Expanding the Capabilities of Constraint-based Metabolic Models for Biotechnology Purposes

by

Kai Hua Zhuang

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Department of Chemical Engineering and Applied Chemistry
University of Toronto

© Copyright by Kai Hua Zhuang, 2012
Expanding the Capabilities of Constraint-based Metabolic Models for Biotechnology Purposes

Kai Hua Zhuang

Doctor of Philosophy

Department of Chemical Engineering and Applied Chemistry
University of Toronto

2012

Abstract

Over the past decade, the constraint-based approach to metabolic modeling has become an important tool for understanding and controlling biology. Unfortunately, the application of this novel approach to systems biology in biotechnology has been limited by three significant technical issues: existing metabolic modeling methods cannot completely model the overflow metabolism, cannot model the metabolism of microbial communities, and cannot design strains optimized for productivity and titer. Three computational methods – the Flux Balance Analysis with Membrane Economics (FBA\textsuperscript{ME}) method, the Dynamic Multi-species Metabolic Modeling (DyMMM) framework, and the Dynamic Strain Scanning Optimization (DySScO) strategy – have been developed to resolve these issues respectively.

First, the FBA\textsuperscript{ME} method, which adopts the membrane occupancy limitation hypothesis, was used to explain and predict the phenomenon of overflow metabolism, an important metabolic phenomenon in industrial fermentation, in \textit{Escherichia coli}. Then, the DyMMM framework was used to investigate the community metabolism during uranium bioremediation, and demonstrated that the simultaneous addition of acetate and Fe(III)
may be a theoretically viable uranium bioremediation strategy. Lastly, the DySScO strategy, which combines the DyMMM framework with existing strain design algorithms, was used to design commodity-chemical producing *E. coli* optimized for a balanced product yield, titer, and volumetric productivity. These novel computational methods allow for broader applications of constraint-based metabolic models in biotechnology settings.
Acknowledgments


In addition, the author of this thesis (Kai Zhuang) would like to acknowledge his supervisor, Radhakrishnan Mahadevan, as well as his collaborators: Melissa Barlett, Goutham Vemuri, Eugene Ma, and Derek Lovley.
# Table of Contents

Abstract .............................................................................................................................................. 2
Acknowledgments ............................................................................................................................ 4
Table of Contents .............................................................................................................................. 5
List of Tables ..................................................................................................................................... 10
List of Figures ..................................................................................................................................... 11
List of Equations ............................................................................................................................... 12

Chapter 1. Introduction ..................................................................................................................... 14
  1.1. The Advent of Metabolic Modeling .......................................................................................... 14
  1.2. The Constraint-Based Reconstruction and Analysis of Metabolism ................................. 16
  1.3. Limitations of the COBRA Models and Methods ................................................................. 19
      1.3.1. The Need for Modeling the Overflow Metabolism ....................................................... 19
      1.3.2. The Need for Community Metabolic Models .............................................................. 21
      1.3.3. The Need for Integrating Bioprocess and Metabolic Models ....................................... 23
  1.4. Thesis Motivations and Objectives ......................................................................................... 25
      1.4.1. Motivations .................................................................................................................. 25
      1.4.2. Objectives .................................................................................................................. 25
  1.5. Thesis Organization ............................................................................................................... 27

Chapter 2. The Economics of Membrane Occupancy and the Respiro-fermentation Phenomenon .......................................................................................................................... 29
  2.1. Introduction ............................................................................................................................ 29
  2.2. Materials and Methods .......................................................................................................... 32
      2.2.1. Membrane Cost of Metabolic Reactions ...................................................................... 32
      2.2.2. Membrane Economics for Prokaryotes ....................................................................... 33
      2.2.3. Calculating the Relative Membrane Costs in E. coli .................................................. 34
      2.2.4. Simulation of the aerobic growth of cytochrome knockout strain .............................. 35
  2.3. Results .................................................................................................................................... 37
      2.3.1. Prokaryotic membrane economics .............................................................................. 37
      2.3.2. Modeling E. coli physiology ....................................................................................... 38
2.3.3. Aerobic physiology in nutrient-excess and nutrient-limiting environments...... 39
2.3.4. Expression of cytochromes and activation of fermentation pathways .................. 43
2.3.5. Expression of Cyd-I under microaerobic growth conditions ................................ 46
2.3.6. Rejection of the Molecular Crowding Theory .................................................... 48

2.4. Discussion .................................................................................................................. 50

2.4.1. How Crowded Is the Cytoplasmic Membrane? ..................................................... 50
2.4.2. Linking Physiology and Morphology in Bacteria .................................................. 50
2.4.3. Membrane Constraints in Eukaryotic Cells ......................................................... 51

2.5. Concluding Remarks .............................................................................................. 53

Chapter 3. The Dynamic Multi-species Metabolic Modeling framework ........ 54

3.1. Introduction ............................................................................................................... 54
3.2. The formulation of the DyMMM framework .............................................................. 56
3.3. Sample Implementation of the DyMMM framework .................................................. 58
3.4. Advantage of the DyMMM framework ..................................................................... 62

3.4.1. DyMMM vs. Monod-kinetic Model ..................................................................... 62
3.4.2. DyMMM vs. Stolyar Model ................................................................................. 63
3.4.3. DyMMM vs. Zomorrodi Model .......................................................................... 64
3.4.4. DyMMM vs. Experimental Studies ..................................................................... 65

3.5. Concluding Remarks .............................................................................................. 66

Chapter 4. Integrative Analysis of Geobacter spp. and Sulfate-Reducing
Bacteria during Uranium Bioremediation ................................................................. 67

4.1. Introduction ............................................................................................................... 67
4.2. Methods .................................................................................................................... 69

4.2.1. Sediment Incubation with Acetate .................................................................... 69
4.2.2. Fe(III) Oxide Addition Experiment .................................................................... 70
4.2.3. Analytical Methods ............................................................................................ 70
4.2.4. Dynamic Metabolic Modeling of Geobacter and SRB ........................................ 70
4.2.5. Core-Pathway Model of SRB ............................................................................ 71
4.2.6. Determination of Model Parameters .................................................................. 72
4.2.7. Simulations of Batch Incubation Experiments .................................................... 73

4.3. Results and Discussion .......................................................................................... 74

4.3.1. Sediment Incubation with Acetate and its Simulation ........................................ 74
Chapter 5. The Design of Long-term Effective Uranium Bioremediation

Strategy using a Community Metabolic Model

5.1. INTRODUCTION

5.2. METHOD

5.2.1. Overview of the Community Modeling framework

5.2.2. Expanded genome-scale metabolic model of Geobacter

5.2.3. Differentiating easy-to-use and hard-to-use Fe(III)

5.2.4. Uranium reduction pathway of Geobacter

5.2.5. Core-pathway model of SRB

5.2.6. Environmental and physiological parameters

5.2.7. Simulations of batch and continuous addition of Fe(III)

5.2.8. Optimization of Acetate and Fe(III) Addition

5.3. RESULTS AND DISCUSSION

5.3.1. Simulation of 2002 Rifle experiment

5.3.2. Batch Addition of Fe(III) does not prolong uranium reduction

5.3.3. Continuous Addition of Acetate and Fe(III) can prolong uranium reduction

5.3.4. Optimizing the rates of acetate and Fe(III) addition

5.4. Concluding Remarks

Chapter 6. DySScO: an efficient strain design algorithm for balanced yield, titer, and productivity
6.3. Results and Discussions ................................................................. 115
  6.3.1. Dynamic Strain Scanning Optimization (DySScO) Strategy ...................... 115
  6.3.2. Designing Succinate-Producing Strains ................................................... 116
  6.3.3. Designing 1,4-Butanediol-Producing Strains ........................................ 120
  6.3.4. Succinate Production in Fedbatch ............................................................ 123
  6.3.5. Balancing Growth Rate and Product Yield ................................................. 126
  6.3.6. Analyses of the Economic Viability of the Strains .................................... 127
  6.3.7. Alternative Approaches ............................................................................ 131
6.4. Concluding Remarks .............................................................................. 132

Chapter 7. Synthesis ..................................................................................... 133
  7.1. A Tale of Tradeoffs .................................................................................. 133
  7.2. Assumptions, Limitations, and Appropriateness ........................................ 136
    7.2.1. From Kinetics to Constraints .................................................................. 136
    7.2.2. The importance of concentration variables ........................................... 137
    7.2.3. Reconnecting fluxes to concentrations .................................................. 138
    7.2.4. Exploring the Optimality Principle ....................................................... 139
    7.2.5. When is the genome-scale model needed? ............................................ 140
    7.2.6. Model Requirements for Environmental Biotechnology ....................... 142
    7.2.7. Parting Thoughts ................................................................................... 145
  7.3. Future Recommendations ........................................................................ 146

Bibliography .................................................................................................. 149

Appendix I. Nomenclature ............................................................................. 162

Appendix II. Major Assumptions and Limitations ......................................... 167

Appendix III. Chapter 2 Supplementary Information .................................... 169
  III-1. Model Sensitivity .................................................................................. 169
  III-2. Comparison with C-13 Flux Analysis ..................................................... 170
  III-3. Construction of FBA\textsuperscript{ME} model ............................................... 170
  III-4. Calculating the Relative Membrane Cost in \textit{E. coli} .................................. 171
  III-5. Estimation of Membrane Protein Content .............................................. 174
  III-6. Evidence for Decrease in Respiratory Efficiency ................................. 175

Appendix IV. Chapter 4 Supplementary Information .................................... 177
IV-1. Determination and sensitivity analysis of Geobacter's maximal uptake rate......177
IV-2. Ideal Batch Reactor Simulations with Low Initial Biomass...............................178
IV-3. Ideal Batch Reactor Simulations with High Initial Fe(III) Concentration..............179
IV-4. Fe(III) amendment simulation with 5mM initial Fe(III) concentration.................180

Appendix V. Chapter 6 Supplementary Information.................................................181
   V-1. Modeling Succinate Inhibition in E. coli.........................................................181
   V-2. The Effect of Fedbatch time on Titer.........................................................182

Appendix VI. List of Model Parameters.................................................................183

Copyright Acknowledgements...............................................................................187
List of Tables

Table 6-1. Knockout Strategies for Succinate Overproduction Identified ...........................................118
Table 6-2. Knockout Strategies for BDO Overproduction Identified .......................................................122
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>The solution spaces of FBA and FBA&lt;sub&gt;ME&lt;/sub&gt;</td>
<td>40</td>
</tr>
<tr>
<td>2-2</td>
<td>Prediction of E. coli Aerobic Growth Physiology</td>
<td>42</td>
</tr>
<tr>
<td>2-3</td>
<td>Prediction of Gene Expression</td>
<td>44</td>
</tr>
<tr>
<td>2-4</td>
<td>Environmental Conditions and Membrane Occupancy</td>
<td>45</td>
</tr>
<tr>
<td>2-5</td>
<td>Predicted cytochrome expression of E. coli grown in oxygen-limited chemostat.</td>
<td>47</td>
</tr>
<tr>
<td>2-6</td>
<td>The Rejection of Molecular Crowding Theory</td>
<td>49</td>
</tr>
<tr>
<td>3-1</td>
<td>The Dynamic Multi-species Metabolic Modeling Framework</td>
<td>61</td>
</tr>
<tr>
<td>4-1</td>
<td>Microbiological and geochemical impact of adding acetate to subsurface sediments</td>
<td>76</td>
</tr>
<tr>
<td>4-2</td>
<td>Model predictions of community dynamics</td>
<td>77</td>
</tr>
<tr>
<td>4-3</td>
<td>Ideal batch reactor model simulations with varying initial ratio of Geobacter and SRB</td>
<td>80</td>
</tr>
<tr>
<td>4-4</td>
<td>Predicted impact of adding Fe(III)</td>
<td>82</td>
</tr>
<tr>
<td>4-5</td>
<td>Experimentally observed impact of adding Fe(III)</td>
<td>82</td>
</tr>
<tr>
<td>5-1</td>
<td>Conceptual model of the subsurface community during bioremediation</td>
<td>86</td>
</tr>
<tr>
<td>5-2</td>
<td>Simulation of the 2002 Rifle Experiment</td>
<td>96</td>
</tr>
<tr>
<td>5-3</td>
<td>Simulation of Batch Fe(III) Amendment</td>
<td>98</td>
</tr>
<tr>
<td>5-4</td>
<td>Simulation of Continuous Fe(III) Amendment</td>
<td>100</td>
</tr>
<tr>
<td>5-5</td>
<td>Optimized Amendment of Fe(III) and Acetate</td>
<td>102</td>
</tr>
<tr>
<td>6-1</td>
<td>A schematic diagram of the DySSCo strategy</td>
<td>110</td>
</tr>
<tr>
<td>6-2</td>
<td>The performance of the succinate-producing strains in batch</td>
<td>117</td>
</tr>
<tr>
<td>6-3</td>
<td>The performance of the BDO-producing strains in batch</td>
<td>121</td>
</tr>
<tr>
<td>6-4</td>
<td>The performance of the succinate-producing strains in fed-batch</td>
<td>125</td>
</tr>
<tr>
<td>6-5</td>
<td>The economic performance of the succinate-producing strains</td>
<td>128</td>
</tr>
<tr>
<td>6-6</td>
<td>The economic performance of the BDO-producing strains</td>
<td>129</td>
</tr>
</tbody>
</table>
List of Equations

<table>
<thead>
<tr>
<th>Equation</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equation 1-1</td>
<td>17</td>
</tr>
<tr>
<td>Equation 2-1</td>
<td>31</td>
</tr>
<tr>
<td>Equation 2-2</td>
<td>32</td>
</tr>
<tr>
<td>Equation 2-3</td>
<td>32</td>
</tr>
<tr>
<td>Equation 2-4</td>
<td>33</td>
</tr>
<tr>
<td>Equation 2-5</td>
<td>33</td>
</tr>
<tr>
<td>Equation 2-6</td>
<td>33</td>
</tr>
<tr>
<td>Equation 2-7</td>
<td>33</td>
</tr>
<tr>
<td>Equation 2-8</td>
<td>33</td>
</tr>
<tr>
<td>Equation 2-9</td>
<td>34</td>
</tr>
<tr>
<td>Equation 2-10</td>
<td>34</td>
</tr>
<tr>
<td>Equation 2-11</td>
<td>34</td>
</tr>
<tr>
<td>Equation 3-1</td>
<td>55</td>
</tr>
<tr>
<td>Equation 3-2</td>
<td>55</td>
</tr>
<tr>
<td>Equation 3-3</td>
<td>55</td>
</tr>
<tr>
<td>Equation 3-4</td>
<td>58</td>
</tr>
<tr>
<td>Equation 3-5</td>
<td>62</td>
</tr>
<tr>
<td>Equation 3-6</td>
<td>62</td>
</tr>
<tr>
<td>Equation 3-7</td>
<td>62</td>
</tr>
<tr>
<td>Equation 4-1</td>
<td>70</td>
</tr>
<tr>
<td>Equation 4-2</td>
<td>71</td>
</tr>
<tr>
<td>Equation 4-3</td>
<td>71</td>
</tr>
<tr>
<td>Equation 5-1</td>
<td>87</td>
</tr>
<tr>
<td>Equation 5-2</td>
<td>88</td>
</tr>
<tr>
<td>Equation 5-3</td>
<td>88</td>
</tr>
<tr>
<td>Equation 5-4</td>
<td>92</td>
</tr>
<tr>
<td>Equation 6-1</td>
<td>107</td>
</tr>
<tr>
<td>Equation 6-2</td>
<td>108</td>
</tr>
<tr>
<td>Equation 6-3</td>
<td>108</td>
</tr>
<tr>
<td>Equation 6-4</td>
<td>111</td>
</tr>
<tr>
<td>Equation 6-5</td>
<td>111</td>
</tr>
<tr>
<td>Equation 6-6</td>
<td>111</td>
</tr>
<tr>
<td>Equation 6-7</td>
<td>112</td>
</tr>
<tr>
<td>Equation 6-8</td>
<td>113</td>
</tr>
</tbody>
</table>
Chapter 1. Introduction

1.1. The Advent of Metabolic Modeling

Biotechnologies are technologies that utilize biological organisms, components, or processes in beneficial ways. Important ancient examples of biotechnologies include agriculture and the fermentative production of food and beverages; some modern examples include the production of biochemicals and biofuels, as well as the bioremediation of sewage and toxic wastes. Today, there are vast public and industrial interests in the development and refinement of biotechnology, fueled by a multitude of social, economical, political, and scientific reasons. One important reason is that biotechnologies are deemed more sustainable in general than non-biological alternatives (Gavrilescu & Chisti, 2005), another reason is that some biotechnologies, such as uranium bioremediation, are functionally superior to non-biological alternatives (Anderson et al, 2003). Recently, advances in laboratory technologies such as high-throughput data acquisition methods have given us unprecedented access to biological information; since the development of modern biotechnologies relies on the manipulation of the biological organisms, components, or processes guided by the detailed scientific understanding of the relevant biology, the accumulation of biological data alone is insufficient — the need to interpret these data in integrative, systemic, and meaningful ways have propelled the exponential growth in the field of computational systems biology.

Over the past decade, the Constraint Based Reconstruction and Analysis (COBRA) approach to systems biology is quickly maturing into a popular tool for investigating large-scale biochemical networks in both academic and industrial settings. The reconstruction of the constraint-based metabolic models from genomic sequences and biochemical data is becoming standardized (Thiele & Palsson, 2010) and the process has been partially automated (Henry et al, 2010). The genome-scale metabolic models of many scientifically and industrially significant organisms, such as Escherichia coli,
Saccharomyces cerevisiae, and Homo sapiens, have been developed. The usages of these models have led to industrially relevant discoveries in areas of metabolic engineering (Brochado et al., 2010; Feist et al., 2010; Yim et al., 2011), bioprocess engineering (Hjersted et al., 2007), and environmental biotechnology (Ahsanul Islam et al., 2010; Mahadevan et al., 2006; Segura et al., 2008; Sun et al., 2009; Zhuang et al., 2011a). For example, Hjersted et al. (2007) used the S. cerevisiae model to screen for potential gene additions that enhance ethanol productivity and identified several alternative genetic engineering strategies with enhanced ethanol productivity (Hjersted et al., 2007). In another example, model-based identification of genetic engineering strategies has led to the design of a commercial product pipeline for the commodity chemical, 1,4-butanediol (Yim et al., 2011). Flux Balance Analysis (FBA), an optimization technique commonly used for the analysis and exploration of these metabolic models, has been particularly instrumental in many of these discoveries (See next section for a description of FBA).

Unfortunately, despite the many successful metabolic modeling efforts, existing metabolic models are limited by several outstanding shortcomings, inhibiting its broader application in biotechnology. In order to understand and appreciate these shortcomings, we must first understand the COBRA approach to metabolic modeling in detail. In the next section, the fundamentals of the COBRA approach as well as a number of COBRA methods are reviewed.
1.2. The Constraint-Based Reconstruction and Analysis of Metabolism

The hallmark of the Constraint-Based Reconstruction and Analysis (COBRA) approach to systems biology and metabolic modeling is its focus on the various constraints of cellular metabolism (Terzer et al, 2009), a trait that distinguishes constraint-based metabolic models from other types of metabolic models (Machado et al, 2012). The standard constraint-based metabolic modeling formulation has two major components — the stoichiometric matrix and the constraints (Price et al, 2004; Terzer et al, 2009; Thiele & Palsson, 2010; Henry et al, 2010; Schellenberger et al, 2011). The stoichiometric matrix is a mathematical representation of the biochemical reaction network, and the constraints represent physicochemical, environmental, and regulatory limitations on the network (Terzer et al, 2009). Generally, the biochemical reaction network is underdetermined — by applying additional constraints, we reduce the size of the solution space (Price et al, 2004; Terzer et al, 2009). Because this approach does not require any kinetic descriptions, a COBRA model can be constructed with minimal experimental data. On the other hand, a variety of physicochemical, environmental, and regulatory information can be incorporated into the model as additional mathematical constraints (Terzer et al, 2009). A standard model reconstruction process has been described (Thiele & Palsson, 2010) and can be automated to aid the manual reconstruction process (Henry et al, 2010). In addition, a number of open-source software and public databases, such as the COBRA Toolbox (Becker et al, 2007; Schellenberger et al, 2011) and the Biochemical Genetic and Genomic database (Schellenberger et al, 2010), have been developed specifically for the reconstruction and analysis of constraint-based metabolic models (Medema et al, 2012).

Since the constraint-based metabolic models are underdetermined mathematically, it offers us a solution space instead of a single solution. The flux balance analysis (FBA) uses linear programming technique to select an optimal point in the solution space for a given objective function (Equation 1-1). The rationale behind this exercise is that under a constant environment, organisms tend to develop an optimal evolutionary strategy to cope with the environment. This optimality principle is a fundamental assumption of
FBA. The mathematical description of FBA is given below (Equation 1-1). Here, \( A \) is the stoichiometric matrix, \( v \) is the vector of metabolic fluxes, \( c^T \) is the objective function vector, \( v_{\text{min}} \) and \( v_{\text{max}} \) are the lower and upper bound of the fluxes:

\[
\text{maximize} \quad c^T v \\
\text{st.} \quad Av = 0 \\
\quad v_{\text{min}} \leq v \leq v_{\text{max}}
\]  

Equation 1-1

The most commonly used FBA objective function is the maximization of biomass flux, which assumes that the organism wants to maximize its growth rate in order to outcompete other organisms. While FBA is useful for predicting wild-type metabolism as well as the metabolic behaviours of organisms under long-term evolution, alternative methods such as the Minimization of Metabolic Adjustment (MoMA) and the Regulatory On/Off Minimization (ROOM) are more useful for the prediction of metabolic behaviours of knockout strains. MoMA minimizes the difference between the metabolic distributions of wild-type and knockout organisms (Segrè et al, 2002), whereas ROOM minimizes the number of significant metabolic changes between wild-type and knockout organisms (Shlomi et al, 2005).

From a metabolic engineering perspective, if we are interested in producing a particular metabolic by-product biologically, then the maximization of flux of interest can be used as the objective function in order to calculate the theoretical maximum yield of this by-product. Unfortunately, the productions of metabolic by-products are often minimal in wild-type organisms because there is a tradeoff between the production of biomass and by-products. To resolve this problem, bi-level optimization algorithms such as OptKnock (Burgard et al, 2003), OptReg (Pharkya & Maranas, 2006), GDLS (Lun et al, 2009), and EMILiO (Yang et al, 2008) are be used to generate “growth-coupled strain designs” in which the production of the by-product of interest is coupled to cellular growth.

A central assumption of the constraint-based approach to metabolic modeling is the notion that the cellular metabolism is at a quasi-steady-state with respect to cell growth.
This quasi-steady-state assumption greatly simplifies the mathematics and enables COBRA models to analyze the metabolic fluxes at a particular steady state without dealing with any reaction kinetics. However, in many biotechnology settings, the reprogramming of the metabolic network in response to external changes is of importance. To address this issue, Mahadevan et al. (2002) coupled the dynamic model of the bioreactor process with the constraint-based model of the cellular metabolism. This method, called the Dynamic Flux Balance Analysis (dFBA), is able to predict the dynamic interactions between the cell and its environment. The dFBA method has been used to investigate the diauxic growth of *E. coli* cells (Mahadevan et al., 2006) and the dynamics control of this organism through a genetic toggle switch (Anesiadis et al., 2008), as well as the production of ethanol by *S. cerevisiae* in fed-batch cultures (Hjersted et al., 2007).

As the constraint-based approach to metabolic modeling continues to mature, the biotechnology industry is beginning to adopt this methodology for bioengineering purposes. However, despite the numerous successes of constraint-based approach, the existing constraint-based models and methods are not without caveats. In the next section, three outstanding limitations of the existing constraint-based approach to metabolic modeling will be discussed; the need to resolve these limitations serve as the fundamental motivation for this PhD thesis.
1.3. Limitations of the COBRA Models and Methods

As mentioned in the previous section, despite the success and popularity of the COBRA approach to metabolism modeling, there are a number of significant issues with the application of constraint-based models and FBA in biotechnology settings. For example, it has been suggested that FBA is not applicable in nutrient-rich environments because it cannot predict the overflow metabolism (Schuster et al, 2008b). In addition, most existing COBRA models and methods (e.g., FBA, ROOM, OptKnock) focus on the prediction and analysis of individual organisms, and are not applicable for biotechnologies that rely on microbial communities. Finally, existing strain-design algorithms (e.g., OptKnock, GDLS, EMILiO) focus on the optimization of the product yield (mol of product per mol of feedstock), which measures the quantity of chemical produced per unit of substrate. These algorithms lack the ability to optimize the volumetric productivity (mmol/L/hr) and titer (mmol/L), which measure, respectively, the quantity of chemical produced per unit of time in a fixed reactor volume and the concentration of chemical product at the end of the bioprocess. These shortcomings greatly reduces the applicability of these algorithms in industrial settings where all three of these design criteria are equally important.

1.3.1. The Need for Modeling the Overflow Metabolism

In nutrient-rich environments, many microorganisms relevant to biochemical production and food processing, as well as most cancer cells, simultaneously utilize the both efficient and inefficient ATP-producing pathways (Pfeiffer & Schuster, 2001; MacLean, 2008; Schuster et al, 2008b; 2008a). This well-known metabolic phenomenon is often referred to as the overflow metabolism, which stems from the hypothesis that when a heterotrophic organism is given abundant nutrient, the limited capacity of its most efficient pathway becomes saturated, and the organism is forced to use inefficient pathway(s) in addition to the efficient pathway, thus sacrificing the ATP yield for an
increased ATP production rate (Molenaar et al., 2009; Pfeiffer & Schuster, 2001; Varma & Palsson, 1994). The most well known type of overflow metabolism is the respiro-fermentation phenomenon whereby an organism simultaneously utilizes both the efficient respiratory pathway and the inefficient fermentative pathway in presence of abundant oxygen. Unfortunately, this “limited-capacity of the efficient pathway” hypothesis is contradicted by both experimental evidence (Vemuri et al., 2006; Vemuri et al., 2007; Chapter 2) as well as theoretical considerations (Molenaar et al., 2009; Chapter 2), and the true mechanistic cause of the phenomenon remains elusive despite decades of research.

Since the “limited-capacity of the efficient pathway” hypothesis essentially claims that there is a constraint on the flux through the most efficient pathway, it is straight-forward to implement in FBA models. For example, the respiro-fermentation phenomenon in E. coli can be predicted if an auxiliary oxygen uptake constraint is included in the FBA formulation (Feist et al., 2007; Varma & Palsson, 1994). However, such auxiliary constraints are essentially “fitting parameters” (Schuster et al., 2008b), leaving us with the vexing question of why isn’t the capacity of the efficient pathways expanded through evolutionary means (Molenaar et al., 2009). Furthermore, this approach failed to predict certain important features of the respiro-fermentation phenomenon (Chapter 2). It has been suggested that FBA’s maximization of the flux through the biomass reaction is incorrect from an evolutionary perspective (Schuster et al., 2008b), thus casting doubt on the validity of FBA in nutrient-rich environments. Given that many biotechnologies, such as industrial fermentation and uranium bioremediation, operate under nutrient-rich conditions, it is obvious that the validity of the constraint-based approach under nutrient-rich conditions must be addressed. To do this, it is necessary to demonstrate that FBA (or similar constraint-based methods) can be used to predict and explain the overflow phenomenon. Some nonlinear objective functions have been explored with some success (Schuetz et al., 2007); unfortunately, these objective functions are difficult to interpret, and do little to enhance our mechanistic understanding of the overflow metabolism. More recently, it has been suggested that the molecular crowding phenomenon (Beg et al., 2007) contributes to the overflow metabolism (Vazquez et al., 2008); however, there is some significant evidence against this theory in E. coli (Chapter 2). Nonetheless, the molecular crowding theory hinted that the overflow phenomenon might be caused by
some mechanistic constraints that are unspecified in the conventional FBA formulation. In any case, there is a need to understand the mechanistic cause of the overflow metabolism and to demonstrate that FBA (or similar methods) is capable of predicting the overflow phenomenon; this need serves as the first motivation for this thesis.

1.3.2. The Need for Community Metabolic Models

Because standard constraint-based metabolic models and standard FBA can only model single organisms, past modeling efforts have focused primarily on pure cultures. Unfortunately, many biotechnology applications require an in-depth understanding of the dynamic community interactions between multiple microbial and metabolic species. As such, this shortcoming limits the usage of constraint-based metabolic models and FBA in these biotechnology applications. For example, acetate-amendment in a uranium-contaminated subsurface environment can promote the simultaneous reduction of U(VI) and Fe(III) by Geobacter species (Zhuang et al., 2011a; Chapter 4 and 5). However, continued acetate injection past 50 days can lead to the dominance of acetate-oxidizing sulfate reducers, which are ineffective in U(VI) reduction, thereby reducing the effectiveness of the bioremediation (Anderson et al., 2003; Vrionis et al., 2005). In another example, cellulose can be converted to ethanol using an artificially-designed microbial co-culture containing cellulolytic and solventogenic Clostridial species (Salimi et al., 2010). It is clear that a systemic understanding of the community ecology is necessary for developing effective biotechnology solutions (McMahon et al., 2007).

There are some pioneering attempts (Stolyar et al., 2007; Vallino, 2003) as well as some more recent attempts (Zomorrodi & Maranas, 2012) at extending the constraint-based modeling approach to the modeling of microbial communities. However, the Vallino model (Vallino, 2003) was far too simplistic for it to be useful for specific applications, and both the Stolyar model (Stolyar et al., 2007) and the Zomorrodi (Zomorrodi & Maranas, 2012) model neglected the dynamic nature of the microbial community. Therefore, methods used in these attempts are inappropriate for modeling complex microbial communities such as communities relevant to subsurface bioremediation.
(Zhuang et al, 2011a) and consolidated bio-ethanol production (Salimi et al, 2010). As such, there is a need to develop community metabolic modeling tools and techniques and to demonstrate their usefulness in biotechnology applications; this need served as second motivation for this thesis.
1.3.3. The Need for Integrating Bioprocess and Metabolic Models

Many microorganisms contain native pathways capable of producing useful chemical compounds (Feist et al, 2010), and synthetic pathways can be genetically inserted into model organisms such as *E. coli* (Yim et al, 2011) and *S. cerevisiae* (Brochado et al, 2010). Unfortunately, from the microorganisms’ perspective, most of these chemicals are metabolic byproducts, and their production is often minimal in wild-type organisms (Feist et al, 2010). To create viable microbial cell factories, the production of these chemicals must be enhanced. In the past, wild-type organisms were exposed to mutagens and selected for desirable phenotypes (Park et al, 2008). The advent of metabolic engineering and systems biology have allowed the rational design of microorganisms in which genes are either amplified or deleted based on systematic consideration of the metabolic network (Feist et al, 2010; Park et al, 2008). Over the past decade, many COBRA methods, such as OptKnock (Burgard et al, 2003), GDLS (Lun et al, 2009), and EMILiO (Yang et al, 2011), have been developed to generate metabolic engineering strategies that couples the production of the desired chemical with the growth of the organism, thus ensuring the growth-based selection for the production of the desired chemical. Unfortunately, in most cases where the desired product is organic, there exists a tradeoff between the growth rate and the product yield because the carbon in the feedstock can be either used to produce either biomass or desired compound. Most of the existing algorithms use a bi-level optimization scheme to select for strains that maximizes the product yield while maintaining the growth rate above a predetermined level.

Previous model-based strain design efforts often set the minimal growth rate to an arbitrarily low level in order to maximize the product yield. However, lowering the growth rate will reduce the biomass concentration, which may lead to a reduction in productivity in spite of the increase in product yield. Unfortunately, because existing strain-design algorithms are based in FBA – which is a static description of the cellular metabolism – this tradeoff cannot be evaluated. For the same reason, the existing algorithms also cannot evaluate the titer potential of a design strain. Given that in
industrial settings, the economic value of a strain depends on the combination of product yield, productivity, and titer, there is a need for a novel strain design strategy capable of optimizing the product yield, the productivity, and the titer of the engineered strain; this need served as the third motivation for this thesis.
1.4. Thesis Motivations and Objectives

1.4.1. Motivations

Based on the discussions in section 1.3 that the existing COBRA models and methods have a number of limitations that prevent their application in biotechnology settings. The present PhD thesis is primarily motivated by the need to address three of these limitations:

1. The need to demonstrate that FBA is applicable for nutrient-rich conditions; specifically, to develop an FBA-based method capable of predicting the overflow metabolism.

2. The need to develop a constraint-based approach for modeling microbial communities.

3. The need for the integrative analysis of the process model and the metabolic model in order to design strains with optimal product yield, productivity, and titer.

1.4.2. Objectives

The objective of this thesis is to develop novel constraint-based metabolic modeling methods in order to fulfill the needs specified in section 1.4.1:

1. To develop an extended FBA formulation capable of predicting the adoption of the overflow metabolism by microorganisms in nutrient-rich conditions, and to explain the mechanistic cause of this phenomenon.

This objective is fulfilled through the development of the Flux Balance Analysis with Membrane Economics (FBA ME) method. Using FBA ME, we demonstrated that the
membrane economics theory can fully explain the mechanistic cause of the respiro-fermentation phenomenon (a type of overflow metabolism) in *E. coli* (Chapter 2).

2. **To develop a constraint-based approach for modeling community metabolism, and to demonstrate the effectiveness of this novel community modeling approach in environmental biotechnology applications.**

This objective is fulfilled through the development and the continued refinement of the Dynamic Multi-species Metabolic Modeling (DyMMM) framework (Chapter 3). We applied the DyMMM framework to the analysis of the interactions between *Geobacter*, *Rhodoferax*, and *sulfate-reducing bacteria* (SRB) during uranium bioremediation (Chapter 3, 4). Using a DyMMM-based community model of *Geobacter* and SRB, we demonstrated that the simultaneous addition of acetate and Fe(III) is a theoretically effective and long-term viable uranium bioremediation strategy.

3. **To develop a novel strain design strategy that optimizes the product yield, productivity, and titer of the engineered strain.**

This objective is fulfilled through the development of the Dynamic Strain Scanning Optimization (DySScO) strategy — a novel strain design method that integrates dFBA model of the fermentation process with existing strain design algorithms in order to evaluate strain designs based on their product yield, productivity, and titer. We were able to demonstrate how the DySScO strategy can be used to design *E. coli* strains optimized for succinate and 1,4-butanediol production.
1.5. Thesis Organization

This document is a PhD thesis submitted to the Graduate Department of Chemical Engineering at the University of Toronto. It has been written in accordance with the section 6.3.4(e) of the University of Toronto Chemical Engineering and Applied Chemistry Graduate Handbook (2011-2012 version). Specifically, the aforementioned section of the handbook specify that “Students may, upon agreement between the student, supervisor(s) and the Graduate Coordinator, submit a collection of publishable papers as their thesis. This collection must at minimum have a coherent topic with an introduction presenting the general theme of the research and a conclusion summarizing and integrating the major findings. The minimum requirement is 3 papers (can be 2 published and 1 submitted) to a good quality, peer-reviewed journal.”

This thesis contains seven chapters:

Chapter 1 introduces the constraint-based metabolic modeling approach to systems biology, and discusses the major knowledge gaps and technical issues that limit its application in biotechnology.

Chapter 2 describes the novel metabolic modeling method called Flux Balance Analysis with Membrane Economics (FBA\textsuperscript{ME}), and the application of this method to the prediction and explanation of the overflow phenomenon in \textit{E. coli}. This chapter is adapted from the journal article “Economics of Membrane Occupancy and Respiro-fermentation” (Zhuang \textit{et al}, 2011b). The work described in this chapter fulfills the \textbf{thesis objective #1} outlined in the Section 1.4.2.

Chapter 3 describes the novel community metabolic modeling framework called the Dynamic Multi-species Metabolic Modeling (DyMMM) framework. This chapter is adapted from the supplementary information section of the journal article “Genome-Scale Dynamic Modeling of the Competition Between \textit{Rhodoferax} and \textit{Geobacter} in Anoxic Subsurface Environments” (Zhuang \textit{et al}, 2011a). It serves the technical foundation for the next three chapters.
Chapter 4 describes the application of the DyMMM framework in the modeling and analysis of the community of Geobacter and sulfate-reducing bacteria during uranium bioremediation. This chapter is adapted from the journal article “Integrative Analysis of Geobacter spp. and sulfate-reducing bacteria during uranium bioremediation” (Zhuang and Barlett et al, 2012). Kai Zhuang and Melissa Barlett contributed equally to this article: Kai Zhuang performed all the modeling work, and Melissa Barlett performed all the experimental work. Melissa Barlett has given her permission for her work to appear in this thesis.

Chapter 5 describes how the community model of Geobacter and sulfate-reducing bacteria is used to aid the design of a novel uranium bioremediation strategy. This chapter is adapted from the journal article “The Design of Long-term Effective Uranium Bioremediation Strategy Using a Community Metabolic Model” (Zhuang et al, 2012). 80% of the work in this article was completed by Kai Zhuang. Eugene Ma has performed the computational optimization of the acetate and Fe(III) addition rates; he has given his permission for this work to appear in this thesis. The works described in chapters 4, 5, and 6 fulfill the thesis objective #2 outlined in the Section 1.4.2.

Chapter 6 describes the novel strain design strategy called “Dynamic Strain Scanning Optimization (DySScO)” that integrates existing strain design strategy with the DyMMM framework. The DySScO strategy can be used to produce metabolic engineering strategies that optimizes for the combination of product yield, titer, and productivity. It is adapted from the article “DySScO: an efficient strain design algorithm for balanced yield, titer, and productivity” (Zhuang et al. Submitted). The work described in this chapter fulfills the thesis objective #3 outlined in the Section 1.4.2.

Chapter 7 provides a synthesis of the previous chapters as well as insights on possible future projects.
Chapter 2. The Economics of Membrane Occupancy and the Respiro-fermentation Phenomenon

2.1. Introduction

Many heterotrophs can produce ATP through both respiratory and fermentative pathways, allowing them to survive with or without oxygen. Since the molar ATP yield (molar ATP yield: mole of ATP produced/mole of substrate consumed) from respiration is about 15-fold higher than that from fermentation, ATP production via respiration is more efficient. Surprisingly, at high catabolic rates, many facultative aerobic organisms employ fermentative pathways simultaneously with respiration, even in the presence of abundant oxygen to produce ATP (Pfeiffer & Schuster, 2001; Vemuri et al., 2006; 2007; Veit et al., 2007; MacLean, 2008; Molenaar et al., 2009). This leads to an observable tradeoff between the ATP yield and the catabolic rate (Pfeiffer & Schuster, 2001; Vemuri et al., 2006). This respiro-fermentation physiology is commonly observed in microorganisms, including *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae* (Molenaar et al., 2009), as well as cancer cells (Vander Heiden et al., 2009). Despite extensive research, the biochemical basis for this phenomenon remains obscure.

One influential theory attributed the utilization of the fermentative pathways to a hypothetical limitation on the respiratory capacity (Sonnleitner & Käppeli, 1986; Majewski & Domach, 1990). This theory suggests that as the respiratory pathway becomes saturated at high substrate influx, the organism may choose to satisfy its ATP demand by fermenting additional substrates, a strategy that offers a fitness advantage at the cost of lowering the ATP yield (Majewski & Domach, 1990; Varma & Palsson, 1994; Pfeiffer & Schuster, 2001). However, overexpressing the genes encoding for the rate-limiting enzymes did not increase the respiratory capacity (Cupp & McAlister-Henn, 1991; Repetto & Tzagoloff, 1991). Furthermore, it is puzzling why the respiratory

---

1 This chapter is adapted from the article “Economics of Membrane Occupancy and Respiro-fermentation” published in Nature Molecular Systems Biology in 2011.
capacity varies with different substrates. Despite this caveat, metabolic models (Palsson, 2000) such as the flux balance analysis (FBA) (Varma & Palsson, 1994; Edwards et al., 2001; Feist et al., 2007) commonly adopt the ‘respiratory capacity limitation’ theory through the introduction of an empirically measured cap on maximal oxygen uptake rate (OUR) (Figure 2-1A,B). In addition to respiration, the tricarboxylic acid (TCA) cycle is actively downregulated in *E. coli*, *B. subtilis*, and *S. cerevisiae* during respiro-fermentation (Vemuri et al., 2006; 2007; Sonenshein, 2007); this implies that the OURs of these organisms at higher catabolic rates are perhaps regulated to be lower than their respective maximal OURs, possibly reflecting an unexplained evolutionary advantage for lowered respiration (Molenaar et al., 2009).

Challenging the conventional assumption that aerobic respiration is always preferred over fermentation (Majewski & Domach, 1990; Varma & Palsson, 1994), a recent theory (Pfeiffer & Schuster, 2001) proposed that while the cellular metabolism maximizes the ATP yield in nutrient-poor environments, it maximizes the catabolic rate and the rate of energy dissipation in nutrient-rich environments. The biochemical basis for this switch in metabolic objective is the prohibitively expensive synthesis costs of respiratory enzymes, particularly during high catabolic rate (Pfeiffer & Bonhoeffer, 2004; Molenaar et al., 2009). This line of reasoning leads to the conclusion that pure fermentation be accompanied with high growth rate. Yet, rapidly growing facultative aerobes also respire. Furthermore, if the catabolic rate is indeed maximized during unlimited growth, it is unclear why the maximum substrate uptake is slower under aerobic conditions than anaerobic conditions (Portnoy et al., 2008). Another theory proposed that the tradeoff between ATP yield and catabolic rate is dependent on the fraction of intracellular volume occupied by respiratory enzymes and glycolytic enzymes, respectively (Vazquez et al., 2008). While the FBA with ‘molecular crowding constraint’ (FBAwMC) (Beg et al., 2007; Vazquez et al., 2008) can predict acetate production to a certain extent, it could not predict the experimentally observed changes in growth rate and yield (Section 2.3.6). Furthermore, FBAwMC cannot predict the production of acetate if the electron transport chain enzymes – membrane-bound enzymes that consume little intracellular volume – are removed from its formulation (Section 2.3.6). Despite these shortcomings, these theories
highlight that the rate of metabolic processes must be accounted for in addition to the metabolic stoichiometry in understanding respiro-fermentative metabolism.

Finally, these aforementioned theories assume that the observed tradeoff between the ATP yield and the catabolic rate is solely caused by the utilization of fermentative pathways. However, experimental evidence (Appendix III-6) suggests that the efficiency of the respiratory pathway itself may be compromised due to the utilization of less-efficient dehydrogenases and cytochromes. Given that there exists a thermodynamic tradeoff between the turnover rate and the energetic efficiency of an enzyme (Meyer & Jones, 1973; Waddell & Repovic, 1997; Pfeiffer & Bonhoeffer, 2002), less-efficient enzymes may be preferred for their increased turnover rate. Based on these observations, we propose a simple, alternative explanation of the respiro-fermentation phenomenon by considering membrane occupancy, which provides a mechanistic explanation to all the observed physiological changes during the transition from respiratory to respiro-fermentative metabolism.
2.2. Materials and Methods

2.2.1. Membrane Cost of Metabolic Reactions

The rate of a metabolic reaction (or flux) when the substrate is present in excess can be described by:

\[ v_i = k_{cat,i} [E_i] \]  
Equation 2-1

Here, \( v_i \) is the rate of the \( i^{th} \) metabolic reaction, \( k_{cat,i} \) is the turnover rate of the enzyme catalyzing this reaction, and \([E_i]\) is the concentration of this enzyme. From Equation 2-1, we can derive that the enzymatic concentration (\( E_i^* \)) required for the \( i^{th} \) reaction to achieve a rate of 1 mmol/gdw/h:

\[ [E_i^*] = \frac{1}{k_{cat,i}} \]  
Equation 2-2

Many metabolic reactions are catalyzed by membrane-bound enzymes (e.g., transporters, cytochromes). For these reactions, we define the specific membrane area (surface area/biomass) required for a particular reaction to achieve the reaction rate of 1 mmol/gdw/h as the membrane cost of the reaction. This membrane cost is directly related to the membrane area consumed per enzyme (\( m_i \)) and the turnover rate of the enzyme:

\[ C_i = m_i [E_i] = \frac{m_i}{k_{cat,i}} \]  
Equation 2-3

It has been found that the intracellular concentration of most metabolites is much higher than the Ks values of the enzymes (Bennett et al, 2009). Therefore, \( k_{cat,i} \) for most membrane-bound enzymes \( v_i \) can perhaps be assumed to be independent of the substrate concentrations. However, \( v_i \) of cytochrome oxidases and glucose transporter are exceptions because oxygen and glucose are extracellular. For these enzymes, the apparent
\( k_{\text{cat},i}^{\text{app}} \) can be related to the metabolite concentration using the following Michaelis-Menten expression:

\[
k_{\text{cat},i}^{\text{app}} = k_{\text{cat},i} \frac{S}{(S + K_s)}
\]

Equation 2-4

For the reactions catalyzed by these enzymes, Equation 2-3 becomes:

\[
C_i = m_i [E_i] = \frac{m_i}{k_{\text{cat},i}^{\text{app}}}
\]

Equation 2-5

It follows from Equation 2-4 and Equation 2-5 that for these reactions:

\[
C_i = \frac{m_i (S + K_s)}{k_{\text{cat},i} S}
\]

Equation 2-6

### 2.2.2. Membrane Economics for Prokaryotes

For a given physiological state, the surface area to volume ratio (\( R_{S/V} \)) and the intracellular density (D) of a prokaryotic cell remain relatively constant; the specific membrane area (\( M_{\text{cyt}} \), area/gdw biomass) of the cytoplasmic membrane can be estimated by:

\[
M_{\text{cyt}} = \frac{R_{S/V}}{D}
\]

Equation 2-7

To maintain membrane integrity, a large portion of the cytoplasmic membrane must be composed of lipids; therefore, only the remaining fraction (\( f_{\text{available}} \)) of the specific membrane area is available for the expression of membrane proteins. In other words, the total membrane cost incurred at the cytoplasmic membrane must be less than or equal to the cytoplasmic membrane budget (\( B_{\text{cyt}} \)).
\[ \sum v_i C_i \leq B_{cyt} \]  
Equation 2-8

Where the cytoplasmic membrane budget \(B_{cyt}\) is given by:

\[ B_{cyt} = f_{\text{available}} \cdot M_{cyt} \]  
Equation 2-9

While it is difficult to measure/determine the cytoplasmic membrane budget and the membrane cost of transmembrane enzymes, the relative membrane cost of these enzymes can be determined from the growth physiology of knockout strains. The relative membrane cost is defined as the fraction of the specific membrane area required for a particular reaction to achieve the reaction rate of 1 mmol/gdw/h; this is mathematically described by:

\[ C_i^* = \frac{C_i}{B_{cyt}} \]  
Equation 2-10

From Equation 2-8 and 2-10, we can derive the relative form of the cytoplasmic membrane constraint, where only the relative membrane costs \(C_i^*\) are unknown:

\[ \sum v_i C_i^* \leq 1 \]  
Equation 2-11

### 2.2.3. Calculating the Relative Membrane Costs in E. coli

As a simplification, our model assumes that only the expression of the glucose transporter and the three cytochromes (Cyo, Cyd-I, Cyd-II) are constrained by the availability of the cytoplasmic membrane. The relative membrane costs of these enzymes were calculated from the growth data of various cytochrome-knockout strains. Because the membrane occupancy constraint is the only constraint where the cost parameters appears in our model, this constraint must be binding \((v_i C_i = 1)\) for us to acquire unique cost parameters. During batch growth, a cell’s metabolism is not expected to be limited by the rate of glucose uptake as glucose concentrations are high. Under these conditions, the binding
requirement \((v_i C_i = 1)\) is automatically satisfied – since the membrane occupancy constraint is the only active/limiting constraint, and hence, must be binding. However, during glucose-limited chemostat growth, the metabolism of the cell can be limited by either the glucose uptake rate or the membrane occupancy constraint. If metabolism of the cell is limited by the glucose uptake rate \((e.g.,\) under conditions where the glucose feed rate is very low), then the membrane may be unsaturated. Therefore, in order to satisfy this requirement, we have chosen data generated from experimental conditions where the membrane is likely to be saturated; in other words, we have implicitly assumed that the membrane occupancy constraint is binding throughout the parameter estimation process. The details of the parameter estimation process and more information on the membrane constraint binding requirement are in the Appendix III-3, III-4, III-5.

The inferred relative cytoplasmic membrane costs under aerobic conditions are:

\[
\begin{align*}
C^*_\text{GUR}: & \quad 0.0556 \quad \text{gdw} \cdot \text{hr / mmol} \\
C^*_\text{Cyo}: & \quad 0.0658 \quad \text{gdw} \cdot \text{hr / mmol} \\
C^*_\text{Cyd-I}: & \quad 0.0427 \quad \text{gdw} \cdot \text{hr / mmol} \\
C^*_\text{Cyd-II}: & \quad 0.0128 \quad \text{gdw} \cdot \text{hr / mmol}
\end{align*}
\]

The FBA\textsuperscript{ME} model of \textit{E. coli} can be found in the BioModels Database (accession number: MODEL1105030000). The model sensitivity to the cost parameters are in Appendix III-1.

2.2.4. \textit{Simulation of the aerobic growth of cytochrome knockout strain}

Portnoy \textit{et al.} (2008) has reported that an \textit{E. coli} strain with Cyo, Cyd-I, and Cyd-II genes deleted is able to grow under aerobic conditions with a glucose uptake rate of 20
mmol/gdw/h and a growth rate of 0.42 h\(^{-1}\). They also found that under aerobic conditions, the expression of the PFL gene is half of the corresponding expression under anaerobic condition. Using FBA\(^{\text{ME}}\), we simulated the growth of this knockout strain by shutting off the Cyo, Cyd-I, and Cyd-II reactions, and constraining the PFL flux to half of its predicted anaerobic value.
2.3. Results

2.3.1. Prokaryotic membrane economics

The lipid membrane is an essential feature of all cells, and hosts various transmembrane proteins such as metabolite transporters and the respiratory enzymes. To maintain membrane integrity, the protein-to-lipid ratio must be kept below a critical level (Molenaar et al., 2009); therefore, only a fraction of the membrane area is available for protein expression. At higher catabolic rates, the prokaryotic cytoplasmic membrane can become saturated with transmembrane proteins. When this happens, all transmembrane proteins must ‘compete’ for the fixed membrane area available for protein expression (see Materials and Methods). Such a constraint on transmembrane protein expression may have significant physiological consequences in prokaryotes, such as *E. coli*, at higher catabolic rates. First, since both substrate transporters and respiratory enzymes are localized on the cytoplasmic membrane in prokaryotes, increased substrate uptake rates necessitates a decrease in the respiratory rate. This decrease in the respiratory rate forces prokaryotes to process the additional substrate through the fermentative pathways, which are not catalyzed by transmembrane proteins, for continued ATP production. Furthermore, since the membrane requirement of an enzyme is inversely related to its turnover rate (see Materials and Methods), the faster, but inefficient respiratory enzymes (such as Cyd-I and Cyd-II in *E. coli*) might be preferred over the slower and efficient enzymes (such as Cyo in *E. coli*), leading to an altered respiratory stoichiometry at higher catabolic rates. Finally, the absence of the respiratory enzymes under anaerobic conditions releases transmembrane space, which can be occupied by glucose transporters, resulting in faster glucose uptake rate (GUR) of *E. coli*.

Unlike the empirically derived uptake constraints used in FBA (Varma & Palsson, 1994; Edwards et al., 2001; Feist et al., 2007), our proposed theory of cytoplasmic membrane economics is mechanistic in nature, and is extensively supported by experimental
evidence. For example, the overexpression of a non-respiratory membrane protein (YidC) in *E. coli* decreased the expression of succinate dehydrogenase and cytochrome oxidases as well as the oxygen consumption rate by 450%; similarly, the overexpression of membrane proteins (YidC, YedZ, and Lepl) alone was capable of inducing aerobic fermentation in the form of acetate production (Wagner *et al.*, 2007). Finally, *E. coli* lacking cytochrome oxidase genes could be adaptively evolved to obtain the same growth rate and GUR under aerobic conditions as under anaerobic conditions (Portnoy *et al.*, 2008), corroborating our proposed theory of a competition for membrane availability between the substrate uptake and respiratory enzymes.

To study the physiological effects of cytoplasmic membrane constraint, we define the fraction of the available membrane required for a reaction to achieve the reaction rate of 1 mmol/gdw/h as the ‘relative membrane cost’ of the reaction. Based on the theoretical and experimental evidence discussed above, we propose that prokaryotic organisms manage the expression of transmembrane proteins economically so that the fitness is maximized while maintaining membrane integrity (see Materials and methods).

### 2.3.2. Modeling *E. coli* physiology

To illustrate that the ‘membrane economics’ theory could satisfactorily explain the physiological changes associated with the respiro-fermentation phenomenon in *E. coli*, we modified the genome-scale metabolic model of *E. coli* (Feist *et al.*, 2007) to include the cytoplasmic membrane constraint. In *E. coli*, the cytoplasmic membrane is shared by a host of proteins, many of which are not directly involved in ATP production. As a simplification, our model assumes that only the expression of the glucose transporter and the three cytochromes (Cyo, Cyd-I, Cyd-II) are constrained by the availability of the cytoplasmic membrane. The relative membrane costs of the glucose transporter and the three terminal cytochrome oxidases (Cyo, Cyd-I, Cyd-II) were calculated from the growth data of various cytochrome- knockout strains (see Materials and methods). The three oxidases have varying turnover rates and energetic efficiency defined in terms of the proton translocation stoichiometry and the associated ATP yield (Portnoy *et al.*, 2008).
Since the ‘relative membrane cost’ of an enzyme is inversely related to its turnover rate, the fast and inefficient Cyd-II (Bekker et al, 2009) has a much lower cost than the slow and highly efficient Cyo. The relative cost of the moderately efficient Cyd-I is similar to that of Cyo under fully aerobic conditions; however, this cost becomes much less than that of Cyo under microaerobic (here, a microaerobic environment is defined as an environment where the oxygen concentration is 0.01 mM) conditions due to Cyd-I’s high affinity for oxygen (Tseng et al, 1996; Bekker et al, 2009).

**2.3.3. Aerobic physiology in nutrient-excess and nutrient-limiting environments**

Simulating the metabolism using the conventional FBA in a nutrient-rich environment requires imposing artificial constraints on certain uptake fluxes, such as glucose and OURs, in order to get meaningful output (Figure 2-1B). This requirement arises because FBA does not account for the tradeoff between catabolic rate and ATP yield—the growth rate increases unboundedly with the catabolic rate (when glucose is the sole carbon source, the GUR is a representation of the overall catabolic rate) in an unconstrained FBA model (Figure 2-1A). Constraining the occupancy of glucose transporters or the cytochrome oxidases on the cytoplasmic membrane, in an otherwise unconstrained FBA model, substantially decreases the solution space (Figure 2-1C). As a result, an ‘optimal solution’ can be found along the line of optimality without any constraints on fluxes (Figure 2-1C). Using ‘relative membrane costs’ calculated from experimental data, the new modeling framework—FBA with membrane economics (FBA\textsuperscript{ME})—predicted that wild-type *E. coli* has a GUR of 10.7 mmol/gdw/h, an OUR of 15.8 mmol/gdw/h, and a specific growth rate of 0.69 per hour during aerobic growth with excess glucose. FBA\textsuperscript{ME} also predicted that under the same growth condition, an *E. coli* knockout strain with no cytochromes has a GUR of 18 mmol/gdw/h and growth rate of 0.42 h\textsuperscript{-1}. These values agree very well with the observed experimental values for *E. coli* grown in batch cultures (Vemuri et al, 2006; Portnoy et al, 2008), which supports our argument that the higher
GUR of *E. coli* during anaerobiosis than under aerobic conditions is due to the absence of the respiratory enzymes.

---

**Figure 2-1. The solution spaces of FBA and FBA\textsuperscript{ME}**

(A) Solution space of an unconstrained *E. coli* model. (B) Yield predicted by FBA. (C) Yield predicted by FBA\textsuperscript{ME}. (D) Energy production pathway predicted by FBA. (E) Energy production pathway predicted by FBA\textsuperscript{ME}. The color in panels (A–C) indicates the growth yield. The color in panels (D, E) indicates the predicted energy production pathway—red for fermentation, blue for respiration through Cyo, and green for respiration through Cyd-II. In panels (B, C), The pink circle indicates the solution with optimal growth yield, and the brown circle indicates the solution with optimal growth rate. The shape of the solution space is different between FBA and FBA\textsuperscript{ME}; the utilization of Cyd-II is predicted by FBA\textsuperscript{ME}, but never predicted by FBA.
We also simulated the aerobic growth of *E. coli* in a glucose-limited chemostat using both conventional FBA and FBA\textsuperscript{ME}. FBA\textsuperscript{ME} successfully predicted the growth rate and yield changes with respect to increasing GUR (Figure 2-2A, B), as well as the aerobic production of acetate (Figure 2-2C) and concomitant repression of oxygen uptake (Figure 2-2D). On the other hand, traditional FBA significantly overestimated the growth rate and yield at higher GURs (this overestimation cannot be explained by varying the growth-associated maintenance energy parameter; Figure 2-2A), and failed to predict the reduction in oxygen uptake at higher GURs (Figure 2-2). Furthermore, FBA\textsuperscript{ME} predicted the selective expression of Cyo and Cyd-II at lower uptake rates (Figure 2-1E and Figure 2-3A, B) as well as the reduction of the TCA cycle activities at higher uptake rates (Figure 2-3C, D). The predictions from FBA\textsuperscript{ME} are in excellent agreement with the gene expression data from a glucose-limited chemostat (Figure 2-3), and the utilization of Cyd-II explains the subtle decrease in respiratory efficiency observed in experiments. Given the simplicity of the constraint we imposed, our model predictions agree surprisingly well with experimental observations, lending strong credibility to the membrane economics hypothesis.

It should be noted that while the FBA\textsuperscript{ME} was able to predict the production of acetate, it did not predict the correct timing of the onset of acetate production using the membrane cost parameters estimated in Section 2.2.3. This is because the timing of the onset is quite sensitive to the cost parameters (Appendix III-1). The accurate prediction of this time would require detailed cost parameters for all membrane proteins, which we did not have. However, even with some variations in the cost parameters, the overall trends were predicted (Appendix III-1).
Figure 2-2. Prediction of *E. coli* Aerobic Growth Physiology

The predicted yield (A), growth rate (B), acetate production rate (C), and oxygen production rate (D) as a function of the GUR. Asterisk indicates experimental measurements obtained by Vemuri et al (2006). The lines refer to the predictions of the FBA$^{ME}$ (red solid), FBA with low GAM (blue dashed), and FBA with high GAM (green dash-dot).
2.3.4. Expression of cytochromes and activation of fermentation pathways

Having validated the concept of membrane occupancy, we studied the coordination of glucose transport with cytochrome oxidase activity, mediated by Cyo, Cyd-I, or Cyd-II at low, medium, or high rates of glucose uptake under aerobic or microaerobic conditions. FBAME indicates that *E. coli* regulates the expression of cytochrome oxidases based on the availability of glucose and oxygen. Under aerobic conditions, at lower GURs (3.2 mmol/gdw/h), the high cost of the efficient Cyo is not penalized because the membrane is not completely saturated—the glucose transporters share the membrane with the Cyo without competition (Figure 2-3A and Figure 2-4A). Consequently, both growth rate and yield increase with GUR (Figure 2-2A, B). As the GUR increases beyond a certain critical level (3.2 mmol/gdw/h), the membrane becomes saturated and any further increase in the GUR requires more glucose transporter proteins to be expressed at the expense of the cytochrome oxidases in the membrane. Thus, at medium GURs (3.2–8 mmol/gdw/h), the low-cost Cyd-II replaces the costly Cyo (Figure 2-3A, B, and Figure 2-4B) at the expense of respiratory efficiency, leading to an increase in growth rate but a decrease in growth yield (Figure 2-2A, B). Once the GUR surpasses a second critical level (8 mmol/gdw/h), the membrane is occupied predominantly by Cyd-II. A further increase in the GUR necessitates some Cyd-II be replaced by the glucose transporters (Figure 2-4C), decreasing the rate of respiration (Figure 2-2D). As a result, a portion of the glucose is metabolized through the fermentative pathways (Figure 2-2C and Figure 2-4C), leading to a significant loss in energetic efficiency and biomass yield (Figure 2-2A). Nonetheless, during this phase, the growth rate continues to increase (Figure 2-2B). Maximal growth rate is reached when the ATP gain from increased GUR can no longer offset the ATP loss from the efficiency reduction (Figure 2-1C and Figure 2-2B).
Figure 2-3. Prediction of Gene Expression

Comparison between FBA predictions and experimentally measured gene expressions of Cyo (A, E), Cyd-II (B, F), succinate dehydrogenase (C, G), and succinyl-CoA synthetase (D, H). In panels (A, B), relative expression is the portion of the available membrane (total=1) used by a protein; this value is directly related to the abundance of the protein on the membrane. Symbols are used in panels (E–H) to distinguish the gene expressions of different subunits (panel (E): star1=cyoA, diamond=cyoB, circle=cyoC, triangle=cyoD; panel (F): star=appB, diamond=appC, circle=appY; panel (G): star=sdhC, diamond=sdhD; and panel (H): star=suchC, diamond=sucD). Gene expression data was obtained from NCBI GEO using the accession number GSE4366.
Figure 2-4. Environmental Conditions and Membrane Occupancy.

The *E. coli* cytoplasmic membrane occupancy by glucose transporters (brown), Cytochrome Cyo (blue), Cyd-I (red), and Cyd-II (green) under aerobic low GUR (A), aerobic medium GUR (B), aerobic high GUR (C), and microaerobic conditions (D). In this figure, the spheres represent the membrane enzymes, and the gray line underneath the spheres represents the cytoplasmic membrane. The arrows represent the metabolic flow.
2.3.5. Expression of Cyd-I under microaerobic growth conditions

Under aerobic conditions, FBAME predicts that Cyd-I is not used because its membrane cost is similar to that of Cyo but it has a much lower energetic efficiency. However, under microaerobic conditions, the membrane cost is inversely related to oxygen affinity of the enzymes (Equation 2-6), making Cyd-I the preferred cytochrome (Figure 2-4D, Figure 2-5) due to its high oxygen affinity (Puustinen et al., 1991; Govantes et al., 2000; Bekker et al., 2009). Conventional FBA cannot predict *E. coli*’s preference for Cyd-I under microaerobic conditions because it cannot take oxygen affinity into account. Using FBAME, we have simulated the oxygen limited chemostat growth of *E. coli* by fixing the growth rate at various dilution rates (0.1 – 0.5 h⁻¹). The C* values for the cytochromes under microaerobic conditions are calculated using Equation 2-5 and Equation 2-10 as well as the measured Kₛ values for the cytochromes (Bekker et al., 2009). In all conditions studied, Cyd-I is preferred at low oxygen concentrations, while Cyo is preferred at high oxygen concentrations (Figure 2-5). This agrees with the findings of a similar oxygen-limited chemostat experiment with *E. coli* strain MC4100 (Figure 2-5, Tseng et al. 1996).

The transcriptional regulation of Cyo and Cyd-I expression is achieved in *E. coli* through ArcA. ArcA represses the expression of Cyo and induces Cyd-I (Tseng et al., 1996). As such, in ArcA knockout strains, Cyo is expressed instead of Cyd-I. Since Cyo is more costly than Cyd-I under microaerobic conditions, this would lead to a decrease in glucose transporter expression. Indeed, a decrease in GUR is indeed observed in ArcA knockout strains grown under microaerobic condition (Nikel et al., 2009).
Figure 2-5. Predicted cytochrome expression of *E. coli* grown in oxygen-limited chemostat.

Predicted cytochrome expression of *E. coli* grown in oxygen-limited chemostat at different dilution rates (A-E) compared to the measured cytochrome expression of *E. coli* grown in oxygen-limited chemostat (F). Cyd-II measurement is not available in the experimental data.
2.3.6. Rejection of the Molecular Crowding Theory

Previously, it has been claimed that FBA with a molecular crowding constraint (FBAwMC) is capable of predicting the acetate overflow in *E. coli* (Vazquez et al, 2008). The FBAwMC theory suggests that fluxes of all metabolic reactions are constrained by the limited intracellular volume. We evaluated the impact of this constraint by introducing the molecular crowding constraint to the iAF1260 model (Feist et al, 2007). (Note: Vazquez et al. used an older model of *E. coli*). We tuned the average crowding coefficient to achieve a growth rate of 0.68. While the FBAwMC was able to predict acetate overflow, it cannot predict growth yield or rate (Figure 2-6). More importantly, FBAwMC cannot predict the acetate overflow if we set the crowding coefficient of the ETC enzymes (specifically, the three cytochromes) or the crowding coefficient of all membrane proteins with metabolic functions (glucose transporters and the ETC enzymes) to zero. For rod-shaped prokaryotic cells such as *E. coli*, the volume of the membrane is relatively small compared to the total volume of the cell, and only a portion of the membrane-bound proteins have metabolic functions. As such, if the acetate overflow is caused by a volumetric crowding, FBAwMC should be able to predict the respirofermentation phenomenon without the inclusion of the membrane proteins into its formulation. The inability of FBAwMC to predict acetate overflow without the inclusion of membrane proteins directly suggests that the available membrane area is the true constraint on cellular metabolism.
Figure 2-6. The Rejection of Molecular Crowding Theory

Intracellular volume constraint cannot explain the respiro-fermentation phenomenon. FBA framework based on such a constraint (FBA\textsuperscript{wMC}) cannot predict the fermentative production of acetate if we set the crowding coefficient of the ETC enzymes or the crowding coefficient of all membrane proteins (transporters and ETC enzymes) to zero. Since the membrane proteins take up negligible intracellular volume, this indirectly suggests that the available membrane area is the true constraint on cellular metabolism.
2.4. Discussion

2.4.1. How Crowded Is the Cytoplasmic Membrane?

Despite the high degree of consistency between our model prediction and experimental observations, as well as various indirect evidence (Wagner et al, 2007), the “coverage” (fraction occupied by proteins) of the cytoplasmic membrane has not been directly confirmed by targeted membrane studies. However, it has been estimated that glucose transporters take up more than 4% of E. coli membrane area (Phillips & Milo, 2009). Assuming 50% of the membrane is composed of lipids (Molenaar et al, 2009), glucose transporters alone occupy more than 8% of the membrane area available for protein expression. Similar estimates suggest that at the highest oxygen uptake rate (18 mmol/gdw/h), about 15% of E. coli’s cytoplasmic proteins are cytochromes while under microaerobic conditions, about 11% (Appendix III-5). Furthermore, estimates based on simulated ATP requirement suggest that about 13% of cytoplasmic proteins are ATP synthase during optimal aerobic growth (Appendix III-5). These estimates suggest an extremely crowded cytoplasmic membrane where various enzymes compete for the available membrane area (Phillips & Milo, 2009). In the future, the measurements of the membrane saturation level and the abundance of various membrane enzymes may serve as a direct confirmation of our hypothesis, and may be useful for the acquisition of more accurate model parameters.

2.4.2. Linking Physiology and Morphology in Bacteria

Our simulations are consistent with the hypothesis that E. coli’s regulatory program has evolved to ensure the efficient utilization of the finite cytoplasmic membrane. The size of E. coli’s specific cytoplasmic membrane available for transmembrane protein expression is directly related to its surface area to volume (S/V) ratio (see Methods), which is greatly affected by the cell morphology; thus, this model allows for the first time, the ability to analyze the relation between morphology and physiology of this organism. For a rod-shape cell, the S/V ratio is very sensitive to the change in its radius, but insensitive to the
change in its length. Throughout the growth phase, the rod-shaped *E. coli* grows primarily by elongation (Begg & Donachie, 1985); this ensures that the same regulatory program remains useful through the growth process despite the necessary morphological changes during binary fission. During energy starvation, *E. coli* becomes spherical and much smaller (Lange & Hengge-Aronis, 1991); the resultant increase in the S/V ratio allows *E. coli* to scavenge for multiple nutrient sources simultaneously using different transporters without sacrificing efficiency.

Morphological control of S/V ratio may be a common strategy in bacteria to obtain the desired rate of respiration. For example, *Geobacter sulfurreducens* is perhaps capable of obtaining much higher rate of respiration from acetate than *Rhodoferax ferrireducens*, a similar iron-reducer, (Esteve Núñez et al, 2005; Risso et al, 2009; Zhuang et al, 2011a) due to its smaller size and the resultant high S/V ratio. The free energy-starved *Dehalococcoides* spp. may have evolved its disk shape in order to maximize the dechlorination rate given the low thermodynamic efficiency associated with the dechlorination process (Jayachandran et al, 2004).

### 2.4.3. Membrane Constraints in Eukaryotic Cells

Given the generic nature of the constraint we proposed, it is only rational that economics of membrane occupancy also be applicable to eukaryotic metabolism since eukaryotes also exhibit aerobic fermentation. Eukaryotes are generally larger than prokaryotes. While the increase in volume accommodates the greater level of complexity, the downside is that it leads to a significant reduction in S/V ratio. Therefore, it is plausible that one reason for compartmentalization of eukaryotic metabolism is to increase the membrane availability. We believe that the concept of competition to occupy transmembrane space by different proteins is also prevalent in eukaryotes as well, although the competing proteins, nature of competition and the direct outcome of the competition are far more complex and warrant further investigation. Two independent observations substantiate this idea. The first is a positive correlation between respiration and the mitochondrial membrane area in yeast (Visser et al, 1995) and higher organisms
(Bicudo & Zerbiniatti, 1995). The second is the implication that glucose uptake in yeast may be limited by the cytoplasmic membrane area (Phillips & Milo, 2009).

In eukaryotes, glucose transporters are located on the cytoplasmic membrane and the cytochromes are located on the mitochondrial membrane. Once transported, glucose is oxidized to pyruvate. The subsequent fate of pyruvate is dependent on whether it is transported into the mitochondria or converted to ethanol (yeast) or lactate (mammals). Analogous to the competition between glucose transporters and cytochromes for space on the cytosolic membrane in prokaryotes, it is reasonable to argue that cytochromes compete with pyruvate transporter for space on the mitochondrial membrane in eukaryotes. We hypothesize that the outcome of this competition plays a large role in controlling respiration and fermentation. Clearly, additional studies are required to determine whether the membrane economics theory alone is sufficient to explain the respiro-fermentation phenomenon in eukaryotes.
2.5. Concluding Remarks

Although it has been long suggested that cellular evolution is governed by non-adjustable mechanistic constraints (Palsson, 2000; Papin et al, 2005; Novak et al, 2006), to date, most metabolic models rely on empirically derived parameters such as glucose and oxygen uptake rate. We showed that complex phenomena, such as the respiro-fermentation in bacteria, could be described by a simple mechanistic constraint on membrane enzyme occupancy. Furthermore, we showed that this simple morphological constraint dictates the regulation of E. coli’s metabolism, thus establishing a direct link between cell morphology and physiology. Finally, we hypothesize that this mechanism might be a critical factor governing eukaryotic metabolism and the evolution of mitochondria.

By demonstrating that FBA with Membrane Economics is capable of explaining and predicting the overflow metabolism in E. coli, we verified the validity of FBA in nutrient-rich conditions. As a result, we can more confidently apply FBA-based methods in many biotechnology applications that require nutrient-rich environments, such as fermentation and uranium bioremediation. Further efforts to elucidate additional fundamental cellular constraints as well as the underlying design principles could significantly improve our understanding of the regulation and evolution of metabolism.
Chapter 3. The Dynamic Multi-species Metabolic Modeling framework

3.1. Introduction

Microorganisms in nature exist in complex communities, either in cooperation or in competition. The composition of the community and the metabolic states of its members are highly sensitive to the ever-changing environment. Furthermore, community activities can modify their environment, which further modifies community composition and behavior. For example, the rumen microbial community composition varies significantly based on the host’s dietary input (Tajima et al., 2001), and the community’s digestive performance is decided by the community’s composition. In addition to natural communities, there are many engineered microbial communities designed for both engineering and scientific purposes such as consolidated bio-production (Salimi et al., 2010). Although several past and recent efforts (Stolyar et al., 2007; Vallino, 2003; Zomorrodi & Maranas, 2012) have sought to extend the constraint-based approach to the modeling of microbial communities, the methods used in these attempts are inadequate for complex microbial communities such as communities relevant to subsurface bioremediation (Zhuang et al., 2011a) and consolidated bio-ethanol production (Salimi et al., 2010). The Vallino model (Vallino, 2003) is too simplistic, and the Stolyar and Zomorrodi models (Stolyar et al., 2007; Zomorrodi & Maranas, 2012) neglect the dynamics of the microbial communities. As a result, the applicability of the constraint-based metabolic modeling approach in both environmental biotechnology as well as technologies such as consolidated bio-production is limited.

In order to extend the constraint-based approach to the modeling of community metabolism, the Dynamic Multi-species Metabolic Modeling (DyMMM) framework was

---

2 This chapter is adapted from the supplementary information of the article “Genome-Scale Dynamic Modeling of the Competition Between Rhodoferax and Geobacter in Anoxic Subsurface Environments” published in Nature ISMEJ in 2011.
developed. The rest of this chapter provides a detailed description of this novel, and delineates its advantages compared to other modeling approaches.
3.2. The formulation of the DyMMM framework

A community metabolic model must account for the metabolic exchanges between species and with the environment, as well as the changes in biomass of the modeled species. The Dynamic Multi-species Metabolic Modeling (DyMMM) framework extends the dynamic flux balance analysis (dFBA) formulation to the community realm. A mathematical formulation of the DyMMM framework describes a community of $N$ microbial species coexisting in an environment containing $M_{EX}$ metabolites.

The growth rate ($dX/dt$) of every microbial species in the community is given by:

$$
\frac{dX_j}{dt} = \mu_j X_j
$$

Equation 3-1

The consumption/production rate ($dS/dt$) of every metabolite in the environment is given by:

$$
\frac{dS_i}{dt} = \sum_{j=1}^{N} v^i_j X_j
$$

Equation 3-2

Here, $X_j$ is the biomass of the $j^{th}$ species in the community; $S_i$ is the concentration of the $i^{th}$ metabolite in the environment. The specific growth rate ($\mu_j$) of the $j^{th}$ microbial species, and $v^i_j$, the specific consumption/production rate of the $i^{th}$ metabolite due to the actions of the $j^{th}$ microbial species, are calculated using the FBA:

Maximize $\mu_j = c^T v_j$

subject to $A_j v_j = 0$

$\nu_j^{\min} \leq v_j \leq \nu_j^{\max}$

Equation 3-3
\[ 1 \leq i \leq M_{\text{ex}} \text{ (number of metabolites)} \]
\[ 1 \leq j \leq N \text{ (number of microbial species)} \]

In Equation 3-3, \( c^T \) is the objective function; in our simulations, growth maximization is used as the objective function. \( A_j \) is the stoichiometric matrix of the \( j^{\text{th}} \) microbial species. \( v_j \) is the reaction flux vector of the \( j^{\text{th}} \) microbial species. \( v_j^{\text{max}} \) and \( v_j^{\text{min}} \) are the flux capacity constraints of the \( j^{\text{th}} \) microbial species based on the corresponding genome-scale model. For the components of \( v_j^{\text{max}} \) and \( v_j^{\text{min}} \) corresponding to external metabolites, the uptake/production constraints \( (V_j^{\text{max}}, V_j^{\text{min}}) \) to these fluxes can be calculated based on the environmental concentration of these metabolites, using either the On/Off method (if the uptake kinetics are not available) or the Michaelis–Menten kinetics method (if kinetics are available). A separate FBA model is used for each microbial species in the community – \( N \) FBA models are required for a community of \( N \) member species – each FBA model maximizes its own objective function.

By integrating the growth rates of all microbial species within the community, as well as the production/consumption rates of all metabolic species in the environment, the DyMMM framework can dynamically predict the temporal changes in metabolite and biomass concentrations in a complex microbial community.

A small sample implementation of the DyMMM framework is included in the next section to further illustrate the simulation process. The DyMMM framework is implemented in MATLAB as an add-on to the COBRA toolbox (Zhuang et al, 2011a; Becker et al, 2007). This allows the user to easily incorporate new species into the DyMMM framework.
3.3. Sample Implementation of the DyMMM framework

A sample implementation of the DyMMM framework is provided here to clarify the algorithm. This sample implementation assumes the existences of a community consisting of two microbial species, A and B, and three substrates, $\text{Met}^1$, $\text{Met}^2$, $\text{Met}^3$ (Figure 3-1). The numbers of microbial and metabolic species are arbitrary in this example. The maximal numbers of microbial and metabolic species depend on the limitation of the computational hardware. After a simulation length (hrs) and a simulation step size (hrs) is given, this particular implementation of the DyMMM framework execute the following routines during each simulation step:

R1: The concentrations of $\text{Met}^1$, $\text{Met}^2$, and $\text{Met}^3$ are used to calculate the upper and lower constraints to $V_{1A}$, $V_{2A}$, $V_{3A}$, $V_{1B}$, $V_{2B}$, and $V_{3B}$, where $V_{ij}$ represent the $i$th reaction flux of the species $j$. Two calculation methods are commonly used:

i. On/Off methods

if $[\text{Met}^i] > 0$, then $-\infty < V_{ij} < \infty$

if $[\text{Met}^i] \leq 0$, then $V_{ij} = 0$

j. Michaelis–Menten kinetics

\[
V_{ij}^{\text{constraint}} \leq \frac{V_{i,\text{max}} \cdot S_i}{S_i + K_i'}
\]

Equation 3-4

R2: The FBA models of species A and B calculates the specific growth rates, $\mu^A$ and $\mu^B$, as well as the reaction fluxes $v_1^A$, $v_2^A$, $v_3^A$, $v_1^B$, $v_2^B$, and $v_3^B$. The models can be described by the following equations.
Maximize $\mu_j = c^T v_j$
subject to $A_j v_j = 0$
$\nu_j^{\min} \leq \nu_j \leq \nu_j^{\max}$

$1 \leq i \leq M_{EX}$ (number of metabolites)
$1 \leq j \leq N$ (number of microbial species)

The constraints to the reaction fluxes (including the uptake constraints) and the reaction network itself forms a viable solution space, and the point in this space where specific growth rate is maximized is selected using linear programming. If the linear programming method provides a solution, then continue to routine R3. If the linear program method fails for species j (indicating that the cells do not have the required nutrients to survive under the conditions at this time), then a special cell-death routine RD is executed to calculate the rate of death $r^d$ of the species j, and $\mu^i$ is set to $r^d$.

R3: The rate of change in the biomass of A and B are calculated with Equation 3-1:

Calculate $\Delta X_j$ by integrating $dX_j/dt$ over the simulation step.

Calculate the new $X_j$ at the end of the simulation step with $X_j^{\text{new}} = X_j^{\text{old}} + \Delta X_j$.

R4: The rate of change in the concentration of external metabolite i is with the Equation 3-2:

Calculate $\Delta S_i$ by integrating $dS_i/dt$ over the simulation step.

Calculate the new $S_i$ at the end of the simulation step with $S_i^{\text{new}} = S_i^{\text{old}} + \Delta S_i$.

---

3 While a FBA model may have multiple equally optimal solutions, they share the same set of exchange fluxes. Since the DyMMM framework focuses on modeling the interactions between these exchange fluxes, it does not matter which one of the optimal solutions the solver picks.
Only the rates of change of external metabolites are integrated since all internal metabolites still follow the internal steady state assumption.

RD: The cell death RD routine is a special user-defined routine that is called upon when the environmental concentrations of substrates are insufficient to sustain the current biomass concentration of the organism. This routine is flexible and can be redefined for each individual organism to reflect its specific mechanism of cell-decay. The RD routine produces the rate of cell death \( r^d \), which should be considered a negative growth rate. If it is unclear how the cell death process occurs, this routine can be left empty. (This approach was adopted in Chapter 4 and 5).

After completion of RD, return to R3.

After routine R4 is completed, time advances one step-size. The routine R1 is initiated again. This continues until the simulation length is reached. At each time point, the flux constraints to each organism varies based on the substrate concentration at that particular time, and leading to dynamic variations in the growth rate. This procedure is illustrated in Figure 3-1.

Note: The step-size parameter is determined by the modeler. Choosing a smaller step-size will increase the accuracy at the cost of increased simulation time.
Figure 3-1. The Dynamic Multi-species Metabolic Modeling Framework

The DyMMM framework consists of four major steps:  
[2] Solve the FBA problems for the member species of the community. This generates the specific growth rates and external reaction fluxes of the member species. If the FBA problem is infeasible for one or more of the member species, then a special cell-death simulation routine is used to generate the specific death rate.  
[3] Calculate dX/dt for each member species. Integrating over this rate to generate the dynamic profile of biomass concentrations.  
[4] Calculate dSi/dt for each member species. Integrating over this rate to generate the dynamic profile of metabolite concentrations.
3.4. Advantage of the DyMMM framework

The usage of the DyMMM framework to model microbial communities holds significant advantages over the traditional Monod-kinetic models, the community metabolic models developed by Stoylar and Zomorrodi (Stolyar et al., 2007; Zomorrodi & Maranas, 2012) mentioned in the introduction (Section 3.1), as well as pure experimental methods.

3.4.1. DyMMM vs. Monod-kinetic Model

The Monod-kinetic model can model relatively simple metabolism quite well. For example, if an organism is treated as having only one limiting substrate, $S$, then the organism’s growth with respect to $S$ can be modeled fairly accurately by:

$$
\mu = \mu_{\text{max}} \cdot \frac{[S]}{[S] + K_s} \quad \text{Equation 3-5}
$$

If an organism is treated as having two limiting substrates, $S_1$ and $S_2$, then the organism’s growth can be modeled with:

$$
\mu = \mu_{\text{max}} \cdot \frac{[S_1]}{[S_1] + K_{s_1}} \cdot \frac{[S_2]}{[S_2] + K_{s_2}} \quad \text{Equation 3-6}
$$

For an organism with $i$ limiting substrates, the organism’s growth is modeled with the General Monod Model:

$$
\mu = \mu_{\text{max}} \cdot \prod_{i=1}^{i} \frac{[S_i]}{[S_i] + K_{s_i}} \quad \text{(Stolyar et al., 2007)}
$$

The issue with the Monod formulation is that if multiple substrates are modeled ($n>1$), and both substrates are available at a low concentration, the growth rate is over-penalized. For example, if each of the substrates are available at a concentration equal to
their half-saturation constant ($K_s$), then the Monod model would predict a growth rate of $1/2^n$ of the maximum because each Monod expression has a value of $1/2$. This is incorrect because if the carbon source uptake rate is at $1/2$ of its maximum, the organism’s requirements for all other substrates are proportionally decreased by $1/2$. The proper reduction in the requirements for these substrates is ensured in the constraint-based model by the mass-balance constraint. Since the DyMMM model consists of constraint-based models of the individual organisms, it does not experience the over-penalization issue of the Monod model.

Moreover, in the Monod model, the yield in the presence of multiple substrates needs to be experimentally determined for each combination, as there are no known mechanisms to compute the yields accurately for different metabolic states. Since $n$ is generally related to the metabolic complexity, the Monod-kinetic model does not scale up well with complex metabolisms. Realistically, prediction accuracy often becomes unacceptable when $i > 2$. Since the DyMMM framework is based on constraint-based models, it is capable of handling significantly more complex metabolisms. For example, the *Geobacter sulfurreducens* model used in this paper contains 727 reactions, 55 of which are exchange reactions ($i = 55$). It is also important to note that the metabolic network of *Geobacter* has many modes of operation such acetate-limiting ammonium-utilization mode, iron-limiting ammonium-utilization mode, acetate-limiting nitrogen fixation mode, iron-limiting nitrogen fixation mode. The DyMMM framework automatically selects the modes of operations based on the substrate concentrations at each time point. This is unachievable for traditional models.

### 3.4.2. DyMMM vs. Stolyar Model

Stolyar *et al.* published the first constraint-based metabolic model of a mutualistic microbial community (Stolyar *et al.*, 2007). In this model, the CBMs of *Desulfovibrio vulgaris* and *Methanococcus maripaludis* are directly connected. The authors suggested that because the species are interdependent, the objective function should be the maximization of the weighted sum of their biomass production fluxes. They tested three
arbitrary sets of weights: 10:1, 1:1, 1:10. The rationale behind this objective function is very weak, and the objective function is inappropriate for most microbial communities. In comparison, the DyMMM framework does not rely on an arbitrary objective function – instead, the FBA problems representing each organism are solved separately, allowing the usage of more established objective functions such as maximization of biomass flux.

Another issue of the Stolyar model is that it cannot predict the dynamic shifts in population and their metabolite concentrations. While this approach may be appropriate when the microorganisms are inter-dependent, it is inappropriate in ecological settings where the community composition is dynamic. Since the DyMMM framework is capable of predicting the community dynamics, it adequately addresses this issue.

### 3.4.3. DyMMM vs. Zomorrodi Model

More recently (after the publication of the DyMMM framework), Zomorrodi et al. have used a bi-level optimization scheme to model community metabolism (Zomorrodi & Maranas, 2012). This approach assumed that there exists an overall objective function for the microbial community; while this assumption may be appropriate for modeling symbiotic communities, its validity is questionable for competitive communities. In a competitive community, the organisms have no cause for cooperating; therefore, there is no need for an overall objective function. In addition, similar to the Stolyar model, the Zomorrodi model neglected the community dynamics – the changes in the community composition and the environmental metabolite concentrations were not modeled. In comparison, the DyMMM framework models the community dynamics, and does not assume an overall community objective. This approach makes the DyMMM framework more appropriate for competitive (Chapter 4, 5) and cross-feeding (Salimi et al., 2010) communities. To successfully model syntrophic communities, it will be necessary to combine the Zomorrodi’s bi-level approach with the DyMMM framework.
3.4.4. DyMMM vs. Experimental Studies

While no model can ever achieve the accuracy of real experimental studies, usage of metabolic modeling has proved to be a valuable analytical tool. The DyMMM framework allows scientists to consider thousands of possible reactions and metabolites simultaneously; this, together with necessary experimental work, allows for a more through and systematic analysis of the underlining causes than otherwise possible. In particular, it is often easier to see the links between system-level phenomena and lower-level characteristics. For example, we used a DyMMM model to help discover that the relative abundance of Geobacter and Rhodoferax species is dependent on how their metabolism interacts with the availability of acetate and ammonium (Zhuang et al., 2011a).

Another benefit of the DyMMM framework is that it allows for complex in silico experiments that would be difficult or costly to set up in a real experimental setting. For example, using a DyMMM model, we studied the effect altering the initial biomass ratio of two subsurface microorganisms, a task that is very difficult to accomplish experimentally (Chapter 4). Similarly, it is often faster and cheaper to perform in silico experiments prior to real experiments in order to filter out incorrect hypotheses and unpromising paths. For example, if a scientist designing an environmental biotechnology strategy wants to study the effect of increasing groundwater flow rate, it is better to model the scenario prior to pumping water into the ground.

Lastly, from an engineering perspective, it is possible to connect the DyMMM model to a mathematical optimization framework in order to find, algorithmically, the best way to engineer a particular system (Chapter 5).
3.5. Concluding Remarks

The DyMMM framework was first developed as a part of Kai Zhuang’s Masters Thesis. The original DyMMM framework was first applied to modeling of the community of Geobacter and Rhodoferax species during uranium bioremediation (Zhuang et al., 2011a), and later on, it was applied to the modeling of an artificially evolved syntrophic community between Geobacter sulfurreducens and Geobacter metallireducens (Zhao et al., unpublished). Subsequently, the DyMMM framework has been modified and expanded in order to model different microbial communities, including a Clostridial community related to consolidated bioethanol and biobutanol production (Salimi et al., 2010), as well as the community of Geobacter and sulfate-reducing bacteria involved in bioremediation (Chapter 4, 5). Most recently, the DyMMM framework was used in conjunction with existing strain design algorithms to form a novel strain design strategy (Chapter 6). The next three chapters describe three distinct applications of this framework.
4.1. Introduction

Stimulating microbial reduction of soluble U(VI) to less soluble U(IV) with the addition of organic electron donors has shown promise as a strategy for preventing the spread of uranium in contaminated groundwater (Lovley, 2003; Wall & Krumholz, 2006; Williams et al., 2011). However, the added electron donors can also promote the activities of other microbial species, possibly hampering the effectiveness of bioremediation. For example, the addition of acetate to sulfate-rich groundwater at a uranium bioremediation study site in Rifle, CO, produced an initial Fe(III) and U(VI) reducing phase, in which Geobacter species predominated, followed by a sulfate-reducing phase during which Fe(III) and U(VI) reduction ceased and acetate-oxidizing sulfate reducers related to Desulfo bacter were more abundant (Anderson et al., 2003; Miletto et al., 2011). The transition from metal to sulfate reduction was accompanied by the increased abundance of sulfate-reducing bacteria (SRB), and took place between 40-60 days in both field and column studies (Anderson et al., 2003; Komlos et al., 2008a; N'Guessan et al., 2008; Vrionis et al., 2005).

Previous studies on the interactions between Fe(III) reducers and sulfate reducers have focused on the competition for electron donors that take place under near steady-state conditions in sedimentary environments. In such environments, acetate and other electron donors that support Fe(III) reduction and sulfate reduction are provided from the relatively slow hydrolysis and fermentation of complex organic matter (Lovley & Chapelle, 1995). Fe(III) reducers have a higher affinity for acetate and other electron donors than sulfate reducers, and, under steady-state conditions, can maintain the

---

4 This chapter is adapted from the article “Integrative Analysis of Geobacter spp. and sulfate-reducing bacteria during uranium bioremediation” published in Biogeosciences in 2012. The experimental (non-computational) work described in sections 4.2.1, 4.2.2, 4.2.3 were performed by Melissa Barlett. These are included in this thesis for completion purposes with the permission of Melissa Barlett.
concentrations of these electron donors too low for sulfate reducers to metabolize (Lovley & Phillips, 1987). If high concentrations of acetate or other electron donors are added to such sediments, then the competition for electron donors is relieved and Fe(III) reduction and sulfate reduction can proceed simultaneously (Lovley & Phillips, 1987).

Thus, it might be expected that the initial responses of *Geobacter* and *Desulfobacter* species to acetate amendments might be more comparable to each other, yet, the consistent observation is that *Geobacter* predominate in the early stages of acetate-driven *in situ* uranium bioremediation. An analysis of the competition between *Geobacter* and *Rhodoferax* species, which are also acetate-oxidizing Fe(III) reducers, demonstrated that although *Rhodoferax* could compete with *Geobacter* at the low rates of acetate production in unamended sediments, the addition of high concentrations of acetate favored *Geobacter* species which grow less efficiently but faster than *Rhodoferax* species (Zhuang *et al.*, 2011a). These results demonstrate that the selective pressures for competition are much different in environments with high concentrations of added electron donors versus environments in which electron donors are slowly provided via fermentation of complex organic matter.

In order to study the dynamic interactions between *Geobacter* and acetate-oxidizing SRB during uranium bioremediation, we have adopted an integrative approach that iteratively combined laboratory sediment experiments with the simulations of a DyMMM model of the *Geobacter* and SRB community. Similar integrative approaches have been used to elucidate many important features of *Geobacter* physiology and ecology in recent years (Izallalen *et al.*, 2008; Mahadevan *et al.*, 2006; 2010; Segura *et al.*, 2008; Sun *et al.*, 2009; Zhuang *et al.*, 2011a). The results of both experimental and modeling approaches suggest that *Geobacter* and SRB populations have little direct interaction and that the late appearance of SRB to added acetate can be attributed to a slower growth rate. This study is an example of how the DyMMM framework can be used to understand microbial interactions in an environmental setting.
4.2. Methods

4.2.1. Sediment Incubation with Acetate

Sediments and groundwater were obtained from a uranium-contaminated site at Old Rifle, CO, that has been described elsewhere (Anderson et al, 2003). Sediments and sterile, anaerobic groundwater were placed in glass bottles that were purged with a 95:5 N₂:CO₂ gas mixture to achieve anaerobic conditions and then sealed with thick butyl rubber stoppers. Sodium acetate was added to provide an initial concentration of 12 mM and the sediments were incubated at 16°C.

Sediments and groundwater were sampled under anaerobic conditions in a N₂-filled glovebag. Acetate, sulfate, ferrous iron and total iron were determined as described below. A mixed sediment/groundwater slurry was immediately fixed for fluorescent in situ hybridization (FISH) in 4% paraformaldehyde/0.5x phosphate buffered saline (PBS; 1X = 7.6 g NaCl, 1.9 g Na₂HPO₄·7H₂O, 0.7 g NaH₂PO₄·2H₂O per L) with 0.1% sodium pyrophosphate and 0.15% Triton-X (Sigma) and stored at 4°C. This solution was vortexed for five seconds and allowed to settle briefly to remove the majority of sediment particles before being added to 0.2µm white polycarbonate filters (GTTP; Millipore, Billerica, MA) and washed with 1% Nonidet (Sigma) solution. Cells were hybridized for three hours as described previously (Lemke et al, 1997; Pernthaler et al, 2001) with a formamide concentration of 35% (v/v). The slides were embedded in Vectashield (Vector Laboratories, Burlingame, CA) and observed with a Nikon epifluorescence microscope. Ten to twenty fields of view for each sample were enumerated.

The number of Geobacter was determined by the number of cells that hybridized the probes GEO3A, GEO3B, and GEO3C (Richter et al, 2007). The number of sulfate reducing bacteria was inferred by determining the number of cells that hybridized the probes SRB385 and SRB385Db (Amann et al, 1990) and subtracting the number of Geobacter. This was done because the SRB385 and SRB385Db probes include the majority of SRB and some other delta-proteobacteria including the Geobacter (Rabus et al, 1996). The probe NON338 was used to account for autofluorescence (Wallner et al, 1993).
4.2.2. Fe(III) Oxide Addition Experiment

After an initial experiment was run as above, bottles were incubated for over 2 months to fully reduce the bioavailable Fe(III) and sulfate in the sediments and groundwater. Once this initial reduction was complete, the groundwater was replaced with fresh groundwater containing natural concentrations of sulfate at about 8-10mM and additional acetate at a concentration of 2mM. The bottles were incubated at 16°C for 17 days. At this point, they were further amended with acetate (5mM) in sulfate-containing groundwater and two of the three bottles were amended with poorly-crystalline Fe(III) oxide at a concentration of 50mM. Acetate in groundwater was added on Days 23 and 27 to maintain concentrations between 2-5mM. The bottles were sampled every 2-3 days through Day 32 and samples were processed as above for geochemistry and bacterial cell numbers.

4.2.3. Analytical Methods

Groundwater samples for acetate and sulfate analyses were filtered through 0.2μm SFCA filters (Corning Inc.; Corning, NY) and measured on a Dionex ICS-1000 (Sunnyvale, CA). Fe(II) and Fe(III) in the water and sediments were determined with the ferrozine method as previously described (Lovley & Phillips, 1987) after extraction in 0.5N HCl for 24 hours.

4.2.4. Dynamic Metabolic Modeling of Geobacter and SRB

A dynamic community metabolic model containing attached Geobacter, planktonic Geobacter, and SRB was used to study the metabolic interactions between Geobacter and SRB. Attached Geobacter were assumed to be the sole reducer of Fe(III) and planktonic Geobacter is assumed to be the sole reducer of uranium (Zhao et al, 2010). This particular model utilizes the genome-scale metabolic model of Geobacter (Zhuang et al,
2011a) and a core-pathway model of SRB (see below). This model been shown to be capable of predicting the microbial activities during the 2002 field bioremediation experiments at the study site (Chapter 5).

**4.2.5. Core-Pathway Model of SRB**

Since the genome-scale metabolic of SRB is not available, we opted to construct a simple stoichiometric model of SRB based on the sulfate reduction stoichiometry (Equation 4-1). The measured biomass yield (4.3 gdw/ mol acetate), the measured maximal uptake rates (3.2 mmol/gdw/hr, 4.2 mmol/gdw/hr), and the saturation constants (0.077 mM, 0.064 mM) for acetate and sulfate found in the literature for *Desulfovibrio postgatei* (Ingvorsen et al, 1984). The stoichiometric model of SRB contains one core metabolic reaction (Equation 4-1) and the exchange reactions for metabolites involved. The exchange reactions for acetate and sulfate are constrained by Equation 4-2 and Equation 4-3. The value for biomass in Equation 4-1 is calculated by multiplying the biomass yield and 1 mol of Acetate.

\[
\begin{align*}
1 \text{ mol Acetate} + 1 \text{ mol SO}_4^{2-} & \rightarrow 2 \text{ mol HCO}_3^- + 1 \text{ mol HS}^- + 4.3 \text{ gdw Biomass} \\
\end{align*}
\]

\[
\begin{align*}
v_{Ac} & \leq \frac{3.2S_{Ac}}{S_{Ac} + 0.077} \\
\end{align*}
\]

\[
\begin{align*}
v_{SO4} & \leq \frac{4.2S_{SO4}}{S_{SO4} + 0.064} \\
\end{align*}
\]
4.2.6. Determination of Model Parameters

Using the methods described in section 2.3, the total concentration of iron in the sediment was determined to be 7.45 mM; about 70% (5.2 mM) of the iron has been determined to be Fe(III) which can be reduced by Geobacter, the remaining 2.25 mM are Fe(II) which is already in a reduced form and cannot be further reduced by Geobacter. Previous work has established that in the Rifle sediment, less than 3% of the Fe(III) are in the easy-to-use amorphous oxide form, the rest are in more hard-to-use forms including Fe(III) silicate, Al-geothite, and magnetite (Komlos et al, 2008a). Based on these experimental data, we calculated the initial concentrations of Fe(II), Easy-to-Use Fe(III), and Hard-to-Use Fe(III) to be 2.25 mM, 0.16 mM, and 5.1 mM respectively. In our model, Easy-to-Use and Hard-to-Use Fe(III) are treated as two different metabolites, whose dynamics are described separately. Our model does not contain the geochemistry, and it is assumed that there is no interchange between the Easy-to-Use and Hard-to-Use Fe(III).

The genome-scale model of Geobacter sulfurreducens has been updated to include separate pathways for the uptakes of Easy-to-Use Fe(III) and Hard-to-Use Fe(III). The uptake kinetics of acetate and Easy-to-Use Fe(III) for Geobacter have been described in a previous model (Zhuang et al, 2011a); however, this model did not include the Hard-to-Use Fe(III) pathway. The saturation constant in the Fe(III) utilization kinetics for the easy-to-use Fe(III) have been previously published to be 1 mM (Zhuang et al, 2011a); here, we identified the $V_{\text{max}}$ and $K_s$ for hard-to-use Fe(III) using the Fe(II) data from the 2002 Rifle experiment (Appendix IV-1) as well as the Fe(II) data from this work (Figure 4-2F). (Appendix IV-1). The $V_{\text{max}}$ for easy-to-use Fe(III) was set to 568 mmol/gdw/hr as previously described (Zhuang et al, 2011a). The $V_{\text{max}}$ for hard-to-use Fe(III) was estimated to be 30 mmol/gdw/hr using Fe(II) data.

Kinetic parameters for Desulfobacter postgatei, which is closely related to the majority of SRB that increase in abundance during the sulfate reduction phase (Miletto et al, 2011), were employed in the SRB modeling with the exception that the published kinetics in our experimental data (Figure 4-1) show that the Rifle SRB’s affinity for $K_s$ is much higher than that of the laboratory Desulfobacter strains during the batch sediment incubation. A
K_s value of 13 mM was estimated by fitting the parameter to the experimental sulfate data (Figure 4-2E).

A death rate of 0.0011 hr^{-1} was chosen for both organisms, which is based on the value of two previous models of sulfate and iron reducers (Bethke et al., 2008; Moosa et al., 2002; 2002). This value can describe the decay of Geobacter (Figure 4-2).

4.2.7. Simulations of Batch Incubation Experiments

To model the laboratory sediment incubation experiment with acetate amendment only, the initial concentrations of acetate and sulfate were set to 12 mM and 10.5 mM, and the initial cell concentrations of both organisms to $3.4 \times 10^4$ cells/ml. These values are representative of the measured experimental conditions during sediment incubation. Additions of acetate and sulfate were made in silico on day 65 to reflect the additions in the laboratory experiment.

Ideal batch reactor models were used to simulate conditions with varying starting proportions of Geobacter and SRB that were impossible to replicate in the lab. The initial cell number for all organisms was assumed to be $10^4$ cells/ml, which is similar to the cell number measured in the batch sediment incubation experiment. These cells were then divided between Geobacter and SRB for varying percentage starts including 100%, 90%, and 50% of each group. The SRB were determined to have twice as much mass/cell as the Geobacter as calculated from the cellular dimensions found in the FISH pictures of each group.

The model used to simulate Fe(III) amendment was initialized with the same conditions as in the acetate-amended batch incubation simulation except no additional acetate and sulfate was added later in the experiment. Two simulations were performed: 5 mM of Easy-to-Use Fe(III) was added on day 45 in one simulation and no Fe(III) was added in the other simulation.
4.3. Results and Discussion

4.3.1. Sediment Incubation with Acetate and its Simulation

The interaction of *Geobacter* species and SRB was evaluated in sediment incubations that simulated conditions in the field experiments, but provided the opportunity to quantify the number of cells in each population over time. With the addition of acetate, there was a rapid growth of *Geobacter* species (Figure 4-1), as previously observed in field studies (Anderson *et al.*, 2003; Vrionis *et al.*, 2005; Williams *et al.*, 2011). Adding acetate also stimulated the growth of SRB, but they grew more slowly (Figure 4-1). After day 24, the number of *Geobacter* species declined, coincident with reduction of ca. 80% of the Fe(III) in the sediment (Figure 4-1). However, the SRB continued to increase in number and became predominant (Figure 4-1). Addition of more acetate and sulfate as they became depleted stimulated additional growth of SRB, but not *Geobacter* species. This pattern of succession from *Geobacter* to SRB has previously been observed in field and column studies (Anderson *et al.*, 2003; Komlos *et al.*, 2008a; Miletto *et al.*, 2011) and the timing of the transition to SRB predominance in the batch studies reported here is similar to that seen in those previous studies.

These results demonstrated concurrent growth of both *Geobacter* species and SRB following the addition of acetate, but it was not possible to elucidate from these studies whether the metabolism of the two metabolic groups had an influence on each other. Therefore, the potential for such interactions was further investigated with a dynamic community metabolic model of *Geobacter* and SRB to investigate the metabolic interactions between these two organisms.

First, the sediment incubation study was modeled in order to check the model validity. The model was able to predict the growth of *Geobacter* species and SRB, and the uptake of acetate and sulfate, the reduction of Fe(III), as well as the evolution of the fraction of cells that are *Geobacter* species (Figure 4-2). Like the sediment incubation experiment, the simulation predicted that a batch addition of acetate would lead to the initial dominance of *Geobacter* species and the latter overtaking of *Geobacter* by the SRB.
The higher biomass yield and the maximum acetate uptake rate of Geobacter resulted in Geobacter growing faster than SRB when Fe(III) was abundant. However, the predicted growth rate of Geobacter decreased drastically once the Easy-to-Use Fe(III) was exhausted; its growth eventually stopped after the exhaustion of all available Fe(III). On the other hand, SRB grew slowly but steadily, overtaking Geobacter species between day 30 and 40. The growth rate of SRB did not vary before and after the exhaustion of Fe(III), suggesting that the microbial reduction of Fe(III) had little effect on the growth of SRB. Thus, despite its relative simplicity, the model was able to predict the shifts in community composition as well as the trends of metabolite utilization observed in the sediment incubations.

It is important to note that while the model was able to predict the general trends, it could not capture finer details. For example, three peaks were observed in the Geobacter and SRB numbers, which could not be predicted individually. It is possible that the peaks were caused by sampling errors due to the highly heterogeneous nature of the sediment. Another possibility is that different strains of Geobacter and SRB, which are known to exist within the Rifle sediment, grew at different rates; since we only modeled one species of Geobacter and SRB respectively, the different peaks could not be predicted. Lastly, the Rifle sediment contains several acetate-consuming species that we did not include in the model; although they are much less influential than Geobacter and SRB (Anderson et al, 2003; Vrionis et al, 2005), they can have a minor influence the community dynamics. Ultimately, these discrepancies between the model and the experimental data can help us improve our model as well as highlighting our knowledge gaps regarding the subsurface community.
Figure 4-1. Microbiological and geochemical impact of adding acetate to subsurface sediments.

A) The number of cells of *Geobacter* and sulfate-reducing bacteria (SRB) as determined by fluorescent *in situ* hybridization of probes GeoA/GeoB/GeoC and SRB385/SRB385Db respectively; B) Fe(III) reduction as indicated by an increase in the proportion of Fe(II) of the total acid-extractable iron in the sediment; C,D) The concentration of acetate (C) and sulfate (D) in the groundwater; The arrow denotes an addition of sterile groundwater to the batches that aimed to increase both acetate and sulfate by 5mM each. Data points are each an average of 5 separate batches; the error bars show the standard deviation.
Figure 4-2. Model predictions of community dynamics.

*In silico* predictions of trends in the fraction of *Geobacter* in the community (A), number of SRB (B), number of *Geobacter* (C), acetate concentration (D), sulfate concentration (E), and Fe(II) concentration (F). Lines are the *in silico* predictions; * are the experimental values. The experimental Fe(II) values are calculated by multiplying the total iron concentration of 7.5 mM with the experimentally measured Fe(II) percentage in total Fe(II) from Figure 4-1.
4.3.2. Simulations with Varying Initial Microbial Fractions

A potentially important consideration in understanding the interactions between *Geobacter* species and SRB following the addition of acetate to groundwater is the composition of the microbial community prior to acetate amendment. For example, previous molecular analysis of the subsurface community prior to the addition of acetate found that about 5% of the Rifle microbial community were *Geobacter* species, whereas sequences that could be attributed to acetate-using SRB were not detectable (Holmes *et al.*, 2005; Miletto *et al.*, 2011; Mouser *et al.*, 2009), possibly giving *Geobacter* an initial numerical advantage over the SRB. To mimic the natural variations in the abundance of *Geobacter* and SRB prior to the addition of acetate, simulations were run with a community that was composed of 0%, 10%, 50%, 90%, and 100% *Geobacter* at the onset. The simulations demonstrated that the abundance of *Geobacter* and SRB prior to acetate addition had very little influence on the community dynamics after the acetate addition. In all the simulations containing *Geobacter*, they quickly became the dominant species after acetate addition, increasing their numbers to about $2 \times 10^6$ cells/ml before the depletion of bioavailable Fe(III) (Figure 4-3, Row 1). SRB grew at a much slower rate than *Geobacter* species, but SRB growth continued for a longer time because of the abundance of sulfate in the system. In all simulations containing SRB, they overtook *Geobacter* in abundance at about day 23 and increased their numbers to about $6 \times 10^6$ cells/ml before the sulfate level became too low to maintain growth (Figure 4-3, Row 1). Acetate, Fe(III), and sulfate were consumed at expected times and quantities comparable to the growth of *Geobacter* and SRB.

Strikingly, the dynamic features of Fe(III) and sulfate reduction in all the simulations containing *Geobacter* and SRB respectively are practically identical. It appears that the dynamics of the simulations containing both organisms is equal to the dynamics of the simulation containing no *Geobacter* plus the dynamics of the simulation containing only *Geobacter*. The difference in the timing of the onset of sulfate reduction between the simulations with 90% and 10% SRB is only about a day. Even with no *Geobacter* present, it still took more than 30 days for the sulfate reduction to become apparent (Figure 4-3, Column 2-5). Acetate never became limiting for any of the simulations,
indicating that SRB were never competitively excluded under these conditions. These simulations clearly demonstrate that the metabolic interactions between *Geobacter* and SRB are minimal following a batch addition of a high dosage of acetate. A similar lack of interaction is expected in the field where the acetate concentration is maintained at a high level due to the continuous addition of acetate (Williams *et al.*, 2011).

The same conclusions can be drawn from simulations initiated with 100 times less initial biomass and simulations initiated with twice as much Fe(III) (Appendix IV-2). However, in the low initial biomass simulation, the onset of sulfate reduction and the time at which the SRB overtook the *Geobacter* was significantly delayed (Appendix IV-2). This delay occurred because it took additional time for SRB to accumulate a sufficiently high biomass to become a major contributor of the community metabolism. Given that the timing of the onset of sulfate reduction is the same in the cases with *Geobacter* and in the case without *Geobacter*, it is clear that the lateness of the onset of sulfate reduction is primarily due to the slow growth rate of SRB and is modulated by the initial abundance of SRB, and it is therefore unrelated to the presence of *Geobacter*. Furthermore, the fact that that the timing of the decrease of *Geobacter* activities is the same in the cases with or without SRB, it is clear that the growth of *Geobacter* is limited by Fe(III) availability alone, and is unrelated to the competition from SRB for acetate.

Thus, both the sediment incubation studies (Figure 4-1) and the simulations (Figure 4-2 and Figure 4-3) suggest that *Geobacter* and SRB are not mutually exclusive during bioremediation. In fact, the results demonstrated that neither organism significantly inhibits the activity of its competitor as long as the common nutrient is abundantly available.
Figure 4-3. Ideal batch reactor model simulations with varying initial ratio of *Geobacter* and SRB.

The community dynamics in the ideal batch reactor is simulated using initial *Geobacter* fractions of 0%, 10%, 50%, 90%, 100%. In the cell number row, green line indicates *Geobacter*, blue line indicates SRB.
4.3.3. Effects of Adding Fe(III) Oxide

The apparent lack of interaction between Geobacter species and SRB suggested that the addition of more Fe(III) oxide to the system might prolong the growth and activity of Geobacter species. A simulation of Fe(III) oxide amendment predicted an increase in Geobacter in response to Fe(III) additions with no changes in SRB growth or sulfate reduction (Figure 4-4). To experimentally test the model predictions, acetate and sulfate, and then acetate, sulfate, and Fe(III) oxide were added to sediments which were previously amended with acetate and incubated until Fe(III) and sulfate were depleted. Geobacter grew in response to the Fe(III) additions in the two bottles where Fe(III) was amended but they did not grow when no Fe(III) was added (Figure 4-5A). SRB grew in response to the additions of acetate and sulfate and the addition of Fe(III) had little effect on their growth (Figure 4-5) and the reduction of sulfate (data not shown). Similarly, simulations showed that if Fe(III) is added to the sediment at the beginning of the experiment, additional Geobacter biomass is produced, leading to faster utilization of Fe(III) but having little effect on the utilization of sulfate (Appendix IV-4). These results were consistent with the model predictions, demonstrating that the addition of Fe(III) oxide can resuscitate the growth of Geobacter, and further suggesting that the activities of Geobacter and SRB are not mutually exclusive during acetate-rich phases of bioremediation.
Figure 4-4. Predicted impact of adding Fe(III).

Predicted impact of adding Fe(III) on the number of *Geobacter* (A) and SRB (B). The different line style and markers distinguish the control simulation where no Fe(III) was added from the simulation where Fe(III) was added. The black arrows indicate the day Fe(III) was added.

Figure 4-5. Experimentally observed impact of adding Fe(III).

Experimentally observed impact of Fe(III) oxide additions on the number of *Geobacter* (A) and SRB (B). The green (diamond style) and red (circle style) lines indicate the data from two bottles in which Fe(III) was added. The blue (star style) line indicates the data from the bottle in which no Fe(III) was added. The black arrows indicate the time Fe(III) was added to both bottles #1 and #2.
4.4. Concluding Remarks

Our results demonstrate that it is possible to predictively model the interaction of growth of Geobacter and SRB when acetate is added to subsurface sediments to promote in situ uranium reduction. It is apparent from both the experimental and modeling approaches that the acetate-oxidizing Geobacter and SRB have little impact on each other as long as acetate is maintained in excess. The growth of Geobacter species is primarily controlled by the availability of Fe(III), and the initial predominance of Geobacter species following the addition of acetate can be attributed to the faster growth of Geobacter species. These studies provide the basis for the modeling-based design of in situ bioremediation approaches which will also have to consider additional complexities, such as rates of acetate delivery to the subsurface via injection wells, and geochemical reactions, such as the reduction of Fe(III) by sulfide, which removes Fe(III), but also generates S°, an alternative electron acceptor for Geobacter species. Through this approach, it is expected that the optimal strategies for the addition of acetate, Fe(III), and possibly other amendments will be identified. In the next chapter, we will look at how the DyMMM model of Geobacter and SRB can be used to aid the design of such a strategy.
Chapter 5. The Design of Long-term Effective Uranium Bioremediation Strategy using a Community Metabolic Model⁵

5.1. INTRODUCTION

Uranium is a common contaminant across many nuclear weapons testing and storage sites around the world. Conventional treatment strategies, including the pump-and-treat technologies and the ground water flushing method of reducing uranium concentration, are very inefficient (Anderson et al., 2003) due to the large volume of contaminated water. Recently, several in situ uranium bioremediation strategies have been field tested in several sites in US, including the sites at Rifle, Colorado (Anderson et al., 2003; Vrionis et al., 2005; Williams et al., 2011) and the sites at Oak Ridge, Tennessee (Van Nostrand et al., 2011). During in situ bioremediation, electron donors are added to the subsurface in order to promote the activities of uranium-reducing microorganisms, such as Geobacter (Caccavo et al., 1994; Mahadevan et al., 2006; 2010; Sun et al., 2009), Anaeromyxobacter (Sanford et al., 2007) and Desulfovibrio (Lovley & Phillips, 1994) species. Unfortunately, in situ addition of electron donors can also promote the activities of other microorganisms that may hinder the bioremediation. For example, during the Rifle field experiments, the addition of acetate to sulfate-rich sites in Rifle, CO, produced an initial Geobacter-driven Fe(III) and U(VI) reducing phase followed by a sulfate-reducing phase driven by sulfate reducing bacteria (SRB), during which Fe(III) and U(VI) reduction ceased (Anderson et al., 2003; Vrionis et al., 2005; Miletto et al., 2011). The little understood microbial ecology during bioremediation is further complicated by the geochemical variations across contaminated sites. Given the complexity of the subsurface systems, powerful modeling tools are needed for the design and optimization of effective site-specific uranium bioremediation strategies.

⁵ This chapter is adapted from the article “The Design of Long-term Effective Uranium Bioremediation Strategy Using a Community Metabolic Model” published in Biotechnology and Bioengineering (Zhuang et al., 2012b). The optimization work described in sections 5.2.8 were performed by Eugene Ma under the guidance of Kai Zhuang. This work is included for completion purpose with the permission of Eugene Ma.
There has been several previous modeling studies of the uranium-contaminated aquifers (Roden & Scheibe, 2005; Scheibe et al, 2006; Yabusaki et al, 2007). Unfortunately, while these studies offer rich descriptions of the geochemistry of the contaminated subsurface, their descriptions of the microbial species are often simplistic. In this study, using the DyMMM framework, we have constructed a community model of the Rifle subsurface by integrating the expanded genome-scale metabolic model of Geobacter sulfurreducens and a core-pathway metabolic model of SRB (Figure 5-1). Previously, the genome-scale model of Geobacter has been coupled to a large-scale reactive transport model (Fang et al, 2011). However, because the uranium bioremediation at Rifle is primarily microbial, we have opted to focus on the modeling of biological phenomenon only.

In the previous Chapter, we used this community model to demonstrate that the transition from metal-reducing phase to sulfate-reducing phase at Rifle is caused by the depletion of bio-accessible Fe(III), and have experimentally verified that Geobacter can be resuscitated during the sulfate-reducing phase through batch addition of Fe(III) (Chapter 4). This study hinted that the simultaneous amendment of Fe(III) and acetate may be, in theory, a viable long-term bioremediation approach. However, both computational (Chapter 4, this Chapter) and experimental evidence (Moon et al, 2010) suggested that batch addition of Fe(III) is insufficient to promote prolonged uranium remediation.

In this Chapter, using the Rifle community metabolic model, we would demonstrate that continuous additions of acetate and Fe(III) to the subsurface could prolong the uranium removal indefinitely. In addition, we have computationally optimized the rates and timings of acetate and Fe(III) additions, and showed that the simultaneous addition strategy can maintain the concentration of soluble uranium below the threshold level mandated by US environmental agencies (Anderson et al, 2003). Interestingly, we found that Geobacter dominance of the community is not required for prolonged uranium removal.
Added acetate is oxidized by both attached and planktonic *Geobacter* as well as SRB. Planktonic *Geobacter* reduces uranium, attached *Geobacter* reduces Fe(III), and SRB reduces sulfate. Two forms of Fe(III) are modeled based on their accessibility by attached *Geobacter*. 

---

**Figure 5-1. Conceptual model of the subsurface community during bioremediation.**
5.2. METHOD

5.2.1. Overview of the Community Modeling framework

The Dynamic Multi-species Metabolic Modeling (DyMMM) framework allows the integration of previously developed Bottom-Up Genome-Scale (BUGS) models as well as traditional kinetic models of individual species into a community model. The detailed description of the DyMMM framework has been given in Section 3.2. Previously, the DyMMM framework has been applied to study the competition between Geobacter and Rhodoferax, two of the iron-reducing species coexisting in Rifle (Zhuang et al., 2011a). Using the DyMMM framework, we constructed a community model of Geobacter and SRB by integrating two genome-scale metabolic models of G. sulfurreducens and a core-pathway metabolic model of SRB (Figure 5-1). The details of the Geobacter and SRB metabolic models are described in the following sections. To assess the relative abundance of Geobacter and SRB in the community, we used a custom metric “Geobacter fraction” that is defined as the ratio of the number of Geobacter cells to the total cell number.

5.2.2. Expanded genome-scale metabolic model of Geobacter

The genome-scale model of G. sulfurreducens has been previously published and experimentally verified (Mahadevan et al., 2006; Segura et al., 2008), and the same acetate kinetic parameters used in the Geobacter and Rhodoferax competition study (Zhuang et al., 2011a) were used here. However, we have expanded the model to include a uranium reduction pathway based on the electron capacitance model (Zhao et al., 2010; 2011) as well as multiple Fe(III) uptake pathways to reflect the existence of multiple Fe(III) types in the subsurface. Two separate genome-scale Geobacter models were used; the attached and planktonic Geobacter are treated as two separate organisms. We assumed that only attached Geobacter cells are able to reduce Fe(III) because Fe(III) exists in precipitate form only in the Rifle sediment (Figure 5-1). We also assumed that only planktonic
Geobacter cells are able to reduce uranium (Figure 5-1); this is a simplifying assumption because we did not know how much uranium is accessible by the attached cells.

5.2.3. Differentiating easy-to-use and hard-to-use Fe(III)

Previous experimental work has established that only 3% of the Fe(III) in the Rifle sediment is amorphous Fe(III) oxide; the rest of the total Fe(III) is made up of Fe(III)-silicate, Al-geothite, and magnetite (Komlos et al, 2008a). The latter three forms of Fe(III) are much more difficult to reduce by Geobacter compared to amorphous Fe(III) oxide because their physical geometry allows for less surface area for iron reduction. To differentiate the easy-to-use Fe(III) from the hard-to-use Fe(III), we included two Fe(III) reduction pathways in the Geobacter model. Although the two Fe(III) uptake pathways are stoichiometrically identical, they have different maximal uptake rates for Fe(III). The uptake parameters for the easy-to-use Fe(III) have been previously published (Zhuang et al, 2011a); here, we have assumed that the $K_s$ for hard-to-use Fe(III) is identical to the $K_s$ for easy-to-use Fe(III), but the maximal uptake rate is much smaller (Equation 5-1, Equation 5-2) The $V_{\text{max}}$ for difficult-to-use Fe(III) was estimated to be 30 mmol/gdw/hr using Fe(II) data (Chapter 4, Appendix IV-1).

$$v_{\text{Fe(III), easy}} \leq \frac{586S_{\text{Fe(III), easy}}}{S_{\text{Fe(III), easy}} + 1}$$  \hspace{1cm} \text{Equation 5-1}

$$v_{\text{Fe(III), hard}} \leq \frac{30S_{\text{Fe(III), hard}}}{S_{\text{Fe(III), hard}} + 1}$$  \hspace{1cm} \text{Equation 5-2}
5.2.4. Uranium reduction pathway of *Geobacter*

The mechanisms of uranium reduction are not yet clearly understood. One hypothesis suggested that electrons can be temporarily stored in reduced cytochromes and subsequently transferred to uranium (Zhao *et al.*, 2011; 2010), although recent evidence suggested that electrons can be directly transferred to uranium via pili (Cologgi *et al.*, 2011). In addition, recently it has been reported that Fe(II) can reduce uranium under specific conditions (Du *et al.*, 2011). In this model, we have lumped all of these different mechanisms into a single uranium reduction equation. To incorporate this lumped reaction into the community model, we expanded the genome-scale metabolic model of *Geobacter* by including a hypothetical reaction for uranium reduction. The uptake of uranium is constrained by Equation 5-3, which reflects the uranium reduction kinetics predicted by the electron capacitance model (Zhao *et al.*, 2011; 2010).

\[
v_{U(VI)} \leq \frac{1.8S_{U(VI)}}{S_{U(VI)} + 0.036}
\]

Equation 5-3

**Modeling the attachment and detachment of *Geobacter***

The attachment and detachment kinetics of *Geobacter* have been previously modeled based on the experimental observation that *Geobacter* primarily existed in planktonic form during uranium reduction (Zhao *et al.*, 2011; 2010). Recent experimental evidence showed that during the iron reduction phase, *Geobacter* primarily existed in planktonic form, whereas during the sulfate reduction phase, *Geobacter* primarily existed in attached form (Shabir *et al.*, Personal Communication). It is possible that during the sulfate reduction phase, the electron acceptor starved *Geobacter* becomes less motile in order to conserve energy. To model this phenomenon, we used different attachment/detachment kinetics for the iron and sulfate reduction phases: we assumed that if *Geobacter* has sufficient energy to grow, then the \( K_{\text{detach}} \) is 0.01 hr\(^{-1}\), and \( K_{\text{attach}} \) is 0.05 hr\(^{-1}\); otherwise, \( K_{\text{detach}} \) is 0.05 hr\(^{-1}\), and \( K_{\text{attach}} \) is 0.01 hr\(^{-1}\). The values are chosen so that the observed
ratios of planktonic and attached *Geobacter* cells during both phases (Shabir *et al.*, Personal Communication) are predicted.

### 5.2.5. Core-pathway model of SRB

The core-pathway model of SRB was described in Section 4.2.5. The same model is used.

### 5.2.6. Environmental and physiological parameters

The acquisition of most of the environmental and physiological parameters has been described in the previous chapter (Chapter 4); we will briefly describe them here. A death rate of 0.0011 hr⁻¹ was chosen for both organisms based on the values of two previous models of sulfate and iron reducers (Bethke *et al.*, 2008; Moosa *et al.*, 2002). This value can accurately describe the decay of *Geobacter* (Chapter 4).

The dilution rate of the system has been previously calculated to be 0.0014 hr⁻¹ (Zhuang *et al.*, 2011a). The rate of acetate addition was calculated to be 0.0028 mM/hr by multiplying this dilution rate with the average target concentration of 2mM used in the 2002 Rifle experiment (Anderson *et al.*, 2003). Based on the 2002 field data (Anderson *et al.*, 2003), we have set the initial concentrations of U(VI) and Fe(II) to 0.8 µM and 50 µM respectively, and have set the influx of U(VI) and Fe(II) to 0.0011 µM/hr and 0.0705 µM/hr respectively.

The average concentration of easy-to-use Fe(III) and hard-to-use Fe(III) has been determined to be 0.16 mM and 5.1 mM respectively (Chapter 4). From column experiment data published by Komlos *et al.* (2008), we have estimated that only about 20% of the hard-to-use Fe(III) is bioavailable under continuous acetate addition.
conditions (Komlos et al., 2008a; 2008b). Therefore, the simulations are initialized with 0.16 mM and 1.02 mM of easy-to-use Fe(III) and hard-to-use Fe(III) respectively.

5.2.7. Simulations of batch and continuous addition of Fe(III)

We simulated a total of four Fe(III) addition scenarios based on the intuitive understanding of the system dynamics. In the first two scenarios, 0.48 mM of easy-to-use Fe(III) and 3.06 mM of hard-to-use Fe(III) were added in batch, on either day 10 or day 60. The concentrations of the Fe(III) added was calculated so that the naturally-existing Fe(III) concentrations are quadrupled. In the third and fourth scenarios, easy-to-use Fe(III) was added continuously at a rate of 0.0224 mM/hr – calculated based on the rate of acetate addition and the Fe(III) reduction stoichiometry of 8 Fe(III) per acetate – starting on either day 10 or day 60. All other parameters remained the same as described in the previous sections.

5.2.8. Optimization of Acetate and Fe(III) Addition

The goal of the optimization problem is to reduce the uranium (VI) concentration below the environmental safety threshold (0.18 µM) rapidly and maintaining it at this low level thereafter, while minimizing the amount of acetate and Fe(III) added. We have formulated the problem as a dynamic nonlinear optimization problem, using an objective function that minimizes the difference between the uranium concentration in the system and a target uranium profile as well as the total concentrations of acetate and Fe(III) in the system. In the target uranium profile (the red dotted line in Figure 5-5J), during the first 250 days, the uranium concentration is first rapidly reduced from 0.8 µM to 0.17 µM, then maintained at 0.17 µM. This portion of the target profile was derived from the predicted uranium profile of a simulated scenario where massive amount of acetate and Fe(III) was added to the system; it was arbitrarily chosen in order to ensure that uranium
is reduced from 0.8 µM to 0.17 µM in a relatively short time. During the next 1750 hours in the target profile, the uranium concentration is maintained at 0.17 µM, a value that is below the environmental safety threshold of 0.18 µM (Anderson et al., 2003). The optimization horizon is 200 days.

The dynamic optimization algorithm is represented in Equation 5-4 below. In the outer optimization loop, the flow rate of acetate and Fe(III) through the injection gallery is hard constrained between a lower and an upper flux value. Similarly, the microbial biomass and the metabolites in the environment are governed by differential equations (Equations 5-1 and 5-2), previously defined earlier in the community modeling framework. The optimization problem is subject to the ordinary differential equations arising from the DyMMM community modeling framework, where a linear programming (LP) problem is solved to obtain the cellular exchange fluxes. In this LP, the growth rate is maximized subject to stoichiometric and mass balance constraints. In this optimization, chemical amendment rates into the system were allowed to change every 10 days.

\[
\begin{align*}
\min W_1[U(VI)_1 - U(VI)_{\text{target}}] + W_2 Ac + &\frac{1}{8} W_2 Fe(III) \\
\text{st.} &\begin{cases}
\text{inj}^\text{min}_k \leq \text{inj}_k \leq \text{inj}^\text{max}_k \\
 dX_j / dt = \mu_j X_j \\
 ds_j / dt = \sum_{i=1}^{N} V_i^j X_j \\
 \max u_j = c^T V_j^i \\
 \text{st.} \begin{cases}
 A_j V_j = 0 \\
 V_j^i,\text{min} \leq V_j \leq V_j^i,\text{max}
\end{cases}
\end{cases}
\end{align*}
\]

Equation 5-4
NOTES: Outer Uranium Minimization Loop

\( W_1 \) and \( W_2 \) represent the weighting factors involved in the optimization objective. For the Fe(III) portion, an additional factor of \( \frac{1}{8} \) is included to depict the stoichiometric relationship of acetate and Fe(III) in the system. \( \ln j_k \) represents the injection gallery of the \( k^{th} \) the chemical amendment species. The two differential equations \( \frac{dX_j}{dt} = \mu_j X_j \) and \( \frac{dS_i}{dt} = \sum_{j=1}^{N} V_{ij} X_j \) were previously described in Equations 1 and 2.

NOTES: DMMM equations

The specific growth rate, \( \mu_j \), is maximized under the relationship, \( \mu_j = c^T V_j \), where \( V_j \) is the specific consumption rate/production rate of \( i^{th} \) metabolite of the \( j^{th} \) microbial species and \( c^T \) is the FBA objective column for maximization of the biomass flux. \( A_j \) is the stoichiometric matrix representation of the chemical reaction network of the \( j^{th} \) microbial species.
5.3. RESULTS AND DISCUSSION

5.3.1. Simulation of 2002 Rifle experiment

Using the DyMMM model of the Rifle community (Figure 5-1), we were able to predict several of the key features in the concentration profiles of acetate (Figure 5-2E), sulfate (Figure 5-2F), and uranium (Figure 5-2J), as well as the dynamic changes in community composition during the 2002 Rifle field experiment (Figure 5-2B). Since the geochemistry of Fe(II) adsorption and precipitation was not modeled, the model predicted the concentration of Fe(II) in both aqueous and solid forms (Figure 5-2H). If we assume that 70% of the Fe(II) produced by *Geobacter* leaves the aqueous phase through either adsorption or precipitation based on the column experiment data (Komlos *et al.*, 2008b), then we can acquire an aqueous Fe(II) concentration profile that is similar to the measured data from 2002 field experiment (Figure 5-2H). Despite the relative simplicity of our model, it is clear that it is capable of satisfactorily predicting field-scale phenomena.

Immediately after the onset of acetate addition, the growth of both *Geobacter* and SRB were acetate-limited – their growth rates increased (Figure 5-2D) as acetate accumulated in the system (Figure 5-2E). Due to its significantly faster acetate uptake rate and slightly higher growth yield, *Geobacter* grew at a much higher rate than SRB (Figure 5-2C, D), and soon dominated the community (Figure 5-2B). However, as the concentration of easy-to-use Fe(III) decreased between day 10 and 13 (Figure 5-2G), *Geobacter* became Fe(III)-limited and its growth rate slowed significantly (Figure 5-2C, D). The growth rate of *Geobacter* decreased further as the concentration of the hard-to-use Fe(III) decreased between day 13 and 45, (Figure 5-2C, D, H). Once both Fe(III) sources were depleted, the *Geobacter* population began to decay. The rate of uranium reduction was slow initially, and increased with the number of planktonic *Geobacter* (Figure 5-2A, C); the reduction of uranium came to a stop as the number of planktonic *Geobacter* decreased due to Fe(III) depletion (Figure 5-2A, C, G, H). It is clear that the halting of both *Geobacter* growth and uranium reduction (Figure 5-2D, J) are due to the depletion of bio-accessible Fe(III) (Figure 5-2G, H).
Unlike *Geobacter*, SRB remained acetate-limited throughout the simulation and grew at a slow but steady rate until day 60 – when the acetate concentration is reduced to a very low level – after which SRB gradually decayed toward a steady state concentration (Figure 5-2C, D, E). Prior to the depletion of the Fe(III) sources, the reduction of sulfate and Fe(III) occurred simultaneously – sulfate reduction continued throughout the simulation, although it only becomes apparent at milli-molar scale after day 45 (Figure 5-2F). This result supports the hypothesis that the apparent delay in the onset of sulfate reduction is perhaps due to the slow growth rate of SRB (Figure 5-2D, Chapter 4).
Figure 5-2. Simulation of the 2002 Rifle Experiment.

The community model was able to accurately predict the results from the 2002 Rifle Experiment (Anderson et al., 2003). The percent of Geobacter that is planktonic (A), the Geobacter fraction (B), the number of microbial cells (C), the growth rates of Geobacter and SRB (D), as well as the concentrations of acetate (E), sulfate (F), easy-to-use Fe(III) (G), hard-to-use Fe(III) (H), Fe(II) (I), and U(VI) (J) are shown. The dotted line in panels E, F, I, J shows the measured chemical concentrations during the 2002 Rifle experiment; the different lines represent data from different test wells. The diamonds in panel B show the Geobacter fraction calculated from measured 16S rRNA data during the 2002 Rifle experiment. Note: Y-axis of Panel C is in log-scale.
5.3.2. Batch Addition of Fe(III) does not prolong uranium reduction

Prior to the addition of acetate, the subsurface microorganisms at Rifle, *Geobacter* included, were limited by the availability of electron donors (Chapelle, 1992). However, our simulation showed that when a large quantity of acetate is added to the subsurface continuously, *Geobacter* soon became acceptor-limited (Figure 5-1). One way to alleviate this acceptor-limitation is to add Fe(III) to the subsurface along with acetate. Moon *et al.* found that *Geobacter* growth was indeed enhanced when Fe(III) is added in batch to a sediment column (Moon *et al.*, 2010), and Zhuang and Barlett *et al.* found that the growth of *Geobacter* can be resuscitated if Fe(III) is added after the depletion of Fe(III) (Chapter 4). However, Moon *et al.* also found the Fe(III) addition had no measurable effect on the reduction of uranium (Moon *et al.*, 2010).

One possible explanation of the Moon experiment (Moon *et al.*, 2010) was that the quantity of Fe(III) added was insufficient as only a very small quantity (0.12-0.15 wt%) of Fe(III) was added. Another possibility is that the hard-to-use Fe(III)-geothite was added in the Moon experiment, leading to slower *Geobacter* activity. To assess these explanations, we simulated two scenarios where much higher quantities of both types of Fe(III) were added in batch either prior to (day 10) or after (day 60) the depletion of naturally existing Fe(III). In both scenarios, acetate was added continuously throughout the experiment. We found that when Fe(III) was added on day 10 (Figure 5-3 A, B), there were more *Geobacter* in the community (Figure 5-3C), leading to faster overall rates\(^6\) of Fe(III) and U(VI) reduction. However, as a consequence, Fe(III) was depleted faster (Figure 5-3A, B), effectively shortening the length of the uranium reduction (Figure 5-3D). On the other hand, when a equal quantity of Fe(III) was added on day 60, *Geobacter* growth and uranium reduction were resuscitated temporarily (Figure 5-3C, D). However, the added Fe(III) was quickly used up within 20 days (Figure 5-3A,B) and *Geobacter* activity halted again (Figure 5-3C, D). It seems that Fe(III) cannot be added blindly at large quantities – higher *Geobacter* population and faster reduction rates may have adverse effects – both the timing and quantity of the addition must be optimized.

\[^6\text{The overall rates of reduction (mM/hr) is the specific rate of reduction (mmol/g/hr) multiplied by the biomass concentration (g/L).}\]

84
Furthermore, continuous or fed-batch addition of Fe(III) may be required for the strategy to remain effective.

**Figure 5-3. Simulation of Batch Fe(III) Amendment**

The simulation suggests that batch amendment of Fe(III) in addition to continuous acetate addition is insufficient to promote long-term bioremediation. The concentrations of easy-to-use Fe(III) (A), hard-to-use Fe(III) (B), the *Geobacter* fraction (C), and the concentration of U(VI) (D) are shown.
5.3.3. Continuous Addition of Acetate and Fe(III) can prolong uranium reduction

Based on the above analysis, we propose that the continuous addition of both acetate and Fe(III) can prolong the uranium bioremediation effect indefinitely. To test our hypothesis \textit{in silico}, we simulated the two scenarios where easy-to-use Fe(III) was added continuously to the subsurface starting on either day 10 or day 60. In both scenarios, acetate was added continuously throughout the experiment. When Fe(III) addition started on day 10 – prior to the depletion of naturally existing Fe(III) – \textit{Geobacter} dominated the community throughout the simulation (Figure 5-4C). Consequently, uranium reduction never waned (Figure 5-4D). When Fe(III) addition started on day 60 – after the depletion of naturally existing Fe(III) and – the \textit{Geobacter} dominance of the community and the reduction of uranium were restored by the Fe(III) addition (Figure 5-4C, D). Importantly, in both scenarios, the uranium concentration was eventually reduced to and maintained at a very low level, clearly demonstrating the potential of continuous Fe(III) addition to complement acetate addition.
Figure 5-4. Simulation of Continuous Fe(III) Amendment

The simulation suggests that continuous amendment of Fe(III) in addition to continuous acetate addition is sufficient to promote long-term bioremediation. The concentrations of easy-to-use Fe(III) (A), hard-to-use Fe(III) (B), the Geobacter fraction (C), and the concentration of U(VI) (D) are shown.
5.3.4. Optimizing the rates of acetate and Fe(III) addition

Despite its predicted effectiveness, the continuous addition of Fe(III) and acetate at fixed rates is inefficient. It is clear that not all Fe(III) added was used by Geobacter, leading to its accumulation in the latter part of the simulation (Figure 5-4A,B). In order to find a bioremediation strategy that is both efficient and effective, we have developed an optimization framework around the Rifle community model that minimizes the difference between the predicted uranium concentration and the target uranium concentration as well as the rate of acetate and Fe(III). The target uranium concentration is formulated by extracting and manipulating uranium reduction dynamics from field-scale studies. The optimized rates of acetate and Fe(III) addition and their effect on bioremediation are shown in Figure 5-5. The rate of Fe(III) addition ranges from 0 mM/day to 0.02 mM/day, with a mean of 0.0071 mM/day (Figure 5-5K). The rate of acetate addition ranges from 0 mM/day to 0.0102 mM/day, with a mean of 0.0036 mM/day (Figure 5-5L). The simulation clearly indicates that this injection profile is capable of maintaining the uranium concentration below the environmental safety threshold (Figure 5-5J). The concentrations of acetate and both Fe(III) types eventually reach zero (Figure 5-5E,G,H), indicating that there is no accumulation of these substrates that might suggest inefficient use.

In previous field-scale experiments (Anderson et al, 2003; Vrionis et al, 2005), uranium reduction was correlated with Geobacter dominance of the community; interestingly, since SRB does not inhibit Geobacter activity during uranium bioremediation (Chapter 4), the strategy does not require the dominance of the Geobacter species. Geobacter concentration is maintained at around $10^9$ cells/L (Figure 5-5C), and Geobacter only accounts for about half of the community at the end of the simulation (Figure 5-5B, C). In a separate simulation where 0.077 mM/day of acetate and 0.077 mM/day of Easy-to-Use Fe(III) was added to the system continuously, the uranium concentration was reduced to and maintained at a near-zero level while the Geobacter fraction eventually leveled to 40% (Figure 5-5B). These results suggests that as long as Geobacter is maintained at a steady state level, it can reduce uranium continuously, even when it is not the dominant member of the community.
Figure 5-5. Optimized Amendment of Fe(III) and Acetate

Fe(III) and acetate addition rates recommended by the optimization algorithm are able to reduce and maintain the U(VI) concentration below the safety threshold indefinitely while minimizing the cost of the amendments. The percent of Geobacter that is planktonic (A), the Geobacter fraction (B), the number of microbial cells (C), the growth rates of Geobacter and SRB (D), the concentrations of acetate (E), sulfate (F), easy-to-use Fe(III) (G), hard-to-use Fe(III) (H), Fe(II) (I), U(VI) (J), as well as the optimized Fe(III) (K) and acetate (L) are shown. Note: Y-axis of Panel C is in log-scale.
5.4. Concluding Remarks

In this work, we used a community metabolic model that contains an expanded genome-scale model of Geobacter and a pathway-scale model of SRB as the basis to design a potentially long-term effective uranium bioremediation strategy. Although the addition of Fe(III) as a supplement to acetate addition has been proposed and tested in the past (Moon et al., 2010), we found that batch addition of Fe(III) is insufficient to maintain Geobacter activity. On the other hand, continuous addition of both acetate and Fe(III) is sufficient to maintain Geobacter activity indefinitely. Most interestingly, we found that continued Geobacter dominance is unnecessary for uranium bioremediation – instead, Geobacter can be maintained at a relatively low concentration as long as there is a continuous supply of Geobacter’s key substrates, namely, acetate and Fe(III).

To our knowledge, this is the first time where a community metabolic model has been used to design an optimized strategy for a complex environmental biotechnology application involving several microbial species. While the results are encouraging, it is important to recognize that we are presenting a proof-of-principle model aimed to delineate a novel approach for environmental biotechnology. In order to implement such optimization-based strategy in the field, the computational algorithm will likely require feedback from on-line sensors capable of detecting environmental changes in real time. In addition, the bio-fouling of the aquifer may be an additional source of concern: the rapid growth of microorganisms stimulated by acetate addition could become inhibitory towards further bioremediation. However, given that bio-fouling has not been reported in the previous field-scale experiments (Anderson et al., 2003; Vrionis et al., 2005; Mouser et al., 2009), and that the mean rate of acetate addition in the computationally-optimized strategy (0.0036 mM/day) is equal to or lower than the acetate addition rate used in past field-scale experiments (Anderson et al., 2003; Vrionis et al., 2005; Mouser et al., 2009), it is unlikely that the proposed strategy will cause the bio-fouling of the aquifer. Furthermore, the current model focuses solely on microbial metabolism; in the future, this model should be coupled with field-scale geochemical models (Fang et al., 2011) in order to take both metabolic and geochemical phenomena into consideration. Lastly, it should be noted that the simultaneous additions of acetate and Fe(III), while theoretically
viable and effective, is impractical due to the immense cost associated with Fe(III) addition. An alternative is to regenerate Fe(III) through the biostimulation of iron oxidizing bacteria as known iron oxidizers such as *Dechloromonas* have been found in Rifle.

While the dynamic community modeling approach presented here is applicable for competitive (Zhuang et al, 2011a) and cross-feeding (Salimi et al, 2010) communities, its applicability is limited for modeling obligatorily mutualistic communities where the organisms rely on each other for survival. Recently, Zomorrodi and Maranas (2012) proposed a static modeling framework that includes a community-level objective function that optimizes the total biomass production rate of the community. To effectively model mutualistic communities, it will be necessary to couple similar community-level objective functions with the dynamic community models.

Nevertheless, the present work represents a significant step forward in the fields of systems biology and metabolic engineering. Together with the work presented in the previous chapter (Chapter 4), we demonstrated that the DyMMM framework is a useful tool for investigating environmental biotechnology problems involving multiple microbial species. These studies (Chapter 4, 5) should represent a template for applying this approach to other complex environmental problems such as dechlorination (Ahsanul Islam et al, 2010) and aromatics degradation (Viñas et al, 2005). In the next Chapter, we will explore how the DyMMM framework can be applied in a very different context – the engineering of a microbial cell factory.
Chapter 6. DySScO: an efficient strain design algorithm for balanced yield, titer, and productivity

6.1. Introduction

Many microorganisms contain native pathways capable of producing useful chemical compounds (Feist et al, 2010), and synthetic pathways can be genetically inserted into model organisms such as *E. coli* (Yim et al, 2011) and *S. cerevisiae* (Brochado et al, 2010). Unfortunately, from the microorganisms’ perspective, most of these chemicals are metabolic byproducts, and their production is often minimal in wild-type organisms (Feist et al, 2010). To create viable microbial cell factories, the production of these chemicals must be enhanced. Traditionally, biochemical production is enhanced through the mechanism of serial mutagenesis and phenotypic selection (Park et al, 2008). The advent of metabolic engineering introduced a new strain development approach in which genes are either amplified or deleted based on systematic consideration of the metabolic network, often with the aid of constraint-based metabolic models (Feist et al, 2010; Park et al, 2008). A caveat of metabolic engineering is that the outcome of strain improvement is limited by its scope: because metabolic engineering is focused on the cellular metabolism, it often neglects the overall bioprocess (Park et al, 2008).

One approach to metabolic engineering is to couple the production of the desired chemical to the growth of the cell, thus ensuring the production of the chemical (Fong et al, 2005; Hua et al, 2006). A number of constraint-based computational algorithms have been developed for identifying growth-coupled strain designs that enhance the product yield. Whereas several of these algorithms, such as OptKnock (Burgard et al, 2003) and GDLS (Lun et al, 2009), focus on identifying gene knockouts only, other algorithms, such as OptReg (Pharkya & Maranas, 2006) and OptReg’LS (Yang et al, 2011) are capable of identifying reaction activation (Jin & Stephanopoulos, 2007) and inhibition

---

7 This chapter is adapted from the article “DySScO: an efficient strain design algorithm for balanced yield, titer, and productivity” (Submitted).
(Nakamura & Whited, 2003) as well. More recent advances include OptForce, which maximizes product yield by knockout, inhibition, and activation targets, relative to a wild-type flux distribution (Ranganathan et al, 2010), and EMILiO, which rapidly identifies the optimal set of modified reactions and their optimal fluxes using a successive linear programming procedure (Yang et al, 2011). The application of such algorithms has led to the development of industrial microbial cell factories (Yim et al, 2011).

Unfortunately, the limitations of metabolic engineering without process engineering are evident in these algorithms: while these algorithms can optimize the product yield of the strain, they cannot optimize the productivity and titer of the strain because they are process-level concepts and cannot be predicted using standard metabolic models. This is problematic because the economic viability of a bioprocess is commonly evaluated by its product yield, titer, and productivity (Meynial-Salles et al, 2008; Stephanopoulos, 2007). Similarly, titer and productivity may also influence other high-level design objectives such as the ecological benefits of biochemical production. Most existing algorithms seek enhancements in product yield and growth rate, with the hopeful assumption that by improving these two attributes, titer and productivity will also improve. However, since the feedstock in a bioprocess can be converted into either biomass or desired product, the growth rate and the product yield cannot be simultaneously maximized. Constrained by this tradeoff, previous strain-design efforts often prioritize product yield optimization by restricting the growth rate to an arbitrarily low level (Burgard et al, 2003; Feist et al, 2007; Pandit & Mahadevan, 2011; Pharkya et al, 2003; Yang et al, 2011). However, a strain with a reduced growth rate would yield lower biomass concentration in bioreactors, which may reduce the volumetric productivity despite the increase in product yield. Recognizing this bias, Feist et al. (2010) have included substrate specific productivity (SSP) as an additional design criterion (Feist et al, 2010). However, whereas SSP increases linearly with growth rate, productivity increases exponentially with growth rate. As such, SSP cannot replace productivity as a design criterion.
Fortunately, it is possible to predict titer and productivity using dynamic flux balance analysis (dFBA) – which incorporates both the process dynamics and the metabolic network (Mahadevan et al, 2002; Hjersted et al, 2007; Oddone et al, 2009; Feng et al, 2012; Hanly & Henson, 2010; Ghosh et al, 2011). By integrating the existing strain-design algorithms with dFBA, we have developed a novel strain-design strategy called Dynamic Strain Scanning Optimization (DySScO). By using product yield, titer, and volumetric productivity as explicit design criteria, DySScO can be used, potentially, to optimize high-level design objectives such as economic viability and ecological benefits. In this article, we present the rationale behind the DySScO strategy as well as the applications of this approach in designing E. coli strains for succinate and 1,4-butanediol productions.
6.2. Methods

6.2.1. An algorithmic description of the DySScO strategy

The DySScO strategy consists of three major phases: scanning, design, and selection. Here, these phases are broken down further and presented as 9 algorithmic steps. Many of the individual steps can be accomplished using a variety of algorithms and tools. For example, Step 6 can be performed using several existing strain design algorithms, such as OptKnock, GDLS, OptReg, OptReg’LS, and EMILiO.

1. Find the production envelope for the desired product.

2. Create N hypothetical strains along the upper bound of the product envelope by fixing their yield. In this work, the production envelope was generated using the COBRA Toolbox for Matlab, and N=10 is used. The first strain had the highest yield, but did not grow; this strain is clearly not a viable solution and was not simulated.

3. Perform dynamic simulations of the behaviors of the hypothetical strains inside bioreactors (e.g., batch, fedbatch) using dFBA. In this work, the DyMMM framework is used for dFBA simulations.

4. The performance of the hypothetical strains are evaluated using product yield (Y), titer (T), and volumetric productivity (P) calculated from the dynamic simulations. The performance of the strains can be ranked using any metric of form:

\[ Z = f(Y, T, P) \]  

Equation 6-1

In our work, we used a simple metric called the consolidated strain performance (CSP), which is calculated as the weighted normalized sum of Y, T, and P:

\[ \text{CSP} = W_1 \frac{Y}{Y_{\text{max}}} + W_2 \frac{T}{T_{\text{max}}} + W_3 \frac{P}{P_{\text{max}}} \]  

Equation 6-2

5. Based on the performance of the hypothetical strains, the growth rate range for the static strain design process is selected.

6. Existing strain design algorithms are used to find a number of high product yield designed strains within the optimal growth rate range. In this work, the GDLS algorithm is used.
7. The dFBA method is used to simulate the dynamic behaviors of the designed strains inside bioreactor (e.g., batch, fedbatch).

8. The performances of the designed strains are evaluated similar to step 4.

9. The best strain design is selected from the set of designed strains based on their CSP.

This process is illustrated in Figure 6-1.
Figure 6-1. A schematic diagram of the DySScO strategy

The DySScO strategy is divided into three phases, which is subdivided into 9 steps. The “scanning” phase includes steps 1-5. The “design” phase include step 6. The “selection” phase includes steps 7-9.
6.2.2. Application of DySScO: Succinate and BDO production in E. coli

The general DySScO steps were followed in order to design both the succinate-producing strain and the BDO-producing strain. The *E. coli* iAF1260 metabolic model is used as the base model; the BDO biosynthesis pathway developed by Genomatica was added to this model in the BDO case (Yim *et al.*, 2011).

6.2.3. dFBA simulations using DyMMM framework

The dynamic multi-species metabolic modeling (DyMMM) framework (Zhuang *et al.*, 2011a) is a previously described extension to dFBA (Mahadevan *et al.*, 2002; Hjersted *et al.*, 2007; Oddone *et al.*, 2009; Feng *et al.*, 2012; Hanly *et al.*, 2012) designed to model the dynamic interactions between multiple microbial species. However, if only one species is present, then the DyMMM framework mathematically reduces to its predecessor dFBA. We have adapted the DyMMM framework for our dFBA simulations of the hypothetical and designed strains because it is much more flexible than the native dFBA function provided by the COBRA Toolbox. Specifically, we have modified the DyMMM to track the volume of the reactor content (by adding Equation 6-3), thus enable the modeling of fedbatch reactors. In brief, the modified DyMMM framework can be mathematically described as:

\[
\frac{dV}{dt} = F_{in} - F_{out} \quad \text{Equation 6-3}
\]

\[
\frac{dX_i}{dt} = \mu_i X_i - \frac{F_{in} X_i}{V} \quad \text{Equation 6-4}
\]

\[
\frac{dS^j}{dt} = \sum_i v_i^j X_i + \frac{F_{in} (S^{i}_{\text{feed}} - S^j)}{V} \quad \text{Equation 6-5}
\]
Here, $V$ is the volume (L) of the reactor, $X_i$ is the biomass (g/L) of the $i^{th}$ microbial species, $S_j$ is the concentration (mM) of the $j^{th}$ metabolite, $F_{in}$ is the rate of flow (L/hr) into the reactor, $F_{out}$ is the rate of flow (L/hr) out of the reactor, and $S_{j, feed}$ is the concentration of the $j^{th}$ metabolite in the feed stream. $\mu_i$ is the growth rate of the $i^{th}$ microbial species, and $v_{i,j}$ is the metabolic flux of the $i^{th}$ microbial species; $\mu_i$ and $v_{i,j}$ are calculated from FBA (Equation 6-6).

### 6.2.4. Simulation Parameters and Reactor Setup

We have assumed that both batch and fedbatch fermentations are carried out in fully anaerobic conditions. We have assumed that all *E. coli* strains modeled have a maximum glucose uptake of 20 mmol/gdw/hr (Portnoy *et al.*, 2008; Chapter 2) and a saturation constant of 1 mM under anaerobic conditions.

For batch simulations, the reactor is assumed to be well mixed at all times. The initial biomass is set to 0.01 g/L, the initial glucose concentration is set to 20 mM, and the initial liquid volume is set to 1 L. The batch time is set to 50 hours.

For fedbatch simulations, the reactor is assumed to be well mixed at all times. The initial biomass is set to 0.01 g/L, the initial glucose concentration is set to 20 mM, and the initial liquid volume is set to 1 L. The fedbatch time is set to 150 hours. The reactor liquid volume is limited to 10 L. An idealized exponential feeding strategy is used where glucose is continuously added to the bioreactor in order to maintain the glucose concentration at 20 mM until the reactor liquid volume reaches 10 L (Paalme *et al.*, 1990). The glucose feed concentration is set to 1000 mM. The rate of glucose addition is governed by the following equation:
Here, $Q_{\text{g}_{\text{lc}} \text{in}}$ is the glucose feed rate (mM/hr), $v_{\text{g}_{\text{lc}}}$ is biomass specific glucose consumption rate (mmol/gdw/hr), $X$ is the biomass concentration (g/L), $V$ is liquid volume (L), $S_{\text{g}_{\text{lc}} \text{feed}}$ (mM) is the glucose concentration in the feed stream, and $S_{\text{g}_{\text{lc}}}$ (mM) is the glucose concentration in the reactor.

The simulation setups are the same for both succinate and BDO cases.

### 6.2.5. Modeling Succinate Inhibition

The product inhibition of *E. coli* metabolism by succinate is modeled using a formulation that is similar to the extended Monod formulation used by Lin *et al.* to model the succinate inhibition of *Actinobacillus succinogenes* (Lin *et al.*, 2008).

$$v_{\text{g}_{\text{lc}}} \leq v_{\text{g}_{\text{lc}} \text{max}} \left(1 - \frac{S_{\text{succ}^*}}{S_{\text{succ}^*}}\right)^k \left(\frac{S_{\text{g}_{\text{lc}}}}{S_{\text{g}_{\text{lc}}} + K_m}\right)$$  \hspace{1cm} Equation 6-8

Here, $S_{\text{succ}^*}$ is the critical succinate concentration at which *E. coli* stops growing. Based on the work of Li *et al.* (2010) on four *E. coli* strains (Li *et al.*, 2010), we estimated $S_{\text{succ}^*}$ to be 80g/L or 678mM of succinate. $k$ is a fitting parameter that determines that shape of the inhibition curve, which we assumed to be at 1.5.

To check the strain’s sensitivity to a stronger inhibition effect, we also simulated the scenario where $S_{\text{succ}^*}$ is halved to 40g/L or 339mM.

### 6.2.6. The Analysis of the Economic Performance of the Strains

The economic performance of the hypothetical and designed strains in batch reactors were analyzed using two metrics: the hourly profit ($Z_h$), which is the profit made per
hour, and the batch profit \((Z_b)\), which is the profit made per batch. These metrics are calculated using the equations below:

\[
Z_h = p_{\text{product}} \times P - p_{\text{feedstock}} \times P/Y
\]

\[
Z_b = p_{\text{product}} \times T - p_{\text{feedstock}} \times S_0^{\text{feedstock}}
\]

Here, \(p_{\text{product}}\) is the price of the product, \(p_{\text{feedstock}}\) is the price of the feedstock, \(P\) is the volumetric productivity, \(T\) is the titer, and \(S_0^{\text{feedstock}}\) is the concentration of the feedstock at the start of the batch. These calculations do not take the downstream costs (e.g., seperation cost) and facility costs into consideration. The price of glucose \((p_{\text{feedstock}})\) is assumed to be $0.13/mol, which is acquired from US Department of Agriculture’s Economic Research Service’s website (http://ers.usda.gov/data-products/sugar-and-sweeteners-yearbook-tables.aspx). The \(Z_b\) and \(Z_h\) of both succinate and BDO strains are evaluated at three price points: when the price of the product is 2X, 3X, and 8X of the price of glucose. The \(Z_h\) formulation is based on the economic formulation of chemical process found in the literature (Li et al, 2009).

6.2.7. Strain Design using GDLS

The GDLS (Genetic Design through Local Search) algorithm (Lun et al, 2009) was used to identify knockout strategies for succinate and BDO over-production in \(E.\ coli\), using the iAF1260 genome-scale model. For each iteration of GDLS, we used a neighborhood size of 2, and a single search path. In addition, we implemented constraints to prevent the local search from cycling back to the previous solution. Each MILP problem (i.e., local search iteration) was given a timeout threshold of 1800 seconds. If the MILP problem did reach the timeout threshold, then GDLS was continued only if a feasible, but not necessarily optimal, solution was identified. In this work, every local search MILP indeed found a feasible solution even if the timeout threshold was met. The MILPs were solved using CPLEX 12.1 and the CPLEXINT interface, with up to 8 parallel threads using 2.4 GHz AMD Opteron processors.
6.3. Results and Discussions

6.3.1. Dynamic Strain Scanning Optimization (DySScO) Strategy

The Dynamic Strain Scanning Optimization (DySScO) Strategy is a novel strain design strategy capable of producing strains with optimized product yield, titer, and volumetric productivity. DySScO consists of three major phases: scanning, design, and selection. First, during the “scanning” phase, by simulating the dynamic behaviors of a large number of hypothetical strains along the upper boundary of the production envelope using dFBA, an optimal range of growth rates can be determined based on the desired design criteria. Then, during the “design” phase, existing strain-design algorithms (e.g. OptKnock, GDLS, EMILiO) can be used to find the potential strains with high product yield within this growth rate range. Lastly, during the “selection” phase, dFBA simulations of these potential strains are performed and the best strain-design is selected based on the desired design criteria.

In order to rank the performance of the strains, any design criterion that is a function of the yield, titer, and volumetric productivity can be used. In our work, we introduced a metric called consolidated strain performance (CSP), which is a weighted sum of the normalized product yield, titer, and volumetric productivity (Equation 6-2). Since the present work is a proof of principle study of the DySScO strategy, we assigned equal weights of 1 to the three design criteria. This weighting system is good for generating strain designs with balanced yield, titer, and volumetric productivity. In real world applications, the weights should be determined on a case-by-case basis based on higher-level design objectives such as the relative economic values or ecological benefits.

In order to demonstrate the necessity and the effectiveness of the DySScO strategy, we applied the strategy to two strain design problems: the designing of a succinate-producing \textit{E. coli} strain and an 1,4-butanediol (BDO) producing \textit{E. coli} strains.
6.3.2. Designing Succinate-Producing Strains

Since succinate is a commodity chemical that is naturally produced by *E. coli*, many previous metabolic engineering efforts have sought to design succinate-producing *E. coli* strains. We applied DySScO to the designing of succinate-producing *E. coli* strain in order to demonstrate the effectiveness of this strategy, and distinguish it from previous strain-design strategies.

During the “scanning” phase, ten hypothetical strains were created *in silico* using the iAF1260 model of *E. coli*. By constraining the succinate flux, these strains were constrained to the upper boundary of the succinate production envelope (Figure 6-2B). dFBA was used to simulate the dynamic behaviors of these strains in batch reactors, and the potential product yield, titer, and volumetric productivity of these strains were calculated from the predicted process dynamics (Figure 6-2A-D). The product yield (mol/mol) is the quantity of product made per unit quantity of feedstock. The titer (mmol/L) is the concentration of the product at the end of the batch. The volumetric productivity (mmol/L/hr) is the quantity of product made per liter of reactor volume per hour. Amongst the ten hypothetical strains, the strain growing at the rate of 0.15 hr\(^{-1}\) has the best overall performance (Figure 6-2A). We rejected the strain growing at the rate of 0.05 hr\(^{-1}\) because it strongly favored yield and neglected productivity (Figure 6-3A). Based on these results, we selected a growth range of 0.1-0.25 hr\(^{-1}\) for the strain design process (Figure 6-2A).

During the “design phase”, the GDLS strain design algorithm (Lun *et al*, 2009) was used to generate a set of knockout strains that coupled succinate production to biomass synthesis (See Section 6.2.1). Three *in silico* strains were selected from the list of strains produced by this algorithm, hereby designated as “YZ1”, “YZ2”, and “YZ3”. YZ1 has a growth rate of 0.16 hr\(^{-1}\), and is very close to the production envelope, while strains YZ2 and YZ3 have growth rates of 0.24 hr\(^{-1}\) and 0.21 hr\(^{-1}\), respectively, and are slightly farther from the production envelope (Table 6-1, Figure 6-2B).
Figure 6-2. The performance of the succinate-producing strains in batch

(A-D) The consolidated strain performance (A), succinate yield (B), volumetric productivity (C), and titer (D) of both the hypothetical strains and the designed strains in batch simulations. The colored bars in panel A shows the relative succinate yield, volumetric productivity, and titer. The red dotted line indicates the range of growth rates selected for the static strain design process. (E) The tradeoff between the succinate yield and the volumetric productivity. The Pareto frontier is outlined by the dotted line connecting the hypothetical strains.
### Table 6-1. Knockout Strategies for Succinate Overproduction Identified

The knockouts are listed by their reaction names in the iAF1260 genome-scale model of *E. coli* metabolism.

<table>
<thead>
<tr>
<th>Succinate Strains</th>
<th>YZ1</th>
<th>YZ2</th>
<th>YZ3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth rate (hr⁻¹)</td>
<td>0.16</td>
<td>0.24</td>
<td>0.21</td>
</tr>
<tr>
<td>Product yield (mol/mol gle)</td>
<td>1.27</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>Knockouts</td>
<td>ALCD2x, GLUDy, LDH_D, PFL, PPKr, TKT2</td>
<td>F6PA, G6PDH2r, ME2, MTHFD, PFL, PYK</td>
<td>ACALD, F6PA, G6PDH2r, GLUDy, ME2, PFL, PYK</td>
</tr>
</tbody>
</table>

The specific knockout strategies used in the strains YZ1, YZ2, and YZ3 are listed in Table 6-1. Many of the knockouts identified in this work (Table 6-1) overlapped with those found in the literature. For example, the knockout of pyruvate formate lyase (PFL), alcohol dehydrogenase (ALCD2x), and D-lactate dehydrogenase (LDH_D) is consistent with the common experimental strategy (Yu *et al.*, 2010) where the pathways for competing fermentation products (e.g., formate, ethanol, lactate) are knocked out. The knockout of the NADP-dependent malic enzyme (ME2) or glucose-6-phosphate dehydrogenase (G6PDH2r) is consistent with previously identified *in silico* strategies (Feist *et al.*, 2010). This overlap is not surprising since DySScO utilizes existing strain-design algorithms to identify the list of potential strains; what distinguishes DySScO is that the “best strain” is selected by ranking them based on an equally weighted consolidated strain performance score, thus achieving a balance between the three design criteria.
During the “selection” phase, dFBA was used to predict the maximum-achievable yield, titer, and productivity of the strains YZ1, YZ2, and YZ3. The performances of these strains are compared with those of the hypothetical strains (Figure 6-2A, Figure 6-5A-C). Both the overall performance (Figure 6-2A, Figure 6-5A-C) and the dynamics of YZ1 (not shown) are similar to that of the hypothetical strain growing at 0.15 hr$^{-1}$. It significantly outperforms YZ2, YZ3 (Figure 6-2, Figure 6-5A-C) as it has a significantly higher volumetric productivity than the 0.1 hr$^{-1}$ hypothetical strain, while having only a slightly lower product yield. As such, we selected YZ1 to be the best strain using the consolidated performance metric.
6.3.3. Designing 1,4-Butanediol-Producing Strains

Previously, Yim et al. from Genomatica Inc. have reported a synthetic pathway in *E. coli* that allows for the fermentation of 1,4-butanediol (BDO), a non-native metabolite (Yim et al., 2011). The OptKnock algorithm was used to recommend a number of growth-coupled BDO-producing strain designs, and a strain design predicted to have a high product yield (0.76 mol/mol) and a low growth rate (0.14 hr$^{-1}$) was chosen and implemented (Yim et al., 2011). In the present work, we added this BDO-producing pathway to the iAF1260 model of *E. coli*, and applied DySScO in hope to find a better strain that is optimized for the consolidated strain performance.

Similar to the succinate case, during the “scanning” phase, ten hypothetical strains were created along the BDO-production envelope. Interestingly, the slope of BDO-production envelope (Figure 6-3B) is much less than that of the succinate envelope (Figure 6-2B), indicating that the tradeoff between the growth rate and the product yield is less severe in the BDO case. As a result, whereas a clear optimal growth rate range is observable in the succinate case (Figure 6-2B, Figure 6-5A-C), the optimal range is much less obvious in the BDO case (Figure 6-3B, Figure 6-6A-C). We rejected the hypothetical strain growing at 0.05 hr$^{-1}$ because it heavily favors product yield over productivity, and selected the large range of 0.1-0.35 hr$^{-1}$ for strain design (Figure 6-6A).

During the “design” phase, two strains were selected from the list of strains produced by GDLS, hereby designated as “YZ4”, “YZ5”. YZ4 has a growth rate of 0.3 hr$^{-1}$ and a product yield of 0.52 mol/mol. YZ5 has a growth rate of 0.35 hr$^{-1}$ and a product yield of 0.51 mol/mol (Table 6-2). We have also recreated the knockout scheme used by Yim et al. (2011) in the iAF1260 model for comparison. Interestingly, this strain based in iAF1260 (designated as YIM1260) behaved differently than the strain reported by Yim et al., which is based on an older model (designated as YIM$^{gm}$). Whereas YIM$^{gm}$ grew at lower rate (0.14 hr$^{-1}$) and had a high yield (0.76 mol/mol), YIM$^{1260}$ grew at higher rate (0.30 hr$^{-1}$) and had a lower yield (0.52 mol/mol).
Figure 6-3. The performance of the BDO-producing strains in batch

(A-D) The consolidated strain performance (A), BDO yield (B), volumetric productivity (C), and titer (D) of the hypothetical strains, the designed strains, and the strain designed by Yim et al. (2011) in batch simulations. The colored bars in panel A shows the relative succinate yield, volumetric productivity, and titer. The red dotted line indicates the range of growth rates selected for the static strain design process. (E) The tradeoff between the BDO yield and the volumetric productivity. The Pareto frontier is outlined by the dotted line connecting the hypothetical strains.
<table>
<thead>
<tr>
<th>BDO Strains</th>
<th>YZ4</th>
<th>YZ5</th>
<th>YIM\textsuperscript{1260}</th>
</tr>
</thead>
<tbody>
<tr>
<td>Growth rate (hr\textsuperscript{-1})</td>
<td>0.3</td>
<td>0.35</td>
<td>0.3</td>
</tr>
<tr>
<td>Product yield (mol/mol gle)</td>
<td>0.52</td>
<td>0.51</td>
<td>0.52</td>
</tr>
<tr>
<td>Knockouts</td>
<td>ALCD2x</td>
<td>ALCD2x</td>
<td>ALCD2x</td>
</tr>
<tr>
<td></td>
<td>PFL</td>
<td>PFL</td>
<td>PFL</td>
</tr>
<tr>
<td></td>
<td>PGI</td>
<td>TKT2</td>
<td>MDH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>LDH\textsubscript{D}</td>
</tr>
</tbody>
</table>

Table 6-2. Knockout Strategies for BDO Overproduction Identified

The knockouts are listed by their reaction names in the iAF1260 genome-scale model of \textit{E. coli} metabolism.

The two BDO strains identified in this work (YZ4 and YZ5, Table 6-2) showed similarities in the predicted flux distributions as YIM\textsuperscript{1260}. Namely, all three strains used pyruvate dehydrogenase and the oxidative TCA cycle, and secreted acetate, all at similar levels. However, unlike YIM\textsuperscript{1260}, in which malate dehydrogenase (MDH) is deleted, YZ4 and YZ5 utilize MDH in the reverse direction. FBA simulations predict that the knockout of MDH in YZ5 reduces BDO yield by 97% while increasing the activity of lactate dehydrogenase (LDH) in the reverse direction. In order to synthesize BDO, LDH must also be deleted, which leads to the construction of strain YIM\textsuperscript{1260} (Table 6-2). Both MDH and LDH activities are coupled to NADH consumption, and their deletion results in a loss of NADH sinks. To restore redox balance, alternative NADH-consuming reactions must become active, and the BDO synthesis pathway includes four NADH-consuming reactions. While channeling excess NADH to the BDO synthesis pathway in this manner improves BDO yield, this is achieved at the cost of lowered growth rate. In contrast, strain YZ5 increases growth rate at the cost of lowered product yield, through reverse MDH activity.
During the “selection” phase, dFBA was used to predict the maximum-achievable yield, titer, and volumetric productivity of the strains YZ4, YZ5, and YIM\textsuperscript{1260}. Their performance is compared with the hypothetical strains (Figure 6-3A, Figure 6-6A-C). YZ5 had better performance than both YZ4 and YIM\textsuperscript{1260}, making it the best of the designed strains; however, it still performed slightly worse than the several hypothetical strains (Figure 6-3A, Figure 6-6A-C). Interestingly, YZ5 is created by knocking out pyruvate formate lyase (PFL), alcohol dehydrogenase (ALCD2x) pathways, which is a subset of the knockouts of YIM\textsuperscript{1260}, which included the knockout of malate dehydogenase (MDH) and lactate dehydrogenase (LDH_D) in addition to pyruvate formate lyase and alcohol dehydrogenase. In other words, YZ5 achieved a slightly better consolidated performance than YIM\textsuperscript{1260} (Figure 6-3A, Figure 6-6A-C) using two less knockouts (Table 6-2).

### 6.3.4. Succinate Production in Fedbatch

During our design of both the succinate and the BDO producing strains, we evaluated the strains’ performances in batch reactors. The DySScO strategy can also be applied to design strains used in fedbatch reactors. However, to do so, it is important to optimize the feeding strategy as it can greatly influence the outcome. Ideally, a bilevel optimization problem can be formulated to simultaneous optimize the strain design and the feeding strategy; however, this is a difficult computational problem. Since our goal was to demonstrate the necessity and effectiveness of a dynamic strain design strategy, we simplified the problem by adopting an idealized feeding strategy in which the glucose concentration is maintained at a high level (20mM) until the reactor becomes full. Using dFBA, we simulated the dynamics of all the hypothetical and designed succinate-producing strains in this ideal fedbatch reactor.

We found that strains’ product yields are similar in both batch and fedbatch reactors (Figure 6-2B, Figure 6-4B). While the titer and productivity are significantly higher in the fedbatch reactor, the trends between strains are similar (Figure 6-2C,D, Figure 6-4C,D). As a result, the consolidated strain performance scores of the strains are similar
between the two reactors. The 0.15 hr\(^{-1}\) strain remains the best hypothetical strain, and YZ1 remains the best designed strain.

However, it is important to note that the titer of the strains is predicted to reach 1200 mM, which is unrealistic as the growth of *E. coli* becomes inhibited at high acid concentrations. It was recently reported that *E. coli* strains NZN111, AFP111, and BL21 can tolerate up to 680 mM of succinate (Li *et al.*, 2010), a concentration similar to the predicted titer of the fast-growing strains, but much higher than the predicted titer of the slow-growing strains (Figure 6-4D). As such, it seems that this inhibitory effect will significantly reduce the consolidated performance of the slow-growing, high yield strains in some situations, making such strains less valuable. However, it should be noted that succinate titer will not reach this level if a constant feeding strategy is adopted instead of an idealized exponential feeding strategy. Nevertheless, if the succinate inhibition kinetics are available, it can be incorporated into the DySScO formulation (Appendix V-1).
Figure 6-4. The performance of the succinate-producing strains in fed-batch

(A-D) The consolidated strain performance (A), succinate yield (B), volumetric productivity(C), and titer (D) of both the hypothetical strains and the designed strains in fed-batch simulations. The colored bars in panel A shows the relative succinate yield, volumetric productivity, and titer. The red dotted line indicates the range of growth rates selected for the static strain design process. The simulations assume the implementation of an ideal feeding strategy where the glucose concentration is maintained at 20 mM until the reactor is full. (E) The tradeoff between the succinate yield and the volumetric productivity. The Pareto frontier is outlined by the dotted line connecting the hypothetical strains.
6.3.5. Balancing Growth Rate and Product Yield

The main motivation behind the development of the DySScO strategy is our recognition that an increased product yield does not necessitate an increase in titer or productivity, a recognition rooted in the known tradeoff between growth rate and product yield (Figure 6-2B, Figure 6-3B). Our simulations confirmed that this basic tradeoff between growth rate and product yield gives rise to three additional tradeoffs: the tradeoff between growth rate and productivity (Figure 6-2C, Figure 6-3C), the tradeoff between growth rate and titer (Figure 6-2D, Figure 6-3D), and the tradeoff between the product yield and productivity (Figure 6-2E, Figure 6-3E).

It is important to note that whereas the growth rate vs. product yield tradeoff and the growth rate vs. titer tradeoff have monotonic shapes (Figure 6-2B,D, Figure 6-3B,D), the yield vs. productivity tradeoff and the rate vs. productivity tradeoff both have a parabolic shape (Figure 6-2C, Figure 6-3C). This means that there is a growth rate at which the volumetric productivity can be maximized, at least theoretically. Since the consolidated strain performance is first-order linear combination of the normalized yield, titer, and productivity, it also has a parabolic shape and has a theoretical maximum.

Equally important to product yield and productivity is the titer. In industrial settings, the titer often governs the downstream process costs such as the cost of separating the desired chemical from the byproducts. In our simulations, the titer of a strain generally mirrors the product yield of the strain (Figure 6-2B,D, Figure 6-3B,D), except at very low growth rates (Figure 6-3B,D). This is because we used a fast growing microorganism and allowed unlimited batch times, allowing all the substrate to be used for fermentation. However, if the strains were limited by their batch time, slower growing strains would be at a disadvantage with respect to titer. As an example, we compared the behaviors of two strains inside a fedbatch reactor, growing at 0.1 hr\(^{-1}\) and 0.25 hr\(^{-1}\) respectively. Given 5 days, the 0.1 hr\(^{-1}\) strain will produce higher titer than the 0.25 hr\(^{-1}\) strain; however, if the batch time is limited to 2 days, the 0.25 hr\(^{-1}\) strain will produce higher titer than the 0.1
hr\(^{-1}\) strain (Appendix V-2). As such, there exists a potential tradeoff between growth rate and titer if batch time is a limiting factor. This tradeoff becomes significant from a batch scheduling perspective, and can be evaluated by observing the product titer dynamics predicted by DySScO (Appendix V-2).

6.3.6. Analyses of the Economic Viability of the Strains

The production of chemicals through biological means from biomass offers many social and ecological benefits in comparison to conventional petrochemicals (Palsson et al., 1981). However, a basic requirement of a successful bioprocess is that it must be economically viable. Since the yield, titer, and productivity of the strains can be calculated from the simulated process dynamics, the economic viability of the overall bioprocess can be evaluated using various financial metrics. In this proof of principle study, we evaluated the economic performance of the strains in the batch reactor using two simple economic metrics – the hourly profit (profit made per hour, Equation 6-9) and the batch profit (profit made per batch, Equation 6-10). These metrics were calculated at three different price points – when the product price is at 2X, 3X, and 8X the price of the feedstock – and compared them to the consolidated strain performance scores (Figure 6-5, Figure 6-6).

The hourly profit is a measure of the economic throughput of bioprocess. It is influenced by both the yield and the productivity of the strain, and is optimized at the medium growth rate range (Figure 6-5D-F, Figure 6-6D-F). On the other hand, the batch profit is a measure of the profit generated per unit of feedstock. It is influenced by the titer and the yield, and is optimized at very high yields (Figure 6-5G-I, Figure 6-6G-I). Whereas the consolidated strain performance (Figure 6-5A-C, Figure 6-6A-C) is completely unaffected by the price point, the hourly profit and the batch profit are greatly affected by the price point. In addition, whereas the batch profit always favors high-yield (slow growing) strains, the hourly profit favors high yield strains at lower price points, and favors high productivity strains at higher price points. This effect is very prominent in the BDO-producing strains: whereas the hourly profit is optimized near 0.15 hr\(^{-1}\) at the 2X
price point, it is optimized near 0.3 hr\(^{-1}\) at the 8X price point (Figure 6-6D-F), which is near the productivity optimal point (Figure 6-3C).

Figure 6-5. The economic performance of the succinate-producing strains

The consolidated strain performance (A, B, C), hourly profit (D, E, F), and profit per batch (G, H, I) of the succinate-producing strains at three different succinate price points. The horizontal dotted line (D-I) indicates the zero profit line.
Figure 6-6. The economic performance of the BDO-producing strains

The consolidated strain performance (A, B, C), hourly profit (D, E, F), and profit per batch (G, H, I) of the BDO-producing strains at three different BDO price points. The horizontal dotted line (D-I) indicates the zero profit line.
It is important to note that there is a tradeoff between the two economic metrics – a strain that maximizes the hourly profit will produce a suboptimal batch profit (Figure 6-5d-I, Figure 6-6D-I). The relative importance of these metrics depends on higher-level design objectives: i.e., whether to maximize economic throughput or the economic return on feedstock investment. Nevertheless, a strain is only economically viable if both the hourly profit and the batch profit are above zero. In a real world engineering setting, additional costs of the bioprocess, such as separation cost, should be included in the analysis. This zero-profit line is indicated by the horizontal dotted line (Figure 6-5D-I, Figure 6-6D-I). As the price point drops, this line rises in relation to both the hourly profit and the batch profit curves, making higher growth rate strains economically unviable. At the price points tested, all the strains we designed (YZ1, YZ2, YZ3, YZ4, YZ5) were viable, although the BDO strains (YZ4, YZ5) produced very little profit at the 2X price point. In addition, the hourly profit of the BDO strains is very sensitive to the price point. For example, a process utilizing YZ5 is barely viable at the 2X price point, but becomes highly profitable at the 8X price point from an hourly profit perspective. This result demonstrates that in general, the economic performance of a strain may be highly sensitive to the price of the product.

One significant caveat of this economic viability analysis is that neither metric took downstream processing costs and equipment depreciation into consideration. Nevertheless, despite the simplicity of the metrics, the results are both useful and illuminating. In particular, strain designs with poor economic performance can be eliminated early in the development cycle. Additionally, in applications where the engineering goal is to maximize profitability, financial metrics can replace the consolidated strain performance as the strain selection criteria in Steps 4 and 9 of the DySSeO algorithm (see Method). The ability to perform such economic assessment is a distinct benefit of the DySSeO strategy.
6.3.7. Alternative Approaches

A key assumption we made in this work is that the flux balance analysis (FBA) method is able to quantitatively predict the growth rate and the by-product secretion rates with sufficient accuracy. An underlining assumption of FBA is that the cellular objective is the maximization of biomass growth. It has been shown that alternative formulations, such as the minimization of metabolic adjustment (Segrè et al, 2002), regulatory on-off minimization (Shlomi et al, 2005), and optimal resource allocation objective (Schuetz et al, 2007; 2012) can improve the predictive capability of the metabolic model under certain conditions. In particular, the minimization of metabolic adjustment and the regulatory on-off minimization methods may be more appropriate for simulating the metabolism of the knockout strains. To adopt these alternative formulations in DySScO, Equation 6-6 can be replaced by the appropriate formulation.

During the development of the DySScO strategy, we have considered a slightly different strategy. First, we would generate a large number of strain designs using exiting algorithms; then, we would evaluate the dynamic properties of these strains using dFBA. Although this approach is slightly more straight-forward than the DySScO strategy, it is much more time consuming. By scanning the production envelope first for an optimal growth rate range, we were able to significantly reduce the time required to find the optimal strain. Our experience is that the advantage of the DySScO strategy is more obvious in cases where the tradeoff between growth rate and product yield is more severe.

Anesiadis et al. (2008) suggested that the tradeoff between growth rate and product yield can be balanced using a genetic toggle-switch: a strain can be developed to grow fast initially (making little or no product) and once sufficient biomass is produced, a genetic switch is triggered to cause the strain to stop growing and focus on chemical production. This is an interesting alternative approach, and should be investigated alongside DySScO in the future.
6.4. Concluding Remarks

The first model-based strain-design algorithm, OptKnock, was developed a decade ago (Burgard et al., 2003). Since then, numerous strain-design algorithms with expanded capacity and improved computational time have been developed (Lun et al., 2009; Pharkya & Maranas, 2006; Ranganathan et al., 2010; Rocha et al., 2010; Yang et al., 2011). Unfortunately, despite the impressive innovations in these algorithms, they are limited by the stoichiometric nature of the metabolic models—without the addition of a process model, these algorithms are incapable of using dynamic design criteria such as titer and productivity. This is a significant limitation from an industrial perspective because the economic value of a strain is evaluated by the combination of yield, titer, and productivity. We addressed this limitation by integrating the dFBA method with the existing strain design algorithms to form the DySScO strategy.

In this proof of concept study, we applied the DySScO strategy to the problems of designing succinate-producing and BDO-producing E. coli strains. We evaluated the tradeoffs between growth rate, product yield, volumetric productivity, and titer, as well as the economic viability of the designed strains. We showed that the economic performance of a strain can be strongly affected by the price difference between the product and the feedstock. As petroleum prices continue to rise, commodity chemical prices are likely to increase as well, making more bioprocesses economically attractive. In the future, it would be interesting to develop a framework that integrates the DySScO strategy with financial projection models, thereby evaluating the economic potential of the strain. This kind of innovation will help broaden the application of model-based strain-design in the industrial context.
Chapter 7. Synthesis

7.1. A Tale of Tradeoffs

A tradeoff is a situation in which the gaining in quality or quantity of something necessitates the losing in quality or quantity of something else. Issues of tradeoffs arise quite commonly in engineering settings, and are often studied through Pareto optimality analysis. The present thesis discusses the resolution of three issues that prevent the broader application of metabolic modeling in biotechnology; interestingly, each of these issues is associated with one or more tradeoffs that stems from the first principles. The investigation of the this topic is of great theoretical interest, and can lead to interesting engineering solutions; as such, a brief discussion of these tradeoffs is warranted.

Chapter 2 of this thesis investigated the hypothesis that Flux Balance Analysis may not be applicable in nutrient-rich environments because it cannot predict the overflow metabolism. It is well known that the overflow metabolism involves a kind of tradeoff between the rate and efficiency of ATP production (Pfeiffer & Schuster, 2001; Schuster et al, 2008b; 2008a), leading to an observable tradeoff between the growth rate and growth yield (MacLean, 2008). Previous discussions of this tradeoff (Pfeiffer & Schuster, 2001; Schuster et al, 2008b) assumed that this tradeoff stems from the limited capacity of the respiratory pathway (Majewski & Domach, 1990; Varma & Palsson, 1994) and the redirection of the excess glucose towards the less efficient fermentative pathway. We suggest that this assumption is incorrect for E. coli. In E. coli, the overflow phenomenon might arise from the limited membrane available for protein expression: at high glucose uptake rates, the number of ETC enzymes on the membrane must be reduced to make room for the additional glucose transporters, leading to a reduced oxygen uptake rate and utilization of the fermentative pathways (Chapter 2). As such, the fundamental reason for this tradeoff appears to be the membrane occupancy constraint, rather than the respiratory capacity constraint.
Chapters 3, 4, and 5 of this thesis focused on the modeling of the community metabolism. Both theoretical considerations as well as experimental evidence (Zhuang et al., 2011a; Chapters 4, 5) demonstrated that the engineering of microbial communities requires the understanding of a tradeoff that stems from the 2nd law of thermodynamics. The 2nd law states that the entropy of any non-equilibrium system is always increasing – this implies that for any organism to escape “entropy death”, the organism must continue to dissipate energy (mostly through chemical reactions). Higher energy dissipation allows faster chemical reactions and a higher overall growth rate, but leaves less energy to be converted into biomass, hence creating the tradeoff between the catabolic rate and the growth yield. In most cases, the organisms with higher catabolic rate will also have a higher growth rate, thus producing a tradeoff between the growth rate and growth yield. Any microbial community containing species that grow at different rates (which is the majority of the communities) will be affected by this tradeoff. For example, we found that the nutrient-rich environment created by acetate addition favors the rate-type Geobacter over the yield-type Rhodoferax and SRB (Zhuang et al., 2011a; Chapter 4). We were able to make use of this tradeoff to develop a novel uranium bioremediation strategy (Chapter 5). It is worth mentioning that the two aforementioned tradeoffs between the growth rate and yield are often mentioned together in the literature (Pfeiffer & Schuster, 2001; Schuster et al., 2008b; MacLean, 2008), and both are referred to as the “rate vs. yield tradeoff”; this can lead to some confusion. It would be useful in the future to make a stronger distinction between these two tradeoffs as they are rooted in very different physical phenomena.

Chapter 6 of this thesis investigated a number of tradeoffs during biochemical production. During biochemical production, the carbon source can be converted to either biomass or product; as such, there is a fundamental tradeoff between the product yield and the growth yield, as exemplified by the production envelope of a given microbial strain (Figure 6-2, Figure 6-3). In the setting of the microbial cell factory, this biological tradeoff between the growth yield and the product yield can potentially create a number

---

8 The growth yield of both Geobacter and SRB are in fact very similar. However, since sulfate-reduction is less favorable than iron-reduction, SRB must have a higher energy efficiency to achieve similar growth yield than Geobacter.
of industrially important tradeoffs: the tradeoffs between product yield, titer, productivity. Unlike the original biological tradeoff, these new tradeoffs are nonlinear and time-dependent in nature, and thus require novel tools to analyze; this lead to the developed of the Dynamic Strain Scanning Optimization (DySScO) strategy (Chapter 6).

Based on the experience and knowledge gained from this thesis, it is clear that biological tradeoffs play important and relevant roles in the engineering of biotechnology. A mechanistic understanding of such tradeoffs can provide answers to previously puzzling questions – such as the explaining of the overflow metabolism in *E. coli* (Chapter 2) – or be leveraged for novel biotechnology applications, such as novel uranium bioremediation strategies (Chapter 4 and 5) and novel strain design strategies (Chapter 6). In the future, it would be very interesting to investigate situations where the aforementioned tradeoffs are present in the biotechnology application, such as the directed evolution of a microbial cell factory.
7.2. Assumptions, Limitations, and Appropriateness

The central framework and the main theoretical perspective of this thesis is the constraint-based approach to metabolic modeling. In this section, we will explore the consequences and implications of two fundamental assumptions in the constraint-based approach, namely, the quasi-steady-state assumption and the optimality principle, including the inability to predict internal and external metabolite concentrations, which is related to all three engineering limitations (Section 1.3) this thesis sought to address. In addition, the limitations and appropriateness of some constraint-based methods, including FBA<sup>ME</sup>, dFBA, and DyMMM, are also discussed (A comprehensive list of model assumptions and limitations is given in the Appendix II), particularly in the context of environmental biotechnology.

7.2.1. From Kinetics to Constraints

In recent years, the constraint-based approach to metabolic modeling and systems biology have gained acceptance as powerful computational framework for biological and evolutionary investigation as well as metabolic engineering. Prior to its development, cellular metabolism is modeled as a system of ordinary differential equations (ODEs) that describes the rate of change of metabolite concentrations (Machado et al., 2012). This dynamic approach was limited to simple metabolic networks (Joshi & Palsson, 1990) because its requirement for kinetic parameters is experimentally prohibitive (Varma & Palsson, 1994; Machado et al., 2012). Using the quasi-steady-state (QSS) assumption, the constraint-based approach reduced the system of ODEs to a mass-balanced flow-network model (Price et al., 2004; Terzer et al., 2009), and thus eliminated the need for kinetic parameters and rate laws. Because this flow-network model is underdetermined, the assumption of the optimality principle (OP) is required to produce a unique solution.
The benefits of the constraint-based approach notwithstanding, some theoretical and practical questions surrounding this successful modeling approach need to be addressed. Firstly, the transformation from a dynamic model to a flow-network model eliminates all concentration variables, which can lead to serious problems. Secondly, the optimality principle must be revisited in light of some recent experimental evidence. Thirdly, despite the improvements in genomic technologies and the standardization of the model reconstruction process, the reconstruction of high-quality genome-scale models still requires significant effort. The appropriateness of genome-scale models in different settings should be addressed. Here, we will provide our perspective on these issues within the settings we are familiar with, including community ecology, environmental biotechnology, and metabolic engineering.

7.2.2. The importance of concentration variables

The concepts of stock and flow are important in the modeling of dynamic systems (Sterman, 2002). A stock variable is a variable that measures the quantity of something that is accumulated over time, whereas a flow variable is a variable that measures the change in the quantity of something over time. In other words, flow variables are the first derivative of stock variables with respect to time. It is essential to distinguish stocks from flows because their interactions can be highly counterintuitive (Sterman, 2002). For example, the volume of water in a tank may increase even though the inflow is reduced, and the debts of a nation can increase even though the deficit is reduced (Sterman, 2002). A model that contains both stock and flow variables can help us understand and explore such counterintuitive phenomena.

A dynamic metabolic model uses both stock and flow variables. Specifically, the metabolic flows are related to metabolite and enzyme concentrations through kinetic rate laws. In contrast, a constraint-based metabolic model contains only flow variables, namely, the metabolic fluxes (Price et al, 2004; Terzer et al, 2009). Given that the changes in metabolite and enzyme concentrations can lead to the reprogramming of the metabolic network, the lack of concentration variables in the constraint-based model is a
significant omission. For example, a recent study evaluated the optimality principle in *Escherichia coli* by comparing the metabolic flux predictions to the C-13 flux measurements, and concluded that the overflow metabolism in *E. coli* is caused by a shift in the cellular objective function (Schuetz *et al*, 2007). The primary rationale in this study is that the constraint-based model cannot predict the overflow metabolism without introducing alternative objective functions. However, two recent studies demonstrated that the introduction of simple constraints that relate the metabolic flows to the enzymatic concentrations allows the constraint-based model to predict the overflow metabolism in *E. coli* without alternative objectives (Vazquez *et al*, 2008; Chapter 2). These demonstrations invalidated the rationale of the first study, which in hindsight, is an artifact of the lack of modeling of the stock variable. Additionally, without concentration variables, engineering metrics such as the volumetric productivity and the titer cannot be calculated (Chapter 6).

### 7.2.3. Reconnecting fluxes to concentrations

Fortunately, a number of methods have been developed to relate the metabolic fluxes to the concentrations of molecules both internal and external to the cell. Three methods we are most familiar with are the dynamic flux balance analysis (dFBA), flux balance analysis with molecular crowding (FBA\textsuperscript{MC}), and flux balance analysis with membrane economics (FBA\textsuperscript{ME}). The dFBA method (Mahadevan *et al*, 2002) focuses on relating the metabolic fluxes to the concentration of metabolites in the environment. Essentially, dFBA connects the constraint-based model of the cell with a dynamic process model of the bioreactor. In this way, the dFBA is able to predict the dynamic response of the metabolic network to environmental changes (Mahadevan *et al*, 2002). Both the FBA\textsuperscript{MC} and the FBA\textsuperscript{ME} methods recognize that concentrations of molecules inside the cell are limited by the real estate of the cell, and relate the metabolic fluxes to this limitation. In FBA\textsuperscript{ME}, the concentrations of the membrane enzymes are constrained by the surface area to volume ratio of the cell (Chapter 2). In FBA\textsuperscript{MC}, the concentrations of all macromolecules in the cell are constrained by the volume of the cell (Beg *et al*, 2007;
Vazquez et al, 2008). The two methods are mathematically similar, and have both been used to investigate the reprogramming of the metabolic network, particularly in the study of the overflow metabolism (Chapter 2; Vazquez et al, 2008).

There are two notable similarities between the aforementioned methods. The first similarity between these methods is that none required the abandoning of the QSS assumption, which states that the rates of change of the concentrations of the internal metabolites are negligible at the timescale of the growth rate. The dFBA dealt exclusively with external metabolites, FBA$^{ME}$ dealt with enzymes, and FBA$^{MC}$ avoided the issue by lumping all macromolecules (enzymes and metabolites) together as “crowding costs”. The QSS assumption might become a hindrance if we were to model the concentration changes of an internal metabolite, such as NADH. Nevertheless, there is undeniable value in the capacity to model the metabolic response to changes in the internal metabolite concentrations.

The second similarity is that all three methods are highly sensitive to the variations in kinetic parameters. This is an important caveat because in our experience, while some parameter values can be found in the literature, their accuracy can be questionable and there are larger variations in the measurement of the same parameter from different sources. Nevertheless, these methods require far fewer parameters than a full-scaled dynamic metabolic model, and do not require detailed knowledge of the enzyme rate laws. In this sense, they strike a balance between model capacity and experimental effort required. Furthermore, even in lieu of well-characterized kinetic parameters, the usage of averaged and cleverly fitted parameters may be sufficient to elucidate important biological insights (Beg et al, 2007; Chapter 2).

### 7.2.4. Exploring the Optimality Principle

The assumption of the *optimality principle* is essential to the generation of unique solutions using the constraint-based approach. In FBA, the maximization of the flux through the biomass reaction is often used as the objective function. Typically, this is
interpreted as the maximization of the growth rate. However, the same objective function
has also been interpreted as the maximization of biomass yield (Schuetz et al, 2007;
2012; Schuster et al, 2008b), which is somewhat confusing. The hidden assumption of
this second interpretation is that substrate uptake rate is the binding constraint, which is
not necessarily true – in FBA\textsuperscript{ME} and FBA\textsuperscript{MC}, the binding constraints are the membrane
occupancy constraint and the molecular crowding constraint respectively. The better
approach is to interpret the maximization of the biomass flux as the maximization of
growth rate, and calculate the biomass yield (Feist & Palsson, 2010) based on the growth
rate and the substrate uptake rates.

The optimality principle in microorganism has been demonstrated in a number of
adaptive evolution experiments (Ibarra et al, 2002; Fong et al, 2003; Lewis et al, 2010) –
when exposed to a single substrate, the organism will evolve towards an optimal
utilization of this substrate. However, in the natural environment, the organism is
exposed to a variety of alternatingly available substrates. The organism must
continuously readjust its metabolic network in order to utilize the available substrates.
Since this metabolic readjustment comes at an energetic cost, it should be minimized. A
recent study compared the C-13 flux measurements to the metabolic flux predictions
using different objective functions, and concluded that metabolism in nature operates
close to but below growth rate optimality (Schuetz et al, 2012). When exposed to
multiple substrates, the microorganism will seek a balance between the maximization of
growth rate for the present condition and the minimization of the adjustment required to
shift to a different condition. This insight is very important to the modeling of
microorganisms in the natural setting, particularly for metabolic generalists such as \textit{E.
coli} (Mahadevan & Lovley, 2008). In this setting, a multi-objective linear program
should be used to calculate metabolism of the microorganism.
7.2.5. When is the genome-scale model needed?

By eliminating the need for enzyme kinetics and rate laws, the constraint-based approach enabled the reconstruction of the genome-scale metabolic models. The development of the constraint-based approach enabled the reconstruction of genome-scale metabolic models (Thiele & Palsson, 2010; Terzer et al, 2009). To date, over 50 genome-scale metabolic models have been developed (Oberhardt et al, 2009), leading to discoveries and applications in a variety of settings (Oberhardt et al, 2009; Gianchandani et al, 2010; Feist et al, 2010; Mahadevan et al, 2010). Unfortunately, the reconstruction of a genome-scale model remains an arduous and time-consuming task (Thiele & Palsson, 2010). In our experience, the manual curation process can take between 6 months to a year, whereas the automated model reconstruction process (Henry et al, 2010) is not yet reliable. In addition, despite the reduction in sequencing cost, the availability of the genome sequence remains a hurdle. In particular, not all species can be cultured in the laboratory. In light of these barriers, alternative models should be examined. For example, both central metabolic models (Varma & Palsson, 1994; Covert et al, 2001) and core-pathway models (Chapter 4, Chapter 5) have been used for both research and engineering purposes. In addition, traditional kinetic models such as the Monod model are still used in many settings. Whether a genome-scale model is required, and which alternative to use if it is not required, must be evaluated in a case-by-case fashion based on the specific application.

In metabolic engineering, since each reaction is a potential engineering target, we would like to have as comprehensive a model as possible. However, since the computational complexity of the strain-design problem typically increases exponentially with model size (Yang et al, 2011), the usage of smaller models can be effective. Alternatively, novel algorithms such as EMILiO can effectively reduce the computational complexity (Yang et al, 2011). It should be noted that it would be useful to integrate the dFBA method with strain design algorithms in order to evaluate dynamic metrics such as the titer and the volumetric productivity (Chapter 6). On the other hand, in a research setting, tradeoff between reaction coverage and model generality must be considered – the more reactions included in a model, the more specific the model becomes. For example, a generic
central metabolic model can potentially be used to represent all organisms, whereas genome-scale models can represent only one organism. If the research goal is to discover a general principle applicable to many organisms, a more generic model can be used instead of the genome-scale model. For example, in a previous study, we have used the genome-scale *E. coli* model and the FBA\textsuperscript{ME} method to demonstrate that the prokaryotic metabolism is constrained by the availability of the cytoplasmic membrane for protein allocation (Chapter 2). We were able to derive the same conclusion using a generic central metabolic model (Covert *et al*, 2001; 2001). On the other hand, if the model is used for contextualizing high-throughput data or discovering emergent properties of the metabolic network, then a genome-scale model is appropriate.

### 7.2.6. Model Requirements for Environmental Biotechnology

Recently, there is a growing interest in the application of the constraint-based approach to the development of environmental biotechnologies. Given our experience and familiarity with this area, we would like to discuss the issues of the concentration variables, the optimality principle, and the appropriateness of the genome-scale models within this context.

First and foremost, in the natural environment, microorganisms exist in communities. For example, the uranium-contaminated aquifer at Rifle Colorado is populated with a wide variety of microorganisms including several *Geobacters*, at least two sulfate-reducing bacteria, *Rhodoferax*, and other species of lesser importance (Anderson *et al*, 2003). Often, one species is wholly reliant on the community for its survival. For example, the *Dehalococcoides* species cannot survive outside the KB1 culture (Ahsanul Islam *et al*, 2010). As such, metabolic models designed for the environmental biotechnology setting should be community models (Zengler & Palsson, 2012; Zhuang *et al*, 2011a). Since the members of the community exchange metabolites through a common environmental pool, the environmental concentrations of metabolites must be modeled. In addition, since different species grows at different rates, and the flow of metabolite in the environment is greatly affected by number of cells, either the number of cells or the
biomass abundance should be modeled as well. As such, methods that model the community as a large flow network containing no concentration variables (Zomorrodi & Maranas, 2012; Stolyar et al, 2007) are inappropriate for this setting.

In our opinion, the best tool currently available for the modeling of community metabolism is the dynamic multispecies metabolic modeling (DyMMM) framework (Zhuang et al, 2011a). The major advantage of this framework is that any constraint-based models of individual organisms can be integrated into a community model. However, currently, the DyMMM framework has several limitations. First, although the DyMMM framework models the concentrations of the metabolites in the environment, it focuses solely on the metabolic processes. However, the metabolite concentrations in the environment can be strongly influenced by geochemical and transport phenomenon. This caveat can be addressed by integrating the DyMMM model with geochemical and reactive transport models (Fang et al, 2011). The second issue of the DyMMM framework is that it assume that each species maximize its growth rate. At the individual organism level, this assumption is appropriate for metabolic specialists such as \textit{Rhodoferax} and \textit{Geobacter} species (Mahadevan & Lovley, 2008). However, as the versatility of the organism increases, the cost of changing between metabolic modes should be considered (Schuetz et al, 2012), and a multi-objective formulation that balances the maximization of growth rate and the minimization of metabolic adjustment should be used instead. At the community-level, this assumption is valid for competitive and cross-feeding communities, but is inappropriate for symbiotic communities. The simulation of symbiotic community requires the development of a community-level objective function (Zomorrodi & Maranas, 2012). One possibility is to maximize the total biomass yield with respect to carbons consumed.

Although a community metabolic model is required for environmental biotechnology, this community model must be composed of individual species. To date, a number of genome-scale models (Mahadevan et al, 2006; Sun et al, 2009; Risso et al, 2009; Flynn et al, 2012; Ahsanul Islam et al, 2010) have been developed for environmental microorganisms. A common issue is that the models are reconstructed based on the genome sequences of laboratory strains, which can differ significantly from the strains in
the natural environment. For example, *Geobacter* is an iron-reducing organism that is capable to reductive remediation of uranium. Although the genome-scale models of *Geobacter sulfurreducens* and *Geobacter metallireducens* are available, neither is detected during the bioremediation process. Instead, a number of similar organisms from the same family are detected. One approach is to develop a pan-genome model of *Geobacter*. However, from our extensive experience in the modeling of this organism, we question the appropriateness of the genome-scale model. In the uranium bioremediation setting, since we are unlikely to genetically modify the organisms in the environment, the focus of our research is on the understanding of the community interactions for improving the bioremediation process. In particular, we need to understand how the growth rates of the organisms change in response to changes in the metabolite concentrations in the environment.

One alternative is to create a kinetic community model using the Monod model of individual species. For the *Geobacter* modeling case, Zhao *et al.* (2011) performed a global sensitivity analysis and found that when the acetate concentration is high (> 3mM), the Monod model prediction is comparable to the genome-scale model. On the other hand, as the acetate concentration decreases, the predictions began to differ. In particular, if two or more substrates are modeled, and both their concentrations are low, the dual-Monod formulation over penalizes the growth rate of the organism. Additionally, the Monod model cannot accurately predict the switching between different metabolic modes. For example, if ammonium is not available in the environment, *Geobacter* will enter a nitrogen fixation mode where its biomass yield decreases and the flux through the respiratory pathway increases (Zhuang *et al.*, 2011a). If the metabolism of the organism is diverse and has many metabolic modes and many substrates, the Monod model would not be able to predict which metabolic mode is active during which conditions.

In our experience, the better alternative to the genome-scale models in the environmental setting is the core-pathway model. The core-pathway model is a constraint-based model containing only a minimal number of stoichiometric equations of the organism. For example, for the *Geobacter* species, which oxidizes acetate and reduces iron and
uranium, the core-pathway model contains two core reactions representing iron and uranium reduction respectively as well as the metabolite uptake and secretion reactions. A nitrogen fixation reaction can also be included. The construction of this model took less than a day, whereas the construction of the original *Geobacter sulfurreducens* took over a year. The core-pathway model is generally applicable to any *Geobacter* species, and its growth rate prediction with respect to changing acetate and iron concentrations is comparable to the genome-scale model. It is superior to the dual-Monod model because the growth rate is penalized twice by the dual-Monod model when both acetate and Fe(III) concentrations are low. As such, the core-pathway model is a valuable alternative to the genome-scale model for modeling environmental microorganisms.

### 7.2.7. Parting Thoughts

From an epistemological perspective, all models are wrong (Sterman, 2002) as they are abstractions of reality containing limitations and assumptions. An essential task of the modeling process is to find the limitations and hidden assumptions of the model and determine whether they interfere with the purpose at hand. In our experience, the process of enumerating the limitations and hidden assumptions of your own model can be both daunting and humbling. (See Appendix II for the result of this enumeration). Nevertheless, if you invest in the effort to do so, you will be rewarded with the knowledge of your model’s limitations and the confidence in knowing that you have used your model appropriately.
7.3. Future Recommendations

The objective of this thesis was to address three limitations associated with the existing metabolic modeling methods in order to broaden its applicability in biotechnology (Section 1.4). To a degree, this objective is satisfied through the development of the FBA\textsuperscript{ME} (Chapter 2), DyMMM (Chapter 3–5), and the DySScO (Chapter 6) methods. However, there remain some limitations to these methods (Section 7.2, Appendix II). In this section, we will make some recommendations for future works.

In Chapter 2, we described the FBA with Membrane Economics method, which was able to predict the overflow phenomenon in \textit{E. coli}, and thus verified the validity of the FBA approach in nutrient-rich environments. We also showed that the membrane occupancy theory is the only theory of overflow metabolism capable of explaining all metabolic characteristics associated with respiro-fermentation in \textit{E. coli}. If we are to prove beyond any doubt that the membrane constraint is the true cause of the overflow metabolism in \textit{E. coli}, it will be necessary to directly observe the cytoplasmic membrane of this organism. A possible experiment is to tag the ETC proteins and the glucose transporters of \textit{E. coli} using traceable markers, and grow \textit{E. coli} in a glucose-limited chemostat at different dilution rates. If visible markers are used, we should be able to estimate the total percent of the membrane covered by the glucose transporters and ETC enzymes – if the membrane occupancy theory is correct, this coverage value should not change after the glucose uptake rate increases above a critical value (when the membrane becomes saturated). In addition, if our theory is correct, it should be able to predict the changes in the abundance of these proteins. A different experiment is to over-express inactive membrane proteins under anaerobic conditions; if our theory is correct, the maximum glucose uptake rate of the organism should decrease linearly with the amount of inactive proteins expressed. These experiments would serve as irrefutable proof for the membrane occupancy theory of overflow metabolism. Another future possibility is to investigate the interaction between the membrane occupancy constraint (Chapter 2) and the volumetric crowding constraint (Beg \textit{et al}, 2007) in eukaryotic cells. It is possible that the membrane occupancy constraint is inactive in eukaryotic cells due to the existence of the
mitochondria, leaving the volumetric constraint the primary constraint on their metabolism. *S. cerevisiae* is an excellent organism for this kind of analysis because its physiology and morphology is well studied and the genome-scale metabolic model is available.

In Chapters 3, 4, and 5, we described the Dynamic Multi-species Metabolic Modeling framework, as well as its application in the investigation of community metabolism during uranium remediation. We also used the community model to show that, in theory, the simultaneous addition of acetate and Fe(III) may be a long-term viable uranium remediation strategy. The field-scale testing of this strategy has been discussed. A successful field-test will irrefutably demonstrate the value of constraint-based metabolic modeling in environmental biotechnology. In the future, it will be useful to couple the DyMMM framework with an online optimization framework to determine the real-time acetate and Fe(III) addition requirements. Furthermore, it will be useful to couple the DyMMM framework with spatial geochemistry model for more detailed investigations. Lastly, it is possible to model mutualistic communities by integrating the bi-level community modeling approach mentioned in section 3.4.3 (Zomorrodi et al., 2012) with the DyMMM framework; the resultant model will have an optimization framework that maximizes the total biomass of the community as the outer layer, and have the DyMMM community model as the inner layer. This kind of model could be a useful tool for studying symbiotic communities.

In Chapter 6, we have described a novel strain-design strategy called Dynamic Strain Scanning Optimization (DySScO). By using product yield, titer, and volumetric productivity as explicit design criteria, the DySScO can potentially optimize the economic value of the designed strains. In Chapter 6, we only demonstrated proof-of-principle application of this strategy by optimizing the metabolic network of the *E. coli* strains designed for producing succinate and 1,4-butanediol. In the future, it would be useful to validate the DySScO strategy experimentally by constructing the recommended strains and measuring their yield, titer, and productivity when grown in batch and fed-batch conditions. In addition, in Chapter 6, the economic values of the strain designs were evaluated using simplified metrics with assumed price values. In the future, the
metric used for this evaluation should be expanded to include up-to-date price information. An interesting possibility would be to integrate DySScO with modeling tools from other disciplines; for example, financial projection models may be used to determine the values of productivity, yield, and titer, thereby truly maximizing the economic value of the strain.

Taking a broader perspective, it will be truly interesting to integrate the metabolic modeling tools presented in this thesis with tools and ideas from other disciplines. An already mentioned possibility is to integrate financial models with the DySScO framework in order to optimize the economic value of the design strains. Another possibility is to use the DyMMM framework to model the global interactions between the biosphere and the atmosphere. For example, it is possible to construct a community model that includes representative species from each environmentally relevant microbial group (e.g., iron reducers, iron oxidizers, methanogens, methanotrophs, sulfate reducers, sulfate oxidizers, firmicutes etc.) and investigate how the global microbial community will respond to anthropogenic increase in CO₂. A third possibility is to integrate metabolic modeling tools with industrial ecology tools such as life cycle analysis in order to produce truly sustainable biotechnologies. Through this kind of large interdisciplinary endeavors, we will discover the true potential of constraint-based metabolic models.


Bethke CM, Ding D, Jin Q & Sanford RA (2008) Origin of microbiological zoning in groundwater flows. Geology. 36: 739-742


Henry CS, DeJongh M, Best AA, Frybarger PM, Linsay B & Stevens RL (2010) High-


Lemke MJ, McNamara CJ & G Leff L (1997) Comparison of methods for the


Mahadevan R, Bond DR, Butler JE, Esteve-Nuñez A, Coppi MV, Palsson BO, Schilling


Pfeiffer T & Bonhoeffer S (2002) Evolutionary Consequences of Tradeoffs between
Yield and Rate of ATP Production. Zeitschrift für Physikalische Chemie 216: 51-63


Zhuang K, Yang L, Mahadevan R (Submitted to Microbial Cell Factories). DySScO: an efficient strain design algorithm for balanced yield, titer, and productivity.

Appendix I. Nomenclature

Metabolic model

A metabolic model is a mathematical model of cellular metabolism. In this thesis, unless otherwise specified, the usage of the term refers to constraint-based metabolic models.

Constraint-based Reconstruction and Analysis

*Abbreviation.* COBRA

*Alternative.* Constraint-based metabolic modeling

COBRA is an approach to metabolic modeling and systems biology that models the cellular metabolism as a flow-network constrained by reaction stoichiometry and other constraints. A COBRA model is also referred to as a constraint-based metabolic model.

Reconstruction

The term “reconstruction” refers to the development process of a constraint-based metabolic model.

Flux

*Alternative.* Metabolic flux

In the context of metabolic models, the “flux” is a measure of the flow rate of metabolite through a metabolic reaction. The term “metabolic flux” can also be used.

Biomass Reaction

The biomass reaction described the process where biomass precursors are used to synthesize biomass.
Flux Balance Analysis

*Abbreviation.* FBA

FBA is a widely used constraint-based method that predicts the metabolic fluxes by maximization a cellular objective. Typically, the flux through the biomass reaction is maximized.

Exchange Reaction

An exchange reaction is a metabolic reaction where the metabolite moves into or out of the cell.

Exchange Flux

The flux through the exchange reaction.

*Unit.* mmol/gdw/hr

Uptake Rate

The rate a metabolite is taken up by the cell. In the context of metabolic models, it is the exchange flux through an uptake reaction.

*Unit.* mmol/gdw/hr

Minimization of Metabolic Adjustment

*Abbreviation.* MOMA

The MOMA algorithm minimizes the difference between the metabolic flux distributions of wild-type and knockout organisms

Regulatory On/Off Minimization

*Abbreviation.* ROOM

The ROOM algorithm minimizes the number of significant metabolic changes between wild-type and knockout organisms.
Optimality Principle

*Abbreviation.* OP assumption

The optimality principle is a fundamental assumption of flux balance analysis. It states that under a constant environment, the metabolic fluxes are optimized to maximize the cellular objective function in this environment.

Quasi-steady-state Assumption

*Abbreviation.* QSS assumption

The quasi-steady-state assumption is a fundamental assumption of constraint-based approach to metabolic modeling. It states that because the time-scale of metabolism is much faster than those of regulation and cell growth, the metabolic network can be assumed to be at a quasi-steady-state.

Biomass Yield

The quantity of biomass produced per unit of substrate.

*Unit.* gdw/mmol

Product Yield

The product yield is the quantity of the desired chemical produced from one unit of substrate.

*Unit.* mol/mol

Productivity

*Alternative.* Volumetric Productivity
The productivity is the quantity of the desired chemical produced per unit of time. In a bioprocess context, productivity generally refers to the volumetric productivity, which is the quantity of the desired chemical produced in one unit of reactor volume during one unit of time.

*Unit.* mmol/L/hr

**Titer**

The titer is the concentration of desired chemical at the end of the bioprocess.

*Unit.* mmol/L

**Flux Balance Analysis with Membrane Economics**

*Abbreviation.* FBA\textsuperscript{ME}

The FBA\textsuperscript{ME} is an extended FBA formulation that considers the protein crowding effect at the cytoplasmic membrane.

**Flux Balance Analysis with Membrane Crowding**

*Abbreviations.* FBA\textsuperscript{MC}, FBAwMC

*Alternative.* FBA with solvent capacity constraint

The FBA\textsuperscript{ME} is an extended FBA formulation that considers the macromolecular crowding effect within the limited cell volume.

**Dynamic Flux Balance Analysis**

*Abbreviation.* dFBA

The dFBA is a dynamic extension to FBA that integrates FBA with a process dynamic model of the bioreactor.

**Dynamic Multispecies Metabolic Modeling framework**

*Abbreviation.* DyMMM
The DyMMM framework dynamic community metabolic modeling framework based on the dFBA formulation.

**Genetic Design through Local Search**

*Abbreviation.* GDLS

The GDLS algorithm is a model-based strain-design algorithm that maximizes the product yield.

**Dynamic Strain Scanning Optimization strategy**

*Abbreviation.* DySScO

The DySScO strategy is a model-based strain design strategy that seeks to maximize the some function of the product yield, titer, and volumetric productivity.
Appendix II. Major Assumptions and Limitations

1. Standard Constraint-based Metabolic Model
   a. Assumptions
      i. The cellular metabolism is at quasi-steady-state
      ii. The cellular metabolism follows the optimality principle
   b. Limitations
      i. It does not contain any concentration variables, which lead to the following consequences:
         1. Cannot model the reprogramming of the cell due to changes in external and internal metabolite concentrations
         2. Cannot model the reprogramming of the cell due to changes in enzyme concentrations
         3. Process-level metrics such as the volumetric productivity and the titer cannot be calculated
      ii. It requires an appropriate objective function for accurate simulations
         1. Even with the objective function, there is the possibility of having multiple equally optimal solutions.

2. Flux Balance Analysis with Membrane Economics
   a. Assumptions
      i. The cellular metabolism is at quasi-steady-state
      ii. The cellular metabolism follows the optimality principle
      iii. The cellular metabolism is constrained by the membrane occupancy constraint
      iv. S/V ratio of the cell is constant as long as the physiological state is unchanged.
      v. In our implementation of the FBA<sup>ME</sup> in Chapter 2, the fermentative pathways are assumed to have no membrane cost. This is a simplification due to the lack of information.
   b. Limitations
      i. The central assumption of the FBA<sup>ME</sup> method still needs to be verified through direct experimental methods. One possibility is to experimentally characterize the abundance of different membrane proteins under different growth conditions.
      ii. FBA<sup>ME</sup> requires the kinetic parameters for all the membrane-bound enzymes in the model.
         1. This can be very difficult to acquire, particularly in lesser-known organisms.
         2. Fitted parameters were used in Chapter 2.
3. The only membrane enzymes modeled in Chapter 2 were the glucose transporter and the cytochrome oxidases. Including additional membrane enzymes may improve the prediction accuracy.

iii. The usefulness of this method for eukaryotic organisms is untested.

3. **Dynamic Multispecies Metabolic Modeling framework**
   
a. **Assumptions**
   
i. The cellular metabolism is at *quasi-steady-state*
   
   ii. The cellular metabolism follows the *optimality principle*
   
   iii. Each individual organism optimize its cellular objective without regarding the overall community
   
   b. **Limitations**
   
   i. DyMMM requires kinetic parameters for selected uptake take reactions, which can be difficult to acquire. Some parameter fitting is required.
      
      1. In chapter 4 and 5, the Vmax of hard-to-use Fe(III) exchange reaction had to be fitted.
   
   ii. DyMMM models used in Chapter 4 and 5 do not model the geochemistry, which is a significant limitation in some environmental settings.
   
   iii. Because of assumption iii, a DyMMM model is inappropriate for the modeling of cooperative or symbiotic communities.
   
   iv. The genetic identities of the members of the community must be known. In most environmental cases, only the more abundant species will be modeled. As such, this approach might miss out important species that are few in number.
   
   v. It is difficult to use the DyMMM model to interpret meta-genomic data.

4. **Dynamic Strain Scanning Optimization strategy**

   a. **Assumptions**

   i. The cellular metabolism is at *quasi-steady-state*

   ii. The cellular metabolism follows the *optimality principle*

   b. **Limitations**

   i. Some knowledge of the bioprocess, including the setup of the bioreactor and the feeding strategy, is required.

   ii. The DySScO strategy is limited by the limitations of the strain design algorithm used.

   iii. The economic metric used in Chapter 6 is very rudimentary.
Appendix III. Chapter 2 Supplementary Information

III-1. Model Sensitivity

In order to analyze the sensitivity of the FBA$^\text{ME}$ model with respect to C$^*$ values, we have performed 50 simulations with C$^*$ values that are randomly generated between 75% and 125% of its original value. The +/- 25% variation is sufficiently large since C$^*$ value of Cyo is only 150% times of the C$^*$ value of Cyd-I, and many independent experiments have clearly demonstrated that $C^{*\text{cyo}} > C^{*\text{cyd-I}} > C^{*\text{cyd-II}}$. (For example, Tseng et al. 1996, Bekker et al, 2009). The 50 simulations showed that the system is robust against up to 25% perturbation in C$^*$ values (Figure III-1). The general trend behind the utilization of cytochromes remains the same - Cyo for low glucose uptake rate, Cyd-II for high glucose uptake rate, and Cyd-I for microaerobic conditions. The critical glucose uptake rates at which alternative metabolic strategies are adopted show minor variations with varying C$^*$ values.

Figure III-1. Fifty simulations were performed with C$^*$ values that are randomly generated between 75% and 125% of its original value. The result shows that the model is robust against parameter variations.
III-2. Comparison with C-13 Flux Analysis

We compared predictions from FBA and FBA\textsuperscript{ME} with the C-13 flux analysis of glucose-limited chemostat growth performed by Vazquez \textit{et al.} (2008). As illustrated in Figure III-2, although both FBA\textsuperscript{ME} and FBA predict the repression of the TCA cycle flux at higher growth rate, the predictions from FBA\textsuperscript{ME} are closer to experimental observations. As demonstrated below, FBA with molecular crowding could not predict TCA cycle repression.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3.png}
\caption{Predicted fluxes through the TCA cycle compared with C-13 flux analysis data.}
\end{figure}

III-3. Construction of FBA\textsuperscript{ME} model

The FBA\textsuperscript{ME} model of \textit{E. coli} is adapted from that of iAF1260 model (Feist 2007) with the following modifications. The iAF1260 model contains multiple version of NADH dehydrogenases (using both menaquinol and ubiquinol) and both Cyo and Cyd-I. It does not contain Cyd-II.

In the FBA\textsuperscript{ME} model, the NADH dehydrogenases and cytochrome oxidases contained in the original iAF1260 are removed, and replaced with reactions NDH-I, NDH-II, CYO, CYD-I, and CYD-II. Only two NADH dehydrogenases are included and are assumed to only use ubiquinol (q8 and q8h2 metabolites in the model), which simplifies the analysis of the ETC system. The glucose exchange reaction, “EX\textsubscript{glc(e)}” in the iAF1260 model (representing all glucose transporters) is replaced with a new version of “EX\textsubscript{glc(e)}” reaction. The new cytochrome oxidase reactions (CYO, CYD-I, CYD-II) and the new “EX\textsubscript{glc(e)}” reaction each contains a pseudo-metabolite called “CMC” which stands for “cytoplasmic membrane cost”. This pseudo-metabolite is used to represent the usage of the cytoplasmic membrane area by transmembrane-proteins. A “cytoplasmic membrane demand” reaction (DM\textsubscript{CMC}) is also added; the upper bound of this reaction is set to the
relative cytoplasmic budget of 1. Together, the CMC pseudo-metabolite and the DM_CMC reaction allow us to implement the membrane budget constraint (below) using COBRA Toolbox.

\[ \sum v_i C_i^* \leq 1 \]

III-4. Calculating the Relative Membrane Cost in *E. coli*

### Membrane Constraint Binding Requirement during Parameter Estimation

Because the membrane occupancy constraint (SI-1) is the only constraint where the cost parameters appears in our model, this constraint must be binding \((C_i v_i = 1)\) for us to acquire unique cost parameters. This implies that only data generated under experimental conditions where the membrane occupancy constraint is the limiting constraint can be used to determine the membrane cost values.

Mathematically, the FBAME can be written as:

- maximize growth rate flux
- such that

\[ \sum S v = 0 \]  \hspace{1cm} (R1)
\[ \sum v_i C_i^* \leq 1 \]  \hspace{1cm} (R2)
\[ v_{\text{glc}} \leq v_{\text{glc}}^{\text{max}} \]  \hspace{1cm} (R3)

Here, R1 represents the standard pseudo-steady-state assumption used in all FBA models – it provides no information regarding the reaction fluxes (See Feist 2007 for more detail). R2 is the membrane constraint, and R3 is the glucose uptake constraint. R3 is used when simulating the glucose-limited growth in a chemostat, and is not used when simulating batch growth where the glucose uptake rate is not limiting.

When R3 is not used (e.g., when we’re calibrating for glucose transporter cost), the binding constraint is automatically satisfied in the optimal solution because R2 is the only constraint that limits the flux through the metabolic network. However, when both R2 and R3 are used (e.g., when we’re calibrating for cytochrome costs), if the glucose availability is very low, the flux through the network may be limited by R3 instead of R2 – in such cases, the binding requirement is not satisfied.
To ensure that the binding requirement is satisfied, we have chosen data generated from batch experiments and from chemostat experiments with sufficiently high glucose uptake rates. Our simulations also confirms that the membrane constraint binding requirement is satisfied for all cases used for parameter estimations.

**Relative Cost of C*<sub>GUR</sub>**

The maximal anaerobic glucose uptake rate of *E. coli* was measured to be 18 mmol/gdw/h (Portnoy 2008) under glucose-excess conditions. We can assume that cytochrome oxidases are not present under anaerobic condition (Tseng *et al.* 1996), and that glucose transporters are the sole occupants of the cytoplasmic membrane (for the purpose of our simplified model). Assuming the cell optimizes its growth rate, we can vary the C*<sub>GUR</sub> parameter until the FBA<sup>ME</sup> model (which contains the membrane budget constraint (Eq SI-1) by definition) correctly predicts the measured glucose uptake rate of 18 mmol/gdw/h. Note: the glucose uptake reaction is left unconstrained to reflect the glucose-excess condition. At C*<sub>GUR</sub>=0.0556 g h/mmol, FBA<sup>ME</sup> predicted an oxygen uptake rate (OUR) of 18 mmol/gdw/hr and a growth rate of 0.44 hr<sup>-1</sup>. In this case, the membrane is predicted to be saturated: \( \sum v_i C^*_i = 0.0556 \times 18 = 1 \).

**Relative Cost of C*<sub>CYO</sub>**

The relative costs of cytochrome oxidases are calculated similarly using experimental data from Bekker *et al.* (2009). In this study, the cytochrome knockouts are grown with a fixed growth rate of 0.15±0.1 hr<sup>-1</sup> in glucose-limited chemostats. For the Cyo strain (∆cyd1∆cyd2), the measured oxygen uptake rate was 6.4±0.4 mmol/gdw/h, and the measured glucose uptake rate was 2.3 mmol/gdw/h.

To determine the Cyo cost, a glucose uptake constraint of 2.3 mmol/gdw/h is used to reflect the glucose-limiting growth condition of this experiment. Since the C*<sub>GUR</sub> is known (from above calculation), and only Cyo exists in this strain, only the C*<sub>CYO</sub> parameter is both unknown and relevant. The Cyd-I and Cyd-II reactions are deleted during this process. We can determine the value of C*<sub>CYO</sub> by manually varying it until the model correctly predicts both the growth rate and the oxygen uptake rate. At C*<sub>CYO</sub> = 0.0658 g h/mmol, FBA<sup>ME</sup> predicted that the Cyo strain would have a growth rate of 0.15 hr<sup>-1</sup> and an oxygen uptake rate of 6.7 mmol/gdw/hr. 6.7 is within one standard deviation of the measured average oxygen uptake rate, which is 6.4±0.4 mmol/gdw/h. In this case, the membrane is predicted to be saturated: \( \sum v_i C^*_i = 0.0658 \times 6.7 \times 2 + 0.0556 \times 2.3 = 1 \).

**Relative Cost of C*<sub>CYD-I</sub>**
For the Cyd-I strain (ΔcyoΔcyd2), the measured oxygen uptake rate was 8.8±1.3 mmol/gdw/h, the measured glucose uptake rate was 3.2 mmol/gdw/h, and the growth rate was again fixed at 0.15±0.1 hr\(^{-1}\).

C\(^{*}\)\(_{CYD-I}\) is similarly determined using FBA\(^{ME}\) by constraining the model to the measured glucose uptake rate of 3.2 mmol/gdw/h, and varying the C\(^{*}\)\(_{CYD-I}\) parameter until the model correctly predicts both the oxygen uptake rate and the growth rate. The Cyd-II and Cyo reactions are deleted during this process. Since C\(^{*}\)\(_{CYD-I}\) is usually not expressed under oxygen-rich conditions, the assumption here is that the cell is able to optimally manage C\(^{*}\)\(_{CYD-I}\) expression when both Cyo and Cyd-II are knocked out. At C\(^{*}\)\(_{CYO}\) = 0.0427 g h/mmol, FBA\(^{ME}\) predicted that the Cyd-I strain would have a growth rate of 0.15 hr\(^{-1}\) and an oxygen uptake rate of 9.62 mmol/gdw/hr. 9.62 is within one standard deviation of the measured oxygen uptake rate, which is 8.8±1.3 mmol/gdw/h. In this case, the membrane is predicted to be saturated: \(\sum v_i C^*_i = 0.0427 \times 9.62 \times 2 + 0.0556 \times 3.2 = 1\).

**Relative Cost of C\(^{*}\)\(_{CYD-II}\)**

For the Cyd-II strain (ΔcyoΔcyd1), the measured oxygen uptake rate was 11.1±0.6 mmol/gdw/h, the measured glucose uptake rate was 6 mmol/gdw/h, and the growth rate was again fixed at 0.15±0.1 hr\(^{-1}\). Unlike the mutant containing Cyo only, which grew at a very low glucose uptake rate, the knockout containing Cyd-II only grew at a high glucose uptake rate of 6 mmol/gdw/h and consumes oxygen at a rate of 11 mmol/gdw/h. At this glucose uptake rate, the “membrane economics theory” predicts that the cell membrane of wild-type *E. coli* is shared by Cyo and Cyd-II. Assuming the genetic programming remain functional in the knockout strain (since the cells were not given time to evolve in the experiment), and expression level of Cyd-II is expected to be the same in the knockout strain as in the wild-type growing at a glucose-uptake rate of 6 mmol/gdw/h.

To determine the Cyd-II cost, a glucose uptake constraint of 6 mmol/gdw/h is used to reflect the glucose-limiting growth condition of this experiment. Both Cyo and Cyd-II reactions are left unconstrained, C\(^{*}\)\(_{CYD-II}\) value is manually varied until 11±0.6 mmol/gdw/h of oxygen is processed through the Cyd-II reaction. At C\(^{*}\)\(_{CYD-II}\) = 0.0128 g h/mmol, FBA\(^{ME}\) predicted that the wildtype would have an oxygen uptake rate of 14.38 mmol/gdw/h. The flux through Cyo is 5.63 mmol/gdw/h, the flux through Cyd-I is 0, and the flux through Cyd-II is 23.12 mmol/gdw/h (11.55 mmol/gdw/h of oxygen, which falls within 11.1±0.6 mmol/gdw/h). In this case, the membrane is predicted to be saturated: \(\sum v_i C^*_i = 5.63 \times 0.0658 + 23.12 \times 0.0128 + 6 \times 0.0556 = 1\).
III-5. Estimation of Membrane Protein Content

We were able to estimate the cellular content of cytochrome oxidases and ATP synthase under several conditions based on experimental measurements and numerical data from BioNumbers database. In this section, BNID refers to the entry ID number in BioNumbers database.

**Number of Polypeptides in E. coli Cytoplasmic Membrane**

The average cell mass of *E. coli* is 0.5 pg/cell (Philips and Milo, 2009), and about 50% of the cell mass is protein (Feist *et al.* 2007). This means that the average protein content of *E. coli* is 0.25 pg/cell, which is similar to the measured values (BNID: 104878, 104879, 104880). The average size of the polypeptides (a monomeric protein) is 40 kDa in *E. coli* (BNID: 105861) or 6.42x10^{-20} g. This means that the total number of polypeptides in *E. coli* is about 3,900,000. About 18% of all polypeptides in *E. coli* is located in the cytoplasmic membrane (BNID: 100019), therefore, the number of polypeptides in the cytoplasmic membrane is about 700,000.

**Cytochrome Content at Maximum Oxygen Uptake**

The maximum oxygen uptake rate (OUR) is 18 mmol/gdw/h in *E. coli* (Feist *et al.* 2007). The kcat of Cyd-II is 70 O_2/enzyme/s, or 252000 O_2/enzyme/h. Assuming only Cyd-II is used at the maximum OUR, then 54000 enzymes are needed. Since Cyd-II is dimeric, the number of polypeptides required is 108000, which is about 15% of all membrane polypeptides.

**Cytochrome Content at Microaerobic Condition**

Under microaerobic conditions, the expression level of Cyd-I can reach above 250 nmol/g protein (Alexeeva *et al.* 2000). Assume *E. coli* protein content is 0.25pg/cell, each cell contains about 38000 Cyd-I enzymes, or 76000 Cyd-I polypeptides, nearly 11% of the all membrane proteins.

**Glucose Transporter Content**

The maximum glucose uptake rate (GUR) is 10.7 mmol/gdw/h in *E. coli* (Feist *et al.* 2007). The Kcat of the PtsG transport system is 756000 Glucose/enzyme/h. Assuming only the PtsG system is used at the maximum GUR, each cell requires 4260 enzymes or 25560 polypeptides (PtsG contains six subunits). This accounts for 4% of all membrane polypeptides. Philips and Milo (2009) has estimated that PtsG accounts for more than 4%
of the membrane surface, or 8% of the available membrane surface. The difference between these two estimates suggests that PtsG takes up more membrane area per polypeptide, which is understandable as oligomers are more complex structurally than monomers.

**ATP Synthase Content at Optimal Growth**

FBA predicts that the flux through ATP synthase reaction during optimal *E. coli* growth is 52 mmol/gdw/h. The kcat of ATP synthase is 1260000 ATP/enzyme/h. This means each cell requires 4141 ATP synthase. Since each ATP synthase contains 22 polypeptides, about 13% of all membrane polypeptides are ATP synthase.

### III-6. Evidence for Decrease in Respiratory Efficiency

Past theories have attributed the observed tradeoff between the ATP yield and the catabolic rate solely to the utilization of fermentative pathways, resulting in lower ATP per glucose metabolized than oxidative phosphorylation. However, chemostat experiments suggest that the decrease in the respiratory efficiency itself may contribute to the overall yield decrease (Figure III-3).

A statistically significant (p = 0.08) decrease in biomass yield was observed between $D=0.3 \text{ h}^{-1}$ and $D=0.4 \text{ h}^{-1}$, with the acetate secretion commencing at $D = 0.4 \text{ h}^{-1}$ at a very low level (Figure SI-6). The decrease in biomass yield from $D = 0.3 \text{ h}^{-1}$ to $D = 0.4 \text{ h}^{-1}$ (0.014 g/mmoll based on average yields) is much larger than the yield decrease due to acetate secretion (0.0039 gdw/mmol, based on stoichiometric calculation using FBA), presenting evidence that the reduction in yield is a consequence of decreased respiratory efficiency rather than mere redirection of carbon from biomass to acetate.

FBA$^{\text{ME}}$ provides mechanistic insight into the decrease in respiratory efficiency. As the membrane became saturated with glucose transporters at high rate of glucose uptake, the costly Cyo was replaced by the cheaper but less efficient Cyd-II, resulting in an increased ATP production rate at the expense ATP yield. Therefore, the outcome of the competition for space on the membrane between Cyo and Cyd-II appears to play a pivotal role in deciding the trade-off between the rate and yield of ATP production. This is evident from gene expression data (Figure 2-3A, B).
Figure III-3. Triplicate glucose-limited chemostat experiments of *E. coli* growth at various dilution rates (Vemuri et al., 2006).
IV-1. Determination and sensitivity analysis of *Geobacter*’s maximal uptake rate

Figure IV-1. Determination and sensitivity analysis of *Geobacter*’s maximal uptake rate (A) and uptake affinity (B) of Hard-to-use Fe(III) using the aqueous Fe(II) concentration data from 2002 Rifle Experiment (Anderson *et al.* 2003). This figure shows that a Vmax value of 30 mmol/gdw/hr and a Km of 1 mM can best predict the production of aqueous Fe(II).
IV-2. Ideal Batch Reactor Simulations with Low Initial Biomass

Figure IV-2. Ideal batch reactor model simulations with varying initial ratio of Geobacter and SRB beginning with cell numbers 100x less than originally examined. In the cell number row, green line indicates Geobacter, blue line indicates SRB.
IV-3. Ideal Batch Reactor Simulations with High Initial Fe(III) Concentration

Figure IV-3. Ideal batch reactor model simulations with varying initial ratio of *Geobacter* and SRB beginning with 5mM Easy-To-Use Fe(III) more than doubling the original amount of total Fe examined. In the cell number row, green line indicates *Geobacter*, blue line indicates SRB.
IV-4. Fe(III) amendment simulation with 5mM initial Fe(III) concentration

Figure IV-4. *In silico* predictions of trends in the number of SRB and the number of *Geobacter* as well as acetate, sulfate, and Fe(III) concentrations starting with 5mM Easy-To-Use Fe(III) simulating an Fe(III) amendment at the start of the experiment.
Appendix V. Chapter 6 Supplementary Information

V-1. Modeling Succinate Inhibition in *E. coli*

In our idealized fedbatch, without succinate inhibition modeling (blue lines), the model predicts the rapid production of biomass (B) as well as succinate (D) until the bioreactor becomes full (A), at which point glucose is no longer added to the system. If succinate inhibition is modeled (red and green lines), the cell growth is much slower (B) and less succinate is produced (D). The cell growth terminates at the critical succinate concentrations (B), the reactor is not filled and the glucose is never fully consumed (D). The final succinate titer is much lower if the succinate inhibition effect is included.

Figure V-1. The liquid volume in the reactor (A), the biomass concentration (B), and the glucose (C) and succinate (D) concentrations of the succinate-producing strain YZ1 modeled with and without succinate-inhibition.
V-2. The Effect of Fedbatch time on Titer

Figure V-2. The predicted succinate production dynamics of the strains growing at 0.1 hr\(^{-1}\) and 0.25 hr\(^{-1}\) are shown. The titer of the strain (illustrated by the circles) is defined as the concentration of succinate at the end of fedbatch time. If the fedbatch time is long (eg. 120 hrs), then the higher yield 0.1 hr\(^{-1}\) strain will have a higher titer. On the other hand, if the fedbatch time is short, the faster growing 0.25 hr\(^{-1}\) strain will have a higher titer.
APPENDIX VI. List of Model Parameters

Chapter 2.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$C^*_\text{GUR}$</td>
<td>0.0556</td>
<td>gdw·hr/mmol</td>
<td>Appendix III-4</td>
</tr>
<tr>
<td>$C^*_\text{Cyo}$</td>
<td>0.0658</td>
<td>gdw·hr/mmol</td>
<td></td>
</tr>
<tr>
<td>$C^*_\text{Cyd-I}$</td>
<td>0.0427</td>
<td>gdw·hr/mmol</td>
<td></td>
</tr>
<tr>
<td>$C^*_\text{Cyd-II}$</td>
<td>0.0128</td>
<td>gdw·hr/mmol</td>
<td></td>
</tr>
</tbody>
</table>

FBA uptake constraints

<table>
<thead>
<tr>
<th>Uptake Type</th>
<th>Value</th>
<th>Unit</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oxygen</td>
<td>18</td>
<td>mmol/gdw/hr</td>
<td>Feist et al. 2007</td>
</tr>
<tr>
<td>Glucose</td>
<td>20</td>
<td>mmol/gdw/hr</td>
<td>Portnoy et al. 2008</td>
</tr>
</tbody>
</table>

Chapter 4 and Chapter 5.

**Geobacter Kinetics**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{\text{max}}^{\text{ac,1}}$</td>
<td>13.3</td>
<td>mmol/gdw/hr</td>
<td>Zhuang et al. 2011a</td>
</tr>
<tr>
<td>$V_{\text{max}}^{\text{ac,2}}$</td>
<td>2</td>
<td>mmol/gdw/hr</td>
<td></td>
</tr>
<tr>
<td>$V_{\text{max}}^{\text{ac,3}}$</td>
<td>2.67</td>
<td>mmol/gdw/hr</td>
<td></td>
</tr>
<tr>
<td>$K_s^{\text{ac,1}}$</td>
<td>0.777</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$K_s^{\text{ac,2}}$</td>
<td>0.0167</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$K_s^{\text{ac,3}}$</td>
<td>0.012</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$V_{\text{max}}^{\text{Fe(III),easy}}$</td>
<td>586</td>
<td>mmol/gdw/hr</td>
<td>Appendix IV-1</td>
</tr>
<tr>
<td>$K_s^{\text{Fe(III),easy}}$</td>
<td>1</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$V_{\text{max}}^{\text{Fe(III),hard}}$</td>
<td>30</td>
<td>mmol/gdw/hr</td>
<td></td>
</tr>
<tr>
<td>$K_s^{\text{Fe(III),hard}}$</td>
<td>1</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>Parameter</td>
<td>Value</td>
<td>Unit</td>
<td>Source</td>
</tr>
<tr>
<td>-----------</td>
<td>-------</td>
<td>------------</td>
<td>-------------------------</td>
</tr>
<tr>
<td>$V_{\text{max}}^{\text{U(VI)}}$</td>
<td>1.8</td>
<td>mmol/gdw/hr</td>
<td>Zhao et al. 2011</td>
</tr>
<tr>
<td>$K_s^{\text{U(VI)}}$</td>
<td>0.036</td>
<td>mM</td>
<td></td>
</tr>
</tbody>
</table>

**Geobacter attachment and detachment during growth phase**

<table>
<thead>
<tr>
<th>$K_{\text{detach}}$</th>
<th>0.01</th>
<th>hr(^{-1})</th>
<th>Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{\text{detach}}$</td>
<td>0.05</td>
<td>hr(^{-1})</td>
<td></td>
</tr>
</tbody>
</table>

**Geobacter attachment and detachment during stationary and death phase**

<table>
<thead>
<tr>
<th>$K_{\text{detach}}$</th>
<th>0.05</th>
<th>hr(^{-1})</th>
<th>Fitted</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{\text{detach}}$</td>
<td>0.01</td>
<td>hr(^{-1})</td>
<td></td>
</tr>
</tbody>
</table>

**SRB Kinetics**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{\text{Ac}}$</td>
<td>3.2</td>
<td>mmol/gdw/hr</td>
<td>Ingvorsen et al. 1984</td>
</tr>
<tr>
<td>$S_{\text{Ac}}$</td>
<td>0.077</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$V_{\text{SO}_4}$</td>
<td>4.2</td>
<td>mmol/gdw/hr</td>
<td></td>
</tr>
<tr>
<td>$S_{\text{SO}_4}$</td>
<td>0.064</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$S_{\text{SO}_4}^{*}$</td>
<td>13</td>
<td>mM</td>
<td>Fitted</td>
</tr>
</tbody>
</table>

$S_{\text{SO}_4}^{*}$ is used for sediment bottle experiments because the sulfate reducers in this experiment seem to have a very high $K_s$ compared to the sulfate reducers in the natural environment.

**Death rates**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit (^{-1})</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_{\text{geo}}$</td>
<td>0.0011</td>
<td>hr</td>
<td>Bethke et al. 2008</td>
</tr>
<tr>
<td>$d_{\text{SRB}}$</td>
<td>0.0011</td>
<td>hr</td>
<td>Moosa et al. 2008</td>
</tr>
</tbody>
</table>

**Simulation of Sediment Bottle Experiment**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[\text{Fe(III)}]_{\text{easy}}^{t=0}$</td>
<td>0.16</td>
<td>mM</td>
<td>Calculated, see Chapter 4.2.6</td>
</tr>
<tr>
<td>$[\text{Fe(III)}]_{\text{hard}}^{t=0}$</td>
<td>5.1</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$[\text{Fe(II)}]^{t=0}$</td>
<td>2.25</td>
<td>mM</td>
<td></td>
</tr>
</tbody>
</table>
### Simulation of 2002 Rifle Experiment

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$[\text{Ac}]_{t=0}$</td>
<td>12</td>
<td>mM</td>
<td>Experimentally measured</td>
</tr>
<tr>
<td>$[\text{SO}<em>4^{2-}]</em>{t=0}$</td>
<td>10.5</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$[\text{Geobacter}]_{t=0}$</td>
<td>$3.4 \times 10^4$</td>
<td>cells</td>
<td></td>
</tr>
<tr>
<td>$[\text{SRB}]_{t=0}$</td>
<td>$3.4 \times 10^4$</td>
<td>cells</td>
<td></td>
</tr>
<tr>
<td>$[\text{Fe(III)}]_{t=0}$</td>
<td>0.16</td>
<td>mM</td>
<td>Calculated, see Chapter 5.2.6</td>
</tr>
<tr>
<td>$[\text{Fe(III)}]_{t=0}$</td>
<td>1.02</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$[\text{Fe(II)}]_{t=0}$</td>
<td>50</td>
<td>µM</td>
<td>Anderson et al. 2003</td>
</tr>
<tr>
<td>$[\text{U(VI)}]_{t=0}$</td>
<td>0.8</td>
<td>µM</td>
<td></td>
</tr>
<tr>
<td>Fe(II) inflow</td>
<td>0.0705</td>
<td>µM/hr</td>
<td></td>
</tr>
<tr>
<td>U(VI) inflow</td>
<td>0.011</td>
<td>µM/hr</td>
<td></td>
</tr>
<tr>
<td>$[\text{Ac}]_{t=0}$</td>
<td>0</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>$[\text{SO}<em>4^{2-}]</em>{t=0}$</td>
<td>7</td>
<td>mM</td>
<td></td>
</tr>
<tr>
<td>Dilution Rate</td>
<td>0.0014</td>
<td>hr$^{-1}$</td>
<td>Zhuang et al. 2011a</td>
</tr>
<tr>
<td>Rate of Acetate Addition</td>
<td>0.0028</td>
<td>mM/hr</td>
<td>Anderson et al. 2003; Zhuang et al. 2011a</td>
</tr>
</tbody>
</table>

#### Chapter 6.

**E. coli Kinetics**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$V_{\text{max}}^{\text{oxygen}}$</td>
<td>0</td>
<td>mmol/gdw/hr</td>
<td>Anaerobic condition</td>
</tr>
<tr>
<td>$V_{\text{max}}^{\text{glucose}}$</td>
<td>20</td>
<td>mmol/gdw/hr</td>
<td>Portnoy et al. 2008</td>
</tr>
<tr>
<td>$K_s^{\text{glucose}}$</td>
<td>1</td>
<td>mM</td>
<td>Assumed</td>
</tr>
</tbody>
</table>

#### Succinate Inhibition

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Unit</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$S_{\text{succe,}^*}$</td>
<td>678</td>
<td>mM</td>
<td>Li et al. 2010</td>
</tr>
<tr>
<td></td>
<td>Batch Simulation</td>
<td>Fedbatch Simulation</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------------------------</td>
<td>-------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>[biomass]_{t=0}</td>
<td>0.01 g/L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[glucose]_{t=0}</td>
<td>20 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V_{t=0}</td>
<td>1 L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V_{max}</td>
<td>10 L</td>
<td></td>
<td></td>
</tr>
<tr>
<td>[glucose]_{feed}</td>
<td>1000 mM</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Batch time</td>
<td>50 hr</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Copyright Acknowledgements

Chapter 2 is adapted from the journal article “Economics of Membrane Occupancy and Respiro-fermentation” (Zhuang et al., 2011b) published Nature Molecular Systems Biology. This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License. This license allows the author (Kai Zhuang) to reuse this work in the thesis.

Chapter 3 is adapted from the supplementary information section of the journal article “Genome-Scale Dynamic Modeling of the Competition Between Rhodoferax and Geobacter in Anoxic Subsurface Environments” (Zhuang et al., 2011a) published in Nature ISMEJ. ISMEJ licensing agreement allows the author to reuse the work in future printed academic work (e.g., this thesis).

Chapter 4 is adapted from the journal article “Integrative Analysis of Geobacter spp. and sulfate-reducing bacteria during uranium bioremediation” (Zhuang et al., 2012a) published in Biogeosciences. This work is licensed under a Creative Commons Attribution License. This license allows the author (Kai Zhuang) to reuse this work in the thesis.

Chapter 5 is adapted journal article “The Design of Long-term Effective Uranium Bioremediation Strategy Using a Community Metabolic Model” (Zhuang et al., 2012b) published in Biotechnology and Bioengineering. The license agreement with Wiley-Blackwell Publishing allows the author (Kai Zhuang) to reuse this work in the thesis.