Development of Steady-State and Dynamic Flux Models for Broad-Scope Microbial Metabolism Analysis

Lian He

Follow this and additional works at: https://openscholarship.wustl.edu/eng_etds

Part of the Bioinformatics Commons, and the Chemical Engineering Commons

Recommended Citation

He, Lian, "Development of Steady-State and Dynamic Flux Models for Broad-Scope Microbial Metabolism Analysis" (2016). Engineering and Applied Science Theses & Dissertations. 159.

https://openscholarship.wustl.edu/eng_etds/159

This Dissertation is brought to you for free and open access by the Engineering and Applied Science at Washington University Open Scholarship. It has been accepted for inclusion in Engineering and Applied Science Theses & Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
WASHINGTON UNIVERSITY IN ST. LOUIS

School of Engineering and Applied Sciences
Department of Energy, Environmental and Chemical Engineering

Dissertation Examination Committee:
   Yinjie Tang, Chair
   Robert Blankenship
   Marcus Foston
   Tae Seok Moon
   Himadri Pakrasi
   Fuzhong Zhang

Development of Steady-State and Dynamic Flux Models for
Broad-SCOPE Microbial Metabolism Analysis

by
Lian He

A dissertation presented to the
Graduate School of Arts & Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2016
St. Louis, Missouri
# Table of Contents

List of Figures ........................................................................................................................................ v

List of Tables ........................................................................................................................................... x

Acknowledgments .................................................................................................................................. xi

Abstract ................................................................................................................................................... xiv

Chapter 1: Introduction to flux analysis techniques .............................................................................. 1

1.1. Flux analysis techniques: flux balance analysis and \(^{13}\)C metabolic flux analysis .............. 2

1.2. Flux analysis in dynamic systems ................................................................................................. 4

1.3. Advances and challenges ............................................................................................................... 6

1.4. This work ...................................................................................................................................... 7

1.5. References .................................................................................................................................. 8

Chapter 2: WUFlux - an open-source platform for \(^{13}\)C metabolic flux analysis of microbes .................................................. 12

2.1. Introduction .................................................................................................................................. 13

2.2. WUFlux Implementation ............................................................................................................... 14

2.3. Results and Discussions ............................................................................................................... 16

2.4. Conclusions .................................................................................................................................. 18

2.5. List of abbreviations ....................................................................................................................... 18

2.6. Availability and requirements ........................................................................................................ 18

2.7. Acknowledgements ........................................................................................................................ 19

2.8. References .................................................................................................................................... 19

Chapter 3: Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on \(^{13}\)C-metabolic flux analysis ......................................................................................... 27

3.1. Introduction .................................................................................................................................. 28

3.2. Materials and Methods .................................................................................................................. 30

  3.2.1. Plasmid and strain construction .............................................................................................. 30

  3.2.2 Bacterial cultivation with [1,2-\(^{13}\)C] glucose ........................................................................ 31
Chapter 5: Experimental analysis and model-based optimization of microalgae growth in photobioreactors using flue gas ................................................................. 76

5.1. Introduction ........................................................................................................... 77
5.2. Materials and Methods ....................................................................................... 78
  5.2.1. Algal cultivation medium and biomass measurement ....................................... 78
  5.2.2. Flue gas treatment using photo-bioreactors ....................................................... 79
  5.2.3. Kinetic model development ............................................................................. 79
  5.2.4. Dynamic optimization framework to profile optimal CO\textsubscript{2} concentrations ................................................................. 80
5.3. Results and Discussions ...................................................................................... 81
  5.3.1 Experimental Analysis of Microalgae growth on flue gas .................................. 81
  5.3.2. Model simulation of algal growth under different flue gas treatment ............... 82
  5.3.3. Optimal CO\textsubscript{2} conditions for microalgal growth ................................ 83
  5.3.4. Experimental verification and model limitations .............................................. 85
5.4. Conclusions ......................................................................................................... 85
5.5. Acknowledgements ............................................................................................. 86
5.6. References .......................................................................................................... 86
Supplemental materials ......................................................................................... 98

Chapter 6: Simulating cyanobacterial phenotypes by integrating flux balance analysis, kinetics, and a light distribution function ......................................................... 101

6.1 Introduction ......................................................................................................... 102
6.2. Materials and Methods ...................................................................................... 104
  6.2.1. Cell cultivation .............................................................................................. 104
  6.2.2. \textsuperscript{13}C-Labelling experiment .................................................................. 104
  6.2.3. Flux balance analysis model .......................................................................... 105
  6.2.4. Simulation of cyanobacterial growth via integrating FBA, kinetics, and cell movements ......................................................................................... 107
6.3. Results ................................................................................................................. 110
6.3.1. Simulation of cyanobacterial optimal growth in a cylindrical PBR ........................................ 110
6.3.2. Simulation of cyanobacterial oxidative pentose phosphate pathway in a cylindrical PBR ........................................................................................................................................................................ 111
6.3.3. Investigation of cyanobacterial photosynthesis efficiency in shake flasks ......................... 112
6.3.4. Model-based investigation of lactate production by engineered cyanobacteria in PBRs ........................................................................................................................................................................ 112
6.4. Discussion ........................................................................................................................................................................ 114
6.5. Conclusions ......................................................................................................................................................................... 116
6.6. Acknowledgements ............................................................................................................................................................. 116
6.7. References ........................................................................................................................................................................... 117
Supplementary Materials ........................................................................................................................................................... 127

Chapter 7: Glycine cleavage powers photoheterotrophic growth of Chloroflexus aurantiacus in the absence of H2............................................................................................................................................................. 137

7.1. Introduction ........................................................................................................................................................................ 138
7.2. Materials and Methods ......................................................................................................................................................... 139
7.3. Results .................................................................................................................................................................................... 140
7.4. Discussion ............................................................................................................................................................................. 145
7.5. Conclusions ........................................................................................................................................................................... 147
7.6. Acknowledgements ............................................................................................................................................................. 147
7.7. References ........................................................................................................................................................................... 148
Supplementary Materials ........................................................................................................................................................... 155

Chapter 8: Summary and future directions......................................................................................................................... 161

8.1. Summary ............................................................................................................................................................................. 162
8.1.1. Flux analysis reveals both carbon and energy metabolisms ................................................................. 162
8.1.2. Cyanobacteria feature high activities of Calvin cycle and low activities of TCA cycle ......................................................................................................................................................................... 163
8.2. Recommended future work ..................................................................................................................................................... 164
8.2.1. Genome-scale 13C-MFA ...................................................................................................................................................... 164
8.2.2. Including hydrodynamic information into integrated genome-scale model to accurately illustrate cell metabolism in subpopulations ................................................................................................................................. 165
8.3. References ........................................................................................................................................... 166

Appendix Chapter 1: $^{13}$C-MFA delineates the photomixotrophic metabolism of *Synechocystis* sp. PCC 6803 under light- and carbon-sufficient conditions .................................................................................. 167

Appendix Chapter 2: Photoheterotrophic fluxome in *Synechocystis* sp. strain PCC 6803 and its implications for cyanobacterial bioenergetics .............................................................................................................. 177

Appendix Chapter 3: Kinetic modeling and isotopic investigation of isobutanol fermentation by two engineered *Escherichia coli* strains ........................................................................................................... 186

Appendix Chapter 4: An ancient Chinese wisdom for metabolic engineering: Yin-Yang 196

Appendix Chapter 5: Elucidation of intrinsic biosynthesis yields using $^{13}$C-based metabolism analysis ......................................................................................................................................................... 206

Appendix Chapter 6: Biofuel production: an odyssey from metabolic engineering to fermentation scale-up ..................................................................................................................................................... 215

Appendix Chapter 7: WUFlux Manual ........................................................................................................... 224
List of Figures

Chapter 2

Figure 2.1: Graphical representation of applying WUFlux for $^{13}$C-MFA. ........................................ 22
Figure 2.2: Comparison of mass isotopomer distribution data in the control (A) and engineered (B) 
E. coli strains determined by WUFlux, METRAN, and INCA. ......................................................... 23
Figure 2.3: Comparison of relative flux distributions in the control (A) and engineered (B) E. coli 
strains determined by WUFlux, METRAN, and INCA................................................................. 24
Figure 2.4: Carbon and energy distributions in both control and engineered E. coli strains....... 25

Chapter 3

Figure 3.1: Fatty acid metabolism of E. coli and the “pull-push-block” strategy for fatty acid 
production. .................................................................................................................................... 51
Figure 3.2: Growth curves of the engineered fatty acid producing strain E. coli DH1 ΔfadE/pA58C- 
TR (A) and the control strain DH1 ΔfadE/pACYCDuet-1 (B) in batch cultures grown on 
glucose under aerobic conditions........................................................................................................ 53
Figure 3.3: Free fatty acid profiles produced by the engineered fatty acid producing E. coli strain 
(A) and the control strain (B). ............................................................................................................ 54
Figure 3.4: Metabolic flux distributions in the central metabolic pathways of the engineered fatty 
acid producing E. coli strain and the control strain during exponential growth on [1,2-$^{13}$C] 
glucose. ............................................................................................................................................. 55
Figure 3.5: Fold changes of transcription levels of the selected genes in the engineered fatty acid 
producing E. coli strain compared to the control strain. ................................................................. 57

Chapter 4

Figure 4.1: A Screenshot of INCA. ........................................................................................................ 71
Figure 4.2: Flux distribution of Synechocystis 6803 under photomixotrophic conditions. .......... 72
Figure 4.3: Flux distribution of Synechocystis 6803 under DCMU-induced photoheterotrophic 
conditions............................................................................................................................................. 73
Figure 4.4: Flux distribution of Synechococcus elongatus UTEX 2973 under photoautotrophic 
conditions............................................................................................................................................ 74
Figure 4.5: Mass isotopomer distribution fittings to the dynamic labeling data of intracellular free metabolites. ............................................................................................................................................. 75

Chapter 5

Figure 5.1: Diagram of the experiment setup. .................................................................................................................. 90
Figure 5.2: Chlorella growth curves. ........................................................................................................................................... 91
Figure 5.3: Effects of CO₂ volume fraction and inhibition constant (KI) on the biomass production. .................................................................................................................................................. 92
Figure 5.4: Simulation of microalgae growth (red lines) under CO₂ (15%) pulses at a frequency of 1 min gas-on/29 min gas-off in a 12 h-12 h light-dark cycle. ....................................................................................................... 93
Figure 5.5: Effect of pulse function on biomass production ........................................................................................................ 94
Figure 5.6: The optimal CO₂ concentration profiles. The model assumed that the cultures were grown under continuous light illumination for 7 days. .......................................................................................................................... 95
Figure 5.7: Simulation of microalgal growth under three CO₂ treatment modes in continuous illumination condition .............................................................................................................................................. 96
Figure 5.8: Effect of flue gas pulse modes on Chlorella growth (without buffer pretreatment). .. 97

Chapter 6

Figure 6.1: Algorithm for simulating cyanobacterial growth and intracellular flux distribution by integrating the flux analysis model, kinetics, and a light distribution function. ............................ 121
Figure 6.2: Simulations of dynamics of cyanobacterial performance in a 60mm-radius cylindrical PBR under 50 μE/m²/s surface light intensity. ................................................................................................................. 122
Figure 6.3: Test of biomass growth performance sensitivity to the mass transfer rate (K_La in h⁻¹), surface light intensity (I₀ in μE/m²/s), and the bioreactor geometry (R in mm). ...................... 123
Figure 6.4: The oxidative pentose phosphate (OPP) pathway in cyanobacteria ...................... 124
Figure 6.5: Comparison of experimental and simulation results of cyanobacterial growth rates in shake flasks. ......................................................................................................................................................... 125
Figure 6.6: Simulation results of D-lactate producing cyanobacteria performance. .................. 126
Figure 6-S1: The influence of time intervals on cyanobacterial growth. ........................................ 129
Figure 6-S2: Probability distribution of cell circulation time in photobioreactors. ................... 130
Figure 6-S3: Cyanobacterial growth with different probability distributions. .......................... 131
Figure 6-S4: Specific growth rates and fluxes through the oxidative pentose phosphate pathway under heterotrophic growth conditions in darkness. ................................................................. 132
Figure 6-S5: Simplified geometry of shake flask. .................................................................................. 133
Figure 6-S6: Determining the photosynthesis efficiency of cyanobacteria growing in 250 ml shake flasks. .................................................................................................................................. 134
Figure 6-S7: Comparison of flux distributions (unit: mmol/g/h) in central metabolism between wild-type, glycogen knockout, and D-lactate producing cyanobacterial strains. .............. 135

Chapter 7

Figure 7.1: Investigation of the key nutrients enhancing the photoheterotrophic growth of strain J-10-fl........................................................................................................................................... 150
Figure 7.2: Strain J-10-fl growth on sodium acetate (NaOAc, 2 g/L) in a minimal medium with or without glycine (0.05 g/L) addition under photoheterotrophic conditions. ......................... 151
Figure 7.3: Mass isotopomer distributions (MIDs) of selected proteinogenic amino acids ([M-57]^+) under photoheterotrophic condition ........................................................................................................ 152
Figure 7.4: Investigation of mass isotopomer distributions of proteinogenic amino acids to reveal CO2 fixation under photoheterotrophic conditions .................................................................................................................. 153
Figure 7.5: Schematic representation of C1 metabolism and CO2 fixation pathways in C.aurantiacus ........................................................................................................................................ 154
Figure 7-S1: Photoheterotrophic growth of J-10-fl strain in acetate-based medium with different concentrations of glycine (A) and serine supply (B). ......................................................... 158
Figure 7-S2: Photoheterotrophic consumption of glycine by strain J-10-fl (with 0.05 g/L glycine supplement) (n=2, 10-day cultures). ............................................................................................................................... 159
Figure 7-S3: The growth of C. aurantiacus OK-70-fl in the acetate medium (without H2). ..... 160
List of Tables

Chapter 1
Table 1.1: Mathematical framework of $^{13}$C MFA and FBA................................................................. 11

Chapter 2
Table 2.1: Comparison of flux estimations from WUFlux, METRAN and INCA................................. 21

Chapter 3
Table 3.1: Exponential growth parameters of the engineered and control E. coli strains grown on glucose in minimal medium........................................................................................................ 49
Table 3.2: Estimated production and consumption of NADPH, NADH, ATP and FADH$_2$ for the E. coli control and engineered strains. ........................................................................................ 50

Chapter 4
Table 4.1: Metabolic network used to measure the flux distribution of Synechococcus 2973 under photoautotrophic conditions........................................................................................................ 70

Chapter 5
Table 5.1: Parameter values used in the model......................................................................................... 88
Table 5.2: Maximum OD$_{730}$ increase rate observed (d$^{-1}$) within four days. ........................................... 89

Chapter 6
Table 6.1: List of parameters used to simulate the growth and metabolic fluxes of cyanobacteria growing in a cylindrical PBR........................................................................................................ 120
Table 6-S1: Parameters used for simulating cyanobacterial growth in shake flasks. ......................... 128

Chapter 7
Table 7-S1: Mass isotopomer distributions of proteinogenic amino acids of Chloroflexus aurantiacus J-10-fl strain grown under photoheterotrophic conditions. ................................. 155
Table 7-S2: Estimation of fractions of proteinogenic alanine in different labelling patterns................. 157
Acknowledgments

In the first year at WashU, I was very fortunate to do my research rotations in two wonderful research groups: Pakrasi lab and Tang lab. With little background on biology, I had the first research experience in microbiology in Professor Pakrasi’s lab. Dr. Anindita Bandyopadhyay and Dr. Lawrence Page served as my mentors, who had taught me many useful experimental skills. Even after my research rotation, postdocs and graduate students in Pakrasi lab have been offering assistances to my research, for which I am truly grateful.

I officially joined Tang lab in 2011, starting research on metabolic flux analysis. It was then a challenging task for me, since I had zero experience with programming in MATLAB. Dr. Tang helped me by sharing his MATLAB code, from which I started to learn computational optimization and algorithms of $^{13}$C MFA. In the following years, Dr. Tang has been inspiring me with passions and great ideas, and he also offered me opportunities to collaborate with many researchers in different fields. I want to thank him for all his instructions and support for my PhD study.

I have enjoyed working with all my labmates. Some of them have graduated and started new adventures. In particular, I would like to thank Dr. Arul Varman (and his family) for not only sharing great ideas on research but also offering kind helps during my tough time in bay area. In the past three years, I have been working with Dr. Le You and Dr. Gang Wu on many projects. Our collaborations yielded much of the work in this dissertation. Other members in Tang lab, including Whitney Hollinshead, Mary Abernathy, Tolutola Oyetunde, and Ni Wan, also contributed to some of my work. I also want to thank the lab alumni, Dr. Yi Xiao, Dr. Xueyang Feng, Huifeng Yu, Xiaotong Duan, and Adrienne Reding.
I want to acknowledge all the collaborators from other universities or institutes. They are Professor Lifeng Peng from Victoria University of Wellington in New Zealand, Professor Yi Yu from Wuhan University in China, Professor Maciek Antoniewicz from University of Delaware, Professor Shulin Chen from Washington State University, and Dr. Seema Singh from Joint BioEnergy Institute. Those research experiences are valuable to me. I also want to acknowledge my committee members, including Professor Yinjie Tang, Professor Robert Blankenship, Professor Himadri Pakrasi, Professor Marcus Foston, Professor Tae Seok Moon, and Professor Fuzhong Zhang, for their valuable suggestions for my research projects.

I would like to thank all my friends. In the first year at WashU, I spent a great time with my friends, Lin Wang, Linjia Mu, and Zhaobang Liu. We supported each other to adapt to the new culture and environment (and survived the qualify exam). I will always remember the best summer trip with Dr. Yi Yu and Dianyi Liu in 2013. I am lucky to have two great roommates and friends, Zheng Luo and Fengbo Zhou, who have helped a lot in my daily life. I have also shared many great times with my friends in the same office: Yi Yu, Arul Varman, Le You, Di Liu, and Wen Jiang.

Last but not the least, I would like to express my sincere gratitude to my family for their love and support. My dissertation is dedicated to my parents and grandparents who encouraged me to take challenges and set off this journey to pursue my PhD study.

Lian He

Washington University in St. Louis

May 2016
Dedicated to my parents and grandparents.
ABSTRACT OF THE DISSERTATION

Development of Steady-State and Dynamic Flux Models for Broad-Scope Microbial Metabolism Analysis

by

Lian He

Doctor of Philosophy in Energy, Environmental and Chemical Engineering

Washington University in St. Louis, 2016

Professor Yinjie Tang, Chair

Flux analysis techniques, including flux balance analysis (FBA) and $^{13}$C-metabolic flux analysis (MFA), can characterize carbon and energy flows through a cell’s metabolic network. By employing both $^{13}$C-labeling experiments and nonlinear programming, $^{13}$C-MFA provides a rigorous way of examining cell flux distributions in the central metabolism. FBA, on the other hand, gives a holistic review of optimal fluxomes on the genome scale. In this dissertation, flux analysis techniques were constructed to investigate the microbial metabolisms.

First, an open-source and programming-free platform of $^{13}$C-MFA (WUFlux) with a user-friendly interface in MATLAB was developed, which allowed both mass isotopomer distribution (MID) analysis and metabolic flux calculations. Several bacterial templates with diverse substrate utilizations were included in this platform to facilitate $^{13}$C-MFA model construction. The corrected MID data and flux profiles resulting from our platform have been validated by other available $^{13}$C-MFA software.

Second, $^{13}$C-MFA was applied to investigate the variations of bacterial metabolism in response to genetic manipulations or changing growth conditions. Specifically, we investigated the central metabolic responses to overproduction of fatty acids in Escherichia coli and the carbon flow
distributions of *Synechocystis* sp. PCC 6803 under both photomixotrophic and photoheterotrophic conditions. By employing the software of isotopomer network compartmental analysis, we performed isotopically non-stationary MFA on *Synechococcus elongatus* UTEX 2973. The $^{13}$C-based analysis was also conducted for other non-model species, such as *Chloroflexus aurantiacus*. The resulting flux distributions detail how cells manage the trade-off between carbon and energy metabolisms to survive under stressed conditions, support high productions of biofuel, or organize the metabolic routes for sustaining biomass growth.

Third, conventional FBA is suitable for only steady-state conditions. To describe the environmental heterogeneity in bioreactors and temporal changes of cell metabolism, we integrated genome-scale FBA with growth kinetics (time-dependent information) and cell hydrodynamic movements (space-dependent information). A case study was subsequently carried out for wild-type and engineered cyanobacteria, in which a heterogeneous light distribution in photobioreactors was considered in the model. The resulting integrated genome-scale model can offer insights into both intracellular and extracellular domains and facilitate the analysis of bacterial performance in large-scale fermentation systems.

Both steady-state and dynamic flux analysis models can offer insights into metabolic responses to environmental fluctuations and genetic modifications. They are also useful tools to provide rational strategies of constructing microbial cell factories for industrial applications.
Chapter 1: Introduction to flux analysis techniques
1.1. Flux analysis techniques: flux balance analysis and $^{13}$C metabolic flux analysis

Computational flux analysis has emerged as an important and unique approach for quantifying the intracellular enzymatic reactions. Among different flux analysis methods, flux balance analysis (FBA) and $^{13}$C metabolic flux analysis (MFA) have received the most interests from the scientific community. Table 1.1 gives general frameworks of these two flux analysis techniques.

Traditional FBA solves constraint-based optimization problems (Orth et al. 2010). Typically, FBA starts with metabolic network reconstruction based on genome annotations, which includes thousands of biochemical reactions (and thus it is referred as the genome-scale FBA model). The biochemical metabolic network is then transformed into a linear system of equations based on reaction stoichiometry. With a defined objective function (for example, maximum biomass growth or maximum ATP production), FBA employs linear programming to characterize the optimal phenotype under steady-state conditions. Flux variability analysis (FVA) is regularly employed to evaluate the robustness of flux predictions. In addition, many more algorithms have been proposed and widely accepted as extensions of FBA, such as minimization of metabolic adjustment (MOMA) (Segre et al. 2002) and regulatory on/off minimization (ROOM) (Shlomi et al. 2005). Both aim to profile flux distributions in mutant strains by minimizing the metabolic responses to gene perturbations between wild-type and genetically engineered strains. Other algorithms, such as k-OptForce (Chowdhury et al. 2014), allow better designs of engineered strains to enhance production of desirable chemicals in microbial cell factories.

$^{13}$C-MFA permits a more rigorous fashion of analyzing the intracellular fluxes by combining both $^{13}$C-labeling experiment and non-linear optimization approach. In the $^{13}$C-labeling experiment, one or multiple $^{13}$C-labeled substrates are supplied into well-defined minimal medium for cell
cultivations. Consequently, the labeled carbons will be incorporated into all the intracellular free metabolites that constitute macromolecules such as proteins and lipids. The labeling patterns, or mass isotopomer distributions (MIDs), of free metabolites or proteinogenic amino acids are directly affected by the flux distribution in cells. To ensure an accurate flux calculation, parallel labeling experiment can be utilized, in which multiple $^{13}$C tracers are used in parallel under the same growth condition (Crown and Antoniewicz 2013). Based on different MIDs revealed by analytical instruments such as GC/MS and LC/MS, one can determine the fluxes computationally.

The computational part of $^{13}$C MFA basically involves simulation of metabolite MIDs by translating successive biochemical processes into mathematical and computational languages. Traditionally, isotopomer mapping matrices (IMM) are employed to allow $^{13}$C-carbons tracing from one metabolite to another (Schmidt et al. 1997). However, such an algorithm requires consideration of all the isotopologues for each metabolite, which substantially increases the number of variables. For example, to simulate a six-carbon metabolite, we need to include $2^6$ variables and correspondingly the same amount of isotopomer balance equations. An alternative but more efficient approach is the element metabolite unit (EMU) algorithm (Antoniewicz et al. 2007; Young et al. 2008), which has been used in most of current $^{13}$C MFA studies. To determine the confidence intervals of flux estimations, an accurate algorithm has been previously proposed (Antoniewicz et al. 2006).

Both flux analysis techniques have proved capable of offering insights into metabolic metabolisms of species in different domains, including bacteria (He et al. 2014; Nogales et al. 2012), yeast (Blank et al. 2005; Feng and Zhao 2013), plant (Mintz-Oron et al. 2012; Williams et al. 2010) and mammalian (Duarte et al. 2007; Quek et al. 2010) cells. For example, integrating both $^{13}$C-MFA and FBA unraveled bacterial adaptations to anaerobiosis (Chen et al. 2011), and analyzing cancer
cell metabolism by using isotopic tracers revealed some unique properties under exponential growth and non-growth phases (Ahn and Antoniewicz 2011). In addition to revealing metabolic flux phenotypes, FBA and $^{13}$C-MFA have been applied to fulfill other diverse research purposes. For instance, genome-scale FBA models can be used for targeting metabolites essential for pathogen survival, and thereby may contribute to effective novel drug development (Kim et al. 2011). $^{13}$C-MFA proves useful to provide valuable metabolic engineering strategies to enhance production of valuable chemicals (Guo et al. 2015). Several review articles have provided extensive and detailed information on applications of flux techniques (Feist and Palsson 2008; Oberhardt et al. 2009; Sauer 2006).

1.2. Flux analysis in dynamic systems

One major limitation which both conventional $^{13}$C MFA and FBA suffer from is their dependence on the assumption of homogeneity and steady state of cell cultures, and thus both fluxes and labeling patterns must remain constant for precise flux measurements. In many cases, however, those assumptions no longer hold. For example, in large scale fermentation system, due to heterogeneous distributions of nutrients, cell cultures cannot be considered as homogeneous. To address these issues, we need innovative dynamic flux approaches. These techniques are likely to permit evaluation of cell metabolism in transient states, non-growth conditions or in rich media, allow integration of flux analysis with kinetic model to simulate large-scale fermentation systems, and provide insights on both intracellular and extracellular domains.

In general, two typical computational ways of performing $^{13}$C dynamic MFA and dynamic FBA are time series-driven and kinetic model-driven approaches (Antoniewicz 2013). In the former approach, time-series data such as the metabolite concentrations over time are combined with
metabolic models to allow process monitoring and control (Leighty and Antoniewicz 2011). As an alternative, integrating FBA or $^{13}$C MFA with kinetic models can capture the dynamic changes of cell metabolisms of batch cultures (Feng et al. 2012).

In another particular case of $^{13}$C MFA, which is often referred as isotopically nonstationary MFA (INMFA), isotopic information of intracellular metabolites exhibits dynamic changes, while the flux distribution remains constant. This situation happens when, for example, all the regular glucose in medium is replaced by fully labeled glucose at the same level. Both IMM-based and EMU-based INMFA frameworks have been proposed to facilitate flux determination in transient states (Jazmin and Young 2013; Shasti and Morgan 2007). INMFA has been frequently used to evaluate autotrophic metabolisms of cyanobacteria and plant cells (Ma et al. 2014; Wu et al. 2015; Young et al. 2011).

INMFA or dynamic $^{13}$C-MFA usually requires knowledge on metabolite abundances, which can be challenging to be accurately determined. Precisely measuring the time-dependent labeling profiles of intracellular metabolites is also not straightforward. Similar to conventional $^{13}$C-MFA, the optimal flux is determined by minimizing the differences between experimental and computational data. However, due to the isotopically nonstationary condition, kinetic equations are required to simulate the dynamic changes in labeling profiles. Following is a general kinetic equation:

$$C_n \cdot \frac{dX_n}{dt} = A_n \cdot X_n + B_n \cdot Y_n,$$
where $X_n$ is a matrix of EMUs to be determined, $Y_n$ is a matrix of EMUs that are previously calculated, $A_n$ and $B_n$ are matrices that describe the metabolic network, and $C_n$ represents free metabolite concentrations.

### 1.3. Advances and challenges

The past few years have seen many great advances in metabolic flux analysis. For instance, FBA model continuously extends its ability of covering the entire genome information. Dynamic flux analysis, although more computationally challenging than classic $^{13}$C MFA, has enjoyed many developments and has been applied in many studies (Ahn and Antoniewicz 2011; van Heerden et al. 2014). $^{13}$C MFA can also characterize the metabolic phenotypes in co-culture systems or microbial communities (Gebreselassie and Antoniewicz 2015; Ghosh et al. 2014), further pushing the boundaries and capacities of flux analysis techniques. More importantly, FBA and $^{13}$C-MFA toolboxes/software have been introduced to allow easy access to flux analysis techniques (Becker et al. 2007; Quek et al. 2009; Weitzel et al. 2012; Young 2014).

On the other hand, we realize there are still many challenges. For example, FBA generally describes the optimal metabolic phenotypes, hence, it often overestimates cell performance. Integrating FBA with multiple modeling approaches or other omics studies may overcome its inherent limitation. In terms of $^{13}$C MFA, several factors may affect the accuracy of flux calculations, for example, metabolic channeling and cell compartmentation (van Winden et al. 2001). Metabolic channeling refers to successive enzymatic reactions without releasing the intermediates to cytosol. If metabolic channeling occurs, one should consider direct transitions of labeled carbons from the initial metabolite to the ultimate one. Cell compartmentation is a general property of eukaryotic cells. To simulate the labeling profiles of intracellular metabolites,
knowledge on their concentrations at local compartments is useful but difficult to precisely determine. Furthermore, modeling either metabolic channeling or cell compartmentation may increase the uncertainties of flux prediction, as more parameters are introduced in the model.

1.4. This work

In this dissertation, we made several attempts to build both $^{13}$C MFA and FBA models for analyzing the metabolic phenotypes of microbes under both steady-state and isotopically dynamic conditions. Generally, constructing a $^{13}$C-MFA model from scratch may take weeks or months, which may be the major reason preventing this technique from being routinely used. To allow a quick $^{13}$C-MFA, we developed a user-friendly and programming-free package for $^{13}$C MFA with inclusion of different bacterial templates. The detailed work will be presented in the next chapter. Applying our techniques allows us to evaluate bacterial responses to different growth conditions or genetic manipulations. In Chapter 3, we examined the central metabolic response to fatty acid overexpression in *E. coli*. Carbon and energy metabolisms in both control and engineered strains were investigated. In Chapter 4, we evaluated cyanobacterial metabolic phenotypes in different growth conditions, i.e., photomixotrophic, photoheterotrophic and photoautotrophic conditions. Particularly, we applied INMFA to analyze the photoautotrophic phenotype of *Synechococcus elongatus* UTEX 2973, a fast-growing cyanobacterium. In Chapter 5 and 6, we sought to apply multiple modeling approaches to simulate cyanobacterial performance in photobioreactors in order to better evaluate the process parameters most affecting their biomass or biofuel production. To this end, we applied Simulink® to simulate the flue gas treatment by cyanobacteria in Chapter 5, and made efforts to integrate a genome-scale FBA model with kinetics, cell movements (simplified hydrodynamic information), and a heterogeneous light distribution function in Chapter 6. In Chapter 7, we applied $^{13}$C tracing technique and evaluated the photoheterotrophic growth of
Chloroflexus aurantiacus, a filamentous anoxygenic phototroph. The last chapter summarizes the dissertation work and shows some recommended future work.

1.5. References


### Table 1.1: Mathematical framework of $^{13}$C MFA and FBA.

<table>
<thead>
<tr>
<th>Mathematical algorithm</th>
<th>$^{13}$C MFA</th>
<th>FBA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\min \sum \left( \frac{MID_{\text{exp}} - MID_{\text{sim}}}{\sigma} \right)^2$</td>
<td>$\max \mu$ (or $c^T v$)</td>
</tr>
<tr>
<td></td>
<td>subject to $S \cdot v = 0$ $lb \leq v \leq ub$</td>
<td>subject to $S \cdot v = 0$ $lb \leq v \leq ub$</td>
</tr>
<tr>
<td></td>
<td>$A \cdot EMU_{\text{unknown}} = B \cdot EMU_{\text{known}}$</td>
<td></td>
</tr>
</tbody>
</table>

In the table, $\mu$ represents the bacterial growth rate; $S$ is the stoichiometry matrix; $v$ is the flux vector; $c$ is a vector defining the objective function; $lb$ and $ub$ represent lower and upper boundaries, respectively; $MID$ represents mass isotopomer distribution; $EMU$ is the elementary metabolite unit; and matrices $A$ and $B$ represent linear function of fluxes, which are used for EMU balances.
Chapter 2: WUFlux - an open-source platform for $^{13}$C metabolic flux analysis of microbes
2.1. Introduction

To facilitate the development of genome scale models, many software has been developed (Lakshmanan et al. 2012). Previously, our research group built a web-based platform named MicrobesFlux (which is available at http://www.microbesflux.org/) (Feng et al. 2012). This platform can automatically draft a metabolic model from the annotated microbial genome in the KEGG database (http://www.genome.jp/kegg/). Based on users’ feedback, we have re-built our system on a commercial server to improve its functionality and robustness. The new MicrobesFlux updates both AMPL optimization software and metabolic network information from the latest version of the KEGG database. This platform now includes 3192 species compared to 1304 species in the previous version. The MicrobesFlux platform only perform FBA analysis, using an objective function (e.g., optimization of biomass yield) to estimate the flux values.

To complement the current platform, we sought to build an open-source MATLAB-based package for performing $^{13}$C-MFA. It is usually a challenging task for inexperienced people to perform $^{13}$C MFA. On the experimental side, the cell cultures have to grow in a minimal (or defined) medium and under (pseudo-) steady state conditions. It is also essential to select proper $^{13}$C tracers and obtain high-quality isotopomer data so as to obtain precise flux calculations. Meanwhile, construction of $^{13}$C-MFA model and flux calculations require knowledge on both metabolic network and computer programming. Therefore, less than 1000 $^{13}$C-MFA papers have been published in the past two decades (Crown and Antoniewicz 2013), many of which are review or method papers. Most $^{13}$C-MFA research studies focus on model species, such as *Bacillus subtilis* and *Escherichia coli*, and only a few $^{13}$C-MFA studies have been performed on non-model microbial species. It is therefore important for microbiologists to have user-friendly and programming free $^{13}$C-MFA tools in order to quickly analyze metabolisms in diverse microbial
systems. To reduce the modeling challenges, $^{13}$C-MFA software with different flux calculation algorithms has been developed, including FiatFlux (Zamboni et al. 2005), INCA (Young 2014), METRAN (Yoo et al. 2008), and 13CFLUX2 (Weitzel et al. 2012). Our laboratory has extensively applied $^{13}$C-MFA to study diverse non-model species for co-utilizing diverse substrates, such as sugars and organic acids. Based on our experiences, we have built an open-source $^{13}$C-MFA platform (WUFlux) to facilitate metabolism analysis in diverse microbes.

2.2. WUFlux Implementation

We chose MATLAB as the programming environment, because it is broadly used by engineers and scientists in both industry and academia. We began with designing a graphical user interface (GUI) by using GUIDE in MATLAB, and subsequently we created functions directly linked to tables, buttons, pop menus and figures on the user interface.

In general, WUFlux contains four major components (Figure 2.1). The first component is ‘Metabolic Reactions’, which is designed to permit modifications (e.g., knocking out reactions, changing boundary conditions, and adding linear constraints) to any model template. Any change in the current model will be automatically saved. Instead of asking users to design the model from scratch, we have included two templates: one template for chemoheterotrophic bacteria and another for photomixotrophic or photoheterotrophic bacteria (with an active Calvin cycle) (He et al. 2014; You et al. 2014). In either template, one or multiple substrates (i.e., glucose, xylose, acetate, pyruvate/lactate, and glycerol) can be chosen.

In ‘Experiment Data’ section, we provide two approaches to import the experimental mass isotopomer distribution (MID) data. Either corrected MID data or raw mass spectrometry (MS) data can be loaded into the software. WUFlux is capable of correcting raw MS data of TBDMS-
derivatized proteinogenic amino acids by employing a previously developed algorithm (Wahl et al. 2004), which promises an accurate data correction. Further, users also need to provide the labeling information of $^{13}$C-labelled substrates applied in the experiments. Since multiple tracers or isotopologues (e.g., 50% [1-$^{13}$C] glucose and 50% [U-$^{13}$C$_6$] glucose) have been frequently used to improve MFA resolutions, the software has been designed to handle such experimental conditions.

The third section ‘Settings’ allows users to customize the optimization parameters (e.g. number of initial guesses and maximum iteration number). Thereafter, the flux calculation is ready to start. To profile the fluxome, we used the element metabolite unit (EMU) algorithm (Antoniewicz et al. 2007) to simulate the MIDs of proteinogenic amino acids. This method largely reduces the number of variables compared to the traditional isotopomer mapping matrices approach (Wiechert et al. 2001). The built-in MATLAB function ‘fmincon’ is employed for non-linear optimization, i.e., minimization of differences experimentally and computationally determined MIDs weighted by measured variances, using ‘interior-point’ as the default algorithm. To avoid local solutions, users need to run different initial guesses of fluxes, so that ‘fmincon’ can find the global optimal solution with the smallest SSR (Sum of Squares Regression) (Figure 2.1).

Monte Carlo method is used in the model for determining the confidence intervals of central metabolic fluxes. Briefly speaking, MID data would be randomly perturbed with normally distributed noises (within average range of measurement errors), and flux profile was then recalculated for multiple times (customizable in WUFlux). The 95% confidence intervals, for example, are consequently determined by the upper and lower 2.5% data via the bootstrap method. Additionally, $\chi^2$ test was applied to determine the goodness of fit, which users can use as the reference to determine whether the final fitting is statistically acceptable. Finally, all the flux values
and confidence intervals are presented in the ‘Results’ panel, which can be exported to an excel file. In particular, a MATLAB figure comparing computationally simulated and experimentally determined MIDs is shown for examining the fitting quality.

2.3. Results and Discussions

Figure 2.1 shows the general procedures of building $^{13}$C-MFA with WUFlux: 1) Choose a suitable template, and modify the metabolic network and constraints, 2) Import the experimental results, 3) Customize the optimization parameters, and 4) Estimate the flux distribution and determine the confidence intervals. The user manual is provided in the Appendix Chapter 7.

As a case study, we applied our software to reproduce the MID data and flux profile of both control and engineered E. coli strains in our previous paper (He et al. 2014), which were calculated by METRAN. The mutant strain was constructed for producing free fatty acids. As shown in Figure 2.2, WUFlux can convert raw MS data into effective MID data, which is in perfect accordance with MID correction results from a well-developed mass correction software (Wahl et al. 2004). We further used WUFlux to characterize the fluxomes of E. coli strains with corrected MID data. The resulting fluxome was then compared with those generated from METRAN (the original data published in reference (He et al. 2014)) and INCA (Figure 2.3 and Table 2.1). In general, the estimated flux values as well as the major changes between the control and engineered strains agree well with published data and optimization results from other software. However, three software platforms did show certain differences in flux estimations. Both WUFlux and INCA exhibited slightly higher fluxes through the TCA cycle and anaplerotic pathways than the results from METRAN. These differences may be caused by the following reasons. First, different optimization algorithms/solvers are employed among three software. In specific, WUFlux relies
on MATLAB built-in function ‘fmincon’ in WUFlux, while INCA employs MATLAB’s ‘lsqnonlin’ function. Second, WUFlux had slightly different choices of MID data (e.g., WUFlux didn’t include the MID data of proline because this amino acid often shows high noise-to-signal ratios). Nonetheless, the flux determination by either software agrees with results and conclusions in our published study.

$^{13}$C-MFA is an important tool to reveal cell energy state, which is important for cell biosynthesis and well-being. In cellular processes, energy molecule ATP is not only used for biosynthesis, but also consumed for diverse non-growth associated activities (such as cell repair and stress responses). $^{13}$C-MFA can calculate total ATP generation from catabolism and ATP usage for biosynthesis. Thus the excess ATP can be assumed as maintenance cost. Here, we demonstrated how to use WUFlux to study energy metabolism by using isotopomer data from the reference (He et al. 2014). In Figure 2.4 A, we plotted the carbon distributions into biomass synthesis, fatty acid production, CO$_2$ loss, and acetate. The results proved that the engineered strain can successfully direct more carbon flow towards fatty acid production, while the control strain used most of the carbons for biomass synthesis. Additionally, we can use flux data to analyze cell energy expenditure. For example, ATP loss per gram of engineered strain was estimated to be two-fold larger than that of control strain (Figure 2.4 B-C), suggesting that overproduction of fatty acid led to a higher energy burden on the host strain. $^{13}$C-MFA can quantify cell energy fluxes and is particularly useful to understand the ATP and cofactor balances in engineered microbial hosts.

Lastly, users may add an ‘energy balance’ equation in WUFlux (i.e., the ATP net production is equal to consumption for biosynthesis). Under such assumption, the P/O ratios may impact flux calculation results. Figure 2.4 D-F illustrated the influence of P/O ratios on flux estimation of the engineered *E. coli* strain. The results show that flux estimation is insensitive to P/O ratios under
‘energy unbalanced’ conditions, when the flux towards ATP maintenance is unconstrained (Figure 2.4 D and E). However, flux values of many pathways and SSR can be significantly affected by P/O ratio under ‘energy balanced’ conditions, when ATP maintenance value is assumed to be zero (Figure 2.4 D and F).

2.4. Conclusions

We have designed and constructed an open-source MATLAB-based $^{13}$C-MFA platform, WUFlux, which is programming-free and readily to modify. By testing WUFlux against another two software, METRAN and INCA, we showed that WUFlux can accurately correct raw MS data and reproduce the flux estimation of previously published flux analysis studies. Because the MATLAB codes of all function files in WUFlux are open to researchers, users can extend or enhance its capabilities. We will also continue to improve its performance in the future. We hope that both WUFlux and MicrobeFlux platforms can provide broad-scope fluxomics functions for characterization of microbial species and engineered microbial mutants.

2.5. List of abbreviations

FBA, flux balance analysis; MFA, metabolic flux analysis; MID, mass isotopomer distribution; MS, mass spectrometry; TBDMS, N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide; and TCA, tricarboxylic acid.

2.6. Availability and requirements

- Project name: WUFlux
- Project homepage: http://13cmfa.org
- Operating systems: Platform independent
• Programming language: MATLAB

• Other requirements: MATLAB 2012b or higher with optimization toolbox, symbolic math toolbox, and statistic toolbox.

• License: WUFlux is freely available.

• Any restrictions to use by non-academics: none

2.7. Acknowledgements

We would like to thank Dr. Le You for her contribution on mass spectrometer data correction in WUFlux manual. The project was funded by NSF DBI Grant #1356669.

2.8. References


Table 2.1: Comparison of flux estimations from WUFlux, METRAN and INCA.

<table>
<thead>
<tr>
<th>E. coli strains</th>
<th>Control Strain</th>
<th>Engineering Strain</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Software</td>
<td>WUFLUX</td>
</tr>
<tr>
<td>G6P == F6P</td>
<td>WUFLUX</td>
<td>70</td>
</tr>
<tr>
<td>F6P + ATP == FBP</td>
<td></td>
<td>82</td>
</tr>
<tr>
<td>GAP == 3PG + ATP + NADH</td>
<td>WUFLUX</td>
<td>169</td>
</tr>
<tr>
<td>3PG == PEP</td>
<td></td>
<td>155</td>
</tr>
<tr>
<td>PEP == PYR + ATP</td>
<td>WUFLUX</td>
<td>122</td>
</tr>
<tr>
<td>PYR == AceCoA + CO2 + NADH</td>
<td>WUFLUX</td>
<td>108</td>
</tr>
<tr>
<td>AceCoA + OAA == CIT</td>
<td>METRAN</td>
<td>36</td>
</tr>
<tr>
<td>CIT == ICIT</td>
<td>METRAN</td>
<td>36</td>
</tr>
<tr>
<td>ICIT == AKG + CO2 + NADPH</td>
<td>INCA</td>
<td>34</td>
</tr>
<tr>
<td>AKG == SucCoA + CO2 + NADH</td>
<td>INCA</td>
<td>25</td>
</tr>
<tr>
<td>SucCoA == SUC + ATP</td>
<td>INCA</td>
<td>21</td>
</tr>
<tr>
<td>SUC == FUM + FADH2</td>
<td>METRAN</td>
<td>28</td>
</tr>
<tr>
<td>FUM == MAL</td>
<td>METRAN</td>
<td>31</td>
</tr>
<tr>
<td>MAL == OAA + NADH</td>
<td>WUFLUX</td>
<td>27</td>
</tr>
<tr>
<td>G6P == PG6 + NADPH</td>
<td>METRAN</td>
<td>29</td>
</tr>
<tr>
<td>PG6 == CO2 + Ru5P + NADPH</td>
<td>WUFLUX</td>
<td>28</td>
</tr>
<tr>
<td>Ru5P == X5P</td>
<td>WUFLUX</td>
<td>13</td>
</tr>
<tr>
<td>Ru5P == R5P</td>
<td>WUFLUX</td>
<td>15</td>
</tr>
<tr>
<td>X5P + R5P == GAP + S7P</td>
<td>INCA</td>
<td>8</td>
</tr>
<tr>
<td>GAP + S7P == E4P + F6P</td>
<td>INCA</td>
<td>8</td>
</tr>
<tr>
<td>X5P + E4P == GAP + F6P</td>
<td>INCA</td>
<td>5</td>
</tr>
<tr>
<td>PG6 == PYR + GAP</td>
<td>INCA</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 2.1: Graphical representation of applying WUFlux for $^{13}$C-MFA.
Figure 2.2: Comparison of mass isotopomer distribution data in the control (A) and engineered (B) *E. coli* strains determined by WUFux, METRAN, and INCA.
Figure 2.3: Comparison of relative flux distributions in the control (A) and engineered (B) *E. coli* strains determined by WUFlux, METRAN, and INCA.
Figure 2.4: Carbon and energy distributions in both control and engineered *E. coli* strains.

A) carbon fates in control strain; B) carbon fates in engineered strain; C) ATP usage for biomass, fatty acid, and maintenance loss; D) the influence of P/O ratios on SSR; E) the influence of P/O
ratios on SSR (‘Energy Unbalanced’); F) the influence of P/O ratios on SSR (‘Energy balanced’).

‘Energy Balanced’ represents the condition when ATP maintenance loss is assumed as zero, and ‘Energy Unbalanced’ represents the condition when ATP maintenance is unconstrained.
Chapter 3: Central metabolic responses to the overproduction of fatty acids in

*Escherichia coli* based on $^{13}$C-metabolic flux analysis
3.1. Introduction

Fatty acids are the precursors to produce transportation fuels and industrial chemicals including surfactants, solvents and lubricants. Fatty acids are conventionally derived from plant oils and animal fats, which causes competition with food supply and environmental concerns. Alternative strategies have recently attracted interest in the production of fatty acids from abundant and inexpensive renewable resources through microbial fermentations (Ranganathan et al., 2012; Steen et al., 2010; Stephanopoulos, 2007).

*E. coli* can serve as an excellent host for fatty acids production due to its fast growth, simple nutrient requirements, well understood metabolic behavior and available genetic tools. However, fatty acid metabolism in *E. coli* is tightly regulated (Steen et al., 2010). Fatty acids synthesized in the wild-type *E. coli* are mainly used to form lipids for cell membrane constitution (Fujita et al., 2007; Magnuson et al., 1993), and only a small amount of free fatty acids are detectable under normal conditions (Steen et al., 2010). Figure 3.1 shows the pathways of fatty acid metabolism in *E. coli*. The synthesis of saturated fatty acid starts with the conversion of acetyl-CoA into malonyl-CoA catalyzed by ATP-dependent acetyl-CoA carboxylase (AccABCD) and the transesterification of malonyl-CoA into an acyl carrier protein (ACP) catalyzed by malonyl-CoA ACP transacylase (FabD), and then followed by cyclic chain elongation. The synthesis of unsaturated fatty acids starts with 3-hydroxydecanoyl-ACP, which is dehydrated by FabA (introducing a double bond to the fatty acid chain) before undergoing unsaturated fatty acid elongation (Feng and Cronan, 2009).

Acyl-ACP thioesterase (TesA) catalyzes the hydrolysis of fatty acyl-ACPs, “pulling” the carbon flux to fatty acid production (Steen et al., 2010; Zhang et al., 2011). Fatty acids can be degraded via the β-oxidation pathway (Fujita et al., 2007). The β-oxidation involves the activation of fatty
acids to fatty acyl-CoAs catalyzed by FadD, followed by cleavage steps in the β-oxidation cycle to yield acetyl-CoA ultimately. The key step in the β-oxidation cycle is the oxidation of acyl-CoA to 2-enoyl-CoA, catalyzed by acyl-CoA dehydrogenase (FadE) (Campbell and Cronan Jr., 2002). “Blocking” the β-oxidation pathway by knocking out fadD or fadE can improve fatty acid accumulation (Lennen et al., 2011; Lu et al., 2008; Steen et al., 2010). Moreover, the fatty acid metabolism is regulated by the global transcription factor fadR. FadR binds to the promoters of several genes in the fatty acid metabolic pathways and controls their expression (Campbell and Cronan Jr., 2002; Magnuson et al., 1993). Specifically, FadR acts as the activator of fabA and fabB genes to “push” fatty acid production, and functions as the repressor of the fad regulon to “block” fatty acid degradation. The “pull-push-block” based metabolic engineering strategies (Figure 3.1) have been reported to produce fatty acids at the level of less than 0.25 g∙g glucose⁻¹ (< 70% of the maximum theoretical yield) in E. coli (Lennen et al., 2010; Li et al., 2012; Liu et al., 2010; Lu et al., 2008; Steen et al., 2010; Zhang et al., 2011; Zhang et al., 2012b; Xu et al., 2013).

Although researchers have created various fatty acid producing strains, most studies target engineering genes in the end fatty acid metabolic pathways. Little attention has been given to the central metabolism genetic interventions for improving fatty acid overproduction until recently. Li et al. (2012) engineered an E. coli strain with deleted acetate production pathway for fatty acid production but no improvement of fatty acid yield was achieved. Fatty acid production places high demands on precursor (acetyl-CoA), reducing power (NADPH) and energy (ATP) molecules (8Acetyl-CoA + 14NADPH + 7ATP → Palmitate (C16:0 fatty acid)). Since these molecules are mainly derived in the central metabolism, it is important to investigate the behavior of the central metabolism as fatty acid overproduction in E. coli are likely to be challenged in balancing these factors (Lennen and Pfleger, 2012). To reveal the metabolic bottlenecks, we used the powerful
$^{13}$C-metabolic flux analysis ($^{13}$C-MFA) tool (Antoniewicz et al., 2007a; Antoniewicz et al., 2007b; Christensen et al., 2002; Stephanopoulos, 1999; Suthers et al., 2007; Young et al., 2008) to characterize the central metabolism of *E. coli* in response to fatty acid production. Specifically, we engineered a fatty acid producing *E. coli* strain through overexpressing *tesA* and *fadR* genes in a *fadE* knockout strain of *E. coli* DH1. The genetic strategy was modified from that in the recent report (Zhang et al., 2012b), which introduced *tesA* and *fadR* genes in separate plasmids into the host *E. coli* strain and produced fatty acids at a high yield. In this study, we integrated the *tesA* and *fadR* genes into a single plasmid so that only one type of antibiotic was required in the cultures and thus the antibiotic toxicity was minimized. We performed tracer ($[1,2-^{13}$C] glucose) experiments and $^{13}$C-MFA for both the engineered strain and the control strain that contained an empty plasmid in the same host *E. coli* strain. We also evaluated the cofactor balance and energy status of both strains. Transcription levels of genes at the key nodes in central metabolic pathways were measured using real time quantitative RT-PCR (qRT-PCR) for additional information. This study aims to provide insights into the metabolism of *E. coli* overproducing fatty acids and guide strain engineering for further improvement of fatty acid production.

3.2. Materials and Methods

3.2.1. Plasmid and strain construction

The plasmid pA58c-TR (Supplementary Figure 1) was constructed from pE8a-fadR and pKS1 using a modified Golden Gate DNA assembly method (Steen et al., 2010; Zhang et al., 2012a; Zhang et al., 2012b) as described in Supplementary Materials and Methods. The constructed plasmid pA58c-TR containing the *tesA* and *fadR* genes was then transformed into a *fadE* knockout strain of *E. coli* DH1 (endA1 recA1 gyrA96 thi-1 glnV44 relA1 hsdR17(rK mK$^+$ λ)) using calcium chloride (Sambrook and Russell, 2001), resulting in *E. coli* DH1 ΔfadE/pA58C-TR (the engineered
fatty acid producing strain). The control strains were constructed by transforming the vector pACYCDuet-1, which has the same backbone as pA58C-TR (pA15 ori, CmR), into the wild-type E. coli DH1 and the E. coli DH1 ΔfadE strains using the above method, resulting in E. coli DH1/pACYC-Duet-1 and E. coli DH1 ΔfadE/pACYC-Duet-1. We compared the growth and 13C labeling patterns of the proteinogenic amino acids of two control strains and found no differences under our cultivation conditions (data are not shown). We then used E. coli DH1 ΔfadE/pACYC-Duet-1 as the control strain in the rest of the study.

3.2.2 Bacterial cultivation with [1,2-13C] glucose

M9 MOPS minimal medium (Supplementary Materials and Methods) with [1,2-13C] glucose was used in the 13C labeling experiments. The first pre-culture inoculated from the glycerol stock was grown for 12 h in 5 mL LB medium. The cells were harvested by centrifugation at 8,000 × g and 4°C for 2 min, washed and used to inoculate the second pre-culture of 5 ml of minimal medium containing [1,2-13C] glucose. Cells from the second pre-culture were harvested, washed and used to inoculate the main culture. The main cultures were carried out in duplicate in 250 mL baffled shake flasks (Kimax, Fisher Scientific) containing 25 mL of minimal medium with [1,2-13C] glucose on a rotary shaker (Big Bill Orbital Shakers, Thermolyne, Thermo Scientific) at 225 rpm. The inoculation fraction was 1% (v/v) in each step. Isopropyl β-D-1-thiogalactopyranoside (IPTG) was supplemented at the final concentration of 0.1 mM to induce tesA gene in 13 hours’ incubation. The fadR gene is under the control of pBAD promoter (Supplementary Figure 1). Since the previous study (Zhang et al., 2012b) using the same promoter showed that the fadR expression could be induced by 7.5-fold in glucose medium, we did not supplement arabinose for pBAD promoter induction in the cultures. Biomass samples were taken at three time points during the exponential growth phase for isotopomer distribution analyses of proteinogenic amino acids. All
cultivations were carried out at 37 °C. The analyses of cell growth, extracellular metabolites and fatty acids are shown in Supplementary Materials and Methods.

3.2.3. Analysis of mass isotopomer distributions of proteinogenic amino acids

GC-MS analyses of proteinogenic amino acids were previously described (Feng et al., 2010). Briefly, the cells were washed twice with 0.9% NaCl and hydrolyzed in 6 mol·l\(^{-1}\) HCl at 100 °C for 24 h. The resulting proteinogenic amino acids were derivatized with \(N\)-Methyl-\(N\)-[tert-butylidimethyl-silyl] trifluoroacetamide (Sigma-Aldrich) in tetrahydrofuran (Sigma-Aldrich) at 70°C for 1 h, and analyzed with GC-MS (Hewlett Packard 7890A and 5975C, Agilent Technologies) equipped with a DB5-MS column (J&W Scientific). The GC-MS program was as follows: the temperature of the column was initially held at 150°C for 2 min, raised to 280°C at 3°C·min\(^{-1}\) and then to 300°C at 20°C·min\(^{-1}\) and held at 300°C for 5 min. One µl of the sample was injected and flowed into the column at 1.2 ml·min\(^{-1}\) at a 1:20 split ratio using helium as the carrier gas. The mass spectra were analyzed using Enhanced Data Analysis software (Agilent Technologies). Mass isotopomer distributions were obtained by integration and corrected for natural isotope abundances (Leighton and Antoniewicz, 2012).

3.2.4. Metabolic flux analysis

\(^{13}\)C-MFA was performed using the Metran software (Yoo et al., 2008), which is based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007a; Young et al., 2008). The E. coli network model used for flux calculations was described previously (Leighty and Antoniewicz, 2012) and is given in Supplementary Table I. In brief, the model included all major reactions of the central carbon metabolism, fatty acid production, amino acid biosynthesis, lumped biomass formation, and transhydrogenation reaction. The model accounts for the exchange of
intracellular and extracellular CO$_2$ (Leighty and Antoniewicz, 2012) and includes G-value parameters to describe fractional labeling of amino acids (Antoniewicz et al., 2007b). One G-value parameter was included for each measured amino acid in each data set. Metabolic fluxes were estimated by minimizing the variance-weighted sum of squared residuals (SSR) between the experimentally measured and model predicted extracellular rates and mass isotopomer distributions using non-linear least-squares regression (Antoniewicz et al., 2006). For combined analysis of parallel labeling experiments, the data sets were fitted simultaneously to a single flux model (Leighty and Antoniewicz, 2012; Crown and Antoniewicz, 2013). Flux estimation was repeated at least 10 times starting with random initial values for all fluxes to find a global solution. At convergence, accurate 68% and 95% confidence intervals were computed for all estimated fluxes by evaluating the sensitivity of the minimized SSR to flux variations (Antoniewicz et al., 2006). Standard deviations of fluxes were determined as follows (Antoniewicz et al., 2006):

$$\text{Flux precision (SD)} = \frac{[\text{flux upper bound 95%} \ - \ \text{flux lower bound 95%}]}{4}$$  \hspace{1cm} (1)

To determine the goodness-of-fit, $^{13}$C-MFA fitting results were subjected to a $\chi^2$-statistical test. In short, the minimized SSR value is a stochastic variable with a $\chi^2$-distribution, where the number of degrees of freedom is equal to the number of fitted measurements $n$ minus the number of estimated independent parameters $p$. The acceptable range of SSR values is between $\chi^2_{\alpha/2}(n-p)$ and $\chi^2_{1-\alpha/2}(n-p)$, where $\alpha$ is a certain chosen threshold value, for example 0.05 for 95% confidence interval (Antoniewicz et al., 2006).

### 3.2.5. qRT-PCR

RNAs were extracted from exponentially growing cells in baffled flasks in duplicate using RNeasy mini kit (Qiagen). Contaminating DNA was removed with RNase-free DNase I (Fermentas). The
purified RNAs were quantified using a NanoDrop 200C spectrophotometer (Thermo Scientific). cDNAs were synthesized using a Revert Aid First Strand cDNA Synthesis Kit (Thermo Scientific) with random hexamer primers following the manufacturer’s protocol supplied. The synthesized cDNAs were diluted 5-fold in nuclease-free water, and 2 µl was amplified using the Power SYBR Green PCR Master Mix (Applied Biosystems) and primers specific to the genes of interest (Supplementary Table II) in a 20 µl reaction system. The reaction for each gene in each sample was performed in triplicates. qRT-PCR assays were carried out on an ABI7500 fast machine with the thermal cycling conditions recommended by the manufacturer. For data analysis, expression levels of the house keeping gene dnaK were used as a control for normalization between samples. Fold changes of genes of interest were calculated as $2^{-\Delta\Delta CT}$ according to Schmittgen and Livak (2008).

3.3. Results

3.3.1. Growth characteristics of the fatty acid producing E. coli

To characterize the growth behavior of the fatty acid producing E. coli and investigate the effect of fatty acid overproduction, both the engineered (E. coli DH1 ΔfadE/pA58C-TR) and the control (E. coli DH1 ΔfadE/pACYCDuet-1) strains were grown in batch cultures on glucose under aerobic conditions. Figure 3.2 shows the growth curves of the two strains and Table I summarizes the growth parameters. Both strains converted glucose to biomass, CO$_2$, acetic acid and fatty acids without any other major by-products. The engineered strain consumed 17.9 g·l$^{-1}$ glucose, with a slightly higher glucose consumption rate compared to the control strain, and produced 3.05 g·l$^{-1}$ of fatty acids at the end of the exponential phase in 23 h, corresponding to the yield of 0.17 g·g glucose$^{-1}$ (≈ 48% of maximum theoretical yield). The control strain consumed 8.2 g·l$^{-1}$ of glucose and produced only 0.2 g·l$^{-1}$ of fatty acids (0.024 g·g glucose$^{-1}$). Free fatty acids produced by the
engineered strain were composed of chain lengths ranging from C12 to C18, whereas the control strain mainly produced C16:0 free fatty acids (Figure 3.3). Acetic acid yield was 10-fold lower in the engineered E. coli strain (0.011 g·g glucose\(^{-1}\)) compared to the control strain (0.11 g·g glucose\(^{-1}\)). The engineered E. coli had a lower specific growth rate (0.26 h\(^{-1}\)) and a lower biomass yield (0.21 g·g glucose\(^{-1}\)) compared to the control strain (0.52 h\(^{-1}\) and 0.43 g·g glucose\(^{-1}\)).

3.3.2. Metabolic flux distributions

Metabolic fluxes (Figure 3.4; Supplementary Table IV) were determined based on the measured mass isotopomer distributions of proteinogenic amino acids (Supplementary Results; Supplementary Table III; Supplementary Figure 2) and production rates of acetic acid and fatty acids (Table 3.1). The measured biomass formation rates (Table 3.1) were not employed as the constraints of the \(^{13}\)C-MFA model; instead, they were estimated by \(^{13}\)C-MFA and then compared to the measured value to validate the metabolic flux results. As can be seen in Figure 3.4, fluxes in the central metabolism of E. coli were redistributed in response to fatty acid overproduction. Firstly, the flux portioning at the acetyl-CoA node was strikingly different between the engineered and control strains. In the engineered strain, the flux into fatty acid production was boosted 7.8-fold, whereas the flux into acetic acid production decreased 10-fold, and the flux into the TCA cycle also decreased slightly. Secondly, the flux through the PP pathway increased 1.5-fold in the engineered strain. Thirdly, the flux through the Entner-Doudoroff (ED) pathway increased 2-fold. The co-regulation of PP and ED pathways is not a coincidence; it has been reported that the alleles encoding zwf (G6P dehydrogenase encoding gene) in the PP pathway is in the same cluster as the eda gene (encoding 6-phosphogluconate dehydrogenase) in the ED pathway (del Castillo et al., 2007; Egan et al., 1992). The up-regulation of the ED pathway in the engineered E. coli enabled a direct supply of Pyr without carbon loss as CO\(_2\) via the oxidative PP pathway (6PG → Ru5P +
CO_2). Fourthly, the flux through Ppc (PEP → OAC), the major anaplerotic flux into the TCA cycle, decreased 1.7-fold in the engineered strain. The decrease of the Ppc flux is beneficial because it enabled more carbon flow from glucose to acetyl-CoA. The reduced Ppc flux could be related to the reduced demand for oxaloacetate based biomass synthesis (Table 3.1). Fifthly, the estimated biomass biosynthesis flux was significantly lower in the engineered E. coli, nearly half of that in the control strain (Figure 3.4), which correlated well with the decreased biomass yield (2-fold; Table I). Lastly, the transhydrogenation flux (NADH → NADPH) was unregulated 1.7-fold.

### 3.3.3. Relative gene transcription levels

To investigate the transcriptional response to the fatty acid overproductions, seven selected genes that are either located at the pathway branches or related to NADPH production were analyzed by qRT-PCR (Figure 3.5). Compared to the control strain, the transcription levels of genes at acetyl-CoA node, ackA (encoding acetic acid kinase) and gltA (encoding citrate synthase), were lower, suggesting reduced metabolic activities in acetic acid production and TCA cycle. In contrast, zwf, encoding the first enzyme in the PP pathway, G6P dehydrogenase, showed higher expression levels, indicating the elevated PP pathway activity. In addition, the transcription levels of icd (encoding isocitrate dehydrogenase) and maeB (encoding malic enzyme) were higher in the engineered strain. Moreover, the transcription levels of udhA (encoding soluble transhydrogenase) and pntA (encoding membrane-bound transhydrogenase) were significantly higher in the engineered strain, increased 1.4- and 2.1-fold, respectively. Taken together, the transcription results suggested that there was an elevated activity in NADPH production pathways in the engineered E. coli, as evidenced by the up-regulation of zwf, icd, maeB, pntA, udhA genes that all encode the enzymes catalyzing relations with NADPH production. Compared to the flux results, the changes of gene transcription levels of ackA, gltA, pntA, udhA and zwf agreed with the changes
of fluxes (Figure 3.4). However, there was inconsistency between the gene transcription level and metabolic flux in the case of maeB, the transcription level of which increased about 4-fold but the corresponding metabolic flux did not change in the same manner, indicating the complex regulations through the interactions of genes, proteins and metabolites at multiple regulatory levels. The increase of maeB transcription was probably induced by stress response mechanisms caused by the imbalance of NAD(H)/NADP(H) ratio in the engineered E. coli due to fatty acid production (Wang et al., 2011). On the other hand, the low malic enzyme flux could be attributed to the low substrate availability in the engineered E. coli due to the redirection of carbon fluxes to fatty acid production and/or post-translational protein modifications. The observation that gene expression is not always correlated with the corresponding metabolic flux has previously been reported (Hua et al., 2007), suggesting that gene transcription analysis alone could not sufficiently predict the metabolic responses of the entire organism to genetic perturbations (Moxley et al., 2009).

3.4. Discussion

3.4.1. Phenotype of the fatty acid producing E. coli

The phenotype of the engineered E. coli is characterized as follows: 1) a typical growth pattern of E. coli was observed with an obvious exponential phase after IPTG induction; 2) the exponential cell growth rate ($\mu = 0.26$ h$^{-1}$) and biomass yield ($Y_{x/g} = 0.20$ g·g$^{-1}$) were significantly lower compared to the control strain ($\mu = 0.52$ h$^{-1}$, $Y_{x/g} = 0.43$ g·g$^{-1}$); 3) during the exponential growth phase, fatty acid production was dominant with little acetic acid secreted; 4) during the stationary phase, cell density remained constant and fatty acid production was insignificant, whereas, acetic acid production showed a remarkable increase.
The engineered *E. coli* strain produced 3.05 g·l\(^{-1}\) and 0.17 g·g glucose\(^{-1}\) of fatty acids corresponding to 48% of the maximum theoretical yield. This result was comparable to the recent study by Zhang et al. (2012b) that used a similar genetic strategy, considering our data were collected in 23 hours instead of after 3 days. The observed chain lengths of the fatty acids produced from this study were dominated by saturated fatty acids C14:0 and C16:0 (Figure 3.3). Compared to the strain constructed by Zhang et al. (2012b), which produced a higher amount of C16:1 fatty acid, the lower unsaturated fatty acid contents observed in our study is consistent with the lower FadR copy number (in a p15A plasmid, Supplementary Figure 1) used in our strain. Like the previous study by Steen et al. (2010), we detected a small amount of 17:0 cyclopropane fatty acid, which was presumably the result of methylation of C16:1 fatty acid induced by free fatty acid accumulation (Cao et al., 2010). In addition, we observed that cell cultures inoculated from different colonies exhibited varied fatty acid productivity (data not shown), which confirmed the self-mutagenesis when *E. coli* was accumulating free fatty acids (Lennen et al., 2011).

### 3.4.2. Central carbon metabolic fluxes

We performed \(^{13}\)C tracer experiments and \(^{13}\)C-MFA for the two strains at each time point during the exponential growth phase of each culture (Supplementary Results; Supplementary Table IV). We obtained statistically acceptable flux estimations in all cases (Supplementary Table V), and the flux results were proven to be reliable as demonstrated by the agreement of the ratios of the estimated biomass fluxes (180/343 ≈ 0.5) and measured biomass yields (143/302 ≈ 0.5) between the engineered and control strains (Supplementary Table VI). We observed that: 1) the mass isotopomer distributions of proteinogenic amino acids and metabolic fluxes were constant among the three time points during the exponential phase, indicating that the metabolism maintained a pseudo-steady state condition; 2) the fluxes were reproducible between the two biological
replicates, indicating constant cellular metabolism of the engineered *E. coli* and constant cultivation conditions maintained during the experiments; and 3) there were notable differences in the fluxes between the engineered and control strains, suggesting altered metabolism of the engineered *E. coli*. The central metabolism in the engineered strain reduced acetate secretion and biomass biosynthesis so that more acetyl-CoA was directed to fatty acid synthesis (One C16:0 fatty acid requires 8 acetyl-CoA). Moreover, the engineered strain reduced Ppc flux (PEP → OAC) so that the PEP pool was retained for acetyl-CoA production. The reduced Ppc flux in the engineered *E. coli* could be related to slower biomass synthesis since the role of Ppc in *E. coli* is to replenish oxaloacetate for biomass synthesis (Sauer et al., 1999). In addition, both the PP and ED pathway fluxes were up-regulated. The up-regulation of ED pathway flux was more efficient for fatty acid production since it not only enhanced Pyr flux by providing a short path and avoiding carbon loss as CO$_2$ via the PP pathway (6PG → Ru5P + CO$_2$), but also produced one NADPH through the shared G6P dehydrogenase catalyzed reaction. Overall, the carbon fluxes of the central metabolism in the engineered *E. coli* were redirected towards acetyl-CoA for fatty acid synthesis.

**3.4.3. Cofactor NADPH metabolism**

The cofactor balancing can be estimated based on the metabolic fluxes (Kind et al., 2013; Sauer et al., 2004). The major pathways supplying NADPH are G6P and 6PG dehydrogenases in the PP pathway, and NADP$^+$-dependent isocitrate dehydrogenase in the TCA cycle. Additionally, NAD(P) transhydrogenase, encoded by *pntAB* and *udhA* genes, can catalyze the reversible conversion between NADH and NADPH to balance the cofactors (Hua et al., 2003; Sauer et al., 2004). The relative contributions of these pathways to NADPH production in the engineered and control strains are shown in Table 3.2. As can be seen, the PP pathway contributed to the supply of NADPH in a large fraction in both *E. coli* strains, which was further up-regulated 1.5-fold in the
engineered strain to increase NADPH production. This result is consistent with previous studies on other *E. coli* strains (Emmerling et al., 2002; Hua et al., 2003; Sauer et al., 2004) as well as *C. glutamicum* for lysine production that requires NADPH as the cofactor (Kind et al., 2013). The fluxes from the PP pathway and TCA cycle in the central metabolism added up to a total NADPH supply of 100 mol·mol glucose^{-1}·h^{-1}, accounting for 50% of the NADPH consumption in the engineered strain (Table 3.2). Thus, there was a 50% gap to be filled to satisfy NADPH demand without perturbing the carbon balances. The transhydrogenation reaction served an excellent role for this purpose. Our flux analysis indicated a significant transhydrogenation flux in the engineered strain (Figure 3.4), accounting for 76% of NADPH consumption for fatty acids and biomass synthesis. Compared to the control strain, the transhydrogenation flux increased 1.7-fold in the engineered strain (Figure 3.4), which was consistent with the up-regulated transcription levels of *pntA* and *udhA* genes (Figure 3.5). Thus, transhydrogenation reaction appeared to play an important role in supplying NADPH for fatty acid production in the engineered *E. coli*. As revealed by the ^13^C-MFA and cofactor balancing analysis, *E. coli* has the ability of replenishing NADPH that is consumed in biomass synthesis and fatty acid production through enhancing the transhydrogenation reaction (*NADP^+ + NADH → NAD^+ + NADPH*) and PP pathway fluxes. This finding is consistent with the previous research that engineering NADPH supply in *E. coli* is unlikely to enhance fatty acid productivity, obtained in a cell free system (Liu et al., 2010).

**3.4.4. Energy ATP metabolism**

Based on the fluxes, ATP formation and consumption for cell growth and non-growth associated cellular maintenance of the two strains were estimated (Table 3.2). The ATP formation fluxes through the glycolysis, TCA cycle and acetic acid secretion pathways added up to 157 and 153 mol·mol glucose^{-1}·h^{-1} in the engineered and control strains, respectively. Another route of ATP
production was via oxidation phosphorylation through the respiratory chain. Assuming the maximum P/O ratio (the moles of ATP formed per oxygen atom: NADH $\rightarrow$ 3ATP, FADH$_2$ $\rightarrow$ 2ATP, Mitsumori et al., 1988), the ATP formation fluxes via oxidation phosphorylation were 664 mol·mol glucose$^{-1}$·h$^{-1}$ in the engineered and 782 mol·mol glucose$^{-1}$·h$^{-1}$ in the control strain. In total, there were 821 and 935 mol·mol glucose$^{-1}$·h$^{-1}$ of ATP formed in the engineered and the control strains, respectively, which were more than enough for cell growth and fatty acid production demands (266 in the engineered and 350 in control were demanded, Table II). The large portions of the ATP produced, 555 mol·mol glucose$^{-1}$·h$^{-1}$ in the engineered *E. coli* and 585 mol·mol glucose$^{-1}$·h$^{-1}$ in the control strain, were potentially consumed for non-growth associated cellular maintenance. When expressed on the basis of dry cell weight (Table I), the cellular maintenance ATP was 154 mmol·g DCW$^{-1}$ in the engineered *E. coli* and 76 mmol·g DCW$^{-1}$ in the control strain. Clearly, the engineered strain required much higher cellular maintenance energy compared to the control strain (Youngquist et al., 2012).

Cellular maintenance covers every cellular reaction involving the consumption of ATP that does not contribute to the net synthesis of biomass and product. Such reactions include the building-up and maintenance of ionic gradients across the membrane and regeneration of degraded macromolecules (Stephanopoulos et al., 1998). Osmotic stress resulting from the culture medium and product accumulation, and disrupted bacterial growth are the major stress factors causing increased cellular maintenance energy (Varela et al. 2004). Previous study found that the overproduction of fatty acids in *E. coli* caused the change in membrane properties due to fatty acid intercalation in the inner or outer membranes, fatty acid accumulation in the periplasm, and change of the composition of membrane lipids (Lennet et al., 2011). These changes incurred secondary effects on the cell including induced membrane stress, compromised membrane integrity, and
reduced cell viability (Lennet et al., 2011). In addition, intracellular fatty acid accumulation could counteract ion diffusion into the cytoplasm. These factors could be likely responsible for the higher cellular maintenance energy observed in the engineered *E. coli* to form and sustain ionic gradients across the cell membrane.

Analysis of energy metabolism (Table II) also revealed that oxidative phosphorylation of NADH played an important role in supplying ATP to support cell growth and maintenance of the engineered *E. coli*. Therefore, it is important to maintain high respiration efficiency (i.e., the P/O ratio) during the cultivation in the future research. To this end, the culture condition such as the glucose concentration should be optimized and the dissolved oxygen concentration in the culture should be precisely monitored and controlled during the cultivation (Noguchi et al. 2004).

3.4.5. Significance of this study and future metabolic engineering strategies

The “pull-push-block” strategies, targeted on engineering genes in fatty acid metabolic pathways, have achieved *E. coli* fatty acid production with less than 70% of the maximum theoretical yield. Further enhancement has proven unpredictable due to the challenges in balancing the precursor acetyl-CoA, cofactor NADPH and energy ATP in the engineered *E. coli* that are mainly derived in the central metabolism (Lennen and Pfleger, 2012). To decipher the metabolic bottlenecks, previously “omics” techniques were employed to compare the profiles of genes and proteins between the fatty acid producing strain and the control strain (Lennen et al., 2011; Zhang et al., 2012b). Unfortunately, only a limited number of genes and enzymes involved in the central metabolism were detected and no information on the reaction rates was provided in these investigations. The overall *E. coli* central metabolism and how it is regulated for fatty acid production remained unclear. In this study, we used $^{13}$C-MFA to quantify the *E. coli* metabolism
and its regulations in response to fatty acid overproduction. $^{13}$C-MFA revealed the flexibility of the central metabolism at the nodes of PP pathway, PEP carboxylation, acetic acid secretion and TCA cycle to accommodate fatty acid overproduction. $^{13}$C-MFA also revealed that the reversible transhydrogenation reaction in E. coli could be significantly regulated to balance the cofactors (NADH and NADPH) to meet with the NADPH requirement for fatty acid production. In contrast, $^{13}$C-MFA found that the ED pathway flux was inherently low in despite of 2-fold upregulation in the fatty acid producing E. coli (Figure 3.4), which appeared to be a rigid node. The upregulation of the ED pathway in the engineered strain is beneficial to fatty acid production since it enhances the supply of pyruvate, the substrate of acetyl-CoA, by avoiding carbon loss as CO$_2$ through the PP pathway while providing one NADPH though G6P dehydrogenase. It is generally believed that rigid nodes are promising targets for metabolic engineering while flexible nodes are poor targets (Stephanopoulos and Vallino, 1991). Therefore, the ED pathway could be a promising target that could be overexpressed in combination with the “pull-push-block” strategy to improve fatty acid production. As demonstrated in this study, the acetate formation pathway was a flexible node and knocking out acetate formation pathway did not improve fatty acid production (Li et al., 2012).

$^{13}$C-MFA based energy balance analysis revealed that a large fraction of ATP supply relied on phosphorylation of NADH via respiration. However, engineered microbial strains often have lower P/O ratio, which limits the high yield of final products (Sauer and Bailey, 1999). To enhance the respiration efficiency, one strategy would be overexpression of NADH dehydrogenase. The aerobic respiratory chain of E. coli functions with either of the two different membrane-bound NADH dehydrogenases, NDH-1 (encoded by nuoA-N) and NDH-2 (encoded by the ndh), coupled with the bd-type or bo-type ubiquinol oxidases (Calhoun et al., 1993). Therefore, overexpressing NDH-1 (or NDH-2) and bo-type oxidase (bd-type is less efficient) could be an effective approach
to increase the P/O ratio. Meanwhile, the dissolved \( \text{O}_2 \) concentration in the culture should be controlled to maintain respiration efficiency since these enzymes are regulated in \( \text{O}_2 \)-dependent manner (Calboun et al., 1993; Noguchi et al., 2004). This study also revealed that the fatty acid producing \( E. \text{coli} \) required high maintenance energy, which is likely due to fatty acid accumulation that interfered with the ion transport and cell membrane function. Overexpressing fatty acid exporter genes and repressing \( \text{fabR} \) (the gene controlling the unsaturated fatty acid biosynthesis according to the ratio of unsaturated to saturated-ACPs in the membrane lipid and restores the lipid composition of the membrane) could reduce the undesired effect of fatty acid accumulation on cell membrane function (Lennen et al. 2011).

3.5. Conclusions

This work is an evident example that \(^{13}\text{C}\)-MFA is a powerful strategy for quantitative investigation of engineered microbes to provide metabolic insights. \(^{13}\text{C}\)-MFA illustrates the high flexibility of the metabolic network of \( E. \text{coli} \) to compensate for external perturbation, and meanwhile identifies the rigid nodes for future metabolic engineering. Moreover, \(^{13}\text{C}\)-MFA based cofactor and energy balance analyses allow the evaluation of the energy status of the cells, which can lead to additional findings. This work provides important information about the metabolic bottlenecks of \( E. \text{coli} \) overproducing fatty acids and strategies which could be exploited in combination of the “pull-push-block” approach for future strain development and bioprocess optimization for the enhancement of fatty acid production.

3.6. Acknowledgements

I would like to thank all the co-authors who have contributed directly in this Chapter: Professor Lifeng Peng from Victoria University of Wellington in New Zealand, Professor Fuzhong Zhang
from Washington University in St. Louis (WUSTL), Professor Maciek Antoniewicz and his student Nikodimos Gebreselassie from University of Delaware, and Dr. Yi Xiao. This project was funded by National Science Foundation and I-CARES program at Washington University in St. Louis. Also, we would like to thank Professor Yehuda Ben-Shahar and his PhD student Xingguo Zheng at WUSTL for their technical assistance for the qRT-PCR experiments.

3.7. References


Li, M, Zhang, X, Agrawal, A, San, KY. 2012. Effect of acetate formation pathway and long chain fatty acid CoA-ligase on the free fatty acid production in E. coli expressing acy-ACP thioesterase from Ricinus communis. Metab Eng 14:380-387.


Table 3.1: Exponential growth parameters of the engineered and control *E. coli* strains grown on glucose in minimal medium. $Y_{X/G}$, biomass yield on glucose; $Y_{F/G}$, fatty acid yield on glucose; $Y_{A/G}$, acetic acid yield on glucose. These values were determined from the mean values of parallel cultivations corresponding to Figure 3.2.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Specific growth rate (h$^{-1}$)</th>
<th>Specific rate (mmol·g DCW$^{-1}·$h$^{-1}$) of</th>
<th>Yield (g·g$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Glucose uptake</td>
<td>Fatty acid formation</td>
</tr>
<tr>
<td>Engineered</td>
<td>0.26</td>
<td>7.2</td>
<td>0.89</td>
</tr>
<tr>
<td>Control</td>
<td>0.52</td>
<td>6.6</td>
<td>0.10</td>
</tr>
</tbody>
</table>
Table 3.2: Estimated production and consumption of NADPH, NADH, ATP and FADH$_2$ for the \textit{E. coli} control and engineered strains.

<table>
<thead>
<tr>
<th>Pathways</th>
<th>NADPH</th>
<th>NADH</th>
<th>ATP</th>
<th>FADH$_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Engineered</td>
<td>Control</td>
<td>Engineered</td>
</tr>
<tr>
<td>Glycolysis</td>
<td>0</td>
<td>0</td>
<td>167</td>
<td>171</td>
</tr>
<tr>
<td>PP pathway</td>
<td>54</td>
<td>80</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>TCA cycle</td>
<td>27</td>
<td>20</td>
<td>141</td>
<td>177</td>
</tr>
<tr>
<td>Amino acid synthesis</td>
<td>-103</td>
<td>-54</td>
<td>17</td>
<td>9</td>
</tr>
<tr>
<td>One-carbon metabolism</td>
<td>0.8</td>
<td>0.4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Fatty acid formation</td>
<td>-23</td>
<td>-177</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Acetic acid formation</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Biomass formation</td>
<td>-46</td>
<td>-24</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Transhydrogenation</td>
<td>90</td>
<td>153</td>
<td>-90</td>
<td>-153</td>
</tr>
<tr>
<td>Oxidative phosphorylation</td>
<td>0</td>
<td>0</td>
<td>-248</td>
<td>-210</td>
</tr>
</tbody>
</table>

*Note: All the flux values are normalized to the glucose uptake rate of 100 mol·h$^{-1}$ for each strain. The excessive NADH and FADH$_2$ were assumed to be converted to ATP via oxidation phosphorylation at the maximum P/O ratio (NADH $\rightarrow$ 3ATP, FADH$_2$ $\rightarrow$ 2ATP) (Mitsumori et al. 1988).
Figure 3.1: Fatty acid metabolism of *E. coli* and the “pull-push-block” strategy for fatty acid production. Overexpression of tesA and fadR pulls and pushes the carbon fluxes to fatty acid production, while knockout of fadE blocks fatty acid degradation. Red-boxed genes are positively regulated by FadR. Green-boxed genes are negatively regulated by the fadR gene product, FadR. Bolded arrows represent reactions enhanced by overexpressing the indicated genes. × represents inactivation of the corresponding pathway by knocking out the indicated gene. ⊿ represents repression or activation exerted by FadR. Full names of genes: accABCD, acetyl-CoA carboxylase; ackA, acetate kinase A; fabA, beta-hydroxydecanoyl thioester dehydrase; fabB, 3-oxoacyl-ACP
synthase I; fabD, malonyl-CoA ACP transacylase; fabF, 3-oxoacyl-ACP synthase II; fabG, 3-oxoacyl-ACP reductase; fabH, 3-oxoacyl-ACP synthase III; fabI, enoyl-ACP reductase; fabZ, (3R)-hydroxymyristoyl acyl carrier protein dehydratase; fadA, 3-ketoacyl-CoA thiolase; fadB, 3-hydroxyacyl-CoA dehydrogenase; fadD, acyl-CoA synthetase; fadE, acyl coenzyme A dehydrogenase; plsB, glycerol-3-phosphate O-acyltransferase; pta, phosphate acetyltransferase; tesA, acyl-ACP thioesterase.
Figure 3.2: Growth curves of the engineered fatty acid producing strain E. coli DH1 \( \Delta \text{fadE/pA58C-TR} \) (A) and the control strain DH1 \( \Delta \text{fadE/pACYCDuet-1} \) (B) in batch cultures grown on glucose under aerobic conditions. Arrows indicate the time points of IPTG addition or sampling for isotopic labeling patterns of proteinogenic amino acids, mRNA and external metabolites analyses. Square, Glucose; Diamond, Biomass; Triangle, Acetic acid; Circle, Fatty acid. Results are from parallel cultivations.
Figure 3.3: Free fatty acid profiles produced by the engineered fatty acid producing *E. coli* strain (A) and the control strain (B). Symbols a, b and c represent the sampling time points for fatty acid analysis as indicated in Figure 3.2.
Figure 3.4: Metabolic flux distributions in the central metabolic pathways of the engineered fatty acid producing *E. coli* strain and the control strain during exponential growth on [1,2-\(^{13}\)C] glucose. Fluxes shown are normalized to glucose uptake rate of 100 for each strain (estimated flux ± SD). Abbreviations: G6P, glucose-6-phosphate; F6P, fructose-6-phosphate; FBP, fructose 1,6-bisphosphate; DHAP, Dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; 3PG, 3-phosphoglycerate; PEP, phosphoenolpyruvate; Pyr, pyruvate; AcCoA, acetyl coenzyme A; Cit, citrate; ICit, isocitrate; AKG, α-oxoglutarate; Suc, succinate; Fum, fumarate; Mal, malate; OAC,
oxaloacetate; Glyox, glyoxylate; 6PG, 6-phosphogluconate; Ru5P, ribulose-5-phosphate; X5P, xylulose-5-phosphate; R5P, ribose-5-phosphate; E4P, erythrose-4-phosphate; S7P, sedoheptulose-7-phosphate; TK-C2, the first two carbon unit of F6P or S7P; TA-C3, the first three carbon unit of F6P or S7P; KDPG, 2-keto-3-deoxy-6-phosphogluconate. The reactions are given in Supplementary Table I.
Figure 3.5: Fold changes of transcription levels of the selected genes in the engineered fatty acid producing *E. coli* strain compared to the control strain. *ackA*, acetate kinase A; *gltA*, citrate synthase; *icd*, isocitrate dehydrogenase; *maeB*, malic enzyme; *pntA*, membrane-bound transhydrogenase; *udhA*, soluble transhydrogenase; *zwf*, G6P dehydrogenase.
Supplementary Materials

All the following supplementary files are available online (DOI: 10.1002/bit.25124).

1. Supplementary Materials and Methods

2. Supplementary Figure 1. Plasmid map of pA58c-TR used in the engineered fatty acid-producing strain. The *E. coli fadR* is under the control of pBAD. A truncated version of the *E. coli* thioesterase gene *tesA* ('tesA; leader sequence deleted) is under the control of P<sub>lacUV5</sub>.

3. Supplementary Figure 2. Fractional abundances of mass isotopomers of Ala (m/z 232), Phe (m/z 336) and Glu (m/z 432) at sampling time a, b and c in the exponential growth phase of the control and engineered *E. coli* strains in each of the [1, 2-<sup>13</sup>C] glucose tracer experiments; P1 and P2 on the horizontal axis denote the two parallel tracer experiments for the control strain; E1 and E2 denote the two parallel tracer experiments for the engineered strain; M0 to Mn represent the mass isotopomers, where n is the number of <sup>13</sup>C atoms.

4. Supplementary Table I. Metabolic network model of *E. coli* used for <sup>13</sup>C metabolic flux analysis.

5. Supplementary Table II. Sequences of the primers used for qRT-PCR experiments in this study.

6. Supplementary Table III. Mass isotopomer distributions of biomass amino acids for the control and engineered *E. coli* strains grown in parallel batch cultures on [1,2-<sup>13</sup>C] glucose.

7. Supplementary Table IV(a). Results of <sup>13</sup>C-MFA for the control and engineered *E. coli* strains grown in parallel batch cultures on [1,2-<sup>13</sup>C] glucose.

8. Supplementary Table IV(b). Results of metabolic fluxes and standard deviations (SD) using combined analysis of <sup>13</sup>C-MFA for the three samples in each of the [1,2-<sup>13</sup>C] glucose tracer experiments.

58
9. Supplementary Table V. Goodness-of-fit analysis for $^{13}$C-MFA of parallel [1,2-$^{13}$C]glucose labeling experiments with *E. coli* control and engineered strains.

10. Supplementary Table VI. Comparison of estimated biomass fluxes and measured biomass yields of the engineered and control *E. coli* strains.
Chapter 4: Characterizing cyanobacterial metabolic flux phenotypes under photomixotrophic, photoheterotrophic, and photoautotrophic conditions
4.1. Introduction

Cyanobacteria are phototrophic prokaryotes that have been studied since 19th century (Stanier and Bazine 1977). They have versatile metabolisms allowing them to adjust to different environmental conditions. Thanks to the advances in metabolic engineering and synthetic biology, they have also been broadly employed as microbial hosts to produce biofuel and other economically valuable products (Lan and Liao 2012; Lindberg et al. 2010; Varman et al. 2013a; Varman et al. 2013b). Compared to plants, cyanobacteria enjoy faster growth rates and hence greater productivities (Dismukes et al. 2008). Unlike other heterotrophic microbes, such as E. coli and Saccharomyces cerevisiae, biomass or biofuel production by cyanobacteria does not suffer from carbon loss. This merit paves the way for attaining high production yields in cyanobacteria. Another unique property of cyanobacteria is their natural abilities to produce alkane/alkene (Coates et al. 2014; Schirmer et al. 2010) and hydrogen (Dutta et al. 2005), further making cyanobacteria promising hosts for biofuel production.

In this chapter, we sought to evaluate their metabolic flux phenotypes under various growth conditions. The resulting flux profiles can allow us to better understand how carbon flows are arranged in the central metabolism of cyanobacteria in order to survive environmental fluctuations, and what strategies we may undertake to engineer them for biofuel production. We chose two wild-type cyanobacterial species, i.e., Synechocystis sp. PCC 6803 and Synechococcus elongatus UTEX 2973. The former is one of the most well-studied and widely used cyanobacterial species. Due to maturely developed genetic engineering techniques for this strain, mutants capable of producing a variety of bio-products have been successfully constructed. The unique trait of Synechococcus elongatus UTEX 2973 is its high growth rate comparable to yeast (Yu et al. 2015). As a matter of
fact, many cyanobacterial species exhibit low growth rates. Such an advantage of *Synechococcus* 2973 may eventually lead to both high yield and production rate of biofuels.

Three growth conditions were evaluated in this chapter: photomixotrophic (*Synechocystis* sp. PCC 6803), photoheterotrophic (*Synechocystis* sp. PCC 6803), and photoautotrophic (*Synechococcus elongatus*) conditions. In particular, since $^{13}\text{CO}_2/\text{NaHCO}_3$ is used as the only carbon source for cyanobacteria under photoautotrophic conditions, traditional $^{13}\text{C}$-MFA is not suitable anymore, because all the carbons will be labeled as $^{13}\text{C}$ under steady-state conditions. To address this issue, INCA (short for isotopomer network compartmental analysis), a platform for isotopically non-stationary flux analysis (Young 2014), was employed to quantify the flux distributions in photoautotrophic conditions (Figure 4.1). To perform isotopically non-stationary $^{13}\text{C}$ MFA, time series data of intracellular metabolites labeling profiles and knowledge on their concentrations are required through experiments (Jazmin and Young 2013; Shastri and Morgan 2007). Ordinary differential equations were used to simulate time series data, and the best fit was determined by minimizing the differences between experimental and computational values.

The resulting flux profiles under different cyanobacterial growth conditions permit detailed examination of how cyanobacteria respond to environmental changes and offer some guidelines for engineering cyanobacteria for biofuel production.

**4.2. Materials and Methods**

**4.2.1. Cultivation conditions and $^{13}\text{C}$ labeling experiment**

We grew *Synechocystis* 6803 in modified BG-11 medium (containing no organic compounds) under photomixotrophic and DCMU ((3-(3,4-dichlorophenyl)-1,1-dimethylurea))-induced photoheterotrophic conditions. Cultures were grown at 30 °C with a light intensity of 50 $\mu\text{E/m}^2/\text{s}$. 

62
In the $^{13}$C-labeling experiment, 5 g/L [1-$^{13}$C] glucose and 2 g/L NaH$^{13}$CO$_3$ were provided in minimal medium under photomixotrophic conditions, and 2 g/L [1,2-$^{13}$C] glucose and 10 $\mu$M DCMU were supplied in medium under photoheterotrophic conditions. In both cases, cultures growing at the exponential phase were sampled at two different time-points. The detailed description can be found in our published articles (You et al. 2014; You et al. 2015).

As for *Synechococcus elongatus* UTEX 2973, they were grown in well-controlled photobioreactors at 38°C surrounded by strong light sources (500 $\mu$E/m$^2$/s). In the labeling experiment, NaH$^{13}$CO$_3$ was pulsed into photoautotrophic cultures growing in a non-labeled medium. The final NaH$^{13}$CO$_3$ concentration was 4 g/L. Biomass samples were taken at different time points (e.g., 5s, 20s, 60s, 2min, etc.), and immediately they were bathed and stirred in liquid nitrogen. Chloroform and methanol (2:1, v/v) were used to extract intracellular free metabolites before they were analyzed by LC-MS.

### 4.2.2. Flux calculations

Steady-state $^{13}$C MFA was applied to calculate the flux distributions of *Synechocystis* 6803 under photomixotrophic and photoheterotrophic conditions, while INCA was used to estimate the photoautotrophic fluxome of *Synechococcus elongatus* UTEX 2973. Cyanobacterial metabolic network includes the glycolysis, Calvin cycle, TCA cycle, glyoxylate shunt, photorespiration pathways, and biomass synthesis (Table 4.1). Both the steady-state or isotopically nonstationary MFA employed element metabolite unit method (Antoniewicz et al. 2007) to simulate the labeling patterns of proteinogenic amino acids or intracellular free metabolites. One hundred initial guesses were tested to find the optimal estimation.
4.3. Results and Discussions

4.3.1. Metabolic flux phenotypes of wild-type *Synechocystis* 6803 under photomixotrophic and photoheterotrophic conditions

Under photomixotrophic growth conditions with sufficient light energy, *Synechocystis* 6803 were provided with both inorganic and organic carbon sources, i.e., glucose and bicarbonate. The resulting flux distribution of *Synechocystis* 6803 growing at the exponential growth phase shows strong activities in the Calvin cycle, while fluxes through the OPP pathway and TCA cycle are weak (Figure 4.2). The anaplerotic pathway was estimated to be active as well, probably due to the reason that malic enzyme reaction serves as a key route for providing pyruvate (Bricker et al. 2004), since high ATP production from photosystems may adversely affect pyruvate kinase activity in glycolysis.

Under photoheterotrophic conditions, *Synechocystis* 6803 has to rely more on organic carbons to obtain NADPH, because NADPH production via photosystems is blocked by DCMU. To address the shortage of NADPH under the stressed conditions, *Synechocystis* 6803 appears to drive appreciable carbon flows from glucose source and central metabolism to the OPP pathway (Figure 4.3), which results in a two-fold higher flux in the OPP pathway than that of glucose uptake rate. Additionally, high activity of the malic enzyme further increases NAD(P)H production in cells. On the other hand, active CO\textsubscript{2} fixations through RuBisCO and phosphoenolpyruvate carboxylase (PEPC) pathways were observed, which is in accordance with our RNAseq experiments (You et al. 2015). We also found that *Synechocystis* 6803 did not show a complete and functional Calvin cycle. Also, RuBisCO and PEPC reactions do not consume any ATP or NADPH directly. Therefore, it is likely that under DCMU-induced photoheterotrophic conditions, CO\textsubscript{2} fixation in cyanobacteria is active at certain levels.
4.3.2. Metabolic flux phenotypes of wild-type *Synechococcus elongatus* UTEX 2973 under photoautotrophic conditions

Figure 4.4 shows the flux distribution of *Synechococcus elongatus* UTEX 2973 under photoautotrophic conditions, which exhibits strong activities in only Calvin cycle. In chemoheterotrophic bacteria, both building blocks and a large amount of energy molecules are obtained through the central metabolism, however, cyanobacteria depends more on photosystems for energy production. After absorbing CO$_2$ through the Calvin cycle, cyanobacteria only need to maintain enough carbon flows in other pathways for producing building blocks. Therefore, fluxes in pathways other than CO$_2$ fixation are generally much smaller.

Interestingly, recent studies on photoautotrophic metabolism of *Synechocystis* 6803 (Xiong et al. 2015; Young et al. 2011) indicate a much stronger OPP pathway than what we observed in *Synechococcus* 2973. It is possible that both the OPP pathway and Calvin cycle can be functional simultaneously in cyanobacterial cultures, which we will discuss in Chapter 6. However, driving the carbon flows in two opposite ways at the same time can lead to suboptimal metabolisms. To obtain a high optimality in the carbon metabolism, *Synechococcus* 2973 employs a better strategy by avoiding as much CO$_2$ loss as possible, i.e., limiting the fluxes through malic enzyme reaction, the OPP pathway, and the TCA cycle.

4.3.3. Plasticity of cyanobacterial pentose phosphate pathway

One interesting finding after our reviewing cyanobacterial flux phenotypes under different growth conditions is the plasticity of cyanobacterial pentose phosphate pathway. The Calvin cycle is an energy-consuming pathway driving the carbon flow from environment towards the central metabolism in cells. The OPP pathway, on the other hand, functions as a source for both NADPH
and biosynthetic precursors at the expense of carbon loss. With sufficient light and carbon sources, the OPP pathway shows very small or even negligible fluxes, while the Calvin cycle boasts high activities (Figure 4.2 and 4.4). Under stressed conditions (specifically, the DCMU-induced photoheterotrophic condition), the OPP pathway is highly upregulated for supporting NADPH production, while the CO₂ fixation is not as strong (Figure 4.3). Therefore, balancing the activities of Calvin cycle and OPP pathway in cyanobacteria embodies a trade-off between the carbon and energy metabolisms, which allows survival under various environmental situations.

4.3.4. Low activities of TCA cycle in cyanobacteria

As Figures 4.2-4 show, the flux distributions under three growth conditions are different, but one common trait is a low activity in the TCA cycle, which can be confirmed by slow changes in the labeling patterns of succinate and malate (Figure 4.5) and also other flux analysis studies (Xiong et al. 2015; Young et al. 2011). In chemoheterotrophic bacteria, a relatively higher flux via the TCA cycle must be sustained in cells to allow sufficient energy production. However, in cyanobacteria, because of abundant ATP and NADPH production through photosystems, it seems unnecessary to keep a strong flux in the TCA cycle. From the perspective of carbon metabolism, diverting a high flux towards TCA cycle will compromise the biomass growth. Hence, maintenance of low activities in cyanobacterial TCA cycle might be a good strategy to achieve high efficiencies in both energy and carbon metabolisms, which is attributed to cyanobacterial metabolic adaption to their physiological requirements (Steinhauser et al. 2012).

However, this low activity in the TCA cycle may unfavorably affect the overall cyanobacterial productivity of chemicals originating from the TCA cycle. Although one previous study shows that driving carbon flows from TCA cycle towards biofuel production has proved successful in
increasing the TCA activity by 2-3 folds (Xiong et al. 2015), the resulting flux is still two orders of magnitude lower than CO₂ fixation via the Calvin cycle. One possible approach of increasing fluxes through the TCA cycle would be to enhance the enzymatic activities and intermediate pool sizes of the TCA cycle.

4.4. Conclusions

In this chapter, we examined the flux profiles of cyanobacteria (Synechocystis sp. PCC 6803 and Synechococcus elongatus UTEX 2973) under different growth conditions. The results reveal that cyanobacteria can organize the flux distributions to adapt to different growth conditions. Although cyanobacteria appear different flux topologies under different growth conditions, TCA cycle is stubborn to change and always shows low activities, while the Calvin cycle and the OPP pathway appear flexible towards variant environments.

4.5. Abbreviations

2PG, 2-phosphoglycerate; 3PG, 3-phosphoglycerate; AceCoA, acetyl-CoA; AKG, α-ketoglutarate; CIT, citrate; DHAP, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; F6P, fructose 6-phosphate; FBP, fructose-1,6-bisphosphate; FUM, fumarate; G6P, glucose 6-phosphate; GAP, glyceraldehyde 3-phosphate; GLC, glycolate; GLX, glyoxylate; MAL, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose 5-phosphate; Ru5P, ribulose-5-phosphate; RuBP, ribulose-1,5-diphosphate; S7P, sedoheptulose-7-phosphate; SBP, sedoheptulose 1,7-bisphosphate; SUC, succinate; SucCoA, succinyl-CoA; and X5P, xylulose-5-phosphate.

4.6. Acknowledgements

I am grateful for Pakrasi Lab for providing both cyanobacterial strains, Synechocystis sp. PCC 6803 and Synechococcus elongatus UTEX 2973, and their kind help on our research. Fluxomics
study of *Synechocystis* 6803 was funded by an NSF Career Grant (MCB0954016) and a NASA Exobiology Grant. I would also like to thank the coworkers, Dr. Le You, Dr. Bertram Berla, and Dr. Himadri Pakrasi, for their invaluable contributions to this project. Profiling the autotrophic flux phenotype of *Synechococcus elongatus* UTEX 2973 was supported by a US DOE grant (DESC0012722). I appreciate the contributions of Mary Abernathy, Whitney Hollinshead, and Ni Wan on this project. Dr. Fangfang Ma and Dr. Doug Allen have also shared valuable ideas with us on flux calculations.

### 4.7. References


Table 4.1: Metabolic network used to measure the flux distribution of *Synechococcus* 2973 under photoautotrophic conditions.

<table>
<thead>
<tr>
<th>#</th>
<th>Reactions (Carbon transitions)</th>
<th>#</th>
<th>Reactions (Carbon transitions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>G6P (abcdef) &lt;-&gt; F6P (abcdef)</td>
<td>21</td>
<td>SBP (abcdefg) -&gt; S7P (abcdefg)</td>
</tr>
<tr>
<td>2</td>
<td>G6P (abcdef) -&gt; 6PG (abcdef)</td>
<td>22</td>
<td>PYR (abc) -&gt; ACA (bc) + CO2 (a)</td>
</tr>
<tr>
<td>3</td>
<td>6PG (abcdef) -&gt; RU5P (bcdef) + CO2 (a)</td>
<td>23</td>
<td>OAA (abcd) + ACA (ef) -&gt; CIT (dcbfea)</td>
</tr>
<tr>
<td>4</td>
<td>F6P (abcdef) &lt;-&gt; FBP (abcdef)</td>
<td>24</td>
<td>CIT (abcdef) &lt;-&gt; ICI (abcdef)</td>
</tr>
<tr>
<td>5</td>
<td>FBP (abcdef) &lt;-&gt; DHAP (cba) + GAP (def)</td>
<td>25</td>
<td>ICI (abcdef) &lt;-&gt; AKG (abcde) + CO2 (f)</td>
</tr>
<tr>
<td>6</td>
<td>DHAP (abc) &lt;-&gt; GAP (abc)</td>
<td>26</td>
<td>AKG (abcde) -&gt; SUC (bcde) + CO2 (a)</td>
</tr>
<tr>
<td>7</td>
<td>GAP (abc) &lt;-&gt; 3PGA (abc)</td>
<td>27</td>
<td>SUC (abcd) &lt;-&gt; FUM (abcd)</td>
</tr>
<tr>
<td>8</td>
<td>3PGA (abc) &lt;-&gt; 2PGA (abc)</td>
<td>28</td>
<td>FUM (abcd) &lt;-&gt; MAL (abcd)</td>
</tr>
<tr>
<td>9</td>
<td>2PGA (abc) &lt;-&gt; PEP (abc)</td>
<td>29</td>
<td>MAL (abcd) &lt;-&gt; OAA (abcd)</td>
</tr>
<tr>
<td>10</td>
<td>PEP (abc) &lt;-&gt; PYR (abc)</td>
<td>30</td>
<td>MAL (abcd) -&gt; PYR (abc) + CO2 (d)</td>
</tr>
<tr>
<td>11</td>
<td>RU5P (abcde) &lt;-&gt; X5P (abcde)</td>
<td>31</td>
<td>PEP (abc) + CO2 (d) -&gt; OAA (abcd)</td>
</tr>
<tr>
<td>12</td>
<td>RU5P (abcde) -&gt; R5P (abcde)</td>
<td>32</td>
<td>ICI (abcdef) -&gt; GOX (ab) + SUC (edcf)</td>
</tr>
<tr>
<td>13</td>
<td>RU5P (abcde) -&gt; RUBP (abcde)</td>
<td>33</td>
<td>GOX (ab) + ACA (cd) -&gt; MAL (abcd)</td>
</tr>
<tr>
<td>14</td>
<td>RUBP (abcde) + CO2 (f) -&gt; 3PGA (cde) + 3PGA (fba)</td>
<td>34</td>
<td>RUBP (abcdef) -&gt; 3PGA (cde) + GLC (ba)</td>
</tr>
<tr>
<td>15</td>
<td>X5P (abcde) &lt;-&gt; GAP (cde) + EC2 (ab)</td>
<td>35</td>
<td>GLC (ab) -&gt; GOX (ab)</td>
</tr>
<tr>
<td>16</td>
<td>F6P (abcdef) &lt;-&gt; E4P (cdef) + EC2 (ab)</td>
<td>36</td>
<td>GOX (ab) + GOX (cd) -&gt; GA (abd) + CO2 (c)</td>
</tr>
<tr>
<td>17</td>
<td>S7P (abcdefg) &lt;-&gt; R5P (cdefg) + EC2 (ab)</td>
<td>37</td>
<td>GA (abc) &lt;-&gt; 2PGA (abc)</td>
</tr>
<tr>
<td>18</td>
<td>F6P (abcdef) &lt;-&gt; GAP (def) + EC3 (abc)</td>
<td>38</td>
<td>CO2.source (a) -&gt; CO2 (a)</td>
</tr>
<tr>
<td>19</td>
<td>S7P (abcdefg) &lt;-&gt; E4P (defg) + EC3 (abc)</td>
<td>39</td>
<td>0.715R5P + 3.624ACA + 1.191G6P + 0.501E4P + 1.2053PGA + 1.002PEP + 1.197PYR + 2.039OAA (abcd) + 1.233AKG + 0.133GAP -&gt; 1.017CO2 + Biomass + 0.683FUM (abcd)</td>
</tr>
<tr>
<td>20</td>
<td>DHAP (abc) + E4P (defg) &lt;-&gt; SBP (cbadefg)</td>
<td>40</td>
<td>CO2.air (a) + CO2 (b) -&gt; CO2 (a) + CO2.out (b)</td>
</tr>
</tbody>
</table>

Note: 1) The letters in brackets represent the carbon transition rules from reactants to products.
Figure 4.1: A Screenshot of INCA.
Figure 4.2: Flux distribution of *Synechocystis* 6803 under photomixotrophic conditions. The figure was reproduced from reference of (You et al. 2014). Relative flux values, which are all normalized to flux of reaction ‘Ru5P→RuBP’, and standard deviations are presented beside the pathways. The estimated glucose consumption rate was 0.24 mmol·g⁻¹·h⁻¹.
Figure 4.3: Flux distribution of *Synechocystis* 6803 under DCMU-induced phototrophic conditions. The figure was reproduced from our previously published data (You et al. 2015). Relative flux values, which are all normalized to glucose uptake rate (0.41 mmol g\(^{-1}\) h\(^{-1}\)), and standard deviations are presented beside the pathways.
Figure 4.4: Flux distribution of *Synechococcus elongatus* UTEX 2973 under photoautotrophic conditions. Relative flux values, which are all normalized to the net CO$_2$ uptake rate, and standard deviations are presented beside the pathways.
Figure 4.5: Mass isotopomer distribution fittings to the dynamic labeling data of intracellular free metabolites. Error bars represent standard measurement errors. M0, M1, and M2 stand for the fractions of non-labeled, singly labeled and doubly labeled metabolites, respectively.
Chapter 5: Experimental analysis and model-based optimization of microalgae growth in photobioreactors using flue gas
5.1. Introduction

CO₂ sequestration from flue gas receives intensive studies due to global warming issues. Typical flue gas discharged from fossil fuel power plants contains 4-14% CO₂, and up to 0.022% NOₓ and SOₓ (Kumar et al. 2011). Besides physical and chemical methods for sequestration of CO₂ from flue gas (Granite and O'Brien 2005), microalgae culture holds great potential for converting flue gas to biomass. Microalgae can capture solar energy more efficiently than plants (Wang et al. 2008), and are also able to synthesize biofuels (such as biodiesel and bio-hydrogen) (Chisti 2007; Li et al. 2008; Schenk et al. 2008). To facilitate the utilization of flue gas, microalgae species, such as Chlorella sp. and Tetraselmis sp., have been tested for their tolerance to CO₂ as well as SOₓ and NOₓ (Kumar et al. 2010). In addition, several microalgae, including Dunaliella tertiolecta (Nagase et al. 2001; Nagase et al. 1997) and Nannochloris sp. (Yoshihara et al. 1996), have the capacity to use NO as their nitrogen source and thus remove it from the flue gas. Different reactor configurations (Kumar et al. 2011; Schenk et al. 2008) and cultivation strategies (Lee et al. 2000; Zeiler et al. 1995) have been studied to improve biomass growth with flue gas, including pH control via addition of alkaline solution, high inoculum size, proper flue gas rate, and optimal nutrition level. Furthermore, kinetic models were applied to analyze influential factors on algal growth using flue gas, including hydraulic residence time, reactor geometry, light intensity, culture temperature, flow rate, and partial pressures of CO₂, NOₓ and CO (Doucha et al. 2005; Vunjak-Novakovic et al. 2005; Westerhoff et al. 2010). For example, an experimental study in combination with mass balance calculations indicated that Chlorella growth attained ~50% decarbonization of flue gas in an optimal photobioreactor (4.4 kg CO₂ produced 1 kg dry weight biomass) (Doucha et al. 2005).
This study focused on model-based optimization for algal growth using flue gases. In general, atmospheric CO\textsubscript{2} (0.039\% by volume fraction) is insufficient to support optimal algal growth (Kumar et al. 2010), while the high concentration of CO\textsubscript{2} in industrial exhaust gases has adverse effects on algal physiology. Therefore, the control of flue gas flow into photo-bioreactors is of practical importance for effective algal CO\textsubscript{2} utilization. To design the optimal strategies for operation of flue gas inflow, we built Monod-based models using MATLAB/Simulink\textsuperscript{®}. The model simulation linked the control of flue gas flow to microalgae growth kinetics, and thus provided guidelines in the bioprocess for maximizing algal growth with flue gases.

5.2. Materials and Methods

5.2.1. Algal cultivation medium and biomass measurement

*Chlorella* sp., *Synechocystis* PCC 6803, and *Tetraselmis suecica* were obtained from the Pakrasi Lab at Washington University in St. Louis. The culture medium to grow *Chlorella* contained 0.55 g·L\textsuperscript{-1} urea, 0.1185 g·L\textsuperscript{-1} KH\textsubscript{2}PO\textsubscript{4}, 0.102 g·L\textsuperscript{-1} MgSO\textsubscript{4}·7H\textsubscript{2}O, 0.015 g·L\textsuperscript{-1} FeSO\textsubscript{4}·7H\textsubscript{2}O and 22.5 µl microelements (containing 18.5 g·L\textsuperscript{-1} H\textsubscript{3}BO\textsubscript{3}, 21.0 g·L\textsuperscript{-1} CuSO\textsubscript{4}·5H\textsubscript{2}O, 73.2 g·L\textsuperscript{-1} MnCl\textsubscript{2}·4H\textsubscript{2}O, 13.7 g·L\textsuperscript{-1} CoSO\textsubscript{4}·7H\textsubscript{2}O, 59.5 g·L\textsuperscript{-1} ZnSO\textsubscript{4}·5H\textsubscript{2}O, 3.8 g·L\textsuperscript{-1} (NH\textsubscript{4})\textsubscript{6}Mo\textsubscript{7}O\textsubscript{24}·4H\textsubscript{2}O, 0.31 g·L\textsuperscript{-1} NH\textsubscript{4}VO\textsubscript{3}). The pH was adjusted to 7-8 with sodium hydroxide solution. BG-11 medium(Rippka et al. 1979) and ASP2 medium(Provasoli et al. 1957) were utilized for growing *Synechocystis* and *Tetraselmis*, respectively. Microalgae stock was maintained in shaking flasks (~100 mL culture, 2.5 Hz) at 30\degree C. Algal growth was monitored by spectrophotometer (Thermal Scientific\textsuperscript{®}, Texas USA) at 730 nm.
5.2.2. Flue gas treatment using photo-bioreactors

Fresh microalgal cultures were inoculated into 200 mL medium in glass bottles. The initial OD$_{730}$ was set to ~0.3. Microalgal growth was supported by fluorescent lamps with a photon flux of 40–50 $\mu$mol·m$^{-2}$·s$^{-1}$ at room temperature (~25 °C). Flue gas was generated by natural gas combustion. It was pumped through a funnel to a condenser tube and then a washing bottle (0.5 L) containing water or water/limestone slurry (buffer solution), before being introduced into the microalgal cultures at an airflow rate of ~250 cm$^3$·min$^{-1}$ per bottle. The volume fraction of CO$_2$ in the flue gas was 5-6 % as measured by a CO$_2$ gas analyzer (LI-COR®, Biosciences, Nebraska USA). A computer control system was used to apply flue gas pulses to algal cultures (Fig. 1). The flue gas pulse included two modes (gas-on: using flue gas; gas-off: with just atmospheric CO$_2$). The flow rate and on-off frequency were controlled by the software coded with Visual Basic®. The actuators were two mass flow controllers (OMEGA Engineering INC, Connecticut, USA) that were connected to a data acquisition card (Measurement Computing Corporation, Massachusetts, USA). Filters (Aerocolloid LLC, Minnesota USA) were used to clean the inflow gases to the mass flow controllers (i.e., removing aerosol particles). The data acquisition card collected the real-time flow rate data that could be stored in the computer. To simulate algal culture using sun light, microalgal cultures were treated with flue gas under light for 12 hours, and then stored in dark aerobically (without flue gas treatment or shaking) for 12 hours (i.e., the light-dark cycle).

5.2.3. Kinetic model development

An un-segregated kinetic model for algal CO$_2$ utilization was developed with the following assumptions: 1) the culture was a well-mixed homogeneous system; 2) CO$_2$ concentration and light intensity were the limiting factors influencing the algal growth; 3) the complex relationship
between CO₂ partial pressure and its equilibrium species H₂CO₃, HCO₃⁻, and CO₃²⁻ was simplified with Henry’s Law (Eq. 2).

\[
\frac{dX}{dt} = \frac{S}{S + K_s + S^2/K_1} \cdot \frac{I}{I + K} \cdot \mu_{\text{max}} \cdot X
\]  

(1)

\[
\frac{dS}{dt} = K_{La} \left( \frac{P}{H} - S \right) - Y_{S/X} \frac{S}{S + K_s + S^2/K_1} \cdot \frac{I}{I + K} \cdot \mu_{\text{max}} \cdot X
\]  

(2)

\(X\) was the biomass concentration, kg·m⁻³; \(S\) was the dissolved CO₂ concentration, mol·m⁻³; \(I\) was the average light intensity, \(\mu\)mol·m⁻²·s⁻¹; \(P\) was the CO₂ partial pressure in the gas phase, Pa; \(\mu_{\text{max}}\) was the maximum specific growth rate of microalgae, h⁻¹; \(K_s\) was the Michaelis-Menten constant of CO₂, mol·m⁻³; \(K_1\) was the inhibition constant of flue gas, mol·m⁻³; \(H\) was Henry’s constant of CO₂, Pa·m⁻³·mol⁻¹; \(K_{La}\) was the mass transfer rate, h⁻¹; and \(Y_{S/X}\) was the yield coefficient, (mol CO₂)/(kg biomass). The average light intensity (I) in photo-bioreactor was calculated by the following equation (Martínez et al. 1997):

\[I = \frac{I_0}{A \cdot X} \left(1 - e^{-AX}\right)\]  

(3)

where \(I_0\) was the surface light intensity, \(\mu\)mol·m⁻²·s⁻¹; and \(A\) was a coefficient with units of m³·kg⁻¹. The parameters and initial conditions used for model simulation were given in Table 5.1 unless otherwise stated.

5.2.4. Dynamic optimization framework to profile optimal CO₂ concentrations

We applied the dynamic optimization approach to find the time-dependent inflow CO₂ concentration profile \(P_{opt}\) that could generate the maximum biomass production (Methekar et al. 2010). Because of the stiff nature of the model equations (i.e., successive sudden changes of the
inlet CO₂ concentrations during algal growth), CVP (control vector parameterization method) was used in this study (Kameswaran and Biegler 2006). Specifically, the entire time span was divided into n discrete time intervals with constant $P_{opt}(i)$ within each time interval ($i=1, 2, …, n$). Eq.1-2 was simulated to find the biomass growth in each time interval using the MATLAB function “ode23s”. MATLAB function “fmincon” was employed to search the optimal $P_{opt}(i)$ ($i=1, 2, …n$) to maximize the final biomass concentration $X_{end}(n)$. Once $P_{opt}(i)$ was determined, n was updated to 2n (each time interval divided by half) and the same optimization procedure yielded new $P_{opt}(i)$ ($i=1, 2, …2n$) and $X_{end}(2n)$. The procedure for searching the new set of $P_{opt}$ was repeated until $(X_{end}(2n) - X_{end}(n))/X_{end}(n) < 0.01\%$. The flowchart of the dynamic optimization procedure was shown in Fig. S1 in the supplementary file. MATLAB and Simulink (Mathworks, Massachusetts USA) were used for model calculations. The Simulink configuration and MATLAB programs were also provided in Fig. S2 and supplementary MATLAB files.

5.3. Results and Discussions

5.3.1 Experimental Analysis of Microalga growth on flue gas

Three different strains were cultivated with flue gas (Table 5.2). The results showed that continuous exposure (12 hours) to flue gas acidified the medium (pH ≈ 5) and highly inhibited microalgal growth. Decreasing CO₂ exposure time (< 6 hours per day) and pre-washing of the flue gas using buffer solution (limestone slurries) only slightly alleviated flue gas stresses on microalgal cells. Comparing algal growth among the three model algal species, Chlorella showed the best growth under flue gas stresses. To overcome flue gas inhibition, we investigated an on-off flue gas input mode in which the flue gas was pulsed into bioreactors at a specific on/off frequency (Fig. 2). The frequency of 1 min gas-on and 29 min gas-off was first applied to support all Chlorella cultures. Such a gas pulse mode reduced the actual exposure time of high concentration CO₂ to the
microalgae, and thus minimized the inhibitory effect of flue gas on microalgal physiologies. For example, 12-hour-per-day on-off flue gas pluses allowed *Chlorella* to generate 20~50% more biomass than shaking flask conditions using atmospheric CO$_2$ during the exponential growth phase.

5.3.2. Model simulation of algal growth under different flue gas treatment

To improve our understanding of the optimal control of flue gas inflow for microalgal growth and reduce experimental efforts, we developed an empirical model to simulate biomass growth with flue gas treatment. Fig. 3 unveiled the effects of CO$_2$ volume fraction and inhibition constant (K$_I$) on the biomass production. The simulation showed that CO$_2$ with a volume fraction ranging from 0.1~1% favored microalgal biomass production. The inhibition coefficient K$_I$ exerted a dramatic influence on algal biomass production. For example, decreasing K$_I$ from 10 mol·m$^{-3}$ to 0.5 mol·m$^{-3}$ reduced the overall biomass production by 60 % (7-day culture) when CO$_2$ volume fraction was ~10%.

The Monod-model also simulated algal growth in the on-off CO$_2$ pulse modes (Fig. 4) in which the cultures were exposed to different CO$_2$ volume fractions of 15% (gas-on) and 0.04% (gas-off, with atmospheric CO$_2$) alternately. Fig. 4 showed the simulated biomass growth, the decrease of average light intensity in the photo-bioreactor due to biomass growth, and variation of dissolved CO$_2$ in the culture medium. Comparing to microalgal growth with atmospheric CO$_2$, the model indicated that the biomass production (in a 7-day culture) could be improved by 35% with 1 min gas-on / 29 min gas-off CO$_2$ treatment when microalgal growth rate was $\mu_{\text{max}} = 0.041$ h$^{-1}$. If microalgal specific growth rate $\mu_{\text{max}}$ was raised to 0.070 h$^{-1}$, the biomass production was increased by 77% in the same CO$_2$ pulse mode. These model results suggested that CO$_2$ pulses more effectively supported biomass growth when $\mu_{\text{max}}$ was high.
To find the optimal CO\textsubscript{2} pulse operation (i.e., the width and the frequency of rectangular pulse), we examined the influence of pulse function on algal growth (Fig. 5). It was clear that a frequent on-off control of flue gas inflow generally promoted microalgal growth. When $\mu_{\text{max}} = 0.041 \text{ h}^{-1}$, the final biomass achieved a maximum of $0.481 \text{ kg} \cdot \text{m}^{-3}$ at the frequency of 10 s gas-on / 7 min gas-off, whereas biomass production dropped to $0.326 \text{ kg} \cdot \text{m}^{-3}$ at the frequency of 380s gas-on / 67 min gas-off (Fig. 5A). We also tested the effects of $\mu_{\text{max}}$, $K_{I}$ and $K_{La}$ on biomass production with different CO\textsubscript{2} pulse functions. First, if $\mu_{\text{max}}$ was raised from $0.041 \text{ h}^{-1}$ to $0.070 \text{ h}^{-1}$ (Fig. 5B), the gas-off duration should be shortened (i.e., a frequency of 10 s gas-on / 5 min gas-off for supporting optimal biomass growth). Second, when the inhibition constant $K_{I}$ dropped from $10 \text{ mol} \cdot \text{m}^{-3}$ to $1 \text{ mol} \cdot \text{m}^{-3}$, an optimal on-off control was achieved at a frequency of 10 s / 9 min (i.e., increase gas-off period, Fig. 5C). Third, reduction of mass transfer coefficient $K_{La}$ from $17 \text{ h}^{-1}$ to $6 \text{ h}^{-1}$ lowered the rate of CO\textsubscript{2} transfer from gas phase to liquid phase and abated CO\textsubscript{2} inhibition to the microalgal physiology. Accordingly, the gas-off period was reduced to 5 min to promote biomass growth (Fig. 5D). In summary, the maximal biomass production required a short period of on-time (a few seconds) and a comparatively longer off-time (5~10 minutes) depending on the severity of CO\textsubscript{2} inhibition and values of $\mu_{\text{max}}$. The off-period could be elongated when flue-gas showed strong inhibition.

5.3.3. Optimal CO\textsubscript{2} conditions for microalgal growth

The dynamic optimization of inlet CO\textsubscript{2} partial pressure was established by control vector parameterization. The results showed that the objective function (maximization of the final biomass production) converged (within 0.01% difference) after dividing the microalgal growth period into 64 time intervals. The simulated optimal CO\textsubscript{2} profiles from dynamic optimization were displayed in Fig. S3. The optimal CO\textsubscript{2} concentration was not constant during microalgal growth,
instead, it should gradually increase to support algal growth during the cultivation. Fig. 6 tested the effect of different model parameters on optimal dynamics of inflow CO$_2$ partial pressure. In general, increasing $\mu_{\text{max}}$ and decreasing $K_{\text{L},a}$ demanded high CO$_2$ concentration to compensate for fast biomass growth and inefficient CO$_2$ transport. On the other hand, decreasing $K_I$ enhanced CO$_2$ inhibition and thus low CO$_2$ concentration should be employed for biomass growth. With the optimal inflow CO$_2$, the biomass production was most influenced by $\mu_{\text{max}}$ (increasing $\mu_{\text{max}}$ from 0.041 h$^{-1}$ to 0.070 h$^{-1}$ resulted in ~80% more biomass growth), while biomass production was insensitive to parameters $K_I$ and $K_{\text{L},a}$. Moreover, the model simulation indicated that the high frequency on-off flue-gas pulses (15% CO$_2$) could support biomass growth almost as well as optimal CO$_2$ conditions (Fig. 7). CO$_2$ pulses could yield over 90% of theoretical biomass growth achieved under optimal CO$_2$ conditions.

Although the dynamic control of inflow CO$_2$ concentration served theoretically as the best way for biomass production, the on-off gas pulse mode still holds many advantages in the scaled-up bioprocess. For instance, direct flue gas treatment is much easier to operate than the dynamic increase of the inflow concentration. From the energy conservation point of view, the flue gas pulses reduce electricity consumption by avoiding continuously pumping flue gases into the photobioreactors or algal ponds. Furthermore, the common photobioreactor design often utilizes feedback control based on algal biomass and CO$_2$ concentrations to adjust inflow CO$_2$. However, such strategy is limited by the time delay of the actuators, unreliable on-line sensors to measure biomass and CO$_2$ concentrations, and sophisticated design of PID (proportional-integral-derivative) control loop. In this study, we have demonstrated that the high frequency on-off flue gas pulses could serve as a cost-effective operation for algal cultivation.
5.3.4. Experimental verification and model limitations

To experimentally verify the effectiveness of on-off control of flue gases for algal culture, we conducted the flue gas treatment with Chlorella using two on-off frequencies (10s gas on / 7 min gas off and 30 min gas on / 30 min gas off). Fig. 8 showed that higher on-off frequency yielded better algal growth than the lower one, and it was also better than the shaking flasks condition (atmospheric CO₂). Therefore, the results qualitatively verified our model, and confirmed that the on-off control of flue gases was able to alleviate flue gas inhibition and promote Chlorella growth.

The model was subject to several limitations. First, the model did not directly account for the influence of toxic compounds SOₓ and NOₓ on algal growth. Second, it over-simplified the chemical reactions and equilibriums in the culture medium including CO₂, H⁺, OH⁻, NH₃, etc. Third, the model did not include CO₂ fluid dynamics, while the actual gaseous mass transfer was not instantaneous and homogenous in the culture medium. Despite these limitations, all kinetic models always represent a compromise between complexity and practical simplicity. In this study, our model simulation still provided useful insights into optimal strategies for algal growth and avoided costly experimental efforts.

5.4. Conclusions

Exposure to continuous flue gas severely inhibited the algal growth. To overcome this problem, we tested an on-off flue-gas treatment to enhance algal growth. The model simulation showed that the frequency of ~10s on-time and 5~9min off-time was an ideal strategy for sustaining optimal algal production, close to theoretical maximum biomass growth. The effectiveness of flue gas control was also experimentally validated. Compared to continuously pumping diluted flue gas or
chemical pretreatment of flue gas, the simple on-off pulse mode can effectively reduce energy and material expenses.

5.5. Acknowledgements

We would like to thank Dr. Himadri Pakrasi for providing microalgae species, Dr. Benjamin Kumfer for helping us measure CO₂ concentration in the flue gas, and Qiaoling Liu for building up the on-off control system. This study was supported by Consortium for Clean Coal Utilization at Washington University and by an NSF Career Grant (MCB0954016).

5.6. References

Methekar R, Ramadesigan V, Braatz RD, Subramanian VR. 2010. Optimum charging profile for lithium-ion batteries to maximize energy storage and utilization. ECS Transactions 25(35):139-146.


Table 5.1: Parameter values used in the model.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Value range</th>
<th>Unit</th>
<th>Reference/Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_{\text{max}}$</td>
<td>maximum specific growth rate</td>
<td>0.041-0.070</td>
<td>h$^{-1}$</td>
<td>(Novak and Brune 1985)</td>
</tr>
<tr>
<td>$K_s$</td>
<td>Michaelis-Menten constant of CO$_2$</td>
<td>0.00021-0.00036</td>
<td>mol·m$^{-3}$</td>
<td>(Novak and Brune 1985)</td>
</tr>
<tr>
<td>$K_i$</td>
<td>inhibition constant of CO$_2$</td>
<td>10</td>
<td>mol·m$^{-3}$</td>
<td>(Silva and Pirt 1984)</td>
</tr>
<tr>
<td>$K$</td>
<td>Michaelis-Menten constant of light intensity</td>
<td>14</td>
<td>$\mu$mol·m$^{-2}$·s$^{-1}$</td>
<td>(Martínez et al. 1997)</td>
</tr>
<tr>
<td>$K_{\text{La}}$</td>
<td>mass transfer rate of CO$_2$</td>
<td>6-17</td>
<td>h$^{-1}$</td>
<td>(Powell et al. 2009)</td>
</tr>
<tr>
<td>$H$</td>
<td>henry’s constant of CO$_2$</td>
<td>3202</td>
<td>Pa·m$^3$·mol$^{-1}$</td>
<td>(Sawyer et al. 2003)</td>
</tr>
<tr>
<td>$Y_{S/X}$</td>
<td>yield coefficient</td>
<td>100</td>
<td>(mol CO$_2$)/(kg biomass)</td>
<td>(Doucha et al. 2005)</td>
</tr>
<tr>
<td>$A$</td>
<td>constant</td>
<td>14.7</td>
<td>m$^3$·kg$^{-1}$</td>
<td>(Martínez et al. 1997)</td>
</tr>
<tr>
<td>$I_0$</td>
<td>surface light intensity</td>
<td>45</td>
<td>$\mu$mol photons·m$^{-2}$·s$^{-1}$</td>
<td>measured</td>
</tr>
<tr>
<td>Atmospheric CO$_2$</td>
<td>atmospheric CO$_2$ concentration</td>
<td>0.04%</td>
<td>volume fraction</td>
<td>assumed in model</td>
</tr>
<tr>
<td>CO$_2$ in flue gas</td>
<td>CO$_2$ concentration in the flue gas</td>
<td>15%</td>
<td>volume fraction</td>
<td>assumed in model</td>
</tr>
<tr>
<td>$X(0)$</td>
<td>initial biomass concentration</td>
<td>0.1</td>
<td>kg·m$^{-3}$</td>
<td>assumed in model</td>
</tr>
<tr>
<td>$S(0)$</td>
<td>initial dissolved CO$_2$ concentration</td>
<td>0.013</td>
<td>mol·m$^{-3}$</td>
<td>assumed in model</td>
</tr>
</tbody>
</table>

Note: Model simulation used $\mu_{\text{max}}$=0.041 h$^{-1}$, $K_s$=0.00021 mol·m$^{-3}$ and $K_{\text{La}}$=17 h$^{-1}$ unless otherwise stated.

$^a$ In the reference, $K_i$=10 mM, and the test range in this study is 0.5-10 mol·m$^{-3}$; $^b$ In the reference, K=1011 lux, which is close to 14 $\mu$mol·m$^{-2}$·s$^{-1}$ (Thimijan and Heins 1983); $^c$ In the reference, H=31.6 atm·M$^{-1}$; $^d$ The experimental results showed that 4.4 kg CO$_2$ was needed for production of 1 kg (dry weight) of biomass; $^e$ The measured light intensity was 40-50 $\mu$mol·m$^{-2}$·s$^{-1}$;
Table 5.2: Maximum OD$_{730}$ increase rate observed (d$^{-1}$) within four days.

<table>
<thead>
<tr>
<th>Strains</th>
<th>12-hour continuous flue gas aeration per day</th>
<th>5–6 hours’ flue gas aeration followed by 5–6 hours’ air aeration per day</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with buffer</td>
<td>without buffer</td>
</tr>
<tr>
<td>Chlorella</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tetraselmis</td>
<td>Very poor growth under continuous flue gas treatment</td>
<td>0.040±0.003</td>
</tr>
<tr>
<td>Synechocystis</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: The increase rate was calculated by the equation $k=\ln(\text{OD}_f / \text{OD}_i) / \Delta t$, where $\text{OD}_f$ and $\text{OD}_i$ are the final and initial optic density at 730 nm, respectively, and $\Delta t$ is the timespan; After treatment, all the cultures were stored in the dark without gas treatment.
Figure 5.1: Diagram of the experiment setup. 1: pure water or limestone buffer solution (12.5kg·m$^{-3}$); 2: microalgae cultures; 3: magnetic stirrer; 4: burner; 5: funnel; 6: condenser tube; 7: filter; 8: mass flow controller (A: flue gas flow; B: air flow); 9: data acquisition card; 10: computer; 11: air pump; 12: flow rate meter; 13: exhaust gas; 14: air; 15: fluorescent lamp; 16: flue gas; 17: iron support.
Figure 5.2: Chlorella growth curves. The flue gas pulses were only in the light period and the frequency was 1 min gas-on / 29 min gas-off. □: flue gas pulses without buffer pretreatment (12 h-12 h light-dark cycle); Δ: flue gas pulses with buffer pretreatment (12 h-12 h light-dark cycle); ◊: cultivation in shaking flasks (12 h-12 h light-dark cycle, with atmospheric CO₂); ■: flue gas pulses without buffer pretreatment (5 h-19 h light-dark cycle); ▲: flue gas pulses with buffer pretreatment (5 h-19 h light-dark cycle); ♦: cultivation in shaking flasks (5 h-19 h light-dark cycle, with atmospheric CO₂).
Figure 5.3: Effects of CO₂ volume fraction and inhibition constant (KI) on the biomass production. The model assumed that microalgae grew in a 12 h-12 h light-dark cycle for 7 days. Red Line: Kᵢ=10 mol·m⁻³; Blue Line: Kᵢ=5 mol·m⁻³; Green Line: Kᵢ=1 mol·m⁻³; Yellow Line: Kᵢ=0.5 mol·m⁻³.
Figure 5.4: Simulation of microalgae growth (red lines) under CO₂ (15%) pulses at a frequency of 1 min gas-on/29 min gas-off in a 12 h-12 h light-dark cycle. CO₂ pulses were only in the light period. Microalgal growth with atmospheric CO₂ was also simulated (cyan lines). (A): biomass growth (red and cyan lines) and average light intensity (green line), $\mu_{\text{max}}=0.041$ h⁻¹. (B): CO₂ concentrations in the culture (blue line) and in the gas phase (green line), $\mu_{\text{max}}=0.041$ h⁻¹. (C): biomass growth (red and cyan lines) and average light intensity (green line), $\mu_{\text{max}}=0.070$ h⁻¹. (D): CO₂ concentration in the culture (blue line) and in the gas phase (green line), $\mu_{\text{max}}=0.070$ h⁻¹.
Figure 5.5: Effect of pulse function on biomass production. The model assumed that microalgae grew under 12 h-12 h light-dark cycle for 7 days. The tested model parameters included (A): \( \mu_{\text{max}}=0.041 \text{ h}^{-1}, K_I=10 \text{ mol} \cdot \text{m}^{-3}, K_{L_a}=17 \text{ h}^{-1} \); (B): \( \mu_{\text{max}}=0.070 \text{ h}^{-1}, K_I=10 \text{ mol} \cdot \text{m}^{-3}, K_{L_a}=17 \text{ h}^{-1} \); (C): \( \mu_{\text{max}}=0.041 \text{ h}^{-1}, K_I=1 \text{ mol} \cdot \text{m}^{-3}, K_{L_a}=17 \text{ h}^{-1} \); (D): \( \mu_{\text{max}}=0.041 \text{ h}^{-1}, K_I=10 \text{ mol} \cdot \text{m}^{-3}, K_{L_a}=6 \text{ h}^{-1} \).
Figure 5.6: The optimal CO$_2$ concentration profiles. The model assumed that the cultures were grown under continuous light illumination for 7 days. The tested model parameters included (1): $\mu_{max}=0.041$ h$^{-1}$, $K_I=1$ mol·m$^{-3}$, $K_{La}=17$ h$^{-1}$; (2): $\mu_{max}=0.041$ h$^{-1}$, $K_I=10$ mol·m$^{-3}$, $K_{La}=17$ h$^{-1}$; (3): $\mu_{max}=0.070$ h$^{-1}$, $K_I=1$ mol·m$^{-3}$, $K_{La}=17$ h$^{-1}$; (4): $\mu_{max}=0.070$ h$^{-1}$, $K_I=10$ mol·m$^{-3}$, $K_{La}=17$ h$^{-1}$; (5): $\mu_{max}=0.041$ h$^{-1}$, $K_I=1$ mol·m$^{-3}$, $K_{La}=6$ h$^{-1}$; (6): $\mu_{max}=0.041$ h$^{-1}$, $K_I=10$ mol·m$^{-3}$, $K_{La}=6$ h$^{-1}$. 
Figure 5.7: Simulation of microalgal growth under three CO2 treatment modes in continuous illumination condition. Blue line: growth with optimal in-flow CO2 concentration (i.e., theoretical maximal biomass growth); Green line: growth with flue gas pulses at a frequency of 1 min / 29 min; Red line: growth with frequent flue gas pulses (A: 10 s / 7 min; B: 10 s / 5 min; C: 10 s / 9 min; D: 10 s / 5 min; E: 10 s / 7 min; F: 10 s / 5 min). Parameters used were: (A): $\mu_{\text{max}}=0.041$ h$^{-1}$, $K_i=10$ mol∙m$^{-3}$, $K_{La}=17$ h$^{-1}$; (B): $\mu_{\text{max}}=0.070$ h$^{-1}$, $K_i=10$ mol∙m$^{-3}$, $K_{La}=17$ h$^{-1}$; (C): $\mu_{\text{max}}=0.041$ h$^{-1}$, $K_i=1$ mol∙m$^{-3}$, $K_{La}=17$ h$^{-1}$; (D): $\mu_{\text{max}}=0.041$ h$^{-1}$, $K_i=10$ mol∙m$^{-3}$, $K_{La}=6$ h$^{-1}$; (E): $\mu_{\text{max}}=0.041$ h$^{-1}$, $K_i=1$ mol∙m$^{-3}$, $K_{La}=6$ h$^{-1}$; (F): $\mu_{\text{max}}=0.070$ h$^{-1}$, $K_i=1$ mol∙m$^{-3}$, $K_{La}=17$ h$^{-1}$. 
Figure 5.8: Effect of flue gas pulse modes on *Chlorella* growth (without buffer pretreatment).

The figure showed the data for *Chlorella* growth during the first 12 hours under different conditions unless otherwise stated. The increase OD$_{730}$ per hour was calculated by (OD$_f$-OD$_i$)/Δt, where OD$_f$ and OD$_i$ were the final and initial optic density at 730 nm, respectively. Δt was the timespan. A: 10s gas-on/7 min gas-off; B: 30 min gas-on/30 min gas-off; C: 5-hour continuous flue gas treatment; D: cultivation in shaking flasks.
Supplemental materials

The objective: find $P_{opt}(t)$ ($0 \leq P(t) \leq 101.325\text{kPa (1 atm)}$) that generated maximum biomass $X_{end}$ in 7 days

Initial conditions: $X(0)=100\text{mg/L}$, $S(0)=13\ \mu\text{mol/L}$; Initial guess: $P_{\text{guess}}=2.03\text{kPa (0.02atm)}$

“fmincon” was used to search constant P and ode23s was used to solve $dX/dt=f(S,X)$; $dS/dt=f(S,X)$

The time span was divided into $N$ intervals

N→ 2N

In each time interval, “fmincon” was used to search $P_{i,\text{opt}}$ ($i=1, 2 \ldots N$) that generated the maximum biomass $X_{end}(N)$. Ode23s was used to solve $dX/dt=f(S, X)$ and $dS/dt=f(S, X)$ in each interval.

$(X_{end}(2N) - X_{end}(N)) / X_{end}(N) < 0.01\%$

Yes

No

$P_{opt}(t)$ is comprised of $P_{opt}(i)$ ($i=1, 2, 3 \ldots N$)

Figure 5-S1: Flow chart of dynamic optimization procedures.
Figure 5-S2: Simulink block diagram of Eq.1-2 for on-off gas control mode. The red box shows the simulation of the on-off gas control mode. The model can be changed to continuous flue gas treatment mode if the red box is substituted with one pulse generator. The Simulink subsystem files will be provided by authors upon the request.
Figure 5-S3: Dynamic optimization loops to determine optimal $P_{opt}(i)$, where $\mu_{\text{max}}=0.041\text{ h}^{-1}$, $K_I=10\text{ mol}\cdot\text{m}^{-3}$, $K_{La}=17\text{ h}^{-1}$. 
Chapter 6: Simulating cyanobacterial phenotypes by integrating flux balance analysis, kinetics, and a light distribution function
6.1 Introduction

In photobioreactors (PBRs), light penetration depth at high cell density can be as short as a few centimeters (Janssen et al. 2003). Thus, during large-size PBR cultivation, cyanobacteria move continuously between the “light zone” (where light is sufficient) and the “dark zone” (where light is substantially shaded). As a consequence, cyanobacterial metabolism in PBRs is spatially and temporally dependent: cells have autotrophic growth in the light zone, and they perform heterotrophic growth in the dark zone by consuming energy-storage compounds. Moreover, PBR performances are also affected by the efficiency of CO$_2$ gas-liquid transfer. To enhance mass transfer, people often use CO$_2$-enriched air in combination with high intensity mixing. Many models have been developed to understand how cyanobacterial physiological dynamics are influenced by the light intensity, CO$_2$ supply, temperature, and geometry of PBRs (Béchet et al. 2013; Cornet et al. 1992; Cornet and Dussap 2009; Takache et al. 2010; Wu and Merchuk 2001). Those kinetic and reactor studies are useful in optimizing PBR design and operations. However, bioprocess modeling is unable to provide an understanding of intracellular enzyme functions and metabolic fluxes in cyanobacteria. To improve engineered microalgae strains’ metabolisms in large-size PBRs, it is necessary to link process models to metabolic models.

On the other hand, metabolic flux analyses (MFA) can quantify in vivo enzyme reaction rates, and thus allow us to investigate the flux phenotypes resulting from complicated gene-protein-metabolite regulations. $^{13}$C-MFA measures carbon fluxes through the central metabolism via $^{13}$C labeling experiments. Alternatively, genome-scale flux balance analysis (FBA) can generate a holistic intracellular flux distribution map (Orth et al. 2010) owing to its extended coverage of genomic information (Lerman et al. 2013). Computational platforms, such as COBRA (Becker et al. 2007) and OptForce (Ranganathan et al. 2010), can predict genetic targets and guide rational
designs of engineered strains. FBA can also be integrated with constraint-based elementary flux mode analysis to identify optimal pathways for bio-productions (Teusink et al. 2009). However, an inherent limitation of traditional GSM is that it predicts only flux distributions that result in maximal yields in an optimal culture condition. They cannot forecast mutant strains’ production titers and rates in dynamic and heterogeneous bioreactors.

In this study, the major goal is to demonstrate multiple-scale modeling approaches by linking cell metabolisms to PBR environmental fluctuations. Specifically, the modeling efforts focus on *Synechocystis* 6803, a most widely used cyanobacterial biorefinery. Appealing traits of this species include amenability to genetic modifications, well-studied genomics, and native genes for biosynthesis of alkanes/alkenes and hydrogen (Bandyopadhyay et al. 2010; Berla et al. 2013; Schirmer et al. 2010)(Bandyopadhyay et al. 2010; Berla et al. 2013; Schirmer et al. 2010)(Bandyopadhyay et al. 2010; Berla et al. 2013; Schirmer et al. 2010). To predict cyanobacterial growth and metabolic flux phenotypes in PBR settings, we integrated a genome-scale cyanobacteria model, iJN678 (Nogales et al. 2012), with growth kinetics, cell movements based on reported PBR hydrodynamics, and a heterogeneous light distribution (Figure 6.1). The model assumption is that heterogeneous PBR conditions affect cyanobacteria, leading to heterogeneous cell metabolisms in different sub-populations. Such an approach can provide biological information ranging from the intracellular domain to the PBR domain, and fill the gaps between systems biology and the PBR process. The multiple-scale modeling is useful for estimating mutant strains’ potentials to achieve the production metrics required for commercialization.
6.2. Materials and Methods

6.2.1. Cell cultivation

*Synechocystis* PCC 6803 was cultivated in a modified BG-11 medium (You et al. 2014) at 30 °C and 180 rpm. We first tested the cyanobacterial growth in different culture volumes. In brief, 50 mL, 100 mL, and 150 mL of cell suspensions were cultivated in 250 mL shake flasks under continuous illumination of ~50 µE/m²/s. We also tested the cyanobacterial growth under different light conditions. Specifically, 15 mL of cultures were grown in 150 mL shake flasks under different light intensities (from ~15 to ~35 µE/m²/s). OD₇₃₀ was used to measure biomass density, and the relationship between the biomass concentration and OD₇₃₀ was 0.45×OD₇₃₀ = Biomass (g/L) (You et al. 2014). We made duplicate cultures of each condition (n=2).

6.2.2. ¹³C-Labelling experiment

¹³C-labeling experiments were performed to determine histidine labeling as evidence of OPP pathway activity under different light conditions. We grew photomixotrophic cultures in BG-11 medium supplied with 2.5 g/L [¹³C] glucose and 4 g/L NaH¹³CO₃ (tracers were purchased from Sigma-Aldrich, Saint Louis, USA). The TBDMS (N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide) method (You et al. 2014) was used to analyze the labeling patterns of proteinogenic histidine. In brief, cells were harvested by centrifugation, and cell pellets were hydrolyzed in 6 mol/L HCl solution at 100 °C for 24 hours. The amino acid solution was air-dried and then derivatized by TBDMS (Sigma-Aldrich, USA) at 70 °C for one hour. A gas chromatograph (GC) (Hewlett-Packard model 7890A; Agilent Technologies, CA) equipped with a DB5-MS column (J&W Scientific, Folsom, CA) and a mass spectrometer (model 5975C; Agilent Technologies, CA) were used for analyzing amino acid labeling profiles. The GC-MS fragment
[M-57]+ contains the complete amino acid backbone, and MS data M+0, M+1, and M+2 represent isotopomers with zero, one, and two $^{13}$C atoms, respectively.

### 6.2.3. Flux balance analysis model

The FBA model was modified from the cyanobacterial model iJN878 (Nogales et al. 2012), which has 843 reactions, including photosynthesis and the central carbon metabolism. A complete list of reactions is provided in Supplementary file 1. The iJN878 model contains a recently discovered $\gamma$-aminobutyrate shunt (Xiong et al. 2014) which converts 2-oxoglutarate to succinate in *Synechocystis* 6803. In our model, two new reactions were added, namely ‘glycogen storage $\rightarrow$ glycogen [c]’ and ‘d-lactate [c] $\rightarrow$ d-lactate [external]’, which were respectively used to simulate glycogen storage/consumption and D-lactate production by an engineered cyanobacterial strain (Varman et al. 2013). The mathematical description of our FBA model is as follows:

$$\begin{align*}
\text{maximize } & \mu \\
\text{subject to } & S \cdot v = 0 \\
& lb \leq v \leq ub \\
& v_{CO_2} \leq f_1(K_{La},[CO_2],Km) \\
& v_{\text{photon}} \leq f_2(l,X,v_{\text{photon,0}})
\end{align*}$$

where $\mu$ represents the specific growth rate, $S$ is the stoichiometric matrix, $v$ represents a vector of flux distribution, and $lb$ and $ub$ represent vectors of the lower and upper boundaries, respectively. Further, $f_1$ is a function of the mass transfer coefficient $K_{La}$, dissolved CO$_2$ concentration [CO$_2$], and half-saturation constant for dissolved CO$_2$, $Km$; $f_2$ is a function of the cell’s local position $l$, biomass concentration $X$, and photon influx on the PBR surface $v_{\text{photon,0}}$. The linear optimization problem was solved by the MATLAB® (2012b) built-in function ‘linprog’ using the ‘simplex’
algorithm. To estimate the flux distribution in engineered cyanobacterial strains, we used the MOMA (minimization of metabolic adjustment) algorithm (Segre et al. 2002), which was solved by the MATLAB built-in function ‘quadprog’ using the ‘interior-point-convex’ algorithm:

\[
\begin{align*}
\text{minimize} & \quad \frac{1}{2} v^T H v - f_{\text{opt}}^T v \\
\text{subject to} & \quad S \cdot v = 0 \\
& \quad lb \leq v \leq ub \\
& \quad v_{\text{CO}_2} \leq f_1(K_{La},[\text{CO}_2],Km) \\
& \quad v_{\text{photon}} \leq f_2(l,X,I_0)
\end{align*}
\]

(2)

where \( H \) is a unit matrix, and \( f_{\text{opt}} \) is the optimal flux distribution of wild-type cyanobacteria. The remaining notations have the same meanings as above.

Moreover, we considered three growth states for cyanobacteria in the FBA model: 1) the ‘light condition’: an autotrophic sub-population in the light zone, 2) the ‘dark condition’: a heterotrophic sub-population in the dark zone, where the photon influx is below 0.4 mmol/g/h (under which the cyanobacterial growth rate is lower than the heterotrophic growth rate in darkness) and glycogen is consumed at a rate of 0.01 mmol/g/h (Knoop et al. 2013) to maintain minimal growth, and 3) the ‘resting condition’: a glycogen-depleted sub-population with no active fluxes in the dark zone. To improve the calculation efficiency, we built a database containing all the flux distributions in response to different photon influxes (Supplemental file 1). By having such a database, we could directly use pre-calculated fluxome from the database according to culture conditions in photobioreactors. Thereby, we did not need to redo flux calculations at each time interval during new simulations.
6.2.4. Simulation of cyanobacterial growth via integrating FBA, kinetics, and cell movements

Figure 6.1 shows our modeling algorithm. To simulate biomass growth as a function of time, we divided the entire time period into finite intervals of 0.002 h (Figure 6-S1 shows that further decreasing the interval period did not change the simulation results). In each time interval, a simplified sinusoid equation (Wu and Merchuk 2002) was used to estimate the cell location in a well-mixed PBR:

\[
l = \frac{r}{2} - \frac{r}{2} \cos\left(\frac{2\pi}{f_r} t\right),
\]

where \( l \) is the shortest distance between the PBR surface and the cell local position, in mm; \( r \) is the radius or thickness of the PBR, in mm; \( f_r \) represents the cyanobacteria circulation frequency, in h; and \( t \) is time, in h. Because cell circulation frequencies in PBRs vary from cell to cell, stochastic effects are induced on a single cell’s metabolism. In fact, the random movements of cells in PBRs have been measured and simulated, and, in the present study, are described by a probability distribution function (Luo and Al-Dahhan 2011). In our model platform, we distinguished cell populations with different circulation times (Figure 6-S2 and 3). Thus, the whole culture was considered to be comprised of twelve sub-populations instead of a plethora of individual cyanobacterial cells. Based on cell locations and the biomass concentrations, we calculated the local light intensity (Katsuda et al. 2000):

\[
\frac{v_{\text{photon}}}{v_{\text{photon,0}}} = \left( \frac{1}{(0.0216 \cdot l + 1)^{1.54} \cdot (0.130 \cdot X \cdot l + 1)^{1.18}} \right),
\]

where \( v_{\text{photon}} \) and \( v_{\text{photon,0}} \) represent the local and surface photon influxes, respectively, in mmol/g/h; and \( X \) is biomass concentration, in g/L.
The CO\textsubscript{2} uptake flux was described by a Michaelis-Menten equation:

\[ v_{CO_2} = v_{CO_2,max} \frac{[CO_2]}{K_m + [CO_2]}, \]

(5)

where \( v_{CO_2,max} \) is the maximum uptake rate of dissolved CO\textsubscript{2}/HCO\textsubscript{3}\textsuperscript{-}, \( K_m \) is the half-saturation constant, and \([CO_2]\) represents dissolved CO\textsubscript{2} concentration. This study assumed that pH was constant at 8.0, and that the dissolved CO\textsubscript{2} and cell culture were homogeneous in PBRs. Because we assumed that cell metabolism was pseudo-steady in each interval (Feng et al. 2012), the FBA model could use linear optimization to profile the intracellular fluxes constrained by light and carbon input fluxes (Equations 4 and 5). The FBA model then predicted the growth rates, glycogen synthesis rates, and CO\textsubscript{2} uptake rates of cell populations with different circulation times in PBRs. Those values were averaged based on the probability distribution function (Figure 6-S2a):

\[ \mu_{app} = \sum_{i=1}^{n} P_i \mu_i \]

\[ v_{CO_2,app} = \sum_{i=1}^{n} P_i v_{CO_2,i} \]  

(6)

\[ v_{glycogen,app} = \sum_{i=1}^{n} P_i v_{glycogen,i} \]

where \( P_i \) is the fraction of \( i \)th cell population (Figure 6-S2a), \( \mu_{app} \) is the apparent specific growth rate in PBRs, \( v_{CO_2,app} \) is the apparent CO\textsubscript{2} uptake rate, and \( v_{glycogen,app} \) is the overall glycogen production rate.

For the kinetic model, we used ordinary differential equations (ODEs) to describe changes in biomass production, glycogen accumulation, dissolved CO\textsubscript{2}/HCO\textsubscript{3}\textsuperscript{-} concentrations, and so forth.
\[
\frac{dX}{dt} = \mu_{\text{app}} \cdot X - K_d \cdot X \\
\frac{d[\text{glycogen}]}{dt} = v_{\text{glycogen,app}} \cdot X - \beta \cdot K_d \cdot X \\
\frac{d[CO_2]}{dt} = K_{La} \cdot ([CO_2] - [CO_2]) + v_{CO_2,\text{app}} \cdot X
\]

The ODEs were resolved in their numerical discrete form (Euler-like integration scheme):

\[
X_{i+1} = X_i + \mu_{\text{app}} \cdot X_i \cdot \Delta t - K_d \cdot X_i \cdot \Delta t \\
[\text{glycogen}]_{i+1} = [\text{glycogen}]_i + v_{\text{glycogen,app}} \cdot X_i \cdot \Delta t - \beta \cdot K_d \cdot X_i \cdot \Delta t \\
[CO_{2,i+1}] = [CO_{2,i}] + K_{La} \cdot ([CO_2] - [CO_{2,i}]) \cdot \Delta t + v_{CO_2,\text{app}} \cdot X_i \cdot \Delta t
\]

where \(i\) and \(i+1\) represent the current and next intervals, respectively; \(\Delta t\) is the time interval (0.002 h); \(K_d\) is the death rate, in h\(^{-1}\); \(\beta\) is the glycogen composition in the biomass, in mmol/g; \([\text{glycogen}]\) is the overall glycogen concentration in the PBR, in mmol/L; \(K_{La}\) is the mass transfer rate of CO\(_2\), in h\(^{-1}\); \([CO_2]\) represents the dissolved CO\(_2\) and HCO\(_3^-\) concentrations, in mmol/L; \([CO_2]^*\) is the combined concentrations of dissolved CO\(_2\) and HCO\(_3^-\) in equilibrium with atmospheric CO\(_2\) (0.039%, v/v), in mmol/L; and \(\mu_{\text{app}}, v_{\text{glycogen,app}},\) and \(v_{CO_2,\text{app}}\) are fluxes determined previously. The updated values of the biomass concentration, dissolved CO\(_2\) concentration, etc., were then used to constrain the FBA model in the next interval. The kinetic parameters are given in Table 6.1. For shake flask conditions, \(K_{La}\) was determined by the following equation: (Nikakhtari and Hill 2005)

\[
K_{La} = 0.032 \cdot N \left(\frac{V}{L}\right)^{0.845}
\]

where \(N\) is the rotation speed, in rpm; \(V\) is the shake flask volume, in mL; and \(L\) is the culture volume, in mL. MATLAB files are provided in Supplementary files 3 and 4.
6.3. Results

6.3.1. Simulation of cyanobacterial optimal growth in a cylindrical PBR

The integrated GSM was first applied to predict cyanobacterial growth in a cylindrical PBR, which was assumed to have a radius of 60 mm and a constant surface light intensity of 50 $\mu$E/m$^2$/s. Although the maximal photosynthetic efficiency in photosynthetic species can reach 4.6% - 6% (Blankenship et al. 2011), not all incident radiation in PBRs can be efficiently used by cyanobacteria, thereby resulting in a lower conversion efficiency (Zhu et al. 2008). Hence, we chose a photosynthesis efficiency of 1.5%, which was within a reasonable range of actual photosynthesis efficiencies of microalgae (Melis 2009). Based on a previous study, the mass transfer rate of CO$_2$ was assumed to be 10 h$^{-1}$ (Kazim 2012). Under such a condition, cyanobacterial biomass concentration could increase from 0.1 g/L to 5 g/L in three weeks, provided that other mineral nutrients are supplied continuously (Figure 6.2a and Figure 6-S1). The modelling results also showed continuous decreases in the growth rate (Figure 6.2b) and intracellular fluxes in the central metabolism (Figure 6.2c-e), which was caused by a continuous decrease in local light intensity over time (Figure 6.2g). As the ‘dark zone’ expanded, some cyanobacteria switched from autotrophic growth to heterotrophic growth in the late growth phase, and eventually became resting cells (Figure 6.2f). The expanding ‘dark zone’ also led to a gradual reduction in glycogen content per gram of biomass, which was the same when all the cells were located in the light zone (Figure 6.2b). This prediction agrees with two previous studies (Aikawa et al. 2012; Hasunuma et al. 2013).

Next, we tested the sensitivity of biomass production to the mass transfer rate, PBR surface light intensity, and PBR diameter. With a light intensity of 100 $\mu$E/m$^2$/s and a moderate mass transfer rate of 15 h$^{-1}$, small PBRs (30mm radius) could produce 20 g/L of biomass in 21 days. Although such productivity has been experimentally observed in small PBRs (Pirt et al. 1983), it can be
hardly achieved in large-size PBRs. As shown by the model, the biomass productivity is highly sensitive to the surface-to-volume ratios of the PBRs, and increasing the PBR diameter reduces biomass productivity dramatically. Hence, to improve biomass production in PBRs, one needs to reduce the surface-to-volume ratios, increase the culture mixing and air flow rate (Lee and Palsson 1994), and maintain a sufficient surface light intensity.

Finally, the simulations also demonstrate that, due to random cell movements in PBRs, single cell fluxome may show stochastic changes (Figure 6-S2 b-c). Additionally, we tested the sensitivity of biomass growth to circulation time. The model indicates that perturbing the circulation speeds of cell subpopulations did not affect total biomass production by PBRs, provided that the CO₂ mass transfer and surface light were constant (Figure 6-S3).

6.3.2. Simulation of cyanobacterial oxidative pentose phosphate pathway in a cylindrical PBR

In cyanobacteria, the oxidative pentose phosphate (OPP) pathway and the Calvin cycle operate in opposite directions: The former generates CO₂ and NADPH, while the latter consumes CO₂ and NADPH. Figure 6.2d shows that the Calvin cycle had a strong flux in the early growth phase, while the OPP pathway remained silent under light-sufficient conditions (Figure 6.4a). In the late growth stage, active fluxes through the OPP pathway appeared (Figure 6.4a) due to the self-shading effect. The activity of the OPP pathway increased concurrently with the glycogen consumption rate in darkness (Figure 6.4a and Figure 6-S4). Thus, an active OPP flux in photoautotrophic cultures is the metabolic response to light deficiency in PBRs. Recent ¹³C-flux measurements also showed positive OPP fluxes in *Synechocystis* 6803 PBR cultures (Xiong et al. 2015; Young et al. 2011). To further confirm our model predictions, we examined the labelling
patterns of histidine by growing *Synechocystis* 6803 with NaH$^{13}$CO$_3$ and [1-$^{13}$C] glucose. When glucose was metabolized via the OPP pathway, non-labeled ribose-5-phosphate was generated from [1-$^{13}$C] glucose (You et al. 2014), which is a precursor to histidine. Therefore, an active OPP pathway was expected to reduce the $^{13}$C-enrichment of proteinogenic histidine. Figure 6.4b shows that the $^{12}$C-concentration of histidine was high under low light conditions, supporting the model prediction that light deficiency leads to an active OPP pathway for C6 sugar utilizations.

**6.3.3. Investigation of cyanobacterial photosynthesis efficiency in shake flasks**

Next, we used the integrated GSM to determine the photosynthesis efficiency of *Synechocystis* 6803 by minimizing the sum of squared errors between experimental and simulated averaged specific growth rates. We simplified the geometry of the shake flasks into a two-dimensional rectangle (Figure 6-S5), and made the local light intensity dependent on the vertical distance from a cell to the light source. The CO$_2$ mass transfer rates in shake flasks were calculated based on Eq. 9. As a consequence, a photosynthesis efficiency of 2.7% (Figure 6-S6) resulted from the best fit of specific growth rates under shake flask cultures (Diamond and circle markers in Figure 6.5). Furthermore, this photosynthesis efficiency was used to simulate the growth of *Synechococcus elongatus* UTEX 2973 (a fast-growing cyanobacterium species) in a column PBR (with 3% CO$_2$ and under 300 $\mu$E/m$^2$/s light intensity) (Yu et al. 2015). The model predicted slightly lower specific growth rate than the experimental value (Square marker in Figure 6.5). This difference is possibly due to an increased photosynthesis activity under high CO$_2$ concentrations (Levitan et al. 2007).

**6.3.4. Model-based investigation of lactate production by engineered cyanobacteria in PBRs**

We further applied the integrated GSM to predict the growth and volumetric D-lactate productivity of engineered cyanobacterial strains, in which a mutated glycerol dehydrogenase was
overexpressed for producing optically pure D-lactate (Varman et al. 2013). The MOMA algorithm was applied to simulate the metabolism in engineered strains (See Methods). Growth-associated lactate production was assumed (i.e., lactate production was proportional to biomass synthesis). First, we tested the relationship between lactate efflux ($v_{lac}$) and specific growth rate ($\mu$) using only the FBA model. Figure 6.6a shows $v_{lac}$ and $\mu$ as functions of the ratio $v_{lac}/\mu$, which denotes the amount of lactate produced per gram of biomass (or mmol lactate/ g biomass). Within a wide $v_{lac}/\mu$ range, from 0.01 to 100 mmol lactate /g biomass, $\mu$ decreased with increasing $v_{lac}/\mu$, but $v_{lac}$ showed a parabolic tendency, peaking at 0.3 mmol/g/h (Figure 6a). Next, we used the integrated GSM to simulate the cyanobacterial growth and D-lactate production in PBRs at different $v_{lac}/\mu$ ratios (Figure 6 b-c). As a result, increasing the $v_{lac}/\mu$ ratio led to lower biomass production, which, however, did not necessarily improve the overall D-lactate production. For example, when $v_{lac}/\mu$ was increased from 1 to 10 mmol/g, lactate production in PBRs remained the same, but biomass production was significantly diminished. Enhancing the lactate pathway (i.e., increase $v_{lac}/\mu$ ratio) can improve lactate production, but excessive overexpression of this pathway may sacrifice biomass growth and impair overall lactate productivity. To resolve this problem, it is desirable to induce the lactate synthesis pathway at late biomass growth phase.

Deleting carbon storage in cyanobacteria is one strategy to redirect carbon flux to product synthesis (van der Woude et al. 2014). However, the integrated GSM shows that such a strategy may not offer significant improvements in final lactate productivity in PBRs (Figure 6.6a, d-e). This finding is consistent with two recent reports: 1) Glycogen knockout did not enhance lactate productivity under nutrient-sufficient growth conditions (van der Woude et al. 2014). 2) Removal of glycogen in an isobutanol-producing cyanobacterium yielded no benefit in production titer and rate (Li et al. 2014). Possibly, glycogen serves as the carbon and energy reservoir to store the energy and carbon.
excess flow in the light zone, and this carbon and energy reservoir can maintain redox homeostasis under stressed growth conditions or in darkness (Gründel et al. 2012). Therefore, deleting glycogen or other carbon storage may impair cyanobacterial survival as well as its resistance to environmental stresses and contaminations.

6.4. Discussion

In this study, a genome-scale FBA model was integrated with information on kinetics, light distribution, and cell movement. Using the integrated GSM, one can simultaneously learn both intracellular information (e.g., flux distributions as functions of time) and extracellular information (e.g., growth curve and nutrient changes in the medium) simultaneously.

In the extracellular domain, the integrated GSM can describe changes in nutrient concentrations, biomass accumulation, and local light intensities. As demonstrated by Figure 6.3, cyanobacteria are intrinsically capable of reaching high biomass concentration in PBRs, however, their performance is usually limited by low light availability and low mass transfer rate. To reduce cell self-shading, high surface-to-volume ratio PBRs equipped with thin panel or hollow fibers have been employed (Lee and Palsson 1994). To improve the mass transfer rate, enhancing aeration rates has proved efficient (Kazim 2012). Better mixing conditions not only lead to better gas transfer rates, but also help maintain more homogeneous conditions for both cells and nutrients.

In the intracellular domain, we observed continuously changing fluxes in the cyanobacterial metabolic network, which were mostly affected by reduced energy and carbon inputs. One interesting finding is the OPP pathway activation as cell cultures get denser. It is a traditional point of view that the Calvin cycle and the OPP pathway are separate systems because the two pathways are reciprocally regulated (Poolman et al. 2003). In this study, the integrated GSM describes that
these two pathways could be employed by two different subpopulations in PBRs at the same time. This simulation explains that a measureable flux through the OPP pathway could be observed in both wild type and engineered cyanobacterial strains via $^{13}$C-based flux analysis (Xiong et al. 2015; Young et al. 2011). In addition, the elementary modes analysis shows that the Calvin cycle and the OPP pathway may function in complementary ways in photoautotrophs, since an active OPP pathway ensures a maintainable flux to triose phosphate synthesis from carbohydrate degradation in low light or darkness (Poolman et al. 2003). Furthermore, we noticed that an active OPP pathway was always present in the D-lactate producing strain, and it became stronger with increased D-lactate production (Figure 6-S7). This indicates that the enhanced OPP activity benefits bio-production by providing more reducing power. In summary, the plasticity of the OPP pathway endows cells with high vitality and energy flexibility (Wasylenko et al. 2015; Wu et al. 2015).

The traditional FBA model usually describes the optimal growth condition, and thus it may not be suitable for suboptimal and heterogeneous cultivation conditions. Our model, on the other hand, is integrated with growth kinetics and a heterogeneous light distribution in PBRs. Hence, the model can not only predict the production yield, titer and rate, but also offer insights into how cells adjust their internal metabolisms to survive under different growth conditions and genetic manipulations. Moreover, the integrated GSM may give more accurate predictions of mutant physiology than GSM alone in bioreactor conditions. For example, the integrated GSM correctly indicates that glycogen knockout may not be an effective strategy to improve PBR lactate production. Lastly, the integrated GSM can reveal real-time variations/dynamics in metabolisms of different subpopulation cells, and thus improve understandings of cellular responses to large-size PBRs.
Nevertheless, our model still has limitations. For example, previous studies have shown that glycogen synthesis could be connected with unknown regulations affecting cyanobacterial viability under stress conditions (Gründel et al. 2012; Xu et al. 2013). However, the integrated GSM may not give the same prediction without further constraints from knowledge of genetic regulations. Additionally, it has been demonstrated that cyanobacteria have circadian behaviors (i.e., their metabolism exhibits day and night rhythms) (Kondo et al. 1993), while our model does not include this property. Finally, some inhibition factors may also influence cyanobacterial growth (e.g., effects of crowding), which are not included in the model. In the future, this model platform should be further improved via additional multi-scale modeling approaches.

6.5. Conclusions

This study demonstrates a genome-scale FBA model integrated with kinetics, cell movements, and a light distribution function. With constraints obtained from bioprocess variables, the integrated GSM can not only simulate the dynamic metabolisms in sub-population cells but also predict PBR overall productivity under light and CO$_2$ conditions. The integration of GSMS with PBR modeling can facilitate the development of new cyanobacterial strains for industrial settings.

6.6. Acknowledgements

I would like to thank Dr. Gang Wu for his contributions to the modeling work and Dr. Le You for valuable discussions for this project. In addition, Ni Wan and Adrienne Redding contributed to cyanobacterial growth experiments directly. We are also grateful for the funding supports from a US DOE grant (DESC0012722) and an NSF Grant (CBET 1438125).
6.7. References


Kazim SA. 2012. Experimental and empirical correlations for the determination of the overall volumetric mass transfer coefficients of carbon dioxide in stirred tank bioreactors: The University of Western Ontario. 48 p.


Table 6.1: List of parameters used to simulate the growth and metabolic fluxes of cyanobacteria growing in a cylindrical PBR.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Significance</th>
<th>Value (Range)</th>
<th>Unit</th>
<th>References/Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{La}$</td>
<td>Mass transfer rate of CO$_2$</td>
<td>$10^a$ (3-15)</td>
<td>h$^{-1}$</td>
<td>(Kazim 2012)</td>
</tr>
<tr>
<td>$K_d$</td>
<td>Death rate</td>
<td>0.0079</td>
<td>h$^{-1}$</td>
<td>(Kim et al. 2011)</td>
</tr>
<tr>
<td>Radius</td>
<td>Radius of PBR</td>
<td>60</td>
<td>mm</td>
<td>Similar to the reactor used in reference (Luo and Al-Dahhan 2011)</td>
</tr>
<tr>
<td>$I_0$</td>
<td>Surface light intensity</td>
<td>50</td>
<td>mmol/g/h</td>
<td>Equivalent to ~50 $\mu$E/m$^2$/s $^b$</td>
</tr>
<tr>
<td>$K_m$</td>
<td>Half-saturation constant of CO$_2$ uptake rate</td>
<td>8</td>
<td>$\mu$mol/L</td>
<td>(Badger and Andrews 1982)</td>
</tr>
<tr>
<td>pH</td>
<td>Medium pH</td>
<td>8.0</td>
<td>unitless</td>
<td>BG-11 medium</td>
</tr>
<tr>
<td>$[X]_0$</td>
<td>Initial biomass concentration</td>
<td>0.1</td>
<td>g/L</td>
<td>Equivalent to an OD$_{730}$ of ~0.2</td>
</tr>
<tr>
<td>$[CO_2]_0$</td>
<td>Initial concentration of dissolved CO$_2$ and HCO$_3^-$ (in equilibrium with air)</td>
<td>0.53</td>
<td>mmol/L</td>
<td>Estimated $^c$</td>
</tr>
</tbody>
</table>

Note: a) The value is used in Figures 6.2, 6.4 and 6.6; b) a photosynthesis efficiency of 1.5% is assumed in Figures 6.2, 6.4 and 6.6; and c) calculation is based on experimental conditions (Supplementary Materials).
Figure 6.1: Algorithm for simulating cyanobacterial growth and intracellular flux distribution by integrating the flux analysis model, kinetics, and a light distribution function.
Figure 6.2: Simulations of dynamics of cyanobacterial performance in a 60mm-radius cylindrical PBR under 50 µE/m²/s surface light intensity. (a) biomass concentration and dissolved CO₂/HCO₃⁻ concentration; (b) glycogen content in biomass and average specific growth rate; (c) flux through the PEP carboxylase reaction; (d) flux through the RuBP carboxylase reaction; (e) fluxes through the TCA cycle (negative flux means the flux direction is reversed); (f) ratios of autotrophic, heterotrophic, and resting sub-populations; and (g) light distribution in the bioreactor as a function of time and relative distance r/R, where r is the local position and R is the radius of the bioreactor. The parameters used for simulation are given in Table 5.1. Abbreviations of metabolites: 3PG, 3-phosphoglycerate; AceCoA, acetyl-CoA; CIT, citrate; MAL, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; and RuBP, ribulose-1,5-bisphosphate.
Figure 6.3: Test of biomass growth performance sensitivity to the mass transfer rate ($K_{La}$ in h$^{-1}$), surface light intensity ($I_0$ in $\mu$E/m$^2$/s), and the bioreactor geometry (R in mm). Each three-dimensional figure shows biomass growth as a function of time and surface light intensity.
Figure 6.4: The oxidative pentose phosphate (OPP) pathway in cyanobacteria. (a) Comparison of simulated fluxes through the OPP pathway between wild-type (black bars) and glycogen-rich (white bars) cyanobacteria strains at different biomass concentrations. ‘*’ means the flux is zero. Compared to the wild type strain, the model assumes that the glycogen-rich strain accumulates five times more glycogen during autotrophic growth and consumes glycogen five times faster during heterotrophic growth. (b) Relative abundance of histidine labeling profile under different *Synechocystis* growth phases. Black bars (low light/early growth phase): light intensity of ~50 $\mu$E/m$^2$/s; white bars (low light/late growth phase): light intensity of ~50 $\mu$E/m$^2$/s; grey bars (high light/late growth phase): light intensity of ~100 $\mu$E/m$^2$/s.
Figure 6.5: Comparison of experimental and simulation results of cyanobacterial growth rates in shake flasks. Diamond markers represent cultures with different volumes under the same light intensity of \( \sim 50 \, \mu E/m^2/s \). The volumes are 50mL, 100mL, and 150mL, respectively, corresponding to diamond markers from top to bottom. Circle markers represent 15 mL cultures growing under different light conditions. The light intensities are \( \sim 15 \, \mu E/m^2/s \), \( \sim 25 \, \mu E/m^2/s \) and \( \sim 35 \, \mu E/m^2/s \), respectively, corresponding to circle markers from top to bottom. The specific growth rates were calculated based on OD\(_{730}\) values in the early exponential growth phase. The square marker represents a reported cyanobacterial growth rate in a mini-PBR (Yu et al. 2015). The parameters used for simulating cyanobacterial growth are given in Table S1.
**Figure 6.6:** Simulation results of D-lactate producing cyanobacteria performance. (a) FBA simulations of D-lactate flux and growth rate as functions of $v_{\text{lac}}/\mu$ (mmol lactate/g biomass). White markers: wild-type strain; black markers: glycogen-knockout strain. (b-c) Simulation of biomass growth (b) and D-lactate production (c) of wild-type cyanobacteria at different $v_{\text{lac}}/\mu$ ratios in a cylindrical PBR. (d-e) Simulation of biomass growth (d) and D-lactate production (e) of glycogen-knockout cyanobacteria at different $v_{\text{lac}}/\mu$ ratios in a cylindrical PBR.
Supplementary Materials

The following Supplemental materials can all be found online (DOI: 10.1186/s12934-015-0396-0). Tables and figures in Supplemental file 2 are presented in the following pages.

Supplemental file 1. Flux distribution database.

Supplemental file 2. Supplementary figures and tables.

Supplemental file 3. MATLAB file for simulating cyanobacterial performance in PBRs.

Table 6-S1: Parameters used for simulating cyanobacterial growth in shake flasks. The remaining parameters used in the model are the same as shown in Table 1 in the article.

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Significance</th>
<th>Value (Range)</th>
<th>Unit</th>
<th>Reference/Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{La}$</td>
<td>Mass transfer rate of CO$_2$</td>
<td>8-55</td>
<td>h$^{-1}$</td>
<td>Calculated based on reference (Nikakhtari and Hill 2005)</td>
</tr>
<tr>
<td>$K_d$</td>
<td>Death Rate</td>
<td>0</td>
<td>h$^{-1}$</td>
<td>n/a</td>
</tr>
<tr>
<td>Thickness</td>
<td>Thickness of the culture</td>
<td>~5, ~15, ~30, ~45</td>
<td>mm</td>
<td>Measured for, 15 mL, 50 mL, 100 mL, and 150 mL cell cultures, respectively, in 250 mL shake flasks</td>
</tr>
<tr>
<td>$X_0$</td>
<td>Initial biomass concentration</td>
<td>0.01</td>
<td>g/L</td>
<td>Measured from the experiment</td>
</tr>
</tbody>
</table>
Figure 6-S1: The influence of time intervals on cyanobacterial growth. We tested different time intervals in the integrated FBA model, and then we plotted the biomass growth as a function of time. The results show that choosing a time interval between 0.001-0.005 h is good for our model.
Figure 6-S2: Probability distribution of cell circulation time in photobioreactors. The distribution is based on the reference (Luo and Al-Dahhan 2011), in which the circulation times follow the normal distribution with the mean and standard deviation being 10 s and 1.3 s, respectively. Instead of using a continuous probability function, we used a discrete probability mass function shown below (Figure a). Figure b and c show the oscillations of specific growth rate and cell distance to the PBR surface in 10 minutes, respectively.
Figure 6-S3: Cyanobacterial growth with different probability distributions. We predicted cyanobacterial growth in a 60mm-radius cylindrical PBR with a constant external photon influx of 50 \( \mu \text{E/m}^2/\text{s} \) and a \( K_{La} \) value of 10 h\(^{-1}\). Different circulation times are tested. The results show that the circulation time does not affect the total PBR biomass productions. The growth curves shown below overlap each other.
Figure 6-S4: Specific growth rates and fluxes through the oxidative pentose phosphate pathway under heterotrophic growth conditions in darkness.
Figure 6-S5: **Simplified geometry of shake flask.** Unlike previous simulations, we simplified the geometry of the shake flasks into a two-dimensional rectangle, and made the local light intensity dependent on the vertical distance from a cell to the light source.
Figure 6-S6: Determining the photosynthesis efficiency of cyanobacteria growing in 250 ml shake flasks.

Best fit is achieved at 2.7%.
Figure 6-S7: Comparison of flux distributions (unit: mmol/g/h) in central metabolism between wild-type, glycogen knockout, and D-lactate producing cyanobacterial strains. Legends: A, wild-type strain; B, glycogen-knockout strain; C, D-lactate producing strain (1 mmol/g); D, glycogen knockout and D-lactate producing strain (1 mmol/g); E, D-lactate producing strain (10 mmol/g); and F, glycogen-knockout and D-lactate-producing strain (10 mmol/g). The model assumes lactate production to be 1mmol/g biomass or 10 mmol/g biomass. Unlike the wild-type strain, lactate-producing cyanobacteria with a $v_{lac}/\mu$ of 1 mmol/g appears to have an active OPP flux (0.01 mmol/g/h). As $v_{lac}/\mu$ is increased to 10 mmol/g, the OPP flux becomes more active (0.04 mmol/g/h). An active OPP flux can be beneficial by providing more NADPH for lactate synthesis.
Additional supplementary information for the integrated genome-scale model

(1) The mass transfer coefficients of CO$_2$ and O$_2$ (Royce and Thornhill 1991).

\[ \frac{K_{La,CO_2}}{K_{La,O_2}} = \sqrt{\frac{D_{CO_2}}{D_{O_2}}} = 0.89, \]

where $K_{La,CO_2}$ and $K_{La,O_2}$ are the mass transfer rates of CO$_2$ and O$_2$, respectively; and $D_{CO_2}$ and $D_{O_2}$ are the diffusion coefficients of CO$_2$ and O$_2$, respectively.

(2) Estimation of dissolved CO$_2$ in equilibrium with ambient air. The parameters are based on the textbook Water Chemistry (Benjamin 2002).

\[ \frac{[H^+][HCO_3^-]}{[H_2CO_3]} = 10^{-6.4} \quad \frac{[H^+][CO_3^{2-}]}{[HCO_3^-]} = 10^{-10.3} \]

\[ [H_2CO_3] = p_{CO_2} \cdot H_{CO_2} = 0.00039 \times 3.4 \times 10^{-2} = 1.3 \times 10^{-5} M \]

\[ [H_2CO_3] + [HCO_3^-] = [H_2CO_3] \left(1 + \frac{10^{-6.4}}{[H^+]} \right) \]

In this study, we assumed that pH was constant as 8.0. Thus the total carbon source available in the medium would be 0.53mM. Following is a table showing $[CO_2/H_2CO_3] + [HCO_3^-]$ at different pHs.
Chapter 7: Glycine cleavage powers photoheterotrophic growth of

*Chloroflexus aurantiacus* in the absence of H$_2$
7.1. Introduction

*Chloroflexus aurantiacus* is a filamentous anoxygenic phototrophic bacterium isolated from hot springs (Hanada and Pierson 2006). It has specialized light-harvesting antenna machines and performs a cyclic photosynthetic electron transport via a type II reaction center (Tang and Blankenship 2013). Its photosystem does not generate NADPH directly, but can convert light energy into ATP via photosynthetic electron transfer. The genome of *C. aurantiacus* strain J-10-fl has been sequenced to facilitate our understanding of its physiology and cellular metabolism (Tang et al. 2011). *C. aurantiacus* shows a versatile carbon metabolism. In aerobic and dark conditions, it can grow on various organic substrates chemoheterotrophically. *C. aurantiacus* switches to photoheterotrophic growth when supplied with acetate under anaerobic and light conditions. In the presence of H$_2$/CO$_2$, *C. aurantiacus* can perform a photoautotrophic growth by fixing CO$_2$ via the 3-hydroxypropionate (3HOP) bi-cycle pathway (Eisenreich et al. 1993; Zarzycki et al. 2009).

In its natural habitat, *C. aurantiacus* consumes organic nutrients (e.g., short-chain fatty acids, acetate, etc.) released from cyanobacteria in the associated microbial mats (Hanada and Pierson 2006; Lee et al. 2014). Its phototrophic metabolisms for coassimilation of organic substrates have been extensively studied. For photoheterotrophic culture of *C. aurantiacus*, the common growth medium requires complex nutrients such as yeast extract or casamino acids (Madigan et al. 1974). In some early work on photoheterotrophic cultures, both acetate and mixed gases of H$_2$ and CO$_2$ were provided (Strauss et al. 1992). However, we found that strain J-10-fl was unable to grow well in acetate-based minimal media with phosphate buffer without amino acids and H$_2$ supplies. To delineate key exogenous nutrients demanded by *C. aurantiacus*, we performed $^{13}$C-tracer experiments. The results enhanced our understanding of *C. aurantiacus* carbon and energy metabolisms, and shed lights on *C. aurantiacus* survival in the ecosystem.
7.2. Materials and Methods

All chemicals and labeled substrates (\(^{13}\)C-sodium acetate and NaH\(^{13}\)CO\(_3\), purity >98\%) were purchased from Sigma-Aldrich (St. Louis, MO, USA). C. aurantiacus strain J-10-fl was grown in a minimal PE medium at 55 °C. One liter medium contained 5 mL phosphate buffer solution, 5 mL basal salt solution, 0.5 g Na\(_2\)S\(_2\)O\(_3\) and 0.5 g (NH\(_4\))\(_2\)SO\(_4\). One liter of phosphate solution contained 75 g KH\(_2\)PO\(_4\) and 78 g K\(_2\)HPO\(_4\). One liter of basal salt solution contained 4.12 g Na\(_3\)EDTA, 1.11 g FeSO\(_4\)·7H\(_2\)O, 24.65 g MaSO\(_4\)·7H\(_2\)O, 2.94 g CaCl\(_2\)·2H\(_2\)O, 23.4 g NaCl and 10 mL tracer solution. One liter of tracer solution contained 11.2 g MnSO\(_4\)·4H\(_2\)O, 2.88 g ZnSO\(_4\)·7H\(_2\)O, 2.92 g Co(NO\(_3\))\(_2\)·4H\(_2\)O, 2.52 g CuSO\(_4\)·5H\(_2\)O, 2.42 g Na\(_2\)MoO\(_4\)·2H\(_2\)O, 3.1 g H\(_3\)BO\(_3\) and 41.2 g Na\(_3\)EDTA. The medium pH was adjusted to 7.5. Commercial RPMI 1640 vitamins solution (100X, Sigma-Aldrich) was added into the sterile medium. In nutrient studies, yeast extract or amino acids were added into medium to test their effects on cell growth. For photoheterotrophic cultivation, the anaerobic cultures (purged with N\(_2\)) were grown in sealed serum bottles (containing 30 mL culture) under continuous illumination (20-30 \(\mu\)mol photons m\(^{-2}\) s\(^{-1}\)). Neither H\(_2\)(g) nor CO\(_2\)(g) was provided in the bottle headspace. Biomass growth was monitored based on optical densities at 600nm.

Strain J-10-fl was cultivated in four tracer media: a) 2 g/L \([1-^{13}\)C]acetate and 0.2 g/L unlabeled yeast extract; b) 2 g/L \([1,2-^{13}\)C\(_2\)]acetate (fully labeled acetate) and 0.05 g/L unlabeled glycine; c) 2 g/L \([1,2-^{13}\)C\(_2\)]acetate, 0.05 g/L glycine, and 0.5 g/L NaH\(^{13}\)CO\(_3\); and d) 2 g/L unlabeled acetate, 0.05 g/L unlabeled glycine, and 0.5 g/L NaH\(^{13}\)CO\(_3\). In each tracer experiment, exponentially growing cells from unlabeled culture were inoculated in the labeled medium at a volume ratio of 1%. The protocol for \(^{13}\)C-metabolite analysis was described elsewhere (Tang et al. 2007; You et al. 2012). In brief, the biomass was hydrolyzed in HCl solution (100 °C), and the resulting amino
acid mixtures were derivatized by $N$-(tert-butyldimethylsilyl)-$N$-methyltrifluoroacetamide (TBDMS). GC-MS equipped with a DB5-MS column (Agilent Technologies, USA) was used to measure amino acid labeling pattern. The mass isotopomer distributions (MIDs) of amino acids were calculated based on five MS fragments (Wahl et al. 2004): [M-159]$^+$ or [M-85]$^+$ (both containing the carbon skeleton of amino acids after loss of their α carboxyl groups), [M-15]$^+$ or [M-57]$^+$ (both containing the whole carbon skeleton of amino acids), and f302 (containing the first and the second carbon of amino acids). The final MS results, M0, M1 and M2, represent unlabeled, singly labeled, and doubly labeled amino acids, respectively (Figure 7.1). We also measured glycine concentrations in the cultures by GC-MS. In brief, culture samples were centrifuged and supernatant was collected. After the supernatant was dried, glycine from the supernatant was derivatized by TBDMS. Meanwhile, we determined the relationship between the glycine standards and corresponding MS abundances, which were used to estimate extracellular glycine concentrations in cultures.

7.3. Results

*C. aurantiacus* exhibits optimal growth in photoheterotrophic conditions when the medium is supplemented with yeast extract. It can also grow well with acetate and H$_2$ in a minimal medium (Strauss et al. 1992). In our experiments, strain J-10-fl supplied with yeast extract reached an OD above 1.0 within six days during photoheterotrophic growth. Without H$_2$ and yeast extract in the minimal medium, however, photoheterotrophic cultures grew poorly. To identify the biomass building blocks that were not effectively synthesized from primary substrate acetate, photoheterotrophic cultures were supplied with $^{13}$C-labeled sodium acetate and unlabeled yeast extract. Figure 7.1 shows the contribution of yeast extract to the synthesis of proteinogenic amino acids. Based on tracer experiments, a high fraction of $^{13}$C-carbon (40%~70%) was incorporated
into several amino acids, such as Ala, Glu, Met, and Asp, suggesting effective synthesis of those amino acids from acetate. Since the labeled carbon of [1-{\textsuperscript{13}}C] acetate (purity >98\%) would be mostly incorporated into amino acids via central metabolic pathways, high unlabeled fractions of amino acids (e.g., Gly, Leu, and Phe) indicated that they were avidly absorbed from exogenous sources.

We then investigated the influences of these highly imported amino acids on the photoheterotrophic growth by replacing yeast extract with different amino acids. We observed the most significant growth enhancement (3~5 folds) when glycine was supplied (Figure 7.2), and a glycine concentration between 0.05 g/L and 0.25 g/L seems to suffice for enhancing the bacterial growth (Figure 7-S1). The influence of inorganic carbon was also investigated, which confirmed that addition of NaHCO\textsubscript{3} (0.5 g/L) in medium slightly promoted photoheterotrophic growth, but much less than glycine did. Therefore, inorganic carbon source was not the rate limiting nutrients to the strain J-10-fl, and glycine played an important role in supporting the bacterial growth. These observations thus intrigued us to further investigate the role of glycine in \textit{C. aurantiacus} metabolism.

To this end, we first traced the fate of glycine in the central carbon metabolism via \textsuperscript{13}C labeling experiments. We cultivated strain J-10-fl photoheterotrophically with [1,2-{\textsuperscript{13}}C\textsubscript{2}] acetate and unlabeled glycine, and then measured the isotopomer distributions of proteinogenic amino acids. Figure 7.3 shows that \textsuperscript{13}C from labeled acetate was incorporated into many key proteinogenic amino acids (e.g., Ala, Asp, and Glu) at significant levels (>90\%), suggesting that glycine was not used as a carbon source for biomass synthesis. The only significant contributions of glycine to proteinaceous amino acids were glycine at 70\% and serine (converted via glycine hydroxymethyltransferase, Caur_2543) at 30\% (Table 7-S1). Substituting glycine with serine in
minimal medium led to only a slight increase in cell growth (Figure 7-S1). Therefore, the profound growth enhancement by addition of glycine appears to be predominately due to factors other than direct conversion to biomass.

Glycine metabolism of *C. aurantiacus* has been previously investigated during photoautotrophic growth (Herter et al. 2001). The researchers have shown that: 1) *C. aurantiacus* does not generate glycine from glyoxylate (a product of the 3HOP pathway), and 2) it exhibits high enzyme activity to produce C1 units from glycine cleavage. C1 units, including 5, 10-methylene-tetrahydrofolate (5, 10-methylene-THF), 5-methyl-THF and 10-formyl-THF, participate in biosynthesis of amino acids and inosine monophosphate. For example, 5, 10-methylene-THF, derived from glycine degradation, can be converted to 5-methyl-THF and 10-formyl-THF. The former contributes to methionine synthesis and the latter histidine synthesis. As for cultures growing with [1,2-13C2] acetate and unlabeled glycine, the most abundant isotopologues of the resulting proteinogenic Met and His have a single unlabeled carbon (Figure 7.3), which suggested that C1 units are mainly synthesized from glycine cleavage under photoheterotrophic conditions. Following is the glycine cleavage reaction (Kikuchi et al. 2008):

\[
\text{Glycine} + \text{THF} + \text{NAD}^+ \leftrightarrow 5, 10\text{-methylene-THF} + \text{CO}_2 + \text{NH}_3 + \text{NADH} + \text{H}^+ 
\]

Glycine cleavage can also be coupled with the serine hydroxymethyltransferase reaction:

\[
2 \text{Glycine} + \text{NAD}^+ + \text{H}_2\text{O} \leftrightarrow \text{Serine} + \text{CO}_2 + \text{NH}_3 + \text{NADH} + \text{H}^+ 
\]

In fact, the glycine consumption rate was higher than its requirement as a carbon source for biomass growth (Figure 7-S2). Therefore, part of glycine must be cleaved and oxidized by the THF-dependent C1 pathway, which contains successive steps that oxidize 5, 10-methylene-THF
to formate or CO$_2$, generating both NADPH and ATP (Fan et al. 2014). The fate of glycine/C1-metabolism and the distribution of C1-metabolic enzymes/genes have been discussed for bacteria and archaea, including *Chloroflexi* (Braakman and Smith 2012). These studies suggest that glycine cleavage and C1 degradation can serve as a key energetic route to produce ATP, NADH and NADPH. Therefore, it is likely that glycine is actively involved in the energy metabolism of *C. aurantiacus* under photoheterotrophic conditions.

In addition to glycine metabolism, we were also interested in investigating the CO$_2$ fixation activity under photoheterotrophic conditions in the absence of H$_2$, as *C. aurantiacus* possesses multiple carbon fixation routes (Tang et al. 2011). To this end, we added NaH$^{13}$CO$_3$ (0.5 g/L) to minimal medium containing unlabeled acetate and glycine (Figure 7.4A-D). A protein BLAST search against either *E. coli* or *Synechocystis* 6803 carbonic anhydrase on NCBI website suggests that the gene encoding carbonic anhydrase that converts H$^{13}$CO$_3^-$ to $^{13}$CO$_2$ is missing in *C. aurantiacus*, and an alternative $^{13}$CO$_2$ source would be from the HCO$_3^-$-CO$_2$ equilibrium in the culture medium (pH=7.5). Although addition of NaHCO$_3$ did not appear to promote J-10-fl growth (Figure 7.2), enzyme activities of CO$_2$ fixation was measurable due to significant $^{13}$C incorporation into proteinogenic amino acids (e.g., ~40% alanine is singly labeled and 5% alanine is doubly labeled, Figure 7.4A and Table 7-S1). As pyruvate is involved in the CO$_2$ fixation pathways and also a precursor to alanine, the labeling patterns of alanine can reflect CO$_2$ fixation routes (Figure 7.5). MS data show that alanine was mostly labeled at 1$^{st}$ position, and a small fraction of alanine was labeled at both 1$^{st}$ and 2$^{nd}$ positions (Table 7-S1). Figure 5B shows the origins of different labeling patterns of alanine. Firstly, unlabeled acetyl-CoA and H$^{13}$CO$_3^-$ can be condensed to [1-$^{13}$C]pyruvate by pyruvate:ferredoxin oxidoreductase (PFOR, Caur_2080) or the 3HOP pathway. After a second H$^{13}$CO$_3^-$ is incorporated in the 3HOP pathway, doubly labeled malyl-CoA is formed,
subsequently resulting in the formation of \([1^{-13}C]\)acetyl-CoA via a cleavage reaction. When the PFOR reaction or the 3HOP pathway converts \([1^{-13}C]\) acetyl-CoA into pyruvate, \([1,2^{-13}C_2]\) pyruvate is generated, which explains the origin of \([1,2^{-13}C_2]\) alanine. Under anaerobic conditions without \(H_2\), the flux of the 3HOP pathway for \(CO_2\) fixation is weak, and \([1,2^{-13}C_2]\) alanine only accounts for \(~5\%\) of the total alanine (Table 7-S1), probably due to the lack of reducing equivalents to power this pathway. As a consequence, most \([1^{-13}C]\) pyruvate would be generated from the PFOR reaction, which also leads to production of \([1^{-13}C]\) serine and \([1^{-13}C]\) glycine (Figure 7.4B and C) after pyruvate is converted into downstream metabolites. Figure 7.4D shows similar labeling patterns between Met (derived from oxaloacetate and C1 unit) and Asp (derived from oxaloacetate), further confirming that most C1 units come from glycine degradation (Figure 7.5A). This phenomenon also indicates that \(^{13}CO_2\) is not preferred to be fixed via the reductive C1 pathway (\(CO_2 \leftrightarrow \) formyl-THF \(\leftrightarrow 5,10\)-methylene-THF) under photoheterotrophic conditions, as 5, 10-methylene-THF can be continuously generated from glycine. We conclude that reducing equivalents from glycine degradation are insufficient to drive appreciable \(CO_2\) fixation as part of promoting heterotrophic growth.

Figure 7.4E shows that \(^{13}C\) enrichment shifts of various amino acids after \(NaH^{13}CO_3\) addition into minimal medium containing \([1,2^{-13}C_2]\) acetate and unlabeled glycine. When strain J-10-fl was grown without \(NaH^{13}CO_3\), 70% proteinogenic glycine and 40% proteinogenic serine were unlabeled. Other key proteinogenic amino acids (i.e., Ala, Asp and Glu) were labeled significantly (\(^{13}C\) enrichments \(~90\%\)). The small fraction of \(^{12}C\) in these amino acids was possibly from the metabolic assimilation of \(^{12}CO_2\) released from glycine degradation. When strain J-10-fl was cultivated with \(NaH^{13}CO_3\), unlabeled glycine and fully labeled acetate, \(^{13}C\) enrichments in all
proteinogenic amino acids were further raised (e.g., ~95% of Glu carbons were $^{13}$C-carbons). This observation confirms photoheterotrophic CO$_2$ co-utilizations by strain J-10-fl.

### 7.4. Discussion

In this study, we attempted to understand why glycine can promote *C. aurantiacus* growth on acetate. Herter and coworkers have previously reported the contribution of glycine to photoautotrophic growth of *C. aurantiacus* strain OK-70-fl with H$_2$/CO$_2$ (Herter et al. 2001). The study indicated that glycine could participate in serine synthesis and contribute approximately half of the C1 units to biomass synthesis. As a comparison, we tested the growth effect of glycine on strain OK-70-fl strain and revealed a similar growth enhancement (by 3~4 folds) on strain J-10-fl during photoheterotrophic growth (Figure 7-S3 and related contents). Whether glycine can promote photoheterotrophic growth in the presence of H$_2$ remains to be tested.

*C. aurantiacus* does not contain a type I reaction center, and its cyclic photosynthetic electron transport system within the type-II reaction center generates ATP. We searched the genome database of strain J-10-fl and found the gene encoding a nickel-dependent hydrogenase (Caur_1188), which may generate NADPH in the presence of H$_2$. However, when *C. aurantiacus* grows on acetate without H$_2$, its NADPH generation could be less efficient. The major routes in central carbon metabolism for producing reducing power are the TCA cycle and oxidative pentose phosphate pathway (OPPP). However, acetate-based metabolism usually does not show strong fluxes through the OPPP, and both PFOR reaction and gluconeogenesis consume reducing equivalents. Additionally, genes encoding transhydrogenase have not yet been reported to exist in the genome of *C. aurantiacus*. Therefore, NADH and NADPH production from glycine cleavage and C1 degradation may increase energy flexibility and thus promote *C. aurantiacus* anaerobic
growth. On the other hand, the THF-dependent C1 degradation pathway in *C. aurantiacus* appears to be influenced by H₂. In our study, glycine cleavage contributes to C1 units for both Met and His synthesis under photoheterotrophic conditions. In comparisons, Herter and coworkers have investigated *C. aurantiacus* phototrophic growth with fully labeled glycine and H₂/unlabeled CO₂ (Herter et al. 2001). Their labeling data of Met and His revealed that 5,10-methylene-THF (the C1 unit for Met synthesis) derived from fully labeled glycine was not converted to formyl-THF (the C1 unit for His synthesis). This phenomenon implied that H₂ may inhibit the C1 unit oxidation, and that formyl-THF synthesis could be formed reductively from CO₂ in the presence of H₂. Lastly, it is possible that glycine may be also involved in other unknown mechanisms promoting *C. aurantiacus* photoheterotrophic growth. For example, a previous report has discovered the marine bacterium *Pelagibacter ubique* possesses all the genes for amino acid biosynthesis, but is still effectively auxotrophic for glycine (Tripp et al. 2009).

An interesting fact is that glycylglycine, a dipeptide of glycine, has been employed in the earliest studies of *C. aurantiacus* (Madigan et al. 1974), which gave the best growth of *C. aurantiacus* compared to other buffers (e.g., Tris, phosphate, and MOPS). Notably, glycylglycine is known to be a good buffer for biological systems since it is relatively non-toxic (Smith and Smith 1949). In the absence of H₂, *C. aurantiacus* grew much better in glycylglycine than in MOPS or other buffers (Figure 7-S3). Glycylglycine could be hydrolyzed at high temperature during medium autoclave and cell incubations (Radzicka and Wolfenden 1996), or degraded by a membrane dipeptidase (Caur_2632). Our isotopic analysis further confirmed that *C. aurantiacus* growing on 100% [1-¹³C] acetate in the presence of unlabeled glycylglycine buffer possessed significantly unlabeled proteinogenic glycine (i.e., ~90% of proteinogenic glycine was completely unlabeled, while only 3% of proteinogenic glutamate was completely unlabeled). Since the culture was grown in minimal
medium containing $^{13}$C-acetate as the sole carbon source, the unlabeled glycine in the biomass must come from unlabeled glycyglycine. All these evidences imply that glycyglycine could be considered as an exogenous source of glycine contributing to growth enhancement for C. aurantiacus.

7.5. Conclusions

In this study, we traced glycine in the central carbon metabolism to answer how it can enhance photoheterotrophic growth of C. aurantiacus in the absence of H$_2$. Our results, together with previous studies and genome annotations, indicate that glycine can be used for producing biomass (mainly glycine and serine), but more importantly glycine molecules are degraded via cleavage reactions, serving as an important route for NAD(P)H production for acetate-grown C. aurantiacus cultures. As it is widely known, glycine is the simplest and also the most abundant amino acid that can be synthesized abiotically on the primitive Earth (Miller 1953). Glycine cleavage and C1 metabolism might be the ancestral energy generation pathways (Braakman and Smith 2012).

7.6. Acknowledgements

We would like to thank Professor Robert Blankenship at Washington University in St. Louis for providing Chloroflexus aurantiacus strain J-10-fl and Professor Madigan at Southern Illinois University for providing Chloroflexus aurantiacus strain OK-70-fl. We would also like to acknowledge Jeremy King from Blankenship group for his kind help on our project. To finish this project, Dr. Joseph Tang, Dr. Le You, Yaya Wang, and Yadana Khin have provided direct and valuable contributions to this work. Finally, we thank the funding support from the NASA astrobiology program (NNX12AD85G) and NSF (DBI 1356669).
7.7. References


Figure 7.1: Investigation of the key nutrients enhancing the photoheterotrophic growth of strain J-10-fl. The medium contained 2 g/L [1-^{13}C] sodium acetate and 0.2 g/L yeast extract. Unlabeled fractions (M0) of proteinogenic amino acids were mainly derived from yeast extract.
Figure 7.2: Strain J-10-fl growth on sodium acetate (NaOAc, 2 g/L) in a minimal medium with or without glycine (0.05 g/L) addition under photoheterotrophic conditions. NaHCO$_3$ (0.5 g/L) was also supplemented to investigate its influence on the bacterial growth. Error bars represent the standard deviations of two biological replicates.
Figure 7.3: Mass isotopomer distributions (MIDs) of selected proteinogenic amino acids ([M-57]+) under photoheterotrophic condition. Strain J-10-fl was grown in a minimal medium supplied with [1,2-\textsuperscript{13}C\textsubscript{2}] acetate (2 g/L) and unlabeled glycine (0.05 g/L). Error bars represent the standard deviations of two biological replicates.
Figure 7.4: Investigation of mass isotopomer distributions of proteinogenic amino acids to reveal CO₂ fixation under photoheterotrophic conditions. Panels (A-C) show the MIDs of alanine, glycine and serine, respectively. Panel (D) displays the comparison of MIDs between Met [M-57]$^+$ and Asp [M-57]$^+$. For Panels (A-D), strain J-10-fl grew in a minimal medium supplied with unlabeled acetate (2 g/L), unlabeled glycine (0.05 g/L) and NaH$^{13}$CO₃ (0.5 g/L). Panel (E) shows the total $^{13}$C enrichments of selected amino acids when J-10-fl was grown using [1,2-$^{13}$C₂] acetate (2 g/L) and unlabeled glycine (0.05 g/L) with or without NaH$^{13}$CO₃ (0.5 g/L) addition. Error bars represent the standard deviations of two biological replicates.
Figure 7.5: Schematic representation of C1 metabolism and CO₂ fixation pathways in *C. aurantiacus*. (A) C1 metabolism of *C. aurantiacus*. Black circles in panel (A) represent carbon atoms that will be donated to or originate from C1 unit carbon atoms. Dashed lines represent the genes that are not annotated in strain J-10-fl. Note: *C. aurantiacus* lacks the annotation of formate dehydrogenase gene, but it may use alternative enzyme (formylmethanofuran dehydrogenase, Caur_0027) for formate oxidation (BERTRAM et al. 1994). (B) Annotated CO₂ fixation pathways in *C. aurantiacus*. The secondary CO₂ fixation route (3HOP pathway) can generate acetyl-CoA of different labeling patterns. Black circles in panel (B) represent ¹³C₀₂ or H¹³CO₃⁻. In both panels, the numbers in the circles represent the positions of carbon atoms in corresponding intracellular metabolites.
Supplementary Materials

Table 7-S1: Mass isotopomer distributions of proteinogenic amino acids of *Chloroflexus aurantiacus* J-10-fl strain grown under photoheterotrophic conditions.

<table>
<thead>
<tr>
<th>Carbon Sources</th>
<th>[1,2-(^{13})C] sodium acetate + glycine</th>
<th>[1,2-(^{13})C] sodium acetate + glycine + NaH(^{13})CO(_3)</th>
<th>sodium acetate + glycine + NaH(^{13})CO(_3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ion Fragments</td>
<td>Fractions</td>
<td>Standard Deviations</td>
<td>Fractions</td>
</tr>
<tr>
<td>Alanine [M-57](^{1})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.04</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.20</td>
<td>0.01</td>
<td>0.10</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.74</td>
<td>0.02</td>
<td>0.87</td>
</tr>
<tr>
<td>'Glycine [M-57](^{1})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.68</td>
<td>0.01</td>
<td>0.56</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.07</td>
<td>0.00</td>
<td>0.14</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.24</td>
<td>0.01</td>
<td>0.29</td>
</tr>
<tr>
<td>'Glycine [M-85](^{1})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.72</td>
<td>0.01</td>
<td>0.68</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.28</td>
<td>0.01</td>
<td>0.32</td>
</tr>
<tr>
<td>'Serine [M-57](^{1})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.26</td>
<td>0.00</td>
<td>0.17</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.12</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.18</td>
<td>0.00</td>
<td>0.13</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.45</td>
<td>0.00</td>
<td>0.60</td>
</tr>
<tr>
<td>'Serine [M-159](^{1})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.27</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.21</td>
<td>0.00</td>
<td>0.16</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.51</td>
<td>0.00</td>
<td>0.64</td>
</tr>
<tr>
<td>'Aspartate [M-57](^{1})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.08</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.22</td>
<td>0.02</td>
<td>0.14</td>
</tr>
<tr>
<td>'M+4'</td>
<td>0.68</td>
<td>0.03</td>
<td>0.81</td>
</tr>
<tr>
<td>Aspartate f302</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.04</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.15</td>
<td>0.01</td>
<td>0.09</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.81</td>
<td>0.01</td>
<td>0.89</td>
</tr>
<tr>
<td>'Glutamate [M-57](^{1})</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.06</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>'M+4'</td>
<td>0.20</td>
<td>0.01</td>
<td>0.12</td>
</tr>
<tr>
<td>'M+5'</td>
<td>0.72</td>
<td>0.02</td>
<td>0.84</td>
</tr>
<tr>
<td>Carbon Sources</td>
<td>[1,2-13C] sodium acetate + glycine</td>
<td>[1,2-13C] sodium acetate + NaH13CO3</td>
<td>sodium acetate + glycine + NaH13CO3</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------------------------------</td>
<td>------------------------------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td>Ion Fragments</td>
<td>Fractions</td>
<td>Standard Deviations</td>
<td>Fractions</td>
</tr>
<tr>
<td>'Histidine [M-57]'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.07</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>'M+4'</td>
<td>0.24</td>
<td>0.02</td>
<td>0.14</td>
</tr>
<tr>
<td>'M+5'</td>
<td>0.51</td>
<td>0.02</td>
<td>0.62</td>
</tr>
<tr>
<td>'M+6'</td>
<td>0.14</td>
<td>0.02</td>
<td>0.18</td>
</tr>
<tr>
<td>'Isoleucine [M-57]'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.04</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.07</td>
<td>0.00</td>
<td>0.03</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.18</td>
<td>0.01</td>
<td>0.11</td>
</tr>
<tr>
<td>'M+4'</td>
<td>0.53</td>
<td>0.01</td>
<td>0.61</td>
</tr>
<tr>
<td>'M+5'</td>
<td>0.17</td>
<td>0.03</td>
<td>0.23</td>
</tr>
<tr>
<td>'Valine [M-57]'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.07</td>
<td>0.01</td>
<td>0.04</td>
</tr>
<tr>
<td>'M+4'</td>
<td>0.25</td>
<td>0.01</td>
<td>0.15</td>
</tr>
<tr>
<td>'M+5'</td>
<td>0.64</td>
<td>0.02</td>
<td>0.79</td>
</tr>
<tr>
<td>'Leucine [M-15]'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+4'</td>
<td>0.09</td>
<td>0.02</td>
<td>0.05</td>
</tr>
<tr>
<td>'M+5'</td>
<td>0.25</td>
<td>0.01</td>
<td>0.20</td>
</tr>
<tr>
<td>'M+6'</td>
<td>0.60</td>
<td>0.05</td>
<td>0.72</td>
</tr>
<tr>
<td>'Isoleucine [M-15]'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.02</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.04</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>'M+4'</td>
<td>0.12</td>
<td>0.02</td>
<td>0.07</td>
</tr>
<tr>
<td>'M+5'</td>
<td>0.26</td>
<td>0.00</td>
<td>0.20</td>
</tr>
<tr>
<td>'M+6'</td>
<td>0.55</td>
<td>0.03</td>
<td>0.69</td>
</tr>
<tr>
<td>'Phenylalanine [M-57]'</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>'M+0'</td>
<td>0.02</td>
<td>0.00</td>
<td>0.02</td>
</tr>
<tr>
<td>'M+1'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+2'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+3'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>'M+4'</td>
<td>0.01</td>
<td>0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+5'</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td>'M+6'</td>
<td>0.07</td>
<td>0.02</td>
<td>0.03</td>
</tr>
<tr>
<td>'M+7'</td>
<td>0.16</td>
<td>0.02</td>
<td>0.09</td>
</tr>
<tr>
<td>'M+8'</td>
<td>0.32</td>
<td>0.00</td>
<td>0.25</td>
</tr>
<tr>
<td>'M+9'</td>
<td>0.37</td>
<td>0.05</td>
<td>0.59</td>
</tr>
</tbody>
</table>

Note: the standard deviations are based on biological replicates.
Table 7-S2: Estimation of fractions of proteinogenic alanine in different labelling patterns.

<table>
<thead>
<tr>
<th>Labeling patterns</th>
<th>variables</th>
<th>Fractions</th>
</tr>
</thead>
<tbody>
<tr>
<td>123</td>
<td>X1</td>
<td>0.55</td>
</tr>
<tr>
<td>123</td>
<td>X2</td>
<td>0.31</td>
</tr>
<tr>
<td>123</td>
<td>X3</td>
<td>0.09</td>
</tr>
<tr>
<td>123</td>
<td>X4</td>
<td>0.00</td>
</tr>
<tr>
<td>123</td>
<td>X5</td>
<td>0.05</td>
</tr>
<tr>
<td>123</td>
<td>X6</td>
<td>0.00</td>
</tr>
<tr>
<td>123</td>
<td>X7</td>
<td>0.01</td>
</tr>
<tr>
<td>123</td>
<td>X8</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Figure 7-S1: Photoheterotrophic growth of J-10-fl strain in acetate-based medium with different concentrations of glycine (A) and serine supply (B).
Figure 7-S2: Photoheterotrophic consumption of glycine by strain J-10-fl (with 0.05 g/L glycine supplement) (n=2, 10-day cultures). Using GC/MS and TBDMS method (See Method and Materials), we measured glycine concentrations in the culture. A standard curve was first made by determining the relationship between the MS abundance and glycine concentration in medium (R²>0.99, Figure 7-S2 A-B). We then used these curves to estimate the concentration of the remaining glycine in medium (Figure 7-S2 C). A brief calculation shows that ~0.3 g/L of biomass would be produced within 10 days (Figure 7-2), which required ~7.5mg/L of glycine. However, ~40 mg/L has been absorbed from the medium, indicating much glycine is not used for biomass synthesis.
Figure 7-S3: The growth of *C. aurantiacus* OK-70-fl in the acetate medium (without H₂). OK-70-fl grew under photoheterotrophic (light condition without O₂) and chemoheterotrophic conditions (dark condition with O₂). 20 mM acetate with or without 2 mM glycine was included in a minimal medium containing (Fig. A) glycylglycine or (Fig. B) MOPS buffer. Error bars represent the standard deviations of at least two biological replicates.
Chapter 8: Summary and future directions
8.1. Summary

In this dissertation, some basic principles of flux balance analysis and $^{13}$C metabolic flux analysis are first presented. In the following chapters, $^{13}$C MFA was applied to different bacterial species in order to investigate their central metabolic responses to genetic modifications or various growth conditions. Although this technique may not be suitable to all the non-model species (for example, *Chloroflexus aurantiacus*, which is shown in Chapter 7), $^{13}$C-based analysis can also allow us to investigate the bottleneck of bacterial growth and help optimize its medium. Further, we used multiple modeling approaches, including genome-scale FBA, kinetics, and cell movement model, to understand the cyanobacterial performance in heterogeneous bioreactors. In the following paragraphs, I will focus on some key information we have learned after applying our flux techniques.

8.1.1. Flux analysis reveals both carbon and energy metabolisms

Except quantifying carbon flows in cell metabolism, flux analysis can also be used to calculate the energy flows throughout the metabolic network. Therefore, by applying flux analysis techniques, we are able to quantify how much energy cells can obtain from the central metabolism and how much they need from other sources, such as oxidation phosphorylation and photosystems.

In Chapter 3, we have used $^{13}$C MFA to quantify the consumption and production of energy molecules in the central metabolism (He et al. 2014). For instance, as Table 3.2 shows, the total amount of NADPH is not enough for both biomass and fatty-acid synthesis at the same time, and the majority of NADPH actually comes from transhydrogenase reaction. In addition, the central carbon metabolism contributes partially to ATP synthesis. To support fatty acid production, more ATP must be synthesized from the oxidative phosphorylation. This finding suggests that
production of high levels of fatty acids relies on an efficient metabolism in oxidative phosphorylation. In either laboratory or industrial settings, oxygen may not be efficiently delivered to cells in media. Hence, the overall fatty acid production is likely to be limited by ATP deficiency. To prove this point of view, people in our lab have experimentally demonstrated that, by overexpressing VHb in *E. coli*, which enhances the efficiency of ATP production under O\textsubscript{2}-limited conditions (Kallio et al. 1994), the overall fatty acid production is improved (data not shown).

Constructing a microbial cell factory is not always straightforward, and many factors may affect the ultimate production of the desirable chemicals. For most of the current research studies, people focus more on driving the carbon flows from substrate towards chemical synthesis. However, a deficient energy metabolism could limit the ultimate productivity. In Appendix Chapter 4, we have more discussions on this issue.

**8.1.2. Cyanobacteria feature high activities of Calvin cycle and low activities of TCA cycle**

In Chapter 4, I have shown the cyanobacterial flux phenotypes under various growth conditions, which feature high fluxes through the Calvin cycle and low ones through the TCA cycle. These phenomena are not surprising, since the Calvin cycle is the major route to direct extracellular carbon sources towards synthesis of intracellular building blocks, and photosynthesis can provide enough ATP and reducing equivalents for cyanobacteria. However, this general flux distribution pattern in cyanobacteria has some implications: 1) the low activities of the TCA cycle can be a crucial limiting factor for producing bio-product originating from the TCA cycle itself; and 2) on the contrary, the cyanobacteria can naturally support a higher productivity of bioproducts whose precursors come from the Calvin cycle. A further analysis of metabolite pool sizes in cyanobacteria (data not shown) further explains those phenomena and confirms our hypotheses: Compared to
metabolite pool sizes in *E. coli*, cyanobacteria show much lower abundancies of intermediates in the TCA cycle, but higher accumulations of intermediates in the Calvin cycle. As is well known, low concentrations of reactants adversely affect the ultimate reaction rates, even if the enzymatic activities are high. Therefore, TCA cycle in cyanobacteria is not an ideal pathway for producing desired chemicals in large quantities.

**8.2. Recommended future work**

Two future directions are recommended here to strengthen what have been achieved in this dissertation. One is to develop genome-scale $^{13}$C-MFA, and the other one is to improve our integrated genome-scale metabolic model.

### 8.2.1. Genome-scale $^{13}$C-MFA

Since conventional $^{13}$C-MFA depends on the labeling patterns of proteinogenic amino acids, most of $^{13}$C-MFA is focused on central carbon metabolism in which the free metabolites are precursors to those proteinogenic amino acids. To expand the capacity of $^{13}$C-MFA, it seems necessary to scale up $^{13}$C MFA. Some efforts have been made to extend the metabolic network from the central metabolism to a bigger one consisting of more than 700 reactions (Blank et al. 2005). With the coverage of the entire genome, researchers were able to reveal the flux redistributions in mutants and evaluate the mechanism sustaining the metabolic network robustness. Therefore, the genome-scale $^{13}$C-MFA can be more informative than traditional $^{13}$C-MFA in terms of providing guidance on cell engineering. Additionally, the genome-scale $^{13}$C-MFA can give more information on energy production and expenditure in cells, allowing us to pinpoint the bottleneck for cell growth or product formation.
To enhance the accuracy of genome-scale $^{13}\text{C}$-MFA, it may be ideal to examine the labeling patterns of metabolites or macromolecules (e.g. fatty acids, nucleotide bases, etc.) in addition to proteinogenic amino acids, since the latter may not offer enough information on fluxes in the peripheral metabolism. In addition, a recent study (Gopalakrishnan and Maranas 2015) also points out that the accuracy of biomass composition measurements is of great importance to genome-scale $^{13}\text{C}$-MFA, since over 60% of reactions in the metabolic network are growth coupled only. If they are not properly measured, the majority of flux calculations will be compromised.

In the aspect of modeling, one can extend the biochemical network in $^{13}\text{C}$-MFA model and define the carbon atom transitions from one molecule to another in each reaction (Gopalakrishnan and Maranas 2015). Another fashion of performing genome-scale $^{13}\text{C}$-MFA is presented in a recent study, in which $^{13}\text{C}$ labeling information and measured inflow/outflow fluxes are used to constrain the genome-scale model (Martín et al. 2015).

### 8.2.2. Including hydrodynamic information into integrated genome-scale model to accurately illustrate cell metabolism in subpopulations

In Chapter 6, we have shown our first attempt to integrate multiple modeling approaches to simulate cyanobacterial performance in photobioreactors. The results yielded both flux changes and growth physiologies. In the model, we used a simplified manner (i.e., a cosine function) to generalize cell trajectories in photobioreactors among different subpopulations, which is reasonably good to predict the culture performance as a whole. However, to accurately describe the cell metabolism of a particular subpopulation, it may be better to incorporate hydrodynamic information that gives accurate predictions of cell trajectories and nutrient concentration distributions in photobioreactors. Computational fluid dynamics is thus required in this
circumstance, which can generate temporal and spatial information of parameters of interest (Luo and Al-Dahhan 2011).

8.3. References


Appendix Chapter 1: $^{13}$C-MFA delineates the photomixotrophic metabolism of *Synechocystis sp.* PCC 6803 under light- and carbon-sufficient conditions


The central carbon metabolism of cyanobacteria is under debate. For over 50 years, the lack of \( \alpha \)-ketoglutarate dehydrogenase has led to the belief that cyanobacteria have an incomplete TCA cycle. Recent in vitro enzymatic experiments suggest that this cycle may in fact be closed. The current study employed \(^{13}\)C isotopomers to delineate pathways in the cyanobacterium *Synechocystis* sp. PCC 6803. By tracing the incorporation of supplemented glutamate into the downstream metabolites in the TCA cycle, we observed a direct in vivo transformation of \( \alpha \)-ketoglutarate to succinate. Additionally, isotopic tracing of glyoxylate did not show a functional glyoxylate shunt and glyoxylate was used for glycine synthesis. The photomixotrophic carbon metabolism was then profiled with \(^{13}\)C-MFA under light and carbon-sufficient conditions. We observed that: (i) the in vivo flux through the TCA cycle reactions (\( \alpha \)-ketoglutarate \( \rightarrow \) succinate) was minimal (<2%); (ii) the flux ratio of CO\(_2\) fixation was six times higher than that of glucose utilization; (iii) the relative flux through the oxidative pentose phosphate pathway was low (<2%); (iv) high flux through malic enzyme served as a main route for pyruvate synthesis. Our results improve the understanding of the versatile metabolism in cyanobacteria and shed light on their application for photo-biorefineries.

**Keywords:** Free metabolites · Glyoxylate shunt · Malic enzyme · Pentose phosphate pathway · TCA cycle

1 Introduction

*Synechocystis* sp. PCC 6803 (hereafter *Synechocystis* 6803) is a naturally transformable cyanobacterium [1] and a model organism for studying photosynthesis [2]. *Synechocystis* 6803 and other cyanobacterial species are promising phototrophic cell factories for synthesis of valuable chemicals and biofuels [3–7]. To explore cyanobacterial metabolism for biotechnology applications, genomics and transcriptomics approaches have been used to study *Synechocystis* 6803 [8]. Complementing these approaches, fluxomics tools (flux balance analysis (FBA) and \(^{13}\)C-metabolic flux analysis (\(^{13}\)C-MFA)) are also powerful in deciphering genome functions and unraveling cell phenotype in phototrophs under photoautotrophic, photoheterotrophic and chemoheterotrophic conditions.

---

"These authors contributed equally to this work."
tomixotrophic, and heterotrophic metabolisms [9-19]. These multi-omics studies have improved our understanding and application of photo-biorefineries [20].

Nevertheless, cyanobacterial metabolism is still not completely resolved. Due to the lack of α-ketoglutarate dehydrogenase, cyanobacteria were thought to have an incomplete tricarboxylic acid (TCA) cycle [21, 22]. This assumption has been employed in most cyanobacterial models so far [11, 16, 17]. Recently, a pair of enzymes from Synechococcus sp. PCC 7002, α-ketoglutarate decarboxylase and succinic semialdehyde dehydrogenase, were found to transform α-ketoglutarate into succinate in vitro [23]. These two enzymes have homologues throughout cyanobacteria. Contemporaneously, Nogales et al. [18] identified an overlapping GABA (γ-amino butyric acid) shunt in silico that could also complete the TCA cycle via GABA and succinic semialdehyde. Such a pathway in cyanobacteria would help explain previous observations that α-ketoglutarate added to cultures of Synechocystis double mutant strain (knockout of succinate dehydrogenase and fumarate reductase) led to accumulation of succinate [24]. However, an in vivo flux from α-ketoglutarate to succinate has not been measured.

Another open question has been whether the glyoxylate shunt was active or ever existed. Glyoxylate shunt activity in some cyanobacterial species was reported [25, 26] and thus were included in the metabolic models of Synechocystis 6803 [11, 17, 27, 28]. But homologues encoding isocitrate lyase and malate synthase have still not been found. Moreover, the oxidative pentose phosphate pathway (OPP pathway) in Synechocystis 6803 has been considered inactive under light conditions [17], but this pathway was recently proved to be highly active in photoautotrophic metabolism [16].

Considering all these controversial conclusions, we were intrigued to re-delineate the photomixotrophic metabolism in Synechocystis 6803. Since the previous application of 13C-MFA to photomixotrophic metabolism was operated in a Synechocystis 6803 culture lacking carbonate and atmospheric CO2 (i.e. CO2-limiting condition) [17], we were motivated to unravel the central metabolism and reconstruct the metabolic network in Synechocystis 6803 under CO2-sufficient conditions. In addition, the application of 13C-MFA requires the attainment of a metabolic steady state. However, in a photobioreactor, cyanobacteria are continuously moving between “light” (near surface of bioreactor) and “dark” (depending on mixing, cell density, and reactor size/geometry) zones. In the light zone, cells fix CO2 and accumulate glycogen, while they may use this storage component (or glucose in the medium) for “heterotrophic” metabolism in the dark zone.

To minimize such heterogeneous growth, this study has performed 13C-MFA experiments using small culture volume (<50 mL) and low biomass density (OD680<0.5) to ensure cell metabolism under light and carbon-sufficient conditions. This approach may provide a better understanding of photomixotrophic metabolism in Synechocystis 6803.

2 Materials and methods

2.1 Photomixotrophic culture

Synechocystis 6803 cultures were grown in a modified BG-11 medium containing no sources of 13C. Ferric ammonium citrate was replaced with ferric ammonium sulfate [29]. 13C was supplied as ~2 g/L NaH13CO3 and 5 g/L glucose (U-13C, or D-13C). The purity of 13C-substrates was >98% (Cambridge Isotope Laboratories, Tewksbury, MA, USA). Inocula from an unlabeled Synechocystis 6803 photautotrophic culture (OD680 = 0.9) was added into 30 mL 13C-labeled medium in 100 mL serum bottles, which were then sealed with rubber septa to prevent atmospheric CO2 intrusion.

All cultures were started with only a 0.5% inoculation volume to minimize the inoculation effect. Cell growth was under continuous illumination (~50 μmol photons/m2/s) on a shaker at 150 rpm at 30°C. Cell density was monitored by a UV-Vis spectrophotometer (GENESYS, Thermo Scientific) at 730 nm. The conversion ratio between OD680 and dry biomass weight was 1 unit OD680 = 0.45 g dry cell weight/L. Total organic carbon analyzer (inorganic carbon measurement mode) with non-dispersive infrared detector (Shimadzu Corporation, Japan) was used to determine sodium bicarbonate concentration in the culture supernatant. Enzyme kits (R-Biopharm, Darmstadt, Germany) were used to measure the glucose concentrations in the culture.

2.2 Isotopic dilution experiments

Isotopic dilution experiments were employed to identify the presence of certain pathways in vivo. To investigate the structure of the TCA cycle, we used glutamate (instead of α-ketoglutarate) as the tracer since cyanobacteria exhibited very low capability to uptake α-ketoglutarate [30]. To examine the presence of the glyoxylate shunt, we used glyoxylate as the tracer. Specifically, unlabeled glutamate (10 mM) or unlabeled glyoxylate (15 mM) was added into 13C-labeled cultures (grown on NaH13CO3 and U-13C glucose) during the exponential growth phase (OD680 = 0.4). After 30 min of incubation with a respective tracer, samples from two biological replicates were harvested and free metabolites were extracted. To identify whether Synechocystis 6803 used glutamate or glyoxylate for biomass synthesis, Synechocystis 6803 cultures were grown with a respective unlabeled tracer (10 mM glutamate or 15 mM glyoxylate), NaH13CO3, and U-13C glucose for 48 h (OD680 reached 0.4). Samples from two biological replicates were then collected to analyze the 13C incorporation into proteinogenic amino acids.
2.3 Metabolite extraction and GC-MS analysis

Isotopomer measurements of free metabolites (TMS-based method) and proteinogenic amino acids (TBDMS-based method) are based on previous reports [31-33]. GC-MS analysis had three technical replicates per biological sample.

Intracellular free metabolites were used to qualitatively characterize functional pathways. Supporting information, Fig. S1A–C illustrate the molecular structure of TMS-derivatized amino acids, succinate, and α-ketoglutarate used for analysis. The fragment [M-15]⁺, minus a methyl group from the TMS group, includes the labeling information of the entire molecule. The [M-16]⁺ fragment, together with [M-43]⁺ or [M-117]⁺ (minus the α-carboxyl group from a metabolite), was used for GC–MS analysis.

Proteinogenic amino acids were used to determine the function and quantify the metabolic fluxes. The mass fragments of ten key amino acids provided sufficient constraints for flux calculations [34–36]. The fragments (M-67)⁺, (M-159)⁺ or (M-85)⁺, and (M-202) were used for flux analysis [37]. In addition, because of overlap peaks and product degradations, several amino acids (proline, arginine, cysteine, and tryptophan) were not analyzed. The isotopic labeling data are shown as mass fractures, i.e. M0, M1, M2, etc., representing fragments containing unlabeled, singly labeled, and doubly labeled metabolites.

2.4 13C-metabolic flux analysis

13C-MFA was used to quantify in vivo fluxes through the central metabolic network in Synechocystis 6803. Photomixotrophic cultures were grown on 1-13C glucose and NaH13CO3. Biomass was collected during the exponential growth phase for proteinogenic amino acids analysis. The metabolic network of Synechocystis 6803 was reconstructed based on tracer experiments and previous reports [16, 38–41] that included glycolysis, the Calvin cycle, complete TCA cycle, glyoxylate shunt, and photorespiration pathways (Supporting information, Table S1). In our 13C-MFA, relative metabolic fluxes through the central metabolism were profiled with the assumption that the Calvin cycle flux from Ru5P to RuBP was 100. The minimization of a quadratic function that calculated the difference between predicted and measured isotopomer patterns solved the relative metabolic fluxes (Supporting information, Table S2). The biomass composition (Supporting information, Table S1) was based on a previous report [9]. Reaction reversibility was characterized by the exchange coefficient, defined as νexch = β(exch/1-exch), where νexch was the exchange flux and β was the exchange constant [42]. In this study, β was equal to 500 and exch ranged from 0 to 1. The forward flux (νfoward) and backward flux (νbackward) in the model were transformed from the νexch and the net flux, νnet, using following equation [43]:

\[
\begin{bmatrix}
ν_{\text{forward}} \\
ν_{\text{backward}}
\end{bmatrix} = \begin{bmatrix}
ν_{\text{exch}} \\
ν_{\text{exch}}
\end{bmatrix} \cdot \begin{bmatrix}
\min(-ν_{\text{net}}, 0) \\
\min(ν_{\text{net}}, 0)
\end{bmatrix}
\]

The optimization for 13C-MFA was performed as follows:

\[
\min (M_{\text{exp}} - M_{\text{sim}}(v))^T (M_{\text{exp}} - M_{\text{sim}}(v))
\]

s.t. \( v \in \{lb, ub\} \)

\[
S \cdot v = 0,
\]

\[
A_1 \cdot X_i = B_1, Y_1 (i=1, 2, 3, 4, 5)
\]

Equation 1 is the quadratic error function that was optimized and \( M_{\text{exp}} \) is the vector of experimentally measured labeling patterns of amino acids. \( M_{\text{sim}} \) is the counterpart of the simulated data as a function of fluxes. \( ν \) is the flux vector that is to be determined. Equation 2 gives the boundary conditions of the flux variables. Equation 3 represents the metabolite balances. Equation 4 represents the elementary metabolic unit (EMU) balance, where \( X_1 \) and \( Y_1 \) represent the unknown and known EMU variables of size i, respectively, and \( A_1 \) and \( B_1 \) are matrices of linear functions of the fluxes [44, 45].

The MATLAB optimization solver “fmincon” was employed to minimize the quadratic error function. To avoid local minima, 100 initial guesses were randomly generated, and the solution set that minimized the objective function was used as the best fit. The Monte Carlo method was used to calculate 95% confidence intervals [46]. The measured isotopomer data was perturbed 1000 times with normally distributed noise within measurement error, and the optimization solver was restarted with the optimal solution. The determination of confidence intervals of the fluxes (95%) was based on 1000 simulations, and confidence intervals were used to calculate standard deviations.

3 Results

3.1 Photomixotrophic biomass growth and metabolic pseudo-steady state

Figure 1 shows the growth curves in serum bottles and shake flasks. Cell doubling times were similar in both containers. The similarity of growth indicates that O2 accumulation in the serum bottle headspace had minimal effect on photomixotrophic growth. During the early growth phase, the specific growth rate in the early exponential phase was 0.079/h. After cultivation in serum bottles for 75 h, cell growth slowed down and the culture pH rose from 9 to 10.

To determine a pseudo-steady state metabolic period for 13C-MFA, biomass samples from serum bottles were
3.2 $^{13}$C-based pathway investigation

Based on isotopic dilution of downstream metabolites after incubation with unlabeled precursors, in vivo enzyme functions were investigated in tracer experiments. Prior to $^{13}$C-glutamate pulse treatment, $\alpha$-ketoglutarate (Fig. 2A), succinate (Fig. 2B), and malate (Fig. 2C) were nearly fully labeled (M5 for $\alpha$-ketoglutarate; M4 for succinate and malate) in $^{13}$C-labeled cultures. After 30-min incubation with unlabeled glutamate, $^{13}$C carbon from glutamate was incorporated into the downstream metabolites of $\alpha$-ketoglutarate. $^{13}$C abundance increased to over 65% in succinate, 90% in $\alpha$-ketoglutarate, and 30% in malate. Mass spectra of these metabolites before and after glutamate addition are shown in Supporting information, Fig. S1. After incubation with unlabeled glutamate for 48 h, all amino acids, except glutamate, from biomass protein remained fully $^{13}$C labeled (Fig. 3 and Supporting information Fig. S2).

Labeled cultures pulsed with unlabeled glyoxylate showed a shift from fully $^{12}$C to $^{13}$C in free glycine and glyoxylate (M2 to M0, Fig. 4A). However, no significant shift was observed in succinate and $\alpha$-ketoglutarate, both of which are downstream metabolites of malate. After the labeled culture was incubated with unlabeled glyoxylate for 48 h, $^{13}$C was only incorporated into proteinogenic glycine (Fig. 4B), while other proteinogenic amino acids remained highly labeled.
3.3 Flux analysis results

$^{13}$C-MFA results are sensitive to model network construction, the labeling patterns of substrates, and the completeness of isotope data. In this study, isotopic dilution experiments were employed to reveal the metabolite network in Synechocystis 6803. Singly labeled glucose and fully labeled bicarbonate were used to generate unique isotope data in amino acids. Via EMU simulations and isotope tracer information from different MS fragments, $^{13}$C-MFA profiled the photomixotrophic metabolism under light and carbon-sufficient conditions. Relative flux distributions, exchange coefficients for reversible reactions, and 95% confidence intervals are shown in Fig. S5 and Supporting information, Table S1. The simulated fluxes fit the isotope data well ($r^2>0.99$, Supporting information, Fig. S3).

$^{13}$C-MFA indicated that Synechocystis 6803 had a high CO$_2$ fixation flux through the Calvin cycle (~100) than the glucose uptake flux (~18) in the early photomixotrophic growth phase. Consistent with this observation, <0.1 g/L glucose was consumed during early growth phase. In contrast, previous $^{13}$C-MFA of photomixotrophic metabolism in Synechocystis 6803 under CO$_2$ limiting conditions found the glucose uptake flux to be ~50 [17]. In another study, when cell culture was dense (OD$_{600}$ up to 20), Synechocystis 6803 utilized significantly more glucose than CO$_2$ for its growth [47]. Therefore, CO$_2$ and light conditions can significantly affect the photomixotrophic metabolism in Synechocystis 6803.

Under photomixotrophic conditions with sufficient light and carbohydrate sources, the flux from CO$_2$ to succinate was not significant (<2% of total CO$_2$ uptake). Most of the flux from α-ketoglutarate went to glutamate biosynthesis. The glyoxylate shunt also did not show a measurable flux (<0.1% of total CO$_2$ uptake). Additionally, the OPP pathway showed a measurable flux (1.9% of total CO$_2$ uptake), which played a minor role in C5 carbon synthesis and NADPH production. The flux through photorespiration, however, was limited to 0.1% of total CO$_2$ uptake. Although the confidence intervals (Supporting information, Table S1) of these anaplerotic reactions (PEP + CO$_2$ → OAA; MAL → CO$_2$ + PYR) were larger than those of other fluxes, malic enzyme still showed significant flux and was the main route for pyruvate synthesis.
4 Discussion

4.1 TCA cycle metabolism

Cyanobacteria are prokaryotes responsible for the conversion of the early atmosphere into our current oxygen-rich atmosphere [48]. Primitive anaerobic prokaryotes developed two separate TCA pathways: the reductive branch (oxaloacetate to succinate) and the oxidative branch (citrate to α-ketoglutarate) (Fig. 2D(ii)). Some anaerobic bacteria, such as Clostridium acetobutylicum, use a bifurcated TCA cycle that terminates at succinate (Fig. 2D(ii)). As atmospheric oxygen levels rose, the two branches linked to complete the TCA cycle. For example, the TCA cycle in facultative anaerobes (e.g., E. coli) can be complete if oxygen is present (Fig. 2D(iii)). A phototrophic bacterium, Chlorobaculum tepidum, employs a reverse TCA cycle [49] (Fig. 2D(iv)).

In our study, the labeling patterns of free metabolites indicated that the pathway for converting α-ketoglutarate to succinate can be complete under glutamate addition conditions. Significant amounts of unlabeled α-ketoglutarate, succinate, and malate (Fig. 2A–C) were observed after unlabeled glutamate was added into 13C-labeled cultures. Since α-ketoglutarate dehydrogenase activity has never been shown to exist in cyanobacteria, we presume that this conversion was accomplished by a newly discovered pathway through succinic semialdehyde [10, 18, 23]. On the other hand, key proteinogenic amino acids, e.g., aspartate (derived from oxaloacetate), alanine (derived from pyruvate), and serine (derived from 3-phosphoglycerate), had very little 13C incorporation (<5%) from glutamate after 2-day incubation with unlabeled glutamate (Fig. 3). These results qualitatively indicated that the flux from α-ketoglutarate towards the complete TCA cycle was very small compared to other fluxes (e.g., fluxes through glycolysis and the Calvin cycle). The low conversion from α-ketoglutarate to its TCA cycle downstream metabolites was also observed in a Synechococcus elongatus PCC 7942 mutant (with an engineered α-ketoglutarate permease), in which α-ketoglutarate was mainly converted into glutamate and glutamine instead of TCA cycle downstream metabolites [30].

Although our 13C-study cannot distinguish whether the conversion of α-ketoglutarate to succinate was via α-ketoglutarate decarboxylase or the GABA shunt [19], this reaction may be notable in cyanobacterial metabolism only under certain conditions (e.g., with the presence of large amount of glutamate or α-ketoglutarate). The poor activity of this reaction may also explain why the previous tracer studies did not observe the conversion of α-ketoglutarate to succinate. These studies used an assay of α-ketoglutarate dehydrogenase activity [21], as opposed to the decarboxylase activity that has been more recently observed to convert α-ketoglutarate to succinate [23].
Although many cyanobacterial species appear to have a complete TCA cycle pathway, it may not be adapted to carry a large flux. A recent FBA model indicates that a complete cyanobacterial TCA cycle via AKG dehydrogenase may reduce biomass growth due to the unnecessary metabolic burden for the synthesis of multi-protein enzymes [19]. For organisms that obtain sufficient reducing equivalents from light reactions, the use of a complete TCA cycle to oxidize carbon is unnecessary. Therefore, the complete TCA pathways in *Synechocystis* 6803 may serve only to regenerate intermediates or fine-tune the metabolic balance under certain photomixotrophic conditions (such as the presence of extracellular glutamate).

### 4.2 The glyoxylate shunt

This study also examined the presence of the glyoxylate shunt and determined its function in *Synechocystis* 6803. Previous metabolic models predicted that *Synechocystis* 6803 contains a bacterial-like glyoxylate shunt [17]. However, *Synechocystis* 6803, and nearly all other sequenced cyanobacteria, lack homologues of known genes that encode isocitrate lyase and malate synthase. Some 13C-MFA [16] and FBA [11, 15] studies have also suggested that the glyoxylate shunt in *Synechocystis* 6803 was incomplete under photautotrophic and photomixotrophic conditions. In our tracer experiments with the addition of unlabelled glyoxylate during the exponential phase, we observed the uptake of glyoxylate and its conversion to glycine (Fig. 4A). However, in the proteinogenic amino acids of 13C-cultures grown with 13C-glyoxylate (Fig. 4B), we did not see significant 13C accumulation in proteinogenic amino acids downstream of malate (i.e. the end-product of the glyoxylate shunt), including alanine and aspartate (Fig. 4B). Statistically, 13C-MFA showed that the in vivo flux through the presumed glyoxylate shunt was essentially zero (Fig. 5). This observation of the glyoxylate shunt was supported by a recent enzymatic test using crude extracts of *Synechocystis* 6803 cells, in which no isocitrate lyase activity was detected [19].

### 4.3 Malic enzyme activity

Under continuous light illumination, the malic enzyme is important for optimal *Synechocystis* 6803 growth. This gene (atb0721) is highly expressed under photomixotrophic conditions compared to photautotrophic conditions [8]. Moreover, 13C-MFA revealed significant malic enzyme flux in *Synechocystis* 6803 under photomixotrophic [16] and CO2 limited photomixotrophic cultures [17]. Previous reports indicated that the malic enzyme reaction (malate → pyruvate + CO2 + NADPH) is instrumental in a carbon concentrating mechanism akin to that in C4 plants, and this enzyme may indirectly transport NADPH between different cell locations. In this study, high malic enzyme activity was also observed when the bicarbonate and reduced carbon source were sufficient. In fact, deletion of malic enzyme gene significantly reduces *Synechocystis* 6803 growth under both photoautotrophic and glucose-based photomixotrophic conditions, while the growth can be recovered by providing pyruvate [50]. Thereby, high flux through malic enzyme (–31) is likely to serve as a key route for pyruvate synthesis when pyruvate kinase is inhibited by ATP, a negative allosteric inhibitor under photosynthetic conditions [50].

### 4.4 The oxidative pentose phosphate pathway

The OPP pathway is an important NADPH synthesis route in heterotrophic organisms. Since photosynthetically light reactions produce significant amounts of NADPH, the OPP pathway becomes futile in phototrophic metabolism. A cyanobacterial mutant (Arw31) of *Synechococcus sp.* strain PCC 7942, that lacks the OPP pathway enzymes, exhibited a similar growth rate to the wild-type strain under photomixotrophic conditions [51]. Moreover, glucose in photomixotrophic cultures has been shown to either increase or have a small effect on key OPP pathway enzyme transcriptions [8, 52, 53]. Taken together, these data indicate that the OPP pathway is dispensable under light conditions. On the other hand, the OPP pathway mutant described above exhibited decreased viability under dark incubations [51]. FBA models also predicted that the OPP pathway was active under light Limited conditions [11]. Our experiments measured a low flux (–1.5) through the OPP pathway under early photomixotrophic growth conditions (i.e. light- and carbon-sufficient culture). However, during the late growth phase (Supporting information, Table S2), the biomass showed a higher unlabelled proteinogenic histidine (M0 fraction). Supporting information, Table S2), indicating that more glucose was directed to OPP for ribose-5-phosphate synthesis (precursor of histidine). These results suggested a flexibility of the OPP pathway in balancing NADPH under different light and carbon conditions.

### 4.5 Limitations of our 13C-MFA techniques for cyanobacterial study

There are several limitations in 13C-MFA techniques for cyanobacterial study. First, 13C-MFA accuracy is highly dependent on metabolic model construction. Nevertheless, incomplete annotations, errors, or inconsistencies are prevalent in cyanobacterial genome databases, rendering it difficult to generate a comprehensive metabolic network for 13C-MFA. Tracer experiments were used here to examine the structure of cyanobacterial metabolic network. Since the key intermediate tracers (e.g. 13C-glutamate and 13C-glyoxylate) are prohibitively expensive, an inverse tracer labeling approach was employed to save experimental costs. A 13C culture background was first built with commonly used 13C-substrates (bicarbonate
and glucose). Unlabeled glutamate or glyoxylate were then added into the \( ^{13}C \)-cultures as tracers. Their incorporation into downstream metabolites were used to determine pathway functions. Since the addition of intermediate tracers may change cell metabolism, the results from isotopic dilution experiments were used to validate the metabolic model.

Second, \( ^{13}C \)-MFA cannot precisely determine certain pathways. For example, the confidence intervals of anaplerotic fluxes are much larger than those of other fluxes. Additional information, such as genetic analyses, is required to validate flux analysis results.

Third, the application of \( ^{13}C \)-MFA requires the attainment of a steady state with minimal labeling changes in the central metabolism. Therefore, metabolism under photoautotrophic or circadian conditions cannot be analyzed with traditional \( ^{13}C \)-MFA. Although an advanced isotopic non-stationary MFA has been developed to capture the transient states of metabolic networks, it is difficult to precisely measure the labeling patterns of low-abundance and unstable free metabolites. It is also difficult to resolve the unexpected labeling kinetics of free metabolites caused by metabolic channeling or light heterogeneity inside of photo-bioreactors [16]. Furthermore, the degradation–regeneration of certain cellular polymers (e.g. cyanophycin) may cause the dynamic exchange of carbons between free metabolites and macromolecules, interfering the non-stationary metabolite labeling [54]. Therefore, this study is only able to profile a steady-state \( ^{13}C \)-MFA under mixotrophic conditions.

Finally, Synechocystis 6803’s metabolism is diverse and affected by both light and carbon conditions, so the observed results in this study pertain solely to photomixotrophic conditions when light and carbon sources are sufficient.

5 Conclusion

This study used \( ^{13}C \)-metabolism analysis to delineate the photomixotrophic metabolism of Synechocystis 6803. \( ^{13}C \)-analyses confirmed the in vivo conversion of \( \alpha \)-keto glutarate to succinate when an additional source was supplied to increase the \( \alpha \)-keto glutarate pool size (e.g. glutamate) while this flux under photomixotrophic conditions is negligible compared to all other fluxes in the model. Glyoxylate was discovered as a potential source for glycolytic synthesis, while the activity of glyoxylate shunt was not observed. Under photomixotrophic conditions, malic enzyme, rather than pyruvate kinase, is a fundamental route for pyruvate synthesis. Oxidative pentose phosphate pathway flux is low when light and inorganic carbon is sufficient. These findings complement information of previous multiple-omics studies, which have shown that \( ^{13}C \)-tools greatly advance the understanding of cellular metabolism. This study also suggests that the photomixotrophic metabolism of cyanobacteria can efficiently incorporate both sugar and CO\(_2\) for biosynthesis, resulting in potentially higher biomass density and productivity.

This research was funded by an NSF Career Grant (MCB-0854016). It was also supported by the Office of Science (BER), U.S. Department of Energy (DE-SC0006870), and the Consortium for Clean Coal Utilization at Washington University in St. Louis to HBF. We thank James Ballard and Katrina Leyden for their writing assistance.

The authors declare no commercial or financial conflict of interest.

6 References


Appendix Chapter 2: Photoheterotrophic fluxome in *Synechocystis* sp. strain PCC 6803 and its implications for cyanobacterial bioenergetics
Photoheterotrophic Fluxome in *Synechocystis* sp. Strain PCC 6803 and Its Implications for Cyanobacterial Bioenergetics

Le You, Lian He, Yinjie J. Tang
Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Missouri, USA

This study investigated metabolic responses in *Synechocystis* sp. strain PCC 6803 to photosynthetic impairment. We used 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU; a photosystem II inhibitor) to block O₂ evolution and ATP/NADPH generation by linear electron flow. Based on ¹³C-metabolic flux analysis (¹³C-MFA) and RNA sequencing, we have found that *Synechocystis* sp. PCC 6803 employs a unique photoheterotrophic metabolism. First, glucose catabolism forms a cyclic route that includes the oxidative pentose phosphate (OPP) pathway and the glucose-6-phosphate isomerase (PGI) reaction. Glucose-6-phosphate is extensively degraded by the OPP pathway for NADPH production and is replenished by the reversed PGI reaction. Second, the Calvin cycle is not fully functional, but Rubisco continues to fix CO₂ and synthesize 3-phosphoglycerate. Third, the relative flux through the complete tricarboxylic acid (TCA) cycle and succinate dehydrogenase is small under heterotrophic conditions, indicating that the newly discovered cyanobacterial TCA cycle (via the γ-amino-butyric acid pathway or α-ketoglutarate decarboxylase/succinic semialdehyde dehydrogenase) plays a minimal role in energy metabolism. Fourth, NAD(P)H oxidation and the cyclic electron flow (CEF) around photosystem I are the two main ATP sources, and the CEF accounts for at least 40% of total ATP generation from photoheterotrophic metabolism (without considering maintenance loss). This study not only demonstrates a new topology for carbohydrate oxidation but also provides quantitative insights into metabolic bioenergetics in cyanobacteria.

Cyanobacteria, which first appeared in shallow marine settings as early as 3 billion years ago (1, 2), are now widely distributed in diverse aquatic and light environments (3–5). They can perform oxygenic photosynthesis and respiration simultaneously in the same compartment (6, 7). Cyanobacteria contain two photosystems to harvest light energy. Photosystem II (PSII) splits water and transports electrons sequentially through plastoquinone (PQ), cytochrome b₆f, plastocyanin, and photosystem I (PSI), forming a linear electron flow (LEF) to produce ATP and NADPH. Alternatively, a cyclic electron flow (CEF) runs around PSII to generate ATP (8, 9). The cyanobacterial CEF involves respiratory electron transport reactions and PSI enzymes. NAD(P)H dehydrogenase complex (NDH-1) oxidizes NADPH and provides electrons for the PQ pool. Then the electrons from PQ flow to PSI and ferredoxin to regenerate NADPH via ferredoxin-NAD⁺ reductase (FRN). The CEF can regulate ATP and NADPH ratios for CO₂ fixation and respiration in plants (10).

Cyanobacteria have a much larger PSII content than plants. For example, the PSI/PSII ratio in *Synechocystis* sp. strain PCC 6803 is about 5, suggesting the significant role of PSI in energy metabolism (6). To decipher photosynthesis mechanisms, researchers often use the herbicide 3-(3,4-dichlorophenyl)-1,1-dimethylurea (DCMU) to block the PQ binding site of PSI so that the LEF is inactivated. Under DCMU treatment, cyanobacteria are unable to grow autotrophically but can recover their growth in the presence of glucose (the photoheterotrophic condition). The metabolic responses to DCMU stress can provide insights into the bioenergetics of photosystems. However, it is still difficult to directly quantify ATP generation from the CEF or LEF experimentally.

This study applied ¹³C-metabolic flux analysis (¹³C-MFA) to analyze the carbon and energy metabolisms in *Synechocystis* PCC 6803 after DCMU impaired its photosynthesis (11–13). *Synechocystis* PCC 6803 is a model cyanobacterium for investigation of the mechanisms of CO₂ fixation, photosynthesis, circadian rhythm, ecological effects, and the evolution of eukaryotic algae and plants. Flux analysis of *Synechocystis* PCC 6803 aimed to reveal how cyanobacteria reorganize their metabolism when only PSI is active. The study also examined whether the recently discovered tricarboxylic acid (TCA) cycle in cyanobacteria (either via the γ-amino-butyric acid pathway [GABA shunt] (14) or via α-ketoglutarate decarboxylase and succinic semialdehyde dehydrogenase (15)) can provide sufficient reducing equivalent if cyanobacterial phototrophic NADPH generation is inhibited. The photoheterotrophic fluxome, together with the transcription analysis, unravels the energy flows of the photosystems and complements the recent ¹³C-MFA studies on cyanobacterial photomixotrophic metabolism (12, 13). This study may also broadly improve our understanding of the evolution of photosystems, the effectiveness of herbicides, and cyanobacterium-based photobiorefineries.

**MATERIALS AND METHODS**

Cultivation conditions. *Synechocystis* PCC 6803 was cultivated with a modified BG-11 medium at 30°C in 250-mL shake flasks (150 rpm) with a working volume of 50 ml (12). Continuous illumination of 50 μmol of photons m⁻² s⁻¹ was supplied. For photoheterotrophic cultures, DCMU (Sigma-Aldrich, St. Louis, MO) was used at a final concentration of 10

---

*Received 26 July 2014*  
*Accepted 11 December 2014*  
*Accepted manuscript posted online 22 December 2014*  
*Editor*: P. de Boer  
*Address correspondence to Yinjie J. Tang, yinjie.tang@wustl.edu.  
*Supplemental material for this article may be found at http://dx.doi.org/10.1128/JB.02149-14.  
*Copyright © 2015, American Society for Microbiology. All Rights Reserved.  
*doi:10.1128/JB.02149-14*
μM (16). All tracer experiments began with a 0.5% inoculation ratio (the initial optical density at 730 nm [OD730] was around 0.002) from a photoautotrophic *Synechocystis* stock. Cell density was monitored by a Genavys UV-visible (UV-Vis) spectrophotometer (Thermo Scientific) at 730 nm. The OD730 (instrument = 0.485 g [dry biomass]/liter) of the culture was measured by an enzyme kit (R-Biopharm, Darmstadt, Germany). To examine the PSI1 functions, we performed chlorophyll a fluorescence induction analysis of both photosynthetic and phototrophic cultures (2 μl/liter glucose) using a fluorometer (FL3500 Photon System Instruments, Czech Republic). The fluorescence yield rose significantly when 4C-MFA was present in *Synechocystis* glauce cultures (see Table S1 in the supplemental material), confirming that PSI1 was unable to capture sufficient photons to energize electrons.

[^1]C glucose and [^13]C bicarbonate were purchased from Cambridge Isotope Laboratories (Tewksbury, MA). Three tracer experiments were carried out to investigate cyanobacterial metabolism under DCMU stress. The first experiment confirmed the inhibition of photoautotrophic protein synthesis by DCMU. We grew *Synechocystis* with 2 g/liter NaH[^13]CO3 in sealed serum bottles (with no source of[^12]CO2). The biomass ratio in the sealed serum bottle was fully labeled. During the early exponential-growth phase (OD730 = 0.3), DCMU and 2 g/liter unlabeled NaHCO3 were injected into the[^13]C-labeled cultures. The biomass was sampled after 1 day incubation with DCMU and unlabeled NaHCO3. The abundance of[^13]C in proteinogenic amino acids was measured in order to determine whether Synechocystis could perform autotrophic biosynthesis under DCMU stress. The second experiment examined cyanobacterial CO2 fixation under DCMU treatment. We cultivated *Synechocystis* PCC 6803 in open air with [U-[^13]C]glucose and DCMU in shake flasks. The biomass was collected after 3 days of cultivation (OD730 = 0.3) for isoprenoid analysis. The third experiment delineated the cyanobacterial metabolic flux profile. *Synechocystis* PCC 6803 was grown in open air with [1,2-[^13]C]glucose and DCMU. Labeled biomass in the exponential-growth phase was collected for isoprenoid analysis, which provided distinct isotopic data for[^13]C-MFA.

RNA sequencing and data analysis. In parallel to the[^13]C-MFA tracer experiments, we examined cyanobacterial gene expression in photoautotrophic, photomixotrophic, and DCMU-treated cultures. Biological triplicates were sampled during the exponential-growth phase. Cell pellets (~10 mg) were frozen with liquid nitrogen immediately after centrifugation. Total RNA was extracted using an miRNeasy mini kit (Qiagen, Valencia, CA). DNA was removed with DNase I (Roche Applied Science, Germany). The RNA integrity numbers of all samples used for sequencing were greater than 8.0. RNA profiling (library preparation, sequencing, and data analysis) was performed by the Genome Technology Access Center at Washington University (https://gtac.wustl.edu/services/). In brief, total RNA was dephased of RNA by using a Ribo-Zero Magnetic Kit (Epicentre, Madison, WI). Subsequently, mRNA was fragmented and reverse transcribed to prepare cDNA libraries using random primers. End repair was performed by adding an A base to each 5-end. The libraries were then ligated to sequencing adapters and were amplified with index primers. The libraries were sequenced using Illumina HiSeq 2000 sequencing machine. Raw image data were converted to base calls using CASAVA. The samples were demultiplexed into individual libraries using FASTQ files by using the proprietary demultiplexer algorithm (Illumina, San Diego, CA). The FASTQ files were aligned with the whole-genome transcriptome using TopHat (version 2.0.8) (17, 18). To compare gene expression levels, pairwise negative binomial tests in edgeR software were performed (19, 20).

Isoprenoid analysis using GC-MS. The labeling of proteinogenic amino acids was measured as reported previously (21). Biomass pellets from 20-mL cultures were harvested by centrifugation and were washed with 0.9% NaCl solution. The pellets were then hydrolyzed with 6 M HCl at 100°C. The resulting mixtures were subsequently air dried and derivatized with N-trim butyldimethylsilyl-N-methylithiofluoroacetamide (TDMS) prior to gas chromatography–mass spectrometry (GC-MS) analysis. We used a gas chromatograph (Hewlett-Packard model 7890A, Agilent Technologies) equipped with a DB-5ms column (J&W Scientific) and a mass spectrometer (5975C, Agilent Technologies). The GC thermal programs have been described in a previous report (21). A published algorithm was used to calculate the M5 data for all derivatized metabolites (22). Due to overlapping peaks or product degradation, several amino acids (proline, arginine, cysteine, and tryptophan) were not used for flux analysis (23). The MS data ([M-57]^+; [M-159]^+ or [M-85]^+, and [M+2]^2) of key amino acids are shown in Table S2 in the supplemental material. Isotoping labeling fractions designated M6, M1, M2, etc., represent fragments containing unlabeled, singly labeled, and doubly labeled amino acids, and so on.

[^1]C-MFA modeling and flux calculation. *[^1]C-MFA* was performed based on proteinogenic amino acids from *Synechocystis* PCC 6803 grown in open air with [1,2-[^13]C]glucose and DCMU. The labeling distribution of key amino acids did not change during the exponential-growth phase (see Table S2 in the supplemental material), indicating the attainment of a pseudo-steady state. Based on previous reports (3, 11, 13, 15), our metabolic network of *Synechocystis* PCC 6803 includes reactions in glycolysis, the Calvin cycle, the TCA cycle, the glyoxylate shunt, photosorption pathways, and biomass synthesis (see Table S3 in the supplemental material). The model also assumed that internally generated CO2 from glucose degradation was exchanged with atmospheric CO2, resulting in a mixture of CO2 pools. By normalizing the glucose uptake rate to a value of 100 (unitless),[^1]C-MFA profiled relative metabolic fluxes through the central metabolism. The relative fluxes were solved by minimizing a quadratic error function that calculated the difference between predicted and measured isoprenoid patterns (see Table S3). The biomass composition was based on a previous report (see Table S3) (24). Since DCMU treatment may increase cyanobacterial glyoxylate content (25),[^1]C-MFA tested the sensitivity of flux distributions to glyoxylate composition. This was done by first assuming 10-fold increase of the glyoxylate content in the biomass composition and then recalculating fluxes by[^1]C-MFA using the same set of labeling information.

Reaction reversibility was characterized by the exchange coefficient (26) and the net flux (26). Optimization for[^1]C-MFA was achieved by the EMU (elementary metabolite units) method (27). The MATLAB optimization solver fmincon was used to minimize the quadratic error function. To avoid local minima, 100 initial guesses were randomly generated, and the solution set that minimized the objective function was used as the best fit. The 95% confidence intervals were calculated using the Monte Carlo method as described previously (28), in which the measured isoprenoid data (M0, M1, M2, etc.) were perturbed with normally distributed noise within measurement errors (±0.005) 500 times, and the optimization solver was restarted with the optimal solution. The confidence intervals of the fluxes were determined based on 500 simulations. The MATLAB program for[^1]C-MFA can be downloaded from the Tang laboratory website (http://tang.ecce.wustl.edu/Tool Development.html).

**RESULTS**

Carbon utilization in DCMU-treated *Synechocystis* PCC 6803 cultures. We added 10 μM DCMU to the photoautotrophic cultures right after inoculation (initial OD730 = ~0.002), and no biomass growth was observed (see Fig. S1 in the supplemental material). We also added DCMU to *Synechocystis* cultures during the exponential-growth phase, and photoautotrophic growth stopped (Fig. 1). If DCMU-treated cultures were supplemented with glucose (Fig. 2), *Synechocystis* growth could be recovered (growth rate [μ], ~0.03 h−1). Next, we investigated *Synechocystis* protein synthesis under DCMU stress. Photosynthetic cultures started with an OD730 of ~0.002 and were grown with NaH[^13]CO3 as the only carbon source. DCMU and unlabeled
NaH\(^{13}\)CO\(_3\) were added during the exponential-growth phase (OD\(_{590}\) ~ 0.3). Biomass samples, collected after 24 h of DCMU-treatment, showed negligible \(^{13}\)C incorporation (proteinogenic amino acids remained mostly \(^{13}\)C labeled [Fig. 1]). This result further confirmed that DCMU blocks protein synthesis under autotrophic conditions (25). In contrast, both \(^{13}\)C and \(^{12}\)C were present in the biomass if DCMU-stressed Synechocystis PCC 6803 was grown with uniformly labeled [U-\(^{13}\)C]glucose and atmospheric \(^{12}\)CO\(_2\) (Fig. 3). Amino acids from glycolysis (e.g., alanine and serine) and the pentose phosphate pathways (e.g., phenylalanine and histidine) had around 10 to 15% \(^{13}\)C enrichment, indicating that RubisCO in the cells still actively fixed atmospheric \(^{12}\)CO\(_2\) under DCMU stress. In addition, aspartate, synthesized from oxaloacetate (OAA), had even higher \(^{13}\)C enrichment (>20%) due to additional \(^{13}\)CO\(_2\) fixation (possibly by the anaerobic reactions).

**Metabolic flux analysis of DCMU-treated Synechocystis PCC 6803.** Synechocystis PCC 6803 was grown in the open air with DCMU and \([12,\,^{13}\)C\(_2\)]glucose. Biomass samples at the mid-exponential-growth phase provided amino acid labeling information for \(^{13}\)C-MFA. There were negligible differences between the isotopomer data sets at two different time points, indicating the attainment of a metabolic steady state (see Table S2 in the supplemental material). The model-fitting quality is shown in Fig. S2 in the supplemental material (\(R^2 > 0.99\); the sum of squared residuals [SSR] is 0.0039). The relative flux distributions, the exchange coefficients for reversible reactions, and the 95% confidence intervals are shown in Fig. 4 and Table S3 in the supplemental material. Our previous Synechocystis \(^{13}\)C-MFA indicated that glucose is used mainly through the glycolysis pathway under photonixotropic conditions (12). However, \(^{13}\)C-MFA of photoheterotrophic cultures revealed that glucose was oxidized to CO\(_2\) by a carbohydrate degradation cycle. Specifically, the oxidative pentose phosphate (OPP) pathway was highly active in oxidizing glucose-6-phosphate (G6P) to pentose phosphates, which were then used to synthesize fructose-6-phosphate (F6P). F6P was diverted to G6P through the reverse reaction of glucose-6-phosphate isomerase (PGI), forming a carbohydrate degradation loop (see Fig. S3A in the supplemental material). A similar glucose degradation cycle was observed recently when Synechocystis PCC 6803 was under atrazine stress (29). In agreement with the observation in Fig. 3, \(^{13}\)C-MFA detected significant CO\(_2\) fixation via RubisCO and phosphoenolpyruvate (PEP) carboxylase. However, 3-phosphoglycerate (3PG), the first metabolite of CO\(_2\) fixation through RubisCO, entered glycolysis instead of the Calvin cycle. Phosphoenolpyruvate carboxylase and malic enzyme, as observed elsewhere under phototrophic (11) and photomixotrophic (12, 13) conditions, remained active under the photoheterotrophic condition.

Recently, the GABA shunt was identified as closing the Synechocystis glycolytic pathway as an alternative C-fixation pathway (30).
FIG 4 Flux distribution and energy flows in *Synechocystis* PCC 6803 under DCMU treatment. The biomass samples from *Synechocystis* PCC 6803 cultures grown with 1,2,3-C\textsubscript{14} glucose and DCMU in open air were collected during the exponential phase for isotope tracer analysis. The estimated relative flux rates are shown beside the pathways, which are normalized to a glucose uptake flux assumed to be 100. The absolute glucose uptake rate was 0.41 mmol/g dry weight of cells/h, and g was 0.038 h\textsuperscript{-1} (Fig. 2; see also Fig. S7 in the supplemental material). The standard deviations are shown after the flux rates (based on the confidence intervals in Table S3 in the supplemental material). The dashed arrows represent biomass flux, and the shaded arrows represent energy flows. Stars highlight some key bidirectional reactions. Abbreviations: 6PG, 6-phosphogluconate; AcCoA, acetyl coenzyme A; AKG, α-ketoglutarate; CIT, citrate; Cyt, cytochrome; DHAP, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; FBP, fructose 1,6-bisphosphate; Fe, ferredoxin; FUM, fumarate; GAP, glyceraldehyde 3-phosphate; GLX, glyoxylate; ICD, isocitrate; PC, plastocyanin; RSP, ribose 5-phosphate; S7P, sedoheptulose-7-phosphate; SUC, sucrose; X3P, xylulose-5-phosphate.

*Synechocystis* TCA cycle (14). Although cyanobacteria have the potential to use these newly discovered enzymes for running the TCA cycle, \(^{13}\)C-MFA found that TCA cycle flux through the GABA shunt and succinate dehydrogenase was negligible. Thus, *Synechocystis* PCC 6803 cannot activate its TCA cycle to generate significant NAD(P)H and ATP under phototrophic conditions. The low TCA cycle flux also indicated that succinate dehydrogenase plays a small role in quinone reduction.

\(^{13}\)C-MFA could deduce how the cyanobacterium ensures its cofactor and ATP balancing, based on the fluxome and biomass growth (Fig. 4). If the glucose uptake is normalized to 100, 489 NAD(P)H molecules can be synthesized in the central carbon metabolism, of which 342 NAD(P)H molecules were used for biomass synthesis (see Tables S3 and S4 in the supplemental material) (30). A total of 654 ATP molecules were required by the central carbon metabolism (112 ATP molecules) and biomass synthesis (542 ATP molecules) (Fig. 4; see also Table S4). A maximum of 369 ATP molecules was produced by the oxidative phosphorylation of excess NAD(P)H (see Table S4). This number is still less than the requirement for biomass growth, and thus the CEF around PSI presumably contributes 285 ATP molecules (without considering maintenance loss) (see Table S4).

Transcriptomics analysis of *Synechocystis* phototrophic metabolism. Photoautotrophic (with CO\textsubscript{2}), photomixotrophic (with CO\textsubscript{2} + glucose), and photoheterotrophic (with CO\textsubscript{2} + glucose + DCMU) cultures were collected at similar cell densities for transcriptome sequencing (RNA-Seq) analysis. All the samples showed high genome-mapping ratios (>90%). The sequence reads matched more than 3,500 of the 3,561 coding genes in the *Synechocystis* PCC 6803 genome. Multidimensional scaling plots compared the similarities of samples obtained under different conditions (see Fig. S4 in the supplemental material).
The negligible distances among the biological triplicates indicated the overall high quality of RNA sequencing. The RNA-Seq results are shown in Fig. 5 and 6, as well as in Table S5 in the supplemental material. Using a 2-fold change in biological triplicates as the cut-off (31), we observed conserved expression of key genes involved in the central carbon metabolic pathways (e.g., the TCA cycle, anaerobic pathways, and glycolysis) among all three conditions. In agreement with the 13C-MFA observation, the expression levels of the PGI gene and genes involved in the OPP pathway were highly upregulated after DMU treatment. TheRubisCO genes (GenBank accession no. BAA1098 and BAA10192), involved in

FIG 5 Pairwise comparisons of RNA expression levels for genes in the central carbon metabolism. The following log, ratios were calculated: [log(photon/microscopic data(DCMU data)) / log(photon/microscopic data(DCMU data); log(photomorphometric data DEVELOPMENT data); log(abundance)]. Abbreviations: GFB, 1,3-bisphosphoglycerate; GLYC, glucose-6-phosphate; SSA, succinic semialdehyde; SUC, succinyl-CoA, succinyl-CoA.
and CO₂ fixation, also remained highly active in DCMU-treated Syncepysis PCC 6803.

As for energy metabolism under DCMU stress, the NDH-1 subunits (GenBank accession no. BAA17787, BAA18722), participating in NADPH oxidation (6), were downregulated due to the NADPH deficiency after PSI inactivation. The succinate dehydrogenase complex (SDH) supplies electrons to the quinone pool, DCMU stress reduced the expression level of the flavoprotein subunit of SDH (BAA17519), which converts succinate to fumarate. Moreover, genes encoding the nitrate transport protein (subunits NrtD and NrtC (BAA10880 and BAA10857, respectively) were downregulated, suggesting a lower rate of nitrogen reduction and assimilation into biomass. On the other hand, most genes associated with CEF around PSI, such as the genes encoding FNR (BAA18459), cytochrome 553 (BAA17354), plastocyanin (BAA10227), and cytochrome b₆f complex (BAA10150, BAA10763, and BAA18236), were upregulated under DCMU stress. The gene encoding ferredoxin (BAA18248), mediating electron transfer around PSI, had a constant high expression level through all three growth conditions. Finally, we observed correlations between the fluxome and corresponding RNA-seq data in key pathways. For example, both ¹³C-MFA and RNA-Seq indicated an active OPP pathway and CO₂ fixation through Rubisco. Comparisons between "omics" data may reveal the transcriptional or posttranscriptional regulation of carbon metabolism by cyanobacterial genes (32).

Flux sensitivity test and validations. ¹³C-MFA often employs a simplified network and takes advantage of published analytical data (such as biomass composition) for microbial characterization. ¹³C-MFA results on model assumptions, flux constraints, and metabolite measurements. To verify our model, we tested the sensitivity of flux topology to modeling constructions. First, we performed in silico deletion of Rubisco-based CO₂ fixation (setting its flux as zero). A three-times-greater SSR was observed (see Fig. S5A in the supplemental material). Next, the in silico deletion of phosphoribulokinase was tested by blocking the conversion from ribulose-5-phosphate (Ru5P) to ribulose-1,5-diphosphate (RuBP), and the model fit quality decreased (see Fig. S5B in the supplemental material). Both tests, together with the observations shown in Fig. 3, validated CO₂ fixation by Rubisco under DCMU stress.

Second, we forced the reversible reaction FeP → G6P to be unidirectional (G6P → FeP). The SSR increased significantly (see Fig. S5C in the supplemental material), and the overall NADP/H production became insufficient for biomass synthesis. Therefore, the carbohydrate degradation cycle, via the reversed FGL reaction and the OPP pathway, is indispensable. Third, DCMU-treated cyanobacteria could accumulate glycerol (25). In the model test, we set a higher glycerol ratio in the biomass equation. With the new biomass composition, the ¹³C-MFA could still fit isotopomer data well (see Fig. S6 in the supplemental material). Compared to the fluxes in Fig. 4, the new flux distribution showed that the flux from G6P toward glycerol increased, and the fluxes to protein synthesis and cell growth decreased (compare Fig. 4 and Fig. S6 in the supplemental material). In spite of these differences in the flux values, the overall metabolic topology—the OPP oxidation cycle, TCA cycle, and Rubisco, etc.—remained the same under different assumptions about biomass composition. Fourth, precise measurement of malic enzyme activity is difficult, since the two pathways (PEP + CO₂ → OAA → MAL → CO₂ + PYP) generate identical labeling patterns in pyruvate (PYP) and OAA. However, ¹³C-MFA revealed small fluxes through the glyoxylate shunt and TCA cycle, which fine-tuned the malate (MAL)/OAA/PYP labeling profiles. In silico deletion of malic enzyme flux resulted in an increase of the SSR (see Fig. SSD in the supplemental material), suggesting that a measurable malic enzyme flux was favored for model fitting. Fifth, we tested the sensitivity of model fitting to the photorespiration flux (RuBP + O₂ → glycolate + 3PG). The SSR rose if the RuBP oxidation flux was forced to be increased (see Fig. S5E in the supplemental material). This result confirmed that cyanobacterial photorespiration was inhibited by DCMU (33). Comparing to gene expression under the photorespiratory conditions, RNA-Seq analysis also verified decreased expression levels of glycolate oxidase (BAA10256 and BAA16857) in DCMU-treated Synechocystis (see Table S5 in the supplemental material). Finally, recent studies found that the cyanobacterial TCA cycle can be completed via either α-ketoglutarate decarboxylase and succinic semialdehyde dehydrogenase (15) or the GABA shunt (14). In our model test (see Fig. S5F in the supplemental material), the SSR increased slightly after in silico deletion of the pathway (α-ketoglutarate -> succinate), indicating that the complete TCA cycle could be functional even though its flux is low (~1%).

**DISCUSSION**

Synechocystis PCC 6803 reorganizes its metabolism after PSII inactivation. First, the activities of the Calvin cycle and the OPP pathway in cyanobacteria are sensitive to the intracellular NADP/H abundance (7). Several scenarios of the pentose phosphate (PP) pathway are shown in Fig. S3 in the supplemental material. Under photomixotrophic conditions, the allosteric glucose-6-phosphate dehydrogenase in Synechocystis is suppressed. Glucose is metabolized mainly through glycolysis (see Fig. S3B) (12, 13). The inactive OPP pathway is also observed in *Rhodopseudomonas palustris* when reducing equivalents are supplied via acetate oxidation (see Fig. S3C) (34). In the photoautotrophic metabolism of Synechocystis PCC 6803, OPP could be employed together with the Calvin cycle (reversed pentose phosphate pathway) to regulate the cofactor balance and CO₂ fixation (see Fig. S3D). Under DCMU stress, Synechocystis metabolism contains a highly active cycle via the OPP pathway to degrade glucose and produce NADPH (see Fig. S3A), while its Rubisco remains for CO₂ fixation to compensate for carbon loss. This metabolic topology does not rely on a fully functional Calvin cycle. A counterpart metabolic feature is also observed in the embryos of plant seeds (35). On the other hand, the TCA cycle and the PEP-glyoxylate cycle are two common metabolic oxidation routes by which microbes generate reducing equivalents (36). Although Synechocystis contains all functional enzymes in the TCA cycle, the flux through the TCA cycle remains minimal after inactivation of its major NADPH generation route (i.e., photosynthesis).

Second, malic enzyme is the main route for supplying pyruvate in *Synechocystis PCC 6803* under continuous light illumination, because pyruvate kinase is inhibited by ATP from photosynthesis (37). The malic enzyme activity has been revealed under photoautotrophic and photomixotrophic conditions (11–13) and with DCMU treatment. RNA-Seq analyses detected the high expression level of malic enzyme (BAA16663) under all three conditions. On the other hand, DCMU can alleviate pyruvate kinase inhibition by blocking LEO-based ATP synthesis, rendering malic en-
zyme less important for pyruvate synthesis. The adverse effect of eliminating malic enzyme (\Delta mtd201) on the growth of DCMU-treated Synechocystis PCC 6803 was shown to be much less significant than that in phototrophic or photoautotrophic cultures (37).

Third, regulation of the ATP/NADPH ratio is important for efficient photosynthesis (8). In DCMU-treated Synechocystis PCC 6803, the oxidative phosphorylation of NAD(P)H cannot produce sufficient ATP to support cell growth. To fulfill the ATP requirement, Synechocystis PCC 6803 upregulates the CEF (Fig. 4). Inside the CEF, NDH-I oxidizes NAD(P)H and supplies electrons to the PQ pool (38–41). SDH in the TCA cycle could also pump electrons to the CEF (6, 42). However, the TCA cycle flux in Synechocystis is small under DCMU stress (Fig. 4), and thus, succinate contributes a minimum number of electrons to the PQ pool. RNA-Seq detected highly upregulated expression of genes associated with PSI (FNR, cytochrome 555, cytochrome b6f complex, and plastocyanin) in DCMU-treated Synechocystis PCC 6803 (Fig. 6). Both 3C-MFA and transcriptional analysis indicate that the CEF pumped a significant amount of ATP to sustain cell metabolism after PSI inactivation (8).

Finally, this study offers evolutionary insight into photosynthetic potential and applications of photofermentation. Photobacterial photosystems originated from anoxygenic phototrophs, which employed either PSI or PSII (43). It is still unclear how these two photosystems were brought together to form cyanobacterial photosynthesis (44). The metabolism of primitive cyanobacterial ancestors might have employed the photoheterotrophic fluxome (as shown in Fig. 4) before the complete development of oxygenic photosynthesis.

Conclusions. Synechocystis PCC 6803 can simultaneously metabolize organic carbons and CO2. This study presents a preliminary picture of cyanobacterial carbon and energy metabolisms under DCMU stress. When photosynthesis is sufficient, the Calvin cycle of Synechocystis PCC 6803 actively fixes CO2. When photosynthesis is limited or inactivated, its OPP pathway can be upregulated to degrade sugar for NADPH generation. Synechocystis PCC 6803 shows a high degree of metabolic flexibility for carbon utilization, making it a robust workhorse for synthesizing value-added products via photofermentation. The knowledge of cyanobacterial fluxomics has broad implications for many research fields, such as herbicide technology, plant physiology, evolution, and bioenergy.

ACKNOWLEDGMENTS

This project is supported by an NSF Career Grant (MCB0954016) and a NASA EXobiology Grant.

We thank Haijun Liu for advice on DCMU experiments and James Ballard for writing assistance.

REFERENCES

Appendix Chapter 3: Kinetic modeling and isotopic investigation of
isobutanol fermentation by two engineered *Escherichia coli* strains
Kinetic Modeling and Isotopic Investigation of Isobutanol Fermentation by Two Engineered *Escherichia coli* Strains

Yi Xiao, Xueyang Feng, Arul M. Varman, Lian He, Huifeng Yu, and Yinjie J. Tang

Department of Energy, Environmental and Chemical Engineering, Washington University, One Brookings Drive, St. Louis, Missouri 63130, United States

ABSTRACT: We constructed an *Escherichia coli* BL21 strain with the Ehrlich pathway (the low-performance strain for isobutanol production). We also obtained a high isobutanol-producing *E. coli* strain JCL260 from the James Liao group (University of California). To compare the fermentation performances of the two engineered strains, we employed a general Monod-based model coupled with mixed-growth-associated isobutanol formation kinetics to simulate glucose consumption, biomass growth, and product secretion/loss under different cultivation conditions. On the basis of both kinetic data and additional $^{13}$C-isotopie investigation, we found that the low-performance strain demonstrated robust biomass growth in the minimal growth medium (20 g/L glucose), achieving isobutanol production (up to 0.95 g/L). It utilized significant amounts of yeast extract to synthesize isobutanol when it grew in the rich medium. The rich medium also enhanced waste product secretion, and thus reduced the glucose-based isobutanol yield. In contrast, JCL260 had poor biomass growth in the minimal medium due to an inflated Monod constant ($K_M$), while the rich medium greatly promoted both biomass growth and isobutanol productivity (~60% of the theoretical isobutanol yield). With the optimized keto-acid pathway, JCL260 synthesized isobutanol mostly from glucose even in the presence of sufficient yeast extract. This study not only provided a kinetic model for scaled-up isobutanol fermentation but also offered metabolic insights into the performance trade-off between the two engineered *E. coli* strains.

1. INTRODUCTION

Biobutanol is a second-generation biofuel that has higher energy density and lower water solubility than ethanol. Acetone–butanol–ethanol (ABE) fermentation is a traditional bioprocess that uses Clostridium acetobutylicum to produce n-butanal, but such a process is restrained by the slow alcohol production rate. To overcome this restriction, the n-butanal pathway derived from *Clostridium* has been reconstructed in fast-growing *Escherichia coli* or yeast strains. Butanol biosynthesis via the *Clostridium* pathway has limitations including low product titer and yield due to the accumulation of toxic metabolites. Another approach is via the keto-acid pathway to produce low-toxicity isobutanol (IB), where the amino acid biosynthesis pathways and the Ehrlich pathway are incorporated for alcohol synthesis. This method shows effective production of higher alcohols because of robust and ubiquitous amino acid pathways.

Table 1 summarizes diverse biobutanol production strategies, including the overexpression of the targeted pathway in different microbial hosts (including photosynthetic microorganisms), the elimination of competing pathways, the system redesign of host metabolism, and the integration of fermentation with in situ product separation. However, few papers have studied the kinetics of engineered microbial hosts for biobutanol fermentation. To apply a newly developed host in the biofuel industry, a kinetic-based model is of practical importance not only for designing optimal scaled-up fermentation but also for understanding the internal metabolic features of microbial hosts in response to various nutrient sources and cultivation conditions. To fill this gap, our lab has created an *E. coli* mutant that produces IB via the Ehrlich pathway. Meanwhile, we have obtained a high-performance *E. coli* strain JCL260 with an optimized metabolism for IB synthesis (offered by the James Liao group). On the basis of fermentation data using both strains, we developed an empirical model to analyze and compare their fermentation kinetics. We also performed $^{13}$C-experiments to investigate the nutrient use of the two mutant strains for the synthesis of biomass and IB.

2. EXPERIMENTAL SECTION

2.1. Pathway Construction. We engineered *E. coli* BL21 (DE3) by heterologous expression of the *kivd* (2-ketoisovalerate decarboxylase) and *adhA* (aldehyde reductase) genes. The two genes were amplified from *Lactobacillus lactis* by PCR with high-fidelity DNA polymerase Pfu (Invitrogen). Primers for *kivd* were 5'-gccagcagatagtcaggtgac-3' and 5'-tgggctttgatttttggc-3'. Primers for *adhA* were 5'-ctcaaggtggtggtttttttgaag-3' and 5'-atgacgctgctgggtttttttgaag-3'. The genes *kivd* (treated with Xhol/KpnI) and *adhA* (treated with KpnI/SphI) were cloned into the pTAC-MAT-Tag 2 expression vector (Sigma-Aldrich) via Xhol/Sphl to create the plasmid pTAC-KA, and then were transformed into *E. coli* BL21 (DE3). This low-performance mutant used its native valine biosynthesis pathway to generate 2-ketoisovalerlate and then converted it to IB by the heterologous Ehrlich pathway (Figure 1). To confirm the expression of *Kivd* and *AdhA*, we performed SDS-PAGE analysis of the recombinant strain and observed the protein bands of Kivd (~60 kDa) and AdhA (~35 kDa). The strain secreted IB, acetate, lactate, ethanol, and a small amount of n-
Table 1. Recent Studies on Biofuel Production by Engineered Microorganisms

<table>
<thead>
<tr>
<th>products</th>
<th>substrates</th>
<th>host cells</th>
<th>titer</th>
<th>research highlights</th>
<th>ref</th>
</tr>
</thead>
<tbody>
<tr>
<td>IB</td>
<td>glucose</td>
<td>E. coli</td>
<td>22 g/L</td>
<td>introduction of a nonfermentative pathway to produce IB; elimination of competing pathways to reduce waste metabolite secretion</td>
<td>6</td>
</tr>
<tr>
<td>IB</td>
<td>CO₂</td>
<td>E. coli</td>
<td>50 g/L</td>
<td>in situ IB removal from the bioreactor using gas stripping</td>
<td>9</td>
</tr>
<tr>
<td>IB</td>
<td>cellulose</td>
<td>Clostridium</td>
<td>0.66 g/L</td>
<td>direct conversion of cellulose to IB using engineered cellulolytic bacterial species</td>
<td>17</td>
</tr>
<tr>
<td>IB</td>
<td>glucose</td>
<td>E. coli</td>
<td>1.7 g/L</td>
<td>a strain optimized for IB production via elementary mode analysis</td>
<td>18</td>
</tr>
<tr>
<td>IB</td>
<td>glucose</td>
<td>E. coli</td>
<td>13.4 g/L</td>
<td>utilization of the NADH-dependent enzyme in lactic acid pathway to alleviate cofactor imbalance</td>
<td>19</td>
</tr>
<tr>
<td>IB amino acids</td>
<td>E. coli</td>
<td>~2 g/L</td>
<td>utilization of protein hydrolyses for higher alcohol synthesis by introducing enzymes for enhanced transamination and deamination cycles</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>IB CO₂</td>
<td>Ralstonia eutropha</td>
<td>140 mg/L</td>
<td>developing an electromicrobial process to convert CO₂ to higher alcohols</td>
<td>20</td>
<td></td>
</tr>
<tr>
<td>butanol</td>
<td>glucose</td>
<td>E. coli</td>
<td>1 g/L</td>
<td>a strain engineered for 1-butanol and 1-propanol production via isoleucine biosynthesis pathway</td>
<td>21</td>
</tr>
<tr>
<td>butanol</td>
<td>galactose</td>
<td>Saccharomyces cerevisiae</td>
<td>2.5 mg/L</td>
<td>overexpression of n-butanol pathway derived from Clostridium</td>
<td>5</td>
</tr>
<tr>
<td>butanol</td>
<td>glucose</td>
<td>E. coli</td>
<td>4.6 g/L</td>
<td>increase of the barrier for the reverse reaction of butyryl-CoA to propionyl-CoA via trans-enol-CoA reductase</td>
<td>3</td>
</tr>
<tr>
<td>butanol</td>
<td>CO₂</td>
<td>E. coli</td>
<td>30 g/L</td>
<td>use of trans-enol-CoA reductase and optimization of NADH and acetyl-CoA driving forces</td>
<td>4</td>
</tr>
<tr>
<td>butanol</td>
<td>CO₂</td>
<td>Sphaerotilus denitrificans</td>
<td>16.5 mg/L</td>
<td>anaerobic production of butanol from CO₂ using CO₂-dependent butanol pathway</td>
<td>22</td>
</tr>
<tr>
<td>butanol</td>
<td>glucose</td>
<td>E. coli</td>
<td>~14 g/L</td>
<td>utilization of a functional reversal of the β-oxidation cycle for the synthesis of alcohols</td>
<td>23</td>
</tr>
<tr>
<td>butanol</td>
<td>CO₂</td>
<td>Sphaerotilus denitrificans</td>
<td>30 mg/L</td>
<td>driving butanol synthesis pathway forward via an engineered ATP consumption</td>
<td>24</td>
</tr>
</tbody>
</table>

Figure 1. Metabolism in the E. coli strains for IB production. R₆, R₂C, R₀, R₆, R₆, and R₆ are shown in the eqs 1–12. IB synthesis consumes one mole NAPDH (by keto-acid reductoisomerase) and one mole NADH (by aldehyde reductase). The cell metabolism removes the redundant NADH by O₂ oxidation or by synthesis of lactate and ethanol.

propanol and methyl-butanol (product profiles similar to those of other IB-producing E. coli strains). Additionally, Professor James Liao from the University of California offered us the E. coli strain JCL260 with plasmids pSA65 and pSA69. This high-performance strain not only contains two plasmids that overexpress the entire IB pathway but also has gene deletions to interfere with the biosyntheses of waste products (acetate, formate, ethanol, succinate, and lactate).

2.2. Culture Conditions. In this study, both shake flask and bioreactor fermentation were used for IB production. Two culture media were used: (a) a minimal medium that contained 2% glucose, M9 salts (Difco), and 10 mg/L vitamin B1; and (b) a rich medium containing the minimal medium with 5 g/L yeast extract. For growing the low-performance strain, 50 mg/L ampicillin was added to the culture media; for growing the JCL260 strain, 25 mg/L kanamycin and 50 mg/L ampicillin were added to the culture media. We performed shake flask cultures for the JCL260 strain (as a positive control to confirm the correct strain and growth conditions). Specifically, 20 mL of culture (20 g/L glucose, rich medium) was placed in a 250-mL flask with a rubber stopper (200 rpm). The initial OD₆₀₀ was ~0.04, and the incubation temperature was 30°C. When OD₆₀₀ reached ~0.8, 0.1 mM IPTG (isopropyl β-D-1-thiogalactopyranoside) was added to induce IB synthesis. During incubation, the rubber stopper was opened briefly (~20 s) every ~12 h to let fresh air into the flask to give the best IB production. Within 55 h of cultivation, we obtained 4.7 g IB/L. To reach a higher titer of IB, additional glucose solution was added into the same culture after 55 h of incubation. The shake flask culture produced ~12 g IB/L using a total of 46 g/L glucose (~63% of the theoretical maximum of IB).

Fermentations were performed in a New Brunswick Bioflo 110 fermentor with a dissolved oxygen (DO) electrode, a temperature electrode, and a pH meter. The 100% DO was defined as the point where the cell free medium was purged by air (~2 L/min) for 15 min. In the oxygen-limited fermentations (air rate = 0 L/min), the DO dropped to 0% during the exponential growth phase. To start each fermentation, 400 mL of culture was inoculated with 5 mL of overnight LB culture (OD₆₀₀ ~3.5) of the recombinant E. coli strain. The cultivation conditions were pH = 7.0 (controlled by adding 2 mol/L NaOH via an autopump), temperature = 30 °C, and stirring speed = 200 rpm. For all fermentations, cells were first grown in aerobic conditions (DO > 50%) before adding 0.2 mM IPTG. After IPTG induction, we imposed two O₂ conditions: (1) in aerobic conditions, air (flow rate: ~1 L/min) was bubbled into the bioreactor to provide O₂ and to remove IB (i.e., gas stripping) from the bioreactor; (2) in oxygen-limited conditions, air was turned off, and the DO was maintained at zero during IB production. For the low-performance strain, we had three fermentations: F1 (minimal medium and aerobic conditions), F2 (minimal medium and oxygen-limited conditions), and F3 (rich medium and oxygen-limited conditions). For JCL260, we had two fermentations (F4: minimal medium
Table 2. Parameters of the Monod Model for E. coli IB Fermentation

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Units</th>
<th>F1</th>
<th>F2</th>
<th>F3</th>
<th>F4</th>
<th>F5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{S}$ Monod constant</td>
<td>g/L</td>
<td>0.32 ± 0.05*</td>
<td>0.32 ± 0.05</td>
<td>0.32 ± 0.05</td>
<td>10 ± 1</td>
<td>10 ± 1</td>
</tr>
<tr>
<td>$K_{a}$ acetate inhibition</td>
<td>g/L</td>
<td>49 ± 11</td>
<td>49 ± 11</td>
<td>1.0 ± 0.2</td>
<td>1.0 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>$\mu_{max}$ specific growth rate</td>
<td>h</td>
<td>0.051 ± 0.004</td>
<td>0.015 ± 0.001*</td>
<td>0.12 ± 0.01</td>
<td>0.12 ± 0.01</td>
<td></td>
</tr>
<tr>
<td>$Y_{SG}$ biomass yield from Glu</td>
<td>g biomass/g Glu</td>
<td>0.18 ± 0.03</td>
<td>0.14 ± 0.01</td>
<td>0.20 ± 0.04</td>
<td>0.38 ± 0.01</td>
<td>0.39 ± 0.03</td>
</tr>
<tr>
<td>$Y_{EC}$ acetate yield from Glu</td>
<td>g acetate/g Glu</td>
<td>0.076 ± 0.007</td>
<td>0.09 ± 0.004</td>
<td>0.3 ± 0.07</td>
<td>0.32 ± 0.01</td>
<td>0.35 ± 0.02</td>
</tr>
<tr>
<td>$Y_{LC}$ ethanol yield from Glu</td>
<td>g ethanol/g Glu</td>
<td>0.26 ± 0.01</td>
<td>0.60 ± 0.05</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$Y_{LO}$ lactate yield from Glu</td>
<td>g lactate/g Glu</td>
<td>0.56 ± 0.01</td>
<td>0.91 ± 0.10</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$Y_{IO}$ IB yield from Glu</td>
<td>g IB/g Glu</td>
<td>0.26 ± 0.05</td>
<td>0.03 ± 0.001</td>
<td>0.19 ± 0.04</td>
<td>0.22 ± 0.01</td>
<td>0.36 ± 0.02</td>
</tr>
<tr>
<td>$\alpha_{AE}$ growth-associated acetate synthesis</td>
<td>g acetate/g biomass</td>
<td>0.62 ± 0.02</td>
<td>0.30 ± 0.01</td>
<td>3.0 ± 2</td>
<td>0.35 ± 0.01</td>
<td>0.51 ± 0.03</td>
</tr>
<tr>
<td>$\alpha_{AE}$ growth-associated ethanol synthesis</td>
<td>g ethanol/g biomass</td>
<td>NA</td>
<td>3.7 ± 0.2</td>
<td>40 ± 2</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$\alpha_{LA}$ growth-associated lactate synthesis</td>
<td>g lactate/g biomass</td>
<td>NA</td>
<td>14 ± 1</td>
<td>14 ± 1</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$\alpha_{IB}$ growth-associated IB synthesis</td>
<td>g IB/g biomass</td>
<td>0.58 ± 0.05</td>
<td>0.078 ± 0.01</td>
<td>0.78 ± 0.06</td>
<td>0.06 ± 0.01</td>
<td>3.3 ± 0.1</td>
</tr>
<tr>
<td>$k_{d}$ cell death rate</td>
<td>h</td>
<td>0.001 ± 0.002</td>
<td>0.01 ± 0.002</td>
<td>0.01 ± 0.002</td>
<td>0.02 ± 0.01</td>
<td>0 ± 0.0001</td>
</tr>
<tr>
<td>$k_{g}$ gas stripping rate</td>
<td>h</td>
<td>0.11 ± 0.02</td>
<td>NA</td>
<td>NA</td>
<td>0.11 ± 0.01</td>
<td>0.11 ± 0.01</td>
</tr>
<tr>
<td>$k_{a}$ acetate production from lactate</td>
<td>g biomass/L</td>
<td>0.013 ± 0.001</td>
<td>0.0034 ± 0.0002</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>$k_{ent}$ yeast extract consumption rate</td>
<td>h</td>
<td>NA</td>
<td>NA</td>
<td>0.55 ± 0.03</td>
<td>NA</td>
<td>0.65 ± 0.08</td>
</tr>
<tr>
<td>$\mu_{max}$ specific growth rate with yeast extract</td>
<td>h</td>
<td>NA</td>
<td>NA</td>
<td>0.48 ± 0.03</td>
<td>NA</td>
<td>0.32 ± 0.03</td>
</tr>
<tr>
<td>$\beta$ non-growth-associated IB production</td>
<td>g IB/g biomass</td>
<td>0.002 ± 0.002</td>
<td>0.012 ± 0.001</td>
<td>0.006 ± 0.00</td>
<td>0 ± 0.0</td>
<td>0 ± 0.0</td>
</tr>
</tbody>
</table>

*Model assuming same values for F1, F2, and F3. **Model assuming same values for F2 and F3. ***Model assuming same values for F4 and the F5. NA: not applicable.

and aerobic conditions; F5: rich medium and aerobic conditions.

2.3. Analytical Methods for Biomass and Metabolites. Culture samples were taken after IPTG induction, i.e., ~3 mL of culture was taken from the bioreactor at each time point for metabolite and biomass analysis. Biomass growth was monitored by optical density OD_{600}. There was a linear relationship between the dry cell weight and OD_{600}. To measure dry biomass weight, biomass samples were harvested by centrifugation, washed with water, and dried at 100 °C until the weights remained constant. Glucose, ethanol, acetate, and lactate were measured using enzyme kits (R-Biopharm). Alcohols could be detected using GC (Hewlett-Packard model 7890A, Agilent Technologies, equipped with a DB-MS column, J&W Scientific) and a mass spectrometer (5975C, Agilent Technologies). The GC–MS detected IB, propional, and methyl butanol. The IB concentration was determined by a modified GC–MS method. Briefly, 400 μL of supernatant was extracted with 400 μL of toluene (Sigma-Aldrich) by 2-min vortex, followed by high-speed centrifugation (16000 x g). The organic layer was then removed for GC–MS analysis using the following parameters: hold at 70 °C for 2 min, ramp to 230 °C at 20 °C min⁻¹, and then hold at 300 °C for 6 min. The carrier gas was helium. The MS scan mode was from m/z 20 to 200. Samples were quantified relative to a standard curve of IB concentrations for MS detection, and methanol was taken as an internal standard.

2.4. 13C-Experiments for Analyzing Nutrient Contributions to IB Productions. In the 13C-experiments, the minimal medium with 2% fully labeled glucose (Cambridge Isotope Laboratories) was supplemented with 1 g/l or 5 g/l yeast extract (Bacto). By measuring 13C-abundance in key metabolites from the engineered strains, we estimated the contribution of yeast extract (nonlabeled) to biomass and IB synthesis in the 13C-glucose medium. Specifically, 5 mL of cultures (with 13C-glucose and yeast extract) were inoculated with 5 μL of overnight LB culture of the engineered strain in a 50-mL Falcon tube with a closed cap (shaking at 200 rpm, 30 °C). The cultures (JCL260 or the low-performance strain) were induced by 0.2 mM IPTG (when OD_{600} = ~0.8), and the samples were taken (at t = ~24 h, middle-log growth phase) for isotopomer analysis of IB and amino acids. The two mass-to-charge peaks (m/z = 74 for unlabelled IB and m/z = 78 for labeled IB) were quantified. Their ratio approximately corresponded to the ratio of IB synthesized from unlabeled yeast extract vs labeled glucose. Concurrently, we did isotopic analysis of proteinogenic amino acids to identify the incorporation of unlabeled carbon from yeast extract into biomass protein. The measurements were based on a GC–MS protocol, using TBDMS (N,N-dimethyl-dimethylsilyl)-N-methyl-(trimethylsilyl)amide, Sigma-Aldrich) to derivatize hydrolyzed amino acids from the biomass. The m/z ions (M–57) from unfragmented amino acids were used for analysis except for those of leucine and isoleucine. Because of overlapping ions with (M–57), the [M–159] was used to calculate the isotopomer labeling information of leucine and isoleucine.

2.5. Model Formulation. We developed a kinetic model to describe the fermentation data after IPTG induction. The model contained six time-dependent process variables: X, ACT, LACT, ETOH, IB, and Glu, which represented the concentrations of biomass, acetate, lactate, ethanol, isobutanol, and glucose, respectively. The biomass growth model consisted of glucose-associated (R_{G}) and yeast-extract-associated (R_{YEXP}) terms. IB production was simulated by a mixed-growth-associated production formulation (eq 5), where β was the non-growth-associated IB production rate. In eq 5, the cell death rate; Y_{IL} was the acetate yield from lactate (equal to 0.67 g ACT/g LACT, based on a 1:1 molar ratio); Y_{YD} Y_{EC}, Y_{EO}, and Y_{IB} were the growth-associated glucose yields to biomass, acetate, ethanol, lactate, and IB. E_{G} was the removal
rate of IB due to gas stripping under aerobic fermentation F1, F4, and F5. In F2 and F3, IB loss was minimal (kib was set to zero). A first-order kinetic parameter (kac) was used to describe acetate production from lactate.

\[
\frac{dX}{dt} = R_x - k_{ac}X + R_{X YE}
\]  

(1)

\[
\frac{dACT}{dt} = R_A + Y_{AL}k_{ac} \cdot LACT \cdot X
\]  

(2)

\[
\frac{dLACT}{dt} = R_L - k_{ac} \cdot LACT \cdot X
\]  

(3)

\[
\frac{dEOH}{dt} = R_E
\]  

(4)

\[
\frac{dIB}{dt} = R_{IB} + \beta \cdot X - k_{IB} \cdot IB
\]  

(5)

\[
\frac{dGlu}{dt} = R_G - \frac{R_X}{Y_{SG}} - \frac{R_A}{Y_{SG}} - \frac{R_E}{Y_{SG}} - \frac{R_L}{Y_{LG}} - \frac{R_{IB}}{Y_{IBG}}
\]  

(6)

In eqs 7–12, Rxp, RAc, RB, RAc, and RIB were the production rates of biomass, acetate, ethanol, lactate, and IB from glucose, respectively.

\[
R_x = \frac{\mu_{\text{max}} \cdot \text{Glu}}{K_x + \text{Glu}} \cdot \frac{X}{1 + \frac{ACT}{K_a}}
\]  

(7)

\[
R_A = \alpha_{AX} \cdot R_X
\]  

(8)

\[
R_E = \alpha_{AX} \cdot R_X
\]  

(9)

\[
R_L = \alpha_{AX} \cdot R_X
\]  

(10)

\[
R_{IB} = \alpha_{AX} \cdot R_X
\]  

(11)

\[
R_{X YE} = \mu_{\text{max}} \cdot \text{Glu} \cdot \frac{X}{1 + \frac{ACT}{K_a}}
\]  

(12)

R_X represented a growth model with Monod constant K_a and maximum specific rate coefficient \( \mu_{\text{max}} \). Since acetate inhibited E. coli growth by decreasing the intracellular pH, a non-competitive inhibition K_A was included in the model.\(^{15}\) The dependence of the glucose-based growth rate on oxygen (i.e., aerobic growth vs anaerobic growth) was implicitly included in the calculation of \( \mu_{\text{max}} \) (i.e., the oxygen conditions affected \( \mu_{\text{max}} \)). \( \alpha_{AX}, \alpha_{E}, \alpha_{L}, \alpha_{X}, \) and \( \alpha_{IB} \) were the growth-associated yields of acetate, ethanol, lactate, and IB, respectively. In the rich medium, the yeast extract was quickly consumed to support biomass growth. The model included a yeast-extract-associated biomass growth rate \( R_{X YE} \) using a two-parameter exponential decay function eq 12. Table 2 summarized model parameters and their units.

For each batch culture, unknown parameters were determined by minimizing the sum of the squares of the differences between the model's predictions and the experimentally observed growth and metabolite profiles.\(^{15}\) The "ode23" command in MATLAB (R2009a, Mathworks) solved the differential equations, while the "fincon" command searched suitable values of parameters. To reduce the risk of having local solutions during the nonlinear parameter estimation, we tested the initial guesses for 30 times within
the range of possible values to identify the global solution. To evaluate the quality of the parameter estimates, we checked the sensitivity of the estimated parameters to the measurement inaccuracies. Fifty simulated fermentation data sets (including both biomass and metabolite data) were generated by the addition of normally distributed measurement noise to the fermentation data set (i.e., randomly perturbed the measured data by 30%). The same data-fitting algorithm found new sets of parameters. From the probability distribution of these parameter distributions, standard deviations of model-fitted parameters were estimated.

3. RESULTS AND DISCUSSION

3.1. IB Fermentation Results. The engineered *E. coli* strains in this study employed the Ehrlich pathway (Figure 1), where 2-ketoisovalerate from valine metabolism is redirected to IB synthesis. For the low-performance strain, we simply overexpressed 2-ketoisovalerate decarboxylase and aldehyde reductase. For strain JCL260, both the Ehrlich pathway and 2-ketoisovalerate synthesis pathway were overexpressed. This strain also had gene deletions involved in byproduct formation to increase pyruvate for IB synthesis; thus, it was reported to produce 22 g/L of IB in 112 h.

This study compared IB fermentation kinetics between the two strains. For the low-performance strain, ethanol and lactate were barely detected in the aerobic conditions (Figure 2). IB titer only reached (0.2 g/L) in F4, because the in situ removal of IB was considerable (the airflow carried IB out of the fermentor). Such gas stripping is an effective strategy to avoid the IB accumulation in the culture that causes the inhibitory effect on alcohol production. In oxygen-limited conditions, the F2 generated 0.95 g/L IB, 1.5 g/L ethanol, 2.2 g/L acetate, and
5.1 g/L lactate, while the lactate was reused in the late fermentation stage (stationary growth phase). With the addition of yeast extract, the F3 had fast biomass growth (Figure 4). The cell density reached a peak (2 g DCW/L biomass) after 7 h of IPTG induction, and glucose was consumed within ~12 h (compared to ~40 h in the F1 and F2). The high rates for biomass growth promoted the IB production rate. It took the F3 15 h to generate 0.6 g/L IB, whereas it took F2 40 h to generate same amount of IB. The addition of yeast extract also resulted in a large amount of growth-associated organic acids (6.0 g/L lactate and 3.0 g/L acetate) and thus decreased IB yield from glucose (0.7 g/L IB and 2.0 g/L ethanol from the F3).

A recent paper reported that JCL260 accumulated 50 g/L IB with in situ gas stripping and intermittent additions of 500 g/L glucose solution in an aerobic batch culture, achieving 68% of the theoretical IB yield. Our study performed shake flask cultures and two aerobic fermentations using JCL260. In the complete minimal medium with 20 g/L glucose (F4, Figure 5), JCL260 had very slow biomass growth and low IB productivity. When yeast extract (5 g/L) was supplemented (F5, Figure 6), IB synthesis was significantly improved, and its titer (residual IB) in the bioreactor reached up to 1.7 g/L (10 times higher than the low-performance strain in the aerobic fermentation). Using the rich medium (20 g/L glucose), the closed shake flask culture (without gas stripping) accumulated 4.7 g/L IB, which was ~60% of the theoretical IB yield. Moreover, JCL260 produced only 1 g/L acetate in aerobic fermentations (2 times lower than the low-performance strain) because of the deletion of phosphotransacetylase (pta) gene.

3.2. Kinetic Modeling of IB Fermentation. The same kinetic model simulated fermentation processes by two IB
produced. Table 2 lists the kinetic parameters obtained by nonlinear parameter fitting. For the low-performance strain, the specific growth rate $\mu_{\text{max,app}}$ (0.015 h$^{-1}$) in the oxygen-limited conditions was lower than that in the aerobic culture conditions (0.051 h$^{-1}$). IB could be synthesized in both growth and stationary phases. The oxygen-limited conditions reduced the growth-associated IB yield but promoted nongrowth-associated IB production (e.g., $\beta = 0.012$ g IB/g biomass h in the F2).
Industrial & Engineering Chemistry Research

the presence of yeast extract, the yeast extract-associated biomass growth rate \( \mu_{\text{YAE}} = 0.48 \, \text{h}^{-1} \) was 1 order of magnitude higher than the glucose-associated growth rates. The addition of yeast extract (F3) also improved the biomass yield coefficient \( (Y_{\text{XS}} = 0.20) \) and the growth associated IB production \( (\tau_{\text{IB}} = 0.78 \, \text{g IB/g biomass}) \). Meanwhile, the yeast extract increased yield coefficients of waste products \( (Y_{\text{W}} = 0.77) \) in the F3. The IB yield coefficient \( Y_{\text{IB}} \) was 0.26 g IB/g glucose under aerobic respiration, higher than \( Y_{\text{IB}} \) under oxygen-limited conditions (F2 and F3).

For JCL260, the fermentation data indicated that the strain had a highly inflated Monod constant \( K_{\text{s}} \) (10 g/L), which caused the biomass growth rate to be slower than that of the low-performance strain. The slow growth led to poor IB synthesis in F4 \( (\tau_{\text{IB}} = 0.06 \, \text{g IB/g biomass}) \). Because of the knockout of the \( \pi \) gene to reduce acetate synthesis, the growth-associated acetate production \( \tau_{\pi} \) in F4 was 0.35 g acetate/g biomass, suggesting that acetate production rate was reduced compared to that of the low-performance strain \( \tau_{\pi} = 0.62 \, \text{g acetate/g biomass in the F1} \). On the other hand, JCL260 still generated acetate after \( \pi \) deletion. The alternate acetate pathways in JCL260 had higher glucose-associated acetate yield \( Y_{\text{A}} \) than that of the low-performance strain under aerobic conditions. This observation was consistent with the fact that JCL260 (the strain with multiple gene knockouts) had a poor respiration rate, and thus a higher fraction of glucose was converted to biomass (i.e., \( Y_{\text{G}} \), also increased) and byproducts rather than degraded to CO₂. When yeast extract was added to the growth media, the growth-associated IB production \( \tau_{\text{IB}} \) was 3.3 g IB/g biomass, which was about 5.7 fold higher than that of the low-performance strain. The addition of nutrients improved the JCL260 biomass growth, the cell energy (such as NADH) generation, and the carbon flux through the IB pathway. In contrast, the low-performance strain had a substantial IB pathway. Therefore, yeast extract only enhanced metabolic overflow to waste metabolites rather than improving IB titers (the F3).

Finally, the continuous flow of air into the bioreactor caused an in situ stripping of IB out of the bioreactor (F1, F4, and F3). Since completely trapping IB from off-gas was technically difficult in this study, we used the model to estimate the total IB production by JCL260 (i.e., \( k_{\text{IB}} = 0 \), Figure 6). The model showed that the total IB could reach 5 g/L in F5 if there was no IB loss by gas stripping. The model prediction of the total IB produced was very close to the IB accumulation (4.7 g/L) in the optimized shake flask culture without gas stripping (using the same rich medium, as described in the Experimental Section), which verified the stripping rate constant \( k_{\text{IB}} \) in the model. Therefore, our model is potentially useful for describing the consolidated bioprocess of simultaneous IB production and recovery.

3.3. Analysis of the Role of Yeast Extract for IB Synthesis. Nutrient supplements play an important role in improving fermentation performance. Rich media have been commonly used for butanol fermentations. In addition to providing the building blocks for biomass growth, \( \text{E. coli} \) can also utilize the Ehrlich pathway to convert protein hydrolysates to higher alcohols. However, the contribution of rich nutrient (yeast extract) to IB production was not quantified. Here, we used \(^1\)C experiments to determine the ratio of carbon utilization from two different sources (nonlabeled yeast extract vs fully labeled \(^1\)C-glucose) under oxygen-limited conditions via GC-MS analysis (Figure 7). For the low-performance strain cultivation with 1 g/L yeast extract, its proteinogenic amino acids (e.g., histidine, leucine, isoleucine, lysine, and proline) were highly imported from exogenous amino acids (>50%, corresponding to the \(^1\)C-dilutions), while IB was mostly labeled with four carbons (m/z = 78, IB came from labeled glucose). When excess yeast extract (5 g/L) was provided, the low-performance strain not only used yeast extract as the building blocks for cell growth but also converted it to IB (<50% IB was nonlabeled). On the other hand, with sufficient yeast extract (5 g/L), JCL260 still mainly used \(^1\)C-glucose for IB production (labeled IB was >90%). In the rich media, JCL260 highly utilized yeast extract for biomass synthesis. It showed much higher \(^1\)C-labeling concentration (~20%) in valine than the low-performance strain (~5%). Higher abundance of \(^1\)C-labeling in valine proved that the overexpression of the keto-acid pathway in JCL260 efficiently enhanced the \(^1\)C-glucose flux toward 2-ketosovalerate (the common precursor for both IB and valine) and reduced the relative valine uptake from the rich media.

Figure 7. Fraction of \(^1\)C carbon in metabolites from the low-performance (A) and JCL260 (B) IB-producing strains. The biomass was grown on fully labeled \(^1\)C-glucose, with 1 g/L (black bar) or 5 g/ L (gray bar) nonlabeled yeast extract (n = 2; GC–MS standard errors <2%). The \(^1\)C fractions (R) of metabolites were based on the following equation:

\[
R = \frac{1}{\pi} \sum_{x=0}^{\pi} (X_M)
\]

where \( \pi \) was the total carbon number of the metabolite (0 ≤ X ≤ \( \pi \)). \( M_x \) was the corresponding \(^1\)C fraction for each isotopomer (\( M_x \) was unlabeled fraction, \( M_x \) was singly labeled fraction, \( M_x \) was doubly labeled fraction, etc.).
4. CONCLUSIONS
This study developed a general empirical model for IB fermentations by two engineered E. coli strains. The model with nonlinear fitted parameters reasonably well described IB accumulation and removal under denoted cultivation conditions. The model results indicated that the two strains displayed a difference in biomass growth behavior and products generation. The comparative study revealed the change of influential kinetic variables in response to the cultivation conditions. Moreover, we quantified the contribution of nutrient sources to product yields via isotopic investigation and proved that the keto-acyl pathway was a rate-limiting step for IB production in the low-performance strain. This study may serve as a springboard for developing useful bioprocess models for fermentation of higher alcohols in the biotechnology industry.

■ AUTHOR INFORMATION
Corresponding Author
*Tel.: 314-935-3441. Fax: 314-935-7211. E-mail: yinjie.tang@seas.wustl.edu.

Author Contributions
Y.X., X.F., and A.M.V. contributed equally to this work.

Notes
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS
This study was supported by an NSF Career Grant (MCB0954016) and the Clean Coal Consortium at Washington University. We thank Professor James Liao for the offer of the JCL260 strain for this study. We also thank Chunlei Mei for her technical assistance.

■ ABBREVIATIONS:
ACAc, acetyl-CoA; AdhA, aldehyde reductase; ALA, 2-acetolactate; DHL, 2,3-dihydroxyisovalerate; G3P, glyceraldehyde-3-phosphate; G6P, glucose-6-phosphate; Glu, glucose; IB, isobutanol; IBA, isobutanol; KIV, 2-ketovalerate; Kivd, 2-ketoisovalerate decarboxylase; P72, pyruvate; Val, valine; PP pathway, pentose phosphate pathway; TCA cycle, tricarboxylic acid cycle

■ REFERENCES
Appendix Chapter 4: An ancient Chinese wisdom for metabolic engineering:

Yin-Yang
An ancient Chinese wisdom for metabolic engineering: Yin-Yang

Stephen G Wu, Lian He, Qingzhao Wang and Yinjie J Tang

Abstract
In ancient Chinese philosophy, Yin-Yang describes two contrary forces that are interconnected and interdependent. This concept also holds true in microbial cell factories, where Yin represents energy metabolism in the form of ATP, and Yang represents carbon metabolism. Current biotechnology can effectively edit the microbial genome or introduce novel enzymes to redirect carbon fluxes. On the other hand, microbial metabolism loses significant free energy as heat when converting sugar into ATP, while maintenance energy expenditures further aggravate ATP shortage. The limitation of cell 'powerhouse' prevents hosts from achieving high carbon yields and rates. Via an Escherichia coli flux balance analysis model, we further demonstrate the penalty of ATP cost on biofuel synthesis. To ensure cell powerhouse being sufficient in microbial cell factories, we propose five principles: 1. Take advantage of native pathways for product synthesis. 2. Pursue biosynthesis relying only on pathways or genetic parts without significant ATP burden. 3. Combine microbial production with chemical conversions (semi-biosynthesis) to reduce biosynthesis steps. 4. Create 'minimal cells' or use non-model microbial hosts with higher energy fitness. 5. Develop a photosynthesis chassis that can utilize light energy and cheap carbon feedstocks. Meanwhile, metabolic flux analysis can be used to quantify both carbon and energy metabolisms. The fluxomics results are essential to evaluate the industrial potential of laboratory strains, avoiding false starts and dead ends during metabolic engineering.

Keywords: ATP, Energy metabolism, Flux analysis, Free energy, Maintenance loss, Semi-biosynthesis

Introduction
In the past decade, molecular biology tools have been developed rapidly and now offer new opportunities for metabolic engineering of microbial hosts [1-6]. These tools include the selection of plasmids with different copy numbers, promoter engineering, codon optimization, synthetic scaffolds, directed evolution or rational design of enzymes, ribosome binding sites editing, and competitive pathways deletion. Advanced genome engineering (e.g., CRISPRs and TALENs) and automation of conventional genetic techniques (e.g., MAGE) provide efficient capabilities for editing genomes and evolving new functions. At the same time, systems biology (e.g., genomics, transcriptomics, and proteomics) can characterize complex cell networks, mine useful genes, discover new enzymes, reveal metabolic regulations, and screen mutant phenotypes. The advent of these powerful tools seems to lead researchers into a new epoch of bioprocess industries using GMMs (genetically modified microorganisms) in the near future. However, that is not the whole story.

The golden age of industrial biotechnology dawned in the early 1940s, driven by the mass production of penicillin and enjoyed a fast growth in the 1950s – 1980s. Microbial bioprocess has produced diverse commodity chemicals (such as ethanol, amino acids, citric acid, and lactate) as well as recombinant proteins and antibiotics in the last century. Those commercial products mainly rely on natural strains or strains with minor genetic modifications (usually only one or few new genes). Since the recent decade, in the hope of producing chemicals at low costs and reducing greenhouse gas emissions, an enormous amount of investment has been devoted to metabolic engineering in many nations. Although modern biotechnologies can engineer microbial platforms to synthesize diverse products in laboratories, there are only a few GMM products that have become commercially promising in the past decade (e.g., artemisinic acid and 1, 4-butanediol). Novel GMMs are also used for chemical manufactures, such as short-chain alcohols and isoprene
[7-9]. Recently, Gevo and Butamax introduce the keto-acid/Ehrlich pathway into yeasts to produce isobutanol [10]. Amyris extend the mevalonate pathway in Saccharomyces cerevisiae for branched and cyclic terpenes (e.g., farnesene) synthesis. However, these companies have not achieved strong net profit margin yet. To date, the industrial-scale biofuel is still ethanol, which is cheaply manufactured from sugar cane in Brazil. In this perspective, we address one of the hidden constraints in microbial cell factories (i.e., energy metabolism).

**The energy losses in microbial cell factories**

Heterotrophic organisms obtain free energy in the form of ATP by breaking organic substrates into CO₂ (Figure 1). Theoretically, oxidation of one mole of glucose to CO₂ ($\Delta_{\text{f}}H^\circ_{\text{reaction}} = -2.8$ MJ/mol) can generate 38 moles of ATP. Hydrolysis of these ATP to ADP ($\Delta_{\text{f}}H^\circ_{\text{reaction}} = -30.5$ kJ/mol) provide $\sim 1.2$ MJ of biochemical energy. Thereby, $\sim$60% of energy from glucose is lost as heat during ATP synthesis (similar to a Carnot heat engine). Besides, cell consumes ATP for diverse maintenance activities, such as nutrient/

Microbial hosts have not evolved towards optimal energy metabolism. Over billions of years of evolution, microbes with a higher growth rate gained a selective advantage when competing for shared energy resources, but these fast growing species have a lower yield of ATP from substrates (e.g., less than 30 ATP/glucose) [12]. The oxidative phosphorylation (P/O) ratio represents ATP generation efficiency through substrate oxidation. Theoretically, three ATP can be obtained from the reduction of one oxygen atom (i.e., P/O = 3) during oxidative phosphorylation. Although slow-growing mammalian cells can achieve P/O values close to 3, bacteria and yeasts often have P/O ratios below 2.5 (note: microbes may dissipate the proton gradient before it can be fully used for charging the ATP synthase). In addition, microbial hosts may lose ATP yield due to byproducts synthesis, membrane leakage, removal of reactive oxygen species, or suboptimal cultivations (insufficient mixing, shear stress, or biofilm formation). Lastly, the electron transport chain for ATP generation and nutrient transporters may compete for membrane and intracellular spaces so that the capacity of the microbial powerhouse cannot be easily upgraded [13,14].

We introduce a terminology "metabolic entropy" to define the free energy in the substrates that is lost through energy metabolism and becomes unavailable for biosynthesis. Metabolic entropy has gained attention from metabolic flux analysis researchers because the objective function of biomass production in FBA (flux balance analysis) always overestimates microbial growth rates. Moreover, FBA predictions highly depend on the assumption of a fixed ATP maintenance coefficient. To address this problem, researchers developed 13C-metabolic flux analysis (MFA) to quantify the microbe metabolic entropy directly via tracer experiments. By examining Bacillus subtilis mutants, 13C-MFA has discovered that the suboptimal cell metabolism is associated with the increased energy usage in the face of environmental and random genetic perturbations [15]. This study suggests that mutating regulatory genes can drive carbon flow towards the desired pathways; however, hijacking carbon fluxes may sacrifice cell energy fitness for adaptive responses under adverse environmental conditions.

The tradeoff between product yield and energy fitness

Traditional metabolic engineering uses plasmids and heterologous enzymes to redirect carbon fluxes. Early studies have shown high copy number plasmids cause significant alterations in cell properties and strongly influence metabolic engineering endeavors [16]. 13C-MFA of E. coli strains revealed higher acetate production and O2 uptake rates in plasmid-containing strains than in the plasmid-free strains [17]. The presence of plasmids can increase cell maintenance, decrease growth rate and change intracellular fluxes, especially suppressing the oxidative pentose phosphate pathway [18]. Similarly, synthetic biology parts (such as novel pathways, protein scaffolds, and genetic circuits) may also increase metabolic entropy if extra nucleic acids and proteins are required to be made by the hosts (note: elongation of one amino acid costs four ATP molecules) [19]. Natural microbes have frugal enzymatic machinery (each native enzyme in a single E. coli cell may only have dozens of molecule copies and places minimal biosynthesis burden on cell metabolism) [20]. During pathway engineering, a large portion of over-expressed enzymes may be inactivated due to protein misfolding. Considerable ATP expenditure for heterologous enzyme over-synthesis can trigger stress responses. For example, 13C-MFA has been used to examine metabolic burdens in E. coli during biosynthesis of recombinant proteins. The results indicate a 25% increase in the total ATP expenditure rate in the highest yielding strain (up to 45 mmol ATP/g CDW/h) [21]. To overcome such an energy limitation, E. coli has to reduce biomass synthesis and enhance oxidative phosphorylation for ATP generation. Besides, engineered microbial hosts often suffer from increased non-growth associated maintenance as well as reduced respiration efficiency (poor P/O ratio) due to membrane stresses [22,23]. If an extended heterologous pathway causes deleterious effects on carbon and energy metabolism, the host will lose the capability to grow in a minimal carbohydrate medium. In this case, rich nutrients, such as yeast extract (producing 1 g of yeast extract consumes >2 g of glucose), have to be supplied to relieve the cell's metabolic burden [24].

Our theory of energy burden can guide strain development to tolerate product stresses. For instance, an isobutanol-tolerant mutant has been isolated after serial transfers; while the final isobutanol productivity of this evolved strain did not show improvement [25]. The export systems (e.g., ABC transporters) have been engineered for recovering cell growth under biofuel stresses [26], while ATP-driven efflux pumps show limited enhancement of short-chain alcohol productivity (~10%) [27]. Interestingly, efflux pumps are very effective when they are introduced into low-performance strains, in which their product titers are well below 1 g/L [26]. These observations explain the fact that cell stress adaptation requires ATP expenditure and induce energy burdens [28]. For the same reason, tolerance engineering often works well on yeast strains for ethanol production because of simple ethanol synthesis pathway and net ATP
generation from glycolysis. For example, engineering transcriptional machinery or up-regulation of the potassium/proton pumps in Saccharomyces cerevisiae can improve both ethanol tolerance and the production titer (>100 g/L) [29,30]. In conclusion, if microbial hosts already have high metabolic burdens, tolerance engineering should focus on regulatory components rather than efflux pumps. For example, a methionine biosynthesis regulator can significantly improve both biofuel tolerance and productivity in Escherichia coli [27]. In yet another case, the inactivation of a histidine kinase may enhance the butanol productivity in Clostridium acetobutylicum by delaying cell sporulation [31].

**Sensitivity analysis of energy penalty on biofuel synthesis**

We employ a genome-scale flux balance model (iJO1366) to simulate the adverse impacts of E. coli energy metabolism on biofuel product yields (Figure 2) [32]. Apart from the intracellular stress caused by enzyme overexpression, the release of large amounts of biofuel molecules (alcohol or fatty acid) will interfere enzymatic reactions in vivo and disrupt the cellular membrane’s integrity, which results in reduced efficiencies of oxidative respiration [25,33]. Thereby, metabolic engineering approaches are effective in redirecting carbon fluxes to biosynthesis only in these low-productivity strains whose energy metabolism are not overloaded. We use FBA to test the penalty of metabolic burdens (such as maintenance cost) and the decrease of P/O ratio on biofuel yields. The simulations show that microbial energy metabolism is usually abundant so that they can support certain amount of metabolic burdens without having apparent biosynthesis deficiency (e.g., without showing a slower growth after mutations). However, cell burden may increase during the routine genetic modifications. When cell powerhouse is unable to afford the increasing ATP expenditure, the biosynthesis yield will have a sudden drop (i.e., “the straw that broke the camel’s back”), forming a “cliff” in Figure 2.

FBA simulations yield two insights into microbial biofuels. First, alcohol (ethanol and isobutanol) producing E. coli strains not only have higher carbon yields (0.67 C-product/C-glucose), but also are insensitive to P/O ratios (Figure 2a, b). Comparing to isobutanol, ethanol production is less sensitive to the metabolic burden (larger energy sufficient zone). Ethanol fermentation, an ancient bioprocess from the beverage industry, does not need additional energy from O₂, lowering its process costs. From a stoichiometric perspective, glycolysis generates two net ATP per glucose, which fulfills the cell energy expenditure. In addition, ethanol synthesis only needs a few native enzymes, and the hosts (e.g., Saccharomyces cerevisiae) are naturally tolerant to alcohols. The entire ethanol synthesis

![Figure 2 Genome-scale FBA models for predicting microbial biofuel yields from glucose.](image-url)

**Figure 2:** Genome-scale FBA models for predicting microbial biofuel yields from glucose. a. E. coli strains produce ethanol (growth rate = 0.05 h⁻¹). b. E. coli strains produce isobutanol (growth rate = 0.05 h⁻¹). c. E. coli strains produce fatty acid (growth rate = 0.05 h⁻¹). d. E. coli strains produce fatty acid (growth rate = 0.05 h⁻¹). We use an E. coli FBA model (iJO1366) to predict production of different biofuels from glucose. Alcohol production is simulated under the aerobic condition (O₂, influx S 1.85 mmol/gDW/hr), while fatty acid is under anaerobic condition (O₂, influx S 1.2 mmol/gDW/hr). The medium conditions and glucose uptake rate (1.8 mmol/gDW/hr) are same for all FBA. Extra metabolic burden is simulated by the costs of both protein overexpression and maintenance energy increase (e.g., 10% extra metabolic burden is equivalent to 10% overexpression of total biomass protein plus proportional increase of non-growth associated ATP loss). For each case, the objective function is set as to maximize the biofuel production. Abbreviations: DW (Dry Weight), FA (Fatty acid), Gic (Glucose), Bi (Isobutanol).
pathway is always inside of the cytosol, and thus they do not have mitochondrial transport limitations. These advantages explain why ethanol fermentation is superior to any other biofuel processes.

Second, energy metabolism may become a critical issue for synthesizing fatty acid-based compounds, which are susceptible to changes in P/O ratio, ATP maintenance loss, and oxygen uptake fluxes. Compared to alcohol production, fatty acid based fuels (such as biodiesel) require longer biosynthetic pathways (more enzymes to overexpress) and considerable ATP usage for product synthesis [34]. Besides, many enzymes in fatty acid pathway are tightly regulated during cell growth, leading to growth associated bioproduction. The simultaneous biomass growth and fatty acid synthesis further slows ATP shortage [35]. Therefore, aerobic fermentation has to be performed to enhance energy metabolism, which reduces product yield and increases the fermentation costs for aeration. Furthermore, the accumulation of fatty acid damages cell membrane and reduces oxidative phosphorylation efficiency. To demonstrate these synergistic effects on fatty acid yields, Figure 2c and d simulate E. coli fatty acid yields responding to P/O ratios and metabolic burdens. As shown in Figure 2c, fatty acid production can achieve a similar yield as ethanol if the host’s biomass growth rate is 0.05 hr⁻¹ and energy maintenance is not high. In reality, fatty acid yield can drop to 50% or less of the theoretical maximum, which is in consistent with the model prediction if we considered a practical biomass growth, extra ATP maintenance, and a low P/O ratio (< 1.5) in FBA (blue star in Figure 2d) [36]. Figure 2d also indicates the high sensitivity of fatty acid yield in response to the P/O ratio (red star in Figure 2d). For instance, one unit change in P/O ratio leads to an abrupt drop in fatty acid yield – from a theoretical maximum to zero.

**Yin-Yang theory in metabolic engineering**

To better understand the limitations of microbial cell factories, we refer to an ancient Chinese philosophy: Yin-Yang. Yin-Yang describes both the bright side and dark side of an object in the world. Yin and Yang oppose each other but are also interdependent. In the case of metabolic engineering, the microbial metabolism is operated by thousands of enzymatic reactions and mass transport processes that involve both carbon (Yang) and energy (Yin) transformations (Figure 1). Through billions of years of evolution and environmental adaptations, biological systems have evolved closely interdependent carbon fluxes for biomass growth and energy fitness, which are similar to the intertwined Yin-Yang forces. Although it is easy to engineer microbial hosts to produce small amounts of diverse products, manufacturing a particular compound with titers and rates beyond the economic break-even point could be limited by suboptimal energy metabolism. In microbial conversions of a substrate to a product, metabolic entropy increases when carbon flux is redirected to the final products (Figure 3a & b). For example, Figure 3c shows the energy loss during conversion of glucose to different biofuels.

To leverage the “Yin-Yang” balance, metabolic engineers tried a few practical approaches to promote energy metabolism and boost productivity. For instance, *Vitreoscilla* hemoglobin (VHb), a soluble bacterial protein, has been used to enhance energy metabolism by improving oxygen delivery, which can significantly improve cell growth and increase chemical production under oxygen-limited conditions [39]. Furthermore, an energy-conserving pathway in E. coli was developed through metabolic evolution for high production of succinate from glucose fermentation [40]. This study shows that inorganic phosphate or phosphoenolpyruvate carboxykinase increases the net production of ATP, compared to the primary mixed acid fermentation pathway via PEPC carboxylase. The extra energy supply allows E. coli to produce succinate close to the theoretical maximum. In another case, an ATP-consuming reaction was introduced into *S. elongatus* PCC 7942 to drive carbon flux from acetyl-CoA to 1-butanol [41]. This study of 1-butanol production further validates that the ATP coupling reaction can make engineered pathways thermodynamically more favorable. To this end, we summarize the following suggestions to overcome the energy roadblocks.

First, a clear understanding of the entire carbon and energy metabolisms in microbial species would help us to conquer the energy limitations. Using E. coli as an example, ATP significantly impacts the product distributions at the pyruvate node [42]. Understanding ATP fluxes can offer rational design for E. coli strains for improving product biosynthesis [40,43]. Flux balance analysis (FBA) and 13C-metabolic flux analysis (MFA) are the only available tools that can quantify energy expenditures. FBA can characterize cell energy metabolism by dividing ATP cost into non-growth associated loss and growth-associated maintenance [22]. Due to the metabolic nature of suboptimal carbon fluxes, FBA, relying on the objective functions, may overestimate the cell potential for biosynthesis capability. 13C-MFA uses tracer experiments to constrain the FBA model so that it can precisely measure enzyme reaction rates. 13C-MFA can profile carbon fluxes through all energy generation/consumption pathways and deduce energy flows in the cell metabolism (ATP and co-factor balancing) [35]. Flux analysis not only allows us to determine the hidden Yin-Yang balance and to design rational engineering strategies, but also to characterize metabolic entropy and identify a strain’s energy potential for further improvement. Although 13C-MFA has not been widely accepted as a routine laboratory measurement tool to assess the engineered microbial hosts, this technology has excellent
Figure 3 Energy fitness and productivities in microbial cell factories. 

a. The trend of metabolic entropy changes (unit: ATP generation per glucose). In optimal metabolism, one mole glucose generates 38 ATP for biosynthesis. Under constraints of P/O ratios and maintenance loss, less ATP can be generated (increase of metabolic entropy). 

b. The transition from carbon limitation to energy limitation with the increase of product yield. In many cases, the energy limitation prevents strains from achieving the yield and titer above break-even point. 

c. Cascade of energy changes (heat of combustion) during biofuel synthesis from glucose. Energy was calculated based on stoichiometry yields (green text, entire bar) and the practical yields (red text, grey bar). Reported yields: ethanol – 96% of theoretical yield [20], isobutanol – 85% of theoretical yield [37], fatty acid – 50% of theoretical yield [35], and H2 (dark fermentation) – 50% of theoretical yield [38].
potential to reveal pathway engineering burdens (i.e., predict “the last straw” in genetic modifications). This tool can informatively tell metabolic engineers and project sponsors what can be done and what cannot be done.

Second, metabolic engineers need to exploit native pathways and avoid extensive pathway reconstruction. In history, many industrial successful cases of improved strain tolerance or productivity just relied on random mutation or evolution, leveraging Natural Selection of mutants for the best ‘Yin-Yang’. Additionally, efforts should aim product synthesis at pathways that do not require significant ATP expenditures (such as ethanol or organic acids). For example, the acetate overproduction pathway in E. coli generates abundant ATP and the engineered strain performs very well even when its oxidative phosphorylation, TCA cycle and competing fermentation pathways are disrupted [43]. When microbial hosts have low-burden biosynthesis pathways, they show robustness in industrial processes. Moreover, artificial synthetic circuits, efflux pumps, or novel pathways should be carefully considered in terms of the energy penalty. By revealing the tradeoffs behind synthetic biology parts via flux analysis approach, metabolic engineers can rationally design their engineering strategies.

Thirdly, although it is difficult to break the Yin-Yang balance in a natural microorganism, synthetic biologists may re-program the carbon metabolism and energy “fitness” by engineering novel microbial systems. Metabolic engineers often apply pathway overexpression to improve the strain productivity. These practices typically encounter adverse metabolic shifts due to energy imbalance. However, the creation of a “minimal or smart” cell can remove unnecessary genes in microbial hosts in effort to reduce cell burden and unlock the biosynthesis regulations [44,45]. Additionally, synthetic biologists try to design and assemble cells using synthetic chromosomes [46]. These artificial biological systems might not necessarily follow the natural Yin-Yang balance evolved over billions of years, so they could have an unusually efficient energy metabolism, and thus achieve product yields close to the theoretical maximum.

Fourth, biological conversion can be integrated with non-living processes to reduce the biosynthesis burden. We can use robust microbial hosts to make simple molecules with high yields and titers, and then convert these molecules into a desired product with a complicated structure via biological and chemical processes. For example, the Keasling Lab achieved the total synthesis of artemisinin with a two-stage semi-synthetic approach. They used the mevalonate pathway in Saccharomyces cerevisiae to synthesize artemisinic acid, followed by a four-step chemical conversion of artemisinic acid to artemisinin [47]. The Zhang Lab has made biopolymers by using engineered E. coli as a first step, to produce a simple molecule mevalonic acid, and then chemically converting it into biopolymers [48]. A significant advantage of these integrated processes is an extremely efficient biocconversion in the first step using a short microbial pathway [49]. For instance, the titer of the semi-product mevalonic acid can reach as high as 88 g/L because its synthesis only requires three steps from the central metabolic node (acetyl-CoA) [48]. In another and more radical approach, an artificial cell-free system containing enzyme cocktails can mimic one or many functions of a biological system. Such systems can be used to synthesize products with near maximum theoretical yields [50,51] since they have no cell maintenance cost.

Lastly, development of non-model microbial workhorses with desired traits in energy metabolism (e.g., photosynthesis) may achieve higher biosynthesis potentials, enabling the design of industrial bio refineries for the production of a broad range of products. In fact, even in the modern era of genomics, it is estimated that > 99% of all bacterial species remain unknown [52]. Some non-model species might have a unique energetics that can facilitate product synthesis. For example, Algenol is developing the engineered cyanobacteria for phototrophic ethanol production from CO2 (http://www.algenol.com/). Moreover, cyanobacterial species have shown faster growth and higher production rate/titer by co-utilization of organic substrates [53]. Cyanobacterial photofermentations, using cheap feedstocks, CO2 and light energy, may facilitate cost-effective and large-scale bio refineries. In fact, Nature is the best synthetic biologist and may have already prepared us excellent chassis that we have not discovered yet. When we try to out-do Nature’s performance, we must first assimilate her lessons of ‘Yin-Yang’.

Conclusions
We have discussed the Yin-Yang concept as the underlying regulatory mechanism in cell metabolism. Biosynthesis of diverse useful products requires sophisticated genetic pathway engineering to steer a high flux to the final product while energy fitness requires the cell metabolism to be wisely changed. Since the powerhouse in microbial cell factory is not limitless, energy shortage eventually leads to metabolic shifts and reduced cell productivity in engineered microbes. The Yin-Yang balance may caution against the assumption that the host metabolism can be modified extensively to produce any desired products. By using fluxomics, we can formulate guidelines to avoid many false starts and dead ends during metabolic engineering. In addition, industrial bioprocess always faces numerous constraints and trade-offs (mass transfer limitations in fermentation, sterilization, strain stability, contaminations, and aeration costs). Feedstock selections, downstream product separation, and waste treatment are critical issues that impact product profitability. Thus, the design-build-
test-learn cycle should cover both strain development and economic analysis. Nevertheless, the Yin-Yang philosophy provides general insights into all biotechnology tradeoffs.

Competing interests
The authors declare that they have no competing interests.

Authors’ contributions
SGW and YJT conceived the initial idea. SGW carried out the modeling studies and reference collections. SGW, JA, QIN, and YJT wrote the paper. All authors read and approved the final manuscript.

Acknowledgement
The authors would like to express their gratitude to NSF Grants (DB: 1356669 and MCB0954016) and DOE (DESC0012722) for their kind support on fluvoxamine research. We also appreciate the writing assistance by Yunci Qi and Mary Abernathy.

Author details
Department of Environmental Engineering and Chemical Engineering, Washington University, St. Louis, MO 63130, USA. 2Fine Chemicals & Biocatalysis Research, BASF Corporation, Tarrytown, NY 10591, USA.

Received: 3 November 2014 Accepted: 2 March 2015
Published online: 20 March 2015

References
43. Cueney TB, Zhou S, Sharmannog KT, Ingram LD. Engineering the metabolism of Escherichia coli W3110 for the conversion of sugar to...
redox-neutral and oxidized products: Homoeacetate production.

44. Fong TS, Church GM. Towards synthesis of a minimal cell. Mol Syst Biol. 2006;2:245.


Appendix Chapter 5: Elucidation of intrinsic biosynthesis yields using $^{13}$C-based metabolism analysis
Elucidation of intrinsic biosynthesis yields using $^{13}$C-based metabolism analysis

Arul M Varman, Lian He, Le You, Whitney Hollinshead and Yinjie J Tang

Abstract
This paper discusses the use of $^{13}$C-based metabolism analysis for the assessment of intrinsic product yields — the actual carbon contribution from a single carbon substrate to the final product via a specific biosynthesis route — in the following four cases. First, undefined nutrients (such as yeast extract) in fermentation may contribute significantly to product synthesis, which can be quantified through an isotopic dilution method. Second, product and biomass synthesis may be dependent on the co-metabolism of multiple-carbon sources. $^{13}$C labeling experiments can track the fate of each carbon substrate in the cell metabolism and identify which substrate plays a main role in product synthesis. Third, $^{13}$C labeling can validate and quantify the contribution of the engineered pathway (versus the native pathway) to the product synthesis. Fourth, the loss of catabolic energy due to cell maintenance (energy used for functions other than production of new cell components) and low P/O ratio (Phosphate/Oxygen Ratio) significantly reduces product yields. Therefore, $^{13}$C-metabolic flux analysis is needed to assess the influence of suboptimal energy metabolism on microbial productivity, and determine how ATP/NAD(P)H are partitioned among various cellular functions. Since product yield is a major determining factor in the commercialization of a microbial cell factory, we foresee that $^{13}$C-isotopic labeling experiments, even without performing extensive flux calculations, can play a valuable role in the development and verification of microbial cell factories.

Keywords: Cell maintenance, Co-metabolism, Metabolic flux analysis, P/O ratio, Yeast extract

Introduction
Recent advances in metabolic engineering have enabled us to engineer microbial cell factories for the efficient synthesis of diverse products, including bulk chemicals, pharmaceutical drugs and biofuels [1,2]. For example, advanced biofuels produced by engineered microorganisms with properties similar to that of petroleum-based fuels, have been reported extensively [3-7]. The emergence of systems biology and synthetic biology has greatly increased the potential of microbial cell factories towards the production of value-added chemicals [8-10]. For economically viable manufacture of bulk and commodity chemicals [11], the product yield is an important consideration. Researchers often include either rich medium or multiple feedstocks in microbial fermentations. Thereby, estimation of the intrinsic product yield is difficult since undefined nutrients may also contribute to the product synthesis (Figure 1). Additionally, new enzymes are often employed to improve microbial productivity [4,12-14], and the separate contributions of the heterologous and native pathways to product synthesis need further validation. Finally, the synthesis of high-energy products (such as biofuels) requires a large amount of ATP and NAD(P)H. Due to suboptimal energy metabolism (e.g., cell maintenance cost), the actual bacterial biosynthesis is often at least three-fold lower than the amount that would be predicted from reaction stoichiometry [15].

Therefore, $^{13}$C-analysis is the recommended method to track the in vivo carbon fluxes from specific substrates to final products. Feeding microbial cultures with $^{13}$C-substrates results in unique isotopic patterns amongst the cell metabolites ($^{13}$C-fingerprints) [16] to delineate metabolic pathways [17]. Integration of $^{13}$C-fingerprints with metabolic modeling can elucidate the intracellular metabolic fluxes (i.e., $^{13}$C-MFA). In the biotechnology field, $^{13}$C-MFA can reveal metabolic responses of microbial hosts to product synthesis and growth conditions [18-20], identify the rigid metabolic nodes that cause
bottlenecks for further rational pathway engineering [21], and perform characterization of novel microbial physiologies [22-25]. In addition to these applications, \(^{13}\)C-MFA may reveal the effect of suboptimal energy metabolism on intrinsic product yields.

**Product yield using rich medium**

Engineered microbes have many metabolic burdens that can inhibit both biomass growth and product synthesis. Since rich media includes both primary carbon substrates (e.g., sugars) and large amounts of nutrients (such as yeast extract), it is commonly used in fermentations to provide diverse nutrients for cell growth and stabilize the production performance of microbes [9,10]. This reduces the culture lag phase and promotes productivity. Multiple studies have revealed that supplementing culture medium with yeast extract or terrific broth — a highly enriched medium that contains yeast extract, tryptone and glycerol as carbon sources — to engineered microbes significantly improves their final biosynthesis yields [26,27]. Since nutrient supplements can provide undefined building blocks for both biomass and product synthesis, it is difficult to precisely calculate the intrinsic product yield from rich-medium fermentation. To overcome this problem, \(^{13}\)C-analysis can gain insights into the carbon contribution from the nutrients to product biosynthesis.

For example, two *E. coli* strains engineered for isobutanol production (i.e., a low performance strain with an Ehrlich pathway [28] and a high performance ICL260 strain with overexpression of both the keto-acid pathway and the Ehrlich pathway [29]) display an increase in isobutanol titer with the inclusion of yeast extract in their culture medium. Using fully labeled glucose and non-labeled yeast extract as carbon sources, \(^{13}\)C-experiments revealed that the low-performance strain derived ~50% of the carbons in the produced isobutanol from yeast extract (Figure 2). On the other hand, ICL260 synthesized isobutanol solely from \(^{13}\)C-glucose and used yeast extract mainly for biomass growth [28]. This observation confirms that overexpression of the keto-acid pathway overcomes bottleneck in the synthesis of isobutanol and effectively pulls the carbon flow from glucose to product. In another work, an E. coli strain was engineered for the conversion of acetate into free fatty acids via the overexpression of both acetyl-coA synthetase and the fatty acid pathways. During acetate fermentation, yeast extract significantly promoted fatty acid productivity, resulting in 1 g/L fatty acids from ~10 g/L acetate [30]. \(^{13}\)C-analysis of the culture with fully labeled acetate and yeast extract has shown that ~63% carbons in the free fatty acids were synthesized from \(^{13}\)C-acetate (Figure 2). Thereby, the intrinsic product yield from a primary substrate in a rich medium could be correctly estimated based on isotopomer analysis.

**Product yield during co-metabolism of multiple carbon substrates**

Algal species are able to utilize both CO\(_2\) and organic carbon substrates. Such mixotrophic metabolisms can alleviate the dependence of algal hosts on light and CO\(_2\) limitations, and thus enable them to achieve high biomass growth rate and product titer [31]. \(^{13}\)C-metabolite analysis has been used to track their photomixotrophic metabolisms in different scenarios. For example, *Synechocystis* sp. PCC 6803 (blue-green algae) is capable of performing photomixotrophic growth. \(^{13}\)C-MFA has shown that CO\(_2\) contributes to 25% of *Synechocystis* biomass yield during its mixotrophic growth with \(^{13}\)C-glucose and \(^{13}\)CO\(_2\) [32]. On the other hand, \(^{13}\)C-analysis has tracked D-lactate synthesis in an engineered *Synechocystis* 6803 [33]. In that study, the lactate production increased substantially during the co-metabolism of both CO\(_2\) and
acetate. Experiments with fully labeled acetate and $^{13}$CO$_2$ determined that nearly all of the lactate molecules were non-labeled and that only the acetyl-CoA-derived proteinogenic amino acids (leucine, glutamate and glutamine) were $^{13}$C-labeled. This observation suggests that acetate entered into TCA cycle and was involved only in biomass growth, while the yield of D-lactate was completely derived from CO$_2$ (Figure 2). This result further indicates that acetate could inhibit the pyruvate decarboxylation reaction and thus direct more carbon flux from pyruvate to lactate. The above study shows the value of $^{13}$C-analysis in improving our understanding of pathway regulations for product synthesis. Since many microbial platforms (including both algal species and heterotrophs) may co-metabolize multiple carbon substrates simultaneously, isotopomer feeding can reveal the contributions of each substrate to the corresponding metabolite pools, and thus predict the potential bottlenecks in biomass or product formations.

**Product yield from alternative pathways**

$^{13}$C-analysis can decipher the yield of products with multiple biosynthesis routes. For example, the autogenic bacterium *Clostridium carboxidivorans* uses syngas (H$_2$, CO and CO$_2$) to generate various chemicals (e.g., acetate, ethanol, butanol, and butyrate) [31]. It contains several routes for CO$_2$ fixation, which includes the Wood-Ljungdahl pathway, the anaerobic pathways, and the pyruvate synthase reactions. $^{13}$C-experiments can identify the relative contribution of each CO$_2$ fixation pathways towards product synthesis. As a demonstration, cultivation of *Clostridium* with labeled $^{13}$CO$_2$ and $^{12}$CO has been shown in Figure 3A. Analysis of the labeling patterns in either alanine or pyruvate could reveal the relative contributions of the different CO$_2$ assimilation reactions to biomass and product synthesis.

Yield of a product form a biosynthesis pathway may suffer losses from side reactions and intermediate degradation/secretion. A statistical analysis on previous metabolic engineering studies observed 20% - 30% yield reduction per engineered enzymatic reaction step (“Rule of Thumb”) [26,27]. To reduce the carbon loss, novel pathways are constantly proposed and engineered into microbial hosts to create a “short-cut” or carbon efficient route from the feedstock to the final product. Whenever heterologous pathways are engineered into a microbe, the actual contribution to the final product of the new pathway versus the native pathway is often difficult to be estimated [35]. In the following example, we demonstrate that $^{13}$C-experiments can determine the relative fluxes through multiple pathways by measuring product labeling. Specifically, 1-butanol could be produced simultaneously from a threonine pathway and a citramalate pathway (a short-cut keto acid-mediated pathway) in *E. coli* [36]. If 1st position $^{13}$C-pyruvate and $^{13}$C-bicarbonate were fed to 1-butanol producing cultures, labeling patterns in 1-butanol can reveal the fluxes through both the routes (Figure 3B). Recently, a non-oxidative glycolytic cycle (NOG) was designed to increase biofuel yield [12]. This NOG pathway starts with fructose
Figure 3 (See legend on next page.)
6-phosphate and undergoes three metabolic cycles to generate acetyl-CoA without losing any carbon. To probe the contribution of NOG pathway to overall cell metabolism, this study has presented a carbon rearrangement map for $^{13}$C-analysis of the NOG pathway function. These examples illustrate the potential of $^{13}$C-analysis to examine the in vivo activity of various novel pathways towards product synthesis.

Product yield influenced by bioenergetic efficiency

The theoretical product yield is generally calculated based on the stoichiometry of product synthesis from a carbon substrate. However, microbial energy metabolism also affects product yield because the synthesis of high-energy chemicals is energetically expensive, consuming large amounts of ATP/NAD(P)H. Cell maintenance (i.e., energy consumed for functions other than the production of new cell material) strongly competes for energy molecules and limits product synthesis. The maintenance energy involves regeneration of macromolecules, futile metabolic cycles, energy spilling reactions, proofreading, cell motility, preservation of chemical gradients, and repairing of cell damage caused by environmental stresses [37,38]. For example, non-growth-associated maintenance in wild type E. coli consumes 7.6 mmol of ATP per gram dry weight per hour [39]. Moreover, oxidative phosphorylation of NADH is a major source for ATP generation (theoretical maximum P/O ratio: 1 NADH → 3 ATPs) [40]. Cytochrome oxidase is a transmembrane protein complex that transfers electrons to O$_2$ and translocate protons across the membrane to establish a proton gradient to power ATP synthase. However, proton translocation through membrane is not always coupled with electron transfer from NADH to O$_2$, which reduces the contribution of oxidative phosphorylation to the establishment of the proton motive force for ATP synthesis [41,42]. Thereby, the actual P/O ratio, which is still in debate, is observed to be below 2.5 [43]. Under metabolic stresses, the respiration efficiency can be further reduced because transmembrane proton gradients for ATP synthesis leak over time, resulting in loss of catabolic energy capture [37,44].

For example, the riboflavin producing Bacillus subtilis has a P/O ratio of 1.3, and a small increase in P/O ratio (from 1.3 to 1.5) could increase riboflavin yields by 20% [45].

The amount of energy from substrate catabolism diverted to non-growth functions varies dramatically depending on different organisms and growth conditions (e.g., during E. coli growth, its energy yield of substrate catabolism could be one-third of the theoretical maximum) [37]. To illustrate the impact of energy efficiency on product yield [46], a small-scale flux balance model related to fatty acid-overproducing strain was built exclusively for this report. This small-scale model employs eight reactions (Table 1) to demonstrates free fatty acid production as a function of non-growth associated ATP maintenance and P/O ratio [47]. The fluxes were resolved by the function below:

$$\max \: v(2) \quad \text{such that } \: A \cdot v = b \quad \text{and} \quad lb \leq v \leq ub,$$

where the objective function is to maximize $v(2)$ (i.e., the relative flux of fatty acid). $A$ is the reaction stoichiometry. $lb$ and $ub$ are upper and lower bound for each reaction flux, $v(i)$. Figure 4A shows the relationship between

<table>
<thead>
<tr>
<th>Flux, v</th>
<th>Reactions</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>v(1) Glucose → 2AcetoCoA + 2ATP + 4NADH</td>
<td>Glycolysis</td>
<td></td>
</tr>
<tr>
<td>v(2) AcetoCoA + 1.5NADPH + 0.85ATP → 0.125 C16:0 fatty acid</td>
<td>Fatty acid synthesis</td>
<td></td>
</tr>
<tr>
<td>v(3) AcetoCoA → 2NADH + NADPH + ATP + FADH2</td>
<td>TCA cycle</td>
<td></td>
</tr>
<tr>
<td>v(4) NADH → NADPH</td>
<td>Transhydrogenation</td>
<td></td>
</tr>
<tr>
<td>v(5) NADH → P/O ATP</td>
<td>Oxidative phosphorylation</td>
<td></td>
</tr>
<tr>
<td>v(6) FADH2 → 0.67FAD/GTP</td>
<td>Oxidative phosphorylation</td>
<td></td>
</tr>
<tr>
<td>v(7) ATP → ATP_maintenance</td>
<td>ATP maintenance (non-growth associated)</td>
<td></td>
</tr>
<tr>
<td>v(8) 6dGlucose + 3.7ATP + 9.5NADPH + 2.5AcetoCoA → 39.7Biomass + 3.1NADH</td>
<td>Biomass formation</td>
<td></td>
</tr>
</tbody>
</table>

Note: Glucose consumption for both biomass growth and product synthesis is normalized to 100. The linear optimizer 'lsqnonlin' function in MATLAB is used for the optimization. The final yield (g fatty acid/g glucose) is calculated as follows: $Y = (v(2)/9.256)/(100-180)$ g C16:0 fatty acid/g glucose.
maximum yield, P/O ratio and ATP maintenance without constraining biomass growth (v(8) × 0) (Table 1). A higher P/O ratio makes the microbial system less sensitive to the increased demand for ATP loss. When the ATP maintenance is low and the P/O ratio is close to 3, the fatty acid yield can reach the theoretical value of 0.36 g fatty acid/g glucose (Figure 4A). In such conditions, eliminating competing pathways or engineering new pathways to avoid carbon loss may be effective to achieve a yield close to the theoretical maximum [48-50]. When ATP consumption for maintenance increases, cells need to use extra carbon substrates for energy generation, thereby decreasing the fatty acid yield significantly. Under these circumstances, one should consider strategies that will either reduce cell maintenance or increase the flux towards ATP synthesis.

In a recent study of an engineered E. coli for fatty acid overproduction [47], 13C-MFA showed that the theoretical ATP/NADPH generation (assuming P/O ratio = 3) from glucose catabolism was much higher than ATP/NADPH consumption for biomass growth and fatty acid synthesis. After optimization of biosynthesis pathway via ‘push-pull-block’ strategies, this engineered strain had a fatty acid yield of only 0.17 g fatty acid/g glucose (Figure 4B) because a substantive fraction of energy yield from glucose catabolism was lost due to the suboptimal energy metabolism. Such high cell maintenance and low P/O ratio in the engineered E. coli are likely caused by the various physiological stresses during biol biofuel production (e.g., changed cell membrane integrity and compositions [51]). Thereby, 13C-MFA not only applies for a better understanding of carbon flux distribution, but also provides a diagnostic analysis of the energy-dependent metabolic capability for product yields. If the microbial metabolism demands a considerable amount of ATP/NAD(P)H for both biosynthesis and cell maintenance, optimal product yield is unlikely to be achieved by overexpressing biosynthesis pathways or by redirecting metabolic fluxes to avoid carbon losses. A more promising approach would be to improve energetic prospect or respiration efficiency, thereby allowing the cells to ‘burn’ substrates more efficiently to satisfy the energy requirement [52,53].

Conclusions
Product yield is one of the main considerations of microbial cell factories [54]. Microbial productivity is not only associated with the efficiency of biosynthesis enzymes, but is also intertwined with the energy metabolism [55]. Simple 13C analysis can characterize the host’s intrinsic production yields under different carbon sources, and determine the contributions of the different pathways to biosynthesis. In addition, 13C-MFA can profile microbial fluxomes and determine the amount of extra substrates that the cell consumes to compensate for ATP losses from diverse cellular processes, which is essential to understand metabolic capability of a microbial host for maximal product yields. In the end, 13C-analysis, using the labeled product as internal standards, can also be employed to correct product measurement noises in fermentation processes due to water loss, product evaporation or degradation [56]. This review paper aims to emphasize the indispensable value of 13C-labeling techniques to the metabolic engineering field as we foresee an extended use of 13C-experiments for the development of microbial cell factories.
Competing interests
The authors declared that they have no competing interests.

Authors' contributions
AWHT wrote the introduction, product yield from rich medium, multiple carbon substrates and parts of yield from alternative pathways. LH wrote the bioreductive efficiency. LJ contributed to the section of product yield from alternative pathways. WH printed the paper. All the authors approved the final manuscript.

Acknowledgement
We thank Prof. Maciej Antoniewicz from University of Delaware for his advice. The manuscript also thanks Katina Leyden for her close reading of this manuscript. This work was supported in parts by the funding from the National Science Foundation (MCB0954016) and the Bridge Funding from Washington University in St. Louis.

Received: 23 December 2013 Accepted: 12 March 2014
Published: 19 March 2014

References
Appendix Chapter 6: Biofuel production: an odyssey from metabolic engineering to fermentation scale-up
Biofuel production: an odyssey from metabolic engineering to fermentation scale-up

Whitney Hollinshed, Lian He and Yinjie J. Tang *
Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO, USA

Edited by:
Andrea Muthopadhyay, Lawrence Berkeley National Laboratory, USA
Reviewed by:
Kesun Ma, University of Waterloo, Canada
Ex-Been Goh, Joint BioEnergy Institute, USA
*Correspondence:
Yinjie J. Tang, Department of Energy, Environmental and Chemical Engineering, Washington University, One Brookings Dr., St. Louis, MO 63130, USA
email: yinjie.tang@wustl.edu

Metabolic engineering has developed microbial cell factories that can convert renewable carbon sources into biofuels. Current molecular biology tools can efficiently alter enzyme levels to redirect carbon fluxes toward biofuel production, but low product yield and titers in large bioreactors prevent the fulfillment of cheap biofuels. There are three major roadblocks preventing economical biofuel production. First, carbon fluxes from the substrate dissipate into a complex metabolic network. Besides the desired product, microbial hosts direct carbon flux to synthesize biomass, overflow metabolites, and heterologous enzymes. Second, microbial hosts need to oxidize a large portion of the substrate to generate both ATP and NADPH to power biofuel synthesis. High cell maintenance, triggered by the metabolic burdens from genetic modifications, can significantly affect the ATP supply. Therefore, fermentation of advanced biofuels (such as biodiesel and hydrocarbons) often requires aerobic respiration to resolve the ATP shortage. Third, mass transfer limitations in large bioreactors create heterogeneous growth conditions and micro-environmental fluctuations (such as suboptimal O2 level and pH) that induce metabolic stresses and genetic instability. To overcome these limitations, fermentation engineering should merge with systems metabolic engineering. Modern fermentation engineers need to adopt new metabolic flux analysis tools that integrate kinetics, hydrodynamics, and 13C-proteomics, to reveal the dynamic physiologies of the microbial host under large bioreactor conditions. Based on metabolic analyses, fermentation engineers may employ rational pathway modifications, synthetic biology circuits, and bioreactor control algorithms to optimize large-scale biofuel production.

Keywords: ATP maintenance, hydrodynamics, metabolic flux analysis, proteomics, synthetic biology

MICROBIAL FACTORIES FOR BIOFUEL PRODUCTION
Extensive research has been performed on the microbial production of biofuels using renewable feedstocks (Figure 1). Carbohydrates, the major carbon sources for biofuel production, can be obtained from either food crops or lignocellulosic biomass. Glycerol, lactate, acetate, and syngas are also used as feedstocks for biofuel synthesis. Moreover, photo-biorefineries are being developed to turn light energy and CO2 into useful chemicals (Lindberg et al., 2010; Lan and Liao, 2012; Oliver et al., 2013). Recently, the Department of Energy has started initiatives to target methane as a cheap resource for “gas-to-liquids” bioconversion in the hope of surpassing Fischer-Tropsch process for liquid fuel production (Conrado and Gonzalez, 2014). Despite the numerous feedstocks that are proposed for biofuel fermentation, production of cheap biofuel has not yet been realized.

Ethanol fermentation by yeast is the most developed biofuel process, but low combustion energy and high purification costs prevent the wide use of ethanol as an economical fuel. Thereby, researchers have engineered microbes to produce new fuels. Advanced biofuel examples include higher alcohols via the keto-acid and the Ehrlich pathway (Atsumi et al., 2008), terpene-based fuels (e.g., isopentenol) from the mevalonate pathway (Withers et al., 2007), fatty acid ethyl esters (Kalscheuer et al., 2006; Steen et al., 2010) and alkanes from fatty acid biosynthesis pathways (Choi and Lee, 2013). These biofuel pathways have been extensively reviewed (Rude and Schirmer, 2009; Feralta-Yahya et al., 2012). Despite the development of these diverse biofuel producers, it is still challenging to commercialize biofuel processes due to the poor microbial productivity in large bioreactors and the low profit margins of biofuels (Zhang, 2009; Lamonica, 2014). As a result, many biotechnology companies have shifted their focus away from advanced biofuels to products with higher commercial value. This review discusses both the metabolic and bioprocess limitations in the scale-up of these biofuel processes and emphasizes the need to integrate systems biology, synthetic biology, and fermentation engineering to optimize metabolic performance in large bioreactors.

METABOLIC ENGINEERING APPROACHES FOR BIOFUEL SYNTHESIS
The microbial host’s metabolism consists of thousands of chemical reactions that control the carbon and energy [e.g., ATP & NAD(P)H] metabolism. The desired biosynthetic pathway for advanced biofuels often requires multiple enzymatic steps. Current molecular biology techniques can effectively alter enzyme
levels to increase the flux toward biofuel synthesis. Common strategies include choice of plasmids and its copy numbers, promoter engineering, codon optimization, synthetic scaffolds, directed evolution/modification of key enzymes, improvement of ribosome binding sites, and knockout/knockdown of competitive pathways (Dueber et al., 2009; Carneiro et al., 2013; Nowroozi et al., 2014). New genetic techniques, such as RNA Interference, CRISPRs or TALENs, offer new capabilities to edit microbial metabolism (Pratt and MacRae, 2009; Jiang et al., 2013; Sun and Zhao, 2013).

To improve the carbon flux to the final products, metabolic engineering often applies two strategies. The first strategy is the “push-pull-block” used to increase the flux toward the biofuel synthesis pathway (Atsumi et al., 2010a; Kind et al., 2013). For example, a threonine-overproducing E. coli strain was engineered to produce 1-propanol via the keto-acid pathway by the “push-pull-block” strategy (Choi et al., 2012): (1) “Pull”—introduce a heterologous feedback resistant threonine dehydratase, (2) “Block”—remove competing metabolic pathways, and (3) “Push”—overexpress acetate kinase and other enzymes in the citramalate pathway to increase carbon flux into the propanol pathway. The second metabolic engineering strategy is to design an alternative biosynthesis pathway that can reduce the loss of carbon by unwanted byproducts. For example, a non-oxidative glycolytic cycle in E. coli has been developed to achieve the complete carbon conversion of sugar into acetyl-CoA (Bogorad et al., 2013). Although these metabolic engineering strategies are effective in increasing the carbon flux toward the desired product, metabolic engineers cannot easily create “biofuel super bugs”. Extensive genetic modifications often increase metabolic burdens on the host and thus interfere with cell growth and product synthesis (Colletti et al., 2011; Pouxt et al., 2014). For example, high copy number plasmids or strong promoters can place a heavy burden on the cell’s growth and negatively affect productivity (Carrier et al., 1998; Jones et al., 2000). Moreover, host cells may incorrectly express or misfold heterologous enzymes, reducing their activities. Low temperature fermentation may be required to ensure the functions of these heterologous enzymes (Chang et al., 2007). In addition, pathway engineering may cause metabolic imbalances and waste product secretions.

An emerging field, synthetic biology, aims to design and construct new biological systems to enhance the capability of engineered microbes (Nielsen et al., 2014). Synthetic biology has been developing genetic circuits that can precisely regulate gene expression in the presence or absence of chemical and environmental inputs (Khalil and Collins, 2010). These synthetic biological devices have been reviewed recently (Way et al., 2014), which include such devices as a toggle switch (two repressors turn each other off), trigger-memory system, and genetic oscillators. Synthetic biology tools have started to be used by metabolic engineers to manipulate fluxes toward biosynthesis pathways at different fermentation stages. For example, a recent study engineered a toggle switch into E. coli that could turn off the TCA cycle and redirect flux toward isopropanol (Soma et al., 2014). Among the synthetic biology tools (Neupert et al., 2008; Topp et al., 2010; Gorochowski et al., 2013), biosensor-regulator systems have particular value for their potential to control a microbial host metabolism according to environmental changes, and thus improve the productivity of microbial hosts (Zhang et al., 2012).

**METABOLIC DILEMMA: CARBON YIELD vs. ENERGY EFFICIENCY**

Current research often focuses on the improvement of carbon flux toward the final product. However, the high demand of energy and reducing equivalents during biofuel synthesis is another important obstacle. First, polymerization of protein and DNA/RNA requires large amounts of ATP (39.1 mmol ATP/g protein; 7.4 mmol ATP/g RNA; and 11.0 mmol ATP/g DNA) (Stephanopoulos et al., 1998). Production of biomass, enzymes for biofuel synthesis, plasmids/mRNA, or synthetic scaffolds...
consumes not only carbon building blocks, but also energy molecules. Second, large amounts of ATP need to be consumed to support cell maintenance processes including energy spilling, microbial motility, cell component repair, and re-synthesis of macromolecules (Hoehler and Jorgensen, 2013). Third, synthesis of biofuel molecules needs ATP and NAD(P)H. For example, fatty acid production requires 7 ATP and 14 NAD(P) to convert acetyl-CoA molecules into one fatty acid (Palmitate, C16:0). The carbon oxidation pathways (such as TCA cycle and oxidative pentose phosphate pathway), oxidative phosphorylation, and transhydrogenase reactions are required to generate sufficient NADPH and ATP for fatty acid synthesis (He et al., 2014; Varman et al., 2014).

Theoretically, 38 ATP molecules are produced from one glucose molecule. Among the 38 ATP molecules, glycolysis and TCA pathways only contribute to 4 ATP, and the remaining 34 ATP are obtained from oxidative phosphorylation, assuming the maximum P/O ratio (1 NADH → 3 ATP and FADH2 → 2 ATP) (Shuler and Kargi, 2002). Under anaerobic conditions, the energy metabolism is inefficient and cells often secrete acetate to overcome the ATP shortage. If the biofuel synthesis requires large amounts of ATP, oxidative phosphorylation becomes a key source for satisfying the ATP demand (i.e., use of aerobic respiration to generate the needed ATP). In addition to the high ATP demand imposed by the biofuel synthesis pathway, metabolic flux analysis studies have revealed that the overexpression of biosynthesis pathways significantly increases ATP maintenance expenditure (Owen et al., 2009), and the metabolic burden in engineered microbial hosts further causes poor respiration efficiency (e.g., P/O ratio ≈ 1.3) (Varma and Palsson, 1994; Sauer and Bailey, 1999). If the hosts suffer from severe ATP limitations, efforts to reduce carbon losses or to increase carbon availability to biofuel synthesis will be futile. In this case, the metabolic bottleneck may shift from carbon limitation to energy limitation (insufficient energy molecules to power biofuel synthesis). Many metabolic engineering approaches to improve carbon efficiency are effective in redirecting carbon fluxes to biofuel in low productivity strains (yield far below theoretical value), but these strategies cannot raise product yields close to stoichiometric predictions if the engineered metabolism is unable to satisfy the overall ATP and NAD(P)H requirements by the microbial hosts. The priorities toward high carbon yield and energy efficiency have to be carefully balanced during strain development of biofuel producers.

**SCALE-UP FERMENTATION OBSTACLES: METABOLIC STRESSES UNDER SUBOPTIMAL CULTURE CONDITIONS**

Fermentation engineering emerged in the early 1940s driven by the mass production of penicillin. Since then, engineers have scaled up fermentation of commodity products such as 1,3-propanediol (Nishikawa and Whited, 2003) and amino acids (Hermann, 2001). Currently, Gevo and Butamax are commercializing isobutanol production from engineered yeast (Nielsen et al., 2014). Industrial isobutanol fermentation is one of the most promising biofuel fermentations as: (1) yeast is a robust industrial host that has a natural tolerance to alcohol; and (2) in situ removal techniques such as gas-stripping have alleviated product toxicity during the fermentation process (Baez et al., 2011). However, large-scale fermentations of other fuels (such as biodiesel) are still underdeveloped.

Maintaining the optimal growth environments in large bioreactors is difficult. In industrial bioreactors (on the scale of 100 m3), the poor mixing/eration can cause temperature and pH fluctuations, O2 limitations, substrate/product inhibitions, and accumulations of wastes (such as acetate) (Enfors et al., 2001). The heterogenous conditions can increase cell stress and ATP maintenance. If gaseous substrates (such as CO2, syngas or CH4) are used for fermentation, gas-liquid mass transfer resistances pose another serious challenge as the gaseous substrates must diffuse across the gas-liquid interface (Blanch and Clark, 1997). In addition, large-scale bioreactors (e.g., a fed-batch bubble column reactor) feed the substrate from the top and aeration from the bottom, creating opposite substrate and O2 gradients (i.e., O2 limitation at the top and substrate limitation at the bottom). This has been reported to cause increased production of waste products, such as formate, lactate, and succinate (Bylund et al., 1998, 2000). Moreover, cell factories synthesizing biofuels are subjected to metabolic burdens due to the drainage of both metabolic precursors and energy for the replication of plasmids, and biosynthesis of heterologous enzymes and products (Carneiro et al., 2013). The suboptimal growth conditions in large bioreactors tend to intensify stress responses, induce metabolic shifts, and alter cell genetic stability. Altogether, most of the engineered "super bugs" struggle to move beyond lab settings.

**13C-MFA: AN INDISPENSABLE TOOL FOR SCALE-UP FERMENTATION ENGINEERING**

Systems biology can characterize biofuel producers and provide guidelines for rational metabolic engineering and optimal fermentations (Figure 2) (Carneiro et al., 2013). "Omics" (transcriptomics, proteomics, and metabolomics) approaches can identify useful mutations, discover gene regulations and enzyme functions, and measure the metabolite pools in response to environmental fluctuations (Pham et al., 2006; Asumi et al., 2010b; Redding-Johanson et al., 2011). Although genome-wide analyses and computational modeling have provided knowledge for the metabolic engineering of microbial hosts (Park et al., 2008), it is not straightforward to correlate "omics" to actual enzyme functions due to complex post-transcriptional regulations (such as allosteric regulation) (Chubukov et al., 2013). For example, increases in mRNA levels may not lead to a corresponding increase in protein levels, while enzyme levels can remain constant despite frequent changes in carbon fluxes (Gigi et al., 1999).

In addition, the existence of isoenzymes and poor enzyme specificity contribute to high metabolic diversity and lead to difficulties in metabolic studies (Schwalb, 2003). To overcome such problems, 13C-MFA (metabolic flux analysis) is used to directly measure enzymatic reaction rates. Compared to other omics studies, 13C-MFA can provide insights into cell's physiology during large-scale fermentations. 13C-MFA uses metabolic reaction stoichiometry and carbon-labeling experiments to precisely estimate metabolite turnover rates (Sauer, 2006; Tang et al., 2009). 13C-MFA in combinations with other in silico metabolic models (flux balance analysis) (Edwards et al., 2001; Orth et al., 2010) can pre-
dict biosynthesis yield, delineate functional pathways, calculate the actual fluxes throughout the metabolic network, validate the function of genetic circuits, and identify metabolic engineering targets. Integration of flux analysis with transcriptomics and proteomics can provide a comprehensive understanding of the genetic regulations of intracellular activities (Zhang et al., 2010; Yoon et al., 2012; Arakawa and Tomita, 2013; Liu et al., 2013).

During the fermentation process, metabolic shifts and genetic mutations are common, which can create subpopulations in the cultures (biofuel producers vs. mutants after loss of biofuel production). Sauer group proposed a “reporter protein”-based 13C-MFA to probe the metabolism of subpopulations (Rühl et al., 2011). This reporter protein, synthesized by a particular subgroup, stores the 13C-labeling information of that subgroup which allows 13C-MFA to identify the different metabolic flux phenotypes. This proof-of-concept approach was tested using engineered E. coli (Shakib et al., 2008; Rühl et al., 2011). Based on the labeling information from hydrolysates (i.e., amino acids) of the green fluorescent protein (GFP). 13C-MFA was able to probe the subpopulation metabolism (GFP producers) during E. coli fermentation. Recently, proteomic analysis of 13C-peptide labeling has also been developed to assist metabolic flux analysis of heterogeneous microbial systems (Mandy et al., 2014). The “reporter proteomics” associated with a specific time of protein expression or a unique species may decipher subpopulation physiology and its dynamic changes in microbial consortia. These studies have paved the way to analyze temporal and spatial microbial phenotypes during a large-scale heterogeneous fermentation.

In addition, 13C-MFA may integrate with microbial kinetics to reveal the changes of cell metabolism throughout its cultivation process. For example, Sauer group have utilized such a dynamic 13C-MFA approach in a fed-batch fermentation to monitor the dynamic changes in intracellular fluxes during the different growth stages. Their results revealed that a riboflavin-producing bacterium’s physiology shifted from an overflow metabolism to an exclusively maintenance metabolism at the late fermentation stage (Rühl et al., 2010). To determine microbial responses to industrial bioreactor configurations and cultivation heterogeneity due to mass transfer limitations, even more novel 13C-MFA approaches are required. For example, metabolic flux analysis in combination with hydrodynamics would provide important insights into cell physiology at different locations inside of a large bioreactor. However, to the best of our knowledge, such metabolic flux analysis techniques have not been fully developed.

**MODERN FERMENTATION ENGINEERING: AN INTEGRATED APPROACH FOR BIOFUEL FERMENTATION MANUFACTURING**

The promise of cheap biofuels has yet to be fulfilled. Currently, alcohols (though of a lower energy density) are more promising biofuels than biodiesel/hydrocarbons, as alcohol synthesis pathways are less dependent on the ATP supply and alcohol fermentation can be conducted under anaerobic conditions. Production of biodiesel and hydrogen require large amounts of ATP, and their processes are usually restricted to aerobic conditions and their productivities are highly sensitive to the P/O ratio (He et al., 2014; Varman et al., 2014). To date, researchers are still unable to create “super bugs” that have both efficient carbon metabolism and frugal energy usage for economical synthesis of advanced biofuels. Thereby, modern fermentation engineering needs to be merged with advanced metabolic engineering to employ the following strategies.

First, it is important to select and develop a proper biofuel chassis with an efficient energy and carbon metabolism. Yeast (e.g., Saccharomyces cerevisiae) is naturally tolerant to alcohols
and acetate, and it is a robust workhorse for industrial alcohol fermentations. However, the energy metabolism of yeast has a lower net ATP production as additional ATP is consumed during the transport of equivalents of NADH from the cytoplasm into the mitochondria (Shuler and Kargi, 2002). In addition, the mitochondrial metabolic transport processes may limit the enzyme-substrate accessibility during biofuel synthesis. Thereby, compartmentalization of biosynthetic pathways in yeast’s mitochondria is required to improve local enzyme concentrations and microenvironments for biofuel synthesis (Avila et al., 2013). On the other hand, a bacterial chassis, E. coli, has faster growth and product synthesis rates, pathways that are easier to modify, and a broader capability to co-utilize carbon sources (including xylose). E. coli has already been engineered to produce higher alcohols, biodiesels and other hydrocarbons that demand large amounts of energy molecules. Another promising chassis is microalgae (i.e., cyanobacteria). Although most studies emphasize their photoautotrophic features, photomixotrophic fermentation has two great advantages for industrial use. (1) Photo-fermentations of organic substrates (sugars) can reach high carbon yields of both biofuel and biomass due to mixotrophic CO₂ fixation and light energy harvesting to generate ATP and NADPH (You et al., 2014). (2) Photo-fermentation by cyanobacteria has no overflow byproducts and is less susceptible to CO₂ light limitations during the bioprocess scale-up. Another possible direction, non-model species, such as thermophiles, are also promising for cheap biofuel production because such species are particularly suited for consolidated bioprocesses (Lin et al., 2014). Finally, it might be possible to construct a new microbial chassis using a synthetic genome in the near future, which may be developed to contain all the advantageous features of the other microbial hosts. These synthetic or non-model microbes may eventually fulfill the promise of cheap biofuels.

Second, the loss of carbon yield due to ATP/NAD(P)/H limitations is often severe in engineered microbial hosts. Therefore, modern fermentation engineering should focus on reducing the metabolic burden and enhancing the prosperity of the energy metabolism. There are several effective strategies to improve NAD(P)/H and ATP availability. (1) Adding nutrient sources (such as yeast extract) into cell culture; (2) overexpressing NADH dehydrogenase to enhance the respiration efficiency (Callahan et al., 1993); (3) using engineered enzymes to balance NADH/NADPH generation and consumption (Berrios-Rivera et al., 2002; Javidpour et al., 2014); (4) employing in situ product recovery/separation to avoid culture inhibitions (Baez et al., 2011); and (5) maintaining the optimal cultivation conditions (such as dissolved O₂ substrate concentrations, etc.).

Third, traditional fermentation engineering aims to understand bioreactor rheology and bio-reaction kinetics, but there should be a deeper understanding of the microbial host in large bioreactors. Fermentation engineers often study process control parameters such as mixing quality, oxygen/nutrient supply, reactor geometry, impeller selection, aeration rate, heat transfer, power-volume ratio, and the necessary utility operation costs (Delvigne et al., 2003). For example, fermentation engineers have designed fed batch algorithms to avoid acetate secretion by E. coli (Shuler and Kargi, 2002). However, future fermentation engineering should also combine process analyses (including computational fluid dynamics, structured kinetic models, and scale-up simulations) with advanced metabolic flux analysis to capture the cell’s behavior and dynamics under the vessel’s specific conditions and provide guidelines for process design and control. Systems analysis has been successfully used as a fermentation engineering technique to reveal metabolic bottlenecks, design optimal culture medium, and monitor physiogical performance under different bioreactor conditions (Becker et al., 2013; Posch et al., 2013). For example, researchers from Genentech have integrated bioprocess models with metabolic flux analysis to study E. coli metabolism in a 1000 L bioreactor for the production of a recombinant therapeutic protein. The integrated modeling approach can be used to optimize process variables and media compositions to unlock the optimal cell metabolism for efficient biosynthesis (Meadows et al., 2010). As expected, future fermentation engineering can significantly benefit from combining metabolic flux analysis with mass transport and hydrodynamics to understand cell physiologies in heterogeneous/dynamic bioreactor environments and identify the bioreactor stress factors.

Fourth, synthetic biology offers new opportunities for fermentation scale-up. Introducing synthetic dynamic control systems can prevent the biosynthesis of unnecessary RNAs/proteins/metabolites, increase the efficiencies of energy and carbon usage, and allow a host to adjust its metabolic flux to minimize “maintenance loss” (Zhang et al., 2012). Metabolic pathways can be regulated via transcription factors that quickly respond to shifts in metabolite levels, and thus fluxes in the biosynthetic pathway could be controlled on the time scale of minutes (Holtz and Keasling, 2010). To build dynamic regulatory capability in a microbial host, biosensor-regulators can be used to promote or repress a biofuel pathway or substrate uptake according to its growth conditions (e.g., quorum sensing) or metabolite concentrations (Dunlop et al., 2010). Recently, a dynamic feedback control system was developed that enabled the host to sense metabolic changes by detecting the toxic intermediate’s concentration (e.g., malonyl-CoA or acyl-CoA), and thereby controlling the expression of biodiesel synthesis genes (Zhang et al., 2012). Such a feedback control allows cells to maintain the concentration of the precursor and intermediate at desirable levels, and thus both genetic stability and productivity in the new strains were improved. In addition, biosensors of feedforward control, will also be beneficial in the scale-up of fermentations, because they allow cells to respond to environmental changes before a deleterious impact on cell metabolism. For example, O₂ limitation leads to acetate secretion and glucose over-consumption during aerobic fermentations. O₂ sensitive promoters (Salmon et al., 2003) can be used to control glucose transporter genes and “shut down” glucose uptake when the cells move to O₂ limiting zones in a large bioreactor (Figure 3). Although bio-sensing regulatory systems can optimize biomass growth and product synthesis at different growth phases/conditions in laboratory settings, applicability of synthetic biology in scale-up fermentation is still unclear. Recently, the stability of genetic circuits (AND and NOR gate) were evaluated under simulated industrial fermentation conditions, and the circuit’s performance (AND gate) deteriorated with...
increased culture volume (Moser et al., 2012). Thereby, it is necessary for fermentation engineers to test and improve synthetic biology circuits under dynamic and heterogeneous cultivation conditions.

SUMMARY
Optimization of biofuel fermentation at the industrial scale is difficult and costly. Biofuel fermentation engineering should integrate with metabolic engineering to tune the expression of multiple heterologous genes, improve the energy metabolism (high P/O ratio and low cell maintenance), and construct sensor-regulator systems to improve cell productivity in industrial bioreactors. Fermentation engineers should have a comprehensive understanding of both the macroscopic (e.g., oxygen level, mixing, and bioreactor controls) and microscopic (intracellular fluxes) parameters, and thus fill the gaps between laboratory studies and industrial applications.

“Scale-down” experiments (large-bioreactor fermentation simulated on small scales) with metabolic flux analysis can be routinely used to diagnose the engineered metabolism, verify synthetic biology circuits, and design optimal fermentation strategies (Figure 2). Most importantly, this broad-scope metabolic knowledge would allow companies to select and focus on “promising” microbial factories with high chances of success.

ACKNOWLEDGMENTS
We would like to thank Prof. Fuzhong Zhang and William Henson for useful discussions on synthetic biology. This work was supported by the funding from National Science Foundation (MCB0954016).

REFERENCES


Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 18 March 2014; accepted: 20 June 2014; published online: 09 July 2014.

Citation: Hollinshead, W., He, L. and Tang, Y. J. (2014) Biofuel production: an odyssey from metabolic engineering to fermentation scale-up. Front. Microbiol. 5:444. doi: 10.3389/fmicb.2014.00444

This article was submitted to Microbial Physiology and Metabolism, a section of the journal Frontiers in Microbiology.

Copyright © 2014 Hollinshead, He and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided that the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.
Appendix Chapter 7: WUFlux Manual
# Table of Contents

1 Requirement.................................................................................................................. 2

2 Installation ....................................................................................................................... 2

3 General procedures of using WUFlux .............................................................................. 3

   3.1 Open WUFlux ........................................................................................................... 3

   3.2 Start a new $^{13}$C MFA model ............................................................................... 4

      3.2.1 Substrate uptake reactions and product formation reactions ...................... 4

      3.2.2 Reactions in the central metabolism .............................................................. 5

      3.2.3 Linear equality and inequality constraints .................................................... 5

   3.3 Input experimental data ......................................................................................... 6

      3.3.1 Mass Isotopomer Distribution ....................................................................... 6

      3.3.2 Labelled substrates ....................................................................................... 7

   3.4 Customize optimization settings ............................................................................ 8

   3.5 Calculate flux distribution ...................................................................................... 8

4 Troubleshooting ........................................................................................................... 10

   4.1 Metabolic reactions ............................................................................................... 10

   4.2 MID data ................................................................................................................. 11

5 Appendix 1. Description of files in WUFlux ................................................................. 13

6 Appendix 2. Abbreviation of metabolites used in WUFlux ........................................... 14
1 Requirement

To use this software, MATLAB 2012b or higher version, together with optimization toolbox, symbolic math toolbox, and statistic toolbox, is required.

2 Installation

Visit the website: http://13cmfa.org/ or http://fluxomics.net. Then Click the icon ‘WUFlux’ to download WUFlux.

In MATLAB, unzip the package you have downloaded, and add the folder into MATLAB working directory.
3 General procedures of using WUFlux

3.1 Open WUFlux.

Simply type 'Main' in the command window. An interface will show up afterwards.

The upper row shows five buttons: Metabolic Reaction, Experimental Data, Start, Results, and Settings. The buttons in the bottom allow users to start a new project, rename a model, open a previous project, and export the current results into an excel file.
3.2 Start a new $^{13}$C MFA model.

To build up an MFA model, click ‘Metabolic Reactions’ first. In the ‘Edit’ panel, choose a microorganism template, *E. coli* (chemoheterotrophic bacteria) or *Synechocystis* 6803 (photomixotrophic or photoheterotrophic) template. Then all the reactions in the central carbon metabolism will show up. Users can modify the template by knocking out certain reactions, changing the lower and upper bounds, and defining the reaction reversibility. All the changes will be automatically saved in the model, so users don’t need to save the project from time to time. The carbon transition rules are presented in the last column of the table, which are included in the model and are not subject to changes.

![Image of metabolic reactions](image)

3.2.1 Substrate uptake reactions and product formation reactions

In the template, we included 6 substrate uptake reactions, i.e., glucose, xylose, acetate, pyruvate, glycerol, and NaHCO$_3$/CO$_2$. Users can choose one or multiple substrates according to their experimental designs. Also, users can add reactions related to product excretion by clicking ‘Add a reaction’, and delete those by clicking ‘Undo’. Note that a new reaction in the central
metabolism cannot be added in the template, since all the carbon transition fates have been pre-defined. In WUFlux, we use a MATLAB built-in function, 'equationsToMatrix', to convert linear equations to the corresponding matrix form. The intracellular metabolites are treated as variables in the model. Users can refer to the following website

http://www.mathworks.com/help/symbolic/equationstomatrix.html?refresh=true for more information regarding how to modify the reactions.

3.2.2 Reactions in the central metabolism

We have provided all the reactions in the central carbon metabolism, including the glycolysis (the Embden–Meyerhof–Parnas) pathway, the oxidative pentose phosphate pathway, the Entner–Doudoroff pathway (in only E. coli template), the TCA cycle, the Calvin cycle (in only Synechocystis 6803 template), the C1 (including 5-methyl-tetrahydrofolate and 5,10-methylene-tetrahydrofolate) metabolism pathway, and the anaplerotic pathway. Amino acids synthesis pathways are also included in the model. The abbreviations of intracellular metabolites are listed in Appendix 2. In the E. coli template, we also provided cofactor balance which can be used for estimating NAD(P)H and ATP consumption and production.

3.2.3 Linear equality and inequality constraints

Additionally, users can add linear equality/inequality constraints. Following table shows several examples.

<table>
<thead>
<tr>
<th>Constraints to be added</th>
<th>How to write in the model</th>
</tr>
</thead>
<tbody>
<tr>
<td>2&lt;v2&lt;10</td>
<td>v2-10&lt;0</td>
</tr>
<tr>
<td>v2-v1+3*v3=9</td>
<td>v2-v1+3*v3-9=0</td>
</tr>
</tbody>
</table>

Page 5
3.3 Input experimental data.

To import the experimental data, click the button ‘Experimental Data’.

3.3.1 Mass Isotopomer Distribution

A table consisting of a list of ion fragments is shown on the left side. Users have three options to input their mass isotopomer distribution (MID) data: first, manually type all the corrected MID data, and check the boxes of MIDs that are chosen for simulation; second, click the button ‘Load MID data’ and then select the excel file ‘MIDtemplate_corrected’, which contains the corrected MID data; third, click ‘Convert Raw MS Data’ and select the excel file ‘MIDtemplate_uncorrected’ containing the raw MS data. The excel templates are provided in the package, and users can simply copy their data into those files. WUFlux will automatically calculate the mean values and standard deviations, if more than two sets of data from the same experiment are provided. The following is a screenshot of the Excel file for uncorrected MID data.
3.3.2 Labelled substrates

On the right side of the ‘Edit’ panel, users can define the substrates used in the $^{13}$C labelling experiments. Here, we define only the $^{13}$C ratio of each carbon atom in the substrate. For example, $[1, 2^{-13}\text{C}]$ glucose can be expressed as $[1, 1, 0, 0, 0, 0]$, which denotes that the first two carbons are labeled as $^{13}$C-carbons and the rest $^{12}$C-carbons. Users do not need to consider the labelling information of other substrates if they are not applied in your model, since their uptake reactions are knocked out in the network.

In the current version, up to three isotopologues of one substrate can be included in the model. For instance, if users use 30% of natural glucose and 70% of $[1, 2^{-13}\text{C}]$ glucose in their experiment, both the ratios and the MID of each isotopologue should be defined as follows:
<table>
<thead>
<tr>
<th>Substrates</th>
<th>Ratio</th>
<th>1st Carbon</th>
<th>2nd Carbon</th>
<th>3rd Carbon</th>
<th>4th Carbon</th>
<th>5th Carbon</th>
<th>6th Carbon</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glucose_1</td>
<td>0.3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Glucose_2</td>
<td>0.7</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Glucose_3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### 3.4 Customize optimization settings.

Before calculating the fluxes, users can modify the ‘Settings’ panel, including number of initial guesses, and algorithm. We use MATLAB built-in function ‘fmincon’ to solve the nonlinear optimization problem. All the parameters, for example, TolFun and MaxFun, have the same meanings as described on the website: [http://www.mathworks.com/help/optim/ug/fmincon.html](http://www.mathworks.com/help/optim/ug/fmincon.html).

In addition, we use Monte Carlo method to calculate the confidence interval of fluxes. Users can manually change the times they want to perturb the data, and the significant level (for example, 0.05 or 0.1) they want to apply in the model.

### 3.5 Calculate flux distribution.

Click ‘Start’ to calculate the intracellular flux distributions. The following window will appear to show the progress of the on-going calculation. All the flux calculations starting from each initial guess will be stored in the model.

![Progress window](image)

Once finished, another window will pop up, saying as follows:
Either click ‘Yes’ to continue calculating confidence intervals, or click ‘No’ to stop the calculation for the time being. The ‘Start’ button will be changed to ‘Continue’ button, and users can continue calculating confidence intervals later.

As shown in the above figure, users will find a table including the best fit of flux distribution and confidence intervals. The right panel shows the correlation of simulated MIDDs and experimental MIDDs. Ideally and theoretically, all the data points should lie on the diagonal line, which means the simulation has a perfect agreement with experiment results. Additionally, χ² test was applied to determine the goodness of fit (shown at the bottom line), which users can use as the reference to determine whether the final fitting is statistically acceptable. The interface also has a ‘Remove
results’ button, which allows users to remove the current flux results. However, once deleted, the results will not be recovered.

The saved data and results will be stored in the ‘.mat’ format, which can be found in the current folder. They can be loaded in MATLAB workspace, allowing users to process the data directly in MATLAB. Users can export all the results to excel files by clicking ‘Export’ button. The file includes the flux distribution, metabolic network information, a comparison between the experimental MID data and the simulated MID data, etc.

4 Troubleshooting

4.1 Metabolic reactions

All the reaction should be written as “PEP + 2*ATP ==> PYR.” Capital letters are required for each metabolite. Here are some examples:

<table>
<thead>
<tr>
<th>Right way of writing reactions</th>
<th>Wrong ways of writing reactions</th>
<th>What is wrong</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEP + 2*ATP ==&gt; PYR</td>
<td>PEP + 2ATP ==&gt; PYR</td>
<td>‘*’ missed in the equation</td>
</tr>
<tr>
<td></td>
<td>PEP + 2*ATP =&gt; PYR</td>
<td>‘==’ should be used in the equation, while the reversibility is defined elsewhere in the table</td>
</tr>
<tr>
<td></td>
<td>PEP + 2ATP ==&gt; pyruvate</td>
<td>‘PYR’ should be used in the model. (See Appendix 2 for abbreviations of all the metabolites used in WUFlux)</td>
</tr>
</tbody>
</table>
4.2 MID data

WUFlux will automatically check the MID data before flux calculation. Following figure shows an example, which says that the data of Alanine [M-57]⁻ is not right, because the summation of M=0, M=1, M=2, and M=3 is not equal to one. Under this circumstance, users should check their data.

![Image of MID data check](image.png)

Also, users may encounter negative MID data sometimes. Usually, the number is very close to zero (e.g. $-2 \times 10^{-5}$), and users can change the number to 0. If, however, the number is not close enough to zero, users should check their MID data.
5 Appendix 1. Description of files in WUFlux

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main.m</td>
<td>The main file of WUFlux.</td>
</tr>
<tr>
<td>Main.fig</td>
<td>The file for designing the interface of WUFlux.</td>
</tr>
<tr>
<td>EMU_Synechocystis6803</td>
<td>The file for calculating flux distribution of <em>Synechocystis</em> 6803 or similar species.</td>
</tr>
<tr>
<td>EMU_Ecoli</td>
<td>The file for calculating flux distribution in <em>E.coli</em> or similar species.</td>
</tr>
<tr>
<td>popcallback.m</td>
<td>The file containing metabolic reactions for different templates.</td>
</tr>
<tr>
<td>utable1callback.m</td>
<td>The file for saving/editing metabolic network in the model.</td>
</tr>
<tr>
<td>utable2callback.m</td>
<td>The file for saving/editing linear equality/inequality constraints in the model.</td>
</tr>
<tr>
<td>utable3callback.m</td>
<td>The file for saving/editing MDF data in the model.</td>
</tr>
<tr>
<td>utable4callback.m</td>
<td>The file for saving/editing substrate labelling information in the model.</td>
</tr>
<tr>
<td>utable5callback.m</td>
<td>The file for saving/editing optimization settings in the model.</td>
</tr>
<tr>
<td>button1callback.m</td>
<td>The file for ‘Add a reaction’ button.</td>
</tr>
<tr>
<td>button2callback.m</td>
<td>The file for ‘Undo’ button.</td>
</tr>
<tr>
<td>button3callback.m</td>
<td>The file for ‘Load MDF data’ button.</td>
</tr>
<tr>
<td>button4callback.m</td>
<td>The file for ‘Convert Raw MS data’ button.</td>
</tr>
<tr>
<td>button5callback.m</td>
<td>The file for ‘Remove Results’ button.</td>
</tr>
</tbody>
</table>
## Appendix 2. Abbreviation of metabolites used in WUFlux

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>3PG</td>
<td>3-phosphoglycerate</td>
</tr>
<tr>
<td>6PG</td>
<td>6-phosphogluconate</td>
</tr>
<tr>
<td>AceCoA</td>
<td>acetyl-CoA</td>
</tr>
<tr>
<td>AKG</td>
<td>α-ketoglutarate</td>
</tr>
<tr>
<td>ALA</td>
<td>alanine</td>
</tr>
<tr>
<td>ARG</td>
<td>arginine</td>
</tr>
<tr>
<td>ASN</td>
<td>asparagine</td>
</tr>
<tr>
<td>CIT</td>
<td>citrate</td>
</tr>
<tr>
<td>CO₂</td>
<td>carbon dioxide</td>
</tr>
<tr>
<td>CYS</td>
<td>cysteine</td>
</tr>
<tr>
<td>DHAP</td>
<td>dihydroxyacetone phosphate</td>
</tr>
<tr>
<td>E4P</td>
<td>erythrose 4-phosphate</td>
</tr>
<tr>
<td>F6P</td>
<td>fructose 6-phosphate</td>
</tr>
<tr>
<td>FBP</td>
<td>fructose-1,6-bisphosphate</td>
</tr>
<tr>
<td>FUM</td>
<td>fumarate</td>
</tr>
<tr>
<td>G6P</td>
<td>glucose 6-phosphate</td>
</tr>
<tr>
<td>GAP</td>
<td>glyceraldehyde 3-phosphate</td>
</tr>
<tr>
<td>GLC</td>
<td>glycolate</td>
</tr>
<tr>
<td>GLN</td>
<td>glutamine</td>
</tr>
<tr>
<td>GLX</td>
<td>glyoxylic acid</td>
</tr>
<tr>
<td>GLY</td>
<td>glycine</td>
</tr>
<tr>
<td>GLU</td>
<td>glutamate</td>
</tr>
<tr>
<td>HIS</td>
<td>histidine</td>
</tr>
<tr>
<td>ICIT</td>
<td>isocitrate</td>
</tr>
<tr>
<td>ILE</td>
<td>isoleucine</td>
</tr>
<tr>
<td>LEU</td>
<td>leucine</td>
</tr>
<tr>
<td>MAL</td>
<td>malate</td>
</tr>
<tr>
<td>MTHF</td>
<td>5,10-Methylenetetrahydrofolate (5,10-CH₂-THF)</td>
</tr>
<tr>
<td>OAA</td>
<td>oxaloacetate</td>
</tr>
<tr>
<td>PEP</td>
<td>phosphoenolpyruvate</td>
</tr>
<tr>
<td>PHE</td>
<td>phenylalanine</td>
</tr>
<tr>
<td>PRO</td>
<td>proline</td>
</tr>
<tr>
<td>PYR</td>
<td>pyruvate</td>
</tr>
<tr>
<td>R5P</td>
<td>ribose 5-phosphate</td>
</tr>
<tr>
<td>Ru5P</td>
<td>ribulose 5-phosphate</td>
</tr>
<tr>
<td>Ru1,5P</td>
<td>ribulose-1,5-diphosphate</td>
</tr>
<tr>
<td>S7P</td>
<td>sedoheptulose-7-phosphate</td>
</tr>
<tr>
<td>SER</td>
<td>serine</td>
</tr>
<tr>
<td>SUC</td>
<td>succinate</td>
</tr>
<tr>
<td>SucCoA</td>
<td>succinyl-CoA</td>
</tr>
<tr>
<td>THR</td>
<td>threonine</td>
</tr>
<tr>
<td>TRP</td>
<td>tryptophan</td>
</tr>
<tr>
<td>VAL</td>
<td>valine</td>
</tr>
<tr>
<td>X5P</td>
<td>xylose 5-phosphate</td>
</tr>
</tbody>
</table>