Application of metabolic modeling and machine learning for investigating microbial systems

Ni Wan
Washington University in St. Louis

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

Part of the Mechanical Engineering Commons

Recommended Citation
https://openscholarship.wustl.edu/eng_etds/334

This Dissertation is brought to you for free and open access by the McKelvey School of Engineering at Washington University Open Scholarship. It has been accepted for inclusion in McKelvey School of Engineering Theses & Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Application of Metabolic Modeling and Machine Learning for Investigating Microbial Systems

by

Ni Wan

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

May 2018
St. Louis, Missouri
Table of Contents

List of Figures ............................................................................................................ v
List of Tables ............................................................................................................. vii
Acknowledgments ........................................................................................................ viii
Abstract ......................................................................................................................... x
Chapter 1. Introduction ................................................................................................. 1
  1.1. Background ........................................................................................................... 2
      1.1.1. Metabolic model ......................................................................................... 2
      1.1.2. Microbial system ......................................................................................... 3
      1.1.3. 13C-based metabolism analysis .................................................................... 4
      1.1.4. Limitations of previous studies and general scopes of current studies .......... 4
  1.2. Dissertation Overview ......................................................................................... 5
Chapter 2. Cyanobacterial Photo-driven Mixotrophic Metabolism and its advantages for
Biosynthesis ....................................................................................................................... 7
  Abstract ......................................................................................................................... 8
  2.1. Introduction ........................................................................................................... 8
  2.2. Light-harvesting apparatuses ............................................................................... 9
  2.3. FBA simulation under various conditions .......................................................... 12
  2.4. Conclusion ........................................................................................................... 14
Chapter 3. Cyanobacterial carbon metabolism: fluxome plasticity and oxygen
dependence ....................................................................................................................... 15
  Abstract ......................................................................................................................... 16
  3.1. Introduction ........................................................................................................... 17
  3.2. Materials and Methods ....................................................................................... 18
      3.2.1. Cultivation conditions and biomass growth for flux and metabolite analysis .. 18
      3.2.2. Isotopomer analysis of proteinogenic amino acids ...................................... 19
      3.2.3. Metabolite pool size determination .............................................................. 19
      3.2.4. Metabolic flux analysis of *Synechocystis* 6803 under dark conditions ....... 20
  3.3. Results and Discussion ....................................................................................... 20
      3.3.1. Fluxomes under dark conditions ................................................................. 21
      3.3.2. FBA simulations ........................................................................................... 25
      3.3.3. Pool size of metabolites .............................................................................. 27
      3.3.4. Investigation of ATP metabolism ............................................................... 28
      3.3.5. Cyanobacterial fluxome plasticity and rigidity ........................................... 31
      3.3.6. Guidelines for cyanobacterial cell factories .............................................. 32
  3.4. Conclusions ........................................................................................................... 33
  Acknowledgement ......................................................................................................... 34
  Supporting information for Chapter 3 ........................................................................ 35
<table>
<thead>
<tr>
<th>Chapter 4. Enhancing fatty acid production in <em>Escherichia coli</em> by <em>Vitreoscilla</em> Hemoglobin overexpression</th>
<th>47</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract ...........................................................................</td>
<td>48</td>
</tr>
<tr>
<td>4.1. Introduction ..........................................................</td>
<td>48</td>
</tr>
<tr>
<td>4.2. Materials and Methods ..............................................</td>
<td>49</td>
</tr>
<tr>
<td>4.2.1. Chemicals and Strains ...........................................</td>
<td>49</td>
</tr>
<tr>
<td>4.2.2. Plasmid construction ............................................</td>
<td>50</td>
</tr>
<tr>
<td>4.2.3. Medium and culture conditions ..................................</td>
<td>50</td>
</tr>
<tr>
<td>4.2.4. Fatty acids, glucose, and acetate measurements .............</td>
<td>51</td>
</tr>
<tr>
<td>4.2.5. FBA Simulation ..................................................</td>
<td>52</td>
</tr>
<tr>
<td>4.3. Results and Discussion ............................................</td>
<td>52</td>
</tr>
<tr>
<td>4.4. Conclusion ............................................................</td>
<td>59</td>
</tr>
<tr>
<td>Acknowledgement ............................................................</td>
<td>59</td>
</tr>
<tr>
<td>Chapter 5. Deciphering <em>Clostridium</em> metabolism and its responses to bioreactor mass transfer during syngas fermentation</td>
<td>60</td>
</tr>
<tr>
<td>Abstract ...........................................................................</td>
<td>61</td>
</tr>
<tr>
<td>5.1. Introduction ..........................................................</td>
<td>61</td>
</tr>
<tr>
<td>5.2. Materials and Methods ..............................................</td>
<td>63</td>
</tr>
<tr>
<td>5.2.1. Strain, medium and culture preparation ......................</td>
<td>64</td>
</tr>
<tr>
<td>5.2.2. Syngas fermentation in serum bottle systems ................</td>
<td>64</td>
</tr>
<tr>
<td>5.2.3. Syngas fermentation in bioreactor systems .................</td>
<td>67</td>
</tr>
<tr>
<td>5.2.4. Isotopomer analysis .............................................</td>
<td>67</td>
</tr>
<tr>
<td>5.3. Results and Discussion ............................................</td>
<td>68</td>
</tr>
<tr>
<td>5.3.1. Importance of yeast extract for P7 growth ..................</td>
<td>68</td>
</tr>
<tr>
<td>5.3.2. Analysis of functional pathways in P7 .........................</td>
<td>72</td>
</tr>
<tr>
<td>5.3.3. P7 growth and bio-production during syngas fermentation</td>
<td>74</td>
</tr>
<tr>
<td>5.3.4. P7 metabolism during syngas fermentation under different gas flow rates</td>
<td>76</td>
</tr>
<tr>
<td>5.4. Conclusion ............................................................</td>
<td>81</td>
</tr>
<tr>
<td>Acknowledgement ............................................................</td>
<td>82</td>
</tr>
<tr>
<td>Chapter 6. Modeling and machine learning of <em>Clostridium</em> fermentations under different syngas compositions</td>
<td>83</td>
</tr>
<tr>
<td>Abstract ...........................................................................</td>
<td>84</td>
</tr>
<tr>
<td>6.1. Introduction ..........................................................</td>
<td>84</td>
</tr>
<tr>
<td>6.2. Materials and Methods ..............................................</td>
<td>85</td>
</tr>
<tr>
<td>6.2.1. Strain, medium and seed cultures .............................</td>
<td>85</td>
</tr>
<tr>
<td>6.2.2. Analyses of cell growth and syngas fermentation products</td>
<td>86</td>
</tr>
<tr>
<td>6.3. Results and Discussion ............................................</td>
<td>86</td>
</tr>
<tr>
<td>6.3.1. Flux balance analysis ...........................................</td>
<td>86</td>
</tr>
<tr>
<td>6.3.2. Empirical kinetic model ..........................................</td>
<td>88</td>
</tr>
<tr>
<td>6.3.3. Machine Learning model ..........................................</td>
<td>92</td>
</tr>
<tr>
<td>6.4. Conclusion ............................................................</td>
<td>98</td>
</tr>
</tbody>
</table>
Chapter 7. Investigation of *Pseudomonas aeruginosa* biofilm metabolism using $^{13}$C tracing..........................................................................................................................99

Abstract..................................................................................................................................................................100

7.1. Introduction................................................................................................................................................100

7.2. Materials and Methods................................................................................................................................101
  7.2.1. Strain Construction.................................................................................................................................101
  7.2.2. Strain, medium and culture preparation...............................................................................................102
  7.2.3. Dynamic labeling in petri dish system.................................................................................................104
  7.2.4. Metabolic flux analysis.........................................................................................................................105
  7.2.5. Proteinogenic amino acids isotopomer analysis..................................................................................106
  7.2.6. Free metabolites isotopomer analysis..................................................................................................107

7.3. Results and Discussion..............................................................................................................................107
  7.3.1. Fluxomes of *Pseudomonas aeruginosa* as planktonic and biofilm.....................................................107
  7.3.2. Variability of central catabolic fluxomes under planktonic and biofilm conditions.............................110
  7.3.3. Dynamic labeling for free metabolites of *Pseudomonas aeruginosa* as planktonic and biofilm........111

7.4. Conclusion..................................................................................................................................................112

Supporting Information for Chapter 7.............................................................................................................114

Chapter 8 Conclusions and Future Directions..............................................................................................119
  8.1. Conclusions.............................................................................................................................................120
  8.2. Future Directions...................................................................................................................................121

References.........................................................................................................................................................123

Appendix Chapter 1. Simulating cyanobacterial phenotypes by integrating flux balance analysis, kinetics, and a light distribution function......................................................................................132

Appendix Chapter 2. Integrating MS1 and MS2 Scans in High-Resolution Parallel Reaction Monitoring Assays for Targeted Metabolite Quantification and Dynamic $^{13}$C-Labeling Metabolism Analysis.................................................................144
List of Figures

Figure 2.1. Photomixotrophic metabolism of cyanobacteria.................................11
Figure 2.2. Cyanobacterial FBA predictions of photoautotrophic and photomixotrophic
growth. ...........................................................................................................13
Figure 2.3. FBA prediction of cyanobacterial lactate production under (a) 
photoautotrophic and (b) photomixotrophic conditions....................................14
Figure 3.1. _Synechocystis_ 6803 growth under different light and carbon conditions 
(n=2) ...........................................................................................................22
Figure 3.2. Fluxome of _Synechocystis_ 6803 in the dark condition .......................23
Figure 3.3. Detection of CO$_2$ fixation in _Synechocystis_ 6803 cultures under darkness…25
Figure 3.4. Ratio of _Synechocystis_ 6803 metabolite pool size to that of _E. coli_........28
Figure 3.5. _Synechocystis_ 6803 growth under aerobic and anaerobic condition after an 8- 
day cultivation .................................................................................................30
Figure 3.6. _Synechocystis_ 6803 fluxomes under different growth conditions...........32
Figure 4.1. Cell growth and fatty acid production from shake tube cultures (10 mL 
culture) ...........................................................................................................54
Figure 4.2. (a) Biomass growth and (b) fatty acid production of FAC, FAR1, and FAR2 
at 72 h post induction. c) Relative RFP expression level of FAR1 and FAR2........57
Figure 4.3. Effect of oxygen uptake flux and maintenance energy (i.e., ATP consumption 
for cell maintenance) on fatty acid yields ....................................................58
Figure 5.1. Growth of P7 cells in the basal medium containing different combinations of 
yeast extract (YE) and carbon sources (Test 1, Table 5.1).................................69
Figure 5.2. Labeling of proteinogenic amino acid (not fragmented, [M-57]$^+$ or [M-15]$^+$) 
after P7 growth in $^{13}$C-labelled substrate for two days (Test 2, Table 5.1)...........71
Figure 5.3. Major metabolic pathways (the Wood Ljungdahl pathway, the TCA cycle, the 
pyruvate metabolism) and carbon transition identified in P7 strain ..............72
Figure 5.4. Dynamic labeling of key metabolites under serum bottle condition (Test 4, 
Table 5.1) ....................................................................................................74
Figure 5.5. Cell growth and medium pH changes during syngas fermentation of P7 strain 
under different gas flow rates (Test 5, Table 5.1) ...........................................76
Figure 5.6. Relative abundance of labeled and un-labelled key metabolites of P7 cell in 
inverse dynamic labeling tests (Tests 6 & 7, Table 5.1) .................................78
Figure 5.7. Enrichment of the un-labelled proteinogenic amino acids in the inverse 
dynamic labelling syngas fermentation ..........................................................80
Figure 6.1. Prediction of optimal rates of biomass growth (figure column 1), acetate 
production (figure column 2), and ethanol production (figure column 3) .............88
Figure 6.2 Experimental results of biomass, acetate and ethanol production under various 
syngas composition .......................................................................................91
Figure 6.3. Sensitivity test of kinetic model about syngas fermentation performance… 92
Figure 6.4. Mean Square Error of Machine Learning model…………………………… 93
Figure 6.5. Comparison of machine learning data fitting with experimental data…… 94
Figure 6.6. Comparison of machine learning predictions with previous study experimental data………………………………………………………………………………… 95
Figure 6.7. Analysis of neural network model performance…………………………… 97
Figure 7.1. Reactor platform…………………………………………………………………… 105
Figure 7.2. Fluxome of *Pseudomonas aeruginosa* as planktonic and biofilm……… 109
Figure 7.3. Comparison of amino acids labeling between wild type and EPS-overexpressing mutant……………………………………………………………………… 111
Figure 7.4. Dynamic labeling for biofilm (A) wild type and (B) planktonic cells………………………………………………………………………………………… 112
List of Tables

Table 4.1. Strains and plasmids used in this study.................................53
Table 5.1. Summary of experimental design for various labeling tests in both serum bottle
cultures and bioreactor cultures...........................................................................66
Table 6.1. Gas compositions used in syngas fermentation of *C. carboxidivorans* P7 a,
b.................................................................................................................................85
Table 7.1. Strains and plasmids used in this study..............................................103
Acknowledgments

First, I would like to thank my advisor Dr. Yinjie Tang for his help and guidelines on my research. He taught me how to find the way to research, and his kindness help gave me a direction for future career.

Second, I would like to thank Dr. Zhiyou Wen for his valuable help with my syngas fermentation research, which leads to the core achievements in my PhD work. I am also grateful to Drs. Philip Bayly, David Peters, Fuzhong Zhang and Spencer Lake kindly serving on my thesis committees, giving their valuable time and precious suggestions in helping improve my research and supporting my future career.

Third, I want to thank Professor James Ballard from the writing center for his great help on improving my writing of almost all my manuscripts. I also want to acknowledge my talented colleagues in the laboratory. The free exchange of ideas and skills is a necessary ingredient of my PhD. I also want to thank all my friends and other colleagues in the department for making my PhD life colorful and enjoyable. And I want to thank grant NSF (CBET 1438042) to support my PhD work.

Last but the most important, I want to give my special thanks to my parents and family for their un-conditional understanding, unreserved support, and endless love.

Ni Wan

Washington University in St. Louis

May 2018
Dedicated to my parents and family.
ABSTRACT OF THE DISSERTATION

Application of metabolic modeling and machine learning for investigating microbial systems

by

Ni Wan

Doctor of Philosophy in Mechanical Engineering and Materials Science

Washington University in St. Louis, 2018

Professor Yinjie Tang, Chair

Metabolic modeling is an important tool to interpret the comprehensive cell metabolism and dynamic relationship between substrates and biomass/bioproducts. Genome-scale flux balance model and $^{13}$C-metabolic flux analysis are metabolic models which can reveal the theoretical yield and central carbon metabolism under various environmental conditions. Kinetic model is able to capture the complex principles between the change of biomass growth and bioproducts accumulation with the time series. Machine learning model is a data driven approach to reveal fermentation behavior and further predict cell performance under complex circumstances. In my PhD study, modeling analysis and machine learning method have been used to exam non-conventional microbial systems. (1) decode the functional pathway and carbon flux distribution in *Cyanobacteria* and *Clostridium* species for bio productions, (2) characterize biofilm physiologies and biodiesel fermentations (engineered *E.coli*) under mass transfer limitations, and (3) optimize syngas fermentations by deciphering and overcoming rate limiting process factor.
Chapter 1. Introduction
1.1. Background

1.1.1 Metabolic model

Metabolic modeling is an important tool to describe microorganism behavior, including microbial biomass growth, high-valued bioproducts and central carbon metabolism. Thereby, the development of model would provide insights into cell growth mechanism and further optimize the design and operation of experiments. “The model is often a simplified description of relationships between observations of the system (responses) and the factors that are believed to cause the observed responses.” were stated in the previous paper (McMeekin et al. 2008). Flux balance model, $^{13}$C-flux balance analysis, kinetic model and machine learning model was applied and developed to study the microbial systems, including *Cyanobacteria*, *E.coli*, *Clostridium* and *Pseudomonas*.

FBA model and $^{13}$C-MFA model are methods which can examine fluxes in metabolic pathways under pseudo-steady state. The difference is that stoichiometry-based FBA could predict theoretical biomass and product yields as the function of carbon substrates (Orth et al. 2010), while $^{13}$C-MFA integrates the $^{13}$C-assisted metabolism analysis with computational modeling to quantify the metabolic flux distribution *in vivo* (Dauner et al. 2001). Kinetic model and machine learning methods can describe the dynamic process. Kinetic model is able to capture dynamic relationship between the environmental changes (e.g. fermentation time, temperature, pH) and biomass growth and bioproducts (He et al. 2015). Machine learning methods can efficiently build predictive models for processes in which fundamental mechanistic understanding is limited (Pappu and Gummadi 2016).
1.1.2 Microbial system

Microorganism can serve as a production facility through optimization process which can produce the high-valued products, including ethanol, butanol, butyric acid. Carbon source was driven by energy molecules to go through the metabolic pathways, thereby forming the bioproducts.

In this study, four strains were included, *Cyanobacteria, E.coli, Clostridium and Pseudomonas*. *Cyanobacteria*, the ancestors of chloroplasts and eukaryotic algae, have powerful photosynthesis capabilities and utilize inorganic carbon CO₂ as well as organic carbon to grow autotrophically, mixotrophically and heterotrophically (Tomitani et al. 2006). Natural products and biofuels (e.g. sugar, isoprene, alcohols, alkanes and hydrogen) and engineered products (e.g. D-lactate) (Ducat et al. 2011; Varman et al. 2013) were the bioproducts derived from cyanobacteria. *E.coli* is widely used as a model host specie for industrial applications, such as biofuels, organic acids, amino acids, sugar alcohols and biopolymers (Chen et al. 2013). *Clostridium carboxidivorans* p7, served as biocatalyst, is commonly used in syngas fermentation, which is an environmental-friendly and economically efficient process to convert the synthetic gas (including CO, CO₂ and H₂) to high-valued products (Shen et al. 2014a). Biofilm, a group of microorganisms (such as *Shewanella oneidensis*, *Pseudomonas aeruginosa*, *Streptococcus mutans*, *Enterococcus faecalis*) stick to each other on a surface, would cause severe environmental problems, including water pollution (Bachmann and Edyvean 2005), soil pollution, metal corrosion and even do harm to human health (Hall-Stoodley et al. 2004).

Deciphering and modeling the function and metabolic features of microbial system would offer guidelines and new insights into optimal bioconversions, and father reveal flux topology between the planktonic cells and biofilm cell/engineered cells.
1.1.3. $^{13}$C-based metabolism analysis

$^{13}$C-based metabolism analysis is widely used to examine metabolic features and further provide the confirmation for the genomics and transcriptomics studies (Zamboni et al. 2009). Two methods were usually applied, one is $^{13}$C fingerprint, which examined the labeling pattern of proteinogenic amino acids under pseudo-steady state (Tang et al. 2012; You et al. 2014), the other is detecting the dynamic labeling change of free metabolites (Hollinshead et al. 2016b). In brief, microorganisms were fed by $^{13}$C carbon source, which led to the labeling of proteinogenic amino acids, organic acids and free metabolites. GC-MS and LC-MS measurement were mostly common used through dynamic pulse trace experiment and $^{13}$C fingerprint to detect abovementioned metabolites. Based on labeling pattern of amino acids and free metabolites, metabolic features and cellular functions have been revealed. To understand the correlation of genome sequencing and relative functional pathway, $^{13}$C techniques were widely used in biology field (Zamboni and Sauer, 2009). This technology is developing along with the improvement of advanced analytical instruments, including LC-MS-MS and LC-TOFMS.

1.1.4. Limitations of previous studies and general scopes of current studies

Metabolic modeling and machine learning is widely spread in various biological systems (Baart and Martens 2012; Bordbar et al. 2014; Dale et al. 2010; Voit 2017). However, the application of metabolic modeling to describe the real system still exists some gaps. Firstly, due to the environmental perturbation, the pathway performance is unable to perfectly match with the genome annotations. Secondly, some bioprocess modeling cannot perfectly provide an understanding of intracellular enzyme functions and metabolic fluxes. Thirdly, equations for the model lacks the completeness to describe the whole structure of the bioprocess. Considering these,
we were intrigued to apply various model to describe the metabolic profiles combining the environmental conditions as well as develop novel model to describe the bioprocess.

My research aims to study the microbial systems using the metabolic modeling and machine learning. Various microbial systems were quantified by the genome-scale model and $^{13}$C-MFA model, and thereby provided theoretically yield of biomass and bioproducts as well as the central metabolism flux. Then genome-scale model and $^{13}$C-MFA got further constrained by the experimental data or maintenance energy to obtain the better description of the pseudo steady state performance. Finally, empirical kinetic model and machine learning model were developed to describe the kinetic process of bioconversion outcomes.

1.2. Dissertation Overview

Metabolic modeling and machine learning is a useful tool to describe and examine the bioprocess. In Chapter 2, Flux balance analysis model compares cyanobacterial photomixotrophic and photoautotrophic carbon metabolism, and further provides the advantages of photomixotrophic for biosynthesis.

In Chapter 3, the flux distribution and energy molecules generation of cyanobacteria under heterotrophic growth was investigated. FBA model further got constrained to predict and verify the fluxome plasticity.

In Chapter 4, since $^{13}$C-MFA study suggested that energy metabolism is the rate-limiting step for fatty acid overproducing, in this case, membrane protein VHB which can provide more oxygen was engineered in E.coli. Experimental results showed that fatty acid production was improved by 70% higher than wild type. FBA model further raised cost-benefit tradeoff between metabolic engineering and metabolic burden based on the experimental results.

Syngas fermentation is an environmental friendly and economical process to convert
synthetic gas to high-valued bioproducts. In Chapter 5 and 6, the carbon metabolism and functional pathway were firstly investigated, then three models, FBA model, empirical kinetic model and machine learning model was applied in this study to describe the syngas fermentation process and further offer the insights into optimize the syngas fermentation process.

In Chapter 7, biofilm and planktonic cell metabolism of *Pseudomonas* PAO1 was revealed by $^{13}$C-MFA. Dynamic pulse trace results further verified the $^{13}$C-MFA results.

Finally, Chapter 8 discussed the major findings and implications of current study in metabolic modeling and machine learning field.
Chapter 2. Cyanobacterial Photo-driven Mixotrophic Metabolism and its advantages for Biosynthesis

Results of this chapter have been published in Front. Chem. Sci. Eng. 2015
Abstract

Cyanobacterium offers a promising chassis for phototrophic production of renewable chemicals. Although engineered cyanobacteria can achieve similar product carbon yields as heterotrophic microbial hosts, their production rate and titer under photoautotrophic conditions are 10 to 100 folds lower than those in fast growing *E. coli*. Cyanobacterial factories face three indomitable bottlenecks. First, photosynthesis has limited ATP and NADPH generation rates. Second, CO$_2$ fixation by ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) has poor efficiency. Third, CO$_2$ mass transfer and light supply are deficient within large photobioreactors. On the other hand, cyanobacteria may employ organic substrates to promote phototrophic cell growth, N$_2$ fixation, and metabolite synthesis. The photo-fermentations show enhanced photosynthesis, while CO$_2$ loss from organic substrate degradation can be reused by the Calvin cycle. In addition, the plasticity of cyanobacterial pathways (e.g., oxidative pentose phosphate pathway and the TCA cycle) has been recently revealed to facilitate the catabolism. Photoautotrophic performance and photomixotrophic performance were compared by flux balance model, thereby proving the advantage of the photomixotrophic growth. The use of cyanobacteria as “green *E. coli*” could be a promising route to develop robust photobiorefineries.

2.1 Introduction

Cyanobacteria are the ancestors of chloroplasts and eukaryotic algae. They have powerful photosynthesis capabilities and are widespread in both land and water habitats (Tomitani et al. 2006). Cyanobacteria can grow under autotrophic, mixotrophic and heterotrophic conditions. Under mixotrophic conditions, cyanobacteria can simultaneously utilize both inorganic carbon sources, such as CO$_2$ via the Calvin cycle, and organic carbon sources, such as glucose and acetate.
Recently, engineering cyanobacteria with exogenous genes to produce industrially important compounds and reduce CO$_2$ emissions has been recently raised (Yu et al. 2013a). For example, the biotechnology company, Algenol, uses cyanobacteria to manufacture ethanol as biofuel, and Photanol uses cyanobacteria to produce lactic acids for bio-plastics. However, fast cyanobacterial growth and biosynthesis rates are only observed in optimal laboratory conditions. Photobioreactor operations suffer major challenges: limited light harvesting efficiency, biomass self-shading, and barriers in CO$_2$ gas-liquid mass transfer. Therefore, photobioreactors cannot achieve high biomass concentrations or desired productivity, and their working volume has to be restricted to ensure substantial surface area for light harvesting. Moreover, CO$_2$ aeration (energy-intensive gas-pumping) adds to the utility costs and the risk of microbial contaminations. *Cyanobacteria* are still difficult to be used for manufacturing products at industrial scales.

On the other hand, many cyanobacteria contain pathways for heterotrophic utilization of organic carbons, such as urea (as both C and N source), sugars, and acetate. For example, *Synechocystis* 6803 autotrophic culture has a doubling time of $>8$ h when the light condition is insufficient. However, its light limiting culture can grow faster with glucose. The organic substrates generally promote cyanobacterial biosynthesis rates, while the cell still maintains CO$_2$ fixation activity. In this review, we discuss the influence of photomixotrophic metabolism on CO$_2$ fixation, carbon fluxes, photosynthesis, nitrogen assimilation, and product synthesis in cyanobacteria. Via a flux balance model $i$JN678 (Nogales et al. 2012), we demonstrate advantageous aspects of photomixotrophic metabolism for product synthesis.

## 2.2 Light-harvesting apparatuses

Cyanobacterial thylakoids form stacks of parallel sheets close to the cytoplasmic membrane. The space between the thylakoids is the PBSs (phycobilisomes). Cyanobacterial growth rate is
driven by light energy (Figure 2.1). Antennas were used to transfer the excitation energy to photosystems (Liu et al. 2013). However, at high light intensity, photoinhibition can occur. The cell scatters excess irradiation using nonphotochemical quenching (such as heat generation) to protect the cyanobacterial reaction center. The mechanisms of photoinhibition have been studied (Aro et al. 1993; Nishiyama et al. 2004). It has been shown that light stress leads to the inactivation of electron transport and the subsequent oxidative damage of the reaction center of PSII. Reactive oxygen species (e.g., singlet oxygen generated during photosynthesis) can also inhibit the repair cycle of the photo-damaged PSII’s reaction center. To reduce photooxidative stresses, cyanobacteria employ an alternative electron flow pathway (i.e., water-water cycle) that uses electrons extracted from water by PSII to re-generate H₂O in PSI. This reaction dissipates excess excitation energy and reduces NADPH yields, but moderately enhances photosynthetic ATP generation and oxygen evolution (Hasunuma et al. 2014). Due to photoinhibition, the surface illumination condition has to be confined to avoid damaging the light-sensitive PSII centers in photobioreactors.
Moreover, cyanobacteria show inefficient use of available light energy in large photobioreactors. The cells at the surface of the incident light have saturated photosynthesis or photo-inhibitions; while the cells at the surface shade those below. By truncating the light harvesting antennas, it is expected to reduce the waste of light energy (Page et al. 2012). Unlike green algae that uses chlorophyll-based antennas, cyanobacterial PBSs are very elegant. Photon excitation and electron transport was affected by the antenna size and arrangements. Genetic modification of PBSs to harvest less light but still retain same photosynthesis activity is difficult.

In summary, the theoretical maximum efficiency of solar energy conversion by photosynthesis is 12%. However, the photon flux utilization by the light-harvesting apparatuses is below 100 mmol/gDW\(^{-1}/h\)\(^{-1}\) (Nogales et al. 2013). Taking into account the known losses in light harvesting, cyanobacterial photosynthesis in optimal bubbled bioreactors can only achieve an efficiency of ~6% (three times below photovoltaic efficiencies of silicon solar cells) (Blankenship...
2.3 FBA simulation under various conditions

To reveal cyanobacterial metabolism and theoretical yield in industrial applications, an FBA model for *Synechocystis* sp. PCC 6803 (iJN678) compares photomixotrophic metabolism and photoautotrophic metabolism (Figure 2.2) as well as theoretically bioproducts (Figure 2.3). The FBA assumes the maximal CO$_2$ uptake rate as 4.0 mmol/gDW$^{-1}$/h$^{-1}$ and the light uptake rate as 50 mmol/gDW$^{-1}$/h$^{-1}$. In autotrophic growth, CO$_2$ and light availability constrains the growth rate below 0.08 h$^{-1}$ (Fig. 2.2(a)). Input of cyanobacterial glucose uptake flux results in larger growth rates (Figs. 2.2 (b–d)). In mixotrophic conditions, biomass growth rate increases with an increase in light influx. If the glucose uptake rate is 2 mmol/gDW$^{-1}$/h$^{-1}$ (Fig. 2.2(d)), *Synechocystis* 6803 can achieve a growth rate around 0.28 h$^{-1}$ (doubling time = 2.5 h; similar to the growth rate of common yeasts) and a high biomass yield (~0.8 g dry cell per g glucose). By increasing glucose uptake to 2 mmol/gDW$^{-1}$/h$^{-1}$, CO$_2$ influx has negligible limitations to biomass growth (Fig. 2.2(d)). This is because the loss of CO$_2$ from glucose catabolism can be re-fixed by the Calvin cycle, forming a closed loop for sugar carbon utilization inside of the cell. Furthermore, we can use the same model to compare *Synechocystis* 6803 lactic acid synthesis under photoautotrophic and photomixotrophic conditions. If the cell growth rates are set as same, the model indicates that co-utilization of a small amount of glucose (0.38 mmol/gDW$^{-1}$/h$^{-1}$) can not only promote lactate production rates, but also allow the strain to produce lactate under light and CO$_2$ limited conditions (Fig. 2.3). These model results indicate that photo-fermentation may ultimately remove the roadblocks for low production titer and rates.
Figure 2.2 Cyanobacterial FBA predictions of photoautotrophic and photomixotrophic growth.

(a) Panel shows the growth under the photoautotrophic condition. The mixotrophic conditions are constrained by glucose uptake rate of (b) 0.38 mmol/gDW\(^{-1}\)/h\(^{-1}\), (c) 1 mmol/gDW\(^{-1}\)/h\(^{-1}\) or (d) 2 mmol/gDW\(^{-1}\)/h\(^{-1}\). The flux distributions are obtained by setting the biomass objective function (Nogales et al. 2012).
Figure 2.3 FBA prediction of cyanobacterial lactate production under (a) photoautotrophic and (b) photomixotrophic conditions.

Growth rates under both photoautotrophic and photomixotrophic conditions are fixed as 0.02 h⁻¹. The photomixotrophic condition is constrained by glucose uptake rate as 0.38 mmol/gDW⁻¹/h⁻¹. The flux results are obtained by maximizing the D-lactate production.

2.4 Conclusion

Cyanobacteria are attractive over E. coli because they require simple nutrients and reduce the competition with food crops for land. However, photosynthesis has limitations, such as cell self-shading, CO₂ membrane transport, photoinhibition, etc. This creates challenges in the design of an optimal photobioreactor for large-scale processes. Thereby, cyanobacterial production process is slow, which increases the risk of contaminations and product losses. Although the photo-fermentation would increase the cost of feedstock, it reduces the costs for photobioreactor operations and product separations. Alternatively, the photo-fermentation with organic substrates can improve product rates and titer by reducing the dependence of productivity on light and CO₂.
Chapter 3. Cyanobacterial carbon metabolism: fluxome plasticity and oxygen dependence

Results of this chapter have been published in Biotechnology and bioengineering 2017, 114, 1593-1602
Abstract

*Synechocystis* sp. strain PCC 6803 has been widely used as a photo-biorefinery chassis. Based on its genome annotation and homology, this species contains a complete TCA cycle, an Embden-Meyerhof-Parnas pathway (EMPP), an oxidative pentose phosphate pathway (OPPP), and an Entner-Doudoroff pathway (EDP). To evaluate how *Synechocystis* 6803 catabolizes glucose under heterotrophic conditions, we performed $^{13}$C metabolic flux analysis, FBA model analysis and metabolite pool size analysis. The results revealed a strong cyclic mode of flux through the OPPP. Small, but non-zero, fluxes were observed through the TCA cycle and the malic shunt. Anaerobic conditions (high and low light) constrained the growth of *Synechocystis* 6803, leading to the finding that oxidative phosphorylation is required for cell growth and that no alternate anaerobic mechanisms for ATP/NADPH generation exist. The pool sizes of intermediates in the TCA cycle, particularly acetyl-CoA, were found to be several folds lower in *Synechocystis* 6803 (compared to *E. coli* metabolite pool sizes), while its sugar phosphate intermediates were several folds higher. Comparing photoautotrophic, photomixotrophic and heterotrophic conditions, the Calvin cycle, OPPP, and EMPP in *Synechocystis* 6803 possess the ability to regulate their fluxes under various growth conditions (plastic), whereas its TCA cycle always maintains at low levels (rigid).

3.1 Introduction

*Synechocystis* sp. strain PCC 6803 (hereafter *Synechocystis* 6803) is a model cyanobacterial species and a key photo-biorefinery host for biofuel and chemical production (Ducat et al. 2011). However, photoautotrophic bio-refineries are difficult to scale up, because
light penetration into large bioreactors is often limited. Additionally, large-scale outdoor cultures are subject to light/dark cycles which require flexible growth strategies. *Synechocystis* 6803 has complete glucose catabolic routes, which can supplement its photosynthetic abilities. Under photomixotrophic conditions, this strain can utilize glucose (via the EMPP and OPPP) and CO$_2$ (via the Calvin Cycle) simultaneously (You et al. 2014). Recent work, based on gene knockouts, enzyme assays and other data, also found that *Synechocystis* 6803 contains a functional Entner-Doudoroff pathway (EDP) under photomixotrophic conditions (Chen et al. 2016). When photosystem II is impaired, this strain employs the oxidative pentose phosphate pathway (OPPP) for NADPH generation, and cyclic electron flow around photosystem I and oxidative phosphorylation for ATP synthesis (Nakajima et al. 2014; Wan et al. 2015). Moreover, *Synechocystis* 6803 can grow heterotrophically in the dark, but requires intermittent light-pulses to activate functional genes, though this light-pulse is not sufficient for photoautotrophic growth (Anderson and McIntosh 1991). Although *Synechocystis* 6803 lacks α-ketoglutarate dehydrogenase, it can use the γ-aminobutyrate acid (GABA) shunt to convert α-ketoglutarate to succinate (minimal measured flux in this study), allowing for a functional TCA cycle (Xiong et al. 2014). *Synechocystis* 6803’s TCA cycle is repressed by photosynthesis (You et al. 2014; Young et al. 2011b), but it is predicted by flux balance analysis (FBA) to play a key role under dark heterotrophic conditions (Nogales et al. 2012).

*Synechocystis* 6803 has the potential to behave like a “green” photosynthetic *E. coli*, as it is able to use glucose, is genetically tractable, and produces a wide range of products (Berla et al. 2013). Previous work has shown that up to 40% of *Synechocystis* 6803’s genes are modulated during regular diurnal cycles, and thus we expect the fluxome to shift under the previously uninvestigated dark condition when an organic carbon source is used for growth (Saha et al. 2016).
We analyzed the plasticity of fluxomic topologies and metabolite pool sizes in *Synechocystis* 6803, and performed genetic modifications to reveal responsive pathways in cyanobacteria for survival under different suboptimal growth environments, including darkness and anaerobic conditions that can occur locally or temporarily in large-scale photo-biorefineries. In conjunction with the advent of synthetic biology, alternative or synthetic metabolic pathways have been inserted into microbes to enhance production of various molecules (Bar-Even et al. 2013; Bogorad et al. 2013; Tai et al. 2016). This study provides an understanding of the basic principles governing flux topology in *Synechocystis* 6803 and will increase the successful, rational engineering of synthetic pathways.

### 3.2 Materials and Methods

#### 3.2.1 Cultivation conditions and biomass growth for flux and metabolite analysis.

*Synechocystis* 6803 was grown in a modified BG-11 medium (with no organic carbon source) at 30°C under two continuous light conditions (~10 and ~50 µE/m²/s). For photoautotrophic and photomixotrophic conditions, *Synechocystis* 6803 was grown with atmospheric CO₂ in 50 mL shaking flasks with a working volume of 10 mL (shaking at 200 rpm). We also grew *Synechocystis* 6803 in an anaerobic chamber (Coy Lab. Products Inc., Ann Arbor, MI) filled with N₂ (90%) and H₂ (10%). The anaerobic cultures were supplemented with either NaHCO₃ (4 g/L) or both NaHCO₃ (4 g/L) and glucose (1 g/L). Carboxylic acids (e.g., acetate, glyoxylate, malate, oxaloacetate, pyruvate, and citrate) or amino acids (glycine or serine) were further tested to promote cell growth under varying conditions. For dark cultures, *Synechocystis* 6803 was grown in aerobic serum bottle (1 g/L glucose) that were wrapped with aluminum foil to maintain complete darkness conditions (shaking at 200 rpm). A pulse of light (~50 µE/m²/s) was given for ~10 min every 24 hours. Two
\(^{13}\text{C}\)-experiments were performed under dark conditions. First, *Synechocystis* 6803 was grown with naturally labeled glucose and NaH\(^{13}\text{CO}_3\) to examine CO\(_2\) fixation under dark conditions. Second, \([1-{^{13}\text{C}}_1]\) glucose (1 g/L) and 4 g/L NaH\(^{13}\text{CO}_3\) were used to quantify the heterotrophic metabolism via \(^{13}\text{C}\) metabolic flux analysis (\(^{13}\text{C}\)MFA). For culture experiments, all chemicals were purchased from Sigma (Saint Louis, MO, USA). Cell density was monitored by a UV-Vis spectrophotometer at 730 nm (one unit OD\(_{730}\) is equal to 0.4 g dry biomass/L). The cell doubling time was estimated during the exponential growth phase. Glucose and acetate concentrations were measured by enzyme kits (R-Biopharm, Darmstadt, Germany).

### 3.2.2 Isotopomer analysis of proteinogenic amino acids.

*Synechocystis* 6803 cultures were supplemented with \([1-{^{13}\text{C}}]\) glucose. The incorporation of labeled carbon into proteinogenic amino acids was measured by GC-MS (Hewlett-Packard model 7890A, Agilent Technologies) and later used in \(^{13}\text{C}\)-MFA. Biomass pellets were hydrolyzed by 6 M HCl at 100°C. The resulting mixtures were subsequently air-dried and derivatized with \(N\)-tert-butyldimethylsilyl-\(N\)-methytrifluoroacetamide (TBDMS) prior to GC-MS analysis (You et al. 2012). A published software was used to analyze and correct amino acid MS data (fragments of [M-57]^+\), [M-159]^+ or [M-85]^+\), and [f302]) (Wahl et al. 2004). Isotopomer labeling fractions (M0, M1, M2, etc.) represent fragments containing unlabeled, singly labeled, doubly labeled amino acids, and so on. Due to overlapping peaks or product degradation, several amino acids (proline, arginine, cysteine, and tryptophan) were not used for flux analysis (Antoniewicz et al. 2007).

### 3.2.3 Metabolite pool size determination.

Metabolite concentrations in *Synechocystis* 6803 under three different light conditions (dark condition, high light ~50 µE/m\(^2\)/s and low light ~10 µE/m\(^2\)/s) and two carbon conditions (autotrophic growth with 4 g/L NaHCO\(_3\), or growth with 1 g/L glucose and 4 g/L NaHCO\(_3\)) were
analyzed using an isotope ratio-based approach (Mashego et al. 2004). In brief, fully labeled *E. coli* metabolites, as internal $^{13}$C-standards, were prepared by growing *E. coli* MG1655 in M9 medium supplemented with 20 g/L U-$^{13}$C$_6$ glucose. *Synechocystis* 6803 cultures (15 mL, n=2) during the exponential phase (OD$_{730} = \sim 0.4$) were mixed with the same amount of biomass of fully labeled and exponentially growing *E. coli*. The isotopic ratio of each metabolite ($^{13}$C-labeled vs. unlabeled) was normalized by biomass to obtain the relative cyanobacterial metabolite pool sizes by benchmarking them against *E. coli*. The full description of this procedure can be found in the Supplementary Method (Part 1).

### 3.2.4 Metabolic flux analysis of *Synechocystis* 6803 under dark conditions.

$^{13}$C-MFA model included the EMPP, the EDP, the Calvin cycle, the TCA cycle, the glyoxylate shunt, photorespiration pathways, and biomass synthesis. Flux calculation was based on isotopomer data from proteinogenic amino acids, which profiled relative fluxes through the central metabolism by normalizing the glucose uptake rate as 100. The relative fluxes were solved by minimizing a quadratic error function that calculated the differences between predicted and measured isotopomer patterns. Reaction reversibility was characterized by the exchange coefficient $exch$ and the net flux $\nu^{net}$ (Dauner et al. 2001; You et al. 2014). We also employed the genome-scale FBA model *i*JN678 to predict cell metabolism (Nogales et al. 2012), which contains more details about biotechnological endeavors, light-limited heterotrophic and photosynthetic metabolism. The full description of these procedures can be found in the Supplementary Method (Part 2).

### 3.3 Results and Discussion
3.3.1 Fluxomes under dark conditions.

*Synechocystis* 6803 growth was reduced under low light (~10 µE/m²/s) or dark conditions, even with the addition of glucose (1 g/L). Under these conditions, *Synechocystis* 6803 had a doubling time of ~20 hr, while photoautotrophic growth (light ~50 µE/m²/s) led to a doubling time of ~14 hr (Figure 3.1). To examine *Synechocystis* 6803’s metabolism under dark conditions, [1-13C] labelled glucose and NaH13CO3 were used for 13C metabolic flux analysis (13C-MFA). The model fitting to experimental data is shown in Figure 3-S1 and Supplementary Table 3-1. The relative flux distributions and standard deviations are shown in Figure 3.2 (upper line), while exchange coefficients for reversible reactions and 95% confidence interval are shown in Supplementary Table 3-2. Under dark conditions, glucose was exclusively utilized through the cyclic sugar phosphate pathway (G6P → pentose phosphates → F6P → G6P), forming a carbohydrate degradation loop. A similar cyclic degradation cycle has also recently been observed when *Synechocystis* 6803 was grown under light conditions and herbicide stresses (using DCMU or atrazine to block its photosystems) (Nakajima et al. 2014; You et al. 2015).
Figure 3.1. *Synechocystis* 6803 growth under different light and carbon conditions (n=2).

White bars are doubling times under light condition of ~50 µE m$^{-2}$ s$^{-1}$: autotrophic growth with ambient air (CTL); autotrophic growth with 4 g/L NaHCO$_3$ (CO$_2$); mixotrophic growth with glucose (GLC). Moreover, the 4g/L NaHCO$_3$ based mixotrophic culture was supplemented with citrate (CIT), α-ketoglutarate (AKG), succinate (SUC), fumarate (FUM), malate (MAL), oxaloacetate (OXA), or pyruvate (PYR), respectively. Black bars represent cell doubling times under low light condition or dark condition: 10 µE m$^{-2}$ s$^{-1}$ light condition (marked as Low) or dark conditions (marked as Dark). For organic carbon supplements, 1g/L concentration was employed.
Figure 3.2. Fluxome of *Synechocystis* 6803 in the dark condition.

The absolute glucose uptake rate was 0.41 mmol/g DCW/h, and specific growth rate $\mu$ was 0.023 h$^{-1}$ (Figure 3-S3). The fluxes were normalized to the glucose uptake rate (represented as 100), and the fluxes are represented as ‘best fit ± standard deviation’. The flux distribution predicted by an FBA model, normalized to a glucose uptake rate of 100, is shown below the $^{13}$C MFA
results in italics. The arrow thickness relates to the degree of flux. The white arrows represent the fluxes towards biomass synthesis. Abbreviation: 3-phosphoglycerate (3PG); 6-phosphogluconate (6PG); acetyl-CoA (AceCoA); dihydroxyacetone phosphate (DHAP); erythrose 4-phosphate (E4P); fructose 1,6-bisphosphate (FBP); fructose 6-phosphate (F6P); glucose 6-phosphate (G6P); glyceraldehyde 3-phosphate (GAP); glyoxylate (GLX); isocitrate (ICT); malate (MAL); oxaloacetate (OAA); phosphoenolpyruvate (PEP); pyruvate (PYR); ribose 5-phosphate (R5P); ribulose-5-phosphate (Ru5P); ribulose-1,5-diphosphate (RuBP); sedoheptulose 1,7-bisphosphate (S17BP); sedoheptulose-7-phosphate (S7P); succinate (SUC); and xylulose-5-phosphate (X5P).

In heterotrophic bacteria, the EMP and TCA cycles are a common strategy for sugar catabolism because of their thermodynamic favorability and balanced cellular demands for energy molecules and pathway intermediates (Bar-Even et al. 2012). However, the TCA cycle in *Synechocystis* 6803 exhibited a very small flux of carbon (relative flux less than 3%) that was mainly used for anabolic purposes. Although the TCA cycle in plant tissue cells (e.g., illuminated leaf) can be highly flexible (operating in a cyclic flux mode, a reverse flux mode from 2-oxoglutarate to citrate, or a branched non-cyclic mode) (Sweetlove et al. 2010), it appears that *Synechocystis* 6803 has very rigid flux through this pathway.

$^{13}$C-MFA also revealed minimal CO$_2$ fixation via RuBisCO under the dark condition. To investigate the CO$_2$ fixation, *Synechocystis* 6803 was grown in the dark with unlabeled glucose and either labelled NaH$^{13}$CO$_3$ or unlabeled NaHCO$_3$ (Figure 3.3). GC-MS analysis showed that the $^{13}$C from the sodium bicarbonate was incorporated into the biomass. Specifically, serine, which can be derived from the 3PG formed by RuBisCO-mediated CO$_2$ fixation, showed a higher $^{13}$C incorporation (2.4%) than that in the control group, which only incorporated naturally occurring $^{13}$C (1.2%). This observation suggests RuBisCO activity, but its overall CO$_2$ fixation flux is likely
limited by the availability of energy molecules (ATP and NADPH) under heterotrophic conditions. A previous study also revealed that RuBisCO can fix CO₂ without a fully functional Calvin Cycle, as might be found in the dark condition (You et al. 2015). NaH¹³CO₃ is also incorporated into aspartate and glutamate (6–8%), derivatives from the TCA cycle, through the PEP carboxylase reaction.

![Figure 3.3. Detection of CO₂ fixation in *Synechocystis* 6803 cultures under darkness.](image)

The cultures were grown on labelled NaH¹³CO₃ and unlabeled ¹²C glucose in the dark (black bar). The control cultures were grown with unlabeled NaHCO₃ and ¹²C glucose in the dark (white bar). Error bars represent the standard deviations of biological triplicates.

### 3.3.2 FBA simulations

We investigated the fluxome of cyanobacteria in darkness based on simulations of FBA model iJN678 (Nogales et al. 2012). The fluxes were estimated by maximizing the biomass accumulation. The FBA fluxes are shown below the ¹³C-MFA measurements in Figure 3.2 and are normalized to a glucose uptake flux of 100. The experimental measurement of glucose uptake rate
was used to constrain the glucose input flux. Similar to predictions based on $^{13}$C-MFA, FBA detected a highly active OPPP, with a flux of 0.76 mmol/g DW/h. Furthermore, the NAD(P)H formed by the OPPP produced 1.5 mmol/g DW/h ATPs via oxidative phosphorylation, indicating oxidative phosphorylation as a major pathway for ATP synthesis. On the other hand, FBA revealed several differing results compared to the $^{13}$C-MFA and the experimental knockout data. For example, the malic enzyme activity is predicted by FBA to be zero under dark conditions, whereas the MFA data demonstrated that it has non-zero flux and is a critically important pathway. The glycolysis pathway was also predicted by FBA to be functional, with flux from G6P to F6P, albeit with low activity. $^{13}$C-MFA demonstrated the stark opposite, with very strong flux in the opposite direction, F6P to G6P. Similarly, the FBA model makes an incorrect prediction that Synechocystis 6803 uses a highly functional TCA cycle under chemoheterotrophic conditions. The discrepancy between the FBA model and MFA model under the heterotrophic condition suggests that FBA models inherently cannot explain unusual growth conditions that lead to unexpected changes in metabolic topology. However, the FBA model still had a good prediction about normal lighted growth conditions, such as autotrophic condition and mixotrophic condition.

To reveal how growth rate responds to variations of glucose and photon influxes, we ran the FBA by setting the maximum CO$_2$ uptake rate as 3.7 mmol/g DW/h (Shastri and Morgan 2005); ranging light intensity from 0~80 mmol/gDW/h; and ranging glucose uptake rate from 0~2 mmol/gDW/h (Figure 3-S2). The specific growth rate of cyanobacteria increased with increased glucose uptake rate and light intensity, achieving a maximum growth rate of ~0.34 h$^{-1}$ (doubling time ~2.02 h) (Figure 3-S2A). Flux through the OPPP and the TCA cycle increased with glucose uptake rate and light intensity. When the non-growth associated maintenance (NGAM) energy is considered, that is, the flux of reaction ATP$\rightarrow$ATP$\rightarrow$ext would be constrained as
\[-1000 \text{ mmol/g DW/h} < v_1 < 6.9 \text{ mmol/g DW/h} \quad (3.1)\]

In this case, the maximum predicted biomass production is further reduced to 0.3 h^{-1} (Figure 3-S2B). Additionally, when we limited the upper bound of the pyruvate dehydrogenase reaction to 0.2 mmol/g DW/h to limit its acetyl-CoA availability, that is, the flux of pyruvate dehydrogenase reaction $\text{pyruvate} \rightarrow \text{AcetylcoA} + \text{CO}_2$, would be constrained as

\[-1000 \text{ mmol/g DW/h} < v_2 < 0.2 \text{ mmol/g DW/h} \quad (3.2)\]

The flux through the TCA cycle and the OPPP decreased and the optimal growth rate was reduced to 0.116 h^{-1} (Figure 3-S2C). This prediction infers that acetyl-CoA is a rate limiting node for cyanobacterial growth.

### 3.3.3 Pool size of metabolites.

*Synechocystis* 6803 under photomixotrophic conditions had a larger metabolite pool size than when under autotrophic or dark heterotrophic conditions. G6P in the glycolysis pathway, 6PG in the OPPP, and S17BP (Sedoheptulose-1,7-bisphosphate) in the Calvin cycle all have drastically larger pool sizes under photomixotrophic conditions (Figure 3.4), again demonstrating the plasticity of certain pathways and their upregulation under different growth conditions. This also suggests that glucose, which serves as both a carbon and energy source, can increase the metabolite pool sizes of sugar phosphates in *Synechocystis* 6803 (Figure 3.4). For most metabolites, pool sizes also had a positive correlation with light intensity. The TCA cycle intermediates, however, are always kept at low and constant levels across growth conditions, demonstrating the rigidity, or constant flux through this pathway in *Synechocystis* 6803. Acetyl-CoA, for example, has a pool size approximately 10 fold smaller than that of *E. coli*. In contrast, *Synechocystis* 6803 metabolites in sugar phosphate pathways, such as G6P and 6PG, exhibited larger pool sizes than *E. coli*. 
Figure 3.4. Ratio of *Synechocystis* 6803 metabolite pool size to that of *E. coli*.

Light intensities of 0, ~10 and ~50 µE m⁻² s⁻¹ are marked as Dark, ~10 and ~50, respectively. Two carbon conditions were used, one with 1 g/L glucose and 4 g/L NaHCO₃ (marked as Mix) and the other with 4 g/L NaHCO₃ (marked as Auto). The z-axis represents relative cyanobacterial metabolite pool sizes.

### 3.3.4 Investigation of ATP metabolism.

Cyanobacterial photosystems convert light energy into NADPH and ATP for growth and other cellular activities (Tamoi et al. 2005). Very few cyanobacterial species in the nature can grow heterotrophically in the dark, which originally led to theories concerning the lack of permeases and transporters or the lack of inducible enzymes for carbon catabolism (Richardson and Castenholz 1987). Under dark heterotrophic conditions, *Synechocystis* 6803 has the ability to
perform glucose catabolism (via cyclic sugar phosphate pathway) coupled with aerobic respiration to produce NADPH and ATP. Based on $^{13}$C-MFA, a total amount of 1.2 mmol/g DW/h ATP is consumed by the central metabolism and biomass synthesis. A strong cyclic sugar phosphate pathway yields 3.1 mmol/g DW/h NAD(P)H that must undergo oxidative phosphorylation for ATP synthesis. The results indicate that oxygen plays an essential role for cyanobacteria growth in dark heterotrophic conditions by converting NAD(P)H to ATP via oxidative phosphorylation.

To test how light intensity and oxygen state influences the energy metabolism, we performed three sets of experiments. First, we compared ATP pool sizes under autotrophic or mixotrophic conditions with normal light (50 $\mu$E/m$^2$/s), low light (10 $\mu$E/m$^2$/s), or no light. The ATP pool size under normal light, photomixotrophic conditions (50 $\mu$E/m$^2$/s) was higher than the other conditions (Figure 3.5). Second, we investigated the effects of oxygen under either normal light (50 $\mu$E/m$^2$/s) or light-limited (10 $\mu$E/m$^2$/s) conditions (Figure 3.5). The data infer that the absence of oxygen affected ATP production, as Synechocystis 6803 grew slower under light sufficient photoanoxic conditions (50 $\mu$E/m$^2$/s), but exhibited no growth under anoxic mixotrophic or autotrophic conditions with a surface light intensity at 10 $\mu$E/m$^2$/s. Under this light condition, FBA predicted ~6.9 mmol/g DW/h ATP from photosynthesis, which were consumed by non-growth associated maintenance. Third, we tested supplementing light-limited (10 $\mu$E/m$^2$/s) anaerobic cultures to enhance growth (Figure 3.5). Supplements included photo-respiration intermediates (glyoxylate or glycine), which can alleviate redox imbalance and amino acid biosynthesis (Allan et al. 2009), and exogenous pigments isolated from spinach (contains both chlorophyll $a$ and chlorophyll $b$: ratio = 2.52) to enhance photosynthesis. However, all tested mixotrophic/autotrophic cultures exhibited poor growth under anaerobic conditions compared to aerobic conditions.
Figure 3.5. *Synechocystis* 6803 growth under aerobic and anaerobic condition after an 8-day cultivation.

*Synechocystis* 6803 was grown with 4 g/L NaHCO₃ either in aerobic conditions (white bars) or in the anaerobic chamber filled with N₂ and H₂ (black bars). The final cell density of cultures under aerobic (control) and anaerobic conditions are compared with varying supplements: glucose (1 g/L), glyoxylate (GLX; 5 mg/L), glycine (GLY; 5 mg/L), and spinach pigment. The initial OD₇₃₀ was ~0.029. The high light intensity was 50 µE m⁻² s⁻¹ and low light intensity was 10 µE m⁻² s⁻¹.

*E. coli* has the ability to obtain ATP under anaerobic conditions via acetate overflow: acetyl-CoA, unable to enter the TCA cycle due to lack of oxygen and redox cofactor regeneration, is converted to acetate (Vemuri et al. 2006). *Synechocystis* 6803 has the gene *slr0453*, hypothetically encoding a phosphoketolase, which should lead to acetate production (Xiong et al. 2015a). However, acetate secretion by *Synechocystis* 6803 under heterotrophic conditions was found to be negligible. Unlike *E. coli*, *Synechocystis* 6803 cannot obtain sufficient ATP solely
from glycolysis, and oxidative phosphorylation is necessary to convert NAD(P)H into ATP. Under anaerobic and light insufficient conditions, *Synechocystis* 6803 may lose the flexibility to balance the ATP requirement for biomass growth.

### 3.3.5 Cyanobacterial fluxome plasticity and rigidity.

Based on our data and previous studies, we compared *Synechocystis* 6803 fluxomes under various conditions (Figure 3.6): dark chemoheterotrophic, photoheterotrophic, photomixotrophic, and photoautotrophic conditions (You et al. 2014; You et al. 2015; Nakajima et al. 2014; Young et al. 2011; Xiong et al. 2015b). The fluxes are shown as the absolute values in separate 3D bar plots and reveal cyanobacterial metabolic responses to environmental changes. In general, the Calvin cycle, the OPPP, and the EMPP showed flexibility in altering their fluxes to adjust to different environments. With sufficient light and inorganic carbon sources, the Calvin cycle showed the highest activity in the metabolic network. Under heterotrophic conditions, however, *Synechocystis* 6803 does not show a fully functional Calvin cycle due to the inactivation of the photosystems, and the OPPP becomes an active mode to supply reducing power. In contrast to the plasticity of the aforementioned pathways, we observed a rigid cyanobacterial TCA cycle, whose fluxes remain low under all the growth conditions, although the flux predictions by the FBA model are higher. This phenomenon suggests that cyanobacterial TCA cycle does not naturally function as an energy source. In an attempt to increase TCA cycle fluxes, *Synechocystis* 6803 was grown with exogenous pyruvate and other TCA cycle intermediates (Figure 3.1, light ~50 µE/m²/s). Only oxaloacetate showed moderate enhancement for growth. This may be due to low activities of the TCA cycle enzymes or the organism failing to uptake these intermediates at a sufficient rate.
Figure 3.6. *Synechocystis* 6803 fluxomes under different growth conditions.

Key reactions in glycolysis, the Calvin cycle, the TCA cycle, and the OPPP were compared: under dark condition (D), photoheterotrophic condition (H; DCMU-induced or atrazine-induced), photomixotrophic condition (M), and photoautotrophic condition (A). Numbers after condition symbol designate independent data sets. References: D, this study; H1, (You et al. 2015); H2, (Nakajima et al. 2014); M1, (You et al. 2014); M2, (Nakajima et al. 2014); A1, (Young et al. 2011b); and A2, (Xiong et al. 2015b). The fluxes are shown in their absolute values in mmol/g DW/h.

3.3.6 Guidelines for cyanobacterial cell factories.

*Synechocystis* 6803 is an organism of interest for use as a biorefinery chassis, and it is important to consider its native metabolic fluxes and intermediate pool sizes when engineering
biosynthesis pathways. Based on our analysis of cyanobacterial metabolism, bioengineering efforts should focus on using sugar phosphate intermediates within the Calvin cycle, OPP, and lower glycolysis pathways. A previous review paper compared chemical titers when different pathways in cyanobacteria were engineered, and the results correspond to our predictions (Angermayr et al. 2015). The results showed high 5.5 g/L titers of ethanol, 2.69 g/L sucrose, 1.14g/L D-lactate, and 1.1g/L D-mannitol, whose precursors are either from the Calvin cycle or lower glycolysis. A very recent FBA model also found that ethanol, isobutanol, 3-methyl-1-butanol, 2-methyl-1-butanol, and propanol were the ideal products to produce in Synechocystis 6803 (Mohammadi et al. 2016). As $^{13}$C-MFA and metabolite pool size analysis found that flux through the TCA cycle was kept at a bare minimum under all growth conditions, these intermediates should be avoided for bioengineering efforts in Synechocystis 6803. The small flux through the TCA cycle leads to the low mg/L titers produced from acetyl-CoA and TCA cycle derivatives, such as free fatty acids or alk(a/e)nes (Yu et al. 2013b). In summary, not all organisms are well suited for the biosynthesis of any products, and development of diverse microbial chassis and use their metabolic strengths for targeted product synthesis will be a promising direction for future microbial cell factories.

3.4 Conclusions

This study improves the current knowledge regarding cyanobacterial fluxomic topology, particularly under dark heterotrophic conditions. Cyanobacteria cannot grow with glucose under anaerobic conditions in the absence of sufficient light. Excessive reducing equivalents need to be removed, and oxygen is required to perform the oxidative phosphorylation. Although previous studies indicate that oxidative phosphorylation does not contribute extensively to support cyanobacterial growth (Montagud et al. 2010), this work implies that oxidative phosphorylation is
essential for cell metabolism in the absence of sufficient light. Cyanobacteria employ the OPP pathway and oxidative phosphorylation as the key pathways for production of energy and reducing equivalents synthesis during dark and heterotrophic conditions. The TCA cycle, on the other hand, provides the building blocks for growth, but maintains a low flux rate. This carbon metabolism is not optimal for biomass accumulation but may shed light on the differences of pathway evolution in cyanobacteria and heterotrophs. The discrepancy between genome scale model predictions and our experimental results demonstrates that many hidden enzymatic constraints or reaction thermodynamics may force microbes to select unpredictable metabolic topologies.

Acknowledgements

We thank Dr. Himadri Pakrasi for useful advice. This work was supported by DOE-BER (DESC0012722 to Y.J.T.), NSF (DBI 1356669 to Y.J.T. and MCB-1331194 to T.S.M.), NSF Graduate Research Fellowship (DGE-1143954 to C.M.I and D.M.D), DOE Office of Science Graduate Fellowship (SCGF2015 to WH), and the Joint BioEnergy Institute, which is supported by the US Department of Energy, Office of Science, Office of Biological and Environmental Research, through contract DE-AC02-05CH11231 between Lawrence Berkeley National Laboratory and the US Department of Energy.
Supporting Information for Chapter 3
Contents: Methods and Experimental details

Two tables (3-S1 – 3-S2)
Three figures (3-S1– 3-S3)

Supplementary Methods

Part S3.1 Metabolite pool size determination.

Metabolite concentrations in *Synechocystis* 6803 under three different light conditions (dark condition, high light ~50 µE/m2/s, and low light ~10 µE/m2/s) and two carbon conditions (autotrophic growth with 4 g/L NaHCO₃, or growth with 1 g/L glucose and 4 g/L NaHCO₃) were analyzed based on an isotope ratio-based approach (Mashego et al. 2004). In brief, fully labeled E. coli metabolites, as internal 13C-standards, were prepared by growing *E. coli* MG1655 in M9 medium supplemented with 20 g/L U-13C6 glucose. *Synechocystis* 6803 cultures (15 mL, n=2) during the exponential phase (OD₇₃₀ = ~0.4) were mixed with the same amount of biomass of fully labeled and exponentially growing *E. coli* in 50mL Falcon tubes. Then the mixed cultures were quenched immediately by placing the falcon tubes in liquid N₂. To avoid freezing the cultures, the solution was rapidly swirled using an electronic thermometer. Within 10 seconds, the sample temperature was quickly dropped to <4 °C (Fu et al., 2015). After the lipid N₂ bath, biomass samples were centrifuged at 4°C (8000 g) and the pellets were extracted using a chloroform-methanol method (Ma et al., 2014). The culture-chloroform-methanol solution was centrifuged at 13,000 g for 5 min at 4 °C to remove cell debris. The supernatant was filtered using Amicon Ultra centrifuge filter tubes (with a 3,000 Da molecular-weight/cut-off, regenerated cellulose membrane; EMD Millipore, Billerica, MA). Samples were then lyophilized and reconstituted in 0.1 mL of acetonitrile-water (6:4, v/v). The aqueous phase from the extraction was transferred into GC vials
and lyophilized. Finally, dried samples were reconstituted in water for analysis using hydrophilic interaction liquid chromatography (HILIC) coupled to electrospray time-of-flight MS. In brief, the separation of metabolites was conducted on a SeQuant Zic-pHILIC column (150-mm length, 2.1-mm interior diameter, and 5-μm particle size; EMD Millipore, Billerica, MA) using an Agilent Technologies 1200 Series HPLC system. A sample injection volume of 1.5 μL was used throughout the analysis. The sample tray and column compartment were set to 4 and 40°C, respectively. The mobile phase was composed of 15 mM ammonium carbonate (Sigma-Aldrich, MO) in water (solvent A) and 15 mM ammonium carbonate in 75% acetonitrile and 25% water (solvent B). A flow rate of 0.2 ml/min was used, unless stated otherwise. Metabolites were separated via gradient elution under the following conditions: 100% B (0 min), 82% B (4.4 min), 72% B (7.7 min), 60% B (9.7 min), 100% B (10.2 min), 100% B (12.5 min), 100% B (13 min, 0.4 mL/min), and 100% B (18.5 min, 0.4 ml/min). The HPLC system was coupled to an Agilent Technologies 6210 series time-of-flight mass spectrometer (for LC-TOF MS) via a MassHunter workstation (Agilent Technologies, CA, USA). A split ratio of 1:4 was used throughout the analysis. Drying and nebulizing gases were set to 10 L/min and 25 lb/in2, respectively, and a drying-gas temperature of 300°C was used throughout the analysis. ESI was conducted in the negative ion mode and a capillary voltage of 3,500 V was utilized. The fragmentor, skimmer, and OCT RT Vpp voltages were set at 100, 50, and 170 V, respectively. The acquisition range was from 70-1000 m/z, and the acquisition rate was 0.86 spectra/s. The MS-TOF was tuned with ESI-L Low concentration tuning mix in the range 50-1700 m/z. Metabolite mass isotopomer distribution data was corrected for background noise. The isotopic ratio of each metabolite (13C-labeled vs. unlabeled) was normalized by biomass to obtain the relative cyanobacterial metabolite pool sizes by benchmarking them against *E. coli*. 
Part S3.2 Metabolic flux analysis.

$^{13}$C-MFA was performed based on $^{13}$C labelled proteinogenic amino acids from *Synechocystis* 6803 grown on $[1^{13}C_1]$ glucose and NaH$^{13}$CO$_3$ under light-pulsed and dark conditions. The labeling distribution of key amino acids did not change during the exponential growth phase, which indicated the attainment of a pseudo-steady state. The metabolic network included the EMP pathway, the ED pathway, the Calvin cycle, the TCA cycle, the glyoxylate shunt, photorespiration pathways, and biomass synthesis. $^{13}$C-MFA profiled relative metabolic fluxes through the central metabolism by normalizing the glucose uptake rate as 100. The relative fluxes were solved by minimizing a quadratic error function that calculated the differences between predicted and measured isotopomer patterns. Reaction reversibility was characterized by the exchange coefficient $\text{exch}$ and the net flux $v_{\text{net}}$ (Dauner et al. 2001; You et al. 2014). $^{13}$C-MFA employed the EMU method for isotopomer data simulations (Antoniewicz et al. 2007b). The MATLAB optimization solver ‘fmincon’ was used to minimize the quadratic error function. To avoid local minima, at least 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 (Zhao and Shimizu 2003), in which the measured isotopomer data (M0, M1, M2, etc.) was perturbed with normally distributed noise within measurement errors (±0.01) and $^{13}$C-MFA was restarted for flux calculations. Based on 1000 simulations generated from perturbed isotopomer data, a confidence interval for each flux can be obtained.

Metabolic pathways are complex and the cells might devise different operational modes for their fluxome. Stoichiometry-based flux balance analysis (FBA), constrained by reaction thermodynamics and inflow/outflow flux measurements, is a useful tool to analyze the influence
of intricate structure of metabolic pathways on cyanobacterial physiologies (Fu 2009; Knoop et al. 2010; Montagud et al. 2010; Montagud et al. 2011; Shastri and Morgan 2005). In this study, we employed the genome-scale FBA model iJN678 (Nogales et al. 2012), which contains more details about biotechnological endeavors, light-limited heterotrophic and photosynthetic metabolism. Experimentally determined glucose uptake rate was used as the input. Maximizing biomass accumulation was employed as the objective for flux determination. The MATLAB optimization solver ‘linprog’ was used to obtain the optimal solution.
Figure 3-S1. Comparison of mass isotopomer distributions (MIDs) between experimental measurements and computational simulations. The data shown in the above figure are [M-57]+, [M-85]+ or [M-159]+, and f302 mass isotopomer distributions for eight amino acids. The fit between predicted and observed isotopomer distributions is reasonably good, with a sum of squared residuals of 0.015.
Figure 3-S2: FBA prediction of Synechocystis 6803 specific growth rate. CO₂ uptake rate upper bound as 3.7 mmol/g DW/h; light intensity ranging from 0~80 mmol/g DW/h; glucose uptake rate ranging from 0~2 mmol/g DW/h. Comparison of the FBA prediction of Synechocystis 6803’s specific growth rate as it varies with glucose uptake rate and photon influx under normal mixotrophic condition (Figure 3-S2A), ‘ATP maintenance’ constrained mixotrophic condition when a 6.9mmol/g/h of NGAM (non-growth associated maintenance) energy is considered (Figure 3-S2B), and ‘pyruvate dehydrogenase’ constrained mixotrophic condition (Figure 3-S2C). See supplementary methods for further information.
Figure 3-S3. Biomass accumulation and glucose consumption in *Syneceocystis* 6803 cultures under dark and heterotrophic conditions. The biomass accumulation (shown as the dried cell weight) and glucose consumption were measured in serum bottles under dark and heterotrophic conditions. Error bars in the figure represent the standard deviations of three biological replicates. Panel A shows that the growth rate $\mu$ is 0.0229 h$^{-1}$. Panel B shows that the biomass yield $Y_{\text{biomass}}/\text{glucose}$ is 0.311 g biomass/g glucose. Thus, the absolute glucose uptake rate is $v_{\text{glucose}} = \mu / Y_{\text{biomass}}/\text{glucose} = 0.41$ mmol glucose/g biomass/h. Panel C shows that glucose consumption curve and growth curve. represents the optimal density of cell cultures measured at 730 nm. Represents the glucose consumption in the cultures. Error bars in the figures are standard deviations of biological triplicates.
Table 3-S1. Simulated and measured mass isotopomer distributions (MID) of proteinogenic amino acids from *Syneceocystis 6803* cultures. Cultures were grown on $^{13}$C glucose and NaH$^{13}$CO$_3$ under dark heterotrophic conditions. Three biological replicates (in exponential growth phase), with one technical replicate each, were used for isotopomer analysis.

<table>
<thead>
<tr>
<th>Amino Acids</th>
<th>MID</th>
<th>Measured data</th>
<th>Standard deviation</th>
<th>Simulated data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>[M-57]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALA</td>
<td>M0</td>
<td>0.8004</td>
<td>0.0027</td>
<td>0.8125</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>0.1419</td>
<td>0.0007</td>
<td>0.1436</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.0352</td>
<td>0.0018</td>
<td>0.0405</td>
</tr>
<tr>
<td></td>
<td>M3</td>
<td>0.0225</td>
<td>0.0002</td>
<td>0.0034</td>
</tr>
<tr>
<td></td>
<td>[M-159]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>0.8975</td>
<td>0.0023</td>
<td>0.9146</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>0.0789</td>
<td>0.0020</td>
<td>0.0757</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.0236</td>
<td>0.0006</td>
<td>0.0096</td>
</tr>
<tr>
<td>GLY</td>
<td>[M-57]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>0.7745</td>
<td>0.0054</td>
<td>0.7828</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>0.1889</td>
<td>0.0052</td>
<td>0.1981</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.0367</td>
<td>0.0003</td>
<td>0.0191</td>
</tr>
<tr>
<td></td>
<td>[M-159]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>0.9514</td>
<td>0.0014</td>
<td>0.9491</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>0.0486</td>
<td>0.0014</td>
<td>0.0509</td>
</tr>
<tr>
<td>VAL</td>
<td>[M-57]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>0.7403</td>
<td>0.0005</td>
<td>0.7432</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>0.1783</td>
<td>0.0003</td>
<td>0.1929</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.0391</td>
<td>0.0002</td>
<td>0.0557</td>
</tr>
<tr>
<td></td>
<td>M3</td>
<td>0.0192</td>
<td>0.0011</td>
<td>0.0075</td>
</tr>
<tr>
<td></td>
<td>M4</td>
<td>0.0141</td>
<td>0.0004</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.0090</td>
<td>0.0009</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>[M-159]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>0.8107</td>
<td>0.0013</td>
<td>0.8366</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>0.1203</td>
<td>0.0012</td>
<td>0.1386</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.0309</td>
<td>0.0004</td>
<td>0.0233</td>
</tr>
<tr>
<td></td>
<td>M3</td>
<td>0.0281</td>
<td>0.0016</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>M4</td>
<td>0.0100</td>
<td>0.0007</td>
<td>0.0001</td>
</tr>
<tr>
<td>LEU</td>
<td>[M-15]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>0.7677</td>
<td>0.0026</td>
<td>0.7652</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>0.1636</td>
<td>0.0054</td>
<td>0.1901</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.0333</td>
<td>0.0045</td>
<td>0.0399</td>
</tr>
<tr>
<td></td>
<td>M3</td>
<td>0.0165</td>
<td>0.0011</td>
<td>0.0044</td>
</tr>
<tr>
<td></td>
<td>M4</td>
<td>0.0094</td>
<td>0.0017</td>
<td>0.0004</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.0065</td>
<td>0.0046</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>M6</td>
<td>0.0031</td>
<td>0.0013</td>
<td>0.0000</td>
</tr>
<tr>
<td></td>
<td>[M-159]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>0.7732</td>
<td>0.0009</td>
<td>0.7940</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>0.1511</td>
<td>0.0012</td>
<td>0.1741</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.0405</td>
<td>0.0007</td>
<td>0.0292</td>
</tr>
<tr>
<td></td>
<td>M3</td>
<td>0.0172</td>
<td>0.0008</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>M4</td>
<td>0.0123</td>
<td>0.0005</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>M5</td>
<td>0.0057</td>
<td>0.0004</td>
<td>0.0000</td>
</tr>
<tr>
<td>LEU</td>
<td>[M-57]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>0.8297</td>
<td>0.0082</td>
<td>0.8353</td>
</tr>
<tr>
<td></td>
<td>M1</td>
<td>0.1172</td>
<td>0.0030</td>
<td>0.1245</td>
</tr>
<tr>
<td></td>
<td>M2</td>
<td>0.0340</td>
<td>0.0012</td>
<td>0.0374</td>
</tr>
<tr>
<td>-----</td>
<td>-------</td>
<td>--------</td>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>M3</td>
<td>0.0191</td>
<td>0.0064</td>
<td>0.0028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[M-159]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>0.8926</td>
<td>0.0055</td>
<td>0.9125</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.0843</td>
<td>0.0080</td>
<td>0.0789</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.0231</td>
<td>0.0026</td>
<td>0.0086</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[M-57]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>0.6520</td>
<td>0.0011</td>
<td>0.6626</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.2255</td>
<td>0.0030</td>
<td>0.2539</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.0519</td>
<td>0.0007</td>
<td>0.0683</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.0148</td>
<td>0.0009</td>
<td>0.0135</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>0.0091</td>
<td>0.0002</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[M-159]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>0.7027</td>
<td>0.0048</td>
<td>0.7117</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.2068</td>
<td>0.0029</td>
<td>0.2427</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.0421</td>
<td>0.0014</td>
<td>0.0405</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.0150</td>
<td>0.0004</td>
<td>0.0047</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>0.0085</td>
<td>0.0003</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F302</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>0.8787</td>
<td>0.0034</td>
<td>0.8733</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.0968</td>
<td>0.0023</td>
<td>0.1113</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.0245</td>
<td>0.0020</td>
<td>0.0155</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[M-57]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>0.0981</td>
<td>0.0093</td>
<td>0.1025</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.7442</td>
<td>0.0083</td>
<td>0.7619</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.0992</td>
<td>0.0013</td>
<td>0.0982</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.0346</td>
<td>0.0014</td>
<td>0.0347</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>0.0238</td>
<td>0.0012</td>
<td>0.0027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[M-159]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>0.1036</td>
<td>0.0083</td>
<td>0.1128</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.7886</td>
<td>0.0084</td>
<td>0.8111</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.0763</td>
<td>0.0015</td>
<td>0.0676</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.0316</td>
<td>0.0019</td>
<td>0.0084</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F302</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>0.8573</td>
<td>0.0049</td>
<td>0.8705</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.1103</td>
<td>0.0032</td>
<td>0.1139</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.0324</td>
<td>0.0028</td>
<td>0.0156</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[M-57]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>0.0850</td>
<td>0.0052</td>
<td>0.1032</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.7579</td>
<td>0.0023</td>
<td>0.7504</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.1173</td>
<td>0.0006</td>
<td>0.1244</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.0250</td>
<td>0.0018</td>
<td>0.0206</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>0.0094</td>
<td>0.0003</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td>M5</td>
<td>0.0053</td>
<td>0.0008</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td></td>
<td>[M-159]^*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M0</td>
<td>0.8331</td>
<td>0.0027</td>
<td>0.8366</td>
<td></td>
</tr>
<tr>
<td>M1</td>
<td>0.1255</td>
<td>0.0014</td>
<td>0.1385</td>
<td></td>
</tr>
<tr>
<td>M2</td>
<td>0.0262</td>
<td>0.0004</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>M3</td>
<td>0.0102</td>
<td>0.0007</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td>M4</td>
<td>0.0051</td>
<td>0.0005</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>Reactions</td>
<td>Best fit</td>
<td>95% confidence interval lower bounds</td>
<td>95% confidence interval upper bounds</td>
<td>Standard deviations</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------</td>
<td>-------------------------------------</td>
<td>-------------------------------------</td>
<td>--------------------</td>
</tr>
<tr>
<td>Glucose == G6P</td>
<td>100.0</td>
<td>100.0</td>
<td>100.0</td>
<td>0.0</td>
</tr>
<tr>
<td>G6P == F6P</td>
<td>-282.3</td>
<td>-308.1</td>
<td>-241.5</td>
<td>17.0</td>
</tr>
<tr>
<td>F6P == FBP</td>
<td>-1.2</td>
<td>-8.8</td>
<td>3.5</td>
<td>3.1</td>
</tr>
<tr>
<td>FBP == DHAP + GAP</td>
<td>-1.2</td>
<td>-8.8</td>
<td>3.5</td>
<td>3.1</td>
</tr>
<tr>
<td>DHAP == GAP</td>
<td>-46.6</td>
<td>-53.8</td>
<td>-36.9</td>
<td>4.3</td>
</tr>
<tr>
<td>GAP == 3PG</td>
<td>25.8</td>
<td>10.8</td>
<td>31.4</td>
<td>5.3</td>
</tr>
<tr>
<td>3PG == PEP</td>
<td>54.4</td>
<td>47.8</td>
<td>64.6</td>
<td>4.3</td>
</tr>
<tr>
<td>PEP == PYR</td>
<td>36.9</td>
<td>20.8</td>
<td>41.6</td>
<td>5.3</td>
</tr>
<tr>
<td>PYR == AceCoA + CO₂</td>
<td>30.4</td>
<td>26.8</td>
<td>36.2</td>
<td>2.4</td>
</tr>
<tr>
<td>Acetate == AceCoA</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>AceCoA + OAA == CIT</td>
<td>3.9</td>
<td>3.5</td>
<td>5.0</td>
<td>0.4</td>
</tr>
<tr>
<td>CIT == ICT</td>
<td>3.9</td>
<td>3.5</td>
<td>5.0</td>
<td>0.4</td>
</tr>
<tr>
<td>ICIT == AKG + CO₂</td>
<td>3.9</td>
<td>3.5</td>
<td>5.0</td>
<td>0.4</td>
</tr>
<tr>
<td>AKG == SUC + CO₂</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>SUC == SucCoA</td>
<td>1.0</td>
<td>0.9</td>
<td>1.2</td>
<td>0.1</td>
</tr>
<tr>
<td>SUC == FUM</td>
<td>0.0</td>
<td>0.0</td>
<td>0.8</td>
<td>0.2</td>
</tr>
<tr>
<td>FUM == MAL</td>
<td>2.4</td>
<td>2.1</td>
<td>3.2</td>
<td>0.3</td>
</tr>
<tr>
<td>MAL == OAA</td>
<td>-2.0</td>
<td>-17.3</td>
<td>-0.7</td>
<td>4.2</td>
</tr>
<tr>
<td>MAL == PYR + CO₂</td>
<td>4.4</td>
<td>3.0</td>
<td>20.1</td>
<td>4.3</td>
</tr>
<tr>
<td>PEP + CO₂ == OAA</td>
<td>15.2</td>
<td>13.5</td>
<td>31.5</td>
<td>4.6</td>
</tr>
<tr>
<td>G6P == 6PG</td>
<td>380.5</td>
<td>339.0</td>
<td>406.3</td>
<td>17.2</td>
</tr>
<tr>
<td>6PG == Ru5P + CO₂</td>
<td>380.5</td>
<td>339.0</td>
<td>406.1</td>
<td>17.1</td>
</tr>
<tr>
<td>Ru5P == RuBP</td>
<td>20.8</td>
<td>18.3</td>
<td>30.8</td>
<td>3.2</td>
</tr>
<tr>
<td>RuBP + CO₂ == 2*3PG</td>
<td>5.4</td>
<td>4.5</td>
<td>7.7</td>
<td>0.8</td>
</tr>
<tr>
<td>X5P == Ru5P</td>
<td>-236.9</td>
<td>-255.9</td>
<td>-206.3</td>
<td>12.6</td>
</tr>
<tr>
<td>R5P == Ru5P</td>
<td>-122.8</td>
<td>-131.7</td>
<td>-108.1</td>
<td>6.0</td>
</tr>
<tr>
<td>GAP + S7P == X5P + R5P</td>
<td>-119.0</td>
<td>-128.4</td>
<td>-103.8</td>
<td>6.3</td>
</tr>
<tr>
<td>E4P + F6P == GAP + S7P</td>
<td>-163.2</td>
<td>-180.2</td>
<td>-138.2</td>
<td>10.7</td>
</tr>
<tr>
<td>E4P + DHAP == S7P</td>
<td>44.2</td>
<td>30.6</td>
<td>52.6</td>
<td>5.6</td>
</tr>
<tr>
<td>GAP + F6P == X5P + E4P</td>
<td>-117.9</td>
<td>-127.4</td>
<td>-102.5</td>
<td>6.4</td>
</tr>
<tr>
<td>6PG == GAP + PYR</td>
<td>0.0</td>
<td>0.0</td>
<td>0.9</td>
<td>0.2</td>
</tr>
<tr>
<td>ICIT == GLX + SUC</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>GLX + AceCoA == MAL</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>RuBP + O₂ == GLX + 3PG</td>
<td>15.4</td>
<td>12.6</td>
<td>24.1</td>
<td>2.9</td>
</tr>
<tr>
<td>GLX == GLY</td>
<td>0.1</td>
<td>0.0</td>
<td>0.5</td>
<td>0.1</td>
</tr>
<tr>
<td>Reaction</td>
<td>Coefficient</td>
<td>Coefficient</td>
<td>Coefficient</td>
<td>Coefficient</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>2*GLX == 3PG + CO₂</td>
<td>7.6</td>
<td>6.2</td>
<td>11.9</td>
<td>1.5</td>
</tr>
<tr>
<td>AKG == GLU</td>
<td>28.0</td>
<td>24.7</td>
<td>32.9</td>
<td>2.1</td>
</tr>
<tr>
<td>GLU == GLN</td>
<td>2.0</td>
<td>1.8</td>
<td>2.4</td>
<td>0.2</td>
</tr>
<tr>
<td>GLU == PRO</td>
<td>1.0</td>
<td>0.9</td>
<td>1.2</td>
<td>0.1</td>
</tr>
<tr>
<td>GLU + CO₂ + GLN + ASP == ARG + AKG + FUM</td>
<td>0.7</td>
<td>0.6</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>OAA + GLU == ASP + AKG</td>
<td>9.2</td>
<td>8.2</td>
<td>10.9</td>
<td>0.7</td>
</tr>
<tr>
<td>ASP == ASN</td>
<td>0.7</td>
<td>0.6</td>
<td>0.9</td>
<td>0.1</td>
</tr>
<tr>
<td>PYR + GLU == ALA + AKG</td>
<td>2.6</td>
<td>2.2</td>
<td>3.0</td>
<td>0.2</td>
</tr>
<tr>
<td>3PG + GLU == SER + AKG</td>
<td>5.2</td>
<td>4.6</td>
<td>6.1</td>
<td>0.4</td>
</tr>
<tr>
<td>SER == GLY + MTHF</td>
<td>3.2</td>
<td>2.8</td>
<td>3.8</td>
<td>0.3</td>
</tr>
<tr>
<td>GLY == MTHF + CO₂</td>
<td>-0.2</td>
<td>-0.3</td>
<td>0.0</td>
<td>0.1</td>
</tr>
<tr>
<td>SER == CYS</td>
<td>0.5</td>
<td>0.5</td>
<td>0.6</td>
<td>0.0</td>
</tr>
<tr>
<td>ASP == THR</td>
<td>2.2</td>
<td>2.0</td>
<td>2.6</td>
<td>0.2</td>
</tr>
<tr>
<td>ASP + PYR + GLU + SucCoA == LYS + CO₂ + AKG + SUC</td>
<td>0.7</td>
<td>0.6</td>
<td>0.8</td>
<td>0.1</td>
</tr>
<tr>
<td>ASP + MTHF + CYS + SucCoA == MET + PYR + SUC</td>
<td>0.3</td>
<td>0.3</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>2*PYR + GLU == VAL + AKG + CO₂</td>
<td>1.3</td>
<td>1.2</td>
<td>1.6</td>
<td>0.1</td>
</tr>
<tr>
<td>AceCoA + 2<em>PYR + GLU == LEU + AKG + 2</em>CO₂</td>
<td>2.0</td>
<td>1.8</td>
<td>2.4</td>
<td>0.2</td>
</tr>
<tr>
<td>THR + PYR + GLU == ILE + AKG + CO₂</td>
<td>1.1</td>
<td>1.0</td>
<td>1.3</td>
<td>0.1</td>
</tr>
<tr>
<td>2*PEP + E4P + GLU == PHE + AKG + CO₂</td>
<td>0.6</td>
<td>0.5</td>
<td>0.7</td>
<td>0.0</td>
</tr>
<tr>
<td>2*PEP + E4P + GLU == TYR + AKG + CO₂</td>
<td>0.4</td>
<td>0.3</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>SER + R5P + 2*PEP + E4P + GLN == TRP + GAP + PYR + GLU + CO₂</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.0</td>
</tr>
<tr>
<td>R5P + MTHF + GLN + ASP == HIS + AKG + FUM</td>
<td>0.3</td>
<td>0.2</td>
<td>0.3</td>
<td>0.0</td>
</tr>
<tr>
<td>0.284<em>G6P+0.495</em>R5P+0.046<em>GAP+0.173</em>DHAP+0.046<em>PYR+3.707</em>AceCoA+0.366<em>MTHF+0.530</em>GLY+0.158<em>PRO+0.387</em>ALA+0.201<em>VAL+0.310</em>LEU+0.171<em>ILE+0.048</em>MET+0.030<em>CYS+0.087</em>PHE+0.058<em>TYR+0.027</em>TRP+0.043<em>HIS+0.105</em>LYS+0.104<em>ARG+0.136</em>GLN+0.110<em>ASN+0.200</em>GLU+0.657<em>ASP+0.199</em>SER+0.165<em>THR==Biomass+0.21</em>FUM</td>
<td>6.6*</td>
<td>5.8</td>
<td>7.8</td>
<td>0.5</td>
</tr>
</tbody>
</table>

Exchange coefficient of Reaction: G6P == F6P | 1.00 | 1.0 | 1.0 | 0.0 |
Exchange coefficient of Reaction: F6P == FBP | 0.00 | 0.0 | 1.0 | 0.3 |
Exchange coefficient of Reaction: FBP == DHAP + GAP | 0.97 | 0.0 | 1.0 | 0.3 |
Exchange coefficient of Reaction: DHAP == GAP | 0.83 | 0.0 | 1.0 | 0.3 |
Exchange coefficient of Reaction: GAP == 3PG | 0.00 | 0.0 | 0.4 | 0.1 |
Exchange coefficient of Reaction: 3PG == PEP | 0.81 | 0.0 | 1.0 | 0.2 |
Exchange coefficient of Reaction: CIT == ICIT | 0.47 | 0.4 | 0.6 | 0.0 |
<table>
<thead>
<tr>
<th>Exchange coefficient of Reaction:</th>
<th>( \text{ICIT} \rightarrow \text{AKG + CO}_2 )</th>
<th>0.61</th>
<th>0.3</th>
<th>0.7</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{SUC} \rightarrow \text{SucCoA} )</td>
<td>0.37</td>
<td>0.4</td>
<td>0.6</td>
<td>0.1</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{SUC} \rightarrow \text{FUM} )</td>
<td>0.26</td>
<td>0.3</td>
<td>0.7</td>
<td>0.1</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{FUM} \rightarrow \text{MAL} )</td>
<td>0.91</td>
<td>0.0</td>
<td>0.9</td>
<td>0.2</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{MAL} \rightarrow \text{OAA} )</td>
<td>0.00</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{X5P} \rightarrow \text{Ru5P} )</td>
<td>1.00</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{R5P} \rightarrow \text{Ru5P} )</td>
<td>0.92</td>
<td>0.0</td>
<td>1.0</td>
<td>0.3</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{GAP + S7P} \rightarrow \text{X5P + R5P} )</td>
<td>1.00</td>
<td>0.6</td>
<td>1.0</td>
<td>0.1</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{E4P + F6P} \rightarrow \text{GAP + S7P} )</td>
<td>1.00</td>
<td>0.3</td>
<td>1.0</td>
<td>0.2</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{GAP + F6P} \rightarrow \text{X5P + E4P} )</td>
<td>0.00</td>
<td>0.0</td>
<td>1.0</td>
<td>0.3</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{SER} \rightarrow \text{GLY + MTHF} )</td>
<td>0.01</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Exchange coefficient of Reaction:</td>
<td>( \text{GLY} \rightarrow \text{MTHF + CO}_2 )</td>
<td>0.00</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

* We constrained the biomass formation rate by minimizing the differences between the predicted \( V_{\text{biomass}} \) and the experimental value (5.59) in the objective function that is expressed as:

\[
\sum \left( \frac{MID_{\text{exp}} - MID_{\text{sim}}}{\delta} \right)^2 + \left( V_{\text{biomass}_{\text{exp}}} - V_{\text{biomass}_{\text{sim}}} \right)^2,
\]

where the subscripts ‘exp’ and ‘sim’ represent the experimental and simulation results, respectively, and \( \delta \) is 0.01.
Chapter 4. Enhancing fatty acid production

in *Escherichia coli* by *Vitreoscilla*

Hemoglobin overexpression

Results of this chapter have been published in Biotechnology and bioengineering 2017, 114, 463-467
Abstract

Our recent \textsuperscript{13}C-metabolic flux analysis (\textsuperscript{13}C-MFA) study indicates that energy metabolism becomes a rate-limiting factor for fatty acid overproduction in \textit{E. coli} strains (after “Push-Pull-Block” based genetic modifications). To resolve this bottleneck, \textit{Vitreoscilla} Hemoglobin (VHb, a membrane protein facilitating O\textsubscript{2} transport) was introduced into a fatty-acid-producing strain to promote oxygen supply and energy metabolism. The resulting strain, FAV50, achieved 70% percent higher fatty acid titer than the parent strain in shake tube cultures. In high cell-density bioreactor fermentations, FAV50 achieved free fatty acids at a titer of 7.02 g/L (51% of the theoretical yield). In addition to “Push-Pull-Block-Power” strategies, our experiments and flux balance analysis also revealed the fatty acid over-producing strain is sensitive to metabolic burden and oxygen influx, and thus a careful evaluation of the cost-benefit tradeoff with the guidance of fluxome analysis will be fundamental for the rational design of synthetic biology strains.

4.1 Introduction

Synthetic biology (SynBio) can produce a broad scope of products, from biofuels to pharmaceutical chemicals, through gene knockouts and heterologous enzyme overexpression. However, extensive pathway modifications may impose considerable burdens on cell metabolism (Glick 1995; Wu \textit{et al.} 2016). With increased steps of genetic manipulation, the metabolic burden from new genetic parts may limit cell production performance and cause metabolic shifts. For example, biosynthesis of fatty acids or related compounds has been a hot field during recent years (Jones \textit{et al.} 2015). SynBio follows Push-Pull-Block strategies to direct carbon flux towards free fatty acids, including introduction of heterogeneous enzymes and knocking out degradation
pathways (Lu et al. 2008; Steen et al. 2010), reversal of a degradation pathway (Dellomonaco et al. 2011), engineering regulators to boost pathway activities (Zhang et al. 2012b), and creating sensor regulator systems to control biosynthesis fluxes (Xu et al. 2014; Zhang et al. 2012a).

Nevertheless, the combination of SynBio approaches still cannot achieve production metrics to meet industrial manufacturing needs (Van Dien 2013). To address this problem, we performed $^{13}$C-MFA of an engineered strain and revealed high ATP consumption for cell maintenance during fatty acid over-production (He et al. 2014). This intracellular energy crisis may become worse when the $O_2$ supply is insufficient during large-scale fermentations. Hence, we propose to promote cell energy metabolism by introducing *Vitreoscilla* hemoglobin (VHb) into the hosts. VHb is a type of soluble protein that can bind $O_2$ at low concentrations and improve the efficiency of bacterial aerobic respiration (Dikshit and Webster 1988; Khosla and Bailey 1988a; Khosla and Bailey 1988b). The ability of wild-type VHb (encoded by *vhb*) to facilitate $O_2$ uptake and biomass growth can be further enhanced by introducing point mutations (amino acid substitutions) (Andersson et al. 2000). Hemoglobin has demonstrated its effectiveness in improving SynBio microbial fermentations. For example, DuPont has successfully employed different bacterial hemoglobin genes to increase carotenoid production by microbial host cells (Cheng et al. 2007).

**4.2 Materials and Methods**

**4.2.1 Chemicals and Strains.**

All chemicals were reagent grade and purchased from either Sigma-Aldrich (St. Louis, MO, USA) or Fisher Scientific (Pittsburgh, PA, USA) unless otherwise noted. Restriction enzymes, Phusion DNA polymerase, and T4 ligase were from New England Biolabs (Ipswich, MA, USA).
The DNA Purification kit, Gel Recovery kit, and Miniprep kit were from Promega (Madison, WI, USA). The DNA sequences of vhb and its mutant vhb20 and vhb50 were based on a previous report (Andersson et al. 2000). All genes were synthesized by GenScript Inc. (Piscataway, NJ, USA) and cloned into pUC57 vector. E. coli DH10B strain was used for plasmid manipulation. The fadE knockout E. coli DH1 strain and the plasmid pA58c-TR were from Dr. Fuzhong Zhang’s lab.

4.2.2 Plasmid construction.

Primers used in this study were synthesized by Integrated DNA Technologies (Coralville, IA, USA). Plasmids pA5c-tesA-VHB-8fadR, pA5c-tesA-VHB20-8fadR, and pA5c-tesA-VHB50-8fadR were constructed by inserting vhb, vhb20, and vhb50 downstream of tesA in plasmid pA58c-TR, respectively. To create the FAC strain, tesA was cloned under the control of a pLacUV5 promoter, giving pA5c-tesA. To create the FAR1 strain, rfp was cloned downstream of tesA in plasmid pA58c-TR. To create the FAR2 strain, rfp was cloned into the pB5k vector to construct pB5k-RFP plasmid, which was co-transformed with pA5c-tesA (Lee et al. 2011). The DNA sequences of all constructed plasmids were validated by sequencing in the Genome Center at Washington University School of Medicine.

4.2.3 Medium and culture conditions.

A modified M9 minimal medium supplied with 2% glucose and appropriate antibiotics was used in this study (Liu et al. 2015). In fatty acid production experiments, all strains were first inoculated into LB medium with appropriate antibiotics. The overnight culture was inoculated 2% v/v into minimal medium containing appropriate antibiotics for adaptation. The overnight minimal medium culture was used to inoculate a 10 mL (shake tube)/20 mL (shake flask) fresh minimal medium with an initial OD of 0.08. Cells were induced with 1mM of IPTG when OD600 reached
0.6. Cell growth and fatty acids production were monitored at different time points. For semi-batch fermentation, 9 mL of LB culture with the engineered strain was incubated overnight and inoculated into 450 ml M9 minimal medium (2% glucose, 30 mg/L chloramphenicol with a supply of Vitamin B12) in the bioreactor (New Brunswick BioFlo 110 fermentor). The fermentation was initiated with the following settings: The incubation temperature was controlled between 35°C ~ 37 °C; the pH of the culture medium was adjusted around 7.2 by automatic addition of ammonium hydroxide (6 mol/L); the airflow rate was kept at ~1.5 L/min, and the average stirring rate was 500 rpm. When OD_{600} of culture reaches 7, 0.1 mM of IPTG (final concentration) was added. Four hours after induction, a glucose stock solution (400 g/L glucose and 12 g/L MgSO_{4}) was intermittently pulsed into the bioreactor to re-supply glucose. After 48 h of induction, a total of 40.72 g/L of glucose was consumed, and the final cultures were harvested for measurement of free fatty acids.

4.2.4 Fatty acids, glucose, and acetate measurements.

Free fatty acid titer was analyzed following a previous protocol (Liu et al. 2015). In brief, 500 μL of cell culture was acidified using 50 μL of concentrated HCl. Fatty acids were extracted twice with ethyl acetate (EtAc) spiked with C19:0 methyl ester (ME) as an internal standard. Fatty acids were then derivatized to fatty acids methyl esters (FAME) using 100 μL of MeOH:HCl (9:1) and 100 μL of TMS-diazomethane (2 M in hexanes). The mixture was incubated for 10 ~ 15 min at room temperature for the reaction to complete. FAME was analyzed using a gas chromatograph-mass spectrometer (GC-MS, Hewlett-Packard model 7890A, Agilent Technologies). Free fatty acids were quantified based on the standard curve of standard FAME mix. Acetate and D-Glucose measurements followed the protocol of commercial kits (r-biopharm, MO, USA)
4.2.5 FBA Simulation.

The genome-scale model iJO1366 (which includes 1366 genes, 2583 reactions, and 1805 metabolites) was employed to simulate fatty acids production in \textit{E.coli} strain (Monk \textit{et al.} 2013). A simplified flux of fatty acid (C16:0) was added as representative of fatty acids production, and the objective function was set to maximize this flux. Default values for the boundary of all fluxes were adopted, except for the following: The upper and lower boundaries of the fatty acid degradation flux were set to zero because \(\Delta fadE\) was knocked out, and the lower boundary of the glucose uptake flux was set based on experimental value. The sensitivity of fatty acid yields to ATP maintenance loss and oxygen influx was tested by FBA. The COBRA toolbox and LibSBML library were employed for genome-scale model manipulation (Bornstein \textit{et al.} 2008; Schellenberger \textit{et al.} 2011), and the Gurobi 5.5 linear solver (Gurobi Optimization Inc.) was utilized for FBA calculation on MATLAB 2012b.

4.3 Results and Discussion

In this work, we chose an \textit{E. coli} DH1 strain with \textit{fadE} knockout (\textit{fadE} encodes an enzyme in the fatty acid \(\beta\)-oxidation pathway) as the host (DH1\(\Delta fadE\) strain) (Steen \textit{et al.} 2010). The control strain (denoted as FA0, Table 4.1) for free fatty acid production carries a plasmid with \textit{tesA} and \textit{fadR} overexpression, and its central metabolism has been investigated by \(^{13}\text{C}\)-MFA in our previous work (He \textit{et al.} 2014). Based on the control strain, we inserted a wild-type \textit{vhb} gene and its two mutants (\textit{vhb20} and \textit{vhb50}) into the same plasmid to generate three strains, FAV1, FAV20, and FAV50, respectively (Table 4.1). Under M9 minimal medium and shake tube conditions, only FAV50 (expressing VHb50, created by His36\(\rightarrow\)Arg and Gln66\(\rightarrow\)Arg in wild type VHb) showed significantly enhanced cell growth (as shown in Fig. 4.1a).
<table>
<thead>
<tr>
<th>Strains</th>
<th>Plasmids</th>
<th>Replication Origin</th>
<th>Overexpressed Operon</th>
<th>Resistance</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>FAV50 8fadR</td>
<td>pA5c-tesA-VHB50-8fadR</td>
<td>p15A</td>
<td>( P_{lacuv5-tesA-vhb50}, P_{BAD-fadR} )</td>
<td>Cm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>FAV20 8fadR</td>
<td>pA5c-tesA-VHB20-8fadR</td>
<td>p15A</td>
<td>( P_{lacuv5-tesA-vhb50}, P_{BAD-fadR} )</td>
<td>Cm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>FAV1 8fadR</td>
<td>pA5c-tesA-VHB-8fadR</td>
<td>p15A</td>
<td>( P_{lacuv5-tesA-vhb}, P_{BAD-fadR} )</td>
<td>Cm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>FA0</td>
<td>pA5c-tesA-8fadR</td>
<td>p15A</td>
<td>( P_{lacuv5-tesA}, P_{BAD-fadR} )</td>
<td>Cm(^R)</td>
<td>Zhang et al. 2012b</td>
</tr>
<tr>
<td>CTL</td>
<td>pA5c-0</td>
<td>p15A</td>
<td>( P_{lacuv5-none} )</td>
<td>Cm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>FAC</td>
<td>pA5c-tesA</td>
<td>p15A</td>
<td>( P_{lacuv5-tesA} )</td>
<td>Cm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>FAR1</td>
<td>pA5c-tesA-RFP-8fadR</td>
<td>p15A</td>
<td>( P_{lacuv5-tesA-rfp}, P_{BAD-fadR} )</td>
<td>Cm(^R)</td>
<td>This study</td>
</tr>
<tr>
<td>FAR2</td>
<td>pA5c-tesA, pB5k-RFP</td>
<td>p15A, BBR1</td>
<td>( P_{lacuv5-tesA}, P_{lacuv5-rfp} )</td>
<td>Cm(^R), Kan(^R)</td>
<td>This study</td>
</tr>
</tbody>
</table>
Figure 4.1. Cell growth and fatty acid production from shake tube cultures (10 mL culture).

(a) Growth curve, (b) fatty acid production, (c) glucose consumption, and (d) acetate production of FAV50, FA0, and CTL strains. CTL, the control strain carrying an empty plasmid; FA0, the strain carrying a plasmid with tesA and fadR overexpression; FAV50, the strain carrying a plasmid with tesA, fadR and vhb50 overexpression. All strains have a DH1(ΔfadE) background. Square, FAV50; triangle, FA0; circle, CTL.

Subsequently, we examined the influence of VHb50 on fatty acid production (Fig. 4.1b). The strains were cultured in minimal medium in shake tubes with 10 mL culture. The control strain CTL (without tesA gene in the plasmid, Table 4.1) demonstrated a growth rate similar to that of strain FA0 (overexpressing only tesA and fadR, Table 4.1) in the exponential phase. However, CTL culture produced a significant amount of acetate and entered the stationary phase earlier.
Strain FAV50 (with \(vhb50\), \(tesA\) and \(fadR\) gene, Table 4.1) grew slower at the beginning, but accumulated 30% more biomass and 70% more free fatty acids with a similar consumption of glucose, compared with FA0 (Fig. 4.1c). Moreover, \(vhb50\) overexpression significantly reduced acetate secretions (Fig. 4.1d). These observations indicate the promotion of energy metabolism can reduce waste product synthesis and improve product yields.

During industrial fermentation, oxygen supply always becomes insufficient when cell density reaches a high level (Garcia-Ochoa and Gomez 2009). To test the performance of VHb under high cell density, we performed semi-batch fermentations on FAV50 in a 1-Liter bioreactor. The culture reached an OD\(_{600}\) of over 50, and the final titer of free fatty acids reached 7.04 g/L, with a yield of 0.173 g FA/g glucose (~51% of the theoretical yield) after two days of semi-batch fermentation. In contrast, overexpression of \(tesA\) and \(FadR\) in DH1(\(\Delta fadE\)) (without using VHb) produced only 4.8 g/L fatty acids in similar semi-fed batch fermentations (Xiao et al., 2016). The titer and yield of our VHb strain is comparable with recent reports on total fatty acid production via systematic modular optimization (Xu et al. 2013) or via a dynamic sensor regulatory system (Xu et al. 2014). These experimental results agree with the \(^{13}\text{C-MFA}\) prediction that energy metabolism is one of the key factors limiting fatty acid production.

We also performed other sets of experiments with FAV50 under better aeration conditions (in a shake tube with 5 mL culture or in a baffled shake flask), and observed less significant improvement in biomass growth or fatty acid production than control strain FA0. Moreover, in shake tube cultures the fatty acid titers for FAV1 and FAV20 were lower than in the control strain FA0. This observation indicates that the metabolic burden from VHb expression may offset the benefits from increased \(O_2\) transfer, while the function of VHb is significant only under microaerobic conditions (Frey and Kallio 2003). To investigate the impact of metabolic burden caused
by protein overexpression on fatty acid production, we replaced the \textit{vhb50} gene in FAV50 with a gene encoding red fluorescent protein (RFP), thus generating the strain FAR1. Thus, the expression level of VHb is comparable with the RFP level in FAR1. RFP was chosen here for several reasons. First, RFP (25-30 kDa) is relatively small, comparable to VHb. Second, expression of RFP does not directly interfere with cell physiology. Moreover, the expression level of RFP can be readily monitored by fluorescence measurement. Specifically, two strains, FAC (overexpressing the \textit{tesA} gene, Table 4.1) and FAR1 (overexpressing the \textit{tesA} and \textit{rfp} genes, Table 4.1), were cultured in M9 medium. Figure 4.2 shows the final titers of biomass growth and fatty acid production at 72 h post induction. Decreased fatty acid titer for FAR1 indicated that overexpression of even a non-toxic, small-size protein can lead to a significant impact on fatty acid productivity. To further explore the metabolic burden of protein overexpression, we increased the RFP expression level by cloning \textit{rfp} under a higher copy number plasmid, generating the strain FAR2 (Table 4.1), which has a five times higher RFP expression than FAR1 (Fig. 4.2c). A dramatic decrease in fatty acid production and cell growth was observed as expected, which further validated the impact of the metabolic burden from heterogeneous gene overexpression. To sum up, our experiments demonstrated various levels of trade-offs between the metabolic burden caused by genetic modifications and the benefits from engineered components (enhanced oxygen flux in this case).
Figure 4.2. (a) Biomass growth and (b) fatty acid production of FAC, FAR1, and FAR2 at 72 h post induction. c) Relative RFP expression level of FAR1 and FAR2.

FAC, FAR1, and FAR2 are strains expressing tesA, with no, medium and high levels of RFP expression, respectively.

SynBio allows the assembly of multiple genetic components in a recombinant host. However, the capability of cell hosts for handling metabolic burden is still hard to be quantified (Wu et al. 2016). In particular, cell energy metabolism has a limited ability to generate ATP. ATP is consumed not only for biomass growth and product synthesis, but also for maintenance of SynBio components (such as plasmid synthesis and enzyme overexpression). The shortage of ATP may lead to undesirable metabolic shifts under sub-optimal cultivations (Wu et al. 2015). To illustrate the effects of oxygen uptake flux and metabolic burden from ATP maintenance loss on fatty acid yields, a genome-scale model (Monk et al. 2013) was employed to simulate cellular physiologies (Fig. 4.3). An apparent trend was that with a decrease in oxygen flux and an increase
in ATP maintenance cost (representing metabolic burden) to certain levels, the yield of fatty acids dropped sharply off a “cliff”, approaching no production (note: the cliff gets steeper as oxygen fluxes decrease). In general, cell metabolism can afford certain metabolic burden without significant decrease of production yield. However, when cells have a higher metabolic burden, the impact of the oxygen supply becomes more significant on fatty acid yields. A small change in oxygen influx may considerably decrease the fatty acid yield (from the blue star to the red star in Fig. 4.3), which is similar to the case for an increase in maintenance energy (metabolic burden). Moreover, when cell physiological status is located on the “metabolic cliff”, its fatty acid production will be highly sensitive to metabolic burdens and the oxygen supply, reducing the reproducibility of strain performance.

![Figure 4.3](image)

**Figure 4.3. Effect of oxygen uptake flux and maintenance energy (i.e., ATP consumption for cell maintenance) on fatty acid yields.**

Color map visualizes the maximal yield of fatty acids under different conditions (oxygen flux and maintenance energy). Red and blue stars represent two strains with the same maintenance energy and different oxygen uptake rates. The default FBA parameters and constraints are taken
from model iJO1366. The FBA assumes a glucose influx $v_{\text{glucose}}$ of 4.5 mmol/gDW/h and a growth rate $\mu$ of 0.04 h$^{-1}$, and a P/O ratio of 1.5.

### 4.4 Conclusion

In summary, this study demonstrates that the introduction of VHb can boost energy metabolism, resulting in enhanced biomass growth and fatty acid titer. The negative effects of metabolic burden on fatty acid production have also been studied by both experiments and FBA simulation, suggesting that SynBio strategies can achieve expected enhancements only if the benefits outweigh metabolic burden. This work expands SynBio strategies for strain development from Push-Pull-Block to Push-Pull-Block-Power. Moreover, this study illustrates that fluxomics studies can provide valuable guidelines for the SynBio and ME communities. Especially, fluxomes can reveal metabolic burdens and the cell energy metabolism, allowing engineers to properly allocate cell resources during strain development (Wu et al. 2016).

### Acknowledgements

This work was supported by the National Science Foundation (DBI 1356669 to YT, MCB1453147 to FZ) and the Human Frontier Science Program (RGY0076/2015 to FZ).
Chapter 5. Deciphering *Clostridium* metabolism and its responses to bioreactor mass transfer during syngas fermentation

Results of this chapter have been published in Scientific Reports, 2017,7
Abstract

This study used $^{13}$C tracers and dynamic labeling to reveal metabolic features (nutrients requirements, pathway delineation and metabolite turnover rates) of Clostridium carboxidivorans P7, a model strain for industrial syngas fermentation, and its implication with bioreactor mass transfer. P7 shows poor activity for synthesizing amino acids (e.g., phenylalanine) and thus, needs rich medium for cell growth. The strain has multiple carbon fixation routes (Wood-Ljungdahl pathway, pyruvate:ferredoxin oxidoreductase reaction and anaplerotic pathways) and Re-citrate synthase (Ccar_06155) was a key enzyme in its tricarboxylic acid cycle (TCA) pathway. High fluxes were observed in P7’s Wood-Ljungdahl pathway, right branch of TCA cycle, pyruvate synthesis, and sugar phosphate pathways, but the cells anabolic pathways were strikingly slow. When syngas flowrate increased from 10 to 20 mL/min, the alcohol productivity was not improved and the labeling rate (~0.03 h$^{-1}$) of key metabolite acetyl-CoA reached to P7 strain’s metabolism limitation regime.

5.1 Introduction

Biological utilization syngas such as $\text{CO}_2$ and CO becomes an important research field due to cheap feedstock and the concerns of global warming. Although photosynthesis cell factories can effective convert $\text{CO}_2$ into biomass, they do not have efficient native pathways for production of extracellular chemicals and thus sophisticated genetic modifications are necessary to develop photo-biorefineries. Unlike photo-biorefinery, syngas fermentation uses native species to converts $\text{CO}_2$, CO and H$_2$ to diverse products such fuels and chemicals (e.g., ethanol, acetic acid, and butanol). Other advantages for syngas fermentation include bioprocess stability and tolerance to inhibitory compounds. Biocatalysts such as Clostridium sp. and Acetobacterium woodii naturally
synthesize alcohol and organic acids and are commonly used in syngas fermentation (Phillips et al. 1993; Wilkins and Atiyeh 2011). Genetic improvements (Alissandratos et al. 2013) and syngas composition optimizations (Fernández-Naveira et al. 2016; Hurst and Lewis 2010; Munasinghe and Khanal 2010) have been attempted to improve efficiency of product synthesis.

Currently, syngas fermentation is still facing challenges because the low solubility of gaseous substrates (CO and H₂) hinders the transport of gas molecules across the gas-liquid interface and diffusion into cells for bioconversion (Fei et al. 2014). The engineering challenges such as gas-to-liquid mass transfer are still restricting industrial syngas fermentation efficiency. Mass transfer can commonly be improved through increasing agitation or gas flowrate. Various reactor designs such as hollow fiber membrane reactor (Shen et al. 2014c), monolithic reactor (Shen et al. 2014b) and rotating packed bed biofilm reactor (Shen et al. 2017) have also been developed to improve mass transfer in syngas fermentation.

In addition to mass transfer limitations, biotransformation of gaseous substrates into metabolites inside the cell can be another bottleneck. There has been a lack of complete knowledge of cell metabolism of syngas fermentation strains. In general, the Wood-Ljungdahl pathway is the key pathway for converting CO/CO₂ into acetyl-CoA in syngas fermentation (Ragsdale 2008; Ragsdale and Pierce 2008). However, the rate and routes of cell metabolism for conversion of acetyl-CoA into cascade metabolites during syngas fermentation has been poorly understood. In order to overcome this shortcoming and enhance the biotransformation of gaseous substrates, it is necessary to perform a thorough investigation of the metabolism of the metabolism pathway in a specific syngas fermentation strain.

Isotope tracer technique has been used to determine the mass transport in bioreactors (Happel et al. 1982; Happel et al. 1986; Happel et al. 1990). This approach can also investigate
cell metabolism by analyzing isotopomer of both proteinogenic amino acids and fast turnover free metabolites (Toya et al. 2007). To investigate whether fermentation process is operated in a mass transfer limitation regime or a metabolism limitation regime, this study designs $^{13}$C-labeling experiments to delineate functional pathways and to identify metabolic rate limiting steps in *Clostridium Carboxidivorans* P7. P7 is a model syngas fermentation strain for producing fuel ethanol from CO/CO$_2$ (Ragsdale and Pierce 2008). This strain has been studied for its microbial physiology and metabolic characteristics (Bruant et al. 2010; Paul et al. 2010; Ukpong et al. 2012) as well as its performance in various bioreactor configurations (Shen et al. 2014b; Shen et al. 2014c). The aim of this work is to elucidate functional pathway for cell growth and syngas conversion under different bioreactor mass transfer scenarios. Via $^{13}$C-fingerprinting of proteinogenic amino acid and fast-turnover metabolites, this study tracks cell adsorption of sugars or syngas (CO/CO$_2$) from culture medium into its biosynthesis pathways (Hollinshead et al. 2016b). The labeling information deciphers cell product synthesis and metabolic responses to bioavailability of carbon substrate (Bosma et al. 1997). In addition, this study reported a reverse-labelling approach following previous report to trace syngas metabolism (Feng et al. 2010). Since continuous flushing $^{13}$CO$_2$ and $^{13}$CO into bioreactor would be prohibitively expensive ($^{13}$CO costs 220$/L, Cambridge Isotope Laboratories, MA), an inverse labeling approach was designed. U-$^{13}$C glucose was used to grow P7 strain so its central metabolites became $^{13}$C-labeled. After organic labelled $^{13}$C glucose were consumed, unlabeled syngas was fed into the culture and time-course samples were taken to monitor un-labelled $^{12}$C entering cell metabolism. The dynamic labeling/un-labelling of metabolites provides new insights into functions of individual central pathway under different mass transfer conditions.

### 5.2 Materials and Methods
5.2.1 Strain, medium and culture preparation.

*C. carboxidivorans* P7 (ATCC-624T) was stored as 1 mL frozen glycerol stocks at -80°C. To recover the cells, the stock culture was inoculated into culture tubes (40 mL) containing 10 mL seed medium, which contained (per liter) 5 g/L glucose and 1 g/L yeast extract (YE) dissolved in basal medium. The basal medium contained 5 g MES (4-Morpholineethanesulfonic acid), 30 mL mineral stock solution, 10 mL of a trace metal solution, 1 mL resazurin sodium salt solution (1% w/v), 10 mL vitamin stock solution, and 10 mL 4% cysteine-sulfide solution as reducing agent. The seed medium (except glucose, vitamin and cysteine-sulfide) was prepared anaerobically and autoclaved at 121°C for 20 min. Glucose stock solution (50%, w/v), vitamin stock solution, and cysteine-sulfide stock solution were then added into the autoclaved liquid through a 0.22-micron filter. The headspace of the culture was pressurized with 10 psi carbon monoxide. All the operations were performed inside an anaerobic chamber (Coy Laboratory Products Inc., MI, USA).

5.2.2 Syngas fermentation in serum bottle systems.

Serum bottles were used to evaluate the necessity of yeast extract in the syngas fermentation of P7 strain and elucidate its central metabolism using $^{13}$C-labelling technology. The aim and design of each set of experiment were summarized in Table 5.1. Specifically, the seed culture in the mid/late exponential phase was inoculated into a 125 mL serum bottles containing 50 mL medium at a 10% inoculum ratio. The medium composition in the serum bottle culture was the basal medium supplemented 1 g/L YE and different combinations of sugars (glucose or fructose) and/or NaHCO$_3$ (Table 5.1). In cases of $^{13}$C labeling tests (Tests 2 and 3, Table 5.1), the seed was first inoculated into serum bottles which only contained basal medium (glucose-free). The bottles was incubated for 12 hours to ensure the residual unlabeled glucose in the inoculum was exhausted to avoid the interference of unlabeled glucose to the labelling results. Then, $^{13}$C-
glucose or $^{13}$C-bicarbonate were injected into cultures before bottles were flushed and pressured with 15 psi syngas containing (v/v) 75% N$_2$, 20% CO, and 5% H$_2$ (Table 5.1). At the end of culture, samples (~15mL) were harvested using falcon tubes then quickly cooled to 0 °C via liquid nitrogen bath (Hollinshead et al. 2016b). The cold samples (~ 0 °C) were centrifuged and the biomass pellets were frozen and stored for subsequent analyses.
Table 5.1. Summary of experimental design for various labeling tests in both serum bottle cultures and bioreactor cultures

<table>
<thead>
<tr>
<th>Experiments</th>
<th>Medium composition</th>
<th>Gas used</th>
<th>Growth condition</th>
<th>Analytical methods</th>
<th>Aims</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serum bottle</td>
<td>5 g/L glucose/fructose and/or 1 g/L NaHCO₃, 1 g/L YE (unlabeled experiment)</td>
<td>Headspace Gas 1 b</td>
<td>Different combinations of the YE and sugars (and/or syngas) were added to the basal medium (as specified in Figure 5.1)</td>
<td>Cell growth (optical density at 660 nm)</td>
<td>Investigate the necessity of yeast extract in P7 cell growth</td>
</tr>
<tr>
<td>(Test 1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum bottle</td>
<td>5 g/L 1-¹³C glucose (or 1, 2-¹³C glucose) 1 g/L Na¹³HCO₃ 1 g/L YE</td>
<td>Headspace Gas 1 b</td>
<td>Inoculation of seed culture into bottle, then add ¹³C glucose and Na¹³HCO₃</td>
<td>GC-MS analysis of proteinogenic amino acids</td>
<td>Investigate the necessity of yeast extract in P7 cell growth</td>
</tr>
<tr>
<td>(Test 2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum bottle</td>
<td>1 g/L Na¹³HCO₃ 1 g/L YE</td>
<td>Headspace Gas 1 b</td>
<td>Inoculation of seed culture into bottle, then add Na¹³HCO₃</td>
<td>GC-MS analysis of proteinogenic amino acids</td>
<td>Identify metabolic pathway and carbon transitions</td>
</tr>
<tr>
<td>(Test 3)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Serum bottle</td>
<td>1 g/L Na¹³HCO₃ 1 g/L YE</td>
<td>Headspace Gas 1 b</td>
<td>NaH¹³CO₃ and Gas1 was added once yeast extract exhausted (OD₆₆₀ ~0.22)</td>
<td>LC-MS analysis of free metabolites</td>
<td>Investigate dynamic ¹³C-labeling from NaH¹³CO₃ incorporation</td>
</tr>
<tr>
<td>(Test 4)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bioreactor</td>
<td>4 g/L glucose 1 g/L YE (unlabeled experiments)</td>
<td>Flushing Gas 2 c</td>
<td>Syngas was aerated after glucose was depleted. Three gas flow rate used (1, 10 and 20 mL/min)</td>
<td>GC-FID analysis of bio-production</td>
<td>Test cell growth and production of carboxylic acids and alcohols under different flow rates</td>
</tr>
<tr>
<td>(Test 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bioreactor</td>
<td>4 g/L U-¹³C glucose 1 g/L YE</td>
<td>Flushing Gas 2 c</td>
<td>¹³C-glucose was fed to the culture, then un-labelled syngas 20 ml/min was aerated</td>
<td>LC-MS analysis of free metabolites</td>
<td>Investigate cells dynamic metabolism using inverse labeling</td>
</tr>
<tr>
<td>(Test 6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bioreactor</td>
<td>4 g/L U-¹³C glucose 1 g/L YE</td>
<td>Flushing Gas 2 c</td>
<td>¹³C-glucose was fed to the culture, then un-labelled syngas 10 ml/min was aerated</td>
<td>LC-MS analysis of free metabolites</td>
<td>Investigate cells dynamic metabolism using inverse labeling</td>
</tr>
<tr>
<td>(Test 7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a The basal medium as described in the seed preparation section was used for all the culture.
b Headspace Gas 1: The serum bottle headspace was pressurized with 75% N₂, 20% CO, and 5% H₂ (v/v) at 15 psi total pressure.
c Aerated Gas 2: The reactor were first flushed with N₂ with labelled glucose and then switched to syngas growth mode by flushing a gas mixture containing 60% CO, 37.5% CO₂, and 12.5% H₂ (v/v). The volumetric gas flow rates were shown in the table.
5.2.3 Syngas fermentation in bioreactor systems.

P7 fermentation were performed using Applikon MiniBio reactor systems. The bioreactor has 250 mL working volume (37°C and 500 rpm agitation), and culture pH was controlled at 6.0 using 1 M NaOH solution. Since continuous flushing of $^{13}$CO and $^{13}$CO$_2$ into bioreactors is prohibitive expensive, we designed an inverse dynamic labeling approaches to reveal the metabolism of P7 cells under different flow rate. In brief, fully labeled $^{13}$C-glucose was first used as a carbon source for cell growth. During this stage, N$_2$ gas was flushed to the vessel and $^{13}$C-glucose was the major carbon source for labeling intracellular metabolites. The exhausting gas from the vessels was passed through a standard MiniBio system condenser then an ice-cold water trap submerged in an ice water bath to trap volatile products. The volumetric gas flow rates used for different sets of the bioreactor cultures were shown in Table 5.1. Once glucose was exhausted in bioreactor, culture was switched to syngas growth mode by switching flushing gas from N$_2$ to a gas mixture containing 50% CO, 37.5% CO$_2$, and 12.5% H$_2$ (v/v) with the same flow rate. After gas switching, pH control was stopped to minimize perturbations from NaOH feeding. To track dynamics of metabolite labeling of P7 cells, broth sample (~20 mL) were harvested at 0s, 30min, 5hr, 24hr, 48hr, 65hr, 88hr and 120 hr, respectively. Each sample were placed in falcon tubes and immediately cooled to ~0 °C by liquid N$_2$ bath for 10s to stop cell metabolism (note: samples need be stirred during liquid N$_2$ bath to avoid being frozen). The quenched samples were then centrifuged (8,000 g) at 4°C and biomass pellets were frozen for further analysis of free metabolites.

5.2.4 Isotopomer analysis.

Analysis of free metabolites followed a previous protocol (Hollinshead et al. 2016b). Briefly, cell pellet was suspended in 1 mL methanol/chloroform solution (7:3 v/v) and shake at
150 rpm at 4 °C. Water was added to the cell-solvent mix to extract cell metabolites. The aqueous phase was filtered through an Amicon Ultra centrifuge filter (3000 Da; EMD Millipore, Billerica, MA), lyophilized, and dissolved in acetonitrile and water (6:4, v/v) solution for LC-MS measurement (Agilent Technologies 1200 Series equipped with a SeQuant Zic-pHILIC column; LC-MS analysis was performed at Lawrence Berkeley National Laboratory). MS distributions of the metabolite were determined based on the ratio of the integrated peak area of the chosen isotopomer to the sum of integrated peak areas of all isotopomers.

The proteinogenic amino acids was measured by GC-MS followed previous protocol (You et al. 2012). Cell biomass was hydrolyzed with 6 M HCl at 100°C, air-dried and then derivatized with N-tert-butyldimethylsilyl-N-methyltrifluoroacetamide for GC-MS analysis. MS data [M-15]+ and [M-57]+ represents entire amino acid, and [M-159]+ or [M-85]+ represents amino acid losing carboxyl group. For leucine and isoleucine, their [M-57]+ signal was overlapped by other peaks and [M-15]+ was analyzed. All MS data were corrected to remove the noise from natural isotopes using published algorithms (Wahl et al. 2004). Labeling fractions (M0, M1, M2...) represent MS fragments with 0, 1, 2 labeled carbons.

5.3 Results and Discussion

5.3.1 Importance of yeast extract for P7 growth.

The first task is to examine the necessity of medium nutrients, particularly yeast extract (YE), on P7 growth (Test 5.1, Table 5.1). The information will inform us the influence of rich medium on labelling test and metabolic flux quantifications. This is because the complex carbon nutrients (i.e., amino acids) in YE can incorporated into biomass, and thus, interfere with a quantitative flux analysis. As shown in Fig. 5.1, P7 barely grew in YE-free medium even carbon sources (glucose (GLU) or fructose (FRU)) were provided. This result was similar to previous report that P7 syngas
fermentation in YE-free medium experienced a very long lag phase (5 days) with minimal cell density achieved over 600 hours (Phillips et al. 2015). YE addition, even without sugars, improved cell growth significantly \( (p<0.01, \text{Fig. 5.1}) \). Such a benefit effect of YE on cell growth indicates that YE can be used in the initial stage of P7 culture to promote rapid cell growth before implementing syngas fermentation stage. Thus, it shortens the syngas fermentation stage and save the gas pumping cost. The cell growth was further improved when sugars (glucose or fructose) were provided. However, supplementation of syngas in the YE- and sugar-containing medium did not further promote cell growth, indicating sugar was the preferred carbon source for P7 cells.

![Figure 5.1. Growth of P7 cells in the basal medium containing different combinations of yeast extract (YE) and carbon sources (Test 1, Table 5.1).](image)

**Figure 5.1.** Growth of P7 cells in the basal medium containing different combinations of yeast extract (YE) and carbon sources (Test 1, Table 5.1).

**X-axis legends:** YE free (Fru/Glu): YE-free and 5 g/L glucose or fructose; YE (carbon free): 1 g/L YE without carbon source; YE (Fru): 1 g/L YE and 5 g/L fructose; YE (Glu): 1 g/L YE and 5 g/L glucose. YE (Glu & syngas): 1 g/L YE, 5 g/L glucose, and 1 g/L NaHCO\(_3\) with syngas mixture in headspace (gas to liquid ration is 2:1).

To further investigate the role of YE for P7 growth, a glucose- or NaHCO\(_3\)- labelling test were performed. Fig. 5.2 shows amino acid labeling results when \(^{13}\)C-glucose (either \(^{1-^{13}}\)C glucose
or 1,2-\textsuperscript{13}C glucose) and NaH\textsuperscript{13}CO\textsubscript{3} were used in the medium (Test 2, Table 5.1). As shown in Fig. 5.2, all amino acids had a significant un-labelled isotopomer, indicating that a large portion of those amino acids were not \textit{de novo} synthesized but rather taken from YE. The labelled aspartate and glutamate isotopomers were significant higher than other labelled amino acids because of \textsuperscript{13}C flux through the TCA metabolites ($p<0.05$, Fig. 5.3). Fig. 5.2 also shows that for all amino acids except methionine, addition of syngas CO and H\textsubscript{2} into \textsuperscript{13}C-glucose culture (either 1-\textsuperscript{13}C glucose or 1,2-\textsuperscript{13}C glucose) did not change labeling of those proteinogenic amino acids (Fig. 5.2A-2J). This observation can be interpreted that \textit{Clostridium} species demonstrate the hierarchy of nutrient utilization (i.e., utilize a preferred carbon source in multiple-substrate medium) (Aristilde et al. 2015). For methionine, presence of CO and H\textsubscript{2} resulted in $\sim$10\% more labelled methionine molecules (note: oxaloacetate (OAA) and 5,10-methyltetrahydrofolate (C1) are its precursors), supporting the fact that CO and H\textsubscript{2} as electron donors facilitate labeled CO\textsubscript{2}, which came from NaH\textsuperscript{13}CO\textsubscript{3}, to enter methyl branch of the Wood-Ljungdahl pathway. Collectively, the results in Fig. 5.2 indicate that P7 has a relatively weak capability for \textit{de novo} synthesizing several key amino acids although its genome contains complete annotations of all amino acid pathways. It is therefore necessary to use rich nutrient (e.g., YE) to support biomass growth and reduce P7 fermentation duration. As a result, a pathway delineations rather than quantitative flux analysis of the P7 cells were performed in this work.
Figure 5.2. Labeling of proteinogenic amino acid (not fragmented, [M-57]⁺ or [M-15]⁺) after P7 growth in ¹³C-labeled substrate for two days (Test 2, Table 5.1).

For all the cultures, basal medium added with 1 g/L Na¹³HCO₃ and 1 g/L YE. The different columns represent different gas composition in the headspace and the labelled glucose in the medium.

A: headspace gas (75% N₂, 20% CO, and 5% H₂ (v/v)) with 1⁻¹³C glucose;

B: headspace gas (N₂) with 1⁻¹³C glucose;

C: headspace gas (75% N₂, 20% CO, and 5% H₂ (v/v)) with 1,2⁻¹³C glucose,

D: headspace gas (N₂) with 1,2⁻¹³C glucose.
Figure 5.3. Major metabolic pathways (the Wood Ljungdahl pathway, the TCA cycle, the pyruvate metabolism) and carbon transition identified in P7 strain.

The cells were grown in basal medium containing Na\textsuperscript{13}HCO\textsubscript{3} with syngas (Test 3, Table 5.1). Cells were harvested at day 6 and subjected to proteinogenic amino acids labelling analysis. Re-citrate synthase (marked as brown), pyruvate carboxylase (marked as blue), fumarate hydratase (Marked as green), and malate dehydrogenase (marked as red, malate \(\rightarrow\) pyruvate + CO2) were annotated as the key enzymes for the TCA pathway. Embden–Meyerhof–Parnas pathway (EMP); phosphoenolpyruvate (PEP); ribulose 5-phosphate (Ru5P).

5.3.2 Analysis of functional pathways in P7.

The central pathways of P7 were delineated via NaH\textsuperscript{13}CO\textsubscript{3} and unlabeled CO and H\textsubscript{2} in the serum bottle cultures (Test 5.3, Table 5.1). The labeling of metabolites in central pathways was
deduced from isotopomer analysis of proteinogenic amino acid. As shown in Fig. 5.3, most leucine was found unlabeled, indicating most of its precursor, Acetyl-CoA, was almost unlabeled. Furthermore, alanine was mainly labeled in its first position (i.e., the loss of carboxyl group of alanine resulted in unlabeled fragment), indicating C2 and C3 carbons of pyruvate (the precursor of alanine) were mostly unlabeled (Fig. 5.3). Oxaloacetate (precursor of aspartate) was labeled with one (39%) and two carbons (15%), while ketoglutarate (precursor of glutamate) was mainly labeled with one carbon (56%). The labeling of pyruvate and oxaloacetate/ketoglutarate confirmed that labeled carbon from NaH\textsuperscript{13}CO\textsubscript{3} was fixed via pyruvate:ferredoxin oxidoreductase (Ccar_01615) and carboxylase (Ccar_18725), while little amount of \textsuperscript{13}C formed Acetyl-CoA. This observation can be interpreted by preferred uptake of CO\textsubscript{2} instead of bicarbonate by the Wood-Ljungdahl pathway (Thauer 1973; Thauer et al. 1975) (note: autotrophic microbes such as algae can use both bicarbonate and gaseous CO\textsubscript{2}).

P7 genome contains an incomplete TCA cycle, missing malate dehydrogenase (catalyze malate \xrightarrow{\longrightarrow} oxaloacetate), succinate-CoA ligase, and ketoglutarate dehydrogenase. In serum bottle cultures with NaH\textsuperscript{13}CO\textsubscript{3}, Fig. 5.4 shows lack of labeling in malate and succinate. A weak TCA cycle is common for anaerobic bacteria. Interestingly, P7 genome lacks Si-citrate synthase, the starting point of the TCA cycle. However, manual search of Re-citrate synthase indicates a gene 2-isopropylmalate synthase (Ccar_06155) identical to the reported Re-citrate synthase (amino acid sequence similarity \textasciitilde79\%). This alternative citrate synthase gene was confirmed by labeling data (Fig. 5.3): \(\alpha\)-carboxyl group of glutamate was unlabeled concurrently with carboxyl group of pyruvate (Tang et al. 2007). Such labeling signature highlights the existence of Re-citrate synthase that causes \textsuperscript{13}C atom transitions to \(\beta\)-carboxyl group of glutamate (Fig. 5.3). In some anaerobic bacteria, citramalate synthase often co-exists with Re-citrate synthase to involve atypical
isoleucine synthesis from acetyl-CoA (Feng et al. 2009). However, BLAST search showed no gene candidate for citramalate synthase and P7 contained a normal threonine dependent isoleucine pathway. This view is supported by isotopic tracing that showed isoleucine was from oxaloacetate and had different labeling patterns from acetyl-CoA derived leucine (Zhuang et al. 2014) (Fig. 5.3).

Figure 5.4. Dynamic labeling of key metabolites under serum bottle condition (Test 4, Table 5.1).

Na\textsuperscript{13}HCO\textsubscript{3} was pulsed with headspace syngas at t=0. Dynamic labeling samples were harvested at 1min, 10min, 30min, 1hr, 2hr, 5hr, 18hr and 47hr and then subjected to free metabolites labelling analysis.

5.3.3 P7 growth and bio-production during syngas fermentation.

In a typical syngas fermentation process, CO, CO\textsubscript{2}, and H\textsubscript{2} must move across the gas-liquid interface and be accessible to microbe cells. Once absorbed, the substrates must be converted into the desired products (such as ethanol) using an efficient metabolic pathway. During “substrate (CO,
CO₂, and H₂) → product (ethanol)” conversion process, a series of sequential transport phenomena occurred, including physical bulk gas-to-liquid mass transfer, the substrate moving across the cell membrane, and enzymatic conversion of the substrate molecules into various metabolites. In general, syngas fermentation is operated in one of two regimes, a gas-liquid mass transfer limitation or a kinetics/metabolism limitation, depending on the bioreactor operational conditions and the cell intrinsic characteristics. Here, syngas fermentation was performed under three gas flow rate scenarios to determine whether the cells are in a mass transfer limitation regime or metabolism regime (Test 5.5, Table 5.1). As shown in Fig. 5.5, glucose was first used as a substrate to grow biomass in a YE-containing medium; the cell growth increased rapidly within the first 12 hours. After glucose was exhausted, the cells ceased growth and the culture was switched to syngas flushing mode. The culture experienced ~24 h lag phase for adapting new substrates (syngas). During this stage, cell acidogenesis ceased and culture pH maintained at around 6.0. After lag phase, the culture resumed its growth and entered acidogenesis phase. This triggers pH drops to ~5.5 and consequently solventogenesis to produce butanol and ethanol (Fig. 5.5). This growth pattern was also reported elsewhere (Ganigué et al. 2016).
Figure 5.5. Cell growth and medium pH changes during syngas fermentation of P7 strain under different gas flow rates (Test 5, Table 5.1).

P7 cells were grown in medium containing unlabeled glucose as a carbon source with flushed by N₂ gas at three flow rates (1, 10 and 20 mL/min). When the glucose was depleted, unlabeled syngas mixture was aerated into reactor. Red line represents the time point switching to the syngas. Data are means of three replicates and the error bars represent the standard deviations.

5.3.4 P7 metabolism during syngas fermentation under different gas flow rates.

An inverse dynamic $^{13}$C-labelling experiment was conducted in P7 syngas fermentation to delineate the dynamic metabolism of free metabolites and proteinogenic amino acids (Tests 6 & 7, Table 5.1). In this inverse labelling test, labelled glucose was used for the initial culture to label the cell metabolites. Upon glucose depletion, un-labelled syngas was used for the cell culture to dilute the labeled metabolites and determine the speed of syngas carbons percolating metabolic network. Fig. 5.6 shows the relative abundance of labelled and un-labelled key metabolites of P7
cells in inverse dynamic labelling tests. After the cultures were switched to unlabeled syngas (flowrate=10 or 20 mL/min), $^{13}$C acetyl-CoA decreased without lag phase (Fig. 5.6A). However, other central metabolites such as free alanine (Fig. 5.6B) had longer lag phase before the unlabeled compound accumulated. The continued change of unlabeled-acetyl-CoA confirmed that the Wood-Ljungdahl pathway has fast responses to gaseous substrates. However, carbons are trapped in acetyl-coA node and require much longer time for cell re-organize its fluxes percolating through downstream pathways. After a lag phase, the fixed carbons in acetyl-CoA began actively synthesizing downstream metabolites. Doubling flowrate from 10 to 20 mL/min did not cause significant difference in acetyl-CoA labeling rates (improve from 0.024 h$^{-1}$ to 0.030 h$^{-1}$, $P$ value > 0.05), indicating cells were in a metabolism limitation regime at the flowrate above 10 mL/min. Fig. 5.6 shows that TCA metabolites and free amino acids (alanine, citrate, glutamate and aspartate) became unlabeled-dominated (M0>50%) with the progression of syngas fermentation, supporting Fig. 5.3 that unlabeled CO/CO$_2$ flew from acetyl-CoA $\rightarrow$ pyruvate (precursor of alanine) $\rightarrow$ oxaloacetate (precursor of aspartate) $\rightarrow$ citrate $\rightarrow$ ketoglutarate $\rightarrow$ glutamate. The low conversion from labeled to unlabeled malate (Fig. 5.6E) supports that P7 operates an incomplete TCA cycle.
Figure 5.6. Relative abundance of labeled and un-labelled key metabolites of P7 cell in inverse dynamic labeling tests (Tests 6 & 7, Table 5.1).

$^{13}$C-glucose was used for the culture in the first 12 h when N$_2$ gas was aerated into the reactor, unlabeled syngas was aerated when $^{13}$C-glucose was depleted. Two flow rates were used (low flowrate=10mL/min; high flowrate=20mL/min).

Data were means of three replicates and error bars represent standard deviations. **Legends:**

□: M0 labeling under low flowrate; △: M0 labeling under high flowrate, ■ $^{13}$C enrichment under low flowrate; ▲ $^{13}$C enrichment under high flowrate. **Solid line:** data from high flow rate samples; **dash line:** data from low flow rate samples.
**Note:** The Wood-Ljungdahl pathway mainly turnover the acetyl group of acetyl-CoA. Thus, the labeling of acetyl group was also estimated in Fig 5.6A: $^{13}$C enrichment in the acetyl group under high (●) and low (◆) flowrates. $k_{LF}$ and $k_{HL}$ are rate coefficients calculated by simulation for acetyl group labeling under low flowrate and high flowrate, respectively. 95% confidence interval was shown in parentheses.

Fig. 5.6 shows that the labeling rates of sugar phosphate metabolites such as glucose 6-phosphate (G6P), glycerate-3-P (G3P), and ribulose 5-phosphate (Ru5P) were slower than labeling of TCA metabolites, with an appreciable amount labelled compounds remaining through entire fermentation period. This result indicates thermodynamic barriers blocking fluxes from pyruvate towards gluconeogenesis and non-oxidative pentose phosphate (PP) pathways. Fig. 5.6 demonstrates a faster synthesis of key metabolites (e.g., glutamate and malate) in central metabolism in bioreactor culture than that in serum bottle culture (Fig. 5.4), possibly due to the enhanced the mass transfer efficiency and better bioavailability of syngas substrates. Moreover, the increase of the un-labelled proteinogenic amino acid proportion after un-labelled syngas flushing for 105h was also determined. As shown in Fig. 5.7, all amino acid species demonstrated low increment of its un-labelled fractions. For example, serine, alanine, and aspartate had 20–35% increment of un-labelled fractions. This observation indicates that syngas carbons were minimally used for anabolism and thus P7 protein (Fig. 5.7) had very slow turnover rates even at high flowrate syngas.
Figure 5.7. Enrichment of the un-labelled proteinogenic amino acids in the inverse dynamic labelling syngas fermentation.

Two gas flow rates were used (Test 6 & 7, Table 5.1). Cells were inoculated with $^{13}$C-glucose and grew until glucose used up, then unlabeled syngas was aerated. Enrichment was defined as the difference between the un-labelled amino acids at beginning of unlabeled syngas aeration (0 hr) and the end of the fermentation (105 hr). Error bars represent the standard deviation of duplicated experimental data.

Overall, the inverse dynamic labelling test of P7 cells indicates that increasing flow rate from 10 to 20 mL/min did not significantly ($P$ value > 0.05) increase the production of free metabolites and proteinogenic amino acids (Figs. 5.5 & 5.6). The results, together with the cell growth performance in Fig. 5.5, demonstrate that the P7 strain was not limited by mass transfer under this flow rate range (10-20 mL/min); instead, P7 intrinsic metabolic capability limited substrate utilization and bio-production. In contrast to photoautotrophic metabolism cyanobacteria whose CO$_2$ fixation rate and flux through central pathways are in the order of minutes (Young et al. 2011a), P7 metabolic conversion rate (in the order of hours) is much slower under syngas
conditions: the Wood-Ljungdahl pathway, right branch of TCA cycle, and pyruvate synthesis showed relatively fast rates, while other biosynthesis pathways have low fluxes even under high flowrate syngas. Since solventogenesis consumes significantly more NAD(P)H than acidogenic metabolism, high syngas mass transfer is necessary for cell membrane to obtain electrons from H₂ and CO for alcohol production and biosynthesis (Amador-Noguez et al. 2011). This leads to conclusion that the energy metabolism rather than carbon metabolism was more influenced by mass transfer limitations in our experiments.

5.4 Conclusion

This study is first to elucidate metabolic pathways and mass transfer under syngas fermentation conditions via ¹³C labeling using P7 cells as model strain. The outcomes offer novel insights. First, complex nutrients (such as yeast extract) are essential for P7 cell growth. Also, the presence of organic carbon (such as sugars) repress syngas utilizations by P7 strain. Second, P7 cells can utilize CO₂ through the Wood-Ljungdahl pathway (Xiong et al. 2016), pyruvate:ferredoxin oxidoreductase, and anaplerotic pathways. Third, P7 strain contains a novel Re-citrate synthase (Ccar_06155). Fourth, only a few pathways in P7 cells are highly active under syngas metabolism with minimal fluxes through protein synthesis. Fifth, the Wood-Ljungdahl pathway can quickly take C₁ substrates after culture switching from glucose medium to syngas conditions, but downstream pathways require much longer time for flux adjustment. In summary, this study bridges the gap between cell metabolisms and bioprocess conditions, which offers broad impact on gas fermentation applications.
Acknowledgements

This work was supported by NSF (CBET 1438042). We are also thankful for Mary Abernathy, Lian He and Di Liu at Washington University and Edward Baidoo from Lawrence Berkeley National Lab for helping with LC-MS analysis.
Chapter 6. Modeling and machine learning of

_Clostridium_ fermentations under different

syngas compositions
Abstract

This study applied metabolic modeling and machine learning to the fermentations of *Clostridium carboxidivorans* P7 under various syngas compositions and flowrates. Flux balance analysis was used to predict syngas fermentation performance under substrate mass transfer constraints, empirical kinetic model was built to describe the dynamic tendency of biomass growth and bioproducts accumulation, while a neural network model could obtain quantitative good fits of fermentation data. This study demonstrates the connection of fermentation operations to both metabolic modeling and machine learning may help engineers understand and design bioprocesses under complex influential factors.

6.1 Introduction

As global energy demand increases, the need for renewable fuels has become important. Research has focused on second generation biofuels/chemicals produced through the conversion of lignocellulosic biomass (Daniell et al., 2012; Latif et al., 2014; Munasinghe and Khanal, 2010). Syngas fermentation is an efficient approach to achieve this environmental friendly and economically process. Previous studies on the effects of gas composition on fermentation performance has been focused on single gas species effect (Hurst and Lewis, 2010; Skidmore et al., 2013) or a gas mass transfer perspective (Chen et al., 2015; Mohammadi et al., 2014; Xu and Lewis, 2012). However, the synergistic effect of multiple syngas components on cell growth and products accumulation has been rarely reported, and modeling of complex syngas fermentation variables is also lacking. In this study, metabolic modeling and machine learning approaches are applied for simulating bioconversion outcomes, which offers guidelines and insights into optimal bioprocessing (Pappu and Gummadi, 2016).
6.2 Materials and Methods

6.2.1. Strain, medium and seed cultures

*Clostridium carboxidivorans* P7 (ATCC-BAA624) was used in this study. Seed culture procedures follow a previous report (Ahmed, 2006). To prepare seed cultures, the strain was grown in anaerobic serum vials using P7 medium containing (per liter) 1 g yeast extract, 5 g MES (4-morpholineethanesulfonic acid), 30 mL mineral stock solution, 10 mL trace metal solution, and 1 mL resazurin sodium (1 wt%). The P7 medium was autoclaved for 20 mins. Then vitamin stock solution (10 mL/L) and a 4% cysteine-sulfide reducing agent (10 mL/L) were added through a 0.20-micron filter under anaerobic conditions. The headspace of seed culture was filled with CO gas. The seed culture was inoculated into an Applikon Mini Bioreactor (10% inoculation ratio). The P7 medium in reactor contained 4 g/L of glucose to shorten syngas fermentation duration (Wan et al., 2017). Reactors were maintained at 37°C with 500 rpm agitation. In the initial glucose-based growth stage, nitrogen gas was sparged into the vessel at 20 ml/min through a micro bubbler and the medium pH was controlled at 6.0 using 1 M NaOH. Once glucose in the medium was exhausted in 24 hours (biomass OD<sub>660</sub>~1.06 ±0.19), the vessel was switched to syngas pulse and the pH control was shut off. The gas compositions and flowrates were shown in Table 6.1.

Table 6.1. Gas compositions used in syngas fermentation of *C. carboxidivorans* P7<sup>a,b</sup>

<table>
<thead>
<tr>
<th>Gas composition</th>
<th>N&lt;sub&gt;2&lt;/sub&gt; (%)</th>
<th>CO (%)</th>
<th>CO&lt;sub&gt;2&lt;/sub&gt; (%)</th>
<th>H&lt;sub&gt;2&lt;/sub&gt; (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>#1</td>
<td>60</td>
<td>20</td>
<td>15</td>
<td>5</td>
</tr>
<tr>
<td>#2</td>
<td>60</td>
<td>0</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>#3</td>
<td>60</td>
<td>40</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>#4</td>
<td>40</td>
<td>20</td>
<td>15</td>
<td>25</td>
</tr>
<tr>
<td>#5</td>
<td>20</td>
<td>20</td>
<td>35</td>
<td>25</td>
</tr>
<tr>
<td>#6</td>
<td>20</td>
<td>40</td>
<td>15</td>
<td>25</td>
</tr>
</tbody>
</table>
Gas flow rate was controlled at 20 ml/min
Gas composition was based on volume percentage

6.2.2. Analyses of cell growth and syngas fermentation products

Growth of P7 strain was monitored by measuring optical density of the culture OD₆₆₀ using a Thermo Fisher Scientific Spectronic 20 Genesys spectrophotometer. Syngas fermentation products (acetate, butyrate, ethanol and butanol) were determined using a Varian 450 gas chromatography (GC) system equipped with a split/splitless injector and flame ionized detector (FID). The compounds were separated using a (60 m x 0.25 mm ID x 0.25 μm) Phenomenex Zebron ZB-1701 column with a 5 m Zebron Z-guard column. Conditions and parameters are as follows: inlet temperature was held at 250°C with an injection volume of 0.1 mL with a 1:20 split ratio. Oven temperature profile was set as follows: 30°C for 16 min and increased to 75 °C at 20 °C/min and hold for 5 min followed by another ramp to 250 °C at 30 °C/min and held for 1 min. Helium was used as carrier gas. The pressure of the gas was maintained at 10 psi for 16 min and increased to 20 psi at 5 psi/min for 12 minutes. The temperature of FID was set at 280 °C. The compounds were quantified by comparing peak areas of the samples to those of known standards. Varian Galaxie Ver. 1.9 software was used to process the data.

6.3 Results and Discussion

6.3.1 Flux balance analysis

Three modeling approaches are used to generalize, compare and predict *Clostridium* metabolic performance under different conditions. First, stoichiometry-based flux balance analysis (FBA) has been used to predict theoretical product yields and global tendency for biosynthesis as
the function of syngas substrates (Milne et al., 2011; Roberts et al., 2010; Senger and Papoutsakis, 2008). An established model iHN637 for Clostridium ljungdahlii (Nagarajan et al., 2013) (containing 785 reactions and 698 metabolites) contained similar metabolic structure and reactions for syngas fermentations, which could serve as a convenient tool to obtain the theoretical productions of biomass growth, acids (e.g., acetate) and alcohol (e.g., ethanol) as a function of inlet gases. Figure 6.1 set H_2 influx as 0 mmol/gDW/h (Figures 6.1A~C), 0.75 mmol/gDW/h (Figures 6.1D~F), and 5 mmol/gDW/h (Figure 6.1G~I). With each fixed H_2 influx, the effect of CO_2/CO influxes on maximal biomass/acetate/ethanol productions was predicted via MATLAB linear programming solver. Theoretically, cell showed no apparent growth when CO uptake rate was below 1 mmol/gDW/h and H_2 uptake was below 0.75 mmol/gDW/h (Figure 6.1A and D), because carbon utilizations were consumed for cell maintenance. As H_2 uptake rate increased from 0 to 5 mmol/gDW/h, cell growth rates increased from 0.012 h\(^{-1}\) to 0.0211 h\(^{-1}\) (i.e., a doubling time of 33 h). Growth rate only increased linearly with the increasing CO uptake. In this case, CO served as both energy and carbon source when H_2 uptake was insufficient (Figures 6.1A~F). Due to limitations in energy metabolism (low CO and H_2 utilizations), the increase of CO_2 uptake capability did not benefit the fermentation (shown as deep blue zones of Figure 6.1). At high H_2 uptake rate (Figure 6.1G~I), increase of both CO_2 (from 0 to 2.5 mmol/gDW/h) and CO (from 0 to 1 mmol/g DW/h) promoted product formations, while CO influx has higher impact on fermentation yields than CO_2. This simulation result indicates that the CO mass transfer is the key factor to control syngas fermentation performance. If CO supply was limited, fermentation outcomes would shift to the blue zones of Figure 6.1.
Figure 6.1. Prediction of optimal rates of biomass growth (figure column 1), acetate production (figure column 2), and ethanol production (figure column 3).

H₂ uptake rate as 0 mmol/gDW/h (A,B,C), 0.75 mmol/gDW/h (D,E,F), 5 mmol/gDW/h (G,H,I), and ranging CO₂ uptake rate from 0~5 mmol/gDW/h, CO uptake rate from 0~5 mmol/gDW/h. Growth rates, acetate production and ethanol production by setting each one as objective function for FBA linear optimization, respectively.

6.3.2 Empirical kinetic model

Based on the experimental results (biomass, ethanol and acetate production) with various syngas compositions and flowrates (Table 6.1 and Figure 6.2). The kinetic model was built to
evaluate the dynamics of fermentations. The kinetic model contained three time-dependent variables: X: cell density, AC: acids (acetic acid) concentration, and AL: alcohols (ethanol) concentration. We assume that dissolved gas concentrations are proportional to their partial pressures and mass transfer coefficients. The biomass growth equation is:

\[
\frac{dX}{dt} = k^* \cdot p_{[CO]} \cdot p_{[CO_2]} \cdot (p_{[CO]} + \frac{p_{[H_2]}}{p_{[CO]} + K_I}) \cdot X - K_d \cdot X \tag{6.1}
\]

where \(p_{[CO]} \cdot p_{[CO_2]}\) are gas fractions (no unit) that represented the carbon contribution to biomass accumulation. In this case, in accordance with experimental data (Composition 2 & 3), the lack of CO and CO\(_2\) led to minimal biomass growth. The term \(\left(p_{[CO]} + \frac{p_{[H_2]}}{p_{[CO]} + K_I}\right)\) represented contributions of energy molecule NADH from CO and H\(_2\) oxidations. \(k^*\) was growth rate constant (hr\(^{-1}\)) which is related with mass transfer coefficients. \(K_I\) represented the CO inhibition on H\(_2\) uptake. Ethanol and butanol production were combined as total alcohol production AL that relied mostly on the energy source:

\[
\frac{dAL}{dt} = B_{alcohol} \cdot X \cdot \left(p_{[CO]} + \frac{p_{[H_2]}}{p_{[CO]} + K_I}\right) \tag{6.2}
\]

Here, \(B_{alcohol}\) is non-growth associated coefficient. Acetic and butyric acid production were combined as total acid production (AC), which was considered as growth-associated. Here \(A_{A/X}\) is the growth associated rate coefficient. \(k_{alcohol}\) described alcohol production from organic acids, and AC production rate can be determined as:

\[
\frac{dAC}{dt} = A_{A/X} \cdot X - k_{alcohol} \cdot X \cdot AL \tag{6.3}
\]

The closed formation for these equations were shown as below:

\[
X = c_1 \cdot e^{\left(k^* \cdot p_{[CO]} + p_{[CO_2]} \cdot \left(p_{[CO]} + \frac{p_{[H_2]}}{p_{[CO]} + K_I}\right) \cdot K_d \right) \cdot t} \tag{6.4}
\]
\[ AL = c_2 + B_{alcohol} \cdot e^{\left(k^* P_{[CO]} P_{[CO_2]} \left( \frac{P_{[CO]}}{P_{[CO]} + K_f} \right)^{-K_d} \right) + t} \left( P_{[CO]} + \frac{P_{[H_2]}}{P_{[CO]} + K_f} \right) \cdot t \quad (6.5) \]

\[ AC = c_3 + e^{\left(k^* P_{[CO]} P_{[CO_2]} \left( \frac{P_{[CO]}}{P_{[CO]} + K_f} \right)^{-K_d} \right) + t} \left( A_A - k_{alcohol} \cdot (c_2 + B_{alcohol} \cdot e^{\left(k^* P_{[CO]} P_{[CO_2]} \left( \frac{P_{[CO]}}{P_{[CO]} + K_f} \right)^{-K_d} \right) + t} \left( P_{[CO]} + \frac{P_{[H_2]}}{P_{[CO]} + K_f} \right) \right) \cdot t \right) \quad (6.6) \]

These equations showed the relationship between the input variables and output variables. Biomass accumulation increased exponentially with the increasing of carbon source and energy source. While, alcohol production was linearly boosted by energy source and biomass accumulation. Acetate production was firstly increasing, when the alcohol production reached

\[ A_A = k_{alcohol} \cdot (c_2 + B_{alcohol} \cdot e^{\left(k^* P_{[CO]} P_{[CO_2]} \left( \frac{P_{[CO]}}{P_{[CO]} + K_f} \right)^{-K_d} \right) + t} \left( P_{[CO]} + \frac{P_{[H_2]}}{P_{[CO]} + K_f} \right) \cdot t \]

(6.7), there would be optimal condition for acetate production. While, with the further increasing of alcohol production, when

\[ A_A < k_{alcohol} \cdot (c_2 + B_{alcohol} \cdot e^{\left(k^* P_{[CO]} P_{[CO_2]} \left( \frac{P_{[CO]}}{P_{[CO]} + K_f} \right)^{-K_d} \right) + t} \left( P_{[CO]} + \frac{P_{[H_2]}}{P_{[CO]} + K_f} \right) \cdot t \]

(6.8), acetate production was repressed by the energy molecules, which boosted the formation of alcohol production.

The ‘ode45’ command in MATLAB was used to solve the differential equations. Figure 6.3 shows the model simulations with different \( k^* \) and gas compositions. In general, biomass accumulation (Figure 6.3A) was enhanced by increased CO, H\(_2\) and CO\(_2\) pressures or higher flow rates. Enhancing CO or H\(_2\) gas pressures also promoted alcohol production (Figure 6.3B) but at the expense of acid consumption (Figure 6.3C). In general, predictions of product curves qualitatively described the fermentation dynamics. However, the kinetic modeling is still difficult
to fit time-course experimental data because of underdetermined influential variables (such as growth inhibitions) and insufficient experimental data for substrate uptakes.

Figure 6.2 Experimental results of biomass, acetate and ethanol production under various syngas composition.

A. Biomass growth under various syngas composition (Table 6.1). B. Ethanol production C. Acetate production.
Figure 6.3. Sensitivity test of kinetic model about syngas fermentation performance.

Cell growth (Figure 6.3A), alcohol production (Figure 6.3B), acid production (Figure 6.3C) was predicted under various k value, CO partial pressure, CO$_2$ partial pressure and H$_2$ partial pressure. Standard represented the standard condition, which is k=10hr$^{-1}$ and syngas content CO=0.25, CO$_2$=0.25 and H$_2$=0.25.

6.3.3 Machine Learning model

Machine learning methods have proved useful in building predictive models for processes in which fundamental mechanistic understanding is limited (Pappu and Gummadi, 2016). A neural network model with one hidden layer was used in this work to study syngas fermentation process. The model was trained with CO, CO$_2$ and H$_2$ fraction, flowrates and time series as the inputs and ethanol, butanol, acetic acid, butyric acid and biomass as the outputs. In this study, 70% of experimental data was randomly split for training the network, 15% was split for network validation, 15% was split to test the accuracy of the network. Training data and test data were scaled as follows:
\[ X_{Norm} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \]  \hspace{1cm} (6.4)

Where \( X_{Norm} \) represents normalized value of each variable, \( X \) is the original value of the each variable before normalization. \( X_{\text{max}} \) and \( X_{\text{min}} \) is the maximum and minimum value of each original variable. Thus, all the variables were scaled between 0~1. 1~50 neurons in the hidden layer were applied to determine the optimal network architecture. The model accuracy of different neuron size was evaluated by mean square error between the simulation and experimental data. As seen in Figure 6.4, 27 neurons were the optimum number of neurons for the neural network trained (it gave the least mean square error). After training model, all the inputs and outputs were re-scaled back. Simulation results showed that the model fitted the fermentation data well (Figure 6.5). To verify the model applicability, previous reported experimental data (Wan et al., 2017) was used and further modeling tests (Figure 6.6) demonstrated same good fitting.

**Mean Square Error**

![Mean Square Error Chart](chart.png)

**Figure 6.4. Mean Square Error of Machine Learning model.**

1~50 neurons were used to test the accuracy of the machine learning model, mean square error of each size of neurons were shown.
Figure 6.5. Comparison of machine learning data fitting with experimental data.

Experimental data for biomass (OD, green dots), ethanol production (g/L, purple dots), acetate production (g/L, black dots), butanol (g/L, red dots) and butyric acid (g/L, blue dots) and model simulation for biomass (OD, green lines), ethanol production (g/L, purple lines), acetate production (g/L, black lines), butanol (g/L, red lines) and butyric acid (g/L, blue lines) was compared. Biological error bar was included in the figure.
Figure 6.6. Comparison of machine learning predictions with previous study experimental data.

Experimental data for biomass (OD, green dots), ethanol production (g/L, purple dots), acetate production (g/L, black dots), butanol (red dots) and butyric acid (blue dots) and model simulation for biomass (OD, green lines), ethanol production (g/L, purple lines), acetate production (g/L, black lines), butanol (red lines) and butyric acid (blue lines) was compared. Biological error bar was included in the figure.

Next, sensitivity tests of CO, CO₂, H₂ gas contents and flow rates were performed using the trained neural network model. For comparison, all input factors and output responses (i.e., fermentation productions) were scaled in the range of 0~1. Simulation demonstrates that the increase of syngas partial pressures promoted fermentations, while excessive CO or CO₂ in syngas could limit biomass growth (Figure 6.7A). Interestingly, there was an optimal flowrate for alcohol production (see * in Figure 6.7A(b)). This can be explained that low flowrates increased mass
transport limitation and high flowrates increased the loss of alcohol from off-gas. Moreover, acetate and butyric acid production was boosted by CO and CO$_2$, but inhibited by excessive H$_2$ and flowrates. The model prediction for butanol was poor because of the relatively small training data size. The trained neural network was also applied on previous reports from other research group (Devarapalli et al., 2017). Comparison of model prediction and experimental data from previous study was shown in Figure 6.7B. In general, the neural network model can capture influential process factors and predict the other study efficiently. Prediction from 0~5 days highly was in accordance with the experimental data. While from 6~8 days, the difference was observed between the experimental data and model prediction. Since our model mainly focused on the first five-days fermentation, the prediction result can still be regarded as reasonably good. The discrepancy of the model prediction can be explained by differences in bioreactor system and operations. Adding new inputs, extending the fermentation duration and increasing the size of datasets in the neural network would improve its accuracy and applicability.
Figure 6.7. Analysis of neural network model performance.

A. Sensitivity test of neural network model. Sensitivity test was conducted by varying the input CO, CO₂, H₂ and flow rates based on the well-trained machine learning model. The fraction of CO, CO₂ and H₂ was from 0~0.4 and flowrates were 10 min/hr or 20 min/hr. These variables were scaled in the range of 0~1. The scaled changes of (a) biomass (b) Ethanol (c) Acetate (d) Butanol and (e) Butyric acid were shown. The asterisk represented the optimal flowrate for alcohol production. B. Artificial Neural Network model prediction of previous study. Orange line represented the artificial neuron network prediction. Black dots represented
the estimated experimental data from previous study. 0.4g/L cell dry weight per OD$_{600}$ was used in this study.

6.4 Conclusion

This study revealed the effect of biomass growth and product distribution with various syngas conditions (flowrate and compositions) on P7 fermentation performances. FBA model predicted optimal cell growth and metabolites productions constrained by CO/CO$_2$/H$_2$. Kinetic model efficiently describes the general tendency of biomass growth and bioproducts accumulation. Considering the limitation in mechanistic models, the neural network model was used to accurately describe the functional relationship between the input elements (CO, CO$_2$, H$_2$, flowrates and time slots) and the process responses (ethanol, acetate, butanol, butyric acid and biomass). These modeling predictions gave us holistic understandings of syngas fermentation and scale-up under complex influential process variables.
Chapter 7. Investigation of *Pseudomonas aeruginosa* biofilm metabolism using $^{13}\text{C}$ tracing
Abstract

This study used $^{13}$C-assisted tracing, microscopy image, and transcription analysis to reveal metabolic features in biofilm of *Pseudomonas aeruginosa* PAO1. First, we grew PAO1 in a steady-state biofilm reactor with labeled glucose and the labeling of proteinogenic amino acids was used to quantify biofilm metabolism during its growth (i.e., $^{13}$C-metabolic flux analysis). We found that the biofilm showed similar fluxome topology to planktonic cells cultured from shake flasks (biofilm cell showed slightly lower respiration rates). Particularly, PAO1 glucose catabolism operated via two major routes: an Entner-Doudoroff-Embden-Meyerhof-Parnas loop and the TCA cycle. Besides, the difference between PAO1 and its biofilm overproducing mutant was also not observable from $^{13}$C-labeling of their proteinogenic amino acids. On the other hand, dynamic labeling of matured biofilm cells indicated that the majority of biofilm cells (>90%) could be metabolic active but their glucose utilization rate was much slower than planktonic cells (turnover time of intracellular glucose-6-P: ~3hrs vs. 5 minutes), possibly due to repressed glycolytic genes (such as *pgi* and *fbp*). In light of physiologically heterogeneous structures, this study suggested those actively-dividing biofilm cells operated identical metabolisms to their planktonic counterparts.

7.1 Introduction

Bacterial biofilm can cause severe environmental pollutions (Bachmann and Edyvean 2005), metal corrosions, and human diseases (Hall-Stoodley et al. 2004). Biofilm formations and structures are highly heterogeneous (O'Toole et al. 2000). Biofilm cells at spatial locations may differ in physiologies (e.g., distinct transcriptomic and proteomic profiles) due to nutrient and oxygen concentration gradients (Williamson et al. 2012). Moreover, extracellular polymer substances (EPS), composed of complex polysaccharides and proteins, fills biofilm matrix and
plays an important role in cell growth and resistances (Mathee et al. 1999). To decipher biofilm features, traditional methods were applied, including cell staining (e.g., crystal violet assay), genetics (PCR/qRT-PCR), protein assays, and structural imaging (e.g., SEM, TEM and x-ray microscopy) (Pantanella 2013). Bacterial mutants are also used to reveal regulatory mechanisms of biofilm cell surviving under aerobic and anaerobic conditions (Ding et al. 2014; Zhang et al. 2014a). However, there is still a lack of knowledge of cell physiologies at metabolic fluxome levels, which focuses on pathway usage and activities for nutrient utilizations and biosynthesis. In this study, *Pseudomonas aeruginosa* PAO1 were used as a model biofilm formation strain to provide new insights into the functions of central metabolic pathways for synthesis of biofilm biomass. Via $^{13}$C labeling of proteinogenic amino acids and pseudo-steady state biofilm reactors, we traced metabolisms of these active-growing biofilm cells (including PAO1 cell and biofilm-overproducing PAO1 mutant). Moreover, dynamic labeling approaches investigated overall glucose catabolic rates in these metabolically-active biofilm cells. The outcome from this study improves our understandings of central pathways in growing biofilm cells.

7.2 Material and Methods

7.2.1 Strain Construction

*YedQ* (previously known as *yhcK*) gene from *E. coli* is cloned from the *pYedQ* plasmid in HindIII/BamHI side of pBBR1MCS-5 plasmid vector to make the pYedQ2 plasmid.

The *S. oneidensis* MR-1 and *P. aeruginosa* PAO1 strains (i.e. MR-1 pYedQ2 and PAO1 pYedQ2) that overexpress c-di-GMP was constructed via plate mating as follows: *S. oneidensis* MR-1 / *P. aeruginosa* PAO1, *Escherichia coli* containing *RK600*, *Escherichia coli* -DH5α- pYedQ2 were grown using LB media overnight in shaking incubator (200 rpm) respectively. Antibiotics such as
chloramphenicol and gentamicin were used in cell culturing when necessary. 1 ml of the culture from each of the three strains was washed two times with fresh LB medium, using centrifugation at 10000 x g for 3 minutes. The cell pellets from each of the three strains was resuspended and mixed at 1:1:1 ratio in 1 ml of LB media. Centrifugation was done to obtain a cell pellet containing cells from all three strains in equal proportions. The cell pellet was transferred to a LB agar plate and incubated in 30 °C for 5 hours. The cells were harvested from the plate via washing with 1 ml of 0.9% sodium chloride solution (saline). 100 ul of cells were spread on a LB agar plate containing 20 ug/ml tellurite and 60 ug/ml gentamicin, and kept in incubator at 30 °C until transposon mutant colonies appeared. Gel electrophoresis was run with the plasmid extracted from the selected mutant as a confirmation that the selected mutant indeed contains the pYedQ₂ plasmid. Swim plate experiments were done in LB agar plates. All swim plates experiments were conducted with five replicates. LCMS was used to verify the increase in intracellular c-di-GMP levels in MR-1 pYedQ₂ and PAO1 pYedQ₂ as compared to MR-1 wild type and PAO1 wild type respectively.

7.2.2 Strain, medium and culture preparation.

In tubing reactor system, *E. coli*, *S. oneidensis* MR-1 and *P. aeruginosa* PAO1 used in this study are described in Table 7.1. Stock cultures were maintained in LB medium with 20% glycerol at -80°C. *S. oneidensis* MR-1 was grown in modified M1 medium (Cao et al. 2011), while *E. coli* and *P. aeruginosa* PAO1 were grown in M9 medium (Klaus et al. 2005). For *S. oneidensis* MR-1 growth in modified M1 medium, 20 mM 13C-labelled sodium lactate was used as an electron donor and the sole carbon source. For *E. coli* and *P. aeruginosa* PAO1 growth in M9 medium, 20 mM 13C-labelled glucose was used as an electron donor and the sole carbon source.
Table 7.1. Strains and plasmids used in this study.

<table>
<thead>
<tr>
<th>Strains / Plasmid</th>
<th>Description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>S. oneidensis</em> MR-1</td>
<td>Manganese-reducing strain</td>
<td>(Myers and Nealson 1988)</td>
</tr>
<tr>
<td><em>E. coli</em> DH5α</td>
<td>Laboratory strain</td>
<td>(Zhang et al. 2014b)</td>
</tr>
<tr>
<td><em>P. aeruginosa</em> PAO1</td>
<td>Laboratory strain</td>
<td>(Stover et al. 2000)</td>
</tr>
<tr>
<td><em>pYedQ</em></td>
<td>Gm(^r); pBBR1MCS3 carrying the YedQ gene</td>
<td>(Gjermansen et al. 2006)</td>
</tr>
<tr>
<td><em>pRK600</em></td>
<td>Cm(^r); oriColE1 RK2-Mob(^+) RK2-Tra(^+); helper vector for conjugation</td>
<td>(Kessler et al. 1992)</td>
</tr>
</tbody>
</table>

MR-1 wild type and MR-1 pYedQ\(_2\) were cultured in modified M1 medium (20 ml) in a flask (100 ml) containing 3-\(^{13}\)C sodium lactate at room temperature with shaking (200 rpm). PAO1 wild type and PAO1 pYedQ\(_2\) were cultured in M9 medium (20 ml) in a flask (100 ml) containing 1,2-\(^{13}\)C-labelled glucose at room temperature with shaking (200 rpm). 2 replicates were done for each of the strains. Cell pellets for each strain were washed two times using 0.9% NaCl solution. Both cell pellets and supernatant were collected for further analysis.

PAO1 and MR-1 biofilm were grown in either tubular biofilm reactor (25 cm of 16” silicon tubing) or glass slides using M9 medium containing 20 mM 1,2-\(^{13}\)C-labelled glucose. In tubular biofilm reactors, the respective media were continuously supplied by Ismatec IPC-N (ISM936C) Low-Speed Digital Peristaltic Pump system (Cole-Parmer, Singapore) at a flow rate of 8 ml/h to the tubular biofilm reactor (Figure 7.1A). The flow system was assembled and prepared as described previously (Sternberg and Tolker-Nielsen 2006). Each tubular reactor was inoculated by injecting 2 ml overnight culture (diluted to an OD\(_{600}\) of 0.01) using a syringe. After inoculation,
the medium flow was stopped for 1 h to allow initial attachment followed by continuous media flow with a flow rate of 8 ml/h. After 4 days of growth, the biofilm samples were harvested for further analysis. In glass slides system, *Pseudomonas aeruginosa* PAO1 and PDO300 was inoculated in M9 medium with 1,2-$^{13}$C glucose overnight to attach on glass slides (Figure 7.1B).

**7.2.3 Dynamic labeling in petri dish system**

To snapshot cell catabolism, dynamic labeling experiments were performed for both planktonic cells and biofilm cells. For planktonic cultures, *P. aeruginosa* PAO1 was inoculated in M9 medium in a shake flask containing 1g/L no-labeled glucose. Once the cells reached late exponential phase (OD$_{600}$ ~0.8 and glucose was close to be used up), fully labeled-$^{13}$C glucose was pulsed into culture (final concentration of 2 g/L $^{13}$C-glucose). Right after $^{13}$C-glucose additions, 15 mL of cell culture at four different sampling points (10s, 1 min, 5 min, and 30 min) was harvested by mixing culture with 5 mL pre-cold M9-ice solutions. Then the samples (n=2) were furtherly quenched with ethanol-dry ice bath (-70 °C). After the culture temperature was dropped to ~0 °C, biomass was centrifuged at 8000 rpm for 1 min (all samples contained ice particles to maintain near 0 °C). The resulting pellets were kept at -80 °C for LC-MS measurement.

To measure catabolic rate in PAO1 biofilm, we used glass slides to grow biofilm because these slides can be quickly harvested and quenched at low temperature. Specifically, glass slides (~1cm$^2$) were placed inside of the petri dishes containing M9 medium (1g/L no-labeled glucose) and *P. aeruginosa* cells (37 °C). After overnight incubations, the glass slides were washed gently using PBS solutions then transferred to fresh M9 medium with 1 g/L no-labeled glucose. The glass slides were cultivated for another 48 hours, and new M9 medium contained 1 g/L no-labeled glucose was used every 24 hours. For dynamic labeling experiments, biofilm samples were washed in 0.9%NaCL solution gently and glass slides with biofilm were transferred and soaked into 25
mL M9 medium contained 1 g/L $^{13}$C fully labeled glucose (for 10s, 1 min, 5 min, 30 min and 3h). Then glass slides were placed in 0.9% NaCL water-ice mixture at different sampling time points to quench cell metabolisms and these slides were stored on the top of dry ice before LC-MS measurements.

Figure 7.1. Reactor platform.
(A) Schematic of tubular biofilm reactor. (B) Schematic of dynamic labeling in petri dish system.

7.2.4 Metabolic flux analysis.

$^{13}$C-metbaolic flux analysis (MFA) of P. aeruginosa biofilm cells was performed based on isotopomer data from proteinogenic amino acids. The software WUflux (He et al. 2016) was used for flux calculations. The MFA model included glycolysis, the OPP pathway, the ED pathway, the
TCA cycle, the glyoxylate shunt, and biomass synthesis. Based on KEGG database, the species lacks phosphofructokinase and thus the reaction (F6P \to FBP) was deleted from the model. $^{13}\text{C}$-MFA profiled relative fluxes by normalizing the glucose uptake rate as 100 units. The relative fluxes were solved by minimizing a quadratic error function that calculated the differences between predicted and measured isotopomer patterns.

7.2.5 Proteinogenic amino acids isotopomer analysis.

The incorporation of labeled carbons into proteinogenic amino acids was measured by GC-MS followed previous protocol (You et al. 2012)(Hewlett-Packard model 7890A, Agilent Technologies) and later used in $^{13}\text{C}$-MFA. Briefly, biomass pellets were hydrolyzed by 6 M HCl at 100°C, then air-dried and derivatized with $\text{N}$-tert-butyldimethylsilyl-$\text{N}$-methyltrifluoroacetamide (TBDMS) prior to GC-MS analysis. A published software was used to correct amino acid MS data (fragments of [M-57]$^+$, [M-159]$^+$ or [M-85]$^+$, and [f302]) (Wahl et al. 2004). Isotopomer labeling fractions (M0, M1, M2, etc.) represent fragments with 0,1,2..labeled carbons in amino acids. Due to overlapping peaks or product degradation, several amino acids (proline, arginine, cysteine, and tryptophan) were not used for flux analysis (Antoniewicz et al. 2007). For measurement of free metabolite labeling, non-labeled culture was pulsed with [U-$^{13}\text{C}$] glucose. Cells were harvested at the certain sampling points, then the incorporation of labeled carbons into free metabolites were traced by LC-MS followed previous protocol (Hollinshead et al. 2016). Briefly, cell pellet was suspended in 1 mL methanol/chloroform solution (7:3 v/v) and shake at 150 rpm at 4 °C overnight. DI water was added to the cell-solvent mix to extract cell metabolites. The aqueous phase was filtered through an Amicon Ultra centrifuge filter (3000 Da; EMD Millipore, Billerica, MA), lyophilized, and dissolved in acetonitrile and water (6:4, v/v) solution for LC-MS measurement (Agilent Technologies 1200 Series equipped with a SeQuant
Zic-pHILIC column). MS distributions of the metabolite were determined based on the ratio of the chosen isotopomer integrated peak area to the sum of all isotopomers integrated peak areas.

7.2.6 Free metabolites isotopomer analysis.

*Pseudomonas aeruginosa* was pulsed with [U-13C] glucose. Cells were harvested at the certain sampling points, then the incorporation of labeled carbons into free metabolites were traced by LC-MS followed previous protocol (Hollinshead et al. 2016b). Briefly, cell pellet was suspended in 1 mL methanol/chloroform solution (7:3 v/v) and shake at 150 rpm at 4 °C overnight. DI water was added to the cell-solvent mix to extract cell metabolites. The aqueous phase was filtered through an Amicon Ultra centrifuge filter (3000 Da; EMD Millipore, Billerica, MA), lyophilized, and dissolved in acetonitrile and water (6:4, v/v) solution for LC-MS measurement (Agilent Technologies 1200 Series equipped with a SeQuant Zic-pHILIC column). MS distributions of the metabolite were determined based on the ratio of the chosen isotopomer integrated peak area to the sum of all isotopomers integrated peak areas.

7.3 Results and Discussion

7.3.1 Fluxomes of *Pseudomonas aeruginosa* as planktonic and biofilm.

To examine the difference of *Pseudomonas aeruginosa*’s planktonic and biofilm metabolism, [1,2-13C] labelled glucose was used in silico tube bioreactor for 13C metabolic flux analysis (13C-MFA). The method of model fitting to experimental data was described in previous study (He et al. 2016). The relative flux distributions and standard deviations are shown in Figure 7.2 (left and upper value) for planktonic and (right and lower value) for biofilm, while exchange coefficients for reversible reactions and 95% confidence interval were also described by previous study (He et al. 2016). Compared with planktonic, most flux values in biofilm cells were differed within 10
units. In general, glucose utilization formed a strong carbohydrate degradation loop, Entner-Doudoroff-Embden-Meyerhof-Parnas (EDEMP) cycle (G6P → 6PG → GAP → F6P → G6P).

Similar result for this strong carbohydrate degradation EDEMP cycle has been recently observed in *Pseudomonas putida* (Nikel et al. 2015). *Pseudomonas* is famous for its ED pathway with minor flux through pentose phosphate pathway and glycolysis (Berger et al. 2014)(Fuhrer et al. 2005). Even though ED pathway cannot be beneficial to ATP generations, it reduced enzymes synthesis cost comparing to the EMP pathway (Stettner and Segrè 2013). On the other hand, the formation of EDEMP cycle in *Pseudomonas* benefits NADPH generations as well as the biosynthesis of F6P, which is the precursor of extracellular polysaccharides (e.g., alginate) as a part of EPS. Similar flux in EDEMP cycle but substantial differences in pentose phosphate pathway and TCA cycle was revealed later (Opperman and Shachar-Hill 2016). Comparing to biofilm cells, planktonic cells showed slightly higher flux through EDEMP cycle and the TCA cycle. The higher fluxes through oxidative cycles could generate more reducing cofactors NAD(P)H to satisfy cell respirations under higher oxygen conditions in shake flasks. In summary, mass transfer limitation might play an important role to affect the planktonic metabolism and biofilm metabolism and biofilm cell is relatively less metabolically active. However, fluxomes showed these actively growing biofilm cells have identical fluxomic topology. These cells locate on the surface of biofilm and thus receive nutrients similar to planktonic cells.
Figure 7.2. Fluxome of *Pseudomonas aeruginosa* as planktonic (left and upper side) and biofilm (right and lower side).

The fluxes were normalized to the glucose uptake rate (represented as 100), and the fluxes are represented as ‘best fit ± standard deviation’. The arrow thickness relates to the degree of flux.
The white arrows represent the fluxes towards biomass synthesis. Abbreviation: 3-phosphoglycerate (3PG); 6-phosphogluconate (6PG); acetyl-CoA (AceCoA); dihydroxyacetone phosphate (DHAP); erythrose 4-phosphate (E4P); fructose 1,6-bisphosphate (FBP); fructose 6-phosphate (F6P); glucose 6-phosphate (G6P); glyceraldehyde 3-phosphate (GAP); glyoxylate (GLX); isocitrate (ICT); malate (MAL); oxaloacetate (OAA); phosphoenolpyruvate (PEP); pyruvate (PYR); ribose 5-phosphate (R5P); ribulose-5-phosphate (Ru5P); ribulose-1,5-diphosphate (RuBP); sedoheptulose-7-phosphate (S7P); succinate (SUC); and xylulose-5-phosphate (X5P).

7.3.2 Variability of central catabolic fluxomes under planktonic and biofilm conditions.

*Pseudomonas aeruginosa* mutant can over-produce EPS that enhanced biofilm formation rates (note: more EPS production caused less biomass yield from glucose). We examined metabolisms for higher biofilm synthesis by comparing amino acid labeling among biofilm and planktonic cells (both wild type and mutant) after they grew with [1,2-13C] labelled glucose in tubing reactor and shake flask, respectively. To broaden our understanding, we also performed similar experiments for *Shewanella oneidensis* MR1 and its biofilm over-producing mutant. Figure 7.3 indicates the difference of proteinogenic amino acid labeling between mutant and wild type from planktonic cultures and biofilm cells was small. The R square of labeling data was >0.99 (within measurement errors). Since the labeling of proteinogenic amino acids mainly records cell metabolism for active biomass synthesis, the similarity of amino acid labeling from different cultues (planktonic cells and biofilm cells) indicates similar fluxomic topologies. Thereby, our data also explained no detectable differences in growth rate bewteen PAO1 and biofilm overproducing PDO in shake flask conditions (Mathee et al. 1999). Besides, these observations can be found in
both *Shewannella oneidensis* and *Pseudomonas aeruginosa* cultures, which suggested the metabolism of growing cells in tubular biofilm reactor was similar to these planktonic cells.

![Diagram](image)

**Figure 7.3. Comparison of amino acids labeling between wild type and EPS-overexpressing mutant.**

**(A) Shewannella oneidensis** planktonic cells and biofilm cells were grown with $3.13C$ lactate in tubing reactor for 72 hrs. **(B) Pseudomonas aeruginosa** planktonic cells and biofilm cells were grown with $1,2.13C$ glucose in tubing reactor for 72 hours. MID represents mass isotopomer distribution, which showed the fraction of labeling M0, M1, M2...(MS fragments with 0, 1, 2 labeled carbons) in each amino acid.

7.3.3 *Dynamic labeling for free metabolites of Pseudomonas aeruginosa as planktonic and biofilm*

Figure 7.4 showed PAO1 free metabolite labeling rates after $[U.13C]$ glucose was pulsed in both biofilm cells and planktonic cells. Results (labeling rates of G6P, first metabolic node after glucose uptake) showed that planktonic cells have much faster response to glucose than biofilm
cells, which proved that planktonic cells are more active than biofilm cells. Compared the final sampling point in biofilm wild type and planktonic cells, labeling of G6P in biofilm cells also reached ~85% (close to the natural labeling abundance, ~90%), which means majority of biofilm cells were metabolic active. Moreover, labeling of free glutamate (downstream products from the TCA cycle) for both planktonic cells and biofilm cells were much slower than that of G6P, which indicated the anabolism was slower than cell catabolism. These studies further indicated overall biofilm cells have slower catabolic rates. In the light of MFA results, we can deduce that majority biofilm cells are active, but with much slower glucose catabolism. On the other hand, a small fraction of biofilm cells may grow as fast as planktonic cells (possibly these surface cells) and these fast growing cells shared same metabolism as planktonic cells.

Figure 7.4. Dynamic labeling for biofilm (A) wild type and (B) planktonic cells.

7.4 Conclusion

This study is first to elucidate insights into metabolic features and growth pattern of biofilm cells and planktonic cells via $^{13}$C labeling. Bacteria within biofilms differ in cell physiology because of nutrients and oxygen diffusion limitations. This study demonstrated that fast growing
biofilm cells (i.e., at the top of the biofilms) may share same metabolic topology to planktonic cells. This finding explains that the dividing cells had different metabolism from these slow-growing cells deep in the biofilms. Due to its faster metabolism than dormant biofilm cells, the dividing cells are more susceptible to killing by environmental stresses or antibiotics. These knowledge bridges the gap between cell metabolisms and biofilm formation, and offers broad impact on biofilm study.
Supporting Information for Chapter 7

Experimental data

Table 7-S1

Figure 7-S1

**Table S1.** Simulated and measured mass isotopomer distributions (MID) of proteinogenic amino acids from PA01 biofilm and planktonic cultures. Cultures were grown on 1,2-\(^{13}\)C glucose under dark heterotrophic conditions.

<table>
<thead>
<tr>
<th></th>
<th>PA01 WT P</th>
<th>PA01 PYEDQ P</th>
<th>PA01 PYHJH P</th>
<th>PA01 WT B</th>
<th>PA01 PYEDQ B</th>
<th>PA01 PYHJH B</th>
</tr>
</thead>
<tbody>
<tr>
<td>alanine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-57</td>
<td>M0 0.4</td>
<td>M1 0.13</td>
<td>M2 0.47</td>
<td>M3 0.01</td>
<td>M0 0.4</td>
<td>M1 0.13</td>
</tr>
<tr>
<td></td>
<td>M2 0.39</td>
<td>M1 0.13</td>
<td>M2 0.48</td>
<td>M3 0.01</td>
<td>M2 0.39</td>
<td>M1 0.13</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M-159</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0 0.45</td>
<td>M1 0.52</td>
<td>M2 0.03</td>
<td></td>
<td>M0 0.45</td>
<td>M1 0.52</td>
</tr>
<tr>
<td></td>
<td>M2 0.44</td>
<td>M1 0.53</td>
<td>M2 0.03</td>
<td></td>
<td>M2 0.44</td>
<td>M1 0.53</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glycine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-57</td>
<td>M0 0.93</td>
<td>M1 0.06</td>
<td>M2 0.01</td>
<td></td>
<td>M0 0.93</td>
<td>M1 0.06</td>
</tr>
<tr>
<td></td>
<td>M2 0.92</td>
<td>M1 0.06</td>
<td>M2 0.01</td>
<td></td>
<td>M2 0.92</td>
<td>M1 0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M-85</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0 0.98</td>
<td>M1 0.02</td>
<td>M2 0.01</td>
<td></td>
<td>M0 0.98</td>
<td>M1 0.02</td>
</tr>
<tr>
<td></td>
<td>M2 0.97</td>
<td>M1 0.03</td>
<td>M2 0.01</td>
<td></td>
<td>M2 0.97</td>
<td>M1 0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>valine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-57</td>
<td>M0 0.17</td>
<td>M1 0.27</td>
<td>M2 0.28</td>
<td>M3 0.27</td>
<td>M4 0.01</td>
<td>M5 0</td>
</tr>
<tr>
<td></td>
<td>M2 0.17</td>
<td>M1 0.27</td>
<td>M2 0.29</td>
<td>M3 0.27</td>
<td>M4 0.01</td>
<td>M5 0</td>
</tr>
<tr>
<td></td>
<td>M3 0.15</td>
<td>M2 0.26</td>
<td>M2 0.28</td>
<td>M3 0.3</td>
<td>M4 0.01</td>
<td>M5 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M-159</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0 0.19</td>
<td>M1 0.47</td>
<td>M2 0.31</td>
<td>M3 0.02</td>
<td>M4 0</td>
<td>M5 0</td>
</tr>
<tr>
<td></td>
<td>M1 0.19</td>
<td>M2 0.47</td>
<td>M2 0.32</td>
<td>M3 0.02</td>
<td>M1 0.19</td>
<td>M2 0.47</td>
</tr>
<tr>
<td></td>
<td>M2 0.17</td>
<td>M1 0.47</td>
<td>M2 0.33</td>
<td>M3 0.02</td>
<td>M2 0.17</td>
<td>M1 0.47</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>leucine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M-15</td>
<td>M0 0.07</td>
<td>M1 0.28</td>
<td>M2 0.41</td>
<td>M3 0.21</td>
<td>M4 0.02</td>
<td>M5 0</td>
</tr>
<tr>
<td></td>
<td>M1 0.07</td>
<td>M2 0.28</td>
<td>M2 0.41</td>
<td>M3 0.21</td>
<td>M4 0.02</td>
<td>M5 0</td>
</tr>
<tr>
<td></td>
<td>M2 0.07</td>
<td>M1 0.28</td>
<td>M2 0.41</td>
<td>M3 0.21</td>
<td>M4 0.02</td>
<td>M5 0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

114
<table>
<thead>
<tr>
<th></th>
<th>M-159</th>
<th>M0</th>
<th>0.16</th>
<th>0.16</th>
<th>0.16</th>
<th>0.13</th>
<th>0.13</th>
<th>0.14</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>0.45</td>
<td>0.45</td>
<td>0.45</td>
<td>0.43</td>
<td>0.44</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>0.35</td>
<td>0.34</td>
<td>0.35</td>
<td>0.4</td>
<td>0.39</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M4</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>isoleucine</th>
<th>M-15</th>
<th>M0</th>
<th>0.1</th>
<th>0.1</th>
<th>0.1</th>
<th>0.08</th>
<th>0.09</th>
<th>0.1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>0.27</td>
<td>0.26</td>
<td>0.28</td>
<td>0.23</td>
<td>0.24</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>0.34</td>
<td>0.32</td>
<td>0.32</td>
<td>0.31</td>
<td>0.33</td>
<td>0.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>0.23</td>
<td>0.22</td>
<td>0.23</td>
<td>0.27</td>
<td>0.25</td>
<td>0.24</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M4</td>
<td>0.06</td>
<td>0.08</td>
<td>0.08</td>
<td>0.11</td>
<td>0.09</td>
<td>0.08</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M6</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M-57</th>
<th>M0</th>
<th>0.14</th>
<th>0.13</th>
<th>0.13</th>
<th>0.13</th>
<th>0.1</th>
<th>0.2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.35</td>
<td>0.27</td>
<td>0.28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>0.36</td>
<td>0.36</td>
<td>0.36</td>
<td>0.4</td>
<td>0.31</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>0.15</td>
<td>0.17</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M-159</th>
<th>M0</th>
<th>0.94</th>
<th>0.94</th>
<th>0.95</th>
<th>0.92</th>
<th>0.94</th>
<th>0.93</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>0.06</td>
<td>0.05</td>
<td>0.04</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M-57</th>
<th>M0</th>
<th>0.97</th>
<th>0.96</th>
<th>0.97</th>
<th>0.96</th>
<th>0.96</th>
<th>0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>0.03</td>
<td>0.04</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M-159</th>
<th>M0</th>
<th>0.22</th>
<th>0.23</th>
<th>0.22</th>
<th>0.19</th>
<th>0.2</th>
<th>0.21</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>0.35</td>
<td>0.32</td>
<td>0.35</td>
<td>0.31</td>
<td>0.32</td>
<td>0.34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>0.29</td>
<td>0.33</td>
<td>0.3</td>
<td>0.34</td>
<td>0.32</td>
<td>0.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>0.13</td>
<td>0.13</td>
<td>0.13</td>
<td>0.16</td>
<td>0.16</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.14</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M-57</th>
<th>M0</th>
<th>0.33</th>
<th>0.28</th>
<th>0.31</th>
<th>0.27</th>
<th>0.28</th>
<th>0.28</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>0.45</td>
<td>0.52</td>
<td>0.48</td>
<td>0.49</td>
<td>0.5</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>0.21</td>
<td>0.19</td>
<td>0.21</td>
<td>0.22</td>
<td>0.23</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>0</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>M-159</th>
<th>M0</th>
<th>0.81</th>
<th>0.76</th>
<th>0.8</th>
<th>0.73</th>
<th>0.74</th>
<th>0.77</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>0.15</td>
<td>0.17</td>
<td>0.15</td>
<td>0.18</td>
<td>0.17</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>0.03</td>
<td>0.05</td>
<td>0.04</td>
<td>0.07</td>
<td>0.07</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>0</td>
<td>0.01</td>
<td>0</td>
<td>0.02</td>
<td>0.01</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M9</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>phenylalanine</th>
<th>M-57</th>
<th>M0</th>
<th>0.82</th>
<th>0.78</th>
<th>0.81</th>
<th>0.74</th>
<th>0.76</th>
<th>0.78</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M1</td>
<td>0.16</td>
<td>0.19</td>
<td>0.17</td>
<td>0.22</td>
<td>0.21</td>
<td>0.19</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M2</td>
<td>0.02</td>
<td>0.03</td>
<td>0.02</td>
<td>0.04</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M8</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>

115
<table>
<thead>
<tr>
<th></th>
<th>M-57</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>M-57</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M5</td>
<td>M0</td>
<td>M1</td>
</tr>
<tr>
<td>aspartate</td>
<td>0.23</td>
<td>0.34</td>
<td>0.31</td>
<td>0.12</td>
<td>0</td>
<td>0.31</td>
<td>0.5</td>
<td>0.2</td>
<td>0.01</td>
<td>0</td>
<td>0.36</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>M159</td>
<td>0.23</td>
<td>0.34</td>
<td>0.31</td>
<td>0.12</td>
<td>0</td>
<td>0.31</td>
<td>0.5</td>
<td>0.2</td>
<td>0.01</td>
<td>0</td>
<td>0.36</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>glutamate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M5</td>
<td>M0</td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td>0.14</td>
<td>0.39</td>
<td>0.36</td>
<td>0.11</td>
<td>0</td>
<td>0.14</td>
<td>0.39</td>
<td>0.36</td>
<td>0.11</td>
<td>0</td>
<td>0.36</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td>lysine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M5</td>
<td>M0</td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td>0.13</td>
<td>0.26</td>
<td>0.29</td>
<td>0.24</td>
<td>0.09</td>
<td>0.12</td>
<td>0.24</td>
<td>0.31</td>
<td>0.24</td>
<td>0.08</td>
<td>0.11</td>
<td>0.12</td>
<td>0.11</td>
</tr>
<tr>
<td>histidine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M5</td>
<td>M0</td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td>0.81</td>
<td>0.09</td>
<td>0.08</td>
<td>0.01</td>
<td>0</td>
<td>0.84</td>
<td>0.08</td>
<td>0.06</td>
<td>0.01</td>
<td>0</td>
<td>0.86</td>
<td>0.07</td>
<td>0.06</td>
</tr>
<tr>
<td>tyrosine</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M5</td>
<td>M0</td>
<td>M1</td>
</tr>
<tr>
<td></td>
<td>0.82</td>
<td>0.07</td>
<td>0.09</td>
<td>0.01</td>
<td>0</td>
<td>0.84</td>
<td>0.07</td>
<td>0.06</td>
<td>0.01</td>
<td>0</td>
<td>0.85</td>
<td>0.07</td>
<td>0.07</td>
</tr>
</tbody>
</table>

116
<table>
<thead>
<tr>
<th></th>
<th>M8</th>
<th>M9</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>M-159</td>
<td>M0</td>
<td>M1</td>
<td>M2</td>
<td>M3</td>
<td>M4</td>
<td>M5</td>
<td>M6</td>
<td>M7</td>
</tr>
<tr>
<td></td>
<td>0.81</td>
<td>0.16</td>
<td>0.02</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.2</td>
<td>0.03</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.16</td>
<td>0.03</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.72</td>
<td>0.21</td>
<td>0.04</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.2</td>
<td>0.03</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.76</td>
<td>0.18</td>
<td>0.03</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Figure 7-S1. Comparison of mass isotopomer distributions (MIDs) between experimental measurements and computational simulations. (A) Biofilm and (B) Planktonic cell. The data shown in the above figure are [M-57]+, [M-85]+ or [M-159]+, and f302 mass isotopomer distributions for eight amino acids. The fit between predicted and observed isotopomer distributions is reasonably good, with a sum of squared residuals of 0.015.
Chapter 8 Conclusions and Future Directions
8.1 Conclusions

In this dissertation, several metabolic modeling including genome scale flux balance analysis, $^{13}$C-metabolic flux analysis model, kinetic model and machine learning model were applied on non-conventional microbial systems. Cyanobacterial can grow photoautotrophically, photomixotrophically and heterotrophically. As revealed in Chapter 2 & 3, FBA model confirmed the mechanism about the photomixtrophic condition and photoautotrophic condition. $^{13}$C-MFA model revealed that pentose phosphate pathway is a good chassis for bioproducts, while the TCA cycle is weak in cyanobacteria under these three growth conditions. However, according to the experimental data, under the complete darkness condition, FBA model showed some limitation. After constrained by the experimental results and $^{13}$C-MFA model results, FBA model showed better performance.

Applying $^{13}$C-MFA model can also offer us the guide about how to improve the bioproducts yield and titer. As revealed in Chapter 4, the recent study suggested that energy metabolism is the rate limiting step for *E.coli* fatty acid production. This model results drive us to find a way to improve energy metabolism, and thereby enhancing the fatty acid production. VHB is a hemoglobin which is reported can improve the oxygen uptake, therefore boost ATP generation. Finally, we successfully engineered VHB into *E. coli* and improve fatty acid production. Furthermore, during the experiments, we also discovered that there exists tradeoff between the metabolic engineering and metabolic burden. FBA model was applied here to explain this phenomenon, the tradeoff between the production and non-growth associated maintenance.

In Chapter 5 and 6, syngas fermentation performance was investigated. The functional pathway was firstly deciphered. Wood-Ljungdahllii pathway can only utilize carbon dioxide but not bicarbonate. Furthermore, dynamic labeling experiments showed that Acetyl-coa is a rate
limiting step in the carbon metabolism. Moreover, three models were compared and simulated the syngas fermentation performance under various syngas compositions and flowrates. FBA model predicted the theoretical yield of bioproducts and biomass growth, suggested that energy metabolism is the limited steps for the biomass and bioproducts. CO can be both energy source and carbon source when H₂ is not sufficient. Kinetic model simulated the global trends of the syngas fermentation, described relationship between biomass, ethanol and acetate production with time series. Machine learning model caught the successfully simulated the experimental data and further applied on performance optimization.

In Chapter 7, the central carbon metabolism of biofilm and planktonic cell was decoded by ¹³C-MFA. A strong EDEMP pathway was discovered in pseudomonas PAO1, which can generate more ATP, the flux in biofilm cells is weaker than planktonic cell, which can be explained by the oxygen transfer limitation. qPCR work further confirmed the flux results. Furthermore, there was no apparent differences between EPS overproducing strain and wild type in proteinogenic amino acids labeling, which suggested that EPS overproducing only affects the secondary metabolism.

8.2 Future directions

In our study, FBA model has its limitations under unusual environmental conditions. We combined the experimental results and ¹³C-MFA results to constrain the FBA model and finally obtained the better results. In this case, more boundary constraints in FBA model need to be investigated by various experimental conditions.

Secondly, machine learning model was applied in syngas fermentation performance. We used the neural network model to trained the model and simulated the experimental data. While, to obtain broad application, experimental data from more various types bioreactors and more environmental conditions (e.g. pH, temperature, medium compositions…) need to be included.
Also, different types of machine learning method, support vector machine, k-nearest neighbor, random forest can be applied in the study. Furthermore, machine learning model can be applied on more biological systems, such as the relationship between genotype, environmental perturbation and phenotype. With the development of the metabolic modeling and machine learning, more and more biological systems could be described and the bottleneck of bacterial growth and bioproducts accumulation could be revealed. Therefore, it would offer the insights into the system optimization.
References


Appendix Chapter 1. Simulating cyanobacterial phenotypes by integrating flux balance analysis, kinetics, and a light distribution function
Simulating cyanobacterial phenotypes by integrating flux balance analysis, kinetics, and a light distribution function

Lian He¹, Stephen G. Wu¹, Ni Wan², Adrienne C. Reding³ and Yinjie J. Tang*⁴

Abstract

Background: Genome-scale models (GSMs) are widely used to predict cyanobacterial phenotypes in photobioreactors (PBRs). However, stoichiometric GSMs mainly focus on fluxome that result in maximal yields. Cyanobacterial metabolism is controlled by both intracellular enzymes and photobioreactor conditions. To connect both intracellular and extracellular information and achieve a better understanding of PBRs productivities, this study integrates a genome-scale metabolic model of Synechocystis 6803 with growth kinetics, cell movements, and a light distribution function. The hybrid platform not only maps flux dynamics in cells of sub-populations but also predicts overall production titer and rate in PBRs.

Results: Analysis of the integrated GSM demonstrates several results. First, cyanobacteria are capable of reaching high biomass concentration (>20 g/L in 21 days) in PBRs without light and CO₂ mass transfer limitations. Second, fluxome in a single cyanobacterium may show stochastic changes due to random cell movements in PBRs. Third, insufficient light due to cell self-shading can activate the oxidative pentose phosphate pathway in subpopulation cells. Fourth, the model indicates that the removal of glycogen synthesis pathway may not improve cyanobacterial bio-production in large-size PBRs, because glycogen can support cell growth in the dark zones. Based on experimental data, the integrated GSM estimates that Synechocystis 6803 in shake flask conditions has a photosynthesis efficiency of ~2.7 %.

Conclusions: The multiple-scale integrated GSM, which examines both intracellular and extracellular domains, can be used to predict production yield/rate/titer in large-size PBRs. More importantly, genetic engineering strategies predicted by a traditional GSM may work well only in optimal growth conditions. In contrast, the integrated GSM may reveal mutant physiologies in diverse bioreactor conditions, leading to the design of robust strains with high chances of success in industrial settings.

Keywords: Glycogen, Multiple-scale modeling, Photobioreactors, Photosynthesis efficiency, Self-shading, Synechocystis 6803

Background

In photobioreactors (PBRs), light penetration depth at high cell density can be as short as a few centimeters [1]. 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 CO₂ gas-liquid transfer. To enhance mass transfer, people often use CO₂-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, CO₂ supply, temperature, and geometry of PBRs [2–6]. 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. 13C-MFA measures carbon fluxes through the central metabolism via 13C labeling experiments. Alternatively, genome-scale flux balance analysis (FBA) can generate a holistic intracellular flux distribution map [7] owing to its extended coverage of genomic information [8]. Computational platforms, such as COBRA [9] and OptForce [10], 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 [11]. 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 [12–14]. To predict cyanobacterial growth and metabolic flux phenotypes in PBR settings, we integrated a genome-scale cyanobacteria model, iJN678 [15], with growth kinetics, cell movements based on reported PBR hydrodynamics, and a heterogeneous light distribution (Fig. 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.

**Results**

**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 μE/m²/s. Although the maximal photosynthetic efficiency in photosynthetic species can reach 4.6–6 % [16], not all incident radiation in PBRs can be efficiently used by cyanobacteria, thereby resulting in a lower conversion efficiency [17]. Hence, we chose a photosynthesis efficiency of 1.5 %, which was within a reasonable range of actual photosynthesis efficiencies of microalgae [18]. Based on a previous study, the mass transfer rate of CO₂ was assumed to be 10 h⁻¹ [19]. Under such a condition, cyanobacterial biomass concentration could increase from 0.1 to 5 g/L in 3 weeks, provided that other mineral nutrients are supplied continuously (Fig. 2a and Additional file 2: Fig. S1). The modelling results also showed continuous decreases in the growth rate (Fig. 2b) and intracellular fluxes in the central metabolism (Fig. 2c–e), which was caused by a continuous decrease in local light intensity over time (Fig. 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 (Fig. 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 (Fig. 2b). This prediction agrees with two previous studies [20, 21].

Next, we tested the sensitivity of biomass production to the mass transfer rate, PBR surface light intensity, and PBR diameter (Fig. 3). With a light intensity of 100 μE/ m²/s and a moderate mass transfer rate of 15 h⁻¹, small PBRs (30 mm radius) could produce 20 g/L of biomass in 21 days. Although such productivity has been experimentally observed in small PBRs [22], 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 ratio, increase the culture mixing and air flow rate [23], and maintain a sufficient surface light intensity.

Finally, the simulations also demonstrate that, due to random cell movements in PBRs, single cell fluxes may show stochastic changes (Additional file 2: Fig. 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 (Additional file 2: Fig. S3).

**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 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 (Fig. 4a). In the late growth stage, active fluxes through the OPP pathway appeared (Fig. 4a) due to the self-shading effect. The activity of the OPP pathway increased concurrently with the glycogen consumption rate in darkness (Fig. 4a and Additional file 2: Fig. 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 [24, 25]. To further confirm our model predictions, we examined the labelling patterns of histidine by growing Synechocystis 6803 with NaH¹³CO₃ and [¹³C] glucose. When glucose was metabolized via the OPP pathway, non-labeled ribose-5-phosphate was generated from [¹³C] glucose [26], which is a precursor to histidine. Therefore, an active OPP pathway was expected to reduce the ¹³C-enrichment of proteinogenic histidine. Figure 3b shows that the ¹³C-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.

**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 (Additional file 2: Fig. S5), and made the local light intensity dependent on the vertical distance from a cell to the light source. The CO₂ mass transfer rates in shake flasks were calculated based on Eq. (9). As a consequence,
a photosynthesis efficiency of 2.7% (Additional file 2: Fig. S6) resulted from the best fit of specific growth rates under shake flask cultures (Diamond and circle markers in Fig. 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% CO₂ and under 500 μE/ m²/s light intensity) [27]. The model predicted slightly lower specific growth rate than the experimental value (Square marker in Additional file 2: Fig. S5). This difference is possibly due to an increased photosynthesis activity under high CO₂ concentrations [28].

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 [29]. 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 (νₘₒ) and specific growth rate (μ) using only the FBA model. Figure 6a shows νₘₒ and μ as functions of the ratio νₑₙ/μ, which denotes the amount of lactate produced per gram of biomass (or mmol lactate/g biomass). Within a wide νₑₙ/μ range, from 0.01 to 100 mmol lactate/g biomass, μ decreased with increasing νₑₙ/μ, but νₑₙ showed a parabolic tendency, peaking at 0.3 mmol/g/h (Fig. 6a). Next, we used the integrated GSM to simulate the cyanobacterial growth and D-lactate production in PBRs at different νₑₙ/μ ratios (Fig. 6b, c). As a result, increasing the νₑₙ/μ ratio led to lower biomass production, which, however, did not necessarily
Fig. 6 Simulation results of d-lactate producing cyanobacteria performance. a FBA simulations of d-lactate flux and growth rate as functions of \( \nu_{dl}/\mu \) (mmol lactate/g biomass). White markers wild-type strain, black markers glyogen knock-out strain. b e Simulation of biomass growth (b) and d-lactate production (c) of wild-type cyanobacteria at different \( \nu_{dl}/\mu \) ratios in a cylindrical PBR. d e Simulation of biomass growth (d) and d-lactate production (e) of glyogen knock-out cyanobacteria at different \( \nu_{dl}/\mu \) ratios in a cylindrical PBR.

improve the overall d-lactate production. For example, when \( \nu_{dl}/\mu \) 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 \( \nu_{dl}/\mu \) 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 [30]. However, the integrated GSM shows that such a strategy may not offer significant improvements in final lactate productivity in PBRs (Fig. 6a, d, e). This finding is consistent with two recent reports: (1) Glycogen knockout did not enhance lactate productivity under nutrient-sufficient growth conditions [30]. (2) Removal of glycogen in an isobutanol-producing cyanobacterium yielded no benefit in production titer and rate [31]. 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 [32]. Therefore, deleting glycogen or other carbon storage may impair cyanobacterial survival as well as its resistance to environmental stresses and contaminations.

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 milieu) simultaneously.
In the extracellular domain, the integrated GSM can describe changes in nutrient concentrations, biomass accumulation, and local light intensities. As demonstrated by Fig. 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 [23]. To improve the mass transfer rate, enhancing aeration rates has proved efficient [19]. 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 [33]. 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 measurable flux through the OPP pathway could be observed in both wild type and engineered cyanobacterial strains via $^{13}$C-based flux analysis [24, 25]. 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 [33]. Furthermore, we noticed that an active OPP pathway was always present in the $\delta$-lactate producing strain, and it became stronger with increased $\delta$-lactate production (Additional file 2: Fig. 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 [34, 35].

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 [32, 36]. 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) [37], 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.

**Conclusion**

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.

**Methods**

**Cell cultivation**

*Synechocystis* PCC 6803 was cultivated in a modified BG-11 medium [26] at 30 °C and 180 rpm. We first tested the cyanobacterial growth in different culture volumes. In brief, 50, 100, and 150 mL of cell suspensions were cultivated in 250 mL shake flasks under continuous illumination of ~50 μE/m$^2$/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/m$^2$/s). OD$_{730}$ was used to measure biomass density, and the relationship between the biomass concentration and OD$_{730}$ was $0.42 \times$ OD$_{730}$ = Biomass (g/L) [26]. We made duplicate cultures of each condition ($n = 2$).

**$^{13}$C-Labelling experiment**

$^{13}$C-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...
[1-13C] glucose and 4 g/L NaH13CO3 (tracers were purchased from Sigma-Aldrich, Saint Louis, USA). The TBDM (N-tert-butyldimethylsilyl-N-methyltrifluoro- acetamide) method [26] was used to analyze the labeling patterns of proteogenic histidine. In brief, cells were harvested by centrifugation, and cell pellets were hydrolyzed in 6 M HCl solution at 100 °C for 24 h. The amino acid solution was air-dried and then derivatized by TBDM (Sigma-Aldrich, USA) at 70 °C for 1 h. A gas chromatograph (GC) (Hewlett-Packard model 7890A, Agilent Technologies, CA) equipped with a DB-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 13C atoms, respectively.

**Flux balance analysis model**

The FBA model was modified from the cyanobacterial model [JN878 [15], which has 843 reactions, including photosynthesis and the central carbon metabolism. A complete list of reactions is provided in Additional file 1. The JN878 model contains a recently discovered α-aminoisobutyrate shunt [38] that converts 2-oxoglutarate to succinate in Synechocystis 6803. In our model, two new reactions were added, namely glycogen storage \(\rightarrow\) glycogen [c] and α-lactate [c] \(\rightarrow\) α-lactate (external), which were respectively used to simulate glycogen storage/consumption and α-lactate production by an engineered cyanobacterial strain [29]. The mathematical description of our FBA model is as follows:

\[
\begin{align*}
\text{maximize } \mu & \quad \text{subject to } S \cdot v = 0 \\
& \quad lb \leq v \leq ub \\
& \quad v_{\text{CO}_2} \leq f_l(K_{lax}, [\text{CO}_2], K_{m}) \\
& \quad v_{\text{photon}} \leq f_l(I, X, v_{\text{photon}}, 0) 
\end{align*}
\]

(1)

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, \(f_l\) is a function of the mass transfer coefficient \(K_{lax}\), dissolved \(\text{CO}_2\) concentration \([\text{CO}_2]\), and half-saturation constant for dissolved \(\text{CO}_2\), \(K_{m}\); \(f_l\) is a function of the cell’s local position \(I\), biomass concentration \(X\), and photon influx on the PBR surface \(v_{\text{photon}}\). 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 [39], which was solved by the MATLAB built-in function ‘quadprog’ using the ‘interior-point-convex’ algorithm:

\[
\begin{align*}
\text{minimize } & \frac{1}{2}v^T H v + f^T v \\
\text{subject to } & S \cdot v = 0 \\
& lb \leq v \leq ub \\
& v_{\text{CO}_2} \leq f_l(K_{lax}, [\text{CO}_2], K_{m}) \\
& v_{\text{photon}} \leq f_l(I, X, v_{\text{photon}}, 0) 
\end{align*}
\]

(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 [40] 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 (Additional file 1). By having such a database, we could directly use pre-calculated fluxome from the database according to culture conditions in PBRs. Thereby, we did not need to redo flux calculations at each time interval during new simulations.

**Simulation of cyanobacterial growth via integrating FBA, kinetics, and cell movements**

Figure 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 (Additional file 2: Fig. S1 shows that further decreasing the interval period did not change the simulation results). In each time interval, a simplified sinusoid equation [41] was used to estimate the cell location in a well-mixed PBR:

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

(3)

where \(I\) 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; \(r_s\) is the 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 [42]. In our model platform, we distinguished cell populations with different circulation times (Additional file 2: Fig. S2–3). Thus, the whole culture was considered to be comprised of twelve sub-populations instead of a plethora of cyanobacterial cells. Based on cell locations and
the biomass concentrations, we calculated the local light intensity [43]:

\[
\frac{v_{\text{photon}}}{v_{\text{photon,0}}} = \left( \frac{1}{0.0216 \cdot (l + 1)^{1.54} \cdot (0.130 \cdot X^2 + 1)^{1.18}} \right)
\]  

(4)

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\(_2\) uptake flux was described by a Michaelis–Menten equation:

\[
v_{\text{CO}_2} = v_{\text{CO}_2 \text{max}} \frac{[\text{CO}_2]}{K_m + [\text{CO}_2]}
\]  

(5)

where \( v_{\text{CO}_2 \text{max}} \) is the maximum uptake rate of dissolved CO\(_2\)/HCO\(_3^-\), \( K_m \) is the half-saturation constant, and [CO\(_2\)] represents dissolved CO\(_2\) concentration. This study assumed that pH was constant at 8.0, and that the dissolved CO\(_2\) and cell culture were homogeneous in PBRs. Because we assumed that cell metabolism was pseudo-steady in each interval [44], the FBA model could use linear optimization to profile the intracellular fluxes constrained by light and carbon input fluxes (Eqs. 4 and 5). The FBA model then predicted the growth rates, glycolysis synthesis rates, and CO\(_2\) uptake rates of cell populations with different circulation times in PBRs. Those values were averaged based on the probability distribution function (Additional file 2: Fig. S2a):

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

\[
v_{\text{CO}_2 \text{app}} = \sum_{i=1}^{n} P_i v_{\text{CO}_2,i}
\]

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

where \( P_i \) is the fraction of ith cell population (Additional file 2: Fig. S2a), \( \mu_{\text{app}} \) is the apparent specific growth rate in PBRs, \( v_{\text{CO}_2 \text{app}} \) is the apparent CO\(_2\) uptake rate, and \( v_{\text{glycogen,app}} \) 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\(_2\)/HCO\(_3^-\) concentrations, and so forth.

\[
\frac{dX}{dt} = \mu_{\text{app}} \cdot X - K_d \cdot X
\]

\[
\frac{d[\text{glycogen}]}{dt} = v_{\text{glycogen,app}} \cdot X - \beta \cdot K_d \cdot X
\]

\[
\frac{d[\text{CO}_2]}{dt} = K_{\text{ld}} \cdot ([\text{CO}_2]^* - [\text{CO}_2]) + v_{\text{CO}_2 \text{app}} \cdot X,
\]

(7)

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

\[
X_{i+1} = X_i + \mu_{\text{app}} \cdot X_i \cdot \Delta t - K_d \cdot X_i \cdot \Delta t
\]

\[
[\text{glycogen}]_{i+1} = [\text{glycogen}]_i + v_{\text{glycogen,app}} \cdot X_i \cdot \Delta t - \beta \cdot K_d \cdot X_i \cdot \Delta t
\]

\[
[\text{CO}_2]_{i+1} = [\text{CO}_2]_i + K_{\text{ld}} \cdot ([\text{CO}_2]^* - [\text{CO}_2]) + v_{\text{CO}_2 \text{app}} \cdot X_i \cdot \Delta t,
\]

(8)

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; [glycogen] is the overall glycogen concentration in the PBR, in mmol/L; \( K_{\text{ld}} \) 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 \( \mu_{\text{app}} \) and \( v_{\text{glycogen,app}} \) and \( v_{\text{CO}_2 \text{app}} \) 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 1. For shake flask conditions, \( K_{\text{ld}} \) was determined by the following equation: [45]

---

**Table 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_d )</td>
<td>Mass transfer rate of CO(_2)</td>
<td>( 10^3(3-15) )</td>
<td>h(^{-1})</td>
<td>[19]</td>
</tr>
<tr>
<td>( \mu )</td>
<td>Death rate</td>
<td>0.0079</td>
<td>h(^{-1})</td>
<td>[46]</td>
</tr>
<tr>
<td>( l_0 )</td>
<td>Radius of PBR</td>
<td>60</td>
<td>mm</td>
<td>Similar to the reactor used in Reference [42]</td>
</tr>
<tr>
<td>( I )</td>
<td>Surface light intensity</td>
<td>50</td>
<td>mmol/g/h</td>
<td>Equivalent to (-50 \mu E/m^2/s)</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>[47]</td>
</tr>
<tr>
<td>( p_{\text{H}} )</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(_{600}) 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 (^a)</td>
</tr>
</tbody>
</table>

---

\(^a\) The values are used in Figs. 2, 3 and 5; \(^b\) a photosynthesis efficiency of 1.5 % is assumed in Figs. 2, 3 and 5; and \(^c\) calculation is based on experimental conditions (Additional file 2)
\[ K_{L}a = 0.032 \times N \left( \frac{V}{L} \right)^{0.445} \]

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 code of the integrated GSM is provided in Additional file 3 and 4.

### Additional files

- Additional file 1. Flux distribution database.
- Additional file 2. Supplementary figures and tables.
- Additional file 3. MATLAB file for simulating cyanobacterial performance in PBRs.

### Author's contributions

YJT and LH initialized this project. LH and SW built the Integrated GSM. LH, NW, and AR performed the experiments. All authors read and approved the final manuscript.

### Author details

1 Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO 63130, USA. 2 Department of Mechanical Engineering and Materials Science, Washington University, St. Louis, MO 63130, USA. 3 Department of Biochemistry and Molecular Biology, College of Wooster, Wooster, OH 44691, USA.

### Acknowledgements

We would like to thank Dr. Le You for valuable discussions for this project. We are also grateful for the funding supports from a US DOE grant (DESC0012722) and an NSF Grant (CBET 1438125).

### Competing interests

The authors declare that they have no competing interests.

Received: 7 October 2015 Accepted: 12 December 2015 Published online: 24 December 2015

### References


### Additional files

- Additional file 1. Flux distribution database.
- Additional file 2. Supplementary figures and tables.
- Additional file 3. MATLAB file for simulating cyanobacterial performance in PBRs.
Appendix Chapter 2. Integrating MS1 and MS2 Scans in High-Resolution Parallel Reaction Monitoring Assays for Targeted Metabolite Quantification and Dynamic $^{13}$C-Labeling Metabolism Analysis
Integrating MS1 and MS2 Scans in High-Resolution Parallel Reaction Monitoring Assays for Targeted Metabolite Quantification and Dynamic $^{13}$C-Labeling Metabolism Analysis

Zhucui Li,†,‡,§ Yujing Li,† Wuju Chen,† Qichao Cao,‡ Yufeng Guo,‡ Ni Wan,† Xiaolong Jiang,† Yinjie J. Tang,† Qinhong Wang,† and Wenging Shui†,‡,§

†Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China
‡Human Institute, ShanghaiTech University, Shanghai 201210, China
§College of Life Sciences, Nankai University, Tianjin 300071, China

Supporting Information

ABSTRACT: Quantification of targeted metabolites, especially trace metabolites and structural isomers, in complex biological materials is an ongoing challenge for metabolomics. Initially developed for proteomic analysis, the parallel reaction monitoring (PRM) technique exploiting high-resolution MS2 fragment ion data has shown high promise for targeted metabolite quantification. Notably, MS1 ion intensity data acquired independently as part of each PRM scan cycle are often underutilized in the PRM assay. In this study, we developed an MS1/MS2-combined PRM workflow for quantification of central carbon metabolism intermediates, amino acids and shikimate pathway-related metabolites on an orthogonal QqTOF system. Concentration curve assessment revealed that exploiting both MS1 and MS2 scans in PRM analysis afforded higher sensitivity, wider dynamic range and better reproducibility than relying on either scan mode for quantification. Furthermore, Skyline was incorporated into our workflow to process the MS1/MS2 ion intensity data, and eliminate noisy signals and transitions with interferences. This integrated MS1/MS2 PRM approach was applied to targeted metabolite quantification in engineered E. coli strains for understanding of metabolic pathway modulation. In addition, this new approach, when first implemented in a dynamic $^{13}$C-labeling experiment, showed its unique advantage in capturing and correcting isotopomer labeling curves to facilitate nonstationary $^{13}$C-labeling metabolism analysis.

Mass spectrometry (MS) coupled to different separation techniques such as liquid chromatography (LC), gas chromatography (GC), and capillary electrophoresis (CE) has become the leading platform for large-scale identification and quantification of metabolites in various biological systems. Nontargeted MS-based metabolomic surveys mainly employ the high-resolution full scanning instruments to profile hundreds of endogenous and drug metabolites. In these studies, metabolite identification is typically ensured by MS/MS analysis of isolated precursors using a data-dependent acquisition (DDA) strategy. However, metabolite quantification relying on MS1 scan data could suffer from low reproducibility due to inadequate sensitivity of MS1 scans for metabolites that are less abundant, prone to matrix interference or have isomeric compounds. Very recently, a data-independent acquisition (DDA) workflow using SWATH-MS technique initially developed for proteomic quantification has been adapted to global metabolomics to allow for sensitive, reproducible, and high-throughput metabolite quantification based on fragment ion intensities in MS2 scans.

MS2-based quantification of small molecules has been primarily implemented in pharmaceutical laboratories for specified drug metabolite analysis. Multiple reaction monitoring (MRM) performed on triple quadrupole instruments is recognised as the “gold standard” MS technique for targeted biomarker quantification due to its high sensitivity and robustness. The MRM technique has been recently adopted in metabolomic studies focusing on a defined subset of metabolites representative of key pathways. This type of targeted metabolomics is a valuable tool for assessing changes of metabolic pathways or networks resulting from genetic mutation, altered gene expression, protein dysfunction or

Received: October 7, 2016
Accepted: December 2, 2016
Published: December 2, 2016
environmental perturbation. For example, Wei and colleagues developed an LC/MS/MRM platform to monitor 205 endogenous metabolites and applied it to targeted metabolite analysis in plasma samples. Nevertheless, MRM assays performed on low-resolution mass spectrometers sometimes are limited by ambiguous metabolite identifications as a result of interfering peaks from complex matrix that resemble the compound of interest under a specific Q1/Q3 transition.

Technological improvements in high-resolution MS have led to new acquisition methods which have challenged the dominance of MRM for targeted MS. One approach, parallel reaction monitoring (PRM), has emerged as a capable alternative to MRM. In a PRM assay, the precursor ion is isolated in Q1 and fragmented in Q2, and all generated fragment ions are monitored in MS2 scans on a high-resolution accurate mass spectrometer. Then fragment ion intensities from MS2 spectra are used for specific precursor quantification. The PRM approach has been successfully implemented in a panel of targeted proteomic studies and demonstrated superior specificity, accuracy, and multiplexing capability compared to the MRM assay.

In contrast, there has been a rare report of PRM application to targeted metabolomic analysis except the most recent research from Zhou et al. and Qiu et al. Both groups developed PRM assays on a hybrid Q-Exactive instrument (Thermo) for measurement of targeted metabolites. Zhou et al. reported monitoring 237 endogenous metabolites for which structural identification solely relied on MS/MS spectral data available in public databases. Qiu et al. constructed PRM assays using chemical standards for quantification of 25 intracellular primary metabolites. It is worth noting that PRM-based quantification is carried out using ion intensities of MS2 fragments, yet the role of MS1 scans has largely been ignored, despite the fact that the respective ion signals of certain precursors could be stronger than those of fragments. Systematic evaluation of MS1 vs MS2 quantification in PRM assays has not been reported for either metabolite or peptide analysis.

Our study, for the first time, has developed an MS1/MS2 combined PRM workflow for targeted metabolite quantification as well as 13C-labeling metabolism analysis on an orthogonal QqTOF instrument (TripleTOF 6600). A total of 65 targeted metabolites include central carbon metabolism (CCM) and shikimate pathway intermediates, and free amino acids. After comparing the sensitivity, reproducibility, and linearity of metabolite quantification based on MS1 versus MS2 ion chromatograms derived from PRM cycles, we concluded that exploiting both MS1 and MS2 scans in PRM assays achieves the best quantitative performance. Furthermore, Skyline can be incorporated into our workflow to process the MS1/MS2 ion intensity data and eliminate noisy signals and transitions with interferences.

The integrated MS1/MS2 PRM approach was applied to targeted metabolite quantification in engineered E. coli strains for understanding of metabolic pathway modulation. In addition, this approach was first implemented in a dynamic 13C-labeling experiment to reveal its capability of accurate measurement of authentic isomers for tracing cell metabolism.

**EXPERIMENTAL SECTION**

Chemicals, details in sample preparation, and data processing can be found in Supporting Information.

**Metabolite Extraction from E. coli Cells.** A wild-type E. coli strain AB2883 (WT) and two engineered strains (TIB01 and TIB02) with improved production of 3-dehydroshikimic acid were grown in LB media aerobically at 30 °C, 250 rpm to produce the seed culture. The seed cultures were then transferred into a 250 mL flask containing the NBS minimal complete medium and grew under 37 °C, 250 rpm until OD600 reached 2.0. Bacteria cultures were diluted 1:1000 and harvested through centrifugation at 4 °C for 3 min at 3500 g. The cell pellets were washed twice with PBS and extracted using a prechilled solvent mixture of methanol/acetonitrile/water (4:4:2, v/v) with 1.0% FA which contains d4-succinic acid as the internal standard. The metabolite extracts were snap frozen in liquid nitrogen, thawed, and then sonicated in a water bath. The freeze-thaw-sonication cycle was repeated twice. Then the samples were incubated at −20 °C for 2 h. After centrifugation at 16000 g for 20 min at 4 °C, the supernatant was harvested and dried out by speed vacuum and stored in −80 °C. Before HILIC-QqTOF analysis, the metabolite extracts were reconstituted in the acetonitrile/water (1:1, v/v) solvent. Each strain was prepared in biological duplicate and metabolite extraction was processed in duplicate for each biological replicate.

**HILIC-QqTOF Analysis.** Metabolite standards or total cell extracts were analysed on a Nexera HPLC system (SHIMADZU, Japan) connected to a TripleTOF 6600 mass spectrometer (AB SCIEX, U.S.A.). LC separation was carried out on a ZIC-HILIC column (100 mm × 2.1 mm, 3.5 μm, Merck, Germany). The mobile phase A was 10 mM ammonium acetate in water, and mobile phase B was acetonitrile. The linear gradient used was as follows: 0–3 min, 90% B; 3–18 min, 90–60% B; 18–25 min, 60–55% B; 25–27 min, 55–50% B; 27–30 min, 50% B, and the column was re-equilibrated for 10 min. The flow rate was 0.2 mL/min and the sample injection volume was 5 μL. Each metabolite standard was also injected into the mass spectrometer through direct infusion to acquire reference MS/MS spectra. Mass spectra were acquired in a negative ion ESI mode with a mass range from 100 to 1000 m/z. The ion source parameters are ion spray voltage, −4500 V; curtain gas, 35 psi; nebulizer gas, 55 psi; heater gas, 55 psi; source temperature, 550 °C. Ramping CE and fixed CE were assessed using standards to acquire optimal CEs for individual metabolite targets.

MS data acquisition for PRM analysis was programmed in the product ion mode, and consisted of one MS1 scan (150 ms) with the mass range from 50 to 900 m/z, followed by targeted MS2 scans (50 ms) under the optimal CE condition across the mass range from 30 to X (X = precursor mass + 10 Da) m/z. The maximal cycle time was 3.25 s for a total of 65 targets. In the dynamic 13C-labeling experiment, total cell extracts sampled at different time points were also analyzed by HILIC-QqTOF in the PRM mode. For certain metabolites having suspected interferences, targeted MS2 scans were performed on specific isotomers of interest.

**RESULTS AND DISCUSSION**

Establishing PRM Assays of 61 Metabolites on TripleTOF 6600 Mass Spectrometer. To develop PRM assays for quantification of 61 intracellular metabolites including 44 central carbon metabolism (CCM) intermediates and 17 amino acids (full name and abbreviation in Table S1), we first analyzed a mixture of the synthetic metabolite standards on the TripleTOF 6600 mass spectrometer in PRM mode. Each metabolite exceeding the MS1 intensity
threshold was subjected to MS/MS fragmentation under default sweeping collision energy (−35 ± 15 eV). However, we noticed that the MS/MS spectra generated under default CE on certain metabolites were of low quality as a result of either insufficient or excessive fragmentation of the precursor ions (Figure S1A). This is different from peptide PRM assays in which the majority of peptides of diverse sequences are fragmented effectively under theoretically predicted CE based on the precursor m/z and charge state. To improve MS/MS spectral quality, we manually optimized CE for each metabolite standard in a way similar to MRM assays on small molecules. For both NADPH and acetyl CoA, an increased CE relative to the default setting gave rise to a greater number of fragment ions as well as diminished precursor peaks (Figure S1B). For another metabolite TPP, two distinct fragment ions at m/z 176.938 and 301.970 of 10-fold higher intensity than the default condition were generated from the precursor subject to a lower CE (Figure S1B). The optimal CE values ranged from −15 to −60 eV for specific metabolites (Table S2).

We then examined the optimized MS/MS spectra to select 1–6 fragments of relatively high abundance and large m/z for target quantification. Skyline software was used here to automatically extract MS2 ion chromatograms after we defined specific transitions for each metabolite precursor. As MS1 full scan was also performed in each cycle of the PRM assay, quantitative data were extracted for a given precursor ion using the same "Transition List" module in Skyline. Therefore, we were able to extract both MS1 and MS2 transitions in a simultaneous data-processing workflow with Skyline.

The reproducibility of MS2-based quantification of 45 tested metabolites was significantly impacted by CE conditions. With optimized CE, fragment peak areas for 38 metabolites showed CVs < 20% in triplicate measurements whereas the number of metabolites with fragment peak area CVs < 20% dropped to 22 under the original CE condition (Figure S1C). It is noteworthy that the variation of CVs has little to do with the data sampling points over extracted ion chromatograms. In fact, a minimum of 18 data points were acquired for all targeted metabolites considering the wide HILIC peak width (1–2 min) and the maximal cycling time including all MS2 acquisitions (>3.25 s). Therefore, we would recommend optimization of CE settings for individual metabolites to ensure sensitivity and reproducibility of targeted PRM analysis, at least on a QqTOF instrument, even though rolling CE or constant CE is typically adopted in data-dependent acquisition for nontargeted metabolomics studies.

Comparison of MS1 and MS2 Quantification in PRM Using Concentration Curves. To compare the quantitative performance of MS1 and MS2 chromatogram extraction from PRM acquisitions, we carried out a series of dilution experiments in both simple and complex matrix. A mixture of 15 metabolite standards were diluted in simplex matrix (50% MeOH) spanning a concentration range from 5 pg to 100 ng on column. The MS1 peak area of the precursor ion and sum of MS2 peak areas of selected fragment ions provided independent quantitative measure of the targeted metabolite at MS1 and MS2 levels. The sensitivity and linear response range of metabolite quantification in simple matrix was...
Figure 2. Concentration curve assessment of MS1 vs MS2 quantification in the PRM assay of metabolites diluted in complex matrix. (A) Concentration curves of RSP by MS1 or MS2 quantification (left), MS1 ion chromatogram (middle), and MS2 ion chromatograms (right) of RSP in the 512-fold diluted cell extract. (B) Concentration curves of Val by MS1 or MS2 quantification (left), MS1 ion chromatogram (middle) and MS2 ion chromatograms (right) of RSP in the 512-fold diluted cell extract. (C) Distribution of R² value for MS1/MS2 PRM quantification of targeted metabolites in serially diluted cell extracts (D). Distribution of CV% for MS1/MS2 PRM quantification of targeted metabolites in cell extracts at different dilute fold (N = 3).

summarized in Table S3, where 9 out of 15 metabolites showed a lower LOQ, when quantified by MS2 than by MS1. For instance, the LOQ of FAD quantification by extracting six MS2 transitions was 5 pg yet the LOQ was 10-fold higher (50 pg) for MS1 chromatogram extraction (Figure 1A). At low concentrations, the MS1 precursor ion chromatogram of FAD was more prone to ion suppression and interferences, leading to increased variability and reduced sensitivity. In contrast, the MS2 fragment ions of FAD appeared to be more robust and selective, thus offering a lower LOQ and a wider dynamic range (Figure 1A). Among the tested metabolites in simple matrix, five showed similar LOQ levels and linear response range between MS1 and MS2 quantification in PRM assays (Table 3). For example, the precursor and fragment ion chromatograms of L-malic acid (Mal) displayed comparable robustness in the low concentration range (Figure 1B). Notably, for 5-ALA, MS1 precursor quantification enabled much better sensitivity than MS2 fragment quantification. The LOQ of 5-ALA quantification with MS2 scans was 20-fold higher than that with MS1 scans, which was attributed to very weak signals of all detected fragment ions of this amino acid (Figure 1C).

PRM quantification of all 61 targeted metabolites was also evaluated in complex matrix by preparing up to 2^12-fold dilution of an E. coli total extract in extraction solvent. Forty-one metabolites of our interest showed better sensitivity, consistency, and wider dynamic range in MS2 versus MS1 quantification (full data in Table S4). In a representative case of d-ribose-5-phosphate (RSP), its MS1 precursor ion response plateaued at the low end of the concentration curve whereas the MS2 ion chromatograms encountered little interference and retained strong signals in the high dilution-fold region (Figure 2A). However, for other metabolites such as Val, very poor fragment ion chromatograms were observed with increase of the dilution fold, thus rendering MS2 quantification ineffective (Figure 2B). It is of our notion that metabolites of larger molecular weight including most sugar phosphates, organic acids, and energy/redox cofactors that tend to generate a greater number of fragments of strong intensity are more amenable to MS2 quantification. In contrast, smaller metabolites especially amino acids producing fewer and weaker fragments are better to be quantified in MS1 scans. Depending on the quantitative reproducibility, sensitivity and linear response range of the total cell extract sample, we classified 50 metabolites to be MS2 quantifiable and 11 to be MS1 quantifiable targets (Table S4).

When each metabolite was measured in its optimal mode (MS1 vs MS2), 89% of all targeted metabolites diluted in complex matrix had a linear regression coefficient (R²) of the concentration curves above 0.995, and 95% of metabolites had r² above 0.990 across 2–3 orders of magnitude (Figure 2C, Table S4). Moreover, the median CV of metabolite
Figure 3. Quantification of isomeric metabolites by PRM MS2 scans. (A) MS/MS spectra of CIT and ICIT. (B) MS1 ion chromatogram of both isomer precursors (left), MS2 ion chromatograms of the unique fragment(s) of CIT (middle) and ICIT (right). (C) MS/MS spectra of Leu and Ile. (D) MS1 ion chromatogram of both isomer precursors (left), and MS2 ion chromatograms of the unique fragments of Leu (middle) and Ile (right). The isomer-specific fragments are underlined in MS/MS spectra.

Quantification in triplicate measurements was below 20% over the entire concentration range down to 20-fold dilution of the original cell extract (Figure 2D). These results indicated that excellent linearity, reproducibility and sensitivity of the PRM assay can be achieved by combining MS1 and MS2 scans.

Quantification of Isomeric Metabolites by PRM MS2 Scans. MS2 scans are particularly useful for differentiation of coeluting structural isomers of identical or very similar molecular weight. Within our target list, two pairs of isomeric metabolites (citrate/isocitrate and leucine/isoleucine) that completely or in most part coelute in HILIC separation, were indistinguishable based on their MS1 precursor chromatograms. However, these isomers could be differentiated with the isomeric-specific fragment ions. For instance, two ions at m/z 72.995 and 117.019 are unique fragment ions derived from isocitrate whereas the abundant ion at m/z 87.009 specifically originates from citrate (Figure 3A). In this way the unique transitions allowed for differentiation and quantification of this isomeric pair in complex matrix even though their MS1 chromatograms were overlapped (Figure 3B). Likewise, Leu and Ile both having their unique fragment ions were quantified with distinct MS chromatograms of specific isomers (Figure 3C,D).

With PRM MS2 scans, strong linear response of these metabolite isomers were obtained in both simple and complex matrix (R² > 0.98 for all four metabolites, see detailed data in Tables S3 and S4). Notably, it is difficult to differentiate these two pairs of isomeric metabolites in traditional MRM assays because MRM typically selects the most abundant fragment ion, which is common to both isomers in this case. Therefore, PRM enables quantification of specific isomers by selecting less abundant and unique transitions that are of adequate sensitivity and free of interferences from the complex background owing to the high-resolution MS2 scans. The previous study by Zhou J et al. also demonstrated the advantage of PRM in increased specificity of metabolite identification and flexibility of postacquisition assay refinement.

Integrated MS1 and MS2 for Metabolomic Study of Engineered E. coli Strains. To further explore the benefits of integrated MS1 and MS2 scans in PRM assays, we performed targeted metabolite analysis in three different E. coli strains of WT, TIM01, and TIM02. Strains TIM01 and TIM02 were
genetically engineered from the wild-type (WT) in the shikimate pathway for improved production of 3-dehydroshikimic acid (DHS). DHS is a key hydroaromatic intermediate in the biocatalytic conversion of glucose into aromatic bioproducts and a variety of industrial chemicals. In addition to the CCM intermediates and amino acids, we established PRM methods for measurement of DHS and three related intermediates (DAHP, DHQ, and SK, full name in Table S1) in its biosynthetic pathway. Then we conducted PRM analysis of all 65 targeted metabolites in the cell extracts from strains WT, TIB01, and TIB02 in four replicates. As MS1 and MS2 ion intensity data were acquired simultaneously in each PRM scan cycle, we extracted the quantitation data for each metabolite from separate MS1 and MS2 scans in Skyline. Evaluation of peak area CV distribution for all targeted metabolites revealed that combining MS1 and MS2 quantification achieved the best precision (Figure 4A). The median CVs of the combined MS1/MS2 PRM assays for all target were 20.3, 18.1, and 13.9% for WT, TIB01, and TIB02 strains, respectively. This result reflects remarkable analytical reproducibility of our approach if considering relatively large variation in cell culture and sample preparation. Thus, it is advantageous to exploit MS1 scan data in PRM assays, which is often underutilized so as to further enhance robustness and sensitivity for targeted metabolome analysis.

The pretreated data matrix from targeted metabolites quantified in three strains was used to develop an unsupervised PCA model (R2 = 0.914, Q2 = 0.944). The result exhibited a well-segregated, closely clustered pattern among different groups in the PCA score plot, indicating a specific molecular signature of each strain that was distinguishable between each other (Figure 4B). Replicates of the same strain were all clustered together, again suggesting good reproducibility of our targeted metabolite quantification. Significantly modulated metabolites in the CCM and shikimate pathways as well as free amino acids in two engineered strains vs WT were then identified according to dual criteria of VIP score in a supervised OPLS-DA model > 1 and p-value in the ANOVA test < 0.01 (Figure 4C). As expected, the final product DHS and its synthetic precursors were all substantially up-regulated in engineered strains, which was consistent with previous reports using similar engineering strategies. A handful of glycolytic intermediates (3-BPG, 2-BPG, PEP) and TCA cycle intermediates (CIT, ICIT) were down-regulated in two engineered strains to varying extent, which might imply effective manipulation of the central metabolic modules to drive carbon sources into the DHS synthetic pathway. Precise quantitation and clustering of 65 metabolic targets including trace metabolites such as redox cofactors in three strains using MS1/MS2-combined PRM assays were presented in Figure 4D. These results reflected fluctuation of intracellular metabolites and the product after pathway engineering, which would be helpful for identifying the biosynthetic bottlenecks and reshaping the fitness of host cells (to be discussed elsewhere).

PRM MS2 Scans Facilitating Dynamic 13C-Labeling Metabolism Analysis. 13C-Metabolic flux analysis (MFA) is a well-established technology, which uses isotopic tracers to determine in vivo reaction rates (fluxes) in the metabolic network. For the first time, we prepared 13C-precursor pools with a practical yield (92%) for each precursor pool with a practical yield (92%) for each
network of cells. Dynamic $^{13}$C-MFA based on transient labeling of intracellular metabolites has shown great potential in control and optimization of biotechnological processes not accessible by conventional MFA.

Dynamic $^{13}$C-MFA relies on time-resolved $^{13}$C-labeling metabolism analysis which refers to the measurement of isotopic labeling enrichment in metabolite pools with a minimal information gain. A big challenge for dynamic $^{13}$C-labeling metabolism analysis is precise measurement of the isotopomer labeling fractions for a large number of intracellular metabolites after the pulse of $^{13}$C substrates. Because the complexity of metabolite isotopomer peaks is substantially expanded in $^{13}$C-labeled metabolic pools, the conventional quantification method of extracting MS1 data for different isotopomers of the same metabolite species would be severely limited due to frequent interferences. For example, in the dynamic $^{13}$C-labeling experiment conducted on the previously described strain TIB01, we observed not only the isotopomer peak at M0 (i.e., unlabeled monoisotop) yet also intense peaks at M2 and M4 positions in the MS1 spectrum of d-RSP measured at time point T0 (Figure 5A). Although the M2 peak could be excluded due to its large mass deviation from the expected isotopomer $m/z$ (43 ppm error), it was very difficult to evaluate the M4 peak which coeluted with all other isotopomers of d-RSP and fell within 3 ppm mass tolerance. Evidently, in this particular sample, both M2 and M4 peaks must have come from ion interferences in the complex background because no $^{13}$C glucose was present in the cell culture at T0. A targeted PRM assay on M0, M2, and M4 and the fully labeled MS5 peaks of d-RSP was conducted on the T0 sample and the panel of $^{13}$C-labeling samples at different time points from 10 s to 2 h. The PRM analysis revealed that the MS/MS spectra of M0 and M5 peaks shared the same pattern yet they were drastically different from those of M2 and M4, confirming the false identity of two isotopomer peaks (Figure 5B). Without PRM MS2 scans, the temporal labeling curves of d-RSP isotopomers were in abnormal shape because the interference ions at M2 and M4 positions largely affected calculation of the labeling fractions of all isotopomers (Figure 5C, left). After removing the interferences with the PRM assay, a regular smooth trend of labeling fraction changes for M0 and M5 peaks were observed, indicating accurate monitoring of the dynamic metabolic flux through d-RSP (Figure 5C, right). Similarly, interference peaks at M2 and M4 positions of another
metabolite di-glucono-1,5-lactone (GL) could be easily flagged and removed by PRM analysis, which ensured correct labeling curve input over the defined time course for dynamic $^{13}$C-MFA (Figure S2).

It was noteworthy that we exploited MS1 data from the PRM assays to determine isotopic labeling fractions of all isomers of any targeted metabolite (e.g., M0–M5 peaks of 8-RSP, M0–M6 peaks of GL), while MS2 data from the same PRM scan cycle validated each isomer with potential interference. Thus, the high-sensitivity PRM MS2 scan allowed for confident discrimination of suspected interferences that have RT profiles and accurate masses very close to the authentic isomers.

**CONCLUSION**

This study demonstrated the unique advantage of integrating MS1 and MS2 scans in high-resolution PRM assays on a QqQTOF platform for multiplexed metabolite quantification. A total of 65 CCM intermediates, amino acids and shikimate pathway-related metabolites were quantified based on either MS1 or MS2 ion intensity data acquired from PRM scan cycles. Concentration curve assessment implicated that exploiting both MS1 and MS2 scans in PRM analysis afforded higher sensitivity, wider dynamic ranges and better reproducibility than relying on either scan mode for quantification. Furthermore, Skyline allowed us to efficiently process these MS1/MS2 ion chromatograms in an integrated mode, and readily eliminate noisy signals and transitions with interfering ions. The high resolution, high mass accuracy and postacquisition refinement options of the PRM assay make it an ideal means for targeted metabolomics. The integrated MS1/MS2 PRM approach developed in our study was applied to targeted metabolite quantification in engineered E. coli cells. This will provide a better understanding of metabolic pathway modulation. In addition, this new approach has shown its significant value in facilitating nonstationary $^{13}$C-labeling metabolomic analysis.

**ASSOCIATED CONTENT**

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.analchem.6b02947.

Table S1: Targeted metabolites in this study and abbreviation. Table S2: Optimized CE values for targeted metabolites analyzed on TripleTOF 6600. Table S3: Validation of standard curves for targeted metabolites diluted in simplex matrix quantified by MS1 or MS2 scans. Table S4: Validation of concentration curves for targeted metabolites diluted in complex matrix quantified by MS1 or MS2 scans. Figure S1: Optimization of CE conditions for targeted metabolites. Figure S2: Distinguishing interferences of GL isomers in dynamic $^{13}$C-labeling metabolomic analysis using PRM MS2 scans (PDF).

**AUTHOR INFORMATION**

Corresponding Author

*Tel.: 86-21-20685595. E-mail: shuiewq@shanghaitech.edu.cn.*

ORCID

Wenqing Shui: 0000-0002-5265-2477

**Author Contributions**

*These authors contributed equally to this work (Z.L. and Y.L.).

**Notes**

The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This work was supported by grants from the Bairenjihua Program of the Chinese Academy of Sciences (W.S.) and the National Natural Science Foundation of China (No. 31401150), the Key Projects in Tianjin Science and Technology Pillar Program (No. 14JCZDSY00062), N.W. was supported by the U.S. NSF Grant (CBET 1438125).

**REFERENCES**