metabolic network reactions needed investigation. Reactions were modified by changing stoichiometry, reaction direction, and adding or remove compounds from the reaction equation. Each modification was tested by running FCA and verifying that the model was within the range of reported yield values. Every model we validated required a significant revision.

In some cases, the model could behave linearly, but provide a prediction that is not within the range of observed values. This case was particularly prevalent in the biosynthesis network. Microbes preform catabolic reactions to obtain energy to manage cell maintenance and for growth. These energy uses are not explicitly accounted for in the metabolic network. There is no clear standard for assigning ATP assembly cost in metabolic reconstructions. Thiele et al. (2010) who published the protocol for metabolic reconstructions, briefly touch on adding GAM and NGAM to the metabolic reconstruction. Near the final stage of the protocol, the authors recommend running a linear regression on chemostat growth experiments to obtain accurate parameters.

Alternatively, they suggest in the absence of growth data, the value can be estimated from the energy required for macromolecular synthesis. However, many authors have identified issues with these methods of estimation and have applied their own methods (Benedict et al.; Goyal et al.; Gonnerman et al.). One recent example was outlined in Goyal et al. as a method combining chemostat data and optimization of parameters using a metabolic model.

Gonnerman et al. performed a sensitivity analysis on the iMG746 biomass reaction. They examined the impact of the exact coefficients of the biomass reaction, adjusted up to 50% smaller or larger, on the predicted growth rate. Changes to the ratio of RNA, DNA, lipid, and protein in the biomass reaction had little effect on the predicted growth rate and product secretion rates. Only the growth associated maintenance, non-growth associated maintenance, proton and sodium pump stoichiometry had a significant impact on the predicted growth rate. Feist et al. performed a similar sensitivity analysis and drew a similar conclusion about the iAF1260 model of *E. coli*.

The ATP assembly cost varied over a wide range in genome-scale metabolic models. We evaluated the significance of each component of the biomass reaction for each model we validated. The ATP assembly cost is the most significant parameter of the biomass objective function. The amino acid composition also varied significantly between models. The variance is likely derived from the method of determining amino acid composition, either directly measuring the concentration of amino acids or estimating the composition from amino acid sequence used in the reconstruction of the microbe. We calculated ATP assembly cost using the energy required to synthesize the enzymes in the model, amino acid maturation, mRNA translation, and transport of phosphate and nitrate. We replaced the original estimation in the biomass objective function and recalculated the growth yield. The predicted growth yield using our assembly ATP assembly cost fell closer to the range of observed yield values than using the original estimation.

The direct use of validated genome-scale metabolic models may be expected to work well for describing microbes in laboratory experiments. Metabolic models are currently validated for laboratory environments, where energy is limited, and under only a narrow range of conditions. This new validation method for genome-sale metabolic models for application to biogeochemical reaction modeling improves FBA predictions of microbial parameters in the natural environment. Genome-scale metabolic models contain an exceptional amount detail about microbial pathways that adds information to biogeochemical modeling. However, some of the information in these models is poorly constrained and estimated. There is a need for clearer standards in the reconstruction process for ATP assembly cost, constraints of reaction direction, and non-gene associated reactions. This new method for validation builds framework using the predictive power of FBA paired with reported yield values, to improve the prediction quality of microbial parameters when simulating microbes in the natural environment.

TABLES

Table 1. Kinetic parameters of methanogenesis and enzymes of *M. barkeri.*

	Parameter			Data source
Methanogenesis	k (mmol·g^{-1}·hr^{-1})	K_S (mM)	χ	
Acetoclastic	7.06	5.0	2	(Smith and Mah)
Methanol	10.85	0.5	2	(Ranalli, Whitmore, and Lloyd)
Enzyme	V_{max} (mmol·g^{-1}·hr^{-1})	K_M (mM)		
Acetate transporter	2.165	0.333		(Button)
Methanol methyltransferase	1332	50		(Sauer, Harms, and Thauer)

Table 2. Genome-scale metabolic models, numbers of metabolites and reactions, number of reactions in catabolic and biosynthesis models, and electron donors and accepters tested using FBA.

Organisms	Model	Ref.	Met,Rxn	C	X	D	A
M. acetivorans	iMB745	8	715,825	111	518	Acetate, Methanol	–
M. barkeri	iMG746	7	718,816	106	501	H2, Ac, Methanol	–
G. metallireducens	iAF987	4	1109,1284	111	863	Acetate	Fe^{+++}
S. oneidensis MR1	iMR1_799	6	744,933	69	624	Lactate, H_2	Fe^{+++}
S. putrefaciens	iW318_789	6	739,918	71	620	Lactate, H_2	Fe^{+++}
S. sp. MR4	iMR4_812	6	755,986	62	630	Lactate, H_2	Fe^{+++}

C, catabolic network; X, biosynthesis network; D, electron donors; A, electron acceptors.

Table 3. Carbon sources for biomass yield predictions using FBA.

Organism	CO_2*	Acetate	Lactate	Methanol
M. acetivorans	N.A.	Y	N.A.	Y
M. barkeri	Y	Y	N.A	Y
G. metallireducens	N.A.	Y	N.A	N.T.
S. oneidensis MR1	Y	Y	Y	N.T.
S. putrefaciens	Y	Y	Y	N.T.
S. sp. MR4	Y	Y	Y	N.T.

*, electron donor is H_2; N.A., not applicable; N.T., not tested.

Table 4. ATP yield of original catabolic network models (mol ATP·mol D^{-1}) and observations.

Organism	H_2	Acetate	Lactate	Methanol
M. acetivorans	–	0.75/0.5[a]	–	0.69/(0~1.0)[a]
M. barkeri	0.187/0.25[a]	0.75/(0~1.0)[a]	–	0.81/(0~1.0)[a]
G. metallireducens	–	0.9/(1.0~1.5)[a]	–	–
S. oneidensis MR1	0[c]/0.5[a]	–	1.25/2.5[a,b]	–
S. putrefaciens	Inf [d]/0.5[a]	–	1.25/2.5 [a,b]	–
S. sp. MR4	0[c]/0.5[a]	–	1.25/2.5 [a,b]	–

[a] (Jin) reported ATP yield values for methanogens and iron respirers. *Mehtanosarcina* produces 1 ATP for every four H_2, *M. acetivorans* gains 0.5 ATP for every acetate degraded, and *M. mazei* yields between 0~1 ATP for every acetate, assumed to be consistent with *M. barkeri*. For iron respirers: four H_2 yield 2 ATP, one acetate yields between 1~1.5 ATP, two lactate yield 3 ATP. ATP yield by oxidation of methanol was assumed to be consistent with acetate ATP yield. [b](Marsili et al.) reported that 1.5 ATP are produced by the electron transport chain and an additional ATP is produced oxidizing lactate to acetate. [c]FBA did not predict synthesis of ATP, but did return a flux distribution. [d]No feasible solution (Inf.) within the metabolic network.

Table 5. Growth yield of original biosynthesis network models (g·mol ATP^{-1})

Organism	CO_2	Acetate	Lactate	Methanol
M. acetivorans	–	Inf[b]/6.2[a]	–	Inf[b]/6.2[a]
M. barkeri	10.61/6.2[a]	2179/6.2[a]	–	9.826/6.2[a]
G. metallireducens	–	59.00/5.3[a]	–	–
S. oneidensis MR1	31.27/5.3[a]	13.4/5.3[a]	31.62/5.3[a]	–
S. putrefaciens	25.70/5.3[a]	13.4/5.3[a]	26.25/5.3[a]	–
S. sp. MR4	25.60/5.3[a]	13.4/5.3[a]	26.14/5.3[a]	–

[a](Jin); [b]No feasible solution (Infeasible, Inf) within the metabolic network.

Table 6. Survey of GAM estimated values use in metabolic models.

Organism	Substrate	GAM(mmol ATP·g^{-1})	NGAM (mmol ATP·g^{-1})	Ref.
E. coli K-12	glucose	59.81	8.39	A
Lactococcus lactis	mannose, galactose, sucrose, lactose	18.15	1	B
M. bakeri	acetate, methanol, H_2/CO_2	65	2	C
M. acetivorans	CO, methanol	65	2.5	D
Methanococcus maripaludis	H_2/CO_2	27.14	7.836	E
Geobacter metallireducens	acetate	79.20	0.81-1000	
Shewanella oneidensis	lactate	220.22	1.03	F
	lactate	0	0-1000	G
Shewanella sp. MR-4	lactate	0	0-1000	G
Shewanella putrefaciens	lactate	0	0-1000	G

A, (Feist et al.) ; B, (Oliveira, Nielsen, and Förster); C, (Gonnerman et al.); D, (Benedict et al.); E, (Goyal et al.); F, (Pinchuk et al.); G, (Ong et al.) removed GAM from the biosynthesis equation in their published models.

Table 7. ATP (mmol ATP·g-1) required for monomer biosynthesis and the cost of biomass assembly.

Organism	Monomer synthesis	Maturation[c]	mRNA[d]	Transport[d]	BAC
E. coli K-12	61.2	28.9	1.39	5.2	35.5
M. acetivorans Δ*MSS*[1]	51.4	28.4	1.39	5.2	35.0
M. barkeri[1]	52.5	28.9	1.39	5.2	35.5
G. metallireducens[1]	45.1	25.8	1.39	5.2	32.4
S. oneidensis MR1[2]	62.8	24.3	1.39	5.2	30.9
S. putrefaciens[2]	62.7	24.3	1.39	5.2	30.9
S. sp. MR4[2]	52.4	24.3	1.39	5.2	30.9

[a] Acetate; [b] Lactate; Δ,Removed reaction from model; [c]Estimated following method of (Arnold and Nikoloski); [d]Additional energy requirements from (Stouthamer).

Table 8. Comparison of ATP required for synthesis of amino acids in iMB745, iMG746, and iAF1260.

Compound name	iMB745 (mmol ATP·g AA^{-1})	iMG746 (mmol ATP·g AA^{-1})	iAF1260 mmol ATP·g AA^{-1})
Alanine	1.11	1.67	1.98
Arginine	2.15	3.11	3.87
Asparagine	0.91	1.43	2.32
Aspartic acid	0.52	0.91	2.09
Cysteine	0.03	0.05	0.02
Glutamic acid	0.64	1.35	2.15
Glutamine	0.50	1.07	1.89
Glycine	3.48	3.81	5.33
Histidine	0.84	1.43	1.62
Isoleucine	3.61	3.45	4.85
Leucine	2.07	2.80	4.46
Lysine	2.78	3.15	5.28
Methionine	1.20	1.20	1.71
Phenylalanine	3.10	3.00	4.49
Proline	1.26	1.37	2.31
Serine	0.00	0.76	1.29
Threonine	1.78	1.92	2.63
Tryptophan	0.93	1.11	1.65
Tyrosine	2.05	2.05	3.16
Valine	2.06	2.51	3.75

Table 9. Maximum flux of loop reactions constrained by flux direction.

Reaction	F/F	R/R	F/R	R/F
Leucine	n.s.	6.75E-04	1000	1000
Isoleucine	n.s.	4.39E-04	0.1048	-4.39E-04
Valanine	n.s.	5.18E-04	5.18E-04	1000

Table 10. ATP yield of validated catabolic network models (mol ATP·mol S^{-1}) and observations.

Organism	H_2	Acetate	Lactate	Methanol
M. acetivorans	–	0.5/0.5[a]	–	0.625/(0~1)[a]
M. barkeri	0.25/0.25[a]	1.0/(0~1)[a]	–	0.75/(0~1)[a]
G. metallireducens	–	1.0/(1~1.5)[a]	–	–
S. oneidensis MR1	0.5/0.5[a]	–	2.5/2.5[a,b]	–
S. putrefaciens	0.5/0.5[a]	–	2.5/2.5[a,b]	–
S. sp. MR4	0.5/0.5[a]	–	2.5/2.5[a,b]	–

[a](Jin); [b](Marsili et al.)

Table 11. Growth yield (with 50% efficiency) of validated biosynthesis network models (g·mol ATP^{-1}) with BAC estimations from this study.

Organism	CO_2	Acetate	Lactate	Methanol
M. acetivorans	–	6.64/6.2[a]	–	7.68/6.2[a]
M. barkeri	4.66/6.2[a]	5.85/6.2[a]	–	6.12/6.2[a]
G. metallireducens	–	6.45/5.3[a]	–	–
S. oneidensis MR1	2.60/5.3[a]	4.71/5.3[a]	5.34/5.3[a]	–
S. putrefaciens	2.60/5.3[a]	4.72/5.3[a]	5.35/5.3[a]	–
S. sp. MR4	2.59/5.3[1]	4.71/5.3[a]	6.01/5.3[a]	–

[a](Jin)

Table 12. Growth yield of validated biosynthesis network models (g·mol ATP^{-1}).

Organism	CO_2	Acetate	Lactate	Methanol
M. acetivorans	–	9.70/6.2[a]	–	7.99/6.2[a]
M. barkeri	7.01/6.2[a]	8.16/6.2[a]	–	8.57/6.2[a]
G. metallireducens	–	8.22/5.3[a]	–	–
S. oneidensis MR1	2.63/5.3[a]	3.38/5.3[a]	3.97/5.3[a]	–
S. putrefaciens	2.62/5.3[a]	3.39/5.3[a]	3.86/5.3[a]	–
S. sp. MR4	2.62/5.3[a]	3.38/5.3[a]	3.86/5.3[a]	–

[a](Jin)

Table 13. Elemental biomass composition from biosynthesis network models.

Organism	C	H	N	O	P	S
E. coli K-12	4.2	7.4	1.0	2.0	6.5E-02	2.3E-02
G. metallireducens	4.4	8.9	1.0	1.6	8.1E-02	2.5E-02
M. acetivorans	3.8	7.2	1.0	2.0	6.4E-02	2.2E-02
M. barkeri	3.5	5.4	1.0	1.3	7.8E-02	2.7E-02
S. oneidensis MR1	4.7	8.9	1.0	1.8	6.7E-02	2.0E-02
S. putrefaciens	4.7	9.5	1.0	1.7	6.7E-02	2.0E-02
S. sp. MR4	4.7	9.5	1.0	1.8	6.7E-02	2.0E-02

FIGURES

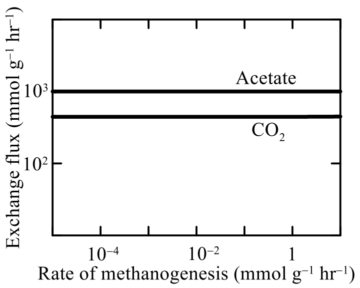

Figure 1. Variations in exchange fluxes of acetate and CO_2 with rates of methanogenesis, predicted by the genome-scale metabolic model of (Gonnerman et al.) and by setting the constraint of sulfite to -10^3 to $+10^3$ $mmol \cdot g^{-1} \cdot hr^{-1}$.

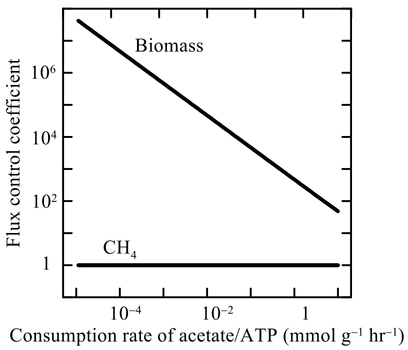

Figure 2. Variations in the flux control coefficients of methane production with acetate consumption, and in the flux control coefficient of biomass synthesis with ATP consumption by *M. barkeri*.

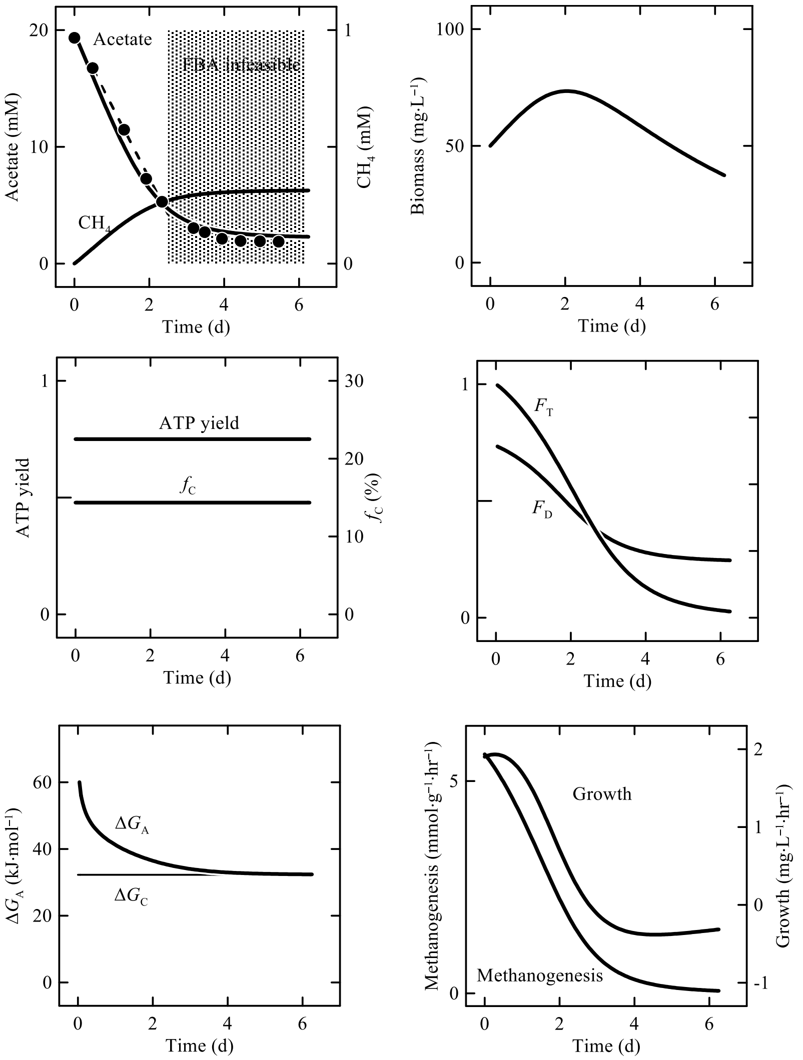

Figure 3. (a)Variations in the concentrations of acetate and methane, and (b) biomass during acetoclastic methanogenesis by *Methanosarcina barkeri*. Experimental data points are from (Fukuzaki, Nishio, and Nagai); solid lines are the simulation results using hybrid method; dashed lines are the simulation results using dynamic FBA (dFBA); shaded region (a) demarks where FBA solution is infeasible using dFBA. (c) ATP yield of the acetoclastic methanogenesis network and fraction of acetate utilized by the biosynthesis network. (d) kinetic factor of acetate and thermodynamic factor for acetoclastic methanogenesis. (e) Amount of energy available and amount of energy saved by *M. barkeri*. (f) Specific growth rate and rate of methanogenesis.

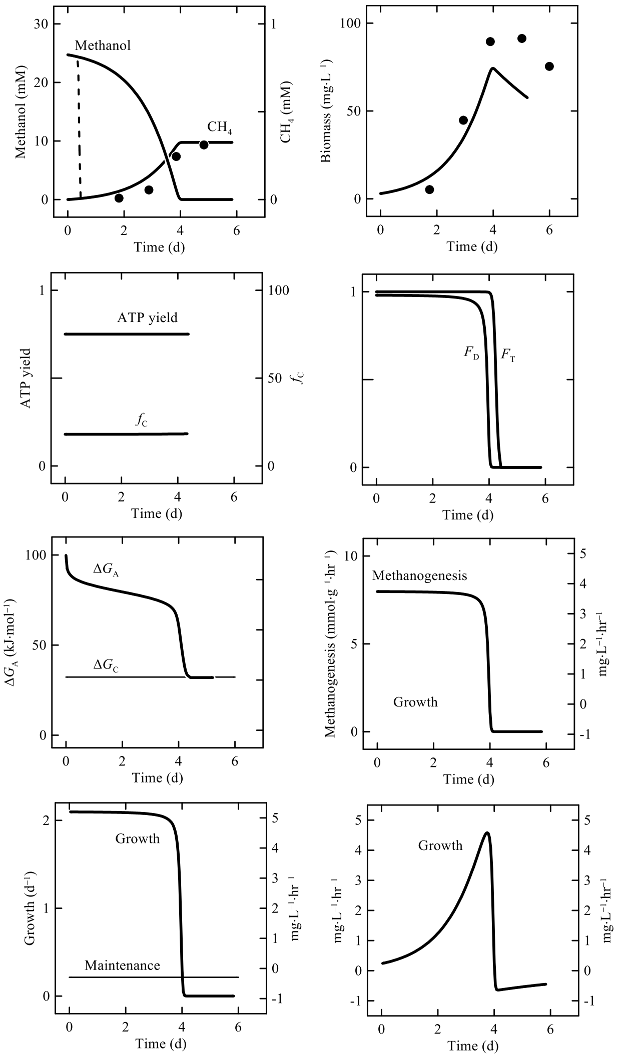

Figure 4. Variations in the concentrations of methanol, methane, and biomass during methanol methanogenesis by *M. barkeri*. Experimental data points are from (Smith and Mah); solid lines are the simulation results using the hybrid method; dashed lines are the simulation results using dynamic FBA (dFBA). (c) ATP yield of the methanol methanogenesis network and fraction of methanol utilized by the biosynthesis network. (d) kinetic factor of methanol and thermodynamic factor for methanol methanogenesis. (e) Amount of energy available and amount of energy saved by *M. barkeri*. (f) Rate of methanogenesis and (g) FBA-predicted growth rate with specific maintenance rate. (h) Net growth rate.

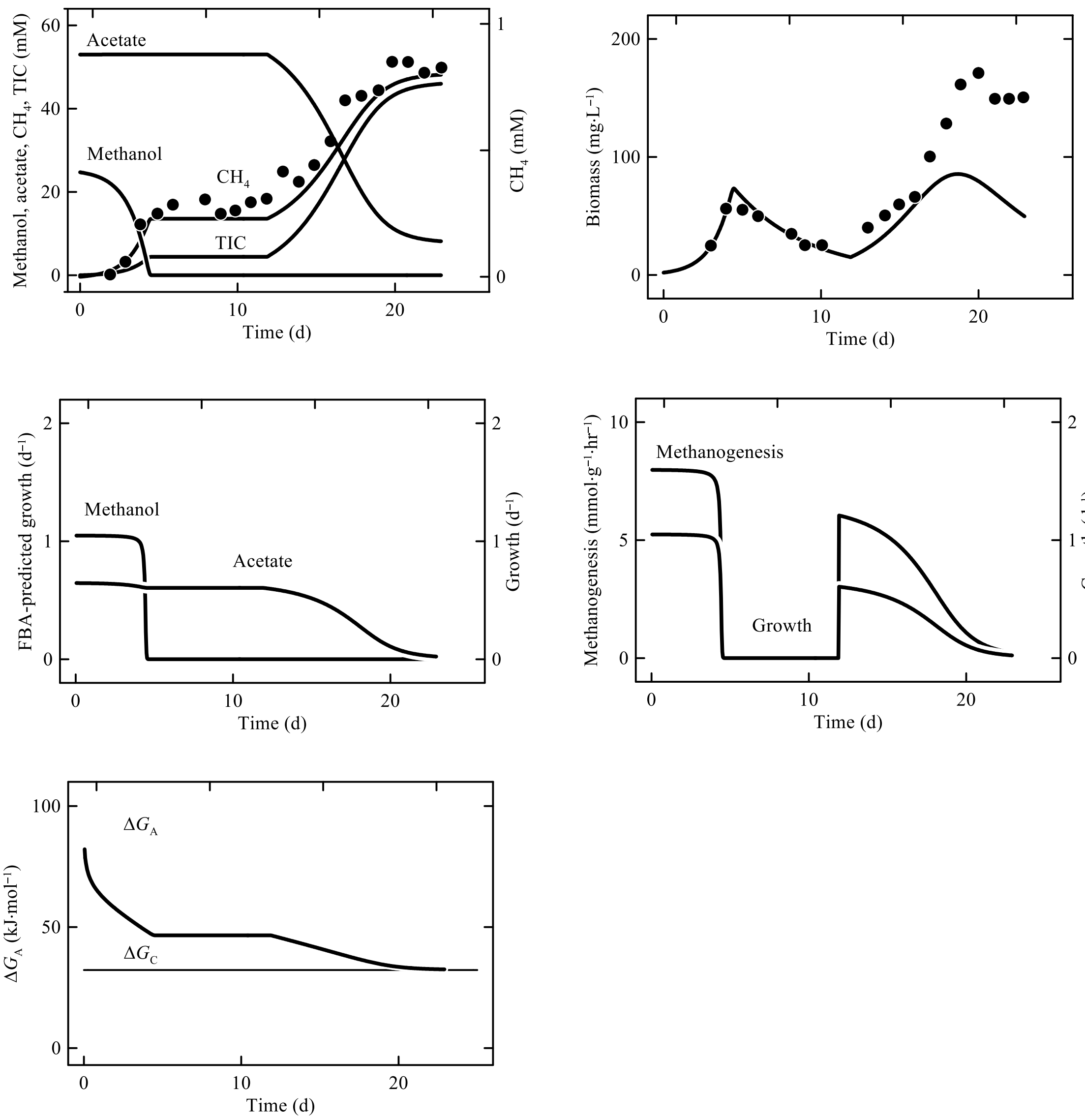

Figure 5. (a) Variations in the concentrations of acetate, methanol, dissolved inorganic carbon, methane, and (b) biomass during diauxic methanogenesis by *Methanosarcina barkeri* using the hybrid method. Experimental data points are from (Smith and Mah). (c) FBA-predicted growth rate for acetoclastic and methanol methanogenesis. (d) Specific growth rate and rate of methanogenesis. (e) Amount of energy available and amount of energy saved by *M. barkeri*.

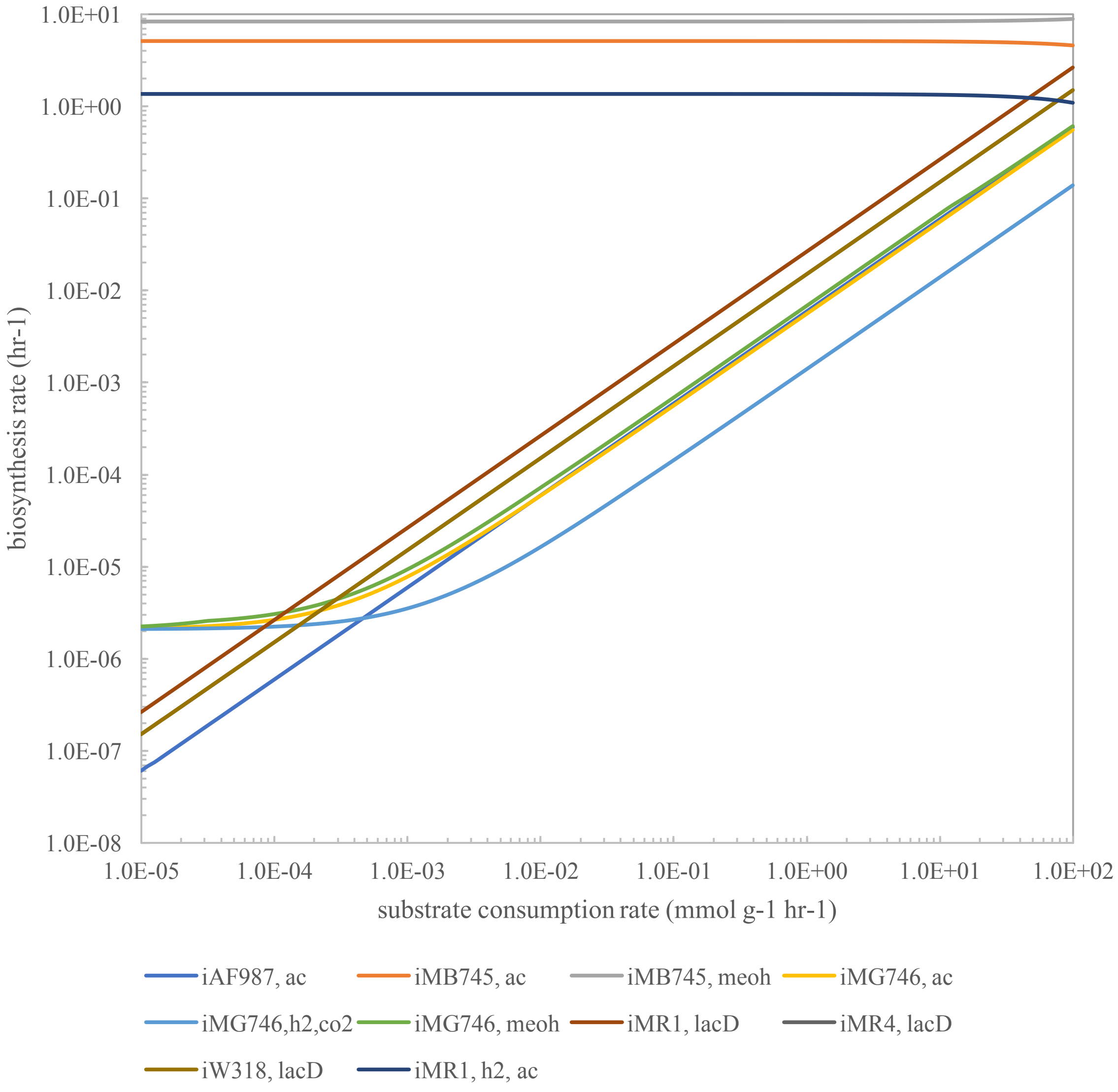

Figure 6. Predicted biosynthesis rates from a range of substrate consumption rates (H_2/CO_2, acetate, lactate, methanol) using published models.

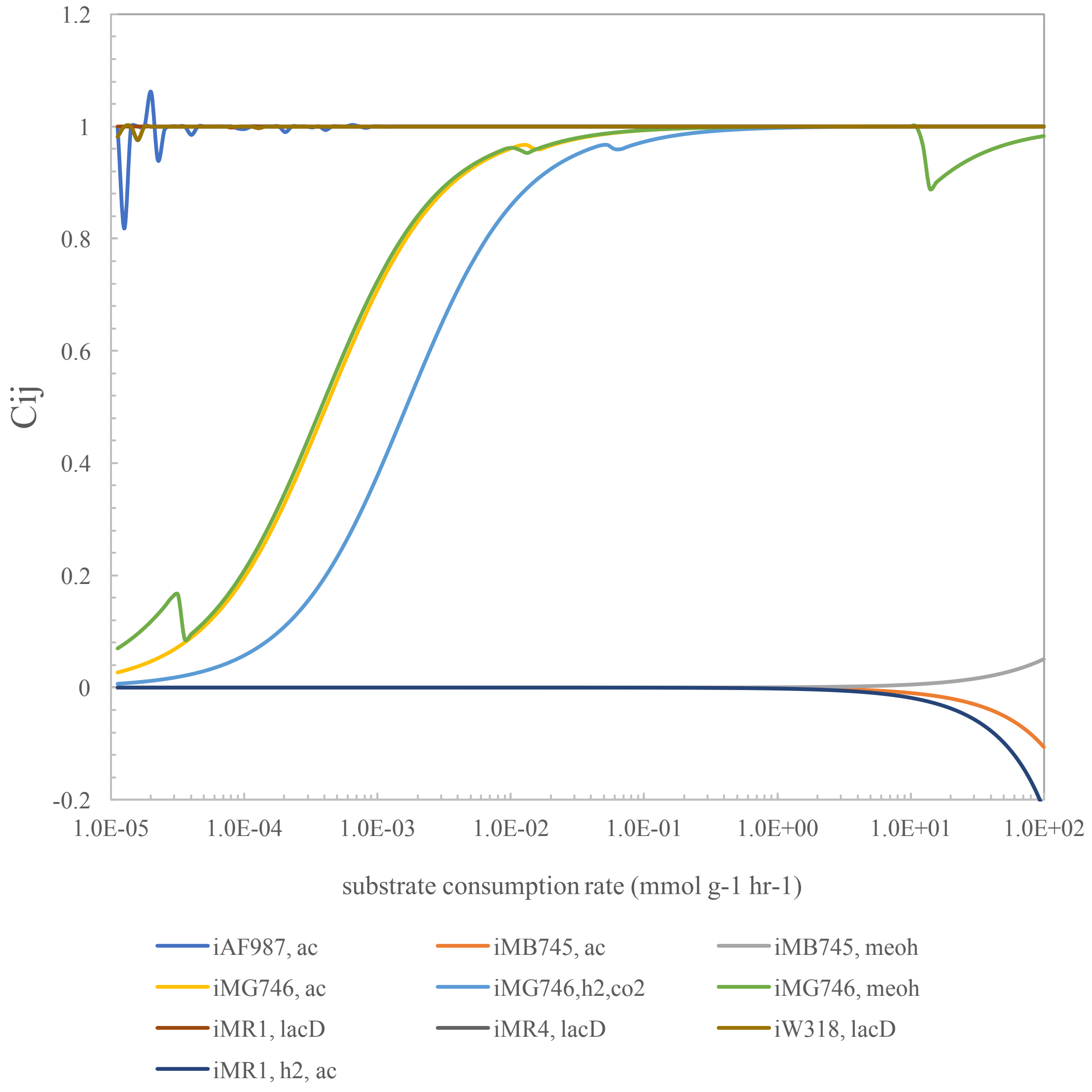

Figure 7. Variations in the flux control coefficients of biosynthesis rate with substrate consumption rates (H_2/CO_2, acetate, lactate, methanol).

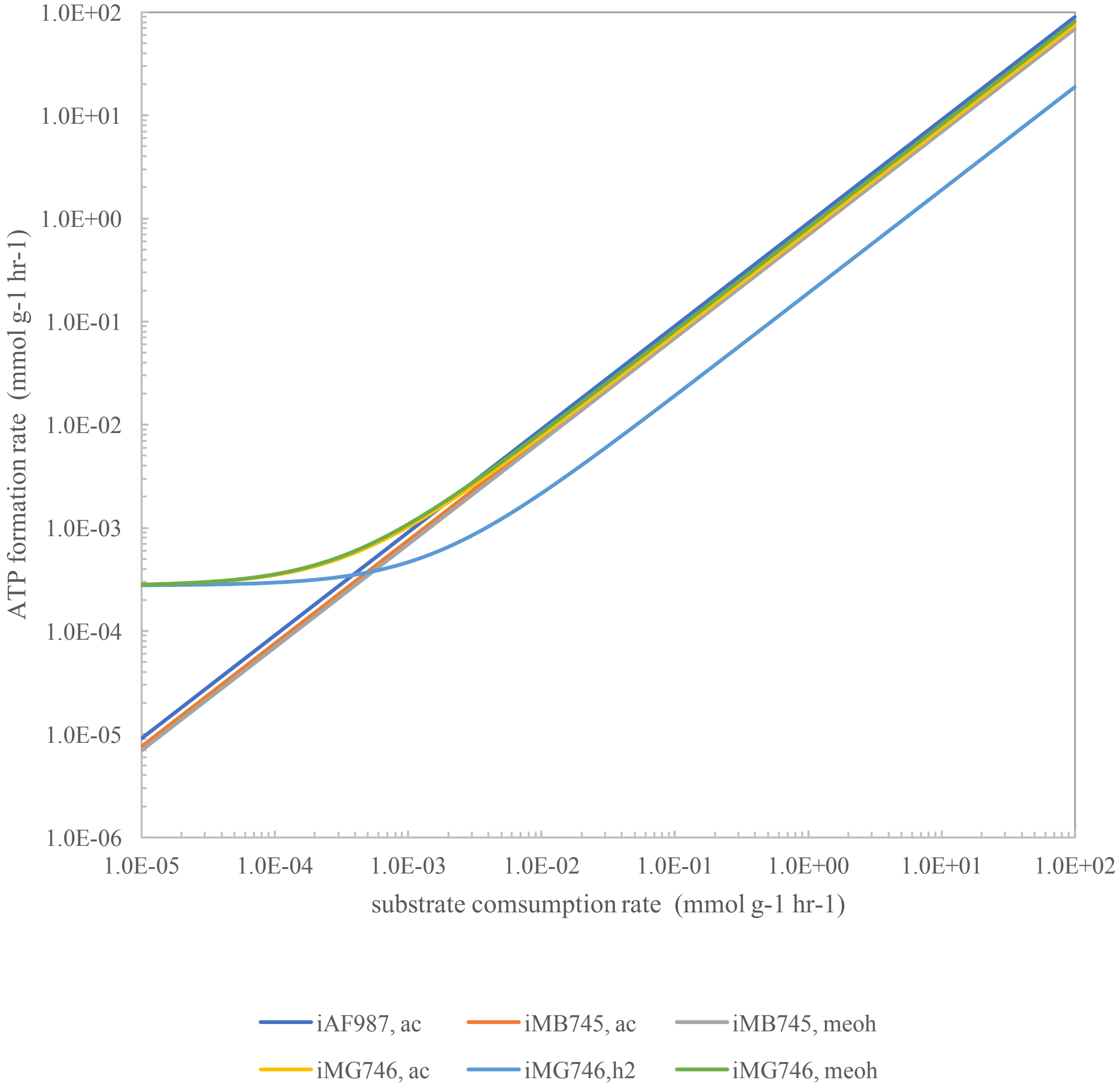

Figure 8. Predicted ATP formation rates from consumption of electron donors (H_2, acetate, lactate, methanol) from the catabolic network.

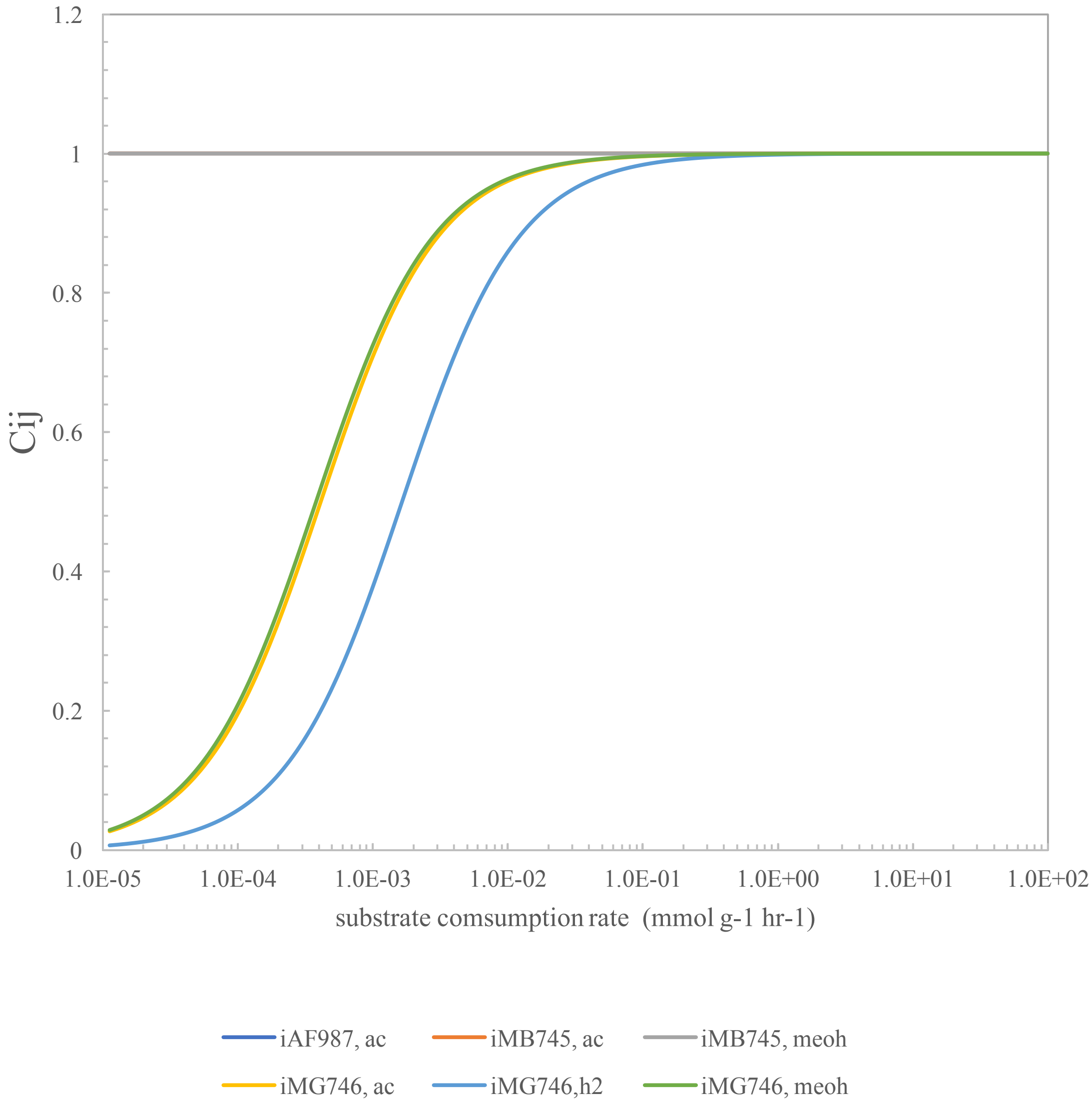

Figure 9. Variations in the flux control coefficients of ATP formation flux with electron donor uptake flux (H_2, acetate, lactate, methanol).

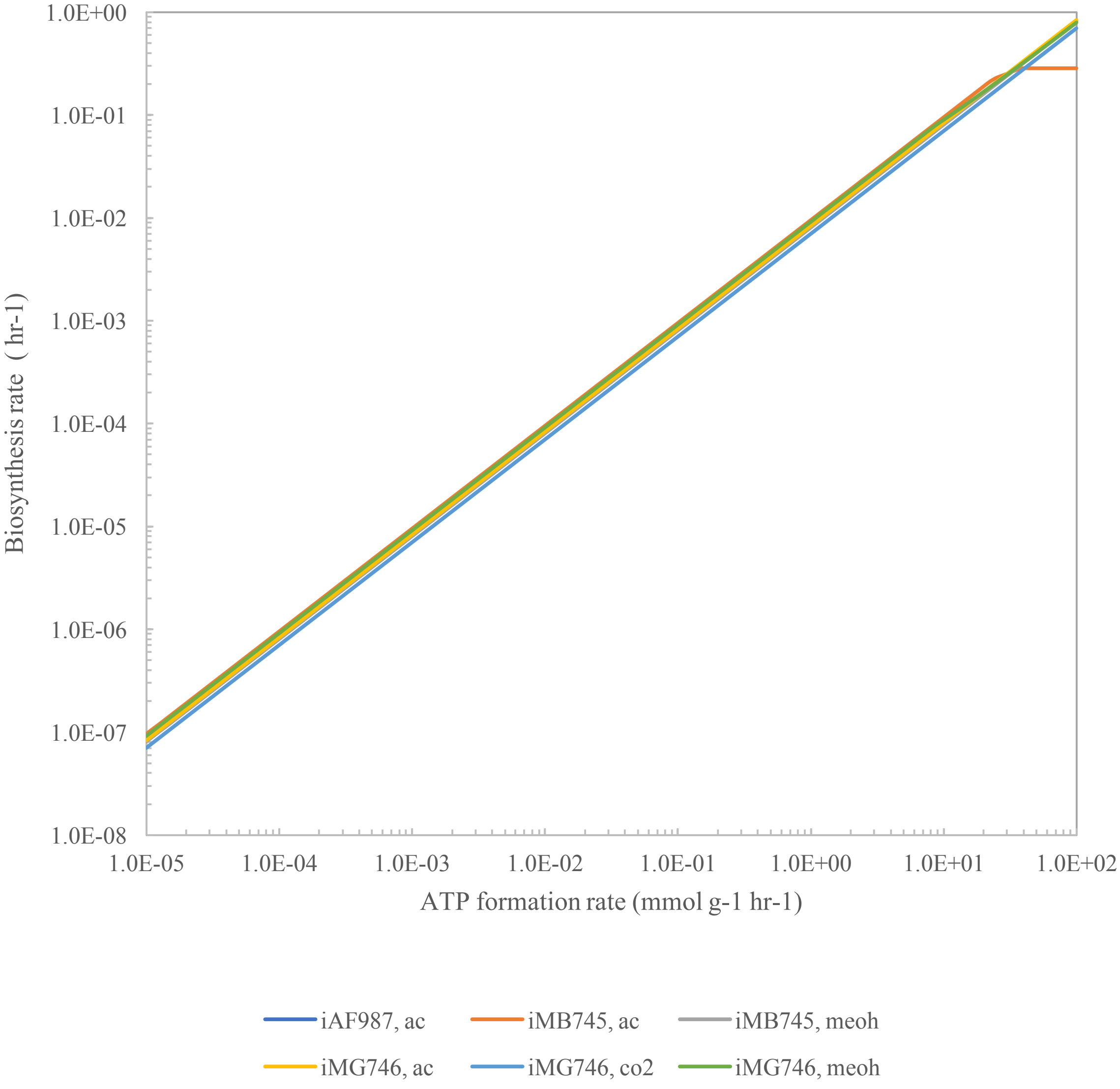

Figure 10. Predicted ATP formation rates from carbon source consumption rates in the biosynthesis network on different carbon sources (CO_2, acetate, lactate, methanol).

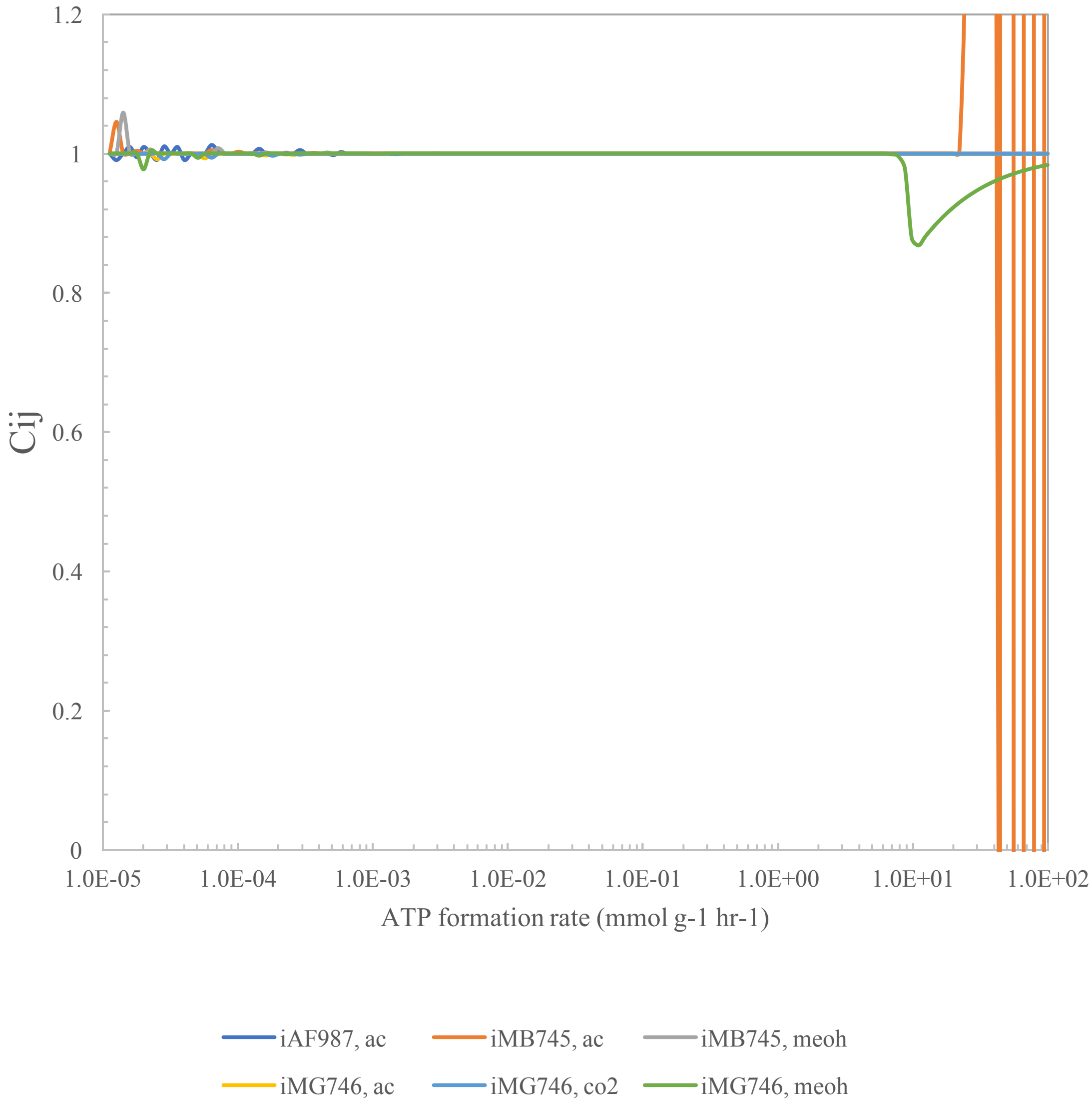

Figure 11. Variations in the flux control coefficients of biosynthesis with ATP flux on different carbon sources (CO_2, acetate, lactate, methanol).

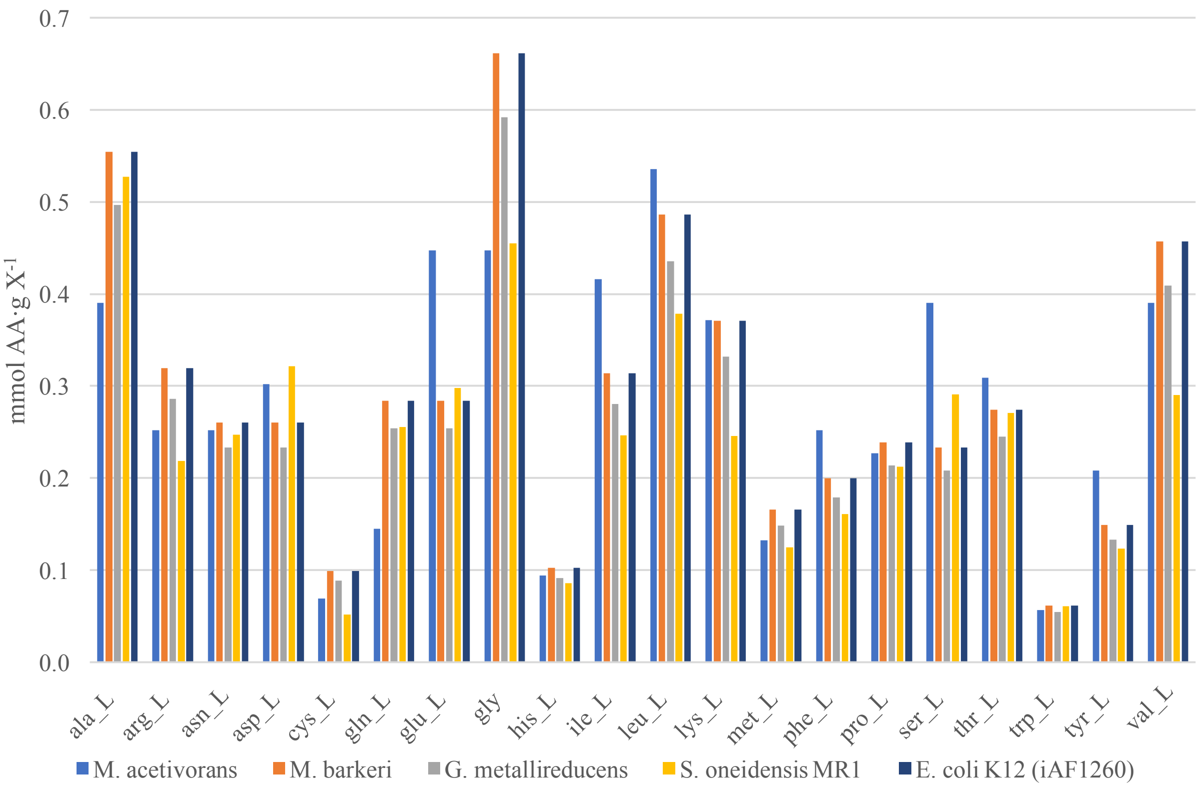

Figure 12. Amino acid abundance (mmol AA·g X^{-1}) in biomass objective function in various models including E. coli. All *Shewanella* models use the same amino acid composition, so *S. oneidensis* MR1 is only shown here.

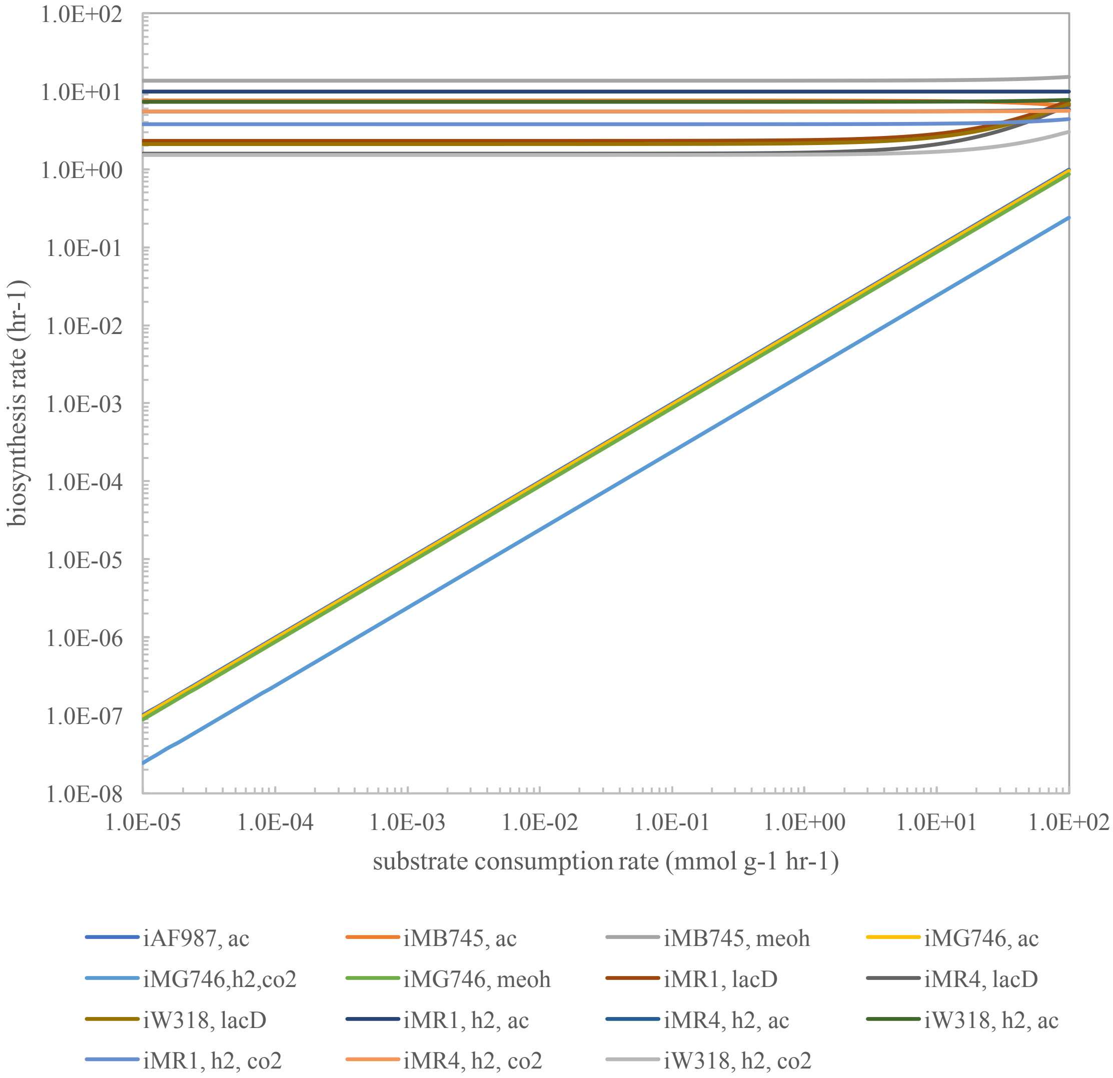

Figure 13. Predicted biosynthesis rates from a range of substrate consumption rates (H_2/CO_2, acetate, lactate, methanol) with modifications to the combined catabolic and biosynthesis network (genome-scale network).

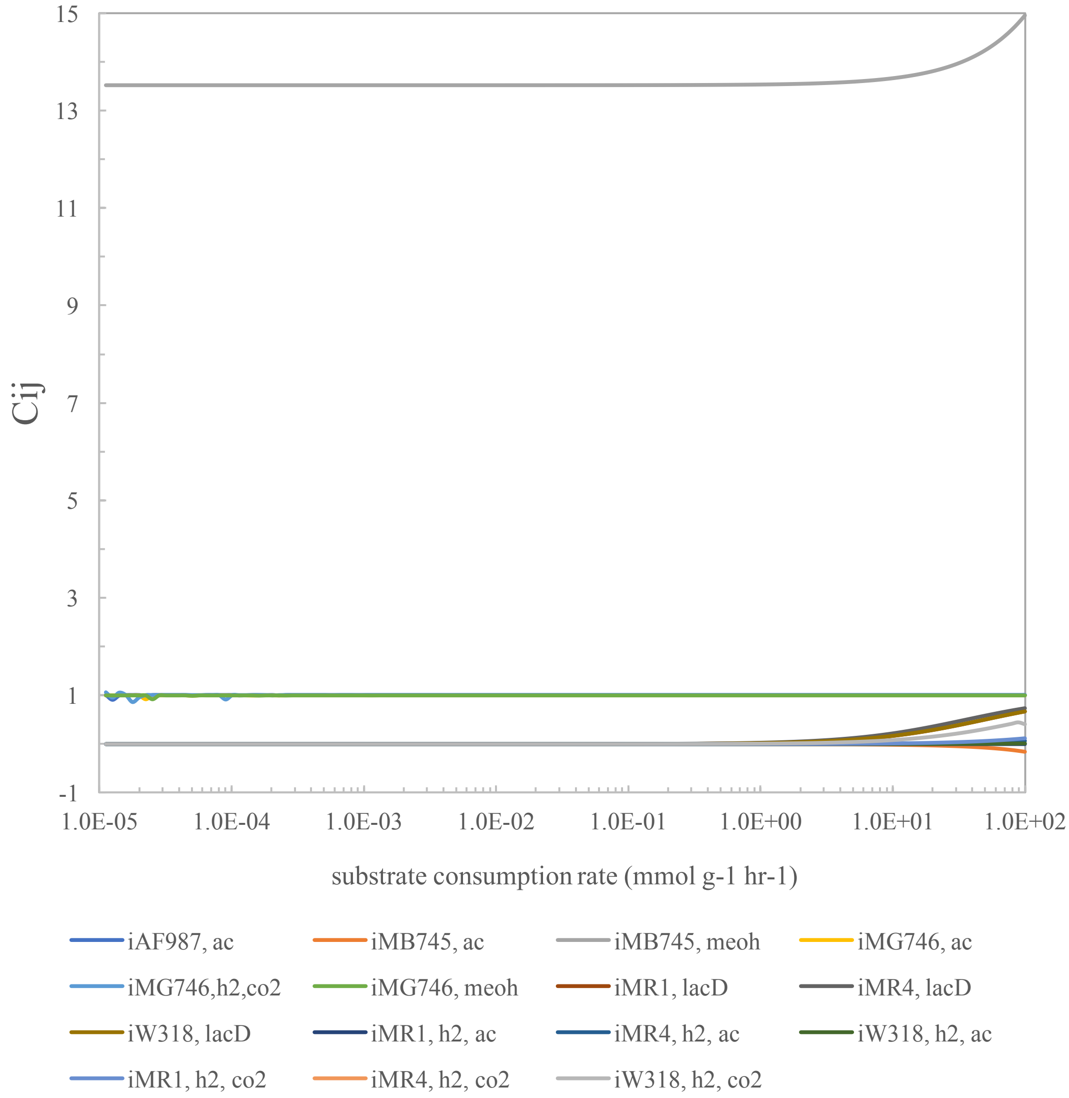

Figure 14. Variations in the flux control coefficients of biosynthesis rate with substrate consumption rates (H_2/CO_2, acetate, lactate, methanol) with corrections to the genome-scale network.

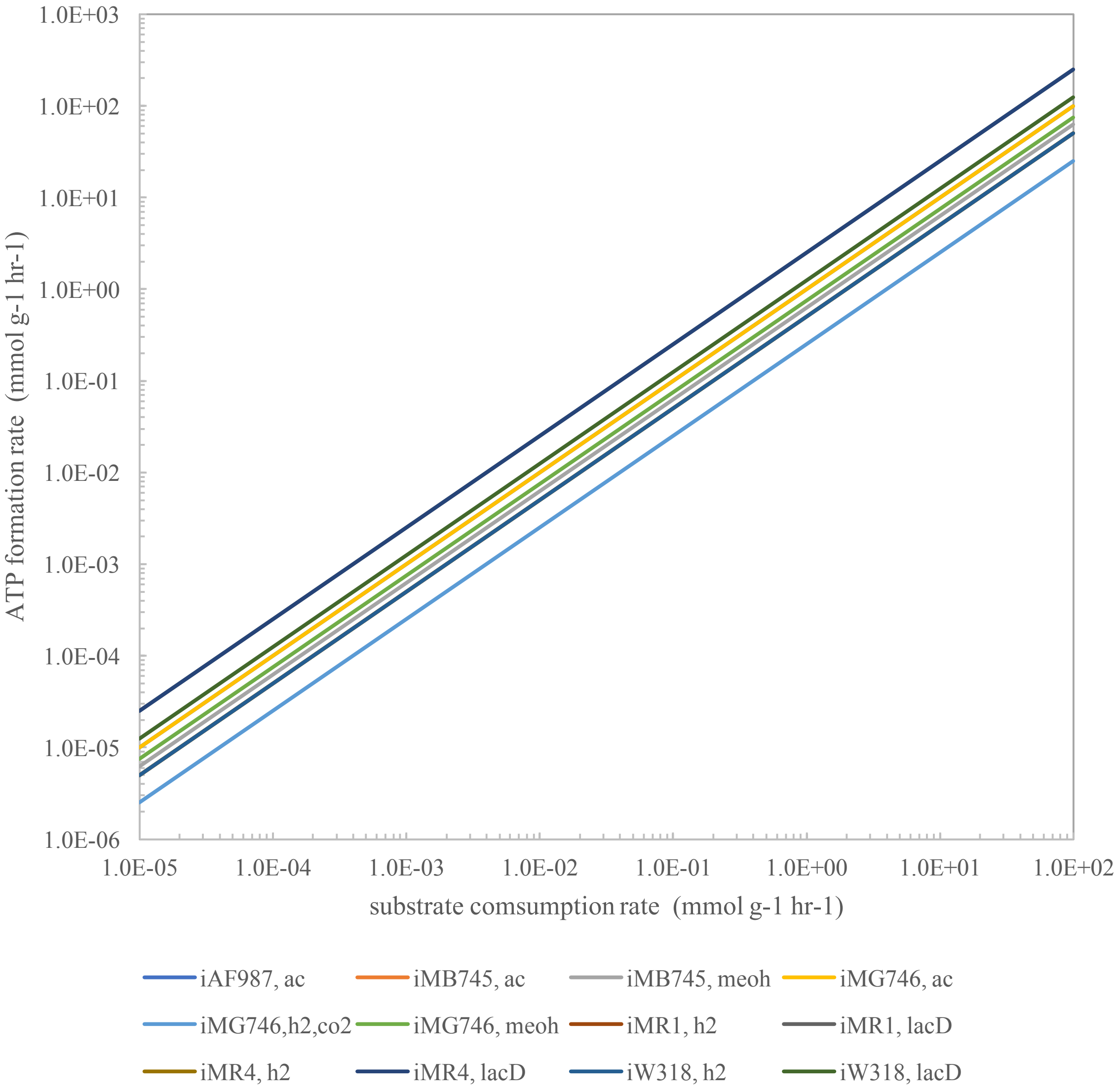

Figure 15. Predicted ATP formation rates from a range of electron donor consumption rates (H_2, acetate, lactate, methanol) from the modified catabolic network.

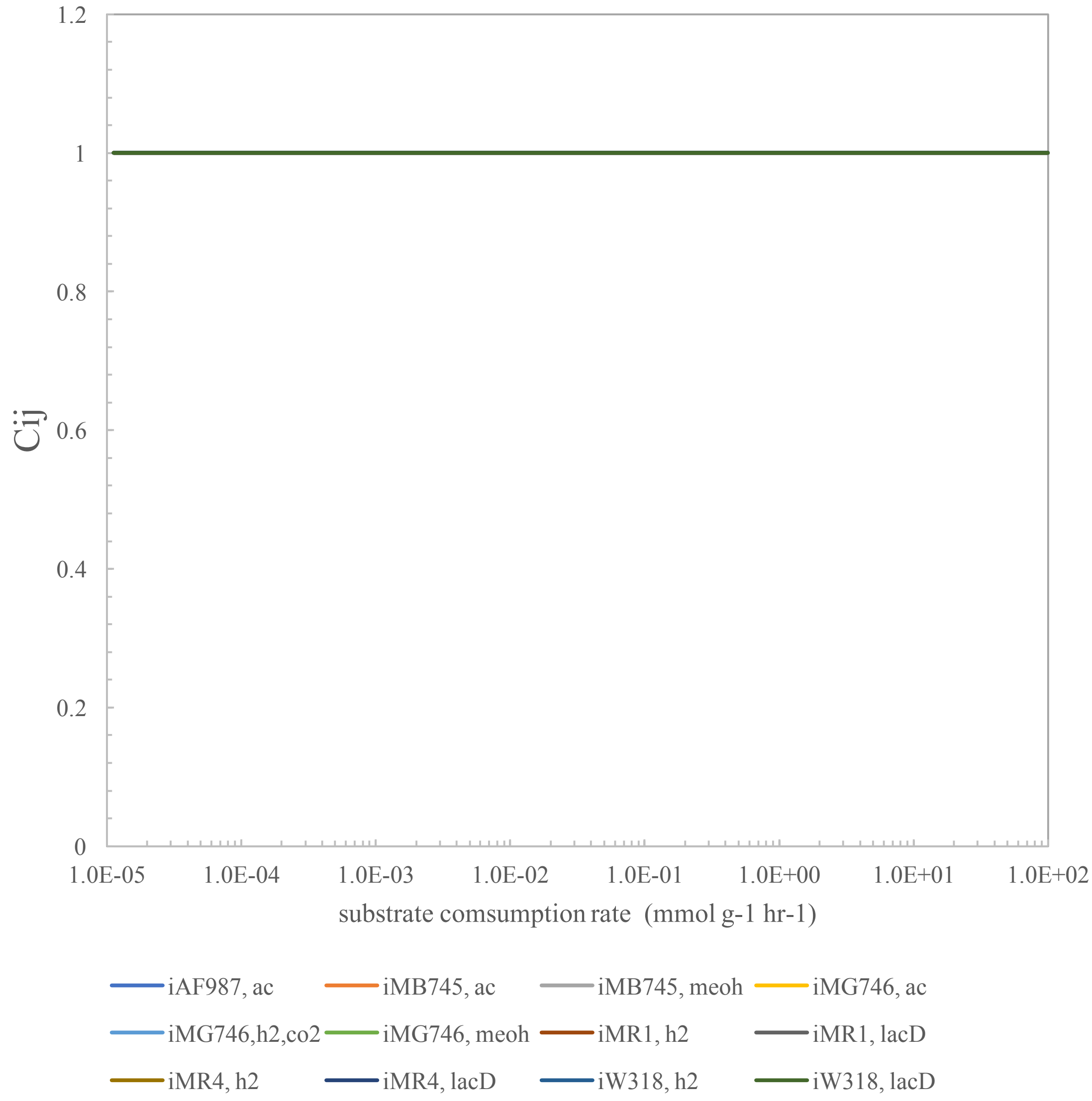

Figure 16. Variations in the flux control coefficients of ATP flux with consumption of electron donors (H_2, acetate, lactate, methanol) in the corrected catabolic network.

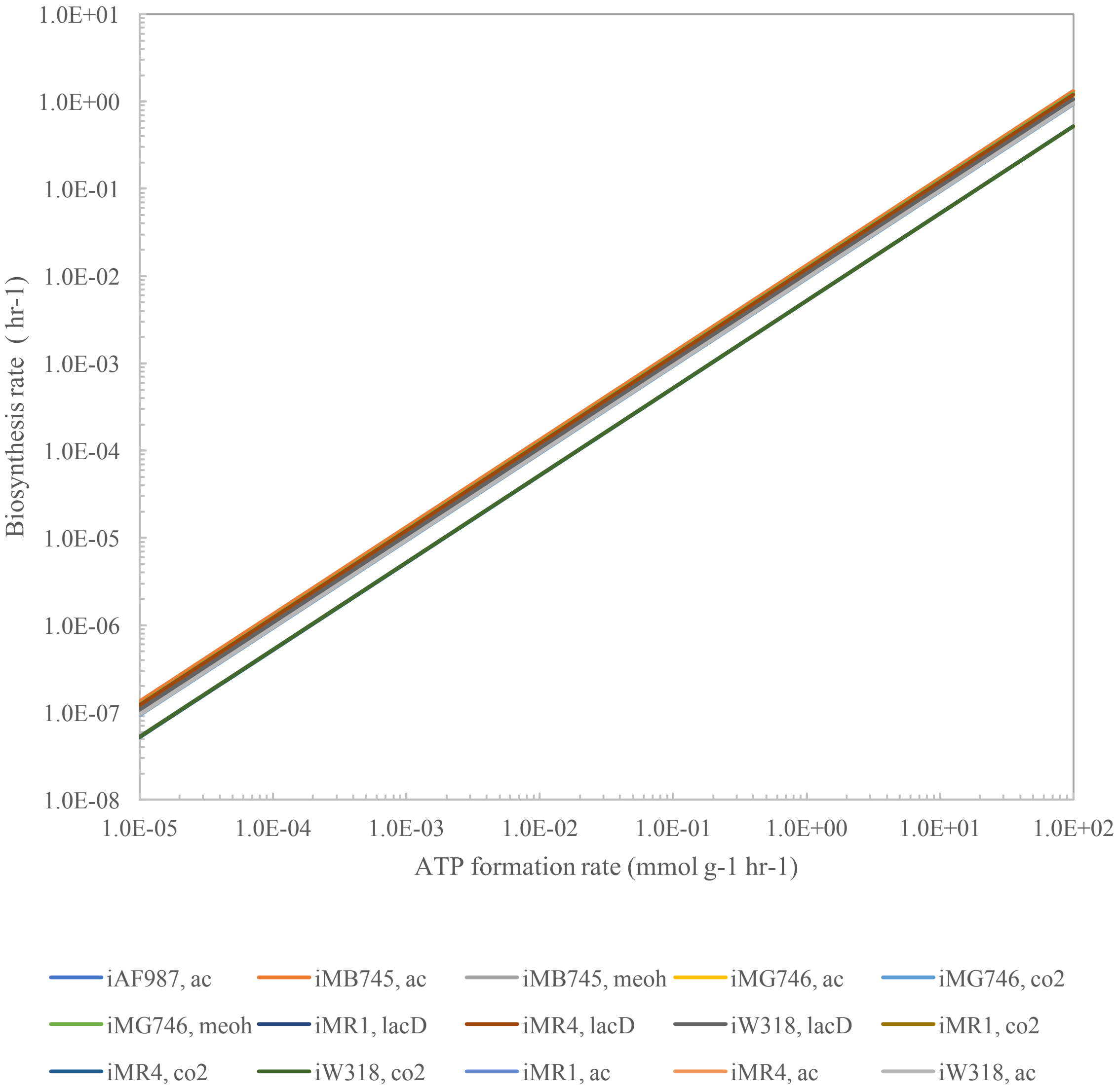

Figure 17. Predicted biosynthesis rates from a range of ATP flux rates from the corrected biosynthesis network on different carbon sources (CO_2, acetate, lactate, methanol).

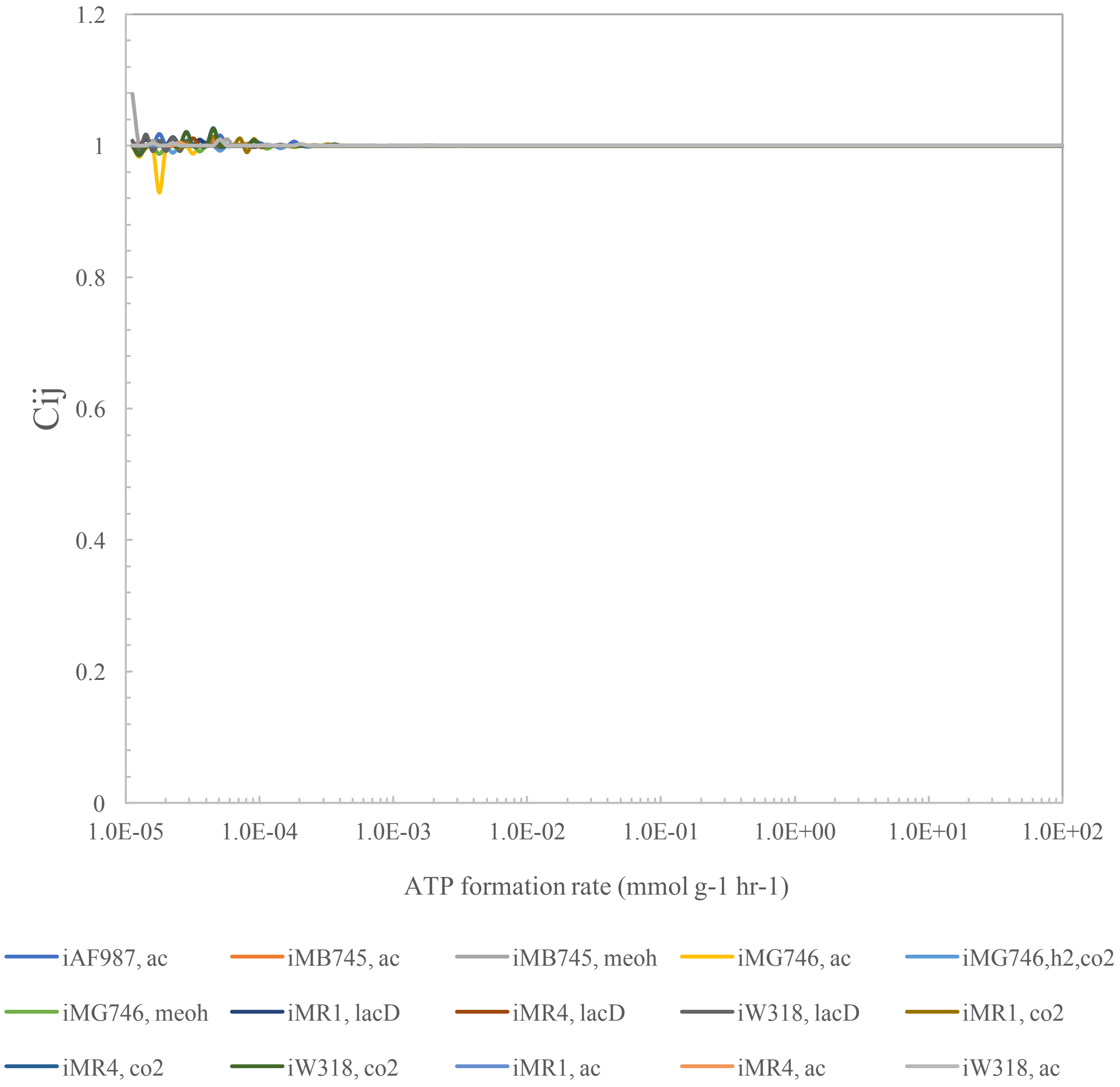

Figure 18. Variations in the flux control coefficients of biosynthesis with ATP flux on different carbon sources (CO_2, acetate, lactate, methanol) with corrected biosynthesis network.

REFERENCES CITED

Arnold, Anne, and Zoran Nikoloski. “Bottom-Up Metabolic Reconstruction of Arabidopsis and Its Application to Determining the Metabolic Costs of Enzyme Production..” *Plant physiology* 165.3 (2014): 1380–1391. Web.

Benedict, M N et al. “Genome-Scale Metabolic Reconstruction and Hypothesis Testing in the Methanogenic Archaeon Methanosarcina Acetivorans C2A.” *Journal of Bacteriology* 194.4 (2012): 855–865. Web.

Bethke, C M. *Geochemical and Biogeochemical Reaction Modeling*. 2nd ed. Cambridge: Cambridge University Press, 2007. Web.

Bethke, C M et al. “The Thermodynamic Ladder in Geomicrobiology.” *American Journal of Science* 311.3 (2011): 183–210. Web.

Bordbar, Aarash et al. “Constraint-Based Models Predict Metabolic and Associated Cellular Functions..” *Nature reviews. Genetics* 15.2 (2014): 107–120. Web.

Butler, Jessica E, Nelson D Young, and Derek R Lovley. “Evolution of Electron Transfer Out of the Cell: Comparative Genomics of Six Geobacter Genomes..” *BMC Genomics* 11.1 (2010): 40. Web.

Button, D K. “Nutrient Uptake by Microorganisms According to Kinetic Parameters From Theory as Related to Cytoarchitecture..” *Microbiology and Molecular Biology Reviews* 62.3 (1998): 636–645. Print.

Cao, Zexing, and Michael B Hall. “Modeling the Active Sites in Metalloenzymes. 3. Density Functional Calculations on Models for [Fe]-Hydrogenase:? Structures and Vibrational Frequencies of the Observed Redox Forms and the Reaction Mechanism at the Diiron Active Center.” *Journal of the American Chemical Society* 123.16 (2001): 3734–3742. Web.

Fang, Yilin et al. “Direct Coupling of a Genome-Scale Microbial in Silico Model and a Groundwater Reactive Transport Model.” *Journal of Contaminant Hydrology* 122.1-4 (2011): 96–103. Web.

Feist, Adam M et al. “A Genome-Scale Metabolic Reconstruction for Escherichia Coli K-12 MG1655 That Accounts for 1260 ORFs and Thermodynamic Information.” *Molecular Systems Biology* 3 (2007): n. pag. Web.

Fell, D A, and K Snell. “Control Analysis of Mammalian Serine Biosynthesis. Feedback Inhibition on the Final Step..” *The Biochemical journal* 256.1 (1988): 97–101. Print.

Flint, H J et al. “Control of the Flux in the Arginine Pathway of Neurospora Crassa.” *The Biochemical journal* 200.2 (1981): 231–246. Print.

Fukuzaki, S, N Nishio, and S Nagai. “Kinetics of the Methanogenic Fermentation of Acetate..” *Applied and Environmental Microbiology* 56.10 (1990): 3158–3163. Print.

Gonnerman, Matthew C et al. “Genomically and Biochemically Accurate Metabolic Reconstruction of Methanosarcina Barkeri Fusaro, iMG746..” *Biotechnology journal* 8.9 (2013): 1070–1079. Web.

Goyal, Nishu et al. “A Genome-Scale Metabolic Model of Methanococcus Maripaludis S2 for CO2 Capture and Conversion to Methane..” *Molecular BioSystems* 10.5 (2014): 1043–1054. Web.

Groen, Albert K et al. “Intracellular Compartment Ation and Control of Alanine Metabolism in Rat Liver Parenchymal Cells.” *European Journal of Biochemistry* 122.1 (1982): 87–93. Web.

Heinrich, R, and T A Rapoport. “A Linear Steady-State Treatment of Enzymatic Chains. Critique of the Crossover Theorem and a General Procedure to Identify Interaction Sites with an Effector..” *European journal of biochemistry / FEBS* 42.1 (1974): 97–105. Print.

Heinrich, R, and T A Rapoport. “A Linear Steady-State Treatment of Enzymatic Chains. General Properties, Control and Effector Strength..” *European journal of biochemistry / FEBS* 42.1 (1974): 89–95. Print.

Henry, Christopher S et al. “High-Throughput Generation, Optimization and Analysis of Genome-Scale Metabolic Models.” *Nature biotechnology* 28.9 (2010): 977–982. Web.

Jin, Q. “Energy Conservation of Anaerobic Respiration.” *American Journal of Science* 312.6 (2012): 573–628. Web.

Jin, Q, and C M Bethke. “A New Rate Law Describing Microbial Respiration.” *Applied and Environmental Microbiology* 69.4 (2003): 2340–2348. Web.

Jin, Qusheng, and Eric E Roden. “Microbial Physiology-Based Model of Ethanol Metabolism in Subsurface Sediments.” *Journal of Contaminant Hydrology* 125.1-4 (2011): 1–12. Web.

Johnson, Eric F, and Biswarup Mukhopadhyay. “Coenzyme F420-Dependent Sulfite Reductase-Enabled Sulfite Detoxification and Use of Sulfite as a Sole Sulfur Source by Methanococcus Maripaludis..” *Applied and Environmental Microbiology* 74.11 (2008): 3591–3595. Web.

Kacser, H, and J A Burns. “Molecular Democracy: Who Shares the Controls?.” *Biochemical Society Transactions* 7.5 (1979): 1149–1160. Web.

King, Zachary A et al. “Next-Generation Genome-Scale Models for Metabolic Engineering..” *Current opinion in biotechnology* 35 (2015): 23–29. Web.

Klier, Christine. “Use of an Uncertainty Analysis for Genome-Scale Models as a Prediction Tool for Microbial Growth Processes in Subsurface Environments.” *Environmental Science & Technology* 46.5 (2012): 2790–2798. Web.

Mahadevan, R et al. “Characterization of Metabolism in the Fe(III)-Reducing Organism Geobacter Sulfurreducens by Constraint-Based Modeling.” *Applied and Environmental Microbiology* 72.2 (2006): 1558–1568. Web.

Marsili, Enrico et al. “Shewanella Secretes Flavins That Mediate Extracellular Electron Transfer..” *Proceedings of the National Academy of Sciences of the United States of America* 105.10 (2008): 3968–3973. Web.

Middleton, R J, and H Kacser. “Enzyme Variation, Metabolic Flux and Fitness: Alcohol Dehydrogenase in *Drosophila Melanogaster*.” *Genetics* 105.3 (1983): 633–650. Print.

Monod, J. “The Growth of Bacterial Cultures.” *Annual Review of Microbiology* 3.1 (1949): 371–394. Web.

Novak, Maja et al. “Experimental Tests for an Evolutionary Trade-Off Between Growth Rate and Yield in E. Coli..” *The American naturalist* 168.2 (2006): 242–251. Web.

Oliveira, Ana Paula, Jens Nielsen, and Jochen Förster. “Modeling *Lactococcus Lactis* Using a Genome-Scale Flux Model .” *BMC microbiology* 5 (2005): 39. Web.

Ong, Wai Kit et al. “Comparisons of Shewanella Strains Based on Genome Annotations, Modeling, and Experiments..” *BMC systems biology* 8.1 (2014): 31. Web.

Orth, Jeffrey D, Ines Thiele, and Bernhard Ø Palsson. “What Is Flux Balance Analysis?.”

Nature biotechnology 28.3 (2010): 245–248. Web.

Pinchuk, Grigoriy E et al. “Constraint-Based Model of Shewanella Oneidensis MR-1 Metabolism: a Tool for Data Analysis and Hypothesis Generation..” *PLoS Computational Biology* 6.6 (2010): e1000822. Web.

Ranalli, G, T N Whitmore, and D Lloyd. “Methanogenesis From Methanol in Methanosarcina Barkeri Studied Using Membrane Inlet Mass Spectrometry.” *FEMS Microbiology Letters* 35.2-3 (1986): 119–122. Web.

Salter, M, R G Knowles, and C I Pogson. “Quantification of the Importance of Individual Steps in the Control of Aromatic Amino Acid Metabolism..” *The Biochemical journal* 234.3 (1986): 635–647. Print.

Sauer, Karin, Ulrike Harms, and Rudolf K Thauer. “Methanol: Coenzyme M Methyltransferase From Methanosarcina Barkeri.” *European Journal of Biochemistry* 243.3 (1997): 670–677. Web.

Scheibe, Timothy D et al. “Coupling a Genome-Scale Metabolic Model with a Reactive Transport Model to Describe in Situuranium Bioremediation.” *Microbial Biotechnology* 2.2 (2009): 274–286. Web.

Simkins, S, and M Alexander. “Models for Mineralization Kinetics with the Variables of Substrate Concentration and Population Density..” *Applied and Environmental Microbiology* 47.6 (1984): 1299–1306. Print.

Sloan, S W. “Lower Bound Limit Analysis Using Finite Elements and Linear Programming.” *International Journal for Numerical and Analytical Methods in Geomechanics* 12.1 (1988): 61–77. Web.

Smith, M R, and R A Mah. “Growth and Methanogenesis by Methanosarcina Strain 227 on Acetate and Methanol..” *Applied and Environmental Microbiology* 36.6 (1978): 870–879. Print.

Stouthamer, A H. “A Theoretical Study on the Amount of ATP Required for Synthesis of Microbial Cell Material..” *Antonie van Leeuwenhoek* 39.3 (1973): 545–565. Print.

Sun, Jun et al. “Genome-Scale Constraint-Based Modeling of Geobacter Metallireducens..” *BMC systems biology* 3 (2009): 15. Web.

Søballe, B, and R K Poole. “Microbial Ubiquinones: Multiple Roles in Respiration, Gene Regulation and Oxidative Stress Management..” *Microbiology (Reading, England)* 145 (Pt 8).8 (1999): 1817–1830. Web.

Tartakovsky, G D et al. “Pore-Scale Simulation of Microbial Growth Using a Genome-Scale Metabolic Model: Implications for Darcy-Scale Reactive Transport.” *Advances in Water Resources* 59.C (2013): 256–270. Web.

Thiele, Ines, and Bernhard Ø Palsson. “A Protocol for Generating a High-Quality Genome-Scale Metabolic Reconstruction.” *Nature Protocols* 5.1 (2010): 93–121. Web.

Torres, N V et al. “Kinetics of Metabolic Pathways. a System in Vitro to Study the Control of Flux..” *The Biochemical journal* 234.1 (1986): 169–174. Print.

Welte, Cornelia, Lena Kröninger, and Uwe Deppenmeier. “Experimental Evidence of an Acetate Transporter Protein and Characterization of Acetate Activation in Aceticlastic Methanogenesis of Methanosarcina Mazei..” *FEMS Microbiology Letters* 359.2 (2014): 147–153. Web.