Investigating Cyanobacteria Metabolism and Channeling-based Regulations via Isotopic Nonstationary Labeling and Metabolomic Analyses

Mary Helen Abernathy
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/eng_etds
Part of the Biomedical Engineering and Bioengineering Commons, and the Chemical Engineering Commons

Recommended Citation
https://openscholarship.wustl.edu/eng_etds/396

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 Engineering and Applied Science Theses & Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
Investigating Cyanobacteria Metabolism and Channeling-based Regulations via Isotopic Nonstationary Labeling and Metabolomic Analyses

by

Mary H. Abernathy

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

December 2018
St. Louis, Missouri
# Table of Contents

List of Figures ..................................................................................................................... vi
List of Tables ........................................................................................................................ viii
List of Abbreviations ........................................................................................................... ix
Acknowledgments .................................................................................................................. x
Abstract ................................................................................................................................. xii

Chapter 1: Introduction .......................................................................................................... 1

1.1 Cyanobacteria as cell factories ....................................................................................... 2
  1.1.1 Inherent challenges facing cyanobacteria ................................................................. 2
  1.1.2 Rewiring cellular metabolism is difficult ................................................................. 4
  1.1.3 Channeling has significant biotechnology impacts ................................................... 5

1.2 INST-MFA as a tool for decoding cellular phenotypes .................................................. 7
  1.2.1 Materials associated with INST-MFA ...................................................................... 11
  1.2.2 Methods for performing INST-MFA ....................................................................... 13
  1.2.3 Consideration of metabolite channeling in INST-MFA .......................................... 19
  1.2.4 Notes on INST-MFA ............................................................................................... 20

Chapter 2: Deciphering cyanobacterial phenotypes for fast photoautotrophic growth via INST-
MFA ....................................................................................................................................... 24

2.1 Introduction ...................................................................................................................... 25

2.2 Results .............................................................................................................................. 27
  2.2.1 Syn.2973 growth and biomass composition ......................................................... 27
  2.2.2 INST-MFA of photoautotrophic metabolism ............................................................ 29
  2.2.3 Synechococcus flux maps in PBR and SF cultures ................................................... 31
  2.2.4 Isotopic metabolite profiling for relative pool size estimation .............................. 32
  2.2.5 Effect of exogeneous organic acids on Syn. 2973’s growth ................................... 33

2.3 Discussion ......................................................................................................................... 35
  2.3.1 Flux comparison among cyanobacterial species ...................................................... 35
  2.3.2 The optimal photoaututrophic metabolism ............................................................ 37
  2.3.3 Metabolic features of Syn. 2973 for bioproduction applications ............................ 39
  2.3.4 Subpopulations, metabolic inactive pools and substrate channeling .................. 40
Chapter 3: Cyanobacteria carboxysome mutant reveals the influence of enzyme compartmentations on cellular metabolisms and metabolic network rigidity

3.1 Introduction

3.2 Results

3.2.1 Growth comparison between the control strain and ΔccmKLMN mutant

3.2.2 Biomass composition analysis

3.2.3 INST-MFA reveals flux topology in ΔccmKLMN mutant by INCA

3.2.4 Change in pool sizes in response to carboxysome knock out

3.3 Discussion

3.3.1 Cyanobacteria maintain rigid flux networks despite differing phenotypes

3.3.2 Differences in the assumed metabolic network of PCC 7002 compared to other cyanos

3.3.3 Using carboxysome-deficient mutants to redirect carbon fluxes

3.3.4 Impact of carboxysome mutations on photomixotrophic metabolism

3.3.5 Insights into channeling-based flux regulation

3.3.6 Conclusions

3.4 Materials and Methods

3.4.1 Strains and Cultivation conditions

3.4.2 INST-MFA and 13C Labeling experiments

3.4.3 Metabolite extraction and analysis

3.4.4 Isotopically nonstationary MFA

3.4.5 Biomass composition analysis

3.4.6 GC-MS analysis of proteinogenic amino acids
3.4.7 Measuring photosynthetic parameters and oxygen evolution ........................................68

Chapter 4: Metabolic engineering of cyanobacteria for photoautotrophic production of heparosan, a pharmaceutical precursor of heparin ........................................................................70
4.1 Introduction .......................................................................................................................71
4.2 Materials and Methods .....................................................................................................74
  4.2.1 Chemicals and reagents ...............................................................................................74
  4.2.2 Microorganisms and culture conditions .......................................................................75
  4.2.3 Cyanobacterial plasmid construction and transformation ...........................................76
  4.2.4 Analysis of heparosan production using LC-MS ..........................................................77
4.3 Results and Discussion .....................................................................................................79
4.4 Conclusions .....................................................................................................................86

Chapter 5: Channeling in native microbial pathways: Implications and challenges for metabolic engineering ......................................................................................................................87
5.1 Introduction .....................................................................................................................88
5.2 Kinetic, thermodynamic and biochemistry evidences on channeling .........................91
  5.2.1 Kinetic and thermodynamic implications ....................................................................91
  5.2.2 Compartmentalization and tunneling .........................................................................91
  5.2.3 In vitro experiments on channeling of cascase enzymes ...........................................93
5.3 In vivo channeling studies ...............................................................................................93
  5.3.1 Approaches to reveal in vivo pathway channeling .......................................................93
  5.3.2 Detection of leaky channels via dynamic labeling of intermediates ................................96
  5.3.3 “Squeeze and Pulse” to investigate pathways of high degrees of channeling .............98
5.4 Modeling/theoretical investigation of pathway channeling ............................................99
  5.4.1 Channeling blocks flux towards thermodynamically favorable pathways .................99
  5.4.2 Channeling facilitates transport of metabolites between enzymes .............................100
  5.4.3 Diffusion and reaction rates along EMPP enzymes ...................................................101
  5.4.4 Channeling alters enzyme-metabolite and enzyme-enzyme interactions ................103
5.5 Implications of metabolite channeling ............................................................................105
  5.5.1 Challenges facing synthetic biology ...........................................................................105
  5.5.2 Applications of synthetic channels to benefit bioproductions ..................................105
  5.5.3 The impact of channeling on genome-phenome mapping ........................................107
5.5.4  The impact of channeling on \(^{13}\text{C}\)-MFA .................................................................................. 108
5.6  Conclusions ............................................................................................................................. 110
Chapter 6: Comparative studies of glycolytic pathways and channeling under \textit{in vivo} and \textit{in vitro} modes ......................................................................................................................... 111
6.1  Introduction ................................................................................................................................ 112
6.2  Materials and Methods ............................................................................................................. 113
   6.2.1  Strain construction .............................................................................................................. 113
   6.2.2  Media composition and growth experiments ....................................................................... 114
   6.2.3  Steady state labeling experiments and GC-MS analysis ...................................................... 114
   6.2.4  Isotopically nonstationary labeling experiment and pool size analysis ............................ 114
   6.2.5  Protein purification ............................................................................................................ 115
   6.2.6  \textit{In vitro} assays of EMPP and EDP enzymes ..................................................................... 115
6.3  Results and Discussion ............................................................................................................ 116
   6.3.1  Comparative studies on isoenzymes’ functions under \textit{in vitro} and \textit{in vivo} conditions .... 116
   6.3.2  The effect of ED pathway overexpression on \textit{in vitro} and \textit{in vivo} glycolytic rates ........ 120
   6.3.3  Transient labeling of \textit{E. coli} EMP intermediates .............................................................. 123
   6.3.4  Broader perspective on glycolysisomes ............................................................................. 125
   6.3.5  Conclusions ........................................................................................................................ 126
Chapter 7: Conclusions and future directions ................................................................................. 127
7.1  Conclusion ................................................................................................................................. 127
7.2  Future directions ....................................................................................................................... 129
   2.2.1  Future directions for cyanobacteria cell factories ............................................................. 129
   2.2.2  Future directions for metabolite channeling ....................................................................... 129
8. References ..................................................................................................................................... 132
9. Appendix 1: Supporting Information for Chapter 2 ................................................................. 155
10. Appendix 2: Deciphering cyanobacterial phenotypes for fast photoautotrophic growth .... 181
11. Appendix 3: Supporting Information for Chapter 3 ................................................................. 195
12. Appendix 4: Supporting Information for Chapter 4 ................................................................. 205
13. Appendix 5: Channeling in native microbial pathways ............................................................ 207
14. Appendix 6: Supporting Information for Chapter 6 ................................................................. 217
15. Appendix 7: Comparative studies of glycolytic pathways and channeling ............................. 222
List of Figures

Figure 1.1: Cyanobacteria as cell factories ................................................................. 2
Figure 1.2: Comparison of biofuel production in cyanobacteria versus E. coli .......... 4
Figure 1.3: Schematic representation of flux control mechanisms and analysis tools .......... 6
Figure 1.4: INST-MFA of steady state metabolism ..................................................... 8
Figure 1.5: INST-MFA workflow .............................................................................. 10
Figure 2.1: Growth performances of cyanobacterial strains ........................................ 28
Figure 2.2: Syn. 2973 flux map determined in PBRs and SF and Syn. 7942 from the PBR .... 30
Figure 2.3: Relative pool sizes of Syn. 2973, Syn. 7942, and Syn. 6803 to E. coli K-12 .......... 33
Figure 2.4: Biomass growth of Syn. 2973 in the presence of carbon sources .................. 34
Figure 2.5: Labeling dynamics for Syn. 2973 and 7942 in PBR conditions for citrate, 3PGA, and malate .............................................................. 42
Figure 3.1: Growth curves in low CO₂, low light and high CO₂, high light conditions .... 51
Figure 3.2: Biomass compositional analysis of Syn. PCC 7002 control and ΔccmKLMN .... 52
Figure 3.3: Syn. PCC 7002 control and ΔccmKLMN flux maps as determined by INCA .... 54
Figure 3.4: Photosynthetic capacity of PSII (Fᵥ/Fₘ) and O₂ evolution ........................... 56
Figure 3.5: Comparison of metabolite pool size between Syn. PCC 7002 control and ΔccmKLMN .............................................................. 58
Figure 3.6: Steady state proteinogenic amino acid labeling from labeled glycerol, acetate and pyruvate to monitor photomixotrophic growth ........................................... 63
Figure 4.1: Schematic featuring the biosynthetic pathway of glycoaminoglycans in Syn.PCC 7942, specifically the expression of heparosan and its downstream applications .... 73
Figure 4.2: Diagrams of recombinant Syn. PCC 7942 (pgp7942) grown in various conditions .............................................................. 75
Figure 4.3: Heparosan producing strain, pgp7942, compared to Syn. PCC 7942 WT under continuous light shake flask conditions ........................................ 83
Figure 4.4: Heparosan producing strain, pgp7942, under continuous
250 photons µmol/m²/s (HL) and 0.2% CO₂ (HC) shake flask conditions ........84

Figure 5.1: Mechanisms for detecting substrate channels ........................................95
Figure 5.2: Types of substrate channels .................................................................98
Figure 5.3: Gibbs free energy of reactions in glycolysis.................................100
Figure 5.4: A simple model for molecule diffusion in cells.................................103
Figure 6.1: Glycolytic pathway map in E. coli...................................................113
Figure 6.2: In vitro reaction rates of the EMPP and EDP.................................121
Figure 6.3: Metabolic flux analysis (determined by WuFlux) in different E. coli strains........122
Figure 6.4: Metabolite pool size comparison to WT E. coli K-12 .........................124
Figure 6.5: Labeling dynamics following ¹³C₆-glucose pulse in different E. coli strains........125
List of Tables

Table 1.1: Brief description of quenching methods .........................................................23
Table 4.1: List of primers and sequences used .................................................................77
Table 4.2: Summary of heparosan production in PCC 7942 and engineering 7942 (pgp7942) ..85
Table 5.1: Evidences of enzyme tunneling and channeling ..............................................89
Table 6.1: Glycolysis genes and reported enzyme kinetics ...............................................117
Table 6.2: In vitro test of EMP isoenzymes via a factorial experiment ...............................119
List of Abbreviations

2PG, 2-phosphoglycolate; 2PGA, 2-phosphoglycerate; 3PGA, 3-phosphoglycerate; 6PG, 6-phosphogluconate; ABC-transporter, ATP binding cassette-transporter; Ac, acetate; AcCoA, acetyl-CoA; ADP, adenosine diphosphate; ADP glucose (ADPG), adenosine diphosphoglucose; αKG, alpha-ketoglutarate; AMP, adenosine monophosphate; ASP, aspartate; ATP, adenosine triphosphate; CIT, citrate; DHAP, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; EDP, Entner-Doudoroff pathway; EMPP, Embden-Meyerhof-Parnas pathway; F6P, fructose 6-phosphate; Fba, fructose bisphosphate aldolase; FBP, fructose 1,6-diphosphate; FUM, fumarate; G1P, glucose 1-phosphate; G3P (GAP), glyceraldehyde 3-phosphate; G6P, glucose 6-phosphate; GA, glycerate; GABA, gamma-aminobutyric acid; GalNAc, N-acetylgalactosamine; GLC, glycolate; GlcN-1P, glucosamine-1-phosphate; GlcN-6P, glucosamine-6-phosphate; Gln, glutamine; GLU, glucose; GLX (GOX), glyoxylate; GLYC, glycogen; ICIT, isocitrate; KDPG, 2-dehydro-3-deoxy-phosphogluconate; MAL, malate; MeOH, methanol; MID, mass isotopomer distribution; MFA, metabolic flux analysis; NAD⁺, nicotinamide adenine dinucleotide; NADH, nicotinamide adenine dinucleotide (reduced); NADP⁺, nicotinamide adenine dinucleotide phosphate; NADPH, nicotinamide adenine dinucleotide phosphate (reduced); OAA, oxaloacetate; OPPP, oxidative pentose phosphate pathway; PEP, phosphoenolpyruvate; Pfk, phosphofructokinase; PPI, pyrophosphate; PPP, pentose phosphate pathway; PYR, pyruvate; R5P, ribose 5-phosphate; Ru5P, ribulose 5-phosphate; RuBP, ribulose 1,5-diphosphate; S7P, sedoheptulose 7-phosphate; SBP, sedoheptulose 1,7-diphosphate; SUC, succinate; TCA, tricarboxylic acid cycle; UDP, uridine diphosphate; UDP glucose (UDPG), uridine diphosphoglucose UTP, uridine triphosphate; X5P, xylulose 5-phosphate; WT, wild type
Acknowledgments

I would like to thank first and foremost my advisor, Dr. Yinjie Tang, for his kindness, patience, financial support, and broad bioengineering training. Dr. Tang imparted on me how to present complex data through clear and engaging insights and memorable story-telling analyses, which will aid me greatly in future endeavors. Secondly, I would like to thank my thesis committee: Dr. Himadri Pakrasi, Dr. Tae Seok Moon, Dr. Marcus Foston, and Dr. Doug Allen for their time and assistance throughout my doctoral studies. I would also like to acknowledge the DOE-BER (DESC0012722, a grant project lead by Dr. Pakrasi and Dr. Costas Maranas) for supporting many aspects of my research. Additionally, I would like to acknowledge the DOE Joint BioEnergy Institute supported by the DOE-BER, through contract DE-AC02-05CH11231, and the scientists at JBEI, Dr. Edward E.K. Baidoo and George Wang. Furthermore, I would like to thank the Proteomics and Mass Spectrometry Core Facility at the Donald Danforth Plant Science Center and its scientists Brad Evans and Jon Mattingly. Many LC-MS analyses were in part enabled by PMSF and major research instrumentation grants from the NSF (DBI-1427621).

I would like to particularly thank Dr. Doug Allen for metabolic flux analysis expertise, mass spectrometry training, countless discussions about my research, and valuable guidance throughout my studies. I would also like to members of his lab (Dr. Fangfang Ma, Dr. Kevin Chu, and Dr. Shrikaar Kambhampati) for their help and mentoring. I am very grateful to Dr. Himadri Pakrasi and his lab for the generous use of their lab equipment for cyanobacterial cultivations and guidance with photosynthetic organisms. Dr. Jingjie Yu, Dr. Justin Ungerer, Dr. Michelle Liberton, Dr. Cory Knoot and Virginia Johnson were all instrumental in helping me throughout my studies and without none of this work would be possible without them. I would also like to thank Dr. Marcus Foston and his laboratory for use of their incubator to complete my
experiments. I have tremendous gratitude towards my peers and lab members at Washington University for their support, collaborations, and assistance: Dr. Lian He, Dr. Whitney Hollinshead, Dr. Ni Wan, Tolutola Oyetunde Jeffrey Czajka, Drew DeLorenzo, Dr. Cheryl Immethun. I am also very appreciative of the EECE office staff who made this process as easy as possible. Finally, I would like to thank my family, friends and dog, Abita, for providing constant encouragement and motivation throughout my time at Washington University.

Mary H. Abernathy

Washington University in St. Louis

December 2018
ABSTRACT OF THE DISSERTATION

Investigating Cyanobacteria Metabolism and Channeling-based Regulations via Isotopic Nonstationary Labeling and Metabolomic Analyses

by

Mary H. Abernathy

Doctor of Philosophy in Energy, Environmental & Chemical Engineering

Washington University in St. Louis, 2018

Yinjie Tang, Chair

Cyanobacteria have the potential to be low-cost and sustainable cell factories for bio-products; however, many challenges face cyanobacteria as biorefineries. This dissertation seeks to advance non-model photosynthetic organisms for biotechnology applications by characterizing central carbon metabolism and its regulations. Cyanobacteria phenotypes for bio-production are examined and their intracellular metabolism is quantified. Using isotopic labeling experiments, phenotypic relationships between biomass composition, central carbon fluxes, and metabolite pool sizes are investigated. Metabolic analyses of cyanobacteria led to new investigations of flux regulation mechanisms via protein spatial organizations or metabolite channeling. Metabolite channeling is further explored as a hypothesis to explain enigmatic labeling patterns and as a method to organize and regulate enzymes for robust central metabolisms. The insights reveal strategies for redirecting central metabolic fluxes for value-added chemicals as well as broad impacts for intracellular modeling approaches.

First, Synechococcus UTEX 2973 was probed with isotopic nonstationary metabolic flux analysis under changing growth conditions. Despite similar genetics to Synechococcus 7942, Synechococcus UTEX 2973’s exhibits a fast growth phenotype with greater carbon fixation driven by higher energy charges, optimal ATP/NADPH ratios, low glycogen production during
exponential growth, and a central metabolism that reduces CO₂ loss. Unusual labeling patterns indicated metabolite channeling as a possible flux regulation mechanism. As cyanobacteria are known to have carboxysomes, a microcompartment that concentrates CO₂ for RuBisCO, it was hypothesized that carboxysome mutants may reveal channeling mechanisms. Carboxysome-free mutants (high CO₂ requiring phenotypes) were found to accumulate metabolites and reach higher steady state ^{13}C enrichment, indicating more homogenous cytoplasms. Carboxysome-free mutants may provide a method for unlocking cyanobacteria flux constraints, reducing catabolic repression, and providing a way to contain genetically modified cyanobacteria.

To ease the constraints of highly regulated and complex metabolic networks, platform or non-model strains can be used to provide a good starting point for small molecules of interest. To take advantage of cyanobacterial native sugar phosphate metabolisms, Synechococcus was engineered for the photoautotrophic production of a high-value polysaccharide, heparosan, which is an unsulfated polysaccharide important for cosmetic and pharmaceutical applications. Via overexpressing two key enzymes, the recombinant strain improves heparosan production by over 50 folds. Synechococcus was also found to naturally synthesize multiple glycosaminoglycans.

Lastly, to further explore metabolite channeling as evidenced by isotopic labeling patterns, we developed cell-free glycolysis pathways and compared their performance with in vivo glycolysis functions in E. coli and its PTS mutants. Enzyme assays, dynamic metabolite labeling and flux analysis further confirmed the hypothesized channel of EMP enzymes where the PTS may be an anchor point to initiate enzyme assemblies. In summary, the outcomes of this thesis provide new insights into non-model phototrophic microbial chassis, reveal flux control mechanisms beyond genetic or transcriptional regulations, and offer practical guidelines for sustainable bio-production via synthetic biology approaches.
Chapter 1: Introduction

This chapter was partially adapted from the following publications:


Abstract

Cyanobacteria are promising platforms for carbon neutral bioproductions. However, cyanobacteria face numerous challenges, such as slower growth than their heterotrophic counterparts, complex carbon and energy metabolisms, and tightly-wired cellular regulation.

Using isotopically nonstationary $^{13}$C flux analysis, which is a powerful tool for precise carbon flux quantification (e.g., metabolic flux analysis of photoautotrophic organisms), we can investigate pathway bottlenecks, a cell’s phenotype, and metabolite channeling. In general, isotopically nonstationary metabolic flux analysis requires three main components: (1) transient isotopic labeling experiments; (2) metabolite quenching and isotopomer analysis using LC-MS; (3) metabolic network construction and flux quantification. Labeling dynamics through cascade metabolites may also identify channeling phenomena in which metabolites are passed between enzymes without mixing with the bulk phase. In this chapter, experimental protocols are outlined to probe metabolic pathways through dynamic labeling. Protocols are described for labeling experiments, metabolite quenching and extraction, LC-MS analysis, computational flux quantification, and metabolite channeling observations.
1.1 Cyanobacteria as cell factories for sustainable production of biofuels, bioplastics, and biopharmaceuticals from light energy and carbon dioxide.

1.1 Cyanobacteria as cell factories

1.1.1 Inherent challenges facing cyanobacteria bioproductions. *E. coli* and yeast are the most widely used microbial cell factory due to their tractability, fast growth, and easy genetic manipulation (Nielson and Keasling, 2016). However, glucose as the sole source of fermentable sugar is costly, and new technology of lignocellulose utilization as a source of feedstock is still very expensive. Cyanobacteria have the potential to be carbon neutral cell factories and are of special biotechnological interest due to their fast growth compared to other photosynthetic organisms, unicellular anatomy, and high photosynthetic efficiency. Since the economic and environmental benefits of CO₂ utilization are vast, engineered cyanobacterial biorefineries were
first developed for sustainable ethanol production in 1999 (Sayre, 2010). More and more genetic parts have been developed for cyanobacterial systems in the past decade, including new promoter systems and genome databases. Currently, metabolic engineers can manipulate cyanobacterial metabolism by eliminating competing pathways, redirecting biosynthesis fluxes, and balancing cofactors. Using CRISPR-Cas (Clustered Regularly Interspaced Short Palindrome Repeats) systems, cyanobacteria genomes can be efficiently and precisely edited for multiple genomic modifications (Scholz et al., 2013; Wendt et al., 2016; Ungerer and Pakrasi, 2016).

Cyanobacteria can naturally produce or have been engineered to produce numerous biofuels (Wang et al., 2013b), bioplastics (Kamravamanesh et al., 2018) and biopharmaceuticals (McPhail et al., 2007; Singh et al., 2005) (Figure 1.1). Figure 1.2 compares *E. coli* and cyanobacteria for biosynthesis of alcohols and fatty acid derived compounds. It is clear that modern molecular biology can engineer cyanobacteria to produce chemicals from CO₂ with similar carbon yields as *E. coli* glucose fermentations. However, *E. coli* still outperforms cyanobacteria with production rates and titers (by ten to hundreds of folds) due to its heterotrophic nature. As engineered pathways become more complicated, cyanobacterial biosynthesis pathways unavoidably face shortages of both energy molecules and building blocks. More knowledge is ultimately needed about how cyanobacteria coordinate their energy and carbon metabolisms to aid biorefineries. Photo-fermentation, increased photosynthetic efficiency, and more insights into cyanobacteria metabolic regulations may help remove some of the roadblocks for low production titer and rates.
Figure 1.2 Comparison of biofuel production in cyanobacteria versus *E. coli*. Yields for cyanobacteria were calculated by milligrams of carbon from the product divided by gram CO$_2$ fixed. It is assumed carbon content was approximately 50% of the biomass. Yields for *E. coli* are calculated by gram of product per gram of glucose used in the culture media. It is assumed that an OD$_{730}$ and OD$_{600}$ of 1 is equal to 0.4 g of dry cell weight (DCW). The data collected is from published reports (Atsumi et al., 2009; Ruffing, 2014; Gao et al., 2012; Trinh et al., 2008; Baez et al., 2011; Tan et al., 2011; Zheng et al., 2012; Wang et al., 2013b; Schirmer et al., 2010).

1.1.2 Rewiring cellular metabolism is difficult. Synthetic biology has made great progress in establishing molecular tools for genome editing and controlling gene expression. However, redirecting metabolic fluxes for effective bioproduction remains difficult because hosts often naturally suppress heterologous pathways and have undesired metabolic shifts, resulting in low titers and production rates. Metabolic fluxes do not always directly correlate with metabolite concentrations or gene/ enzyme expressions. In reality, cell metabolism is highly regulated and coordinated through multiple interactions between genes, enzymes and metabolites (Figure 1.3). Substrate level regulations modulate the catalytic activity of enzymes via metabolite-protein interactions, namely allosteric inhibition (Piazza et al., 2018). In general, metabolite-protein interactions are highly conserved across evolutionarily divergent species for the maintenance of metabolic homeostasis (Reznik et al., 2017), and this regulation is responsible for fast metabolic
dynamics under changing growth environments (Khodayari and Maranas, 2016). Additionally, gene-protein interactions affect metabolic fluxes by regulating enzyme expressions. Transcriptional regulation is a commonly employed mechanism used to orchestrate metabolic responses (Utrilla et al., 2016); however transcriptional fold changes do not always correlate with metabolic fluxes (Schwender et al., 2014). Less experimentally explored, enzymes may dynamically associate via pairwise interactions to form “pearl necklace like” channels or via non-pairwise interactions to form enzyme clusters (Sweetlove and Fernie, 2018). Unlike transcription or protein expressions that often show significant variations, channeling provides flux constraints under environmental/genetic perturbations (Fischer and Sauer, 2005; Fischer et al., 2004). Because small metabolite diffusion is considered sufficiently fast for enzyme kinetics (Abernathy et al., 2017b), channeling may increase reaction rates, but it is more likely channeling functions as a regulatory role in controlling metabolic fluxes and protecting reaction intermediates from competing reactions or harsh cytosol environments (Sweetlove and Fernie, 2018). Although it has been recognized that enzymes can organize spatially and temporarily to shuttle intermediates without releasing them into the cytosol, it is difficult to validate weak transient interactions between enzymes under in vitro (cell free) conditions (Zhang, 2011), thus proximity channeling in bacteria remains controversial (Wheeldon et al., 2016). As for cyanobacteria, carboxysome microcompartmentalization and photosystems co-localization on membranes leads to heterogeneous distributions of enzymes inside the cell. It is likely that central metabolite intermediates are passed without being released into the bulk cytosol phase, as revealed by our paper (Abernathy et al., 2017a).

1.1.3 Channeling has significant biotechnology impacts. Metabolite channeling is integral to biotechnology applications because channeled pathways would constrain and repress
heterologous pathways (van Winden et al., 2001). Metabolite channeling increases the robustness of native pathways under environmental and genetic perturbations (Fischer et al., 2004; Fischer and Sauer, 2005) and may prevent native pathways from releasing intermediates as the precursors for engineered pathways.

Figure 1.3 Schematic representation of flux control mechanisms and analysis tools.

Understanding channeling-based regulation can help resolve inferior efficiencies in heterologous pathways by making proper co-localizations of engineered enzymes (Dueber et al., 2009; Lee et al., 2013). Moreover, the heterogeneous distribution of intracellular enzymes may also confound kinetic modeling and multiple-omics analyses. Since metabolic flux analysis (MFA) relies on the assumption that all metabolites are homogenously distributed in the cell and no diffusion limitations exist, metabolite channeling may cofound metabolic flux values or be revealed during MFA. Understanding the scope and the rules dictating the formation of metabolite channeling in native pathways may improve genome-to-phoneme mapping, pathway engineering, and metabolic modeling (van Winden et al., 2001; Antoniewicz et al., 2007). Greater cognizance of channeling will enable bio-systems engineers to add an extra level of control on cellular
metabolic fluxes to current methods that rely mostly on gene copy numbers and transcriptional and translational regulations.

1.2 Isotopically nonstationary metabolic flux analysis as a tool for decoding cellular phenotypes

$^{13}$C-metabolic flux analysis (MFA) provides quantitative information about carbon flux through metabolic networks. Traditional $^{13}$C-MFA relies on the labeling of proteinogenic amino acids, which have slow turnover rates and require isotopically steady state tracer experiments. On the other hand, transient labeling dynamics of fast-turnover central metabolites can provide snapshots of cell metabolism and offer rich insights into pathway activities. Using experimentally measured labeling dynamics of free metabolites, isotopically nonstationary MFA (INST-MFA) has been developed (Antoniewicz, 2015a). Unlike isotopically steady state MFA (cells are cultured with a specific $^{13}$C-tracer for a long period), INST-MFA requires measuring the change of free metabolite or amino acid labeling after a cell’s culture has been switched from non-labeled medium to labeled medium (e.g., pulse $^{13}$C-substrates into non-labeled cultures). Since metabolite labeling speed is correlated to flux rate and metabolite pool sizes, the knowledge of labeling dynamics of key metabolites can be used to decipher metabolic fluxes and pool sizes through a metabolic network.
**Figure 1.4 INST-MFA of steady state metabolism.** Left pane: after $^{13}$C-substrate pulse ($C_{IN}$), the flux rate through a key metabolite ($C_A$) determines the labeling rate of $C_A$ as seen below. Right pane: in the top panel, the shaded circle represents the measurement time points for INST-MFA, while the non-shaded circle represents measurements taken at isotopic steady state. The dotted line represents metabolic fluxes at steady state. An example of an isotopic labeling experiment is shown in the bottom pane. When labeled CO$_2$ is introduced into an algal culture, Calvin cycle metabolites become progressively labeled. The labeling dynamics for 3PGA is illustrated. Once isotopic steady state is reached, the metabolites will be fully labeled.

INST-MFA has been applied to photoautotrophic organisms that can only utilize CO$_2$ because steady state CO$_2$ labeling provides no unique isotopomer data to gain flux information (Young et al., 2011). To perform INST-MFA, transient labeling data are collected immediately following the introduction of a $^{13}$C-tracer for multiple time points (Figure 1.4; shaded circle). Due to fast metabolite turnover rates, a protocol to rapidly and precisely quench cell metabolism is required. The metabolite labeling patterns are commonly analyzed using LC/MS (Bennette et al., 2011). The labeling dynamics can be used to quantify metabolic fluxes, assuming: 1) the
cells are at metabolic steady-state; 2) there is no enzyme preference between $^{12}$C and $^{13}$C atoms; 3) the cell is homogeneous (no diffusion limitations exist) (Zamboni et al., 2009). For flux estimation, a non-linear least squares regression is solved where the objective function is to minimize the sum of squared residuals (SSR) between the measured (m) and predicted mass isotopomer distributions ($x(t)$), pool size ($C$), and metabolic rates ($r$) (Antoniewicz, 2015a).

\[ \text{Equation 1.1:} \quad \min SSR = \sum \frac{(x(t) - x_m(t))^2}{\sigma_x^2} + \sum \frac{(r - r_m)^2}{\sigma_r^2} + \sum \frac{(C - C_m)^2}{\sigma_C^2} \]

s. t. \( C \times \frac{dx}{dt} = f_{\text{isotopomer model}}(v, x(t)) \)

\[ R \times v = r \]

\[ S \times v = 0 \]

The metabolic network and fluxes ($v$) are subject to constraint by the stoichiometry matrices ($S$ and $R$) and external metabolic rates ($r$). Metabolite pool sizes ($C$) can be measured or fitted using isotopomer data (Young, 2014). The elementary metabolite unit (EMU) framework can be used for effective isotopomer calculations (Antoniewicz et al., 2007; Young et al., 2008). INCA (Young, 2014) and OpenMebius (Kajihata et al., 2014) are excellent dynamic flux analysis softwares, which resolve fluxomes experimentally determined time-dependent labeling data. The workflow of INST-MFA is shown in Figure 1.5.
Figure 1.5 INST-MFA workflow. The workflow consists of three major parts: (1) experiment design, and isotopic labeling experiments (ILE); (2) metabolite extraction and dynamic labeling data analysis; (3) metabolic network construction and quantification of fluxes.

Isotopic nonstationary labeling provides an enhanced understanding of exchange fluxes and metabolite pool sizes (Schaub et al., 2008) and is particularly useful for studying mammalian cells and other higher organisms (Wiechert and Nöh, 2013; Niklas et al., 2011; Nargund et al., 2014). INST-MFA provides reliable and experimentally determined quantitative measurements of metabolic fluxes (Schaub et al., 2008; Wahl et al., 2008). INST-MFA is also useful for elucidating hidden pathway constraints such as metabolic reflux (Nargund et al., 2014), pathway bottlenecks, and metabolite channeling (Hollinshead et al., 2016). Metabolite channeling creates two metabolite pools: a fast turnover pool (inside the channel) and an inactive pool (outside the channel). In this case, INST-experiments may generate unsequential labeling along cascade metabolites (i.e., downstream metabolites are labeled faster than their precursors).
1.2.1 Materials associated with INST-MFA

1 Transient isotope labeling experiments

1.1 Cell culture at metabolic steady-state: Minimal media cultures in their exponential growth phase at metabolic steady-state. Prepare regular (for seed cultures) and minimal cell media (see Note 1) specific to the strain being cultured. The volume of culture needed for the experiment is dependent on the strain’s cell density, number of experimental time points and sampling procedure (see Note 2). Follow specific media instructions for storage conditions.

1.2 $^{13}$C-tracer: Prepare a stock solution of the $^{13}$C-tracer (e.g., NaH$^{13}$CO$_3$ or U-$^{13}$C-glucose) using equation 2 below, where $C_{\text{culture}}$ is the desired final culture concentration (see Note 3), $C_{\text{stock}}$ is the concentration of the stock solution being prepared, $V_{\text{culture}}$ is the volume of the culture(s), and $V_{\text{stock}}$ is the volume of stock solution being pulsed (see Note 4).

\[ C_{\text{culture}} \times V_{\text{culture}} = C_{\text{stock}} \times V_{\text{stock}} \]

Equation 1.2:

1.3 Quenching supplies: Minimal (no carbon) media or phosphate buffered saline and dry ice/liquid N$_2$ (see Note 5).

1.4 Bench-top centrifuge with refrigeration options: Pre-chill centrifuge prior to the experiment (make sure the centrifuge temperature close to 0 °C). Pre-chilled centrifuge tubes (50 mL or 15 mL tubes) are used.

2 Analytical determination of Mass Isotopomer Distributions

2.1 Extraction solution: The choice of extraction solution depends on the species and the metabolites of interest (see Note 6). Additionally, the volume of extraction solution can be optimized.

2.2 De-ionized water.
2.3 Centrifuge filters for LC/MS analysis: Amicon Ultra centrifuge filters-3kDa (EMD Millipore, USA), Costar Spin-X centrifuge tube filters-0.45 μm cellulose acetate membrane (Corning, USA) or centrifugal filter nylon-0.45 μm (VWR, USA) are recommended depending on instrument technique and metabolites of interest.

2.4 Vortexer/Lyophilizer/vacuum

2.5 Liquid chromatography-mass spectrometer (LC/MS): Two examples of LC settings are following. All solvents and reagents used should be of HPLC grade. (a) Reverse phase-ion pairing: 10mM tributylamine and 15mM acetic acid in water (mobile phase A), 100% HPLC grade methanol (mobile phase B), Phenomenex Synergi Fusion-RP column or an alternative column. (b) Hydrophilic interaction liquid chromatography (HILIC): 10 mM ammonium carbonate in water (mobile phase A) and 10 mM ammonium carbonate in 70% acetonitrile and 30% water (mobile phase B), SeQuant ZIC-pHILIC column (EMD Millipore) or an alternative HILIC column. High-resolution mass spectrometers (e.g. TOF or Orbitrap) and/or triple-quadrupole instruments for multiple reaction monitoring (MRM) analysis are recommended for reliable metabolite identification. A MS analysis software will be needed along with other LC/MS consumables.

2.6 Internal standards: Approximately 1 μg/mL analytical metabolite standards should be prepared to aid in peak identification. Additionally, standards should be spiked into biological samples and analyzed to observe matrix effects.

3 Metabolic flux analysis
3.1 Computer and software: Equipped with the ability to perform isotopically nonstationary metabolic flux analysis and run appropriate software, such as INCA (isotopomer network compartmental analysis) (Young, 2014).

3.2 Biomass composition analysis: Dried biomass washed with 0.9% saline solution, analytical instruments and their associated reagents. Common measurements are total protein content, amino acid composition, lipid measurement, and carbohydrate measurement. National Renewable Energy Laboratory developed procedures for standard biomass analysis (http://www.nrel.gov/bioenergy/biomass-compositional-analysis.html) (optional).

3.3 Growth rate analysis: a growth curve for the species of interest measured at the appropriate absorbance using a plate reader or UV/Vis spectrometer (optional).

3.4 Carbon uptake and excretion measurements: a steady state culture, a HPLC, GC, GC-MS and/or enzymatic kits, appropriate solvents, consumables, and sampling equipment.

1.2.2 Methods for performing INST-MFA

1 Transient isotope labeling experiments and fast-cooling protocol

1.1 Prepare cultures. Depending on the strain’s cell density and number of time points, prepare proper amount of non-labeled cultures (see Note 2). To capture the labeling dynamics where the slope is changing rapidly (Ma et al., 2014) (i.e., the earlier time points), numerous time-course samples have to be taken after $^{13}$C-pulse experiments (see Note 7).

1.2 Prepare $^{13}$C tracer stock. For example, NaH$^{13}$CO$_3$ is used to probe photoautotrophic cultures, and U-$^{13}$C-glucose is often used to study heterotrophic organisms (see Note 8).
1.3 Set up for tracer experiment. Before \(^{13}\)C-pulse, obtain dry ice or liquid nitrogen and have the quenching solution and centrifuge tubes pre-chilled in the appropriate apparatus. Then, deliver the \(^{13}\)C-substrate into the culture and begin timing.

1.4 Metabolite quenching. At the measured time point, sample the aliquot of culture. If using a quenching solution, the culture should be quickly mixed in with the quenching solution (e.g., an icy cold buffer) then the sample tubes should be submerged in liquid nitrogen (or dry ice) to cool cultures to \(\sim 0 \, ^\circ\text{C}\) as fast as possible (Hollinshead et al., 2016) \((\text{see Note 9})\). If the quenching temperature is not low enough, central metabolic reactions may be active during sampling processes. Therefore, it is important to quickly quench cell metabolism at \(\sim 0 \, ^\circ\text{C}\) using both liquid nitrogen and cold buffer.

1.5 Centrifuge cell culture to pellet the quenched biomass at \(0 \, ^\circ\text{C}\). Use appropriate speed (e.g., \(6000 \times g\)) and time to pellet cell biomass, while trying to keep centrifuge time short. Cell pellets can be stored at \(-80 \, ^\circ\text{C}\) before extraction.

2 Analytical determination of Mass Isotopomer Distributions via LC/MS

2.1 Extract metabolites from cell biomass. While on ice, add internal standard to cell pellet or biomass \((\text{see Note 10})\) if trying to quantify metabolites. Mix 1 mL of cold extraction solution \((\text{see Note 6})\) with biomass then transfer to a new vial. Shake for 4 hours to overnight at \(\sim 0 \, ^\circ\text{C}\). After shaking in extraction solution, add 0.5 mL of ddH\(_2\)O, vortex, and spin down at 5,000 x g for 2 min. Pipet out methanol-water layer into a new vial. Repeat extraction with ddH\(_2\)O. Filter methanol-water layer using centrifuge filter and transfer filtered sample into a new vial.
2.2 Concentrate metabolite extraction. Dry down water-methanol samples using lyophilizer/vacuum or N₂ gas. Re-suspend the sample in ddH₂O with 100 μL ddH₂O (or ddH₂O-acetonitrile solution, if also looking at water insoluble compounds using the HILIC-MS method).

2.3 LC/MS Analysis (see Note 11). Filter all mobile phases before LC/MS analysis. A brief description of reverse phase ion-pairing LC/MS operation is as follows: A recommended injection volume of 10 μL and a column flow rate of 0.2-0.3 mL/min may be used. An LC gradient as follows is suggested: 0% of mobile phase B (0 min), 8% B (8 min), 22% B (18 min), 40% B (28 min), 60% B (32 min), 90% B (34 min), 0% B (39 min), and 0% B (49 min) (Young et al., 2011; Ma et al., 2014). The mass spectrometer usually operates with an ESI source under negative ionization mode for analysis of organic acids and sugar phosphates. In targeted metabolite analysis by MRM mode, first optimize the collision energy and define MRM transitions for each metabolite using its chemical standard.

A brief description of HILIC- LC/MS protocol developed by the Joint Bioenergy Institute (Hollinshead et al., 2016) is as follows: A HILIC column, such as a SeQuant ZIC-pHILIC (2.1mm x 150mm and 5-μm particle size), should be used. For the ZIC-pHILIC column, the recommended flow rate and injection volume are 0.2 mL/min and 1-2 μL. Metabolites are separated via gradient elution under the following conditions: 100% B (0 min), 60% B (9 min), 60% B (11.8 min), 100% B (12 min), and 100% B (22 min). Additionally, optimize LC and MS conditions with standard solutions.

2.4 Data analysis. The metabolite labeling pattern is analyzed as mass isotopomer distributions (MID). For example, M+0 of 3-Phosphoglyceric acid (3PGA) represents the fraction of unlabeled 3PGA, whereas M+3 represents the fraction of a fully labeled 3PGA (Figure 1.4). A
metabolite has $2^n$ isotopomers, where $n$ represents the number of carbons. To determine the fractions of all isotopomers, high resolution MS methods have to be employed for distinguishing authentic labeled isotopomers from ion interferences. Internal standards and technical replicates are essential for confident identification of each metabolite, integration of all isotopomers (M+0, M+1, M+2, etc.), and an assessment of analytical variations. Metabolite identification is usually based on LC retention time, high-accuracy mass measurement, or specific MRM transitions. Mass spectral libraries such as METLIN or the NIST Mass Spectral Database may be used for compound structural assignment when standards are not available (Jazmin and Young, 2013; Schultz et al., 2013). Next, the MID values can be calculated based on the isotopic envelop detected for each metabolite. The M+0 distribution will be the peak area of M+0 divided by the sum of all the isotopomer peak areas for that metabolite. In principle, the sum of all the MID percentages for a metabolite should be equal to one. Biological replicates are needed for deriving the standard deviation for each MID distribution. More accurate MID results can be acquired if using specific algorithms to normalize the natural isotope occurrence (Niedenführ et al., 2016). However, this type of normalization is not normally required for LC/MS data processing as the natural abundances of labeled isotopes are much lower than the experimental variation.

3 Metabolic flux analysis

3.1 INST-MFA platform. OpenMebius (Kajihata et al., 2014), INCA (Young, 2014), or another appropriate platform can be utilized. INCA integrates transient balances using a custom linear ODE solver that involves computing the matrix exponential of each state-transition matrix. INCA allows for the compartmentalization of metabolites, fitting of metabolite pool sizes, and use of dilution parameters if one observes dilution from unlabeled sources, such as glycogen,
biomass reflux, or metabolite channeling. OpenMebius is an open source software, but requires pool size information for precise flux estimation.

3.2 Metabolic network reconstruction. First, build a network based on the genome of the species of interest and biochemical literature. Useful databases are KEGG, BRENDA, and BioCyc Database Collection. The equations should contain information on reaction stoichiometry, reversibility, and atom transitions. The metabolic network should contain enough information to resolve simulated measurements based on all experimental measurements while also avoiding unnecessary reactions that can lead to overfitting the model. Linear pathways may be grouped to reduce computational costs (see Note 12). Energy and redox balancing reactions are often omitted from the model.

3.3 MID data and biomass equations. Flux calculation requires MID data and a biomass equation. The biomass equation represents a proportioned sink (in millimoles per gram of biomass) of all the metabolites required for biomass growth (Antoniewicz, 2015a; Shastri and Morgan, 2005). The flux values are affected by biomass equations. Although a standard biomass equation is often used as an approximate for other bacterial species, a non-model species may have different biomass compositions (such as amino acid composition, lipids, and carbohydrates) (Pramanik and Keasling, 1997). For example, cyanobacteria tend to have high protein content due to the abundance of its photosystems.

3.4 Substrate uptake, extracellular metabolite excretion and growth rate. Additional measurements of inflows and outflows offer constraints for precise flux analysis. Carbon uptake is expressed in units (mmol/gDCW/h). Extracellular metabolite excretions constrain the outflows of metabolic network. Measured growth rate (h⁻¹) can be used to validate flux model predictions. These constrains will improve the accuracy of the model.
3.5 Estimation of fluxes and pool sizes. INST-MFA model obtains flux solutions by minimizing the difference between measured and simulated mass isotopomer distributions. It is recommended to restart the flux estimation numerous times from different random starting points to obtain a global solution (Crown and Antoniewicz, 2013). The goodness-of-fit is based on the sum of squared residuals (SSR) and the degrees of freedom. The confidence intervals of flux values are determined through either parameter continuation or Monte Carlo methods. Parameter continuation is performed using a flux coupling analysis and predictor-corrector method of minimized computational expense. The Monte Carlo method introduces random errors from a normal distribution and flux measurements are repeated with perturbed data sets. In general, parameter continuation is a more efficient method for calculating flux confidence intervals (Antoniewicz et al., 2006). Moreover, pool size measurement is important for dynamic flux analysis. However, precise pool size analysis is difficult because of metabolite leakage during sample processing and the requirement of internal standards. To overcome this problem, the INCA platform fits the metabolite pool sizes based on isotopomer data. In this case, INST-MFA can also generate a set of predictions of metabolite pool sizes in addition to flux values.

3.6 Model Assessment. One should look at reactions with large confidence intervals, which indicate the flux cannot be determined precisely. Additional measurements of metabolite pool sizes may improve the flux determinations. A poor squared residual or weighted residual for an individual metabolite could mean there is an error with the measurement or there is a mistake in the metabolic network. Furthermore, it is important to check whether the regression residuals are distributed evenly. Statistical analysis can reveal the importance of a particular measurement for determining certain flux values.
1.2.3 Consideration of metabolite channeling in INST-MFA. During channeling, metabolites can be passed through a pathway without mixing with the bulk phase. Channeling may change molecule labeling orientations or bypass the denoted pathway reactions. Certain metabolic nodes of a channeled pathway must be leaky for connecting other pathways that share the same intermediates. At a leaky node, the intermediate can diffuse outside the channel and form an inactive cytosol pool, which will have a slower turnover rate. In this case, dynamic $^{13}$C-pulse experiments may be used to identify a “channeling signature”: downstream metabolites are labeled faster than their precursors (Hollinshead et al., 2016). Dynamic $^{13}$C labeling can elucidate leaky channels by observing whether there is any unsequential order for labeling cascade metabolites. To date, there is little knowledge of how microbes channel their native pathways. Metabolic flux analyses (including INST-MFA) are highly sensitive to the effect of channels on molecule orientation and pathway preferences (Williams et al., 2011). Any observation of channeling in dynamic labeling experiments should provide caution during construction of metabolic pathways and analysis of labeling data.

Channeling establishes microcompartmentation of certain metabolites, which must be considered in MFA. If unsequential labeling of cascade metabolites occurs due to channeling effects (presence of multiple pools of same metabolites), it may cause issues obtaining an accurate and best-fit solution. To improve model fitting quality, dilution parameters are used to account for inactive metabolite pools outside of channels (Young et al., 2011). While the introduction of these dilution factors helps reduce model fitting errors, it also introduces addition degrees of freedom into the model. This may increase undetermined parts of the metabolic network (van Winden et al., 2001). In INST-MFA, dilution parameters can be introduced into the model (Kelleher and Masterson, 1992; Antoniewicz et al., 2007). For example, let $C_B$ be the
active pool of metabolite B (the pool involved in the cascade reaction) and \(C_{B'}\) be the inactive pool of B. \(C_{B,\text{bulk}}\) is the total sampled pool (mix of \(C_B\) and \(C_{B'}\)). Then three reactions can be added to the model in such form: (1) \(C_B \rightarrow C_{B,\text{bulk}}\) (2) \(C_{B'} \rightarrow C_{B,\text{bulk}}\); (3) \(C_B \rightarrow C_{B'}\).

1.2.4 Notes on INST-MFA

1. Minimal cell media should not include any carbon sources which the cell may uptake in addition to the \(^{13}\text{C}\)-tracer. While a rich media may be required to analyze the cell at the desired phenotype, this will create difficulties in analyzing the results later due to multiple nutrition compounds in the culture medium interfering with the intracellular analytes (Madji-Hounoum et al., 2016). If using rich media, cells are often washed with a saline solution before extraction (Murphy et al., 2013).

2. Depending on a strain’s cell density, different volumes of cell broth may be needed to obtain quality analytical data. For example, typically 5 mL of log phase \(E.\ coli\) \((\text{OD}_{600} = 1)\) is sufficient for LC/MS samples (Oldiges, 2004), while 25 mL of log phase cyanobacteria \((\text{OD}_{730} = 0.4)\) may be needed for improved LC/MS peak quality.

3. The final concentration of the substrate added can be estimated from literature uptake values and from the longest time point planned. To achieve metabolic steady-state, the cell should be saturated with labeled substrate during the entire experiment so as to not affect the uptake and growth rates (Nöh et al., 2006).

4. It is best to perform all actions as swiftly as possible: 1) Make sure to have the appropriate tools to inject the entire \(^{13}\text{C}\)-stock solution into culture as quickly as possible; 2) High concentration \(^{13}\text{C}\)-substrate stock solution is preferred but may be limited by substrate solubility.
5. INST experiments require proper quenching methods (Faijes et al., 2007; Chen et al., 2014; Millard et al., 2014). In this paper, we focus on a fast-cooling method (Table 1.1), which allows for quenching with high time precision and minimal metabolite leakage. Other methods may also be applicable for rapid quenching procedures.

6. The extraction solution should be optimized to the species and metabolites of interest based on peak height, area, or ion count (Bennette et al., 2011; Bajad et al., 2006). Common extraction mixtures are methanol/water, chloroform/methanol, acetonitrile/methanol/water (Prasad and Ferenci, 2003; Madji-Hounoum et al., 2016). Additionally, if one is interested in nucleotide triphosphates, it has been shown that the addition of 0.1M formic acid in acetonitrile improves extraction (Rabinowitz and Kimball, 2007; Bennett et al., 2008) (after which ammonium hydroxide is combined with the extract to prevent metabolite degradation through neutralization). The amount of extraction solution used is typically between 0.5 and 1.5 mL and is also open to optimization.

7. One can optimize the sampling time points to improve computational quality (Nöhl et al., 2006; Nöhl et al., 2007). In general, it is desired to take more time points during the period of fast labeling changes. One group calculates sampling times through the parameterization of the time points (Nöhl et al., 2006) (refer to cited publication for more details).

8. Parallel labeling experiments can provide temporal resolution of certain fluxes and also improve flux precision to a certain point (Antoniewicz, 2015b). INCA contains a feature that allows a user to search for optimal parallel tracer combinations that minimize the optimality criterion (Young, 2014; Möllney et al., 1999).

9. Use a digital thermometer as the stirring apparatus to help mixing, prevent freezing and inform when the cell culture is quenched (< 0 °C).
10. Possible internal standards: uniformly isotopically-labeled metabolites, PIPES (parent mass-301) (Ma et al., 2014), Naphthyl phosphate (Schaub et al., 2008), N-acetyl-glutamine (Bajad et al., 2006), 2-amino-3-bromo-5-methylbenzoic acid (Baran et al., 2010).

11. Hydrophilic interaction chromatography (HILIC) can achieve separations of common polar metabolites, such as low abundant metabolites, organic compounds, and amino acids (Bajad et al., 2006), while reverse phase ion-pairing chromatography is particularly good for measuring sugar phosphates (Ma et al., 2014; Young et al., 2011; Arrivault, 2009). A triple-quadrupole (QqQ) MS instrument can detect a pre-selected set of targeted metabolites by MRM, which offers supreme sensitivity and accurate quantification, while high-resolution MS provides highly specific analysis of low concentration intracellular metabolites (Young et al., 2011; Schultz et al., 2013). A common soft ionization technique used in MS analysis is electrospray ionization (ESI), which is particularly useful for chemically unstable metabolites, such as nicotinamide adenine dinucleotide phosphate (NADPH), adenosine triphosphate (ATP), and phosphoenolpyruvate (PEP) (Ma et al., 2014). Depending on metabolites of interest and available instruments, multiple combinations of LC and MS systems could be considered.

12. The oxidative pentose phosphate pathway with atom transitions (represented by abcedf) can be grouped as glucose-6-phosphate (G6P) (abcdef) → ribulose-5-phosphate (RU5P) (bcdef) + CO₂ (a) or if 6-phosphogluconic acid (6PG) MID is obtained, the reaction can be broken down into 2 separate reactions: G6P (abcdef) → 6PG (abcdef), 6PG (abcdef) → RU5P (bcdef) + CO₂ (a).
Table 1.1 Brief description of quenching methods

<table>
<thead>
<tr>
<th>Quenching method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Methanol solution</td>
<td>Use a pre-chilled methanol solution (&lt; -50 °C, either pure methanol or a 40%-75% methanol and water solution can be utilized)</td>
</tr>
<tr>
<td>Fast-cooling method</td>
<td>Use pre-chilled minimal media or buffer solution (&lt; -10 °C) while bathing culture in dry ice or liquid N₂.</td>
</tr>
<tr>
<td>Fast filtration</td>
<td>Set up a vacuum filtration system that can pull liquid through rapidly (&lt;5s) with a nylon filter (25 mm).</td>
</tr>
<tr>
<td>Rapid Sampling Unit</td>
<td>Use a sampling valve aiewicznd port injector to rapidly sample from bioreactor; may still a quenching solvent such as cold methanol or hot ethanol</td>
</tr>
<tr>
<td>(Nöh et al., 2007; Link et al., 2015)</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 2: Deciphering cyanobacterial phenotypes for fast photoautotrophic growth via isotopically nonstationary metabolic flux analysis

This chapter has been reproduced from the following publication:


Abstract

Background: Synechococcus elongatus UTEX 2973 is the fastest growing cyanobacterium characterized to date. Its genome was found to be 99.8% identical to S. elongatus 7942 yet it grows twice as fast. Current genome-to-phenome mapping is still poorly performed for non-model organisms. Even for species with identical genomes, cell phenotypes can be strikingly different. To understand Synechococcus 2973’s fast-growth phenotype and its metabolic features advantageous to photo-biorefineries, $^{13}$C isotopically nonstationary metabolic flux analysis (INST-MFA), biomass compositional analysis, gene knockouts, and metabolite profiling were performed on both strains under various growth conditions.

Results: The Synechococcus 2973 flux maps show substantial carbon flow through the Calvin cycle, glycolysis, photorespiration and pyruvate kinase, but minimal flux through the malic enzyme and oxidative pentose phosphate pathways under high light/CO$_2$ conditions. During fast growth, its pool sizes of key metabolites in central pathways were lower than suboptimal growth. Synechococcus 2973 demonstrated similar flux ratios to Synechococcus 7942 (under fast growth conditions), but exhibited greater carbon assimilation, higher NADPH concentrations, higher
energy charge (relative ATP ratio over ADP and AMP), less accumulation of glycogen, and potentially metabolite channeling. Furthermore, *Synechococcus* 2973 has very limited flux through the TCA pathway with small pool sizes of acetyl-CoA/TCA intermediates under all growth conditions.

**Conclusions:** This study employed flux analysis to investigate phenotypic heterogeneity among two cyanobacterial strains with near-identical genome background. The flux/metabolite profiling, biomass composition analysis, and genetic modification results elucidate a highly effective metabolic topology for CO$_2$ assimilatory and biosynthesis in *Synechococcus* 2973. Comparisons across multiple *Synechococcus* strains indicate faster metabolism is also driven by proportional increases in both photosynthesis and key central pathway fluxes. Moreover, the flux distribution in *Synechococcus* 2973 supports the use of its strong sugar phosphate pathways for optimal bio-productions. The integrated methodologies in this study can be applied for characterizing non-model microbial metabolism.

### 2.1 Introduction

To achieve metrics required for commercialization, cyanobacterial photobiorefineries must have comparable biosynthesis capability to commonly used heterotrophic organisms. Recently, *Synechococcus elongatus* UTEX 2973, a cyanobacterium, was isolated, whose growth rate reaches a doubling time of 2 h under high light and high CO$_2$ conditions (Yu et al., 2015). In comparison, *Synechococcus elongatus* PCC 7942 exhibits a doubling time of ~5 h although its genome sequence is 99.8% identical to *Synechococcus* 2973 (55 single nucleotide polymorphisms and a large 188 kb inversion between *Synechococcus* 2973 and 7942) (Yu et al.,
To understand how cyanobacteria achieve maximal growth rates, this study aims to decipher *Synechococcus* 2973 flux topology under diverse growth conditions.

Metabolic flux analysis (MFA) can provide a quantitative description of the metabolic network, link genome profiling to phenome analysis, and reveal pathway regulations through comparative studies. Currently, the cyanobacterial strain, *Synechocystis* sp. PCC 6803 (doubling time ~8 h), is considered the model cyanobacterium whose metabolism has been extensively profiled by flux analysis tools (Young et al., 2011; Xiong et al., 2015; Yang et al., 2000; You et al., 2014). *Synechocystis* 6803 has significant flux through malic enzyme and oxidative pentose phosphate pathways (OPPP) under the photoautotrophic and photomixtrophic conditions. It also operates a cyclic TCA cycle via the GABA shunt. To profile *Synechocystis* 6803 photoautotrophic metabolism, $^{13}$C-pulse experiments and isotopically nonstationary metabolic flux analysis (INST-MFA) were developed (Young et al., 2011). Using the software package, INCA, mass isotopomer data from dynamic labeling experiments can be used quantify fluxes without the need to precisely determine metabolite pool sizes (which are fitted as parameters to account for transient labeling data) (Young, 2014), and is therefore more convenient than other flux profiling methods (Yuan et al., 2008). In the current study, INST-MFA, gene knockouts, and metabolite analysis were performed to obtain insights into the physiology and metabolic regulations of *Synechococcus* 2973 under different bioreactor conditions. Meanwhile, aspects of biomass composition were measured to reveal changes in macromolecule tradeoffs that correlate to cell growth and bioreactor conditions (Pramanik and Keasling, 1997). The outcome highlights the advantages and hurdles of establishing *Synechococcus* 2973 as a new platform organism for bio-production.
2.2 Results

2.2.1 *Synechococcus* 2973 growth and biomass compositions. In optimal photobioreactor (PBR) conditions, *Synechococcus* 2973 exhibited a growth rate of 0.33 ± 0.05 h\(^{-1}\) and *Synechococcus* 7942 grew at 0.14 ± 0.02 h\(^{-1}\) (Figure 2.1a). There was no statistically significant decrease in growth rates in *Synechococcus* 2973 OPPP pathway mutants (the *pgl* deletion mutant showed slightly better growth in diurnal light conditions). *Synechococcus* and mutants were also analyzed under suboptimal diurnal conditions and same light irradiation as seen in Figure 2.1b. Diurnal cycles greatly impaired all strains’ growth, while *Synechococcus* 2973 still demonstrated slightly faster growth than *Synechococcus* 7942. Based on cell growth and biomass composition analysis, under continuous light conditions, *Synechococcus* 2973 assimilated ~12.2 mmol-C/gDCW/h in the PBR but only ~6.7 mmol-C/gDCW/h under the shaking flask (SF) conditions; while *Synechococcus* 7942 assimilated ~5.1 mmol-C/gDCW/h in the PBR. The estimated values of carbon fixation were further confirmed by the measured net CO\(_2\) uptake rate using gas chromatography methods (Figure S1) (Mueller et al., 2017).

The composition of *Synechococcus* 2973 biomass was measured in both continuous light PBR and SF conditions (Supplementing File: Appendix 1). The total protein (~50%) and amino acid levels did not vary significantly between PBR and SF conditions; however protein from the PBR had more glutamate/glutamine than that from the SF (p-value 0.054 using two-tailed equal variance Student’s t-Test). *Synechococcus* 2973 contained 9 ± 1 % lipids and 1.5 ± 0.5 % glycogen in the PBR and 13 ± 2 % lipid and 6.0 ± 1.0 % glycogen in SF conditions. For comparison, *Synechococcus* 7942 biomass composition was quantified under optimal PBR conditions and comprised 11 ± 1 % lipids, 13 ± 4 % glycogen, and 41 ± 0.4 % protein. Optimal PBR conditions were determined by Yu et al., 2015. Proteinogenic glutamate/glutamine content
in *Synechococcus* 7942 was significantly lower (p-value <0.05) than that in *Synechococcus* 2973 biomass regardless of conditions. Fatty acid compositions of lipids in both strains were predominantly C16:1 and C16:0 fatty acids (85%), followed by C18:1 and C18:0. Relevant to photosynthetic light-harvesting, less chlorophyll a was produced in PBR conditions (4.4 ± 0.5 µg/mL/OD$_{730}$ for 2973; 5.2 ± 0.6 µg/mL/OD$_{730}$ for 7942) compared to SF (7.4 ± 0.3 µg/mL/OD$_{730}$ for 2973, p-value <0.01). Increased chlorophyll levels in SF cells could compensate for insufficient light conditions.

**Figure 2.1 Growth performances of cyanobacterial species.** (A) Growth rate of *Synechococcus* 2973, *Synechococcus* 7942 and the *Syn.* 2973 Δzwf mutant in PBR and SF conditions under continuous light. Doubling times in hours is reported below for each strain. †*Synechococcus* 7942 was grown at 300 µmol photons•m$^{-2}$•s$^{-1}$. Standard deviations are a result of 3-5 biological replicates. *p*-value <0.02 between *Syn.* 2973 and 7942 PBR conditions using two-tailed equal variance Student’s t-Test. (B) Diurnal growth curve and rates of *Synechococcus* strains under SF conditions. 12-hour diurnal growth curve of *Synechococcus* 2973 WT, *Synechococcus* 7942 WT, *Synechococcus* 2973 Δzwf and *Synechococcus* 2973 Δpgl. Standard deviations are based on three biological replicates.
2.2.2 INST-MFA of photoautotrophic metabolism. INST-MFA relies on measurement of transient labeling data of central metabolites. To quench time-course metabolism after $^{13}$C-pulse, we used ice-cold media and liquid N$_2$ bath to quickly freeze time-course samples (Fu et al., 2015). The approach avoided traditional cold methanol quenching strategy that causes significant metabolite leakage for gram-negative bacteria (i.e., cyanobacteria) (Bolten et al., 2007) and resulted in improved LC-MS peak quality (Figure S2) (Millard et al., 2014). The INCA software package (isotopomer network compartmental analysis) was used to calculate flux values based on labeling data (Young, 2014) (Figure 2.2 and Figures S3 & S4). The fitted model for the *Synechococcus* 2973 in the PBR was statistically acceptable (Sum of square residuals SSR=713, with an accepted range of [597 740]). The fitted model for SF had a worse fit (SSR=981, with an accepted range of [382 471]). The fitted model for *Synechococcus* 7942 in the PBR was statistically acceptable (SSR=616, range of [543 679]). SSRs were improved if biomass composition information and CO$_2$ uptake rate were used to constrain the fluxes. Flux-partitioning across conditions and strains was compared through relative flux values that were normalized to CO$_2$ uptake rate and are plotted in Figure 2.2 (Confidence intervals and dilution factors are described in Tables S3-S5; Figure S3 & S4 show isotopomer fitting quality for individual metabolites). Additionally, labeling dynamics for citrate and malate are compared between *Synechococcus* 2973 and 7942 in Figure 2.5. $^{13}$C-enrichment for key metabolites is also presented in Figure 2.5; Figures S5 & S6. The unusual labeling pattern of certain metabolites may be indicative of substrate channeling (or the presence of metabolic inactive pools in bulk cytoplasm). While transient isotopic labeling experiments are not direct evidence for substrate channeling, it is one explanation for the observation of non-sequential labeling and genome-to-phenome differences.
Figure 2.2 *Synechococcus* 2973 flux map determined in PBRs and SF and *Synechococcus* 7942 from the PBR. *Syn.* 2973 PBR values are listed first, followed by SF in italicized values and the third values are *Syn.* 7942 PBR. Net fluxes are normalized to net uptake of 100 mol CO₂. Mean flux values and 95% confidence intervals are given in Table S3, S4, and S5. Estimated net CO₂ uptake rates (mmol/gDCW/h) were 12.2 (PBR), 6.7 (SF), and 5.1 (PBR-7942). Green represents the pentose phosphate pathways and Calvin cycle, orange arrows represent photorespiration pathways, purple arrows represent glycolytic pathways, blue represents the TCA cycle, red arrows represent anaplerotic pathways and black arrows represent biomass synthesis. Dotted lines represented pathways with enzymes that are not annotated for *Synechococcus elongatus* on the KEGG Genome database.
2.2.3 Synechococcus flux maps in PBR and SF cultures. Figure 2.2 provides a comparison of the flux distributions between *Synechococcus* 2973 PBR and SF cultures and *Synechococcus* 7942 PBR cultures. Although the cells have significantly different CO₂ fixation rates and biosynthetic fluxes (i.e., *Synechococcus* 2973 PBR cultures contained more proteins in biomass, while *Synechococcus* 2973 SF and *Synechococcus* 7942 PBR cultures had higher carbon allocation to glycogen), the carbon partitioning through fluxes in central intermediates was similar after normalization to CO₂ uptake rate. *Synechococcus* did not exhibit flux through OPPP regardless of the culturing platform or strain. Approximately 12~13% of the fixed carbon was directed to the TCA cycle. *Synechococcus* had no flux from αKG to FUM, resulting in a linear TCA pathway that is consistent with those in leaves or other photosynthetic tissues (Young et al., 2011; Xiong et al., 2015; Ma et al., 2014). The flux distribution explained the absence of a number of key genes in *Synechococcus* 2973 genome, including succinate-CoA ligase, malate dehydrogenase, α-ketoglutarate dehydrogenase, and GABA shunt genes. Although cyanobacteria are thought to employ a cyclic TCA cycle via GABA shunt for respiratory energy production (Xiong et al., 2014), a complete TCA cycle seems non-essential for the *Synechococcus elongatus* strains possibly due to evolutionary non-necessity or energy molecule balance (Rubin et al., 2015). Based on genome annotation, *Synechococcus elongatus* synthesizes fumarate from aspartate via purine metabolism, which can then be converted into malate and succinate (dynamic labeling data shown in Figure S5) providing a possible source of several organic acids and bypassing. On the other hand, *Synechococcus* 2973 had highly active anaplerotic fluxes through PEP carboxylase for the synthesis of TCA cycle intermediates, but its malic enzyme fluxes were minimal (less than ~1 % of total fixed CO₂ in both the PBR and SF conditions). Apart from respiratory fluxes, photorespiration was highly elevated in the PBR relative to SF and
Synechococcus 7942 PBR conditions, demonstrating its role in balancing carbon and energy fluxes under a particular oxidative state when surplus light is available. Additionally, *Synechococcus* 7942 PBR demonstrated increased fluxes through transaldolase reactions and SBPase.

2.2.4 Isotopic metabolite profiling for relative pool size estimation. We analyzed the change of metabolite concentrations in different cyanobacterial cultures (Bennett et al., 2009). Because of metabolite leakage for gram negative cyanobacterial cells, the precise determination of cyanobacterial pool sizes is difficult. Therefore, we used isotopic ratio method to benchmark the change of cyanobacterial metabolite pools against intracellular metabolites from an *E. coli* culture (as standards). Figure 2.3 shows the relative ratio of metabolite pool sizes normalized by the amount of biomass. Comparing to *E. coli*, cyanobacterial R5P/Ru5P and other sugar phosphates were abundant, whereas the pool sizes of TCA metabolites were significantly low (Figure 2.3c). *Synechococcus* 2973 had generally smaller metabolite pools than *Synechocystis* 6803 (with the exception of R5P/Ru5P) and similar metabolite pools compared to *Synechococcus* 7942 (Figure 2.3c) with the exception of UDP-glucose (i.e., *Synechococcus* 7942 contained more UDP-glucose for glycogen synthesis) (Figure 2.3b). Notably, *Synechococcus* 2973 had the highest NADPH concentration among the three cyanobacterial species (Figure 2.3a). In addition, *Synechococcus* 2973 PBR cultivations decreased central metabolite concentrations (mainly sugar phosphates) relative to SF cultures (Figure 2.3d). Lastly, comparing two growth conditions of *Synechococcus* 2973 (Figure 2.3d), the SF cells contained more ADP and AMP than PBR cells, suggesting decreased energy charges (relative ATP concentration to overall adenosine phosphate concentrations) under suboptimal SF condition.
Figure 2.3 Relative pool sizes of *Synechococcus* 2973, *Synechococcus* 7942, and *Synechocystis* 6803 to *E. coli* K-12. A ratio of 1 indicates the same metabolite concentrations (normalized to gram biomass) between cyanobacteria and *Escherichia coli*. A ratio greater than 1 indicates a larger pool size in cyanobacteria strain than *Escherichia coli*. Standard deviations are based on three cyanobacteria biological replicates. (A). Average relative pool size of energy molecules compared to *E. coli* in PBR conditions. (B). Average relative pool size of UDP glucose to *E. coli* in PBR conditions. ADP glucose ratio is relative to *Synechococcus* 6803 (normalized to 10) due to the lack of ADP glucose in the *E. coli* control. (C). Relative metabolite pool size of *Synechococcus* 2973, *Synechococcus* 7942 and *Synechocystis* 6803 in PBR conditions to *E. coli* K-12. (D). Relative metabolite pool size normalized to free glutamate between the shaking flask and PBR conditions for *Synechococcus* 2973.

2.2.5 Effect of exogenous organic acids on *Synechococcus* 2973’s growth. *Synechococcus* 2973 shows strong sugar phosphate pathways, but its TCA cycle is incomplete with limited flux and low metabolite pools. We hypothesized that the addition of exogenous organic substrates
may alleviate such biosynthesis bottlenecks. However, cell growth was not enhanced after supplying OAA, malate, αKG or organic acids (pyruvate and acetate) in either PBR or SF conditions (Figure 2.4a). Nevertheless, cells demonstrated the capability to incorporate exogenous nutrients into proteinogenic amino acids (especially for OAA, Figure 2.4b). Moreover, unlabeled malate from culture medium could be significantly incorporated into aspartate, suggesting an uncharacterized malate dehydrogenase to synthesize oxaloacetate (the precursor of aspartate). In summary, *Synechococcus* 2973 shows poor capability to co-utilize organic acids, and its anabolism may limit the conversion of these nutrients into cyanobacteria biomass (Yan et al., 2011).
Figure 2.4 Biomass growth of *Synechococcus* 2973 in the presence of carbon sources. (A). Growth rate (h⁻¹) of *Synechococcus* 2973 in two conditions when fed with 4 g/L sodium bicarbonate and 6 mM of malate (MAL), acetate (AC), citrate (CIT), OAA, alpha-ketoglutarate (aKG), and a mix. Standard deviation is calculated from biological duplicates. There was no significant difference between growth rates. (B). Percent $^{12}\text{C}$-enrichment in amino acids through provision of 4 g/L of $^{13}\text{C}$-bicarbonate and 6 mM of unlabeled organic acids in SF cultures (labeling for 48 h) compared against the control of just 4 g/L of $^{13}\text{C}$-bicarbonate. The amino acids were chosen because of their direct relation to key metabolic precursors (Ala $\rightarrow$ PYR, Ser $\rightarrow$ 3PGA, Phe $\rightarrow$ PEP/E4P, Glu $\rightarrow$ aKG, Asp $\rightarrow$ OAA). [M-57] represents unlabeled amino acids. Note: *Synechococcus* 2973 used unlabeled organic carbon sources to synthesize proteinogenic amino acids, the [M-57] % of amino acids increased compared to the control sample that was only fed with $^{13}\text{C}$-bicarbonate. Standard deviations are a result of biological duplicates *represents a p-value of <0.02 using two-tailed equal variance Student’s t-Test.

2.3 Discussion

2.3.1 Flux comparisons among cyanobacterial species. *Synechococcus* 2973 and 7942 have a number of phenome differences. The CO₂ uptake rate in *Synechococcus* 2973 was more than twice the amount of *Synechococcus* 7942. *Synechococcus* 2973 showed up to a 3.5-fold greater $^{13}\text{C}$-enrichment over *Synechococcus* 7942 for metabolites such as FBP and S7P (Figure S6) at 40 seconds, indicating robust Calvin cycle fluxes. From the normalized flux map, there are differences in the carbon partitioning as seen in Figure 2.2 and Figure S7. *Synechococcus* 2973 flux through fructose-biphosphate aldolase and phosphatase was greater while the flux towards glycogen and the malic enzyme were less. Flux from F6P to G6P, which is the key step towards OPPP and sugar storage metabolism, was significantly decreased in *Synechococcus* 2973 under both conditions. Malate and citrate were labeled rapidly in *Synechococcus* 2973 due to the enhanced PEPC (phosphoenolpyruvate carboxylase) reaction. Unlike *Synechocystis* 6803 that maintains a functional GABA shunt (Xiong et al., 2014), *Synechococcus elongatus* apparently does not contain the genes to operate a complete TCA cycle (as annotated by the KEGG Genome database), but based on labeling data presented here may utilize a linear pathway from OAA to
make citrate and subsequently aKG. The GABA shunt contains one decarboxylating reaction and the absence of this shunt reduces CO₂ loss. Moreover, a succinate dehydrogenase deletion mutant was created, and the mutant showed no defected growth under continuous light conditions, which confirmed the absence of fluxes from aKG to fumarate. Unlike Synechocystis 6803, Synechococcus 2973 pyruvate is nearly exclusively made from flux through pyruvate kinase, while scarcely using malic enzyme (MAL → PYR + CO₂). Pyruvate kinase is a key step for producing pyruvate, but its activity in some cyanobacterial species can be inhibited by high ATP/ADP ratios during photosynthesis (Bricker et al., 2004). For example, significant amount of pyruvate in Synechocystis 6803 was derived from malate degradation (Young et al., 2011), and overexpression of pyruvate kinase in cyanobacteria can improve the production of lactate and alcohols (Oliver and Atsumi, 2015; Angermayr et al., 2014). However, Synechococcus 2973 has naturally evolved to overcome this pyruvate synthesis bottleneck, de-coupling pyruvate kinase inhibition from high photosynthesis rates. Furthermore, as shown in Figure 2.2 and genome scale modeling (Nogales et al., 2012), strong photorespiration is essential for optimal photosynthesis and maximal Synechococcus 2973 growth despite its net carbon loss. Synechococcus 2973 exhibits twice the O₂ evolution rate of Synechococcus 7942 (Mueller et al., 2017). Such high oxygen levels power linear electron flow in photosynthesis but can create an overly reduced state that must be balanced by photorespiration or cyclic electron flow if the Calvin cycle is operating at a high level (Kramer and Evans, 2011).

OPPP can generate NADPH and oxidize sugar phosphates as a means to derive more reduced compounds. The activation of the OPPP at the same time as the Calvin Cycle generates a futile cycle that reduces the efficient use of carbon and energy, which is why in higher plants the two are coordinately regulated by multiple mechanisms including redox state and pH.
(Buchanan, 1980; Werdan et al., 1975). Due to lack of sufficient regulation or Calvin cycle bottlenecks, slow growing *Synechocystis* 6803 demonstrated \(~12\%\) CO₂ loss through OPPP under continuous light (Young et al., 2011), while *Synechococcus* 2973 does not exhibit significant flux through the OPP pathway. To further confirm this point, we constructed *Synechococcus* 2973 zwf and pgl deletion mutants by disrupting the gene encoding G6P dehydrogenase and 6-phosphogluconolactonase respectively, so that G6P could not be converted to 6PG (Figure S2e & f). The mutants showed similar growth rate to the wild type strain under continuous light and diurnal bioreactor conditions (Figure 2.1), indicating the OPPP did not provide a benefit to the organism and was futile for photoautotrophic metabolism in *Synechococcus* 2973 (Scanlan et al., 1995). These mutants and metabolic insights from the quantification of central carbon fluxes provide new strategies for the redirection of carbon flux for the bioengineering of photosynthetic organisms. The observation that *Synechococcus* 2973 exhibits a similar growth phenotype to *Synechococcus* 7942 under diurnal conditions emphasizes the necessity for optimal bioreactor conditions.

### 2.3.2 The optimal photoautotrophic metabolism (reduced pool sizes, enhanced energy levels, and repartitioning of biomass composition)

Metabolite concentrations are tied to input/output fluxes and can provide insight on biochemical level regulation (Schwender et al., 2015; Buescher et al., 2015). The lack of correlation between pool sizes and absolute fluxes has been shown in literature (Huege et al., 2011; Ma et al., 2014), as well as here between *Synechococcus* 2973, *Synechococcus* 7942 and *Synechocystis* 6803. Compared to SF and *Synechocystis* 6803, the concentrations of a number of central metabolites from glycolysis and TCA, in particular, in *Synechococcus* 2973 were reduced in the PBR conditions. Less metabolite accumulation in central metabolism under PBR conditions could indicate reduced feedback or
higher rates of metabolite turnover. From literature, it is known as cell growth rates and carbon influxes increase, the fraction of anabolic enzymes is increased to pull central metabolites towards macromolecule synthesis (You et al., 2013). Compared to the heterotroph *E.coli*, cyanobacterial acetyl-CoA/organic acid concentrations were much lower (Figure 2.3c). Such observations, together with cultivation experiments using organic acids, indicated that the TCA cycle and its associated anabolic pathways in *Synechococcus* 2973 operate at lower rates than in the *E.coli* chassis, rendering difficulties in producing chemicals from cyanobacterial TCA cycle (Xiong et al., 2015). However, in *Synechococcus* 2973, the photosynthetic capacity generates sugar phosphates readily (FBP phosphatase/aldolase, fructose-bisphosphatase, transketolase and RuBisCO reactions) and can enhance its growth (Liang and Lindblad, 2016; Uematsu et al., 2012; Bernstein et al., 2016). Its strength in producing sugar phosphate building blocks can be potentially applied for biotechnological productions (Markou and Nerantzis, 2013).

Additionally, the higher energy charge in the PBR than SF for *Synechococcus* 2973 may benefit anabolic metabolism and cell tolerance to stress conditions (Srivastava et al., 2013). Cyanobacterial biosynthesis has been closely tied to photosynthesis and changing bioreactor conditions (Carrieri et al., 2015). In SF cultures, there was an increase in lower energy molecules (NADP⁺, ADP, and AMP) inhibiting anabolism, possibly resulting in higher levels of some central metabolites. The PBR grown *Synechococcus* 2973 showed the highest NADPH content due to enhanced light-harvesting that can facilitate organic carbon assimilation and biomass growth (Zhou et al., 2016; Nogales et al., 2012). *Synechococcus* 2973 adjusts its PSII/PSI ratio throughout growth although it was found to have a slightly lower PSII/PSI ratio than *Synechococcus* 7942 while exhibiting greater O₂ evolution (Mueller et al., 2017). Additionally, it may present with the most optimal ATP/NADPH ratio to sustain metabolism while not
exceeding ATP demand (Kramer et al., 2011). This demonstrates its natural advantage in energy metabolism regulation and a key to unlocking cyanobacteria growth constraints. While CO₂ fixation rate is the obvious key factor for phenotype divergence, CO₂ fixation rate is influenced by optimal energy balance, biomass distribution/central carbon fluxes, and compartmentalization of photosynthetic reactions.

*Synechococcus* 2973 PBR cells produce little glycogen (~8 times less than *Synechococcus* 7942 PBR cells and 10 times less than the reported value for *Synechocystis* 6803 (Iijima et al., 2015)). When less carbon is allocated to make glycogen, more can be partitioned toward biomass synthesis, resulting in greater protein levels in *Synechococcus* 2973 than 7942 (Stal and Moezelaar, 1997). On the other hand, *Synechococcus* 2973 produced significantly more glycogen under suboptimal SF circumstance, possibly leading to longer cell doubling times.

### 2.3.3 Metabolic features of *Synechococcus* 2973 for bioproduction applications

The effective use of microbial chassis relies on the performance of cell central metabolism. Understanding and exploring native pathways with high metabolic strengths is a promising direction for future microbial cell factories (Nielsen and Keasling, 2016). Our fluxome results indicate that *Synechococcus* 2973 is advantageous for synthesis of targeted products from its sugar phosphate pathways under optimal bioreactor conditions. Recently, researchers have engineered *Synechococcus* 2973 to produce ~9 g/L of sucrose under potassium chloride stress (Song et al., 2016). This engineered strain also demonstrates excellent potential to produce valuable polysaccharide products. On the other hand, it has been observed that while cyanobacteria can easily adapt their photosynthetic operations, there is a high number of essential metabolic genes involved in photoautotroph growth. This presents additional challenges for mutant strain generation or extensive genetic modification (Nielsen and Keasling, 2016).
Therefore, *Synechococcus* 2973 may not be efficiently engineered using numerous heterologous enzymes or rewiring fluxes from its weak acetyl-CoA/TCA pathways for broad-scope bio-productions.

### 2.3.4. Subpopulations, metabolic inactive pools, and substrate channeling

Modeling results indicated the presence of unlabeled pools that required additional dilution factors for isotopic data fitting. For SF cultures, an increase in dilution factors was necessary to improve fitting, still resulting in an unaccepted SF fit. The dilution of the metabolite active pool by unlabeled metabolites infers phenotypic heterogeneity, i.e., the presence of inactive/non-growth cells or the presence of pre-existing sources of carbon that are more slowly turned over after $^{13}$C-bicarbonate pulses (Mohr et al., 2013). Inactive cells may have biomass degradation and induce reflux of unlabeled carbon into the metabolic network (Nargund et al., 2014). Inactive subpopulations or the presence of metabolic inactive pools intracellularly can be increased by poor light transmittance due to suboptimal culture conditions, cell self-shading and/or insufficient mixing conditions (He et al., 2015). This may explain why *Synechococcus* 2973 does not exhibit significant growth advantages in SF conditions.

In addition, metabolite channeling can pass metabolite intermediates between enzymes without diffusion in the bulk volume of the cell. Channeling can significantly increase pathway efficiency (Wheeldon et al., 2016) and in bacteria is often associated with microcompartments (i.e., carboxysomes) or organized enzyme metabolons (Conrado et al., 2008; Jandt et al., 2013). For fast growing *E.coli* species, glycolysis channeling (glycolysisosome) has been evident (Shearer et al., 2005). Substrate channeling can be inferred from transient labeling experiments and isotope dilutions (Wheeldon et al., 2016). When tracing $^{13}$C-enrichment from 3PGA to downstream metabolites in *Synechocystis* 6803, it was previously found that certain downstream
metabolites could be labeled faster than precursors (e.g., PEP was labeled faster than 3PGA) (Young et al., 2011). In this study, we observed $^{13}$C-enrichment might not completely follow expected precursor-product relationships (Figure S6). *Synechococcus* 2973 showed rapid $^{13}$C-enrichment to S7P over its precursor FBP/F6P, while *Synechococcus* 7942 showed an increase in enrichment to SBP over FBP/F6P. Because cyanobacteria possess bifunctional FBP/SBPase enzymes, the preference for FBP aldolase reactions over trans aldolase reactions may indicate channeling systems. The possible channeling routes in the Calvin cycle, especially due to microcompartmentation by carboxysomes, may provide advantages for CO$_2$ fixation. Upon further analysis, *Synechococcus* 2973 also shows evidence for channeling of glycolysis intermediates towards the TCA pathway with faster citrate M+4 distribution over time that exceeds the $^{13}$C-enrichment levels of 3PGA (Figure 2.5). These channeling data also suggest a heterogenetic distribution of intracellular enzymes and metabolites. It remains unclear whether 0.2% differences at the genome level or other regulation levels, such as metabolite channeling, are responsible for spatial organization of enzymes that confer benefits in growth and metabolism in *Synechococcus* 2973.
Figure 2.5 Labeling dynamics for *Synechococcus* 2973 and 7942 in photobioreactor conditions for Citrate, 3PGA, and Malate. (A). The mass isotopomer distribution for citrate (M+0) and (M+4) fractions as a function of time. Experimentally measured MIDs are the data points, while the line represents the fitted data from INCA. (B). The mass isotopomer distribution for malate (M+0) and (M+4) fractions as a function of time. Experimentally measured MIDs are the data points, while the line represents the fitted data from INCA. (C). The total $^{13}$C-enrichment of 3PGA compared to citrate as a function of time from experimentally measured MIDs. (D). Substrate channel scheme proposed based on labeling data from *Synechococcus* 2973. Experimentally measured MIDs with error bars represent standard deviations from biological duplicates.

2.3.5 Conclusions. *Synechococcus* 2973 demonstrates efficient carbon fixation and small fluxes towards carbon loss and transient storage pathways. Under optimal growth conditions, photosynthesis generates abundant NADPH for rapid biomass synthesis. Under suboptimal conditions, cell anabolism cannot digest the influx from CO$_2$ fixation, causing an accumulation of central carbon metabolites. *Synechococcus* 2973 demonstrates diverse phenotypes and
metabolic regulations based on bioreactor conditions, which can be used for broad photosynthetic bioengineering. Moreover, metabolite channeling appears to be involved in intracellular movement that can subvert diffusional limitations and improve reaction thermodynamics and pathway efficiency (Noor et al., 2014). INST-MFA of cyanobacteria is not only important for genome-to-phenome mapping but also crucial for the rational application of platform photo-biorefineries. Currently, INST-MFA is still challenging, and flux results may be influenced by substrate channeling or subpopulations. Sorting out these factors will require advanced labeling experiments (Allen et al., 2012; Mandy et al., 2014) to complement computer modeling and labeling experiments. Therefore, mutant creations and biomass composition analysis were performed to offer information resources to aid this study.

2.4 Methods

2.4.1 Cultivation conditions and transient labeling experiments. *Synechococcus* 2973 was grown in BG-11 medium (pH 8.0-8.5) at 38 °C with 3% CO₂ aeration (2000 mL/min) and continuous 500 μmol photons•m⁻²•s⁻¹ under photobioreactors (Supporting photo 1). *Synechococcus* 2973 was also grown in shaking flasks (atmospheric CO₂, 38 °C and continuous and diurnal 100 μmol photons•m⁻²•s⁻¹ light, 250 rpm). *Synechococcus* 7942 was grown in identical conditions, except in its optimal PBR conditions, where the light condition was 300 μmol photons•m⁻²•s⁻¹. Higher light irradiance inhibits *Synechococcus* 7942 growth. For the isotopic pool-size experiment, *Synechocystis* 6803 was grown in identical photobioreactors at 30 °C, 3% CO₂ aeration and continuous 300 μmol photons•m⁻²•s⁻¹ light. For each labeling experiment, cultures were diluted to an OD₇₃₀ 0.05 in BG-11 medium without citrate (so citrate could not be used a carbon source or affect bulk citrate labeling dynamics) and grown to OD₇₃₀
0.6 (exponential phase) before the experiment was initiated. Prior to the experiment, the 3% CO₂ aeration was replaced with air (0.04% CO₂) and a 2 mL aliquot of NaH¹³CO₃ (>98% purity, Sigma Aldrich, St Louis) was injected into each PBR or shaking flask (SF) for a final concentration of 4 g/L of NaH¹³CO₃ to saturate the cell with labeled ¹³CO₂. Each identical culture in the PBR and SF was quenched at different time intervals after the ¹³C-pulse from 20 seconds to 2 hours. These time-courses samples were analyzed to capture labeling dynamics of metabolites. Timed biological duplicates were used to generated standard deviations for experimentally measured MID values; however a minimum standard deviation of 3% was used in INCA model.

2.4.2 Metabolite quenching, extraction, and analysis. The "fast-cooling" via cold solvents followed by centrifugation is reliable for recovering intracellular labeled metabolites (Fu et al., 2015). After a ¹³C-pulse, samples were quenched by mixing with ice-cold minimal medium (<-5 °C, without a carbon source) in a falcon tube, which was immediately soaked in liquid N₂ to sustain cold temperature (To avoid ice formation, the culture sample was rapidly swirled for several seconds). Then the quenched biomass samples (<0 °C) were harvested via refrigerator centrifuge (3 min, 6000 rpm). For LC-MS analysis, the pelleted samples were extracted with a chloroform-methanol method (Ma et al., 2014). Ion-pairing LC-MS/MS was performed at the Proteomics and Mass Spectrometry Facility, Donald Danforth Plant Science Center, St. Louis. The labeling of energy molecules (e.g., NADH and ATP), organic acids and acetyl-CoA were quantified using hydrophilic interaction liquid chromatography (HILIC) coupled to electrospray time-of-flight MS (Hollinshead et al., 2016) at Joint Bio-Energy Institute, CA.

2.4.3 Estimation of relative pool sizes of metabolites. Changes of metabolite pool sizes in cyanobacteria were analyzed by an MS isotopomer ratio approach (Bennett et al., 2008) (using
fully labeled cell extracts as internal standards for semi-quantitative metabolomics (Wu et al., 2005)). Specifically, E. coli K-12 was cultured with uniformly-labeled $^{13}$C-glucose and $^{13}$C-sodium bicarbonate in an M9 minimal media. The estimation of metabolite pool sizes was performed by mixing a known amount of labeled E. coli biomass with unlabeled cyanobacteria cultures (three biological replicates). Then, the mixtures were quenched by a liquid N$_2$ bath and extracted for HILIC-MS analysis (Figure S2c). The isotopic ratio of each metabolite (labeled vs. unlabeled) was normalized by the amount of E. coli and cyanobacterial biomass, respectively.

2.4.4 Biomass composition analysis. The cells were harvested during exponential growth by centrifugation, and cell pellets were washed with 0.9% NaCl and ddH$_2$O and then freeze-dried. Protein and amino acid compositional analysis was performed by the Molecular Structure Facility, University of California (Davis, CA). Carbohydrates, lipids, chlorophyll a, and ash were measured by previously reported methods (details in supporting information: Appendix 1).

2.4.5 GC-MS analysis of proteinogenic amino acids. To confirm the capability of Synechococcus 2973 to uptake organic acids, unlabeled seed cultures were inoculated (4% inoculation ratio) into the medium with fully labeled NaH$^{13}$CO$_3$ and unlabeled substrates (6mM) including acetate, TCA cycle intermediates (e.g., malate and citrate) and pyruvate in shaking flasks or PBRs for cultivation of 48 hours. The labeling in proteinogenic amino acids was then analyzed using a TBDMS method (You et al., 2012).

2.4.6 Isotopically Nonstationary MFA. MFA is estimated based on isotope labeling dynamics of the free metabolites. The MFA model included the Calvin cycle, pentose phosphate pathway, TCA cycle, the glyoxylate shunt, anaplerotic pathways, and photorespiration pathway. A list of reactions in the network model with their atom transitions is provided in Table S1. A lumped
biomass equation based on biomass composition analysis and biochemical equations was used to fit the model to biomass production (Table S2). The INCA platform (Young, 2014) used a custom MATLAB ODE solver to fit over 890, 600, and 800 individual mass fragments for *Synechococcus* 2973 PBR, SF, and *Synechococcus* 7942 to obtain fluxes in the assumed cyanobacterial network, respectively. INCA provided a goodness of fit via a chi-square statistical test as well as confidence intervals of all estimated parameters (Model formulation details are included in the Supporting Information).

### 2.4.7 Construction of the zwf, pgl and the succinate dehydrogenase deletion mutants.

Three mutants were built, which served to independently validate the related aspects of metabolism described by the modeling process. The Δzwf mutant was constructed by insertional mutagenesis of the first enzyme of the OPPP, glucose-6-phosphate 1 dehydrogenase. The Δzwf mutant was characterized via growth rate (Figure 2.1). The mutant activity was confirmed via IP-LC-MS/MS by the absence of the downstream product of its disrupted gene, 6-phosphogluconate (Figure S2d,e,f). The Δpgl and Δsdh mutant was constructed by a cpfI based CRISPR system was used to delete succinate dehydrogenase from start codon to stop codon from the chromosome. Details are provided in the Supporting Information.
Chapter 3: Cyanobacteria carboxysome mutant reveals the influence of enzyme compartmentalization on cellular metabolisms and metabolic network rigidity

This chapter has been reproduced from the following manuscript currently undergoing preparation:

Abernathy MH, Czajka J, Allen DK, Cameron J, Tang Y. Cyanobacteria carboxysome mutant reveals the influence of enzyme compartmentalization on cellular metabolisms and metabolic network rigidity. [Currently in preparation]

Abstract

Cyanobacterial carboxysomes encapsulate carbonic anhydrase and ribulose-1,5-bisphosphate carboxylase/oxygenase and are key organelles that promote CO₂ concentration and fixation. Genetic deletion of the major structural proteins encoded within the ccm operon in Synechococcus sp. PCC 7002 (ΔccmKLMN) disrupts carboxysome formation and significantly affects cell physiology. In this study, we employed both metabolite pool size analysis and isotopically nonstationary metabolic flux analysis (INST-MFA) to examine metabolic regulation in cells lacking carboxysomes. Under high CO₂ environments, the ΔccmKLMN mutant had similar growth rates as the control strain and maintained a similar flux distribution through the central metabolism, with the exceptions of moderately elevated protein synthesis and photorespiration activity. Metabolite analyses indicated that the ΔccmKLMN strain had larger pool sizes of pyruvate, UDPG, and aspartate as well as higher excretions of malate and succinate. Under photomixotrophic conditions, both the control strain and the ΔccmKLMN mutant metabolized acetate and pyruvate. Provision of acetate promoted carboxysome mutant growth when light and CO₂ were insufficient. The results suggest that the ΔccmKLMN mutant has minimal changes in flux and instead
reorganizes its metabolism through significant changes in intracellular metabolite pool concentrations. The removal of microcompartments may loosen the flux network regulation and allow for redirection of central metabolites to competing pathways (e.g., lactate production). This study provides important insights into both metabolic regulation via enzyme compartmentation and the compensatory responses in cyanobacterial phototrophic metabolism.

3.1 Introduction

Cyanobacteria are promising production systems for renewable resource demands because they are photosynthetic, genetically tractable, and capable of rapid growth. Photosynthetic organisms assimilate carbon using ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO). However, in addition to its slow carboxylation kinetics, RuBisCO also oxygenates ribulose-1,5-bisphosphate (RuBP) producing 2-phosphoglycolate (2PG), a compound that is toxic to cells at significant levels (Kern et al., 2011). Cyanobacteria have evolved CO$_2$ concentrating mechanisms (CCM) around RuBisCO, encapsulating the enzyme within clathrate structures comprised of multiple proteins that form microcompartments referred to as carboxysomes. Carboxysomes result from protein-protein interactions between structural proteins and RuBisCO, which allows the formation of an icosahedral shell that also encases carbonic anhydrase (CA) (Cameron et al., 2013). These polyhedral structures concentrate and increase CO$_2$ availability to RuBisCO (Cai et al., 2009; Dou et al., 2008), without saturating CA (Mangan and Brenner, 2014).

Bacteria, as simple systems, are presumed to have little or no spatial subcellular organization, and thus the cytosol is generally assumed to be homogenous (Sweetlove and Fernie, 2018). However, evidence of metabolon and metabolite channeling in bacteria exists (Ishikawa et al., 2004; Broddrick et al., 2016; Abernathy et al., 2017a), indicating that enzymes may be more
organized than previously suspected. Carboxysome micro-compartmentalization and photosystem co-localization on membranes are an obvious exception, and the advanced spatial organization confers enhanced growth and productivity. RuBisCo has also been shown to operate apart from carboxysomes and organized near the thylakoid membranes along with Calvin cycle enzymes (Agarwal et al., 2009; Cameron et al., 2013). Central intermediates of metabolism are likely passed between enzymes without being released into the cytosol (Abernathy et al., 2017b). The channeling of metabolites between enzymes could also uniquely contribute to metabolic regulation (Sweetlove and Fernie, 2018). In eukaryotes, channeling is often associated with cellular organelles and membrane compartments, such as the mitochondria or peroxisomes (Graham et al., 2007). While less thoroughly studied, the formation of metabolons and enzyme proximity channeling may also widely present in bacterial pathways (Abernathy et al., 2017b).

Cyanobacterial carboxysomes effectively organize RuBisCO for CO₂ fixation and are the starting point of carbon metabolism in cyanobacteria. By removal of cyanobacterial carboxysomes, this study investigates how the redistribution of enzymes in space or disruption of metabolons can affect cellular metabolism and the metabolic flux network. Isotopic experiments are regularly used to estimate metabolic fluxes and pathway channeling (Shearer et al., 2005; Williams et al., 2011; Zhang et al., 2017). In cyanobacteria, isotopically nonstationary metabolic flux analysis (INST-MFA) uses time-dependent mass isotopomer distributions (MID) of free metabolites to estimate fluxes through central metabolism (Young et al., 2011; Abernathy et al., 2017a; Hendry et al., 2017; Xiong et al., 2015a). The time-dependent labeling experiments can also identify heterogeneous intracellular environments (i.e. metabolically inactive pools) (Abernathy et al., 2017b). In this study, ∆ccmKLMN in conjecture with ¹³C-MFA offers an effective approach to understand Synechococcus 7002 metabolism in the absence of an innate microcompartment that
might offer the anchoring point for cascade enzymes (Abernathy et al., 2018). Specifically, \( \Delta ccmKLMN \) mutant was grown under 1% v/v CO\(_2\) environments (hereby referred to as high CO\(_2\) conditions), and its phenotype was measured for flux reorganization, metabolite concentrations, biomass composition, labeling dynamics of cascade metabolites, and photosynthetic capabilities. The results provide insight into cyanobacterial metabolism and overlooked metabolons for flux regulations.

### 3.2 Results

#### 3.2.1 Growth comparison between the control strain and \( \Delta ccmKLMN \) mutant

The *Synechococcus* PCC 7002 control strain and the \( \Delta ccmKLMN \) mutant were grown at high CO\(_2\) (1% v/v) and continuous light (125 \( \mu \text{mol photons m}^{-2}\text{•s}^{-1} \)), which provided enough biomass for analyses while still maintaining phenotypic growth differences (Figure 3.1). The growth rates in the exponential phase were 0.11 ± 0.005 h\(^{-1}\) and 0.085 ± 0.005 h\(^{-1}\) at 1% CO\(_2\) and 125 \( \mu \text{mol photons m}^{-2}\text{•s}^{-1} \) for control and \( \Delta ccmKLMN \) mutant respectively. In the CO\(_2\) high concentration, the addition of NaHCO\(_3\) did not significantly promote mutant growths. Under low CO\(_2\) conditions (0.2% v/v CO\(_2\)), inclusion of 50 mM NaHCO\(_3\) raised culture pH to ~9 and significantly reduced growth rate for the control strain. When the CO\(_2\) concentration was below 0.25%, the \( \Delta ccmKLMN \) mutant was unable to grow, and the presence of 50 mM NaHCO\(_3\) led to significant cell death (i.e., the drop of OD\(_{600}\) during incubation as seen in Figure 3.1). Under photomixotrophic conditions, the control strain did not utilize glycerol to promote cell growth. The mutant showed slight growth with glycerol under low CO\(_2\) circumstances (0.2% v/v CO\(_2\)). These observations revealed the deleterious impact on biomass productions if the cyanobacterial ability to concentrate CO\(_2\) in microcompartments is removed.
Figure 3.1 Growth as a function of CO$_2$ and light for the control and $\Delta$cemKLMN strains. Low CO$_2$, low light conditions are 0.2% CO$_2$ and 25 µmol m$^{-2}$•s$^{-1}$. High CO$_2$, high light conditions are 1% CO$_2$ and 125 µmol m$^{-2}$•s$^{-1}$. Glycerol at a concentration of 0.2% (w/w) and sodium bicarbonate (NaHCO$_3$) at a concentration of 50 mM were added to cultures to investigate photo-mixotrophic growth rates.

3.2.2 Biomass composition analysis. The biomass compositions (Figure 3.2) were measured to reveal the differences in carbon partitioning between the control strain and the $\Delta$cemKLMN mutant. The fatty acid profile was compositionally similar between the two strains, with the exception of statistically significant (p-value <0.05) increases of 18:2 cis-9,12 and 18:1 cis-11 fatty acids and a decrease of 18:1 cis-9 fatty acid in the $\Delta$cemKLMN mutant. Protein-based amino acid profiles were also similar with a except for slightly more leucine and lysine and slightly less serine and threonine in the mutant. The biomass of the $\Delta$cemKLMN strain exhibited a higher protein fraction, 51.2 ± 1.7 wt% protein of dry cell weight (DCW), compared to the control strain (43.7 ± 0.4%). The total glycogen/carbohydrate and lipid content were comparable with the
ΔccmKLMN mutant comprised of 10.4 ± 3.1 wt% carbohydrate and 7.3 ± 3.0 wt% lipids relative to the control strain that produced 12.3 ± 3.4 wt% carbohydrate and 6.8 ± 4.9 wt% lipids.

**Figure 3.2** Biomass compositional analysis of *Synechococcus* PCC 7002 control and ΔccmKLMN mutant. A.) Fatty acid profile found in each strain as a percent of total dry cell weight. B.) Amino acid profile found in each strain as mole percent. C.) Macromolecule biomass composition as a percentage of dry cell weight. *indicates a significant difference between PCC 7002 control and ΔccmKLMN (p-value < 0.05). Both control and mutant strains were cultivated in high CO₂, high light conditions (1% v/v CO₂ and 125 µmol photons m⁻²s⁻¹).

**3.2.3 INST-MFA reveals flux topology in ΔccmKLMN mutants revealed by INCA.**

Using the biomass constraints as a sink reaction in the metabolic model, isotopically nonstationary metabolic flux analysis was performed with the control and ΔccmKLMN mutant strains under high CO₂ and high light conditions. Figure 3.3 is scaled with net fluxes normalized to a net CO₂ uptake of 100 using a network derived from annotated genes found in the KEGG database. The flux
through the Calvin cycle was large, about 7 moles of carbon per 100 moles assimilated are lost through decarboxylation reactions, and about 11% of the carbon flux is directed to the TCA cycle. The INST-MFA model produced statistically acceptable sum-of-squared-residual fits of 549 and 502 with 95% confidence ranges of [425, 550] and [429, 552], respectively for the carboxysome mutant and control strain.

Based on the INST-MFA analysis, the flux through some central metabolic pathways (including the glycolysis, the TCA and the Calvin cycle pathways) were statistically indistinguishable between the ΔccmKLMN strain and the mutant (despite the differences in outflow fluxes towards synthesis of biomass and extracellular metabolites). This result indicated that the mutant metabolism was able to conserve flux topology after removal of its innate carbon concentration mechanism under high light and high CO₂ conditions.
Figure 3.3 *Synechococcus* PCC 7002 control and ΔccmKLMN flux maps as determined by INCA (Young, 2014) under high CO₂, high light conditions. Net fluxes are normalized to a CO₂ uptake of 100 mmol of CO₂ and reported as mmol/gDCW/h. PCC 7002 ΔccmKLMN flux values are reported in blue, and PCC 7002 control flux values in red. Percentages of total ¹³C enrichment at steady state of some metabolites are represented by bar graphs. ¹³C enrichment was calculated at isotopic steady state using the formula $\frac{1}{N} \sum_{i=1}^{N} M_i * i$. Standard deviations are based off biological replicates (n=2).
Metabolic regulations for photorespiration. Photorespiration occurs when RuBisCO oxygenates RuBP producing 2PG and 3-phosphoglycerate (3PGA). 2PG is metabolized to glyoxylate and is then either decarboxylated to formate or processed by conversion to glycine, serine, hydroxypyruvate, and finally glycerate, which re-enters the Calvin cycle as 3PGA. The KEGG genome annotation of the photorespiratory pathway in PCC 7002 is not complete, thus, the extent to which additional pathways are utilized remain unclear. In the cyanobacterial control strain, photorespiration is minimized because of the carboxysome and efficient CO\(_2\) concentrating mechanism, where a measureable net flux of 6.5 moles of 2PG were estimated to be produced per 100 C-moles of biomass under high light/high \(\text{CO}_2\) condition. Best estimates of photorespiratory flux for the mutant indicated a net flux of 7.6 moles of 2PG per 100 C-moles of biomass. This increase in photorespiration is consistent with elevated amounts of free RuBisCO, increased \(\text{CO}_2\) compensation points (Marcus et al., 1992), and decreased fluorescence measurements (\(F_V/F_M\)) values under high \(\text{CO}_2\) conditions (Figure 3.4). The \(F_V/F_M\) ratio measures the normalized quantum efficiency of photosystem II (PSII); here, the variable fluorescence (\(F_V\)) was slightly increased in the control strain compared to the \(\Delta ccmKLMN\) mutant while the maximum fluorescence (\(F_M\)) remained similar, resulting in 1.15-fold higher \(F_V/F_M\) and thus higher photosynthetic efficiency of PSII. The \(\Delta ccmKLMN\) mutant was also found to have a lower net \(O_2\) evolution (\(O_2\) generation through photosynthetic water splitting and \(O_2\) uptake through aerobic respiration and photorespiration) than the control strain (397 ± 68 \(\mu\)mol \(O_2\)/mg chl \(a/hr\) vs 496 ± 74 \(\mu\)mol \(O_2\)/mg chl \(a/hr\)). Under dark conditions, respiratory \(O_2\) uptake was found to be slightly higher in the \(\Delta ccmKLMN\) mutant and represented a larger percentage of net \(O_2\) evolution. In cyanobacteria, aerobic respiration during continuous light conditions is limited due to a small TCA cycle flux and sufficient energy generation from photosynthesis. Therefore, the majority of respiratory \(O_2\) uptake
is assumed to be photorespiration and in this case, accounted for about 30 ± 7% of O₂ evolution in the ΔccmKLMN strain and 20 ± 3% in the control strain. The Mehler reactions may also account for some O₂ uptake dependent upon on several factors including carbon content and light (Allahverdiyeva et al., 2011). Furthermore, in the carboxysome mutant, glycerate, a downstream photorespiration intermediate, was found to have a larger metabolite pool and to be labeled quicker, indicating higher flux to glycerate; however this was not the case for 2PG, which was found to be labeled faster but also had a smaller metabolite pool (Figure 3.5). Carbon fixation is closely tied to PSII activity as acts as a sink for reducing equivalents, but photorespiration may act as a substitute allowing for sustained electron flow. Thus, our results support the hypothesis that photorespiration may play a critical role in maintaining photosynthetic efficiency during cell stresses.

### Table 3.2.4: Change in pool sizes in response to carboxysome knockout.

<table>
<thead>
<tr>
<th>Metabolite</th>
<th>ΔccmKLMN</th>
<th>PCC 7002 Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fv/Fm</td>
<td>0.33 ± 0.01*</td>
<td>0.38 ± 0.01*</td>
</tr>
<tr>
<td>Fv</td>
<td>0.21 ± 0.01</td>
<td>0.23 ± 0.01</td>
</tr>
<tr>
<td>Fm</td>
<td>0.62 ± 0.02</td>
<td>0.61 ± 0.01</td>
</tr>
<tr>
<td>Fo</td>
<td>0.42 ± 0.01</td>
<td>0.38 ± 0.00</td>
</tr>
</tbody>
</table>

*Fv/Fm is the maximum fluorescence, Fo is the minimum fluorescence, and Fv is the variable fluorescence.

Figure 3.4 Photosynthetic capacity of PSII (Fv/Fm) and O2 evolution of Synechococcus PCC 7002 control and ΔccmKLMN. Technical (n=2) and biological replicates (n=3) were used for standard deviations. *p<0.001. Percentages represent the respiratory O₂ uptake (as measured in the dark) as a fraction of the total net O₂ evolution. F₀ is the minimum fluorescence, Fₘ is the maximum fluorescence, and Fₐ is the variable fluorescence.

3.2.4 Change in pool sizes in response to carboxysome knockout. This study further investigated ΔccmKLMN compensatory responses at metabolite levels (Heise et al., 2015). Pool
sizes were measured qualitatively using fully $^{13}$C labeled *E. coli* cell extracts as an internal standard (Figure 3.5a). The $\Delta ccmKLMN$ mutant was found to dramatically increase its intracellular pyruvate concentrations (by 18 folds compared to the control), suggesting a severe flux congestion at pyruvate node. Moreover, the mutant had significantly higher levels of intracellular UDPG, aspartate, and glycerate and less ADPG. UDPG labeled rapidly in the $\Delta ccmKLMN$ mutant (Fig. S2), which may be important for sucrose synthesis. Sucrose is often excreted by *Synechococcus* when stressed and may reflect the loss of a carbon concentrating mechanism causing a metabolic shift from the competing glycogen pathway (Qiao et al., 2018). The metabolite pool size of RuBP was significantly decreased and the levels of 2PG and 3PGA were slightly decreased compared to the control. There was also an observed pool size increase and increased $^{13}$C-enrichment (Fig. S2) in the free amino acid aspartate, which quickly exchanges with OAA, the product of PEP. However, the flux through PEP was not elevated in comparison to the control strain as previously suggested (Hackenberg et al., 2012). Finally, secreted metabolites were also analyzed in the supernatant. The $\Delta ccmKLMN$ mutant was found to excrete organic acids including succinate and malate (Fig. 3.5b). Enzymic-based spectroscopic assay (R-BioPharm, Darmstadt) determined that the $\Delta ccmKLMN$ mutant produced extracellular succinate (2.8 ± 0.3 mg/L) (Table S1). The excreted organic acids may have a minor impact on the carboxysome deficient mutant flux estimates.
Figure 3.5 Comparison of metabolite pool size between *Synechococcus* PCC 7002 control and ΔccmKLNM. A.) Relative pool size comparisons between the two strains. A y-axis of 1 represents equal metabolite concentrations while a y-axis > 1 indicates the factor of pool size in ΔccmKLNM greater than PCC 7002 ctrl. The relative pool size was normalized using a $^{13}$C-labeled internal standard and cell density. B.) Secretion of succinate and malate as measured in the supernatant as a function of peak intensity, counts per second (cps) and normalized to a $^{13}$C labeled internal standard.

3.3 Discussion

3.3.1 Compensatory response in cyanobacteria conserves flux networks. Carboxysomes are key microcompartments in cyanobacteria for carbon concentration and assimilation. The loss of carboxysomes significantly impairs cell growth under atmospheric and low CO$_2$ conditions. Under high concentration CO$_2$ (1% v/v), however, the ΔccmKLNM mutant maintains flux ratios similar to the control strain despite large transcriptional changes seen in carboxysome mutants under high CO$_2$ conditions (Hackenberg et al., 2012). At a metabolite level, significant changes in steady-state metabolite pool sizes, metabolic fluxes, and biomass composition were noted when
comparing the control to the ∆ccmKLMN strain. Macromolecule partitioning in cyanobacteria was changed with the ∆ccmKLMN mutant containing a higher relative protein content. A slight decrease in carbohydrate content of the ∆ccmKLMN strain may represent a shift in carbon allocation during disruption of the CCM. The loss of the CCM changes the carbon supply and energy-balance, leading to secretion of succinate and malate and an accumulation of intracellular metabolites such as UDPG, aspartate, glycerate, and pyruvate. The CCM is closely tied to allosteric metabolites (Burnap RL et al., 2015), and a loss of carboxysomes may cue metabolite shifts towards free amino acids (i.e. aspartate) and related carbon skeletons (i.e. pyruvate). This causes changes in reaction thermodynamics due to metabolic imbalances. The control strain was also found to have a slightly higher [ATP]/[NADPH] than the carboxysome-free mutant (Fig. S3) and higher photosynthetic capability of PSII (Fig. 3.4). The higher metabolite pool of NADPH in the ∆ccmKLMN mutant indicates saturated electron acceptors, which lowers net O₂ evolution (Wang et al., 2016). Photorespiration was found to be slightly elevated in the ∆ccmKLMN mutant as expected (Fig. 3.3 and Fig. 3.4), which would require more ATP and could result in lower [ATP]/[NADPH]. While transcriptional and translational control must play an important regulatory and adaptive role, it has been proposed that many aspects of metabolic regulation in cyanobacteria can exhibit control at the level of individual metabolite concentrations and energetic source-sink balances (Cano et al., 2018). Thus, despite changes in the spatial organization, complex compensatory responses eventually lead to a highly conserved flux network.

3.3.2 Differences in the assumed metabolic network of PCC 7002 compared to other cyanobacterial species. PCC 7002 had subtle differences in the metabolic network of the model cyanobacterium Synechocystis 6803 (Young et al., 2011). In the TCA cycle, PCC 7002 lacks fumarate hydratase which interconverts malate and fumarate such that fumarate is solely
generated from purine metabolism (Figure 3.3). PCC 7002 also lacks the GABA (γ-aminobutyric acid) shunt and produces succinate through the intermediate succinic semialdehyde (Zhang et al., 2016). Additionally, PCC 7002 genome annotation lacks phosphoketolase, phosphate acetyltransferase, and acetate kinase. Thus PCC 7002 cannot generate acetyl-coA directly from the pentose phosphate pathway. In contrast, *Synechocystis* 6803 has an active phosphoketolase shunt that splits xylulose-5-phosphate (Xu5P) to acetyl phosphate and glyceraldehyde-3-phosphate (G3P); acetyl phosphate can then be converted into acetyl-CoA or acetate. *Synechocystis* 6803 may utilize this shunt to enhance acetyl-coA synthesis and increase the ratio of phosphoglycerate kinase/phosphoglycerate mutase, an important branch node for photosynthetic organisms (Knoop et al., 2013; Xiong et al, 2015b; Chwa et al., 2016).

### 3.3.3 Using carboxysome-deficient mutants to redirect carbon fluxes to important precursor metabolites.

Cyanobacterial flux can be redirected if central carbon sinks or competing pathways are altered (Xiong et al., 2015a). Blocking the synthesis of glycogen, which can accumulate up to 30% of a cyanobacteria’s dry cell weight (Ungerer et al., 2018), is considered a key metabolic strategy for redirecting carbon fluxes. However, knocking out the production of glycogen does not typically increase product titers due to significantly decreased growth rates (Li et al., 2014). In other studies, knocking out glycogen synthesis resulted in about 6% more carbon being directed to the TCA and 30% more flux being directed to UDP-glucose (Hendley et al., 2017). Here, without blocking competing pathways, the Δ*ccmKL MN* mutant was able to increase UDPG pool sizes (by two folds) as well as UDPG synthesis rates (Fig. 3.5 and Fig. S3). The sugar unit is the precursor for sucrose or other valuable polysaccharides like heparosan (Sarnaik et al., 2019). Interestingly, the Δ*ccmKL MN* mutant under high CO₂ and normal nitrogen conditions has significantly elevate intracellular pyruvate and excretes organic
acids through overflow metabolism. Under N-deprivation, knocking out glycogen synthesis elevates pyruvate and other organic acid excretion but causes significantly inhibits growth, making it a poor metabolic engineering strategy (Benson et al., 2016; Jackson et al., 2015). In the ΔccmKLMN mutant, pyruvate accumulation was high despite unchanged central flux network around pyruvate (Figure 3.3 and 3.5). This pyruvate accumulation could be from the malate shunt (and thus the aspartate storage pool) or pyruvate kinase and may represent a different strategy to alter fluxes for biotechnological applications. For example, the ΔccmKLMN mutant had a 24% increase in lactate productivity (uM lactate/h) and a 125% increase in lactate titer (mM lactate) compared to a engineered cyanobacteria lactate producing strain (Clark et al., 2018).

3.3.4 Impact of carboxysome mutation on photomixotrophic metabolism. To probe photomixotrophic growth in the WT and carboxysome mutant, labeled glycerol, acetate, and pyruvate was used for isotopic experiments (Fig. 3.6). Both strains incorporated acetate and resulted in labeled Leu (derived from acetyl-CoA) and Glu (derived from ketoglutarate), which indicates that organic carbons can be mainly incorporated to the right branch of the TCA cycle. The incorporation of the 13C-carbon acetate labeling into leucine but not isoleucine indicates PCC 7002 uses a threonine dependent pathway rather than the citramalate pathway for isoleucine synthesis (Wu et al., 2010). Under the high light/CO2 condition, a mixotrophic environment had minimal impact on ΔccmKLMN mutant growth. Under low CO2 conditions, acetate supplementation rescued growth of the ΔccmKLMN mutant and resulted in the incorporation of acetate carbon into biomass (revealed by isotopic labeling of amino acids) (Figure 3.6a & b). In addition, the ΔccmKLMN mutant also incorporated glycerol into biomass under low CO2 conditions. This is consistent with previous literature indicating that glycerol aids in PCC 7002
growth at atmospheric conditions (0.04% v/v CO₂) by increasing reducing equivalents (NADPH) in the cell (Ludwig and Bryant, 2012). Neither strain utilized significant amounts of labeled glycerol under high CO₂ conditions when with sufficient light sources indicating that the assimilation of CO₂ was the major source of carbon for metabolism. Finally, pyruvate can be co-utilized by both strains for synthesis of some amino acids (such as leucine, isoleucine and glutamate). However, pyruvate inhibits \( \Delta ccmKLMN \) mutant growth under low light and low CO₂ conditions. In summary, tests of photomixotrophic conditions suggest that loss of carboxysomes has a higher impact on cyanobacterial photomixotrophic metabolism under suboptimal growth conditions.
Figure 3.6. Steady-state proteinogenic amino acid labeling from labeled glycerol, acetate and pyruvate to monitor photomixotrophic growth of PCC 7002 control and ΔccmKLNM. GC-MS data for proteinogenic amino acids under photomixotrophic conditions with 25 mM labeled carbon. Standard deviations are estimated based on two biological replicates (n=2). Percent $^{13}$C increase is graphed. Labeling of amino acids (A) and growth rate change (B) from mixotrophic cultures under low light/CO$_2$. Labeling of amino acids (C) and growth rate change (D) from mixotrophic cultures under high light/CO$_2$. Growth rate ($\Delta$OD$_{600}$/day) change (B & D) is normalized to the photoautotrophic growth rate of PCC 7002 control.

3.3.5 Insights into channeling-based flux regulation. Presence of metabolon or metabolite channeling may cause a heterogeneous mixture of intracellular metabolites. In INST-MFA, dilution parameters are typically introduced to aid in model fitting due to observations that
metabolites never reach 100% isotopic enrichment (Antoniewicz et al., 2007; Boyle et al., 2017). These fitted dilution or G parameters for multiple metabolites represent unlabeled biomass reflux or heterogeneous metabolite pools that arise from enzyme localization and/or metabolite channeling. Removal of carboxysomes resulted in higher steady state $^{13}$C enrichment in sugar phosphate metabolites (S7P, 3PGA, RuBP; Figure 3.3) and thus lower inactive pool percentages (Tables S1 and S2). This may indicate more homogeneous metabolite pools and the breakage of metabolite channels present in the cyanobacteria. Recent literature has focused on metabolite channeling as a way for cells to regulate fluxes and cell homeostasis either with dynamic sequential protein-protein interactions or enzyme clusters (Zhang et al., 2017; Sweetlove et., 2018; Abernathy et al., 2017). Protein-protein interactions and localization of enzymes is a form of control for cells (Stipanuk et al., 2018). Calvin cycle enzymes may channel intermediates through enzyme co-localization, as evidenced through ultra-centrifugation and immunoelectron microscopy studies (Agarwal et al., 2009). Genetic removal of carboxysomes could disrupt native metabolite channels and cause a more homogeneous environment as observed from higher degree of $^{13}$C enrichment and higher metabolically active pools in sugar phosphate metabolites. In contrast, the $\Delta ccmKLMN$ mutant had lower $^{13}$C-enrichment for malate and citrate. The INCA based modeling suggests that flux results are sensitive to not only pathway topology and reversibility, but also isotopic dilutions due to heterogeneous intracellular mixture of metabolites. Carboxysome-free mutants may serve as a method of containment for the genetically engineered microbial cell factories (Clark et al., 2018) and as a method for loosening metabolic rigidity, allowing for flux redirection towards pyruvate, TCA intermediates and UDPG; Thus, increasing the mutant’s ability to be engineered for biomanufacturing or mixotrophic growth. Deciphering these innate channels and associated
flux regulations can improve both metabolic modeling and metabolic engineering effectiveness (Abernathy et al., 2017b).

### 3.3.6 Conclusions

Carboxysome removal in PCC 7002 alters the spatial organization of CO$_2$ assimilation. Enzyme organization and protein-protein interactions affect flux networks but the quantitative details remain poorly understood. This study offers new insights into metabolic rigidity and flexibility complex/multiple-level metabolic controllers. Regulation at transcriptional and translational levels (e.g., fine-tuning ribosome production) (Li et al., 2018) as well as channeling serve to minimize cellular impact and reorganization under genetic or environmental perturbations (He et al., 2017). Future studies are needed to elucidate the rules dictating the formation of metabolons or metabolite channeling in central pathways, and as such will enable bio-designers to add an extra level of metabolic control to current methods that rely mostly on gene copy number, transcriptional and translational regulations.

### 3.4 Materials and Methods

#### 3.4.1 Strains and Cultivation Conditions

Both *Synechococcus* PCC 7002 control strain and the *Synechococcus* PCC 7002 ΔccmKLNM mutant were cultured and maintained in A+ medium without citrate (18 g/L NaCl, 0.6 g/L KCl, 1 g/L NaNO$_3$, 5 g/L MgSO$_4$•7H$_2$O, 0.05 g/L KH$_2$PO$_4$, 0.37 g/L CaCl$_2$, 0.03 Na$_2$EDTA•2H$_2$O, 1 g/L Tris HCl, 0.555 mM H$_3$BO$_3$, 2.3 μM ZnCl$_2$, 0.21 μM MoO$_3$, 3 μM NH$_4$Fe(SO$_4$)$_2$•12 H$_2$O, 22 μM MnCl$_2$, 0.012 μM CuSO$_4$, 0.05 CoCl$_2$, 0.0004 g/L vitamin B12) with appropriate antibiotics (100 μg/mL kanamycin or 25 ug/mL gentamicin respectively) (Stevens et al., 1973). Concentrated glycerol and sodium bicarbonate stocks were filtered and added to denoted cultures at a final concentration of 0.2% w/w and 50 mM.
respectively. The same temperature (37 °C), shaking speed (200 rpm) and culture volumes (20 mL of medium in 125 mL shake flasks) were used in all experiments. Cultivations occurred at either 0.2% v/v CO\textsubscript{2} with continuous 25 μmol of photons m\textsuperscript{-2}\textbullet{s}\textsuperscript{-1} or 1% v/v CO\textsubscript{2} with 125 μmol of photons m\textsuperscript{-2}\textbullet{s}\textsuperscript{-1}. Optical density was measured at 730 nm (OD\textsubscript{730}). The \textit{Synechococcus} PCC 7002 mutants were generated previously (Gordan et al., 2016) and obtained from the Pfleger group (University of Wisconsin-Madison). Briefly, the control strain was generated by inserting RbcL-GFP in the SYNPCC7002_A2842 neutral site along with a gentamicin resistance cassette. The \textit{Synechococcus} PCC 7002 ccmKLMN mutant was generated by replacing \textit{ccmKLMN} with a kanamycin resistance cassette along with RbcL-GFP (Gordan et al., 2016). The GFP in control and ccmKLMN mutant were used for the observation and confirmation of carboxysomes inside of cells.

3.4.2 INST-MFA and \textsuperscript{13}C Labeling experiments. INST-MFA was conducted at 1% v/v CO\textsubscript{2} with 125 μmol of photons m\textsuperscript{-2}\textbullet{s}\textsuperscript{-1}. For each experiment, cultures were inoculated from a seed culture to an OD\textsubscript{730} 0.05 in A+ medium and grown to exponential phase. Prior to the labeling experiment, a 2 mL aliquot of saturated NaH\textsuperscript{13}CO\textsubscript{3} (> 98% purity, Sigma Aldrich, St Louis) was injected into the medium for a final concentration of 4 g/L of NaH\textsuperscript{13}CO\textsubscript{3}. Following the \textsuperscript{13}C-pulse, the cultures were quenched at different time intervals (30, 60,120, 180, 300, 150, 600, 900, 3600 s) with ice cold medium and liquid nitrogen bath (Abernathy et al., 2017). The time-courses samples were centrifuged for 5 minutes (8000 x g, ~0 °C). Biomass pellets were then frozen at -80 °C until metabolite extraction and analysis. Timed biological duplicates were used to generate standard deviations for experimentally measured MID values. Relative metabolite pool sizes were analyzed by an MS isotopomer ratio approach using fully labeled cell extracts as internal standards, specifically, fully labeled \textit{E. coli} K-12 (cultured with uniformly labeled \textsuperscript{13}C glucose and \textsuperscript{13}C sodium bicarbonate). The estimation of metabolite pool sizes was performed by mixing a known
amount of $^{13}$C labeled *E. coli* biomass with unlabeled cyanobacteria (three biological replicates). Then, the mixtures were extracted for LC-MS analysis. The isotopic ratio of each metabolite (labeled vs. unlabeled) was normalized by the amount of *E. coli* and cyanobacterial biomass, respectively (Abernathy et al., 2017).

3.4.3 Metabolite Extraction and analysis. The pelleted samples were extracted with a chloroform–methanol method (Ma et al., 2014). Supernatant aliquots were taken for analysis by lyophilizing 5 mL of medium. All samples were resuspended in 1 mL of ddH$_2$O for LC-MS/MS. Ion-pairing LC–MS/MS was performed at the Proteomics and Mass Spectrometry Facility, Donald Danforth Plant Science Center, St. Louis. Succinic acid was quantified using an UV-method enzyme kit from R-Biopharm (cat No. 10176281035; Darmstadt, Germany) with A+ medium as controls.

3.4.4 Isotopically Nonstationary $^{13}$C-MFA. A metabolic network was constructed with carbon atom transitions and included the Calvin cycle, pentose phosphate pathway, TCA cycle, anaplerotic pathways, and a grouped photorespiration pathway (Table S1). A lumped biomass equation based on biomass composition analysis and biochemical equations was used to fit the model to biomass production. The INCA platform used a custom MATLAB ODE solver to fit the experimentally determined free metabolite labeling data for *Synechococcus* PCC 7002 control and the ΔccmKLMN mutant (Young 2014). INCA provided a goodness of fit via a Chi square statistical test as well as confidence intervals of all estimated parameters. Dilution parameters from INCA model were further studied to account for labeling dilutions due to metabolically inactive pools (Young et al., 2011).
3.4.5 **Biomass composition analysis.** The cells were harvested during exponential growth by centrifugation, washed with 0.9% NaCl and ddH₂O and then freeze-dried. Protein and amino acid compositional analysis was performed by the Molecular Structure Facility, University of California (Davis, CA). Total carbohydrate and lipid amounts were measured by previously reported methods (Abernathy et al., 2017). Fatty acid relative profile analysis was performed by MicrobialID. The biomass measurements were used to generate a biomass described using equations described elsewhere (Abernathy et al., 2017). Chlorophyll, pigments, RNA, DNA, and inorganics were estimated from previously published literature (Hendry et al., 2017).

3.4.6 **GC–MS analysis of proteinogenic amino acids.** To assess the capability of *Synechococcus* 7002 to uptake glycerol, acetate, or pyruvate, unlabeled seed cultures were grown at 1% CO₂ until exponential phase and used to inoculate A+ medium at 5% v/v containing either 25 mM of labeled glycerol-2-¹³C, labeled acetate-¹³C₂, or labeled pyruvate-3-¹³C. The cultures were transferred to either 1% CO₂ or 0.2% CO₂ and, after 3 doubling times, the labeling in proteinogenic amino acids was analyzed using a TBDMS method (You et al., 2013).

3.4.7 **Measuring photosynthetic parameters and oxygen evolution.** Cells were harvested during log growth and adjusted based on cell numbers. The fluorescence parameters (Fₐ/Fₘ) of PSII were measured using a fluorescence fluorometer, FL-200 (Photon Systems Instruments, Brno, Czech Republic). Cells were dark adapted for 90 seconds before F₀ and Fₘ were measured using red LED measuring light. Using a custom-built Clark-type electrode, light induced oxygen evolution was measured for 2 minutes at 37 °C following the addition of 10 mM NaHCO₃ as described by a previous protocol (Ungerer et al., 2018). For respiratory oxygen uptake, oxygen consumption was measured for 2 minutes under dark conditions. Chlorophyll *a* content was measured by removing
one milliliter of culture, extracting with methanol, and measuring the resulting absorbance at 665 and 652 nm as determined by a previous method (Porra et al., 1989).
Chapter 4: Metabolic engineering of cyanobacteria for photoautotrophic production of heparosan, a pharmaceutical precursor of heparin

This chapter has been reproduced from the following publication:


*Contributed equally as co-authors; I aided in strain engineering, performed data analysis, performed lab-scale cyanobacteria cultivations, and drafted the manuscript.

Abstract

Heparosan is an unsulfated polysaccharide potentially important for its wide range of cosmetic and pharmaceutical applications, particularly as the precursor for the extensively used anticoagulant, heparin. Generally sourced from animals, commercially available heparin may encounter various immunological and contamination risks. Thus, safe and sustainable microbial platforms could serve as an alternative heparin source. *Synechococcus*, due to their fast photoautotrophic growth, strong sugar phosphate metabolisms and generally regarded as safe (GRAS) nature, may serve as photo-biorefineries for manufacturing heparosan. In this study, we have synthesized an integrative plasmid *pUPm48* for cloning *galU* and *PmHS2* genes in *Synechococcus elongatus* PCC 7942. The engineered recombinants (*pgp7942*) exhibited significant production of heparosan under different culture conditions, where the products were present in both supernatant and cell biomass. The maximum yield of $0.7 \pm 0.2 \, \mu g/g$-DCW (dry cell weight) and a titer of $2.8 \pm 0.3 \, \mu g/L$ was achieved by *pgp7942* under shake flask and continuous light conditions. The analysis also found PCC 7942 encodes a promiscuous
uridyltransferase for UDP-glucose synthesis and naturally produces multiple
glycosaminoglycans including chondroitin sulfate (CS). This study demonstrates for the first-
time cyanobacteria as a promising photoautotrophic refinery for producing a high-value
polysaccharide commonly from animals.

4.1 Introduction

Glycosaminoglycans (GAGs) are a class of saccharides consisting of repeating units of
amino sugars and uronic acids and have increasingly enticed research and commercial interests
due to their wide range of physiological functions (Chavaroche et al., 2013; Peterson et al., 2008;
Cress et al., 2014; Badri et al., 2017). These GAGs, heparan sulfate, chondroitin sulfate, and
hylauronic acid (Figure 4.1), are naturally produced across a range of organisms and play key
roles in moisture retention, cell adhesion and proliferation. Currently GAGs are mainly extracted
from food animal tissues (approximately 100 mg per kg of tissue) (Vaidyanathan et al., 2017)
and are widely used in pharmaceutical and cosmetic applications. However, immunological
reactions and interspecies disease transfers are a major problem, and a contaminated global
supply of heparin (an important anticoagulant) in 2007-8 resulted in 100 deaths (Chavarache et
al., 2013). De novo chemical synthesis of heparin has been attempted and one ultra-low
molecular weight heparin (<1.5 kDa, fondaparinux) has been commercially successful, but yields
are low and biological activity is challenging to replicate as a result of polymerization and
complex sulfation (Liu and Linhardt, 2014).

The precursor of heparin and heparan sulfates (HS) (Wang et al., 2011; Suflita et al.,
2015) is heparosan (HS-0S), an unsulfated polysaccharide consisting of a linear copolymer of
repeating units α-1,4 linked D-glucosamine (GlcNAc) and β-1,4-D-glucuronic acid (GlcUA) (Jin
et al., 2016). Heparosan has also been used to enhance the efficacy of protein-based therapeutics by forming drug conjugates and could serve as a replacement for polyethylene glycol coatings for drug delivery applications (Lane et al., 2017). In recent years, the biosynthesis of GAGs from microbial production platforms has emerged as an alternative to extraction from animal sources (He W. et al., 2015; Englaender et al., 2017). In vitro chemoenzymatic synthesis from in vivo synthesized heparosan could serve as an effective alternative to animal derived heparin (Wang et al., 2011; Wang et al., 2013a; Wu et al., 2015). As a result, efficient heparosan synthesis becomes the primary step towards heparin production (Wang et al., 2011). Heparosan is naturally produced by pathogenic bacteria, such as E. coli K5, Pasteurella multocida, Avibacterium paragallinarum as a part of polysaccharide capsules, which impart virulence to these organisms (Wang et al., 2011; Jin et al., 2016). Heparosan synthesis by genetically modified non-pathogenic Bacillus subtilis and E. coli BL21 has been successfully achieved by cloning kfiC and kfiA genes from E. coli K5, yet molecular weight and homogeneity were difficult to control (Jin et al., 2016, Zhang et al., 2012). Nevertheless, Pasteurella multocida possesses a PmHS2 gene that encodes for a dual functional heparosan synthase that replaces the two genes, kfiC and kfiA, (Chavaroche et al., 2012) and results in smaller-molecular-weight polysaccharide chains (DeAngelis and White, 2004).
Figure 4.1. Schematic featuring the biosynthetic production pathway of glycosaminoglycans in Synechococcus PCC 7942, specifically the expression of heparosan and its downstream applications. Genes in red indicate genes that are not annotated in KEGG database. Genes in blue are naturally present in Synechococcus 7942. Genes in yellow boxes were heterologously expressed in recombinant Synechococcus 7942 for the production of heparosan. 6PG, 6-phosphogluconic acid; AcCoA, acetyl CoA; ADP, adenosine diphosphate; F6P, fructose-6-phosphate; G1P, glucose-1-phosphate; G6P, glucose-6-phosphate; GalNAc, N-acetylgalactosamine, GlcN-1P, glucosamine-1-phosphate; GlcN-6P, glucosamine-6-phosphate; Gln, glutamine; Glu; Glutamic acid; PPi; pyrophosphate, PYR, pyruvate; TCA, tricarboxylic acid; UDP, uridine diphosphate; UTP, uridine triphosphate;

This study aimed at engineering cyanobacteria using PmHS2 from P. multocida (Figure 4.2a). Being GRAS (generally recognized as safe), cyanobacteria can serve as photosynthetic platform for production of valuable chemicals (Yu et al., 2013; Varman et al., 2013). Compared to eukaryotic photoautotrophs, cyanobacterial hosts have established genomic data, molecular manipulation tools, low cost of culture mediums, and generally faster growth (Lu, 2010). Moreover, cyanobacteria possess strong metabolic fluxes through its sugar phosphate pathways and high nucleotide sugar pool sizes (Abernathy et al., 2017a). They naturally synthesize
complex polysaccharides like glycans for colonization, symbiosis, protection, and food reservation (Singh et al., 2005; Kehr and Dittmann, 2015; Philippis et al., 1998). However, to the best of the authors’ knowledge, the production of GAGs with specific disaccharide repeating units has not been previously reported in cyanobacteria. Here, we selected a model cyanobacterum *Synechococcus elongatus* PCC 7942 (PCC 7942) for the photosynthetic synthesis of heparosan to demonstrate their potential as a new chassis for the synthesis of value-added carbohydrates and high-value pharmaceuticals (Figure 4.1).

### 4.2 Materials and Methods

#### 4.2.1 Chemicals and reagents. T4 DNA ligase, restriction enzymes, and shrimp alkaline phosphatase were purchased from New England Biolabs. PrimeSTAR Max DNA polymerase (2X) high fidelity PCR master-mix was purchased from *Clontech* (DSS TaKaRa Bio India Pvt Ltd.). *NucleoSpin*® Gel and PCR Clean-up kit was purchased from Macherey-Nagel (MN, India). Plasmid extraction Miniprep kit was purchased from *GeneAll* (AllianzBio, Mumbai, India).
Figure 4.2. Diagrams of recombinant Synechococcus PCC 7942 (pgp7942) grown in various conditions for the production of heparosan. (A) Representation of the synthetic operon construct, pUPm48, for homologous recombination into NSI of chromosome. (B) PCC 7942 transformants, pgp7942, obtained after complete segregation of the cloned gene were analysed through colony PCR and agarose gel electrophoresis. galU amplicon in lane 1, PmHS2 amplicon in lane 2, entire operon (galU+PmHS2) has been amplified using NSI specific primers in lane 3; followed by >1kb DNA ladder in lane 4. Primers used are listed in Table 4.1. (C) Variable growth conditions to grow pgp7942 and wild-type (WT) PCC 7942. The strains were grown under two distinct shake flask conditions; 250 photons μmol/m²/s (HL) and 0.2% (v/v) CO₂ (HC) (condition 1), 100 photons μmol/m²/s (LL) and 0.04% (v/v) atmospheric CO₂ (LC) with variations in media composition (condition 2). Pgp7942 cells were scaled up under natural diurnal light conditions with irradiance of maximum 1200 ± 200 μmol/m²/s and temperature 33 ± 10 °C with atmospheric CO₂ (0.04%) (condition 3).

4.2.2 Microorganisms and culture conditions. Construction and amplification of recombinant plasmids were performed in E. coli Top10F (ThermoFisher Scientific) cells. Cells were grown at 37 °C in Luria Bertani (LB) broth supplemented with 100 μg/mL spectinomycin (HiMedia) (Sarnaik et al., 2017). For PCC 7942 cultivation, the seed cultures were grown with continuous...
illumination of $55 \pm 2 \, \mu\text{mol/m}^2\text{s}$ and $28 \pm 0.5 \, ^\circ\text{C}$ temperature. BG-11 medium (supplemented with $30 \, \mu\text{g/ml}$ of spectinomycin in case of transformants) was used. Experiments were performed under several laboratory conditions in shake flasks (50 mL culture volumes) at a speed of 250 rpm at atmospheric (0.04% (v/v)) (LC) and 0.2% (v/v) CO$_2$ (HC) at 37 $^\circ\text{C}$ and continuous light environments (100 photons $\mu\text{mol/m}^2\text{s}$ (LL) and 250 photons $\mu\text{mol/m}^2\text{s}$ (HL)). Low nitrogen cultures were prepared with BG-11 medium with $1/5 \, \text{NaNO}_3$ (0.15 g/L) to promote carbohydrate accumulation, while 30 mM of glycerol was added to cultures for photomixotrophic growth (Figure 4.2c). Both cell pellet and supernatant samples were analyzed for heparosan. All experiments were performed in duplicates. Cultures were induced with 1 mM of isopropyl $\beta$-D-1-thiogalactopyranoside (IPTG) at early exponential growth. Culture growth was monitored by recording optical densities of the cultures at 730 nm using a UV-Vis spectrophotometer and measuring the dry cell weight (DCW) of the lyophilized cell pellets.

**4.2.3 Cyanobacterial plasmid construction and transformation.** The plasmid, $pUPm48$, was constructed using the $pAM2991$ vector for transformation of PCC 7942. Initially, the plasmid $pU48$ was constructed by cloning the $galU$ gene amplified from $E. \text{coli}$ MG1655 (UTP-glucose-1-phosphate uridyltransferase, NC_000913.3) genomic DNA between restriction sites, $EcoRI$ and $BamHI$ of the $pAM2991$ vector (Primers; $UE\_F$ and $UAB\_R$) (Table 4.1). The restriction $AflII$ site was introduced in the construct using the $galU$ gene reverse primer ($UAB\_R$) for the introduction of another gene into $pU48$. The plasmid $pUPm48$ was constructed by cloning the $PmHS2$ gene (Heparosan synthase B, AY292200.1) amplified from $Pasteurella \text{multocida}$ genomic DNA between the restriction sites $AflII$ and $BamHI$ sites of the $pR48$ vector (Primers; $PA\_F$ and $PB\_R$) (Table 4.1) (Figure 4.2a). PCC 7942 transformants ($pgp7942$) were developed based on homologous recombination strategy using $pUPm48$ through natural transformation. The
transformation protocol was followed as demonstrated by Clerico et al. Colonies obtained on plates were passaged at least four times to get stable transformants (Sarnaik et al., 2017). Gene integration in cyanobacterial genome was confirmed by colony PCR using neutral site primers \textit{SP48\_F} and \textit{SP48\_R} (Table 4.1) (Figure 4.2b). Agarose gel electrophoresis of the colony PCR samples show bands corresponding to 1 kbps (lane 1, \textit{galU} gene), 2.2 kbps (lane 2, \textit{PmHS2} gene) and 3.5 kbps (lane 3, neutral site primers).

\textbf{Table 4.1. List of primers and sequences used.} The underlined sequences are the restriction sites for cloning the genes.

<table>
<thead>
<tr>
<th>Primer names</th>
<th>Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>\textit{UE_F}</td>
<td>5’-GACTGAATTCAATGGCTGCCATTAATACGAA-3’</td>
</tr>
<tr>
<td>\textit{UAB_R}</td>
<td>5’-ACTTGGATCCCTAAGTTACTTCTTAATGCCCATCTC-3’</td>
</tr>
<tr>
<td>\textit{PA_F}</td>
<td>5’-AAGTCTTAAGATGGAAGAGAAAAAAGAGATG-3’</td>
</tr>
<tr>
<td>\textit{PB_R}</td>
<td>5’-AATCGGATCTCCAAACAAATAAAAAGGTAACAG-3’</td>
</tr>
<tr>
<td>\textit{SP48_F}</td>
<td>5’-GTCTTTTCGACTGAGCCTTCG - 3’</td>
</tr>
<tr>
<td>\textit{SP48_4}</td>
<td>5’-CAGGCAGCCATCGGAAGC - 3’</td>
</tr>
</tbody>
</table>

\textbf{4.2.4 Analysis of heparosan production using LC-MS}

\textbf{Materials for digestion of samples and LC-MS analysis.} Unsaturated disaccharide standards of CS (\textit{\Delta}UA-GalNAc; \textit{\Delta}UA-GalNAc4S; \textit{\Delta}UA-GalNAc6S; \textit{\Delta}UA2S-GalNAc; \textit{\Delta}UA2S-GalNAc4S; \textit{\Delta}UA2S-GalNAc6S; \textit{\Delta}UA-GalNAc4S6S; \textit{\Delta}UA2S-GalNAc4S6S), unsaturated disaccharide standards of HS (\textit{\Delta}UA-GlcNAc; \textit{\Delta}UA-GlcNS; \textit{\Delta}UA-GlcNAc6S; \textit{\Delta}UA2S-GlcNAc; \textit{\Delta}UA2S-GlcNS; \textit{\Delta}UA-GlcNS6S; \textit{\Delta}UA2S-GlcNAc6S; \textit{\Delta}UA2S-GlcNS6S), and unsaturated
disaccharide standard of HA (ΔUA-GlcNAc), where ΔUA is 4-deoxy-α-L-threo-hex-4-enopyranosyluronic acid, were purchased from Iduron (UK). Actinase E was obtained from Kaken Biochemicals (Japan). Various polysaccharide lyases were obtained for digestions. Chondroitin lyase ABC from Proteus vulgaris was cloned and expressed in E. coli and purified in the Linhardt lab. Recombinant Flavobacterial heparin lyases I, II, and III were also expressed and purified by the Linhardt lab using E. coli strains provided by Jian Liu (College of Pharmacy, University of North Carolina). 2-Aminoacridone (AMAC) and sodium cyanoborohydride were obtained from Sigma-Aldrich (St. Louis, MO, USA). All other chemicals were of HPLC grade. Vivapure Q Mini H strong anion exchange spin columns were from Sartoriou Stedim Biotech (Bohemia, NY, USA).

**GAG extraction, sample desalination and digestion.** Lyophilized pellets were re-suspended with 15 mL of ddH₂O, then sonicated for 10 min in an ice bath using a Misonix Sonicator 3000 (frequency 20 kHz) with microtip at 40% magnitude, followed by centrifugation at 3220 relative centrifugal force (RCF) for 10 min. The supernatant was then collected and freeze-dried. Dried supernatant and extracted pelleted samples were re-dissolved in 400 μL ddH₂O and desalted by passing through a 3 kDa molecule weight cut-off spin column. The casing tubes were replaced before 200 μL of digestion buffer (50 mM ammonium acetate containing 2 mM calcium chloride adjusted to pH 7.0) was added to the filter unit. Recombinant heparin lyase I, II, III (pH optima 7.0–7.5) and recombinant chondroitin lyase ABC (10 mU each, pH optimum 7.4) were added to each sample and mixed well. The samples were all placed at 37 °C for 12 h, after which enzymatic digestion was terminated by removing the enzymes by centrifugation. The filter unit was washed twice with 300 μL distilled water and the filtrates containing the disaccharide products were dried via vacuum centrifuge.
2-Aminoacridone (AMAC) labelling and LC-MS analysis. The dried samples were AMAC-labelled by adding 10 μL of 0.1 M AMAC in DMSO/acetic acid (17:3 (v/v)) incubating at room temperature for 10 min, followed by adding 10 μL of 1 M aqueous sodium cyanoborohydride and incubating for 1 h at 45 °C. A mixture containing all 17-disaccharide standards prepared at 0.5 ng/μL was similarly AMAC-labeled and used for each run as an external standard. After the AMAC-labeling reaction, the samples were centrifuged and each supernatant was recovered for LC-MS analysis. LC was performed on an Agilent 1200 LC system at 45 °C using an Agilent Poroshell 120 ECC18 (2.7 μm, 3.0 × 50 mm) column. Mobile phase A was 50 mM ammonium acetate aqueous solution, and the mobile phase B was 100% methanol. The mobile phase passed through the column at a flow rate of 300 μL/min. The gradient was 0-10 min, 5-45% B; 10-10.2 min, 45-100% B; 10.2-14 min, 100% B; 14-22 min, 100-5% B. A triple quadrupole mass spectrometry system equipped with an ESI source (Thermo Fisher Scientific, San Jose, CA) was used a detector in multiple reaction monitoring (MRM) mode. The MS parameters were set at negative ionization mode with a spray voltage of 3000 V, a vaporizer temperature of 300 °C, and a capillary temperature of 270 °C.

4.3 Results and Discussion

Controlled polymerization of heparosan is an important prerequisite to obtain functionally active molecule. Among various natural producers of heparosan, mammals use an active hetero-complex of glycosyltransferases EXT1 and EXT2 yielding a polymer of on average 170 kDa, whereas E. coli K5 synthesizes a polymer of 10-20 kDa using synchronous activity of two enzymes KfiA (glucosaminyl transferase) and KfiC (glucuronyl transferase). On the contrary, Pasteurella expresses a dual functional glycosyltransferase PmHS1 and its cryptic homolog
*PmHS2*. Functionally, *PmHS1* has higher affinity towards short oligosaccharides while *PmHS2* exhibits more affinity towards UDP-sugars which are limiting precursor molecules (Chavaroche et al., 2013). The gene, *PmHS2*, is a glycosyltransferase responsible for the step-wise addition of UDP-glucuronic acid (UDP-GlcUA) and UDP-N acetyl glucosamine (UDP-GlcNAc) for heparosan polymerization (Chavaroche et al., 2012). Based on these reports, the heparosan synthase, *PmHS2* was selected for expression of HS-0S in cyanobacteria. (Figure 4.1).

Being GRAS organisms, cyanobacteria are known to be efficient producers of complex polysaccharides that can be extended to pharmaceuticals (Abed et al., 2009). However, natural production of heparosan or specific GAGs for biotechnology applications has not yet reported in cyanobacteria (Philippis et al., 1998; Kehr and Dittmann, 2015). Therefore, cyanobacterium PCC 7942 was explored as the host for production of heparosan. Successful functionality of the *PmHS2* gene product in cyanobacteria requires a large UDP-glucuronic acid pool (Figure 4.1), mitigated by expression of *galU* (UDP-glucose pyrophosphorylase). Thus, for HS-0S production in cyanobacteria, the vector construct, *pUPm48*, with the genes *galU* and *PmHS2* under the *P_{trc}* promoter (Figure 4.2a) was able to be successfully integrated into the PCC 7942 genome to create the transformants, *pgp7942* (Figure 4.2b).

UDP-glucuronic acid is one of the essential monomeric precursors and is typically synthesized from UDP-glucose. While many cyanobacteria do not have an annotated UDP-glucose pyrophosphorylase that catalyzes the conversion of glucose-1-phosphate to UDP-glucose, some cyanobacteria encode other enzymes relevant to UDP-glucose production and consumption. For example, *Synechocystis* sp. PCC 6803 (PCC 6803) possesses a gene *CugP* (Uniprot ID CUGP_SYNY3) that encodes for UTP-glucose-1-phosphate uridyltransferase, which synthesizes UDP-glucose, same as *galU* (Maeda et al., 2014). Protein BLAST (*BLASTp*)
of this non-

of this non-\textit{galU} type uridyltransferase exhibited 99\% sequence similarity with mannose-1-phosphate guanylyltransferase (BAM54903.1) from PCC 6803 (Figure S1). Its \textit{NTP\textunderscore transferase} functional domain belonging to \textit{Glyco\textunderscore transf\textunderscore GTA\textunderscore type} superfamily was verified from Conserved Domain database at NCBI (domain architecture ID 11440233). \textit{BLASTp} of this domain displayed 79\% sequence homology with mannose-1-phosphate guanylyltransferase (ABB58003.1, gene: SYNPCC7942\textunderscore RS10005) from PCC 7942 (Figure S2). We hypothesize that the gene RS10005 encodes a promiscuous enzyme that can also synthesize UDP-glucose in PCC 7942. This bi-specificity of mannose-1-phosphate guanylyltransferase (gene \textit{RS10005} from PCC 7942) requires further functional annotation. The heterologous gene, \textit{galU}, assists in UDP-glucose formation for a sufficient pool of UDP-glucuronic acid for heparosan production than native RS10005 alone

Cyanobacteria tend to synthesize polysaccharides under various physical and chemical stress conditions (Philippis et al., 1998). Therefore, to improve heparosan (HS-0S) production from \textit{pgp7942}, cells were grown under different environmental conditions (Figure 4.2c), such as high CO\textsubscript{2} (HC), photomixotrophic and nitrogen limitations conditions. The strain \textit{pgp7942} produces significantly higher yields of HS-0S under regular nitrate concentrations than the WT (Figure 4.3a) with the highest titer 2779 ± 285 ng/L and a productivity of 278 ± 29 ng/L/day under HL and HC. Glycerol was added to the medium in LC, LL conditions, but did not result in significant improvement of HS-0S production in \textit{pgp7942} (Figure 4.3a), with titers of 0.7± 0.2 μg/L and 0.6± 0.1 μg/L, with and without glycerol respectively under LC, LL conditions. Cultures when grown in BG\textsubscript{11} with low nitrate concentration (300 ppm) exhibited the smallest HS-0S titer in \textit{pgp7942} and significantly impacted HS-0S and GAG titers. This observation contrasts to cyanobacterial glycogen or polyhydroxybutyrate (PHB) accumulation under nitrogen
starvation as seen in other studies (Monshupanee et al., 2013). The deleterious effect can be explained by a depletion of building blocks for product synthesis, such as glutamine, and a depletion of light harvesting proteins, such as phycobilisomes, during nitrogen limitation (Espinosa et al., 2007).

LC-MS analysis showed that all PCC 7942 cultures synthesized varying types of GAGs and that heparosan, in addition to other GAGs, was also secreted into the supernatant (Figure 4.3b). Cyanobacteria are known to produce an extracellular matrix (ECM) to serve as protection from the outside environment. The ECM is composed of complex heteropolysaccharides assembled and exported through multiple pathways and then are attached to the cell surface or freed into the surrounding environment as released polysaccharides (Ahmed et al., 2014).

According to phylogenomic analyses, *Synechococcus* has been found to possess the least amount of proteins serving the ECM due evolutionary loss, which may reduce cross-talk and competition for substrates compared to other cyanobacteria species (Pereira et al., 2015). Wild-type *Synechococcus* was found to synthesize chondroitin sulfate (CS) (1.2 ± 0.2 μg/L after 120 h under HC, HL conditions) and hyaluronic acid (HA) (0.3 ± 0.1 μg/L after 120 h under HC, HL conditions). The transformant, *pgp7942*, produced similar amounts of HA at 0.4 ± 0.1 μg/L after 120 h under HC, HL conditions, but less CS at 0.5 ± 0.2 μg/L, indicating possible intermediate competition. Higher titers of GAGs and higher supernatant to pellet ratios were obtained in *pgp7942* cultures grown under continuous 250 μmol/m²/s of light and 0.2% CO₂, most likely correlated to biomass increase and secretion of product under optimal growth conditions.

Compared to shake flask cultures of recombinant *B. subtilis* (Jin et al., 2016), the titer and yield reported here is much lower. HS-0S yields on biomass and carbon substrate from *pgp7942* have reached 0.72 ± 0.15 μg/g-DCW (0.83 μg/g-C, using an estimated CO₂ uptake of 1 mmol/g-
DCW/h), while engineered *B. subtilis* achieved at 90 mg/g-C. This is not surprising since heterotrophic bacteria can grow into much higher cell density and use sugar instead of CO$_2$ to directly synthesize polysaccharides.

**Figure 4.3. Heparosan producing strain, *pgp7942*, compared to *Synechococcus 7942* WT under continuous light shake flask conditions.** Shaking flask conditions vary in light from 100 photons µmol/m$^2$/s (LL) to 250 photons µmol/m$^2$/s (HL) and atmospheric CO$_2$ (0.04% (v/v)) (LC) to 0.2% (v/v) CO$_2$ (HC), but are held at a temperature of 37 °C, induced with 1 mM of IPTG, and have an rpm of 250 for 50 mL culture volumes. Low N indicates BG-11 was used with 1/5 of the standard amount of nitrogen typically used. Standard deviations are based off $n = 2$. (A) Titer of heparosan (HS-0S) in both cell pellet and extracellular media. Conditions where the supernatant was not analysed but estimated: *pgp7942* LC/LL + 30 mM glycerol; *pgp7942* LC/LL + low N; WT HC/HL (B) Titer of total glycosaminoglycans (GAGs) in both cell pellet
and extracellular media. (C) Break down of GAG titer in the cell pellet only as heparin sulfate (HS), chondroitin sulfate (CS) and hyaluronic acid (HA).

Moreover, WT and pgp7942 cultures were analysed over time for 10 days under HL and HC conditions. Productivity (Table 4.2) and titer in pgp7942 increased with dry cell weight, and following IPTG induction, HS yield was improved over CS and HA in pgp7942. HS-0S yield was highest during early exponential growth and levelled off after 72 h of growth at 0.72 ± 0.15 µg/g-DCW (Figure 4.4). These time series cultures displayed significant improvement in HS-0S titer over WT (> 30 ng/L). Genetic modification nor heparosan accumulation negatively affected the growth of the HS-0S producing strain.

![Graph A](image1.png) ![Graph B](image2.png) ![Graph C](image3.png)

**Figure 4.4. Heparosan producing strain, pgp7942, under continuous 250 photons µmol/m²/s (HL) and 0.2% (v/v) CO₂ (HC) shake flask conditions.** Red arrow indicates time of induction with 1 mM IPTG. (A) Titer of heparosan (HS-0S) as measured in the lyophilized cell pellet. (B) Yield of glycosaminoglycans, heparan sulfate (HS), chondroitin sulfate (CS), and hyaluronic acid (HA), produced in the pellet of strain, pgp7942. (C) Growth curve for the heparosan producing strain, pgp7942, and the wild type (WT).
Table 4.2. Summary of heparosan production in PCC 7942, WT and engineered 7942 (pgp7942). Highest heparosan and biomass productivity in the cell pellet are reported for each condition. Standard deviation is based off duplicates. All experiments conducted at 33 °C or 37 °C. *Experiment carried out under natural diurnal conditions in environmental laboratory.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Condition description</th>
<th>Carbon source</th>
<th>Light</th>
<th>Heparosan Productivity (µg/L/day) (n=2)</th>
<th>Biomass productivity (at mid-log phase) (g/L/day)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pgp7942</td>
<td>Shaking flask, HC/HL</td>
<td>0.2% CO₂</td>
<td>250 µmol/m²/s (continuous)</td>
<td>0.21 ± 0.03</td>
<td>0.59 ± 0.01</td>
</tr>
<tr>
<td>WT</td>
<td>Shaking flask, HC/HL</td>
<td>0.2% CO₂</td>
<td>250 µmol/m²/s (continuous)</td>
<td>0.01 ± 0.01</td>
<td>0.57 ± 0.02</td>
</tr>
<tr>
<td>pgp7942</td>
<td>Shaking flask, LC/LL</td>
<td>0.04% CO₂ (atmospheric) + 30 mM glycerol</td>
<td>100 µmol/m²/s (continuous)</td>
<td>0.07 ± 0.03</td>
<td>0.35 ± 0.11</td>
</tr>
<tr>
<td>pgp7942</td>
<td>Shaking flask, LC/LL</td>
<td>0.04% CO₂ (atmospheric)</td>
<td>100 µmol/m²/s (continuous)</td>
<td>0.06 ± 0.00</td>
<td>0.44 ± 0.01</td>
</tr>
</tbody>
</table>

This study observed that PCC 7942 can synthesize multiple GAGs (Figure 4.3b) with WT naturally producing chondroitin sulfate (CS). Following induction of the genes galU and PmHS2, pgp7942 produced more HS-0S than CS (Figure 4.4b), indicating competition between polysaccharide synthesis pathways (Figure 4.3c). Interestingly, PCC 7942 does not have an annotated gene for the conversion of UDP-N-acetylglucosamine to UDP-N-acetylgalactosamine, one of the essential precursors for CS (Figure 4.1), nor have an annotated gene for CS synthase. Since cyanobacteria are able to produce a wide range of GAG-like compounds, such as
extracellular polymeric substances (EPS) and lipopolysaccharides (LPS), it is likely that there are unannotated genes and promiscuous enzymes responsible for complex carbohydrates. For example, *E. coli* K4 possess a CS synthase (*WP_000025667*) (Ninomiya *et al.*, 2002) and BLASTp showed 27% homology to PCC 7942’s cellulase synthase. Further gene function studies are required to identify a CS pathway in PCC 7942.

**4.4 Conclusions**

PCC 7942 recombinants, *pgp7942*, were successfully developed for the photoautotrophic production of heparosan through homologous recombination of the genes *galU* and *PmHS2*. *pgp7942* cells were cultivated under different environmental conditions yielding maximum heparosan production (2.8 µg/L) with high light and high CO₂, which was 10³ folds higher than that in wild type strain. PCC 7942 release as well as store GAGs and hence, could serve as efficient biosynthetic platforms for continuous production of glycans like heparosan at a commercial scale. Further research efforts are needed for polymer characterization and regulation; however this study presents proof-in-concept for cyanobacteria as a platform for complex carbohydrate synthesis.
Chapter 5: Channeling in native microbial pathways: Implications and challenges for metabolic engineering

This chapter has been reproduced from the following publication:


Abstract

Intracellular enzymes can be organized into a variety of assemblies, shuttling intermediates from one active site to the next. Eukaryotic reaction compartmentalization within mitochondrion and peroxisomes and substrate tunneling within multi-enzyme complexes have been well recognized. Intriguingly, the central pathways in prokaryotes may also form extensive channels, including the heavily branched glycolysis pathway. In vivo channeling through cascade enzymes is difficult to directly measure, but can be inferred from in vitro analysis, reaction thermodynamics, transport/reaction modeling, analysis of molecular diffusion and protein interactions, or steady state/dynamic isotopic labeling. Channeling presents challenges but also opportunities for metabolic engineering applications. It rigidifies fluxes in native pathways by trapping or excluding metabolites for bioconversions, causing substrate catabolite repressions or inferior efficiency in engineered pathways. Channeling is an overlooked regulatory mechanism used to control flux responses under environmental/genetic perturbations. The heterogeneous distribution of intracellular enzymes also confounds kinetic modeling and multiple-omics analyses. Understanding the scope and mechanisms of channeling in central pathways may improve our interpretation of the robust fluxomic topology throughout the metabolic network and lead to better design and engineering of heterologous pathways.
5.1 Introduction

Metabolic flux is controlled by both enzyme kinetics and substrate diffusions. Under *in vitro* enzyme biochemistry, purified enzymes are often dissociated, and metabolites are well-mixed; so enzymatic reactions have no spatial effect, and the homogenous system can be described by Michaelis–Menten models. However, macromolecular crowding in cells has been shown to promote enzyme associations and structural organization within the cytoplasm (Mourão et al., 2014; Spitzer, 2011). Meanwhile, fluidity in the cytoplasm has been found to be closely related to metabolic activity (Parry et al., 2014). Since *in vivo* environments may significantly impede enzyme kinetics (Minton, 2001), enzyme may organize to pass the product of one reaction site to an adjacent site without releasing into the bulk phase. Such mechanisms not only improve substrate diffusion and reaction-equilibrium, but also benefit other cellular functions, including protection of unstable intermediates, forestallment of substrate competition or inhibition among different pathways, and maintenance of a stable enzymatic microenvironment (Zhang, 2011).

Several mechanisms of multi-enzyme channeling for catalyzing cascade reactions have been recognized (Wheeldon et al., 2016). Intramolecular tunnels, as evidenced in tryptophan synthase, connect two active sites (Miles, 2001). Activated chemical swing arms can pass intermediates between active sites as seen in fatty acid synthases and pyruvate dehydrogenase complexes. Moreover, electrostatic guidance uses complementary charges between protein residues spanning two active sites to direct intermediates as confirmed in malate dehydrogenase and citrate synthase (Wu and Minteer, 2015). Additionally, eukaryotes can compartmentalize enzymes in cellular organelles—i.e., the TCA cycle is carried out in the mitochondrion. On the other hand, evidence has shown that cascade enzymes in central and secondary pathways in
prokaryotes, as well as eukaryotes, are also organized extensively (Table 5.1). For example, glycolytic enzymes may form glycolysisosomes, which suggests intracellular enzymes are not randomly distributed in the cytosol. This hypothesis is not yet generally accepted because central metabolism enzymes often have high copy numbers and, metabolite diffusion is believed to be sufficiently fast (Wheeldon et al., 2016). This chapter reviews evidences and scopes of natural channels in the microbial metabolism as well as methods to measure metabolite channeling through cascade enzymes. We further discuss how channeling influences metabolic flux regulations in heavily branched central pathways. The validation and understanding of natural metabolite channeling through cascade enzymes will offer new guidelines for genome-to-phenome mapping, metabolic modeling, and bio-systems engineering.

<table>
<thead>
<tr>
<th>Enzyme reactions</th>
<th>Detection method</th>
<th>System employed</th>
<th>Species Studied</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyfunctional enzymes forming molecular tunnels</td>
<td>Kinetic studies, enzyme cross linking, LC-MS/MS, SDS-PAGE, and X-ray structural analysis</td>
<td>in vitro</td>
<td>Purified enzyme</td>
<td>(Huang et al., 2001; Miles, 2001; Dunn, 2012)</td>
</tr>
<tr>
<td>OPPP (oxidative pentose phosphate pathway)</td>
<td>Isotopic dilution of $^{14}$C-labeled OPPP intermediates by unlabeled intermediates</td>
<td>in vitro</td>
<td>Purified yeast enzymes, yeast extract, pea and soybean nodule extracts</td>
<td>(Debnam et al., 1997)</td>
</tr>
<tr>
<td>Glycolysis to lactate</td>
<td>Isotopic dilution with $^{13}$C labeled substrates and steady-state MFA</td>
<td>in vivo</td>
<td>Human Cell line AGE1.HN</td>
<td>(Niklas et al., 2011)</td>
</tr>
<tr>
<td>TCA cycle</td>
<td>Isotopic dilution using [4-$^{13}$C]glutamate and [2-$^{13}$C]acetate and observing labeled carbon orientation using $^{13}$C NMR</td>
<td>in vivo</td>
<td>Saccharomyces cerevisiae</td>
<td>(Sumegi et al., 1990)</td>
</tr>
<tr>
<td>Calvin cycle</td>
<td>Isotopic dilution using $^{13}$CO$_2$ and INST-MFA</td>
<td>in vivo</td>
<td>Synechocystis sp. PCC 6803</td>
<td>(Young et al., 2011)</td>
</tr>
<tr>
<td>Pathway/Enzyme Pathway</td>
<td>Methodology</td>
<td>Organism</td>
<td>Reference</td>
<td></td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------</td>
<td>----------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>EMPP (Isotopic dilution with challenging substrates)</td>
<td>in vivo</td>
<td>Escherichia coli</td>
<td>(Shearer et al., 2005)</td>
<td></td>
</tr>
<tr>
<td>TCA cycle (from malate to citrate)</td>
<td>LC-MS/MS and SDS-PAGE; Site-directed mutagenesis; kinetic modeling</td>
<td>in vivo, in vitro, and in silico</td>
<td>Recombinant enzymes from E. coli or Saccharomyces cerevisiae</td>
<td>(Bulutoglu et al., 2016; Vélot and Srere, 2000)</td>
</tr>
<tr>
<td>TCA cycle (fumarate to citrate; malate to citrate)</td>
<td>isotopic dilution, NMR (13C and 1H) and steady-state MFA</td>
<td>In vivo</td>
<td>Arabidopsis thaliana</td>
<td>(Williams et al., 2011)</td>
</tr>
<tr>
<td>TCA cycle (from ketoglutarate to succinate/fumarate)</td>
<td>isotopic labeling with 13C glutamate or 13C acetate; 13C NMR analysis</td>
<td>In vivo</td>
<td>Saccharomyces cerevisiae</td>
<td>(Sumegi et al., 1993)</td>
</tr>
<tr>
<td>3PGA to PEP</td>
<td>Isotopic dilution using 13CO2 and INST-MFA</td>
<td>in vivo</td>
<td>Synechocystis sp. PCC 6803</td>
<td>(Huege et al., 2011)</td>
</tr>
<tr>
<td>PEP to PYR to lactate</td>
<td>Isotopic dilution with [U-13C]glucose and INST-MFA</td>
<td>in vivo</td>
<td>Chinese Hamster Ovary-K1 cell line</td>
<td>(Nicolae et al., 2014)</td>
</tr>
<tr>
<td>Glycolysis</td>
<td>Proteomic analysis of mitochondrial fractions and 13C-glucose studies</td>
<td>in vitro</td>
<td>Arabidopsis thaliana</td>
<td>(Giegé et al., 2003)</td>
</tr>
<tr>
<td>PCB-degradation pathway</td>
<td>Kinetic studies compared between channeled complex and individual complexes</td>
<td>in vitro</td>
<td>Burkholderia xenovorans</td>
<td>(Baker et al., 2009)</td>
</tr>
<tr>
<td>CoA channeling through fatty acid synthase</td>
<td>Substrate pool size quantification and comparative enzyme kinetics</td>
<td>in vitro</td>
<td>Purified chloroplasts from spinach and pea shoots</td>
<td>(Roughan, 1997)</td>
</tr>
<tr>
<td>Fatty acid β-oxidation cycle</td>
<td>Protein crystallization, X-ray diffraction, and homology modelling</td>
<td>in vitro</td>
<td>Escherichia coli</td>
<td>(Ishikawa et al., 2004)</td>
</tr>
<tr>
<td>Polyketide pathway</td>
<td>Kinetic analysis of modular enzymes</td>
<td>in vitro</td>
<td>Saccharopolyspora erythraea</td>
<td>(Wu et al., 2001)</td>
</tr>
<tr>
<td>Electron transport chain complexes I, III and IV</td>
<td>Enzyme activities in the presence of specific electron donors and inhibitors along for genetically modified supercomplexes</td>
<td>in vivo</td>
<td>Mitochondria from mouse fibroblasts</td>
<td>(Lapuente-Brun et al., 2013)</td>
</tr>
<tr>
<td>Photosynthesis (light-harvesting complexes)</td>
<td>Spectroscopic analyses of isolated protein complexes and genetically modified complexes</td>
<td>in vitro</td>
<td>Chlamydomonas reinhardtii</td>
<td>(Iwai et al., 2010)</td>
</tr>
</tbody>
</table>
5.2. Kinetic, thermodynamic and biochemistry evidences on channeling

5.2.1 Kinetic and thermodynamic implications. A cell’s cytoplasm is very crowded with macromolecules constituting up to 40% of the matrix volume, which significantly increases viscosity and diffusional heterogeneity (Zhang, 2011). Consequently, the formation of enzyme associations for metabolite channeling has been hypothesized to overcome many barriers in metabolic pathways (Atkinson, 1977). First, low water concentration strongly inhibits a reaction’s equilibrium and rate. If insufficient free water exists in the cell to support optimum metabolic reaction rates (Hall and Minton, 2003), cells may link enzymes to channel metabolites directly through the pathway. Secondly, concentrations of key intracellular metabolites are often smaller than the Michaelis constant values of the catalyzing enzymes (Fendt et al., 2010), and these metabolite nodes may limit pathway fluxes (Bennett et al., 2009). Channeling improves enzyme conversion of low-concentration substrates while not inhibiting reaction rates. Lastly, many pathways have thermodynamic bottlenecks. For example, the oxidation of malate to oxaloacetate in the TCA cycle has a large positive $\Delta G^\circ$, but the reaction becomes thermodynamically favorable if malate$\rightarrow$oxaloacetate$\rightarrow$citrate can be treated as one step (Noor et al., 2014). Therefore, channeling is an important mechanism to overcome kinetic and thermodynamic obstacles in central metabolism.

5.2.2 Compartmentalization and tunneling. Compartmentalization of multiple enzymes and substrate tunneling within a multi-functional protein complex are well-studied. Cellular compartmentalization is a common characteristic of eukaryotes. Their organelles encapsulate enzymes or DNA by membranes. Compartments of enzyme complexes can be mitochondrion, peroxisomes, chloroplasts, the endoplasmic reticulum, and the Golgi apparatus, which has been extensively explored (Huttanus and Feng, 2017). Moreover, polyfunctional enzyme subunits may
contain elegant tunnels for intermediate conversions. X-ray 3-D structural analyses provide physical evidence of direct enzymatic conversion among active sites without diffusion.

Tryptophan synthase was studied by the X-ray structural analyses, which revealed the three-dimensional arrangement of the enzyme subunits to facilitate intramolecular tunneling of the intermediate indole (Huang et al., 2001). Transient-state kinetic approaches confirmed direct formation of L-tryptophan from indole 3-glycerol phosphate without releasing indole into the bulk phase. Similarly, tRNA-dependent amidotransferase, 4-hydroxy-2-ketovalerate aldolase/acylating, acetaldehyde dehydrogenase, cytidine triphosphate synthetase, pyruvate dehydrogenase, carbamoyl phosphate synthetase, glutamine phosphoribosylpyrophosphate amidotransferase, and asparagine synthetase contain multiple active sites (Huang et al., 2001; Weeks et al., 2006). These active sites are spatially close (ranging from ~25 Å to ~100 Å in length), which diminishes transit time and protects reactive intermediates from contacting with the external medium. Tunneling can also sequester reactive intermediates from hostile external environments. For example, acetyl-CoA synthase/carbon monoxide dehydrogenase from *Moorella thermoacetica* forms a molecular tunnel with a hydrophobic cavity of 138 Å, which controls the Wood-Ljungdahl pathway and reduces CO toxicity (Weeks et al., 2006). In addition to tunneling, other mechanisms may facilitate substrate transport between enzyme reaction sites. For instance, multiple subunits within fatty acid synthase shuttle intermediates for growing fatty acid chains to various catalytic sites via protein-protein interactions (Mashek et al., 2007).

Similarly, pyruvate dehydrogenase complexes employ chemical swing arms to pass intermediates between dehydrogenase and acetyltransferase to transform pyruvate into acetyl-CoA (Perham, 2002).
5.2.3 **In vitro experiments on channeling of cascade enzymes.** *In vitro* analyses of enzyme complexes and interactions have been performed by cross-linking/LC-MS, transient measurement of reaction rates, isotope labeling, kinetic modeling, SDS-PAGE, and detection of cascade resistance to a competing side reaction or a reaction inhibitor (Wheeldon et al., 2016). Recently, recombinant citrate synthase (CS), mitochondrial-malate dehydrogenase (mMDH), and aconitase were found to adopt a metabolon structure (Bulutoglu et al., 2016). Site-directed mutagenesis of conserved arginine residues located in the positively charged electrostatic channel connecting mMDH and CS active sites led to impaired bioconversion of malate. This explicates why high TCA cycle fluxes can operate in a viscous mitochondrial matrix. Additionally, a three-enzyme system (malate hydrogenase, citrate synthase, and lactate dehydrogenase) was used to mimic the *in vivo* TCA pathway malate → oxaloacetate → citrate and found enzyme co-immobilization increased citrate production rates compared to the soluble system (Srere et al., 1973). Some purified enzymes can naturally aggregate to form stable channels under *in vitro* conditions (even in the presence of salt and detergent stresses). For example, the Kohl group studied tissue extracts and purified enzymes from the pentose phosphate pathway (PPP) (from yeast and plant cells), which channeled [1-14C] glucose into 14CO2 without influence by the addition of unlabeled intermediates, such as glucose 6-phosphate (G6P), 6-phosphogluconolactone, or 6-phosphogluconate (Debnam et al., 1997). They also found a lesser degree of channeling of G6P than 6-P-gluconate. This may reflect the fact that G6P has more than one metabolic fate (sharing between PPP and glycolysis).

5.3 **In vivo channeling studies**

5.3.1 **Approaches to reveal in vivo pathway channeling.** Intracellular enzymes along central pathways, like the Embden-Meyerhof-Parnas pathway (EMPP), may co-localize to form
channels (Table 5.1), such as glycolysisosomes (Shearer et al., 2005). This opinion is still highly controversial because it is difficult to investigate enzyme interactions intracellularly. For *in vivo* studies, three common approaches are used. First, genetic engineering can be used to enhance channeled pathways or disrupt enzyme interactions (Figure 5.1c.). For example, to investigate channeling of TCA cycle enzymes in mitochondria, citrate synthase and malate dehydrogenase in *Saccharomyces cerevisiae* were mutated (Vélot and Srere, 2000). The mutation in citrate synthase did not change its kinetic efficiency but caused deficiency in the TCA cycle. Such an inhibitory phenotype could be overcome by malate dehydrogenase overexpression. This study inferred that enzyme interactions are essential for the TCA cycle functioning. Secondly, channeling can be characterized by measurement of molecule movement through channels or identify co-localization of enzymes. *In vivo* diffusivities of small solutes and macromolecules can be observed through photo bleaching of fluorescently labeled macromolecules, microscopy and time-resolved anisotropy (Mika et al., 2010; Verkman, 2002). The observation of metabolite turnover without aqueous-diffusive barriers in crowded cellular environments infers metabolite channeling from one enzyme to an adjacent enzyme. Third, $^{13}$C or $^{14}$C experiments are used to detect elusive channeling phenomena because channeling occurrence may create a redistribution of C labeling (Niklas et al., 2011). In the TCA cycle of *S. cerevisiae*, the $^{13}$C fractions of the second and third carbon in oxaloacetate from $[4-^{13}$C]$\alpha$-ketoglutarate were not randomized, which reveals possible channeling of succinate (a symmetric metabolite) in the enzyme reactions of $\alpha$-ketoglutarate $\rightarrow$ succinate $\rightarrow$ fumarate (Sumegi et al., 1993). In another example, when 1$^{st}$ position labeled glucose is used as a substrate, the channeled EMPP could bypass unlabeled triose phosphate input from the PPP and the Entner–Doudoroff pathway (EDP), altering the isotopomer composition of pyruvate and alanine (Tang et al., 2012). Therefore, $^{13}$C-metabolite flux analysis
(MFA) can offer a noninvasive strategy for the detection of metabolite channeling. Based on the network structure and labeling strategies, $^{13}$C-MFA models may include the cytosolic pathway bypasses or subcellular compartmentations to characterize channeled flux (Williams et al., 2011). On the other hand, $^{13}$C-MFA often relies on the labeling of amino acids. Without extensive isotopomer information of pathway intermediates, $^{13}$C-MFA cannot provide conclusive evidence of channeling in central pathways.

Figure 5.1. Mechanisms for detecting substrate channeling. (A) By using challenging substrates, the Kohl group estimated the extent of EMPP channeling (Shearer et al., 2005). (B) To investigate a pathway with a high degree of channeling, first, a reaction step can be weaken in the channel (i.e., knockdown an enzyme in the pathway) to squeeze metabolites out of the channeling. Following, a $^{13}$C-pulse can detect labeling dynamics (Hollinshead et al., 2016). (C) By mutating an assembly peptide of an enzyme, individual enzyme activity will not be affected, yet the pathway flux may be impaired (Vélot and Srere, 2000).
5.3.2 Detection of leaky channels via dynamic labeling of intermediates. Intracellular pathways may form imperfect channels to release metabolites that connect to other pathways (Figure 5.2). In pathways with leaky channels, intermediates appear in two separate compartments: a channeled pool (with high metabolite turnover rates) and a cytosol pool (with much lower metabolite turnover rates). Under this assumption, dynamic labeling using $^{13}$C (or $^{14}$C) can identify these leaky nodes. For example, $^{13}$C analyses of photoautotrophic metabolism in *Synechocystis* 6803 found channeling of sugar phosphate pathways due to faster labeling in downstream products (PEP: phosphoenolpyruvic acid) than its upstream metabolite (3PG: 3-P-glycerate) right after the cells were fed with $^{13}$CO$_2$ (Huege et al., 2011). Another study saw metabolite channeling in Calvin cycle intermediates with fructose-6-P, ribose-5-P, and glyceraldehyde-3-P being less labeled than downstream intermediates, sedoheptulose-7-P, ribulose-5-P and ribulose-1,5-P (Young et al., 2011). In other words, isotopically nonstationary experiments may reveal leaky metabolite nodes in a channel because its metabolically inactive cytosol pool dilutes the bulk metabolite labeling rate. Although the observation of non-sequential labeling of cascade metabolites after the $^{13}$C-substrate pulse can be used as a “channeling signature”, such method has limitations. First, turnover rates of central metabolites are so rapid that labeling among cascade metabolites reaches steady state instantaneously after $^{13}$C-pulse (Link et al., 2013). Secondly, if the pathway intermediates are perfectly channeled without a metabolically inactive pool, channeling signatures will not be observed.

Kohl et al. proposed an alternative approach based on the redistribution of isotopes in the presence of challenging metabolites (Shearer et al., 2005). In his study of EMPP enzyme channeling, *E. coli* was engineered so that the mutants could uptake both $^{14}$C-glucose and unlabeled sugar phosphate intermediates (called challenging substrates). Based on different $^{14}$C
enrichment rates of downstream metabolites with or without the presence of a challenging substrates, they built a two-compartment model (channel pool vs. cytosol pool) to quantify the fraction of the total flux through the channel as seen in Figure 5.1a. The authors found that the addition of unlabeled sugar phosphates in the labeled culture medium have different degrees of dilutions on the end-product $^{14}$CO$_2$. They found that the glycolysis intermediate fructose-bisphosphate (FBP) was perfectly channeled so the addition of unlabeled FBP did not affect $^{14}$CO$_2$ evolution from $^{14}$C-glucose. The challenging substrate G6P, on the other hand, diluted the $^{14}$CO$_2$ labeling by ~30%, while unlabeled fructose-6-P (F6P) diluted the $^{14}$CO$_2$ labeling by >60%. Thus, G6P and F6P were both channeled to a lesser degree than FBP. This observation is consistent with the factors that G6P and F6P are shared metabolites between the EMPP and PPP. Such leakage is essential for metabolic network connections. Moreover, fructose-bisphosphate aldolase (FBA) is a thermodynamic bottleneck in the EMPP (Flamholz et al., 2013), and the high degree of FBP channeling is essential to overcome this barrier.
Figure 5.2. Types of substrate channels. (A) Example of a perfect channel. Reactions are said to be perfectly channeled if their transient time, $\tau$, which is the time it takes for a reaction to reach steady-state flux, approaches zero (B) Example of a leaky channel where intermediates are also exchanged with the bulk phase. In this case, its transient time, $\tau$, is an indicative of the extent of channeling and is between zero and free diffusion. Leaky channels are necessary for connecting different metabolic pathways that share intermediates.

5.3.3 "Squeeze and Pulse" to investigate pathways of high degree of channeling. The use of challenging substrates by Kohl group have two potential drawbacks. First, the cell cannot naturally uptake challenging substrates without engineering transporters and secondly, the mutant may have catabolite repressions. As such, Hollinshead et al. (2016) designed a two-step approach to examine channeling without the use of challenging substrates. In the first step, genetic tools are used to down-regulate a reaction step, creating a bottleneck in the pathway as seen in Figure 5.1b. Such genetic modification aims to squeeze metabolites into cytosol at the targeted node. This may decrease the pathway flux, making the measurements of metabolite turnover easier. The second step is to pulse $^{13}$C-substrate into cell culture, then time-course samples are collected to capture the labeling dynamics of the cascade metabolites. If a pathway is channeled, the labeling of the precursor metabolites will be slower than their product metabolites because a fraction of the precursor metabolites are squeezed outside of the channel to become
less reactive. For example, a \textit{pfk} \textit{E. coli} mutant was created to decrease the degree of channeling in the EMPP. Then $^{13}$C-pulse experiments found G6P/F6P had slower labeling rates than their product PEP (Hollinshead et al., 2016), supporting the hypothesis of channeling in the EMPP (i.e., glycolysisosomes) by Kohl group.

5.4. Modeling/ theoretical investigation of pathway channeling

5.4.1 Channeling blocks flux towards thermodynamically favorable pathways. The EMPP is important for cell growth, thus many efforts have been proposed to bypass this pathway to increase catabolism. Figure 5.3 indicates multiple thermodynamic bottlenecks in cascade enzymes (highlighted in Figure 5.3a). Alternatively, the EDP requires less enzymatic proteins, and its cumulative $\Delta G^\circ$ is more favorable than the EMPP (Flamholz et al., 2013). However, \textit{E. coli} prefers to use the EMPP rather than its native EDP, even when the EDP is highly overexpressed (Hollinshead et al., 2016). This dilemma can be explained by the possible change of intrinsic Gibbs free energy by EMPP channeling. In a channeled pathway, the local concentrations of a reactant and a product can be altered, leading to a more negative $\Delta G$ which can drive a reaction forward. As illustrated by Figure 5.3b, we present the $\Delta G$ of the pathway from glyceraldehyde-3-P to 3PG (enzymes \textit{gapdh} and \textit{pgk}) as a function of concentrations of reactants and products. The $\Delta G$ of these two reactions in \textit{E. coli} is close to zero (Park et al., 2016), represented as red circles in Figure 5.3b. However, the value can be decreased exponentially with a higher concentration of reactant and a lower concentration of product. In addition, the $\Delta G$ of a reaction can be determined by forward and backward fluxes (Flamholz et al., 2012; Park et al., 2016). It is safe to assume that the forward flux is much higher than the backward flux in a channeled pathway, rendering $\Delta G$ a very negative number. Taken together, it
is plausible that a channeled pathway is more thermodynamically favorable than a non-channeled pathway, altering pathway preference.

### Figure 5.3. Gibbs free energy of reactions in glycolysis.

A Cumulative Gibbs free energy ($\Delta G$) in EMPP and EDP. The highlighted regions indicate the thermodynamic bottlenecks. The error bars are the 95% confidence interval of estimated $\Delta G$. (B) Contour plots of $\Delta G$ as a function of metabolite concentrations in gapdh and pgk reactions (13PG: glyceraldehyde 1,3-bisphosphate; GAP: glyceraldehyde 3-phosphate; 3PG: 3-P-glycerate). The x and y axes are the concentrations of metabolites in log scale, and isolines marked $\Delta G$ values. The red circles are the estimated $\Delta G$ in E. coli cells. $\Delta G$ is calculated by eQuilibrator (Flamholz et al., 2012) (assuming a pH of 7.2 and an ionic strength of 0.25 M) and metabolite concentrations are based on a previous report (Bennett et al., 2009). In $\Delta G$ equation above, $J^+$ and $J^-$ represent forward and backward fluxes.

### 5.4.2. Channeling facilitates transport of metabolites between enzymes.

Diffusion is assumed to be fast enough for reaction rates in small bacterial cells without the need for metabolite channeling (Wheeldon et al., 2016). But others have found that enzymes’ active site orientation, size, distance (Bauler et al., 2010), and electrostatic interactions (Eun et al., 2014) influence fast
metabolite conversions. In this current study, a diffusion model examined the concentration gradient of a metabolite (c) between two cascade enzymes (A, B) with a distance R (Figure 5.4). If the enzymes are close to each other (within 20 nm), diffusion was fast enough to diminish concentration differences between A and B, regardless of D and k. As distance between enzyme A and B increases, the metabolite concentration at the location of enzyme B becomes more heavily influenced by the diffusion coefficient. Based on biochemistry experiments, most enzymes catalyze their reactions to a rate (average $k_{cat}/K_M$ ~$10^5$ M$^{-1}$s$^{-1}$) that is 1,000~10,000 times slower than the diffusion-rate for collisions of proteins and metabolites ($10^8$ M$^{-1}$s$^{-1}$) (Bar-Even et al., 2011). However, such in vitro-measured kinetic parameters may not correlate with in vivo enzyme activity. For example, in vitro reconstruction of the EMPP in a cell-free system observes protein aggregations and significant loss of kinetic activities at enzyme concentrations of a few micro-moles (Liu et al., 2017). In contrast, the bacterial cytoplasm is highly crowded with non-uniform macromolecules and “bound” water. At this high volume fraction, intracellular suspensions exhibit properties that are characteristic of glass-forming liquids and provide a glassy dynamic to the cytoplasm (Parry et al., 2014). Such a high viscosity and tortuosity causes transport barriers via the Stokes-Einstein law. If metabolite diffusion coefficients are reasonably reduced by 10-fold (Mika et al., 2010) by crowding effects, the metabolite concentration around the enzyme B is reduced by ~85%, and thus limit flux towards enzyme B.

5.4.3 Diffusion and reaction rates along EMPP enzymes. An in silico study was conducted to relate enzyme copy number or concentration (enzyme concentration estimated as copy number/$N_A \times$ cell volume (0.7 $\mu$M$^3$)) to diffusion and distance between entities. Relative concentration profiles (Figure 5.4b) were determined based on relative time and positioning using the equation presented in Figure 5.4a. The variables x, c and t are dimensionless after they
are normalized to the distance between A and B \( R \), the initial concentration of A \( C_0 \), and a characteristic time \( R^2/D \), respectively. The resulting equation is irrelevant to \( R \) and \( D \), and here, we also assumed that the relative concentration is one at \( x=0 \) and zero elsewhere, and the concentration gradient is zero at \( x=R \). To achieve 99% of the initial concentration near enzyme B, it was found to take roughly two relative time units. Figure 5.4c shows the estimated time for one metabolite or protein to travel in a specific distance. A previous review suggests that 1 nM of molecules in cells corresponds to 1 \( \mu \)m average intermolecular distance, and that 100 nM corresponds to a distance of 0.25 \( \mu \)m (Grima and Schnell, 2008). For enzymes involved in glycolysis, their copy numbers are generally high (between \( 10^4 \) and \( 10^5 \)) (Voges et al., 2015), which gives a concentration of \( \sim 100 \) nM (between 24 nM and 240 nM) for each individual enzyme. Under such circumstances, the average time for a NBD-glucose molecule (fluorescently labeled deoxyglucose used to monitor glucose uptake) to travel from one enzyme to another one is only 2.5 millisecond (Figure 5.4c). As displayed in Figure 5.4d with the dotted-line, 2.5 millisecond is smaller than reaction turnover time of most enzymes (except \( \text{tim, pgk, pgm, pgl} \) in the glycolysis pathway (Flamholz et al., 2013), further suggesting diffusion poses few bottlenecks. However, if the distance between two enzymes increases 4-fold to 1 \( \mu \)m, the diffusional time increases to 40 millisecond, leading to increased diffusion barriers in the EMPP (except \( fba \)). Further, if a pathway contains multiple steps, cumulative distance among all enzymes will be increased. For instance, the EMPP contains 9 enzymatic steps from G6P to pyruvate. If the model assumes an average distance of 0.25 \( \mu \)m and a diffusion coefficient of 50 \( \mu \)m\(^2\)s\(^{-1}\), 142 millisecond is needed for conversion from G6P to pyruvate, and the time for diffusion accounts for 14.2% of the total time (the sum of diffusion time plus \( 1/k_{\text{cat}} \) of EMPP enzymes). However, through channeling interactions, if the average distance among enzymes is shortened to 0.1 \( \mu \)m, the time fraction for
diffusion drops to 2.6%. This shows even a loose channel can promote fluxes through a long pathway.

**Figure 5.4. A simple model for molecule diffusion in cells.** (A) The governing equation in dimensionless form, where $D$ (µm²s⁻¹) is the diffusion coefficient, and $C₀$ (mM) is the initial concentration. (B) Concentration profile as a function of relative distance and time. The color represents the relative level of concentration and the isolines are numbered with corresponding concentrations. The model describes the diffusion of a metabolite or a protein A to another protein B, while B is assumed to be motionless. The distance between them is $R$ (µm). (C) A table showing how long it may take for free metabolites or proteins to reach 99% of initial concentration at B. The diffusion coefficients are from a reference (Mika and Poolman, 2011) (D) Average time of two reaction events for enzymes in the glycolysis pathway. The values of $k_{cat}$ are from a reference (Flamholz et al., 2013). The dotted line indicates that diffusion in this case is slower than most reaction turnover times.

### 5.4.4 Channeling alters enzyme-metabolite and enzyme-enzyme interactions

Proteins or macromolecules can be considered to be motionless when compared to free metabolites. For example, it may take 2 min for ribosome-loaded mRNAs to travel 1 µm, which is four orders-of-magnitude longer than a glucose molecule. This finding is consistent with the phenomenon of
macromolecular crowding in cells (Ellis, 2001). If enzyme copy numbers are low, the average distance between cascade enzymes may be increased, and metabolite channeling is more likely to be beneficial to multiply the chances of enzyme-reactant collisions. Secondary metabolism in plants has been shown to form channeled pathways to improve low-concentration intermediates (Jandt et al., 2013; Jørgensen et al., 2005; Roze et al., 2011). In addition, metabolite channeling could also prevent promiscuous side reactions. Promiscuous enzymes (e.g., hydrolase) are widely present in microbes as many cross-reactions are being discovered through gene knock outs of critical enzymes (Khersonsky and Tawfik, 2010). Metabolic channeling can help define the products of these promiscuous enzymes (Schwab, 2003). While catalytic promiscuity may introduce unwanted side reactions, channeling may effectively repress side reactions by concentrating enzymes and their desired substrates within a pathway (Piedrafita et al., 2015). Therefore, another benefit for metabolite channeling is to improve pathway specificity. Finally, intracellular proteins are constantly being broken down by diverse proteases (Oda, 2012).

Proteolysis is important for metabolic regulation by deactivating enzymes, transcription factors, and receptors. In general, proteolysis can be highly promiscuous. To protect a wide range of essential enzymes, channeling may reduce the chance of these enzymes being freely exposed to non-specific intracellular proteases so that cell only removes unnecessary or abnormal proteins. In eukaryotes, enzymes can protect themselves by concentrating in compartments; in prokaryotes, compartments are limited, consequently localization of enzymes could be a natural defense against enzyme proteolysis (Gottesman, 1996; Meyer et al., 2014).
5.5 Implications of metabolic channeling

5.5.1 Challenges facing synthetic biology. In metabolic engineering applications, novel pathways with beneficial energy and carbon yields have been proposed to improve bio-productions. However, heterologous pathways typically have inferior efficiency relative to native pathways. Unlike native pathways, the engineered pathways are likely unable to overcome intracellular crowding effects, diffusion limitations, promiscuous side reactions or optimal regulation control. One of the challenges in metabolic engineering is glucose catabolite repression, which blocks E. coli utilization of lignocellulosic sugars. Disruption of the presumed glycolysis channel (such as knocking out the phosphotransferase system or a key gene in the EMPP) decreases biomass growth but improves cell capability to co-utilize C5 sugars (Gonzalez and Antoniewicz, 2017; Hollinshead et al., 2016). This finding strongly implies that metabolite channeling may block EMPP enzymes from co-utilizing intermediates produced from other pathways (e.g., xylose degradation pathway). Such catabolite repression may be reduced if the channeled flux is disrupted, allowing the products of C5 sugars to diffuse from the cytosol into the pathway channel for metabolic conversion.

5.5.2 Applications of synthetic channels to benefit bioproductions. Metabolic engineers have attempted to increase the intrinsic inefficiency of metabolic reactions through enzyme localization and compartmentalization of metabolic pathways. Compartmentalization of pathways has been deployed in eukaryotes by tagging enzymes to specific cellular organelles (Avalos et al., 2013; Sheng et al., 2016). This approach offers benefits such as inhibition of competing reactions, pathway stabilization, substrate concentration, and reducing the release of toxic byproducts (Huttanus and Feng, 2017). It also represents a starting point for innovative bacteria engineering tactics such as targeting bacterial microcompartments like carboxysomes or
heterologously expressing microcompartments (Chen and Silver, 2012; Parsons et al., 2010; Worsdorfer et al., 2011).

Another channeling strategy involves the engineering of protein interaction surfaces or direct protein fusion of interacting partners (i.e., protein or DNA/RNA scaffolds). Scaffolds reduce diffusion time, mitigate toxic metabolite inhibition and make heterologous enzymes more stable (Baek et al., 2013; Moon et al., 2010; Sachdeva et al., 2014; Wang and Yu, 2012). For example, co-localization of three mevalonate biosynthetic enzymes via a protein scaffold achieved improvement in product titer even with low enzyme expression (Dueber et al., 2009; Wang and Yu, 2012). DNA scaffolds have also been developed by orderly locating enzymes on genomic or plasmid DNA to increase the proximity of enzymes, which has shown to facilitate microbial threonine (Lee et al., 2013) and 1,2-propanediol (Conrado et al., 2012) productions. RNA scaffolds with distinct protein-docking sites have been designed to offer precise enzyme spatial organization. Thus, in vivo RNA assemblies can be used to engineer biological pathways with multidimensional architectures, such as hydrogen production (Delebecque et al., 2011).

A particularly interesting area of study that benefits from channeling strategies is flavonoid production in engineered microorganisms. It is well established that flavonoid production in plants is facilitated by the presence of metabolons (Jørgensen et al., 2005). However, flavonoid isolation from naturally-producing plant material is difficult, making heterologous flavonoid production in recombinant organisms like E. coli attractive. Nonetheless, cofactor and substrate supply tend to be the limiting steps in flavonoid bio-production in E. coli. For heterologous flavonoid production in recombinant microorganisms, pathway modulation and enzyme co-localization have been proved valid strategies. Catechin flavonoid production in E. coli was enhanced using combinatorial protein scaffolds with catechin biosynthesis (Zhao et al.,
To simulate the multi-enzyme complexes involved in flavonoid production in plants, a translational protein fusion between two pathway enzymes was used to increase cyanidin 3-**O**-glucoside formation over tandem transcriptional expression of the two enzymes (Yan et al., 2008).

Heterologous expression of microcompartments and co-localization of enzymes are still difficult to employ widely in the metabolic engineering field. The synthetic channel places metabolic burdens on the host and may also affect protein folding. The formation of capsids can impair cell growth and productivity. Moreover, microcompartments and metabolons are constrained by global cell regulations (Chen and Silver, 2012). Current synthetic channeling studies focus mainly on cell productivity without knowledge of the structure, mechanism, or kinetics of engineered channels. It is often unclear whether improved cell productivity is from concentrated metabolite pools, enzyme protection, or elimination of metabolite inhibition. To produce heterologous pathways as efficient as native pathways, we need more efforts to understand the delicate mechanisms of natural channels in central pathways.

**5.5.3 The impact of channeling on genome-phenome mapping.** There have been many studies on the mechanisms regulating metabolic fluxes in microbial systems. Most research focuses on transcriptional regulations and DNA-to-protein regulations under diverse conditions. Protein and metabolite interactions (allosteric regulation) were also proposed to enable instantaneous metabolic response and pathway modulation to maintain metabolic homeostasis (Link et al., 2013). Computational studies have been carried out to relate fluxes to data from the metabolome and proteome (Liebermeister and Klipp, 2006; Noor et al., 2016). Most models assume that metabolic fluxes are determined by metabolite concentrations and protein abundance. Intuitively, gene/protein expressions and metabolite concentrations should have strong influences over their
functional outputs—i.e., fluxes. However, much evidence has shown that neither transcription/enzyme levels nor allosteric regulation are sufficient to explain the flux variations in response to various growth conditions (Blank et al., 2005; Chubukov et al., 2013; Hackett et al., 2016). Even among different heterotrophic microbial species, the substrate catabolic fluxes in different bacterial families or their mutants show much smaller variations than their genome or transcriptional differences (Tang et al., 2009; Wu et al., 2016). This leaves unresolved the question of what are the hidden forces that maintain robustness and stability of flux ratios under genetic or environmental perturbations (Fischer and Sauer, 2005; He et al., 2017). Recent E.coli studies show surprisingly simple regulatory correlations, where only a few global transcription factors (e.g., Crp and Cra) and intracellular metabolites (e.g., cyclic adenosine monophosphate, FBP, and fructose-1-P) appear to coordinate microbial central metabolism (Kochanowski et al., 2017). This leaves a strong hypothesis that native bacterial pathways form well-organized channels to direct metabolic fluxes via several key regulation nodes. A wide distribution of pathway channels is beneficial for cell metabolism to minimize enzyme synthesis costs and to remember an optimal fluxome to organize nutrient influxes throughout the metabolic network. From this point of view, channeling can cause enzyme reaction rates to vary non-linearly with enzyme or metabolite concentration (Wortel et al., 2014) and thus mislead multi-omics investigations of metabolic regulations (Chubukov et al., 2013).

5.5.4 The impact of channeling on $^{13}$C-MFA. $^{13}$C-MFA deciphers cell metabolism via the labeling of intermediate metabolites, which relies on the assumptions of homogenous cell environments and the absence of diffusion limitations. Additionally, $^{13}$C-MFA assumes that scrambling reactions’ net and exchange fluxes (such as succinate to fumarate) are symmetric (Wiechert et al., 2001). For a complex pathway, channeling can provide alternative isotopic
labeling routes. For example, if $[3^{13}\text{C}]$propionate is used as a substrate, malate will be symmetrically labeled in the absence of metabolic channeling. In reality, malate is asymmetrically labeled in *Saccharomyces cerevisiae*, revealing the features of the channeling in TCA cycle enzymes (Sumegi et al., 1990). Therefore, precise $^{13}\text{C}$-MFA must consider the channeling influences on molecule orientation and pathway preferences (e.g., addition of alternate flux channels in the MFA model). One $^{13}\text{C}$-MFA study found that the assumption of channeling bypasses improved MFA model fit and provided a more accurate description of the metabolic network (Williams et al., 2011). However, the use of $^{13}\text{C}$-MFA to investigate pathway channeling relies on precise measurements of labeling positions of key metabolites using both MS and NMR as well as goodness of fit by realistic models. It may fail in highly connected networks or autotrophic metabolism if model or measurement resolution is not sufficient.

On the other hand, isotopically nonstationary flux analysis is a new approach to profile fluxes by tracing isotopic dynamics along cascade metabolites. Metabolite labeling patterns can be confounded by the presence of metabolically inactive pools outside of pathway channels. To account for such phenomena, dilution parameters are used to describe a lack of isotopic equilibrium due to the existence of two or more metabolite pools (Antoniewicz et al., 2007). Dilution factors improve the goodness of fit, but introduce additional degrees of freedom into model. This may increase undetermined parts of the metabolic network (van Winden et al., 2001). Dilution parameters, as a channeling factor, provide quantifications of concentration heterogeneity among intracellular metabolites.
5.6 Conclusions

Intracellular proteins can be organized in a variety of assemblies and functional modules. Channeling through central pathways not only improves flux efficiency and reduces promiscuous reactions, but also maintains flux rigidity against genetic modifications and harsh environmental conditions (He et al., 2017). Characterization of in vivo channeled pathways still lacks standard tools, particularly for longer pathways. To obtain clear evidence of metabolite channeling, imaging, LC-MS, genetic engineering, protein structural analysis, and $^{13}$C-labeling needs to be integrated. The knowledge of channeling represents a new type of constraint to improve computational predictions of cell physiologies and metabolic engineering strategies. To date, many puzzling issues remain regarding the mechanisms for channeling formation, the impact of channeling on kinetic and metabolic modeling, and the evolitional/genetic methods to create channels for metabolite conversions. Channeling offers an explanation for why non-model species with native pathways are much more robust for industrial bio-production. Understanding of channeling may allow for the development of more robust synthetic strains with highly effective heterologous pathways.
Chapter 6: Comparative studies of glycolytic pathways and channeling under *in vitro* and *in vivo* modes

This chapter has been reproduced from the following publication:


*Contributed equally as co-authors; I contributed to the metabolic flux analysis, pool size analysis, $^{13}$C isotopically nonstationary labeling experiments, and drafting of the manuscript.

**Abstract**

This study constructed cell-free glycolytic enzyme systems and compared them to their *in vivo* functions in *E. coli*. Under *in vitro* conditions, flux regulation followed enzyme concentrations and kinetics. In *E. coli*, only one of the isozymes of phosphofructokinase (PfkA) and fructose-bisphosphate aldolase (FbaA) facilitate Embden-Meyerhof-Parnas (EMP) flux, but under *in vitro* assays, these isozymes were interchangeable. Additionally, *in vitro* introduction of the Entner–Doudoroff (ED) pathway improved glycolysis rates, while *in vivo* overexpression of the ED pathway could not capture significant flux unless its phosphotransferase system (PTS) was knocked out. Lastly, *in vivo* $^{13}$C-experiments revealed that the labeling order of EMP pathway intermediates was not strictly cascade, indicating intracellular metabolites were not well mixed. These enigmatic observations cannot be fully explained by thermodynamics or substrate level regulations. This paper supports the long-time conjecture that EMP enzymes are channeled, and the PTS may be an anchor point to initiate enzyme assemblies.
6.1 Introduction

In light of diverse regulatory and metabolite channeling theories, this study re-examined \textit{E. coli} glycolytic pathways: the Embden-Meyerhof-Parnas (EMP) pathway, the pentose phosphate (PP) pathway and the Entner-Doudoroff (ED) pathway. Theoretically, the ED pathway would increase glycolytic flux in microbial hosts for desired products (Khodayari and Maranas, 2016; Liu et al., 2013); thus mutation efforts are employed to improve both native ED pathway usage (Hollinshead et al., 2016) and glucose catabolism in \textit{E.coli} (Charusanti et al., 2010; Huerta-Beristain et al., 2017). However, the ED pathway in \textit{E.coli} is always inferior to the EMP pathway despite its favorable thermodynamics, less enzymatic steps, and comparable enzyme activities (Flamholz et al., 2013; Peng and Shimizu, 2003). Studies on pathway regulations in bacteria often focus on reaction thermodynamics, enzyme kinetics, allosteric inhibitions, or enzyme expressions, while protein-protein interactions remain unfamiliar. To investigate flux controls among competing pathways for glucose catabolism, this study has designed both \textit{in vitro} and \textit{in vivo} systems in conjunction with reported channeling theories (Huege et al., 2011; Young et al., 2011). Via cross-talk and comparisons among enzyme expressions and turnover of cascade metabolites under \textit{in vivo} vs. well-mixed \textit{in vitro} conditions, we aim to obtain new insights into glycolytic regulations. Moreover, the phosphotransferase system (PTS) is the main glucose transporter in \textit{E. coli} (Gosset, 2005) and plays a key role in phosphorylating glucose to glucose-6-phosphate (G6P), which is the branching point for the three glycolytic pathways (Flores et al., 2002) (Figure 6.1). According to chemotactic assembly theory, the PTS would be a starting point of EMP channeling (Zhao et al., 2018). Thus, by using PTS mutants, isotopic labeling and cell free systems, this study revisits the flux re-distributions at a major flux control point in \textit{E. coli} mutants. The outcomes support recent views about \textit{in vivo} flux regulations via channeling (Sweetlove and Fernie, 2018).
Figure 6.1. Glycolytic pathway map in *E. coli*. Both upper EMP and ED enzymes are represented. The enzymes in red represent the enzymes that were also present in the cell-free system. The ED pathway enzymes are underlined. The boxes represent metabolic pathways branching off glycolysis at the G6P node.

6.2 Materials and Methods

6.2.1 Strain Construction. Wild type (WT) *E. coli* (K-12 BW25113) and its mutant strains are listed in Table S1. The plasmid pED was derived from plasmid pBbE5c-YFP (50-70 copies per cell) with the YFP gene replaced with *E. coli* *edd* and *eda* genes, as reported in our previous paper (Hollinshead et al., 2016). Briefly, the strain (∆ptsG + ED) was constructed as follows: competent cells were made from the ∆ptsG mutant using chilled transformation and storage solution. Heat shock transformation was used to transform the ∆ptsG strain with the purified plasmid pED. The
resulting strain was plated on kanamycin (20 µg/mL) and chloramphenicol (30 µg/mL) Luria-Bertani (LB) plates to isolate the transformed cells.

6.2.2 Media composition and growth experiments. All experiments were started with LB seed cultures with appropriate antibiotic. Next minimal M9 medium consisting of 11 g/L of 5x M9 salts (Sigma Aldrich, St. Louis, MO), 2 mM MgSO$_4$·7H$_2$O, 0.1 mM CaCl$_2$·2H$_2$O, 1 mg/L of thiamine hydrochloric acid, 3 mg/L of FeSO$_4$·7H$_2$O, and trace minerals was inoculated (1% v/v) with 10 g/L glucose. All cultivations were aerobic 5 mL cultures on a rotary shaker at 250 rpm and 37 °C. Mutant cultures were supplied with 20 µg/mL kanamycin and 30 µg/mL chloramphenicol as needed. To induce plasmid genes, 0.05 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added at early exponential phases.

6.2.3 Steady state labeling experiments and GC-MS analysis. _E. coli_ cultures were inoculated in fresh M9 minimal medium (0.5% v/v) and were grown with 10 g/L 1-$^{13}$C glucose (Sigma Aldrich, St. Louis, MO) (n=2) until isotopic steady state. Amino acid extraction and GC-MS analysis were performed as described previously during exponential growth phase (You et al., 2012). The fragments [M-15]$^+$ or [M-57]$^+$ and [M-159]$^+$ or [M-85]$^+$ were used for isotopic tracing. Using a previously reported algorithm, the natural isotopic abundance was corrected for the derivatized amino acids (Wahl et al., 2004). Based on mass isotopomer distribution (MID) data from labeled amino acids, flux ratios among glycolytic pathways (Figure 6.3) were computed using software WUFlux (He et al., 2016). Standard deviations were calculated based on the measurement errors of MID data.

6.2.4 Isotopically nonstationary labeling experiments and pool size analysis. _E. coli_ WT and mutant cultures were inoculated from overnight M9 cultures in fresh M9 minimal medium with 2
g/L glucose and allowed to grow for approximately 4 hours. Cultures in exponential growth phase were pulsed with $^{13}$C$_6$-glucose for a final concentration of 4 g/L. To measure the $^{13}$C-incorporation into metabolites over time, cultures were quenched at different time points (from 5 s to 30 s) using ice-cold media and a liquid nitrogen bath. All time points were completed in duplicates. Then biomass was harvested through centrifugation, and the pelleted samples were extracted with a chloroform-methanol method before LC-MS analysis (Ma et al., 2014). The labeling of central metabolites were quantified at Joint Bio-Energy Institute (Berkeley), using hydrophilic interaction liquid chromatography (HILIC) coupled to electrospray time-of-flight MS, as described in our recent paper (Hollinshead et al., 2016). For metabolite pool-size analysis, WT E. coli was grown with $^{13}$C$_6$ glucose and served as the internal standards. Mutant strains were grown with M9 minimal medium with 2 g/L unlabeled glucose and harvested at exponential phase. Fully labeled WT cultures and unlabeled culture were mixed and metabolites were extracted for LC-MS analysis. The relative pool sizes between WT and mutant strains were calculated based on isotopomer enrichment ratios.

6.2.5 Protein purification. Enzymes (Table S2) were individually overexpressed and purified from recombinant E. coli strains as described in our recent paper (Liu et al., 2017). Freshly purified protein was concentrated using Amicon Ultra-15 centrifugal filter devices (Millipore, Darmstadt, Germany). Concentrated protein was applied to a PD-10 column (GE Healthcare) equilibrated with storage buffer (100 mM phosphate buffer, 10% glycerol, pH 7.6) for buffer exchange. Protein concentrations were measured with a Pierce BCA protein assay kit (Thermo Fisher Scientific, MA). Proteins were stored at -80 °C after flash freezing in liquid nitrogen.

6.2.6 In vitro assays of EMPP and EDP enzymes. In all in vitro assays, 0.1 to 1.0 μM of each protein component, 0.5 mM ATP, 2 mM ADP, 2.5 mM NAD$^+$ were added to a reaction buffer
containing 50 mM sodium phosphate buffer (pH 7.6), 1 mM potassium chloride, 1 mM Tris (2-carboxyethyl) phosphine (TCEP), 2 mM magnesium chloride, 0.2 mM 2,3-bisphosphoglyceric acid (2,3-BPG). For EDP related assays, 2.5 mM NADP$^+$ and 0.2 mM FeSO$_4$ were added as well. Reactions were initiated by adding 5 mM glucose and incubated at 37 °C. Upon adding glucose to initiate the reactions, all contents were quickly mixed, and the increase in absorbance at 340 nm due to NAD(P)H formation was measured in a microplate reader (as a convenient way for tracking glycolytic activity). In brief, for individual enzyme activity, reactions were measured spectrophotometrically and the unit (U) or amount of purified protein that converted one μmol of substrate into product per minute was determined.

6.3 Results and Discussion

6.3.1 Comparative studies on isozymes’ functions under in vitro and in vivo conditions. E. coli phosphofructokinase, Pfk, has two isozymes that convert F6P to FBP. Under in vivo conditions, PfkA has higher copy numbers than PfkB (Table 6.1), and knocking out pfkA significantly impairs E. coli glucose metabolism (Table 6.2). In contrast, knocking out the isoenzyme gene pfkB has a negligible effect on E. coli growth (Robinson and Franenkel, 1978; Long et al., 2016). Under in vitro conditions here, purified PfkB exhibits twice the enzyme activity (U/mg) as PfkA and removal of PfkB has minimal impact on the in vitro glycolysis reaction rates (Table 6.2). Following Pfk, the isozymes, fructose-bisphosphate aldolase-A and B (FbaA and FbaB) reversibly cleave FBP (Figure 6.1). Under in vivo conditions, fbaA knockouts are lethal to E. coli, while ΔfbaB does not affect cell growth (Table 6.2). Under in vitro conditions, FbaA has a higher activity than FbaB and the removal of FbaA slowed down but did not stop the glycolytic reaction. As expected, all tested isozymes in Table 6.2 were interchangeable under in vitro conditions. For example, in the absence of PfkB or FbaA, in vitro EMP flux could be recovered if excessive PfkA
or FbaB were added. We also tested the double deletion of Pfk or Fba and different crowding effects under *in vitro* conditions (Table 6.2). Removal of both PfkA and FbaA isozymes caused the highest reduction of the EMP flux, while removal of PfkA and FbaB showed minimal impact. In summary, *in vitro* systems represent homogenous environments, and multiple systems level interactions are lost (Graham et al., 2007). Therefore, pathway function correlates well with activities and concentrations of individual enzymes. On the other hand, intracellular systems may not be ‘well mixed’ and enzyme-enzyme interactions or enzyme clustering may play a key role in directing EMP fluxes (Sweetlove and Fernie, 2018; Shearer et al., 2005). For example, it has been reported that Pfk and Fba enzymes form channels via electrostatic forces to confer catalytic advantages with minimal enzyme usage (Matic et al., 2001), and metabolic flux can be highly nonlinear to enzyme expression levels (Siddiquee et al., 2004).

**Table 6.1. Glycolysis genes and reported enzyme kinetics.** Gene location in *E. coli* (taxid:679895) was determined through NCBI nucleotide blast. The standard condition for $\Delta G$: pH = 7.2, ionic strength = 0.25 M, metabolite concentrations = 1 M. $^a$ (Taniguchi et al., 2010), $^b$ The actual $\Delta G$ was calculated by eQuilibrator (Flamholz et al., 2012) using reported metabolite concentrations (Bennett et al., 2009), $^c$ (Flamholz et al., 2013).

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Reaction</th>
<th>Gene</th>
<th>Gene location</th>
<th>In vivo conc. (nM)$^a$</th>
<th>$\Delta G$ (standard condition) (kJ/mol)</th>
<th>$\Delta G$ (actual, kJ/mol)$^b$</th>
<th>$k_{cat}$ (1/s)$^c$</th>
<th>$K_{M}$ (mM)$^c$</th>
<th>Measure d enzyme activity (U/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper glycolysis</td>
<td>Glc-&gt;G6P</td>
<td><em>glk</em></td>
<td>2501940-2502905</td>
<td>410</td>
<td>0.08</td>
<td>2.12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>G6P-&gt;F6P</td>
<td><em>pgi</em></td>
<td>4223686-4225335</td>
<td>169</td>
<td>2.5</td>
<td>-0.3</td>
<td>120</td>
<td>0.28</td>
<td>NM</td>
</tr>
<tr>
<td>Pathway</td>
<td>Gene</td>
<td>Start</td>
<td>End</td>
<td>Km</td>
<td>Vmax</td>
<td>Km'</td>
<td>Vmax'</td>
<td>118</td>
<td>37</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
<td>-------</td>
<td>-----</td>
<td>----</td>
<td>------</td>
<td>-----</td>
<td>------</td>
<td>-----</td>
<td>----</td>
</tr>
<tr>
<td>F6P-&gt;FBP</td>
<td>pfkB</td>
<td>1800627-1801556</td>
<td>167</td>
<td>0.16</td>
<td>0.71</td>
<td>1.86</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FBP-&gt;GAP+ DHAP</td>
<td>fbaA</td>
<td>3063524-3064603</td>
<td>676</td>
<td>5.5</td>
<td>-0.5</td>
<td>10.5</td>
<td>0.17</td>
<td>1.97</td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>fbaB</td>
<td>2170991-2172043</td>
<td>ND</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DHAP-&gt;GAP</td>
<td>tpiA</td>
<td>4100668-4101435</td>
<td>768</td>
<td>5.5</td>
<td>-0.5</td>
<td>9000</td>
<td>1</td>
<td>NM</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PPP</td>
<td>G6P-&gt;6PGL</td>
<td>zwf</td>
<td>1929096-1930571</td>
<td>113</td>
<td>-5</td>
<td>0.0</td>
<td>174</td>
<td>0.174</td>
<td>7.11</td>
</tr>
<tr>
<td>ED only</td>
<td>6PGL-&gt;6PG</td>
<td>pgl</td>
<td>794042-795037</td>
<td>ND</td>
<td>-23.2</td>
<td>-31.2</td>
<td>1900*</td>
<td>0.02*</td>
<td>NM</td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lower glycolysis</td>
<td>GAP-&gt;BPG</td>
<td>gapA</td>
<td>1857028-1858023</td>
<td>2380</td>
<td>3.9</td>
<td>-1.9</td>
<td>268</td>
<td>0.89</td>
<td>4.71</td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<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></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
ND: no data available. NM: not measured. 
Source organism is *E. coli* unless noted by *

**Table 6.2. In vitro test of EMP isoenzymes via a factorial experiment.** In *vitro* rate changes (%) based on the control condition after the removal or addition of Pfk and Fba isoenzymes. The control condition is all EMP enzymes at a concentration of 100 nM. In *vivo* rate changes were based on a recent report (Long et al., 2016). Confidence intervals (95% CI) were determined through reiterations of curve fitting calculations resulting in corresponding lower and upper bounds (LB, UB).

<table>
<thead>
<tr>
<th>PfkA (nM)</th>
<th>PfkB (nM)</th>
<th>FbaA (nM)</th>
<th>FbaB (nM)</th>
<th><em>In vitro</em> rate change (95% CI: LB, UB)</th>
<th><em>In vivo</em> rate change</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0% (-2%, +103%)</td>
<td>-75%</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>-42% (-34%, -51%)</td>
<td>-8%</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>-71% (-61%, -83%)</td>
<td>-100% (lethal)</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>-5% (-2%, -9%)</td>
<td>~0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>---------</td>
<td>------------------</td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>-71%</td>
<td>(-55%, -73%)</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>-37%</td>
<td>(-34%, -40%)</td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>-13%</td>
<td>(-2%, -26%)</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>-42%</td>
<td>(-37%, -49%)</td>
</tr>
<tr>
<td>500</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>-26%</td>
<td>(-21%, -35%)</td>
</tr>
<tr>
<td>1000</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>+144%</td>
<td>(+131%, +165%)</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0</td>
<td>500</td>
<td>+113%</td>
<td>(+103%, +122%)</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0</td>
<td>1000</td>
<td>+161%</td>
<td>(+157%, +163%)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Control plus 225g/L polyethylene glycol 3550</td>
<td>+103% (-19%, +110%)</td>
</tr>
</tbody>
</table>

### 6.3.2 The effect of ED pathway overexpression on *in vitro* and *in vivo* glycolytic rates.

To explore the flux control of glycolytic pathways, the competing native ED pathway was overexpressed (ED pathway enzymes underlined; Figure 6.1). Reaction thermodynamics and kinetics of the ED reactions have advantages over the EMP pathway, and the ED pathway requires fewer reaction steps and much lower protein synthesis costs (Flamholz et al., 2013). In a cell-free system, we tested glycolytic enzymes with concentrations ranging from 100 to 1000 nM (enzymes included in *in vitro* reactions are listed in Table S2) (Liu et al., 2017). Although the ED pathway demonstrated catabolic rates lower than the EMP pathway at the same enzyme concentrations (Figure 6.2a, Table 6.1), *in vitro* overexpression of ED enzymes could significantly increase glucose conversions (Figure 6.2b). As expected, *in vitro* introduction of the ED enzymes could
also recover glucose catabolic rates in the absence of PfkA or FbaA, two key enzymes for *in vivo* EMP pathway control.

**Figure 6.2. In vitro reaction rates of the EMPP and EDP.** (A). Reaction rates with protein concentrations at 1000 nM. Reactions were assumed to be first-order processes and the rate constants, k (1/hr), were calculated based on absorbance curves and minimizing the sum of squared residuals. (B). Reaction rates with varied enzyme concentrations.

*E. coli* metabolism is highly regulated, namely through transcriptional and allosteric regulation, which is closely tied to catabolite repressor/activator proteins and carbon substrates. Despite bypassing transcriptional regulation via plasmid engineering, overexpressing the native ED pathway does not increase flux through the ED pathway, and the engineered strain still prefers to utilize its EMP pathway rather than its ED pathway (Figure 6.3). One possible explanation is that EMP channels protect their intermediates from competing reactions such as the ED pathway. To test this hypothesis, presumed channels must be disrupted while still maintaining enzyme functions. In *E. coli*, the PTS uses PEP to phosphorylate glucose into G6P and is thought to be the first step of EMP channeling (Zhao et al., 2018; Flores et al., 2002). By knocking out the PTS, the
presumed channel could be disrupted, facilitating flux redistributions from the EMP towards the competing ED pathway. Thereby, we constructed PTS-knock-out mutants to examine their in vivo flux redistributions via metabolic flux analysis (Figure 6.3). The $\Delta$ptsG mutant showed less flux through the EMP pathway, while its PP pathway flux was significantly elevated. After overexpression of native edd and eda in the $\Delta$ptsG mutant, the ED pathway effectively hijacked PP pathway fluxes from the 6PG node and became a dominating glycolysis route. Intracellular metabolite analyses further confirmed the flux redistributions (Figure 6.4). The $\Delta$ptsG mutant had much smaller pool sizes in upper EMP pathway metabolites (FBP, GAP/DHAP and PGA) but larger PEP concentrations than WT strain. Furthermore, overexpression of the ED pathway drained the G6P and F6P pools, and the $\Delta$ptsG or ED pathway overexpression had slight impact on the concentrations of TCA cycle metabolites.
**Figure 6.3. Metabolic flux analysis (determined by WUFlux) in different *E. coli* strains** (A). WT (B). WT with an overexpressed ED pathway (ED) (C). a PTS mutant (ΔptsG) and (D). a PTS mutant with an overexpressed ED pathway (ΔptsG + ED). Maximal growth rates (GR) are given as hr⁻¹ based on biological triplicates, presented at the top right of every flux map. Flux map arrows represent the flux distribution throughout the central pathways. The table represents the percentage of G6P flux being directed along the EMP, PP or ED pathways. Experimental flux ratios were determined by ¹³C-MFA. Standard deviations were calculated based on the measurement errors of MID data.

**6.3.3 Transient labeling of *E. coli* EMP intermediates.** To explore heterogeneous metabolite pools in the EMP pathway, we used isotopically nonstationary ¹³C-labeling, which can probe the bulk pool’s labeling dynamics and be used to investigate substrate channeling (Sweetlove and Fernie, 2018). In our transient study, we observed faster FBP, DHAP, and 3PGA labeling (bulk M+6 MID and M+6 counts) than their upstream precursors G6P and F6P (Figure 6.5a). This reversal of the labeling order could be explained by the channeling theory: G6P is not uniformly mixed in the cytosol, and instead, the portion of G6P outside of the enzyme channel becomes less metabolically active, contributing largely to the M+0 pool during dynamic labeling experiments. Such dilution of intracellular metabolite labeling has been reported by isotopic nonstationary flux analysis of bacterial metabolisms (Huege et al., 2011; Young et al., 2011; van Winden et al., 2001). Conversely in the ΔptsG and ΔptsG+ED mutant, labeling between G6P, F6P, and FBP was strictly sequential (labeling enrichment: G6P>F6P>FBP), along with a decrease in the FBP pool (Figure 6.4). Moreover, ΔptsG and ΔptsG+ED mutants had much slower labeling in FBP than other cascade metabolites, PGA and DHAP, forming a "V" shape in Fig. 6a. The higher labeling of FBP downstream products was consistent with observed elevated bypass fluxes via PP pathway and ED pathway in these mutants. During dynamic ¹³C labeling of ΔptsG mutants (Figure 6.5b), FBP contained much more M+3 isotopomer (3 labeled carbons) than its precursor F6P due to the
increase of reversed Fba reactions from C3 sugar phosphates (PGA/DHAP→FBP). This result was consistent with the report of the gluconeogenesis in PTS mutants (Flores et al., 2005).

Figure 6.4. Metabolite pool size comparison to WT E. coli K-12. Pool size analysis of PTS mutant (ΔptsG) and a PTS mutant with an overexpressed ED pathway (ΔptsG + ED) compared to the WT. A ratio of 1 means both strains had the same pool size per biomass (n=3).
Figure 6.5. Labeling dynamics following $^{13}$C$_6$-glucose pulse in different E. coli strains (A). On the left axis, the fully labeled or M+6/M+3 mass isotopomer ratio is plotted for WT, ΔptsG, and ΔptsG + ED. Mass isotopomer ratios are the relative abundance of mass isotopomers where M+0 is fully unlabelled, M+3 is three labeled carbons (fully labeled for PGA and PEP) and M+6 is six labeled carbons (fully labeled for FBP). Cultures were pulsed with $^{13}$C$_6$-glucose for 5 s. The right y-axis is the absolute M+6 cps. (n=2) (B). Mass isotopomer distribution of M+3 in F6P and FBP (n=2).

6.3.4 Broader perspective on the glycolysisosomes. The common view of cell metabolism and metabolic modeling assumes enzymes and metabolites are homogenously distributed throughout the cell, but amidst surmounting observations, this assumption fails. In this study, both in vitro and in vivo glycolytic systems were compared. We demonstrated knocking out the PTS in E. coli loosens the metabolic flux network and changes reaction directions towards competing pathways. Although this work did not directly prove that intracellular enzymes self-organize to form subtle metabolons$^{36}$, dynamic labeling experiments provide reasonable in vivo channeling tests for systems perturbed at possible channeling flux control nodes (Sweetlove and Fernie, 2018; Abernathy et al., 2017b; Shearer et al., 2005). Intuitively, Malate $\rightarrow$ Oxaloacetate $\rightarrow$ Citrate in the TCA cycle of eukaryote cells have been proved to be highly channeled (Wu and Minteer, 2015; Bulutoglu et al., 2016). In this study, labeling rates among PEP, citrate, malate, and succinate in
the wild type *E. coli* were compared in Figure S2. After pulsing cells with $^{13}$C$_6$-glucose, PEP carboxylase led to amount of $^{13}$C$_3$ (M+3) labeled C4 TCA metabolites (e.g., malate and oxaloacetate), while pyruvate dehydrogenase produced fully labeled Acetyl units (M+2). We expected that citrate synthase would generate $^{13}$C$_5$ (M+5) citrate. However, no M+5 citrate could be detected (only M+2 or M+3) by LC-MS. Such enigmatic observations were consistent with confounding labeling patterns due to metabolite channeling (Abernathy et al., 2017b), and in turn could introduce errors into $^{13}$C-based flux analysis.

**6.3.5 Conclusions.** Proximity channeling through metabolic pathways cannot be measured intracellularly via typical tools (such as protein imaging and protein structural analysis) due to weak and possibly transient interactions. Therefore, the hypothesis of metabolite channeling through the EMP pathway is still controversial (Graham et al., 2007; Shearer et al., 2005). This study compared *in vitro* and *in vivo* glycolytic activities, which elucidated different performances of glycolysis enzymes under two modes and offered explanations of intracellular EMP functions using the channeling theory. The observed differences in metabolite isotopic dilutions may be due to disruption of metabolite channels and provides a general *in vivo* approach for probing organisms for substrate channeling. However, this study still cannot answer many remaining questions regarding the structural formation of glycolysisosomes or potential mechanisms for metabolite channeling (Sweetlove and Fernie, 2018; Wheeldon et al., 2016). Novel *in vivo* tools are needed to further understand and validate proximity channeling in native pathways.
Chapter 7: Conclusions and future directions

7.1 Conclusions

Cyanobacteria hold enormous potential as cost-effective biocatalysts for the direct conversion of CO₂ to value-added chemicals. However, their scaled industrial use has been limited due to bioprocess and economic feasibilites. The bioproduction of small molecules, such as biofuels, bioplastics, or some bioactive compounds, requires genetic engineering of cyanobacteria, effective redirection of carbon and robust strains for outdoor cultivations. To rationally engineer these photosynthetic organisms for efficient bioproduction requires further knowledge about photosynthesis and carbon assimilation efficiencies, metabolic regulation, and genome to phenome mapping.

This thesis used isotopically nonstationary metabolic flux analysis, which quantifies cellular carbon metabolism, to investigate desired cyanobacterial phenotypes for biotechnology applications. In the second Chapter, the fast growing cyanobacteria, *Synechococcus* UTEX 2973, was probed, which revealed ideal carbon partitioning for biomass growth, fast reincorporation of carbon, and considerable photorespiration for properly balancing increased photosynthetic rates. The fast growth of *Synechococcus* UTEX 2973 allows cyanobacteria to compete with its heterotrophic counterpoints, aiding in productivity rates for scaled processes. Increased carbon assimilation in *Synechococcus* UTEX 2973 did not require substantial central carbon metabolism rewiring. The robust central metabolism flux ratios discovered in the differing phenotypes prompted us to explore metabolite channeling as a flux regulation mechanism. In eukaroytes, channeling is often associated with membranes; thus in Chapter 3, we examined a carboxysome-deficient mutant with INST-MFA and pool size analysis. We found by knocking out the
carboxysome, we could redirect fluxes to important metabolites for many biotechnology applications by possibly creating a more homogeneous cytoplasm. The carboxysome-deficient mutant was found to accumulated UDP glucose and pyruvate, which can aid biotechnology applications as demonstrated in Clark et al., 2018 through lactate production. Additionally, these high CO₂ requiring phenotypic strains represent a method of containment for genetically modified cyanobacteria. Despite the drastically different cyanobacterial phenotypes examined, many carbon fluxes were found to be robust, while metabolite pool sizes were found to change more drastically, and labeling dynamics revealed possible substrate channels. High flux through sugar phosphate pathways and large nucleotide sugar pools in cyanobacteria motivated us to demonstrate proof-in-concept photoautotrophic production of a high-value saccharide, heparosan, in cyanobacteria.

In Chapter 5, metabolite channeling was further explored in silico and reviewed across species, demonstrating the prevelance of substrate channeling in nature. Using a model system (E. coli) in Chapter 6, this thesis further elucidated channeling in central catabolic pathways to try and advance our understanding of this important metabolic controller. It was found the phosphotransferase system may anchor glycolytic channels, restricting flux to other catabolic pathways, and many enzymes in the glycolytic pathways elucidated different performances under in vivo and in vitro modes. While genomic and transcriptomic modifications are typically the dominating metabolic engineering strategies, protein-protein interactions (e.g., metabolite channeling), enzyme-metabolite interactions and catabolic regulation in cyanobacteria still represent a relatively unexplored field.
7.2 Future directions

7.2.1 Future directions for cyanobacteria engineering. This thesis revealed the importance of metabolite and channeling regulation, showed further challenges that face photomixotrophic growth, and translated our results into engineering strategies for cyanobacteria. While cyanobacteria require simple nutrients, grow fast and can be genetically manipulated, their complex metabolic regulation and inherent challenge of converting oxidized CO$_2$ to a more reduced form may mean these generally regarded as safe organisms (GRAS) are better suited for the production of a high-value products rather than commodity chemicals. Learning more about how these organisms regulate their metabolism and other cellular processes will help propel cyanobacterial biorefineries. Photofermentation is one way to increase cyanobacteria productivity and titer using cheap carbon sources, such as glycerol. However under different conditions, cyanobacteria strains utilized organic substrates in a variety ways demonstrating the need for further investigation in photofermentations and catabolic repression in cyanobacteria. Furthermore, while many metabolic engineering strategies redirect carbon fluxes by knocking out or down competing pathways, carboxysome-deficient mutants represent another strategy for the redirection of fluxes that does not impair growth or require nitrogen deficient conditions to produce polysaccarides. Finally, using the fast growing cyanobacteria strain, productivites can be enhanced or engineering strategies can be transferred to other model cyanobacteria strains. Concentrated efforts are needed to fully understand robust platform strains for a rational bioproduct in order to facilitate stable scaled bioprocesses.

7.2.2 Future directions for metabolite channeling. Channeling improves bioconversion rates, stabilizes low copy number enzymes, and reduces side products from promiscuous enzymes. It
has been hypothesized to control fluxes through metabolic pathways and facilitate cascade reactions in a crowded environment. Channeling is a natural evolution outcome; yet despite experimental evidence, there are many unresolved questions. Superior modeling and thermodynamic approaches need to be developed that integrate all the metabolite controllers represented in Figure 1.3. Little knowledge exists about how bacteria spatially organize their enzymes; however recent research has found that heterologous enzyme expression is influenced by its location on the genome, although it is not clear why (Englaender et al., 2017). During pathway engineering, overexpression of a gene cluster is in general more efficient than expressing multiple enzymes from different sources. Although it is common for a gene cluster to translate a biosynthesis pathway, genes of cascade enzymes in central metabolism are often not located next to each other but still may form channels. Thus, it is unclear whether gene location or other factors (e.g., pH, ion strength, electrostatic interactions, metabolites, or membranes) coordinate enzymes. To test this hypothesis, genetic engineering can disrupt the presumed channels via plasmid or chromosome relocations. Structure-guided enzyme mutations can be performed to weaken enzyme-enzyme interactions without affecting enzyme activity to monitor channel effects. Chemical cross-linking combined with mass spectrometry may examine protein assemblies and reveal both the structure and the interactions of cascade enzymes (Liu et al., 2015; Tan et al., 2016).

Additionally, genome scale modeling can assess channeling using experimental $^{13}$C-MFA and transcriptomic data. Predicted intracellular fluxes from genome scale modeling can be used as a theoretical baseline for comparing against experimentally determined pathway fluxes via $^{13}$C-MFA to identify pathways that are showing higher flux than expected based upon expression activity. This analysis will be useful for considering the coordinated effects of regulatory
mechanisms that can lead to the identification of pathways exhibiting channeling. Finally, metabolic engineers often build heterologous enzymes or synthetic pathways to redirect metabolic fluxes towards the desired products (Bogorad et al., 2013), but natural channeling mechanisms may cause design flaws. Increases in enzyme copy numbers do not always aid in the expression of a heterologous pathway (Englaender et al., 2017) nor does uptake of a significant intermediate aid in heterologous pathway production (He et al., 2017). These observations leave additional questions for metabolic engineers in regards to integration of channeling knowledge for strain development and computational strain design.
8. References


Arrivault S. Use of reverse-phase liquid chromatography, linked to tandem mass spectrometry, to profile the Calvin cycle and other metabolic intermediates in Arabidopsis rosettes at different carbon dioxide concentrations. Plant J. 2009;59:826.


Kamravamanesh D, Kovacs T, Pfugl S, Druzhinina I, Kroll P, Lackner M, Herwig C. Increased poly--hydroxybutyrate production from carbon dioxide in randomly mutated cells of
cyanobacterial strain *Synechocystis* sp. PCC 6714: Mutant generation and characterization. *Biores. Tech.* 2018;266:34-44.


Ma F, Jazmin LJ, Young JD, Allen DK. Isotopically nonstationary $^{13}$C flux analysis of changes in *Arabidopsis thaliana* leaf metabolism due to high light acclimation. *Proc Natl Acad Sci USA.* 2014;111:16967–16972.


Mangan NM, Brenner MP. Systems analysis of the CO$_2$ concentrating mechanism in cyanobacteria. eLife. 2014;02043.


Ungerer J, Lin PC, Chen HY, Pakrasi HB. Adjustments to photosystem stoichiometry and electron transfer proteins are key to the remarkably fast growth of the cyanobacterium *Synechococcus elongatus* UTEX 2973. *mBio.* 2018;9(1):e02327-17.


### 9. Appendix 1: Supporting Information for Chapter 2

Deciphering cyanobacterial phenotypes for fast photoautotrophic growth via isotopically nonstationary metabolic flux analysis

**Biomass composition analysis.** Protein and amino acid compositional analysis was performed by the Molecular Structure Facility, University of California (Davis, CA) and measured by hydrolysis and oxidation with performic acid. For lipid measurement, biomass was mixed with chloroform-methanol-water (2:1:0.8) and shaken (200rpm) for 3 h. After the complete extraction of lipids from the biomass, the resulting chloroform layer was dried to determine the weight of the lipid residue. The fatty acid composition was then analyzed via GC-MS after transesterification. Glycogen was analyzed colorimetrically at 680nm as a reducing sugar based on a previously published anthrone protocol (Seifter et al., 1949). Chlorophyll *a* was extracted with 100% methanol in the dark for 24 h. The concentration was determined spectrophotometrically at 665 and 652 nm according to previous literature (Porra et al., 1989). Other photosynthetic pigments were not accounted in the lumped biomass equation due to minimal biomass contribution (Schledz et al., 2001). Ash weight of the biomass samples was calculated from the mass difference after complete combustion. Total carbohydrate was measured based on sulfuric acid hydrolysis protocols (www.nrel.gov/biomass/analytical_procedures.html), which was performed by Dr. Wei Liao group at Michigan State University. Nucleotide contents were estimated based on mass balance (Shastri and Morgan, 2005).

**Ion-pairing LC-MS/MS method:** For LC-MS analysis, the pelleted samples were extracted with a chloroform-methanol method (Ma et al., 2014). The aqueous phase from the extraction was transferred into GC vials, lyophilized, and stored at -80 °C. Dried samples were reconstituted in water for LC-MS/MS analysis. Ion-pairing LC-MS/MS was performed on a Shimadzu HPLC linked to an ion-trap triple quadrupole MS/MS system (4000 QTRAP and 6500 QTRAP, AB Sciex Instruments) at the Proteomics and Mass Spectrometry Facility, Donald Danforth Plant Science Center, St. Louis, MO. The mobile phases utilized for a gradient elution were 10 mM tributylamine (Acros, Belgium) and 11 mM acetic acid (A) and methanol (B). A Synergi Hydro-RP column (80 Å, 150 x 2.00 mm, Phenomenex Inc., CA) was used with the following gradient profile: 0% B (0-5 min), 0-45% B (5-50 min), 45-90% B (50-52 min), 90% B (52-57 min), 90-0% B (57-59 min), 0% B (59-64 min). Negative ionization mode and multiple-reaction monitoring mode were used, and all data acquisition and analysis was performed on the Analyst software (AB/MDS Sciex).

**Mutant Construction.** For the *zwf* mutant, a plasmid was designed to target a 1000 bp region (extending from nucleotides 41-1041) of the *zwf* open reading frame for replacement with a kanamycin resistance cassette by double homologous recombination. A circular polymerase
extension cloning (CPEC) method (Quan and Tian, 2011) was used to generate an upstream (1000 bp)-kanamycin-downstream (1000 bp) construct in pBR322. Sequencing and restriction digestion were used to confirm the resulting plasmid (pSL2480), which was introduced into Synechococcus 2973 cells by tri-parental conjugation as previously described (Yu et al., 2015). To create the cargo strain, the pSL2480 plasmid was first transformed into competent HB101 E. coli cells containing the pRL623 helper plasmid. Overnight cultures (100 µL each) of the cargo strain and the conjugal strain (pRL443) were washed with distilled water and mixed with 200 µl of washed Synechococcus 2973 cells. Cell mixtures were plated on BG-11 + 5% LB (v/v) agar plates containing 10 µg/ml kanamycin. Plates were placed under continuous illumination (~100 μmol photons•m⁻²•s⁻¹) at 38 °C. Mutant colonies appeared within 4-5 days and segregation was confirmed by PCR.

A cpf1 based CRISPR system was used to delete 6-phosphogluconolactonase and succinate dehydrogenase from start codon to stop codon from the chromosome (Ungerer and Pakrasi, 2016). In brief, oligos targeting the gene’s coding region were cloned into the aarI sites on pSL2680 to target the nuclease to the chromosomal region containing the gene of interest. Next a region 2kb in length containing 1kb of upstream and downstream sequences were assembled into the resulting plasmid using Gibson assembly. The resulting plasmid was introduced into Synechococcus 2973 cells by tri-parental conjugation. After repatchings onto selective media, patches were sequenced and a patch that was positive for the deletion was started in BG-11 without antibiotics and grown to an OD₇₃₀ of 1.0. Serial dilutions were plated to obtain single colonies. Colonies that were cured of the editing plasmids were then sequenced to verify the deletion of 6-phosphogluconolactonase and succinate dehydrogenase.

**INST-MFA model formulation:** The INCA model listed in Supporting Table 1 and its results listed in Supporting Table 3, 4 and 5, along with the lumped biomass equation generated from Supporting Table 2 are based on the following:

I. The INCA platform is convenient because it does not need to estimate metabolite pool sizes; however, the platform has to solve numerically stiff ODE equations for isotopomer changes through very different time scales of metabolite labeling (Nöh et al., 2007). Therefore, it is computationally expensive for flux calculations and confidence interval estimations.

II. Metabolite levels and fluxes were assumed to be constant throughout the experiment and were not perturbed by the addition of sodium bicarbonate. The effect of isotope discrimination was assumed to be negligible.

III. Succinate and fumarate are symmetrical molecules, and this was accounted for within the INCA model.

IV. It is unknown whether the TCA cycle in Synechococcus is closed by the GABA shunt (Xiong et al., 2014) or the oxoglutarate dehydrogenase pathway (Zhang and Bryant, 2011); thus the reaction network was modeled as AKG $\rightarrow$ SUC + CO₂, lumping both shunts into one closed TCA reaction. The SSR of the model was not changed when this reaction was removed, obtaining a p-value of $6.5\times10^{-1}$ based on the chi-square statistical test with one degree of freedom. This indicates that the flux was not significantly different from zero.
However, the sequential labeling of free glutamate and succinate suggests a low, yet measurable GABA shunt in *Synechococcus* 2973 (Supporting Figure 4).

V. Although there is limited to no evidence for a bacterial glyoxylate shunt in *Synechococcus*, we included this pathway in the model to be consistent with the previous work. Addition of the glyoxylate shunt only slightly changes model fitting quality, while the resulting flux of glyoxylate shunt was still close to zero ($p < 0.001$, based on chi-square statistical test with two DOFs).

VI. To test the presence of OPPP, we removed these two equations (G6P→6PG, 6PG→Ru5P +CO2) from the model. This resulted in reduced set of DOFs. Without the OPPP, the model fitting quality was not significantly changed. This indicates very little flux through OPPP and was not a consequence of the choice of measurements because oxidation that produces CO2 and a pentose phosphate would be unique to the labeling in the Calvin cycle. This was additionally confirmed by OPPP knock out mutant, Δzwf.

VII. Confidence intervals were evaluated using two different methods due to numerical stiffness problems that were encountered. The approaches gave similar results and therefore served to confirm each other. Numerical stiffness problems are expected due to label accumulation processes occurring over a large range of time scales. To solve the entire non-stationary isotopomer reaction network with high numerical precision is much more difficult than solving for steady state isotopomer balance equations (Nöh et al., 2007). Thus, in addition to parameter continuation, we used the Monte Carlo method to examine to confidence intervals based on potential measurement timing errors (Zhao and Shimizu, 2003). Specifically, the sample quenching time points were perturbed by normally distributing noise within measurement timing variations (±3 seconds). INCA was then restarted for flux calculations. By repeatedly running INCA using perturbed timing of isotopomer data, the resulting distribution of flux values allowed for the estimation of flux standard errors when measuring the rapid turnover of intracellular metabolites subjected to manual errors as seen in Supporting Table 3. Additionally, while INCA-based MFA did not require the pool size estimation, if we added estimated pool sizes (MAL, PEP, AcCoA, SUC, CIT) from isotopomer ratio analysis (Bennett et al., 2008), the flux results from PBRs had improved SSR and facilitated determination of unresolved confidence intervals (Supporting Table 3). As expected, net fluxes remained relatively similar to flux results without pool size constraints.

VIII. The uneven distribution of the residuals in the shaking flask model demonstrated that they were less consistent and of reduced quality (Supporting Figure 3). Dilution factors (Young et al., 2011) were required for several metabolites that were significantly less labeled than their downstream products, including: 3PGA, F6P, FBP, and SBP (Figure 5). In INCA platform, dilution parameters are often required to describe the lack of equilibrium between labeled and unlabeled pools or labeling dilution from unknown resources. The dilution factors were more crucial for the shaking flask conditions, which represented a higher degree of inactive metabolite pools. Besides the inactive pools prescribed for the PBR, an additional dilution for DHAP, CIT, PEP, and MAL improved the SF-based model fitting. Poor quality of SF flux results is not surprising because sub-optimal and light-limiting SF conditions causes sub-populations of cyanobacteria with different metabolisms (He et al., 2015).

IX. Several assumptions were made when quantifying the biomass components. Glycogen was assumed to have a molecular weight of 666.58 g/mol, chlorophyll a was assumed to have
a molecular weight of 893.5 g/mol, lipids were assumed to have a molecular weight of 270 g/mol, DNA was assumed to have a molecular weight of 487 g/mol, and RNA was assumed to have a molecular weight of 499.5 g/mol.

**Biomass Composition Analyses of *Synechococcus 2973* and *Synechococcus 7942***. Standard deviations are reported from biological duplicates to quadruplicates. † or * indicates p-value <0.05 using two-tailed equal variance Student’s t-Test.

<table>
<thead>
<tr>
<th>Biomass components (unit: % of gDCW)</th>
<th><strong>Syn. 2973- Photobioreactor</strong></th>
<th><strong>Syn. 2973- Shaking Flask</strong></th>
<th><strong>Syn. 7942- Photobioreactor</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Protein</td>
<td>52.7 ± 2.5</td>
<td>49.4 ± 2.1</td>
<td>40.9 ± 0.4</td>
</tr>
<tr>
<td>Lipid</td>
<td>9.3 ± 0.6</td>
<td>13.2 ± 2.4</td>
<td>10.7 ± 0.6</td>
</tr>
<tr>
<td>Glycogen</td>
<td>1.5 ± 0.5††</td>
<td>6.0 ± 1.0*</td>
<td>12.6 ± 4.3†</td>
</tr>
<tr>
<td>Other sugars</td>
<td>6.5 ± 0.7</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Ash</td>
<td>7 ± 1</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Chlorophyll a %</td>
<td>1.2 ± 0.2</td>
<td>1.0 ± 0.3</td>
<td>1.2 ± 0.1</td>
</tr>
<tr>
<td>Chlorophyll a (μg/mL/OD$_{730}$)</td>
<td>4.4 ± 0.5*</td>
<td>7.4 ± 0.3*</td>
<td>5.2 ± 0.6</td>
</tr>
<tr>
<td>RNA/DNA</td>
<td>21.8ª</td>
<td>16.9ª</td>
<td>21.1ª</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Amino Acid profiles (unit: mmol/gDCW)</th>
<th><strong>Syn. 2973- Photobioreactor</strong></th>
<th><strong>Syn. 2973- Shaking Flask</strong></th>
<th><strong>Syn. 7942- Photobioreactor</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Glycine</td>
<td>0.44 ± 0.02</td>
<td>0.48 ± 0.07</td>
<td>0.34 ± 0.00</td>
</tr>
<tr>
<td>Proline</td>
<td>0.21 ± 0.01</td>
<td>0.24 ± 0.04</td>
<td>0.16 ± 0.00</td>
</tr>
<tr>
<td>Alanine</td>
<td>0.62 ± 0.03</td>
<td>0.67 ± 0.08</td>
<td>0.49 ± 0.01</td>
</tr>
<tr>
<td>Valine</td>
<td>0.31 ± 0.01</td>
<td>0.35 ± 0.06</td>
<td>0.25 ± 0.01</td>
</tr>
<tr>
<td>Leucine</td>
<td>0.46 ± 0.02</td>
<td>0.53 ± 0.08</td>
<td>0.37 ± 0.01</td>
</tr>
<tr>
<td>Isoleucine</td>
<td>0.23 ± 0.01</td>
<td>0.27 ± 0.04</td>
<td>0.19 ± 0.01</td>
</tr>
<tr>
<td>Methionine</td>
<td>0.08 ± 0.00</td>
<td>0.08 ± 0.01</td>
<td>0.06 ± 0.01</td>
</tr>
<tr>
<td>Amino Acid</td>
<td>Syn. 2973-Photobioreactor</td>
<td>Syn. 2973-Shaking Flask</td>
<td>Syn. 7942-Photobioreactor</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------------------------</td>
<td>-------------------------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>Cysteine</td>
<td>0.04 ± 0.00</td>
<td>0.04 ± 0.01</td>
<td>0.03 ± 0.00</td>
</tr>
<tr>
<td>Phenylalanine</td>
<td>0.20 ± 0.01</td>
<td>0.24 ± 0.04</td>
<td>0.15 ± 0.00</td>
</tr>
<tr>
<td>Tyrosine</td>
<td>0.16 ± 0.00</td>
<td>0.18 ± 0.03</td>
<td>0.13 ± 0.00</td>
</tr>
<tr>
<td>Histidine</td>
<td>0.06 ± 0.00</td>
<td>0.08 ± 0.011</td>
<td>0.04 ± 0.00</td>
</tr>
<tr>
<td>Lysine</td>
<td>0.21 ± 0.01</td>
<td>0.22 ± 0.04</td>
<td>0.16 ± 0.00</td>
</tr>
<tr>
<td>Arginine</td>
<td>0.27 ± 0.01</td>
<td>0.28 ± 0.04</td>
<td>0.21 ± 0.00</td>
</tr>
<tr>
<td>Glutamate and Glutamine</td>
<td>0.60 ± 0.05*</td>
<td>0.44 ± 0.05</td>
<td>0.36 ± 0.00*</td>
</tr>
<tr>
<td>Aspartate and Asparagine</td>
<td>0.45 ± 0.02</td>
<td>0.49 ± 0.07</td>
<td>0.41 ± 0.01</td>
</tr>
<tr>
<td>Serine</td>
<td>0.27 ± 0.02</td>
<td>0.30 ± 0.04</td>
<td>0.21 ± 0.00</td>
</tr>
<tr>
<td>Threonine</td>
<td>0.29 ± 0.01</td>
<td>0.32 ± 0.05</td>
<td>0.22 ± 0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fatty acid profiles (wt%)</th>
<th>Syn. 2973-Photobioreactor</th>
<th>Syn. 2973-Shaking Flask</th>
<th>Syn. 7942-Photobioreactor</th>
</tr>
</thead>
<tbody>
<tr>
<td>C14:1</td>
<td>0.3 ± 0.1</td>
<td>1.1 ± 0.4</td>
<td>2.5 ± 0.1</td>
</tr>
<tr>
<td>C14:0</td>
<td>1.2 ± 0.5</td>
<td>0.7 ± 0.07</td>
<td>2.5 ± 0.3</td>
</tr>
<tr>
<td>C16:1</td>
<td>32.3 ± 0.1</td>
<td>39.9 ± 1.4</td>
<td>35.0 ± 0.1</td>
</tr>
<tr>
<td>C16:0</td>
<td>53.3 ± 1.3</td>
<td>56.3 ± 0.6</td>
<td>51.4 ± 0.1</td>
</tr>
<tr>
<td>C18:1</td>
<td>9.4 ± 0.1</td>
<td>1.8 ± 0.7</td>
<td>4.8 ± 0.4</td>
</tr>
<tr>
<td>C18:0</td>
<td>2.9 ± 0.9</td>
<td>0.3 ± 0.3</td>
<td>3.1 ± 0.1</td>
</tr>
</tbody>
</table>

NA: measured values from the biomass analysis service center had high noises.

*Estimated based on biomass composition and the reference (Shastri and Morgan, 2005). Previous study indicates that flux result is not sensitive to the measured nucleotide composition of the cells (Pramanik and Keasling, 1997).

Table S1. A complete list of reactions and atom transitions for the Synechococcus elongatus UTEX 2973 metabolic network.
Glycolysis and the Oxidative Pentose Phosphate pathway

G6P (abcdef) $\leftrightarrow$ F6P(abcdef)
G6P (abcdef) $\rightarrow$ 6PG (abcdef)
6PG (abcdef) $\rightarrow$ Ru5P (bcdef) + CO2 (a)
F6P (abcdef) $\leftrightarrow$ FBP (abcdef)
FBP (abcdef) $\leftrightarrow$ DHAP (cba) + GAP (def)
DHAP (abc) $\leftrightarrow$ GAP (abc)
GAP (abc) $\leftrightarrow$ 3PGA (abc)
3PGA(abc) $\leftrightarrow$ 2PGA (abc)
2PGA (abc) $\leftrightarrow$ PEP (abc)
PEP (abc) $\leftrightarrow$ PYR (abc)

The Citric Acid Cycle and Amphibolic reactions

PYR (abc) $\rightarrow$ AcCoA(bc) + CO2 (a)
OAA (abcd) + AcCoA (ef) $\rightarrow$ CIT (dcbfe)
CIT (abcdef) $\leftrightarrow$ ICIT (abcdef)
ICIT (abcdef) $\rightarrow$ AKG (abcde) + CO2 (f)
AKG (abcde) $\rightarrow$ SUC (bcde) + CO2 (a)
SUC (abcd) $\leftrightarrow$ FUM (abcd)
FUM (abcd) $\leftrightarrow$ MAL (abcd)
MAL (abcd) $\leftrightarrow$ OAA (abcd)
MAL (abcd) $\rightarrow$ PYR (abc) + CO2 (d)
PEP (abc) + CO2 (d) $\leftrightarrow$ OAA (abcd)
ICIT (abcdef) $\rightarrow$ SUC (dcef) + GLX(ab)
GLX (ab) + ACA (cd) $\rightarrow$ MAL (abcd)

Calvin-Benson-Bassham Cycle
Ru5P (abcde) ↔ X5P(abcde)
Ru5P (abcde) ↔ R5P (abcde)
Ru5P (abcde) → RuBP (abcde)
RuBP (abcde) + CO2 (f) → 3PGA (cde) + 3PGA (fba)
X5P (abcde) ↔ GAP (cde) + EC2 (ab)
F6P (abcdef) ↔ E4P (cdef) + EC2 (ab)
S7P (abcdefg) ↔ R5P (cdefg) + EC2(ab)
F6P (abcdef) ↔ GAP (def) + EC3 (abc)
S7P (abcdefg) ↔ E4P (defg) + EC3 (abc)
DHAP (abc) + E4P (defg) → SBP (cbadefg)
SBP (abcdefg) → S7P (abcdefg)

**Photorespiration**

RuBP (abcde) → 3PGA(cde) + 2PG (ba)
2PG (ab) → GLC (ab)
GLC (ab) → GLX (ab)
GLX (ab) + GLX (cd) → GA (abd) + CO2 (c)
GA (abc) ↔ 2PGA (abc)

**Glycogen Synthesis and breakdown**

G6P (abcdef) ↔ G1P(abcdef)
G1P (abcdef) ↔ GLYC (abcdef)

**Biosynthesis Syn. 2973-Photobioreactor**

161
Table S2. Equations used to form the lumped biomass equation. The equations represent the formation of amino acids, nucleotides, photosynthetic pigments, and lipids to central metabolites. Using measured and literature values for amino acids, carbohydrates, lipids, nucleotides and pigments, the stoichiometric molar quantity for central metabolites per 1 kg of DCW was determined.

<table>
<thead>
<tr>
<th>Amino Acids (Shastri and Morgan, 2005)</th>
<th>Carbohydrates and lipids (Shastri and Morgan, 2005)</th>
<th>Nucleotides (Yang et al., 2000) and pigments</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLU + ATP + 2<em>NADPH =&gt; PRO + NADP</em> + ADP</td>
<td>G6P =&gt; Glycogen</td>
<td>R5P + 1.2<em>ASN + 0.25</em>SER + 2<em>GLN + 0.25</em>GLY + FTHF+ 0.5<em>CO₂ + 2.3</em>H₂O + NADPH + 0.75<em>NAD + 8</em>ATP =&gt; DNA + 0.75<em>FUM + 2</em>GLU + THF + 8<em>ADP + 0.75</em>NADH + NADP*</td>
</tr>
<tr>
<td>PYR + GLU =&gt; ALA + AKG</td>
<td>7<em>AcCoA + 12</em>NADPH =&gt; C14 + 7<em>CoA + 12</em>NADP*</td>
<td>R5P + 1.2<em>ASN + 2.1</em>GLN + 0.54<em>GLY + 1.1</em>FTHF + 0.54<em>CO₂ + 2.2</em>H₂O + 0.79<em>NAD + 8.2</em>ATP =&gt; RNA + 0.75<em>FUM + 2.1</em>GLU + 1.1<em>THF + 8.2</em>ADP + 0.79*NADH</td>
</tr>
<tr>
<td>2*PYR + GLU =&gt; VAL + CO₂ + AKG</td>
<td>7<em>AcCoA + 12</em>NADPH =&gt; C14 + 7<em>CoA + 12</em>NADP*</td>
<td>8Glu+4GAP+4PYR=&gt;Chl a</td>
</tr>
<tr>
<td>Chemical Reaction</td>
<td>Balanced Reaction</td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>--------------------</td>
<td></td>
</tr>
<tr>
<td>$2^\text{*PYR} + \text{AcCoA} + \text{GLU} \Rightarrow 2^\text{*CO}_2 + \text{AKG} + \text{LEU} + \text{CoA}$</td>
<td>$8^\text{*AcCoA} + 14^\text{*NADPH} \Rightarrow C16 + 8^\text{*CoA} + 14^\text{<em>NADP}^</em>$</td>
<td></td>
</tr>
<tr>
<td>$\text{THR} + \text{NH}_3 + \text{PYR} + \text{GLU} + 6^\text{*NADPH} \Rightarrow \text{ILE} + \text{AKG} + \text{CO}_2$</td>
<td>$8^\text{*AcCoA} + 14^\text{*NADPH} \Rightarrow C16 + 8^\text{*CoA} + 14^\text{<em>NADP}^</em>$</td>
<td></td>
</tr>
<tr>
<td>$\text{ASP} + \text{SUCCoA} + \text{CYS} + 5^\text{*MeTHF} + \text{ATP} + 2^\text{*NADPH} \Rightarrow \text{MET} + \text{SUC} + \text{CO}_2 + \text{AKG}$</td>
<td>$8^\text{*AcCoA} + 15^\text{*NADPH} \Rightarrow C17 + 8.5^\text{*CoA} + 15^\text{<em>NADP}^</em>$</td>
<td></td>
</tr>
<tr>
<td>$\text{CHM} + \text{GLU} \Rightarrow \text{PHE} + \text{AKG} + \text{CO}_2$</td>
<td>$8.5^\text{*AcCoA} + 15^\text{*NADPH} \Rightarrow C17 + 8.5^\text{*CoA} + 15^\text{<em>NADP}^</em>$</td>
<td></td>
</tr>
<tr>
<td>$\text{CHM} + \text{GLU} + \text{NAD}^* \Rightarrow \text{TYR} + \text{AKG} + \text{CO}_2 + \text{NADH}$</td>
<td>$9^\text{*AcCoA} + 16^\text{*NADPH} \Rightarrow C18 + 9^\text{*CoA} + 16^\text{<em>NADP}^</em>$</td>
<td></td>
</tr>
<tr>
<td>$\text{R5P} + \text{GLN} + 2^\text{*ATP} + 2^\text{<em>NAD}^</em> \Rightarrow \text{HIS} + \text{GLU} + 2^\text{*PP} + \text{AICAR} + 2^\text{*NADH} + 2^\text{*ADP}$</td>
<td>$9^\text{*AcCoA} + 16^\text{*NADPH} \Rightarrow C18 + 9^\text{*CoA} + 16^\text{<em>NADP}^</em>$</td>
<td></td>
</tr>
<tr>
<td>$\text{ASP} + \text{SUCCoA} + \text{GLU} + \text{PYR} + 2^\text{*NADPH} + \text{ATP} \Rightarrow \text{LYS} + \text{AKG} + \text{CoA} + \text{CO}_2 + \text{SUC}$</td>
<td>$9^\text{*AcCoA} + 16^\text{*NADPH} \Rightarrow C18 + 9^\text{*CoA} + 16^\text{<em>NADP}^</em>$</td>
<td></td>
</tr>
<tr>
<td>$2^\text{*GLU} + \text{AcCoA} + \text{CP} + \text{ASP} + 2^\text{*ATP} + \text{NADPH} \Rightarrow \text{ARG} + \text{AKG} + \text{FUM} + \text{Ac} + \text{CoA} + 2^\text{<em>ADP} + \text{NADP}^</em>$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{SER} \Rightarrow \text{GLY}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{AKG} + 2^\text{<em>NH}_3 + \text{NADPH} + \text{ATP} \Rightarrow \text{GLN} + \text{NADP}^</em> + \text{ADP}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{OAA} + \text{GLU} \Rightarrow \text{ASP} + \text{AKG}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{NH}_3 + \text{ASP} + \text{ATP} \Rightarrow \text{ASN} + \text{ADP}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{SER} + \text{FTHF} + \text{GLY} \Rightarrow \text{MeTHFA}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table S3. Net fluxes and dilution parameters determined by $^{13}$C INST-MFA for the *Synechococcus 2973* photobioreactor model. Relative mean values are net fluxes relative to a net CO$_2$ uptake rate of 100 (actual uptake rate was estimated as 12.2 mmol/g-DW/h and those mean values are reported as well). Mean parameter estimates and 95% confidence bounds using INCA’s parameter continuation method are shown below. Dilution parameters represent the fraction of active pool in the cell. Although INCA platform does not need pool size measurement, including the pool size data can improve the model confidence interval. For example, * indicates a net flux where the 95% confidence intervals were not resolved. Constraining several pool size measurements with estimated values (MAL, PEP, AcCoA, SUC, CIT) resulted in the best fit model and resolved the confidence interval issue and are noted by *. In addition, average flux values and their flux ranges were also calculated via Monte Carlo method as stated above (randomly perturbed sampling time within measurement errors, n=50). The Monte Carlo based flux estimation gives the possible flux ranges due to imprecise timing during sample harvesting processes. In the Figure 2 of the main text, we reported the flux results without constraining the pool sizes.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Relative mean values</th>
<th>Confidence Intervals 95%</th>
<th>Monte Carlo based flux estimation</th>
<th>Absolute mean values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pool size</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>not constrained</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G6P &lt;-&gt; G1P</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0, 0.0</td>
<td>0.0 ± 0.0</td>
</tr>
<tr>
<td>G1P &lt;-&gt; GLYC</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0, 0.0</td>
<td>0.0 ± 0.0</td>
</tr>
<tr>
<td>G6P &lt;-&gt; F6P</td>
<td>-0.1</td>
<td>-0.1</td>
<td>-0.9, -0.1</td>
<td>-0.1 ± 0.06</td>
</tr>
<tr>
<td>G6P -&gt; 6PG</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0, 0.09</td>
<td>0.0 ± 0.06</td>
</tr>
<tr>
<td>6PG -&gt; RU5P + CO2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0, 0.09</td>
<td>0.0 ± 0.06</td>
</tr>
<tr>
<td>Reaction</td>
<td>Value 1</td>
<td>Value 2</td>
<td>Value 3</td>
<td>Value 4</td>
</tr>
<tr>
<td>---------------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
<td>-----------</td>
</tr>
<tr>
<td>F6P &lt;-&gt; FBP</td>
<td>-39.9</td>
<td>-39.4</td>
<td>-42.6, -34.5</td>
<td>-41.0 ± 0.8</td>
</tr>
<tr>
<td>FBP &lt;-&gt; DHAP + GAP</td>
<td>-39.9</td>
<td>-39.4</td>
<td>-42.6, -34.5</td>
<td>-41.0 ± 0.8</td>
</tr>
<tr>
<td>DHAP &lt;-&gt; GAP</td>
<td>-78.6</td>
<td>-77.7</td>
<td>-80.3, -75.3</td>
<td>-79.6 ± 2.9</td>
</tr>
<tr>
<td>GAP &lt;-&gt; 3PGA</td>
<td>-197.4</td>
<td>-195.2</td>
<td>*-197.3, -188.8</td>
<td>-199.9 ± 7.2</td>
</tr>
<tr>
<td>3PGA &lt;-&gt; 2PGA</td>
<td>25.0</td>
<td>25.0</td>
<td>23.7, 26.6</td>
<td>25.0 ± 2.3</td>
</tr>
<tr>
<td>2PGA &lt;-&gt; PEP</td>
<td>29.3</td>
<td>28.8</td>
<td>28.9, 29.6</td>
<td>29.7 ± 1.3</td>
</tr>
<tr>
<td>PEP &lt;-&gt; PYR</td>
<td>22.0</td>
<td>19.7</td>
<td>21.0, 22.2</td>
<td>20.9 ± 0.8</td>
</tr>
<tr>
<td>R5P &lt;-&gt; X5P</td>
<td>-78.5</td>
<td>-77.6</td>
<td>-81.1, -76.4</td>
<td>-79.5 ± 2.9</td>
</tr>
<tr>
<td>R5P &lt;-&gt; R5P</td>
<td>-37.5</td>
<td>-37.1</td>
<td>-38.8, -36.5</td>
<td>-37.8 ± 1.6</td>
</tr>
<tr>
<td>R5P &lt;-&gt; RUBP</td>
<td>116.0</td>
<td>114.7</td>
<td>112, 119.6</td>
<td>117.3 ± 4.4</td>
</tr>
<tr>
<td>RUBP + CO2 -&gt; 3PGA + 3PGA</td>
<td>107.9</td>
<td>107.1</td>
<td>*105.5, 109.2</td>
<td>110.3 ± 2.9</td>
</tr>
<tr>
<td>X5P &lt;-&gt; GAP + EC2</td>
<td>-78.5</td>
<td>-77.6</td>
<td>-81.1, -76.4</td>
<td>-79.5 ± 2.9</td>
</tr>
<tr>
<td>F6P &lt;-&gt; E4P + EC2</td>
<td>39.8</td>
<td>39.4</td>
<td>38.7, 41.1</td>
<td>40.3 ± 1.4</td>
</tr>
<tr>
<td>S7P &lt;-&gt; R5P + EC2</td>
<td>38.7</td>
<td>38.3</td>
<td>37.7, 40.0</td>
<td>39.2 ± 1.43</td>
</tr>
<tr>
<td>F6P &lt;-&gt; GAP + EC3</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0, 2.7</td>
<td>0.0 ± 0.8</td>
</tr>
<tr>
<td>S7P &lt;-&gt; E4P + EC3</td>
<td>0.0</td>
<td>0.0</td>
<td>-2.7, 0.0</td>
<td>0.0 ± 0.8</td>
</tr>
<tr>
<td>DHAP + E4P -&gt; SBP</td>
<td>38.7</td>
<td>38.3</td>
<td>*35.3, 43.6</td>
<td>39.5 ± 1.2</td>
</tr>
<tr>
<td>SBP -&gt; S7P</td>
<td>38.7</td>
<td>38.3</td>
<td>*35.3, 43.6</td>
<td>39.5 ± 1.2</td>
</tr>
<tr>
<td>PYR -&gt; AcCoA + CO2</td>
<td>12.1</td>
<td>11.7</td>
<td>11.7, 12.7</td>
<td>12.2 ± 0.8</td>
</tr>
<tr>
<td>OAA + AcCoA -&gt; CIT</td>
<td>2.7</td>
<td>2.3</td>
<td>2.3, 3.1</td>
<td>2.6 ± 0.5</td>
</tr>
<tr>
<td>CIT &lt;-&gt; ICI</td>
<td>2.7</td>
<td>2.3</td>
<td>2.3, 3.1</td>
<td>2.6 ± 0.5</td>
</tr>
<tr>
<td>ICIT &lt;-&gt; AKG + CO2</td>
<td>2.3</td>
<td>2.3</td>
<td>2.3, 2.8</td>
<td>2.3 ± 0.1</td>
</tr>
<tr>
<td>AKG -&gt; SUC + CO2</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0, 0.6</td>
<td>0.0 ± 0.0</td>
</tr>
<tr>
<td>SUC &lt;-&gt; FUM</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0, 0.8</td>
<td>0.3 ± 0.5</td>
</tr>
<tr>
<td>FUM &lt;-&gt; MAL</td>
<td>2.0</td>
<td>1.6</td>
<td>1.6, 2.3</td>
<td>1.9 ± 0.5</td>
</tr>
<tr>
<td>MAL &lt;-&gt; OAA</td>
<td>2.0</td>
<td>-0.3</td>
<td>1.3, 2.4</td>
<td>0.5 ± 0.9</td>
</tr>
<tr>
<td>Reaction</td>
<td>Flux (μmol·g DW⁻¹·h⁻¹)</td>
<td>SD (μmol·g DW⁻¹·h⁻¹)</td>
<td><em>i</em> (μmol·g DW⁻¹·h⁻¹)</td>
<td>Error (μmol·g DW⁻¹·h⁻¹)</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------</td>
<td>----------------------</td>
<td>------------------------</td>
<td>--------------------------</td>
</tr>
<tr>
<td>MAL → PYR + CO2</td>
<td>0.0</td>
<td>1.9</td>
<td>*0.0, 5.0</td>
<td>1.4 ± 1.2</td>
</tr>
<tr>
<td>PEP + CO2 → OAA</td>
<td>5.1</td>
<td>7.0</td>
<td>*5.1, 9.9</td>
<td>6.7 ± 1.3</td>
</tr>
<tr>
<td>ICIT → GLX + SUC</td>
<td>0.4</td>
<td>0.0</td>
<td>0.0, 0.9</td>
<td>0.3 ± 0.5</td>
</tr>
<tr>
<td>GLX + AcCoA → MAL</td>
<td>0.0</td>
<td>0.0</td>
<td>*0.0, 0.3</td>
<td>0.0 ± 0.0</td>
</tr>
<tr>
<td>RUBP → 3PGA + 2PG</td>
<td>8.1</td>
<td>7.6</td>
<td>6.8, 10.6</td>
<td>9.1 ± 2.6</td>
</tr>
<tr>
<td>2PG → GLC</td>
<td>8.1</td>
<td>7.6</td>
<td>6.8, 10.6</td>
<td>9.1 ± 2.6</td>
</tr>
<tr>
<td>GLC → GLX</td>
<td>8.1</td>
<td>7.6</td>
<td>6.8, 10.6</td>
<td>9.1 ± 2.6</td>
</tr>
<tr>
<td>GLX + GLX → GA + CO2</td>
<td>4.2</td>
<td>3.8</td>
<td>3.1, 5.6</td>
<td>4.7 ± 1.1</td>
</tr>
<tr>
<td>GA ↔ 2PGA</td>
<td>4.2</td>
<td>3.8</td>
<td>3.1, 5.6</td>
<td>4.7 ± 1.1</td>
</tr>
</tbody>
</table>

\[
0.404*R5P + 3.132*AcCoA + 0.023*G6P + 0.358*E4P + 0.519*3PGA + 0.715*PEP + 3.234*PYR + 1.477*OAA + 0.760*AKG + 0.144*GAP -> Biomass + 0.522*FUM + 1.850*CO2
\]

Table S4. Net fluxes and dilution parameters determined by \(^{13}\)C INST-MFA for the *Synechococcus 2973* shaking flasks. Values are relative to a net CO₂ uptake rate of 100 (actual
uptake rate was estimated as 6.7 mmol/g-DW/h). Median parameter estimates and 95% confidence bounds using INCA’s parameter continuation method are shown. Due to suboptimal cultivation conditions, we did not perform further confidence interval analysis using fixed metabolite pool size or Monte Carlo methods.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Mean value</th>
<th>LB 95%</th>
<th>UB 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>G6P &lt;-&gt; G1P</td>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>G1P &lt;-&gt; GLYC</td>
<td>0.0</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>G6P &lt;-&gt; F6P</td>
<td>-0.2</td>
<td>-1.25</td>
<td>0.23</td>
</tr>
<tr>
<td>G6P -&gt; 6PG</td>
<td>0.0</td>
<td>0.00</td>
<td>1.13</td>
</tr>
<tr>
<td>6PG -&gt; RU5P + CO2</td>
<td>0.0</td>
<td>0</td>
<td>1.02</td>
</tr>
<tr>
<td>F6P &lt;-&gt; FBP</td>
<td>-37.8</td>
<td>NaN</td>
<td>-35.65</td>
</tr>
<tr>
<td>FBP &lt;-&gt; DHAP + GAP</td>
<td>-37.8</td>
<td>NaN</td>
<td>-35.65</td>
</tr>
<tr>
<td>DHAP &lt;-&gt; GAP</td>
<td>-74.3</td>
<td>-75.34</td>
<td>-73.57</td>
</tr>
<tr>
<td>GAP &lt;-&gt; 3PGA</td>
<td>-186.7</td>
<td>-189.93</td>
<td>-184.76</td>
</tr>
<tr>
<td>3PGA &lt;-&gt; 2PGA</td>
<td>30.1</td>
<td>29.33</td>
<td>NaN</td>
</tr>
<tr>
<td>2PGA &lt;-&gt; PEP</td>
<td>30.1</td>
<td>30.13</td>
<td>NaN</td>
</tr>
<tr>
<td>PEP &lt;-&gt; PYR</td>
<td>23.4</td>
<td>22.20</td>
<td>24.68</td>
</tr>
<tr>
<td>RU5P &lt;-&gt; X5P</td>
<td>-74.1</td>
<td>-75.59</td>
<td>-73.34</td>
</tr>
<tr>
<td>RU5P &lt;-&gt; R5P</td>
<td>-35.4</td>
<td>-36.19</td>
<td>-35.06</td>
</tr>
<tr>
<td>RU5P -&gt; RUBP</td>
<td>109.5</td>
<td>107.99</td>
<td>11.52</td>
</tr>
<tr>
<td>RUBP + CO2 -&gt; 3PGA + 3PGA</td>
<td>108.7</td>
<td>107.96</td>
<td>NaN</td>
</tr>
<tr>
<td>X5P &lt;-&gt; GAP + EC2</td>
<td>-74.1</td>
<td>-75.59</td>
<td>-73.34</td>
</tr>
<tr>
<td>F6P &lt;-&gt; E4P + EC2</td>
<td>37.6</td>
<td>37.20</td>
<td>38.32</td>
</tr>
<tr>
<td>S7P &lt;-&gt; R5P + EC2</td>
<td>36.7</td>
<td>36.14</td>
<td>37.27</td>
</tr>
<tr>
<td>F6P &lt;-&gt; GAP + EC3</td>
<td>0.0</td>
<td>-1.06</td>
<td>27.62</td>
</tr>
<tr>
<td>S7P &lt;-&gt; E4P + EC3</td>
<td>0.0</td>
<td>-27.62</td>
<td>1.06</td>
</tr>
<tr>
<td>DHAP + E4P &lt;-&gt; SBP</td>
<td>36.5</td>
<td>11.48</td>
<td>37.55</td>
</tr>
<tr>
<td>Reaction</td>
<td>Rate 1</td>
<td>Rate 2</td>
<td>Rate 3</td>
</tr>
<tr>
<td>--------------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>SBP -&gt; S7P</td>
<td>36.5</td>
<td>11.48</td>
<td>37.55</td>
</tr>
<tr>
<td>PYR -&gt; AcCoA + CO2</td>
<td>15.3</td>
<td>14.93</td>
<td>NaN</td>
</tr>
<tr>
<td>OAA + AcCoA -&gt; CIT</td>
<td>2.2</td>
<td>2.20</td>
<td>NaN</td>
</tr>
<tr>
<td>CIT &lt;-&gt; ICIT</td>
<td>2.2</td>
<td>2.20</td>
<td>NaN</td>
</tr>
<tr>
<td>ICIT &lt;-&gt; AKG + CO2</td>
<td>2.2</td>
<td>2.20</td>
<td>NaN</td>
</tr>
<tr>
<td>AKG -&gt; SUC + CO2</td>
<td>0.0</td>
<td>0.00</td>
<td>NaN</td>
</tr>
<tr>
<td>SUC &lt;-&gt; FUM</td>
<td>0.0</td>
<td>0.00</td>
<td>NaN</td>
</tr>
<tr>
<td>FUM &lt;-&gt; MAL</td>
<td>1.4</td>
<td>1.37</td>
<td>NaN</td>
</tr>
<tr>
<td>MAL &lt;-&gt; OAA</td>
<td>1.0</td>
<td>-0.23</td>
<td>2.18</td>
</tr>
<tr>
<td>MAL -&gt; PYR + CO2</td>
<td>1.2</td>
<td>0.0</td>
<td>2.84</td>
</tr>
<tr>
<td>PEP + CO2-&gt; OAA</td>
<td>4.6</td>
<td>3.40</td>
<td>6.95</td>
</tr>
<tr>
<td>ICI -&gt; GLX + Suc</td>
<td>0.0</td>
<td>0.00</td>
<td>0.07</td>
</tr>
<tr>
<td>GLX + AcCoA -&gt; MAL</td>
<td>0.8</td>
<td>0.0</td>
<td>1.36</td>
</tr>
<tr>
<td>RUBP -&gt; 3PGA + 2PG</td>
<td>0.8</td>
<td>NaN</td>
<td>1.43</td>
</tr>
<tr>
<td>2PG -&gt; GLC</td>
<td>0.8</td>
<td>NaN</td>
<td>1.43</td>
</tr>
<tr>
<td>GLC -&gt; GLX</td>
<td>0.8</td>
<td>NaN</td>
<td>1.43</td>
</tr>
<tr>
<td>GLX + GLX -&gt; GA + CO2</td>
<td>0.0</td>
<td>0.00</td>
<td>0.93</td>
</tr>
<tr>
<td>GA &lt;-&gt; 2PGA</td>
<td>0.0</td>
<td>0.00</td>
<td>0.93</td>
</tr>
<tr>
<td>0.426<em>R5P + 4.846</em>AcCoA +</td>
<td>2.5</td>
<td>2.53</td>
<td>2.53</td>
</tr>
<tr>
<td>0.090<em>G6P + 0.417</em>E4P +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.580<em>3PGA + 0.834</em>PEP +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.701<em>PYR + 1.313</em>OAA +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.868<em>AKG + 0.175</em>GAP -&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Biomass + 0.536*FUM +</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.142*CO2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Dilution Parameters**

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Rate 1</th>
<th>Rate 2</th>
<th>Rate 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>R5P</td>
<td>0.86</td>
<td>0.77</td>
<td>1.00</td>
</tr>
<tr>
<td>F6P</td>
<td>0.55</td>
<td>0.46</td>
<td>0.65</td>
</tr>
<tr>
<td>3PGA</td>
<td>0.63</td>
<td>0.54</td>
<td>NaN</td>
</tr>
</tbody>
</table>
Table S5. Net fluxes and dilution parameters determined by $^{13}$C INST-MFA for the *Synechococcus 7942* photobioreactor model. Values are relative to a net CO$_2$ uptake rate of 100 (actual uptake rate was estimated as 5.1 mmol/g-DW/h and those absolute mean values are reported as well). Median parameter estimates and 95% confidence bounds using INCA’s parameter continuation method are shown.

<table>
<thead>
<tr>
<th>Reaction</th>
<th>Relative mean value</th>
<th>Confidence Intervals 95%</th>
<th>Absolute mean values</th>
</tr>
</thead>
<tbody>
<tr>
<td>PS not constrained</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G6P $\leftrightarrow$ G1P</td>
<td>0.0</td>
<td>0.0, 0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>G1P $\leftrightarrow$ GLYC</td>
<td>0.0</td>
<td>0.0, 0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>G6P $\leftrightarrow$ F6P</td>
<td>-0.6</td>
<td>-9.9, -0.62</td>
<td>-0.04</td>
</tr>
<tr>
<td>G6P $\rightarrow$ 6PG</td>
<td>0.0</td>
<td>0.0, 9.3</td>
<td>0.005</td>
</tr>
<tr>
<td>6PG $\rightarrow$ R5P + CO2</td>
<td>0.0</td>
<td>0.0, 9.3</td>
<td>0.005</td>
</tr>
<tr>
<td>F6P $\leftrightarrow$ FBP</td>
<td>-16.3</td>
<td>-24.4, -16.2</td>
<td>-2.0</td>
</tr>
<tr>
<td>FBP $\leftrightarrow$ DHAP + GAP</td>
<td>-16.3</td>
<td>-24.4, -16.2</td>
<td>-2.0</td>
</tr>
<tr>
<td>DHAP $\leftrightarrow$ GAP</td>
<td>-77.3</td>
<td>NaN, -77.2</td>
<td>-4.0</td>
</tr>
<tr>
<td>GAP $\leftrightarrow$ 3PGA</td>
<td>-194.2</td>
<td>NaN, -194.1</td>
<td>-10.1</td>
</tr>
<tr>
<td>3PGA $\leftrightarrow$ 2PGA</td>
<td>30.5</td>
<td>27.8, 32.3</td>
<td>1.6</td>
</tr>
<tr>
<td>2PGA $\leftrightarrow$ PEP</td>
<td>30.5</td>
<td>30.5, 32.3</td>
<td>1.6</td>
</tr>
<tr>
<td>PEP $\leftrightarrow$ PYR</td>
<td>22.2</td>
<td>20.4, NaN</td>
<td>1.2</td>
</tr>
<tr>
<td>RU5P $\leftrightarrow$ X5P</td>
<td>-76.7</td>
<td>NaN, -76.7</td>
<td>-4.0</td>
</tr>
<tr>
<td>RU5P $\leftrightarrow$ R5P</td>
<td>-36.4</td>
<td>NaN, -36.4</td>
<td>-1.9</td>
</tr>
<tr>
<td>Reaction</td>
<td>Rate 1</td>
<td>Rate 2</td>
<td>Rate 3</td>
</tr>
<tr>
<td>---------------------------------------</td>
<td>--------</td>
<td>----------</td>
<td>--------</td>
</tr>
<tr>
<td>RU5P -&gt; RUBP</td>
<td>113.1</td>
<td>113.0, 123.5</td>
<td>5.9</td>
</tr>
<tr>
<td>RUBP + CO2 -&gt; 3PGA + 3PGA</td>
<td>113.1</td>
<td>113.0, 123.5</td>
<td>5.9</td>
</tr>
<tr>
<td>X5P &lt;-&gt; GAP + EC2</td>
<td>-76.7</td>
<td>NaN, -76.7</td>
<td>-4.0</td>
</tr>
<tr>
<td>F6P &lt;-&gt; E4P + EC2</td>
<td>38.8</td>
<td>38.8, NaN</td>
<td>2.0</td>
</tr>
<tr>
<td>S7P &lt;-&gt; R5P + EC2</td>
<td>37.9</td>
<td>37.9, NaN</td>
<td>2.0</td>
</tr>
<tr>
<td>F6P &lt;-&gt; GAP + EC3</td>
<td>-23.1</td>
<td>-38.9, NaN</td>
<td>-0.1</td>
</tr>
<tr>
<td>S7P &lt;-&gt; E4P + EC3</td>
<td>23.1</td>
<td>NaN, 38.9</td>
<td>0.1</td>
</tr>
<tr>
<td>DHAP + E4P -&gt; SBP</td>
<td>61.0</td>
<td>15.5, 76.9</td>
<td>2.0</td>
</tr>
<tr>
<td>SBP -&gt; S7P</td>
<td>61.0</td>
<td>15.5, 76.9</td>
<td>2.0</td>
</tr>
<tr>
<td>PYR -&gt; AcCoA + CO2</td>
<td>15.1</td>
<td>15.1, NaN</td>
<td>0.8</td>
</tr>
<tr>
<td>OAA + AcCoA -&gt; CIT</td>
<td>2.7</td>
<td>2.6, 4.5</td>
<td>0.2</td>
</tr>
<tr>
<td>CIT &lt;-&gt; ICI</td>
<td>2.7</td>
<td>2.6, 4.5</td>
<td>0.2</td>
</tr>
<tr>
<td>ICIT &lt;-&gt; AKG + CO2</td>
<td>2.7</td>
<td>2.6, 4.5</td>
<td>0.2</td>
</tr>
<tr>
<td>AKG -&gt; SUC + CO2</td>
<td>0.0</td>
<td>0.0, 1.9</td>
<td>0.03</td>
</tr>
<tr>
<td>SUC &lt;-&gt; FUM</td>
<td>0.0</td>
<td>0.0, 1.9</td>
<td>0.03</td>
</tr>
<tr>
<td>FUM &lt;-&gt; MAL</td>
<td>1.7</td>
<td>1.7, 3.6</td>
<td>0.1</td>
</tr>
<tr>
<td>MAL &lt;-&gt; OAA</td>
<td>0.0</td>
<td>-1.8, 1.2</td>
<td>0.1</td>
</tr>
<tr>
<td>MAL -&gt; PYR + CO2</td>
<td>1.7</td>
<td>0.0, 3.7</td>
<td>0.02</td>
</tr>
<tr>
<td>PEP + CO2 -&gt; OAA</td>
<td>6.4</td>
<td>5.0, 8.4</td>
<td>0.3</td>
</tr>
<tr>
<td>ICIT -&gt; GLX + SUC</td>
<td>0.0</td>
<td>0.0, 0.002</td>
<td>0.0</td>
</tr>
<tr>
<td>GLX + AcCoA -&gt; MAL</td>
<td>0.0</td>
<td>0.0, 0.04</td>
<td>0.0</td>
</tr>
<tr>
<td>RUBP -&gt; 3PGA + 2PG</td>
<td>0.0</td>
<td>0.0, 0.3</td>
<td>0.002</td>
</tr>
<tr>
<td>2PG -&gt; GLC</td>
<td>0.0</td>
<td>0.0, 0.3</td>
<td>0.002</td>
</tr>
<tr>
<td>GLC -&gt; GLX</td>
<td>0.0</td>
<td>0.0, 0.3</td>
<td>0.002</td>
</tr>
<tr>
<td>GLX + GLX -&gt; GA + CO2</td>
<td>0.0</td>
<td>0.0, 5.0</td>
<td>0.001</td>
</tr>
<tr>
<td>GA ↔ 2PGA</td>
<td>0.0</td>
<td>0.0, 5.0</td>
<td>0.001</td>
</tr>
<tr>
<td>-----------</td>
<td>-----</td>
<td>---------</td>
<td>-------</td>
</tr>
<tr>
<td>0.465<em>R5P + 3.789</em>AcCoA + 0.188<em>G6P + 0.275</em>E4P + 0.461<em>3PGA + 0.550</em>PEP + 2.698<em>PYR + 1.154</em>OAA + 0.805<em>AKG + 0.218</em>GAP → Biomass + 0.528<em>FUM + 1.369</em>CO2</td>
<td>3.3</td>
<td>3.3, 3.3</td>
<td>0.17</td>
</tr>
</tbody>
</table>

**Dilution Parameters**

<table>
<thead>
<tr>
<th>Comp</th>
<th>Value</th>
<th>Range</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>PEP</td>
<td>0.77</td>
<td>0.71, 0.82</td>
<td>0.77</td>
</tr>
<tr>
<td>RuBP</td>
<td>0.64</td>
<td>0.57, 0.70</td>
<td>0.49</td>
</tr>
<tr>
<td>3PGA</td>
<td>0.74</td>
<td>0.70, 0.79</td>
<td>0.72</td>
</tr>
<tr>
<td>S7P</td>
<td>0.57</td>
<td>0.51, 0.63</td>
<td>0.47</td>
</tr>
<tr>
<td>SBP</td>
<td>0.54</td>
<td>0.46, 0.58</td>
<td>0.48</td>
</tr>
<tr>
<td>FBP</td>
<td>0.51</td>
<td>0.44, 0.54</td>
<td>0.49</td>
</tr>
</tbody>
</table>
Figure S1 Comparison of CO₂ uptake rate measured by gas chromatography headspace/chlorophyll a content and estimated from the biomass composition. Total CO₂ was calculated from previously described methods as the sum of the dissolved and gaseous CO₂ in a culture (Xiong et al., 2015). Standard error bars were based on measured standard deviations from biomass composition of 5 biological replicates; *represents a p-value of <0.02 using two-tailed equal variance Student’s t-Test.
Figure S2. *Synechococcus* 2973 metabolites analyzed by LC-MS. A) FBP- the sample was extracted by methanol quenching method. 1 e4 represents its intensity (cps). B) FBP- the sample was extracted by liquid nitrogen quenching. 4.5 e4 represents its intensity (cps). C) PEP- metabolite pool size estimation by extracting a mixture of labeled *E. coli* and non-labeled *Synechococcus* 2973 cultures. (D-F) are chromatograms of 6PG (6PG standard, 6PG from *Synechococcus* 2973, and 6PG from *Synechococcus* 2973 Δzwf).
Figure S3. A. The reported sum of square of residuals (SSR) for isotopomer fitting of individual metabolite from INCA modeling simulations (without adding the pool sizes as constraints). The PBR had a smaller SSR (713), while the SF model had much higher SSR (981).
Figure S4. Dynamic labeling patterns from INCA for all metabolites measured using IP-LC-MS/MS or HILIC-MS for *Synechococcus 2973* under PBR conditions. Experimentally measured MIDs with error bars representing standard measurement errors from biological duplicates. PGA (185), 6PG (275), PEP (167), G6P (259), R5P (229), S7P (289), SBP (269), SUC (118), MAL (134), F6P (259), CIT (192), FBP (339), G1P (259), 2PG (155), are all the metabolites listed with their nominal masses of M0 mass isotopomer shown in parentheses.
Figure S5. The average percent $^{13}$C-enrichment of Citrate, Malate, Succinate, and free Glutamate in the photobioreactor conditions for *Synechococcus* 2973 and *Synechococcus* 7942 as a function of time. Average $^{13}$C-Enrichment is calculated from experimentally measured MID$e$s where $N$ is the number of carbons, $M_i$ is the MID of the $i$th isotopomer, average $^{13}$C-Enrichment is $\frac{1}{N} \cdot \sum_{i=1}^{N} M_i \cdot i$, and error bars represent average standard deviations from biological duplicates.
Figure S6. Average $^{13}$C-enrichment of metabolite intermediates in the CBB cycle and PPP, in *Synechococcus* 2973 and *Synechococcus* 7942. Average $^{13}$C-Enrichment is calculated from experimentally measured MIDs where $N$ is the number of carbons, $M_i$ is the MID of the $i$th isotopomer, average $^{13}$C-Enrichment is $\frac{1}{N} \cdot \sum_{i=1}^{N} M_i \cdot i$, and error bars represent average standard deviations from biological duplicates.

Figure S7. Net fluxes fold changes of *Synechococcus* 2973 over *Synechococcus* 7942 net fluxes. Fold change was calculated as the normalized net flux of 2973 over the normalized net flux of 7942 minus 1. Metabolic pathways and enzymes in the dark blue follow the y-axis scale on the left, while metabolic pathways and enzymes in light blue follow the y-axis scale on the right.
Supporting Photo 1. Multi-cultivator system used to cultivate fast growing cyanobacteria, *Synechococcus 2973*. The system is from Photon Systems Instruments (MC 1000) and uses cool white LEDs with adjustable irradiance.
10. Appendix 2: Deciphering cyanobacterial phenotypes for fast photoautotrophic growth via isotopically nonstationary metabolic flux analysis

Doi: 10.1186/s13068-017-0958-y
Deciphering cyanobacterial phenotypes for fast photoautotrophic growth via isotopically nonstationary metabolic flux analysis

Mary H. Abernathy, Jingjie Yu, Fangfang Ma, Michelle Liberton, Justin Ungerer, Whitney D. Hollinshead, Saratram Gopalakrishnan, Lian He, Costas D. Maranas, Himadri B. Pakrasi, Doug K. Allen and Yinfie J. Tang

Abstract

Background: Synechococcus elongatus UTEX 2973 is the fastest growing cyanobacterium characterized to date. Its genome was found to be 99.8% identical to S. elongatus 7942 yet it grows twice as fast. Current genome-to-phenome mapping is still poorly performed for non-model organisms. Even for species with identical genomes, cell phenotypes can be strikingly different. To understand Synechococcus 2973's fast-growth phenotype and its metabolic features advantageous to photo-biorefineries, [13]C isotopically nonstationary metabolic flux analysis (INST-MFA), biomass compositional analysis, gene knockouts, and metabolite profiling were performed on both strains under various growth conditions.

Results: The Synechococcus 2973 flux maps show substantial carbon flow through the Calvin cycle, glycolysis, photorespiration and pyruvate kinase, but minimal flux through the malic enzyme and oxidative pentose phosphate pathways under high light/CO2 conditions. During fast growth, its pool sizes of key metabolites in central pathways were lower than suboptimal growth. Synechococcus 2973 demonstrated similar flux ratios to Synechococcus 7942 (under fast growth conditions), but exhibited greater carbon assimilation, higher NADPH concentrations, higher energy charge (relative ATP ratio over ADP and AMP), less accumulation of glycogen, and potentially metabolite channeling. Furthermore, Synechococcus 2973 has very limited flux through the TCA pathway with small pool sizes of acetyl-CoA/TCA intermediates under all growth conditions.

Conclusions: This study employed flux analysis to investigate phenotypic heterogeneity among two cyanobacterial strains with near-identical genome background. The flux/metabolite profiling, biomass composition analysis, and genetic modification results elucidate a highly effective metabolic topology for CO2 assimilation and biosynthesis in Synechococcus 2973. Comparisons across multiple Synechococcus strains indicate faster metabolism is also driven by proportional increases in both photosynthesis and key central pathway fluxes. Moreover, the flux distribution in Synechococcus 2973 supports the use of its strong sugar phosphate pathways for optimal bio-productions. The integrated methodologies in this study can be applied for characterizing non-model microbial metabolism.

Keywords: [13]C labeling experiments, Channeling, Glycogen, Metabolites, Photobioreactor, Energy charge
Background
Efforts toward a sustainable bio-based economy have focused on phototrophic hosts for the production of chemicals from CO₂ and light. Cyanobacteria are of special biotechnological interest due to their metabolic flexibility, unicellular anatomy, and high photosynthetic efficiency. Additionally, cyanobacteria can use flue gas from power plants to mitigate CO₂ emissions that contribute to climate change [1]. To achieve metrics required for commercialization, cyanobacterial photo-biorefineries must have comparable biosynthesis capability to commonly used heterotrophic organisms. Recently, Synechococcus elongatus UTEX 2973 was isolated, whose growth rate reaches a doubling time of 2 h under high light and high CO₂ conditions [2]. In comparison, under optimal growth conditions, S. elongatus PCC 7942 exhibits a doubling time of ~ 5 h although its genome sequence is 99.8% identical to Synechococcus 2973 (55 single nucleotide polymorphisms and a large 188 kb inversion between Synechococcus 2973 and 7942) [2]. To understand how cyanobacteria achieve maximal growth rates, this study describes Synechococcus 2973 flux topology under both optimal and suboptimal growth conditions.

Metabolic flux analysis (MFA) can provide a quantitative description of the metabolic network, link genome profiling to phenomenology analysis, and reveal pathway regulations through comparative studies. Currently, the cyanobacterial strain, Synechocystis sp. PCC 6803 (doubling time ~ 8 h), is considered to be the model cyanobacterium whose metabolism has been extensively profiled by flux analysis tools [3–6]. Synechocystis 6803 has significant flux through malic enzyme and oxidative pentose phosphate pathways (OPPP) under the photoautotrophic and photomixotrophic conditions. It also operates a cyclic TCA cycle via the γ-aminobutyric acid (GABA) shunt, which forms succinate through the intermediates glutamate and succinate semialdehyde, despite the missing enzyme from 2-oxoglutarate to succinic-CoA [7]. To profile Synechocystis 6803 photoautotrophic metabolism, 13C-bicarbonate pulse experiments and isotopically nonstationary metabolic flux analysis (INST-MFA) were developed [3]. Using the software package, INCA, mass isotopomer data from dynamic labeling experiments can be used to quantify fluxes without the need to precisely determine metabolite pool sizes (which are fitted as parameters to account for transient labeling data) [8], and is therefore more convenient than other flux profiling methods [9]. In the current study, INST-MFA, gene knockouts, and metabolite analysis were performed to obtain insights into the physiology and metabolic regulations of Synechococcus 2973 under different bioreactor conditions. Meanwhile, aspects of biomass composition were measured to reveal changes in macromolecule tradeoffs that correlate to cell growth and bioreactor conditions [10]. The outcome highlights the advantages and hurdles of establishing Synechococcus 2973 as a new platform organism for bioproduction. The comparative studies among metabolisms of different strains may also offer new insights into flux dependency on adaptive evolution.

Results
Synechococcus 2973 growth and biomass compositions
In optimal photobioreactor (PBR, 500 μmol photons/m² s continuous light) conditions [2], Synechococcus 2973 exhibits very rapid growth (0.33 ± 0.05 h⁻¹; Fig. 1a). For comparison, maximal growth rate of Synechococcus 7942 was 0.14 ± 0.02 h⁻¹ at 300 μmol photons/m² s continuous light due to photo-inhibition at greater light intensities. Synechococcus strains were also grown under suboptimal light conditions in shake flasks (100 μmol photons/m² s continuous light or diurnal light irradiation; Fig. 1b). Suboptimal cultivations greatly impaired cyanobacterial photosynthesis, while Synechococcus 2973 still demonstrated moderately faster growth rate than Synechococcus 7942. Based on growth and biomass composition analyses, Synechococcus 2973 assimilated ~ 12.2 mmol-C/gDCW/h in the PBR but only ~ 6.7 mmol-C/gDCW/h under the shaking flask (SF) conditions. The estimated values of carbon fixation were further confirmed by the measured net CO₂ uptake rate using gas chromatography methods (Additional file 1: Figure S1) [11]. This study additionally considers two Synechococcus 2973 mutants (Δωf and Δpgl) whose genes from the oxidative pentose phosphate pathway (OPPP) were removed. The two mutants showed unimpaired growth rates compared to the wild-type strain in our tested conditions (Fig. 1).

The composition of Synechococcus 2973 biomass was measured in both PBR and SF conditions under continuous light (Additional file 1: Part 1). The total protein was ~ 50% of cyanobacterial biomass, while protein from the PBR had more glutamate/glutamine than that from the SF (p value 0.054 using two-tailed equal variance Student’s t Test). Synechococcus 2973 contained 9 ± 1% lipids and 1.5 ± 0.5% glycogen in the PBR and 13 ± 2% lipid and 6.0 ± 1.0% glycogen in SF conditions. For comparison, Synechococcus 7942 biomass composition was quantified under optimal PBR conditions and comprised 11 ± 1% lipids, 13 ± 4% glycogen, and 41 ± 0.4% protein. Proteinogenic glutamate/glutamine content in Synechococcus 7942 was significantly lower (p < 0.05) than that in Synechococcus 2973 biomass regardless of conditions. Fatty acid compositions of lipids in both strains were predominantly C16:1 and C16:0 fatty acids (85%), followed by C18:1 and C18:0. Relevant to photosynthetic light-harvesting, less chlorophyll a was
produced in PBR conditions (4.4 ± 0.5 µg/mL/OD_{290} for 2973; 5.2 ± 0.6 µg/mL/OD_{290} for 7942) compared to SF (7.4 ± 0.3 µg/mL/OD_{290} for 2973, p < 0.01). Increased chlorophyll levels in SF cells may be a cellular adjustment to compensate for insufficient light conditions.

INST-MFA of phototrophic metabolism
INST-MFA relies on measurement of transient labeling data of central metabolites after a \(^1\)C-pulse. To quench metabolism, we used ice-cold media and liquid \(N_2\) bath to quickly freeze time-course samples. The approach avoided a traditional cold methanol quenching strategy that causes significant metabolite leakage for gram-negative bacteria (i.e., cyanobacteria) [12] and resulted in improved LC-MS peak quality (Additional file 1: Figure S2a, b) [13]. The INCA software package (isotopomer network compartmental analysis) [8] was used to calculate flux values (Fig. 2) based on labeling data and the metabolic network (Additional file 1: Tables S1, S2). The fitted model for the *Synechococcus* 2973 in the PBR was statistically acceptable [Sum of square residuals SSR = 713, with an accepted range of (597 740)], whereas the best fit of the SF model could not be completely explained by measurement errors (SSR = 981, with an accepted range of [382 471]). The fitted model for *Synechococcus* 7942 in the PBR was statistically acceptable (SSR = 616, range of [543 679]). Biomass composition information and CO\(_2\) uptake rate were used to improve the accuracy of the flux maps. Flux-partitioning across conditions and strains was compared through relative flux values that were normalized to CO\(_2\) uptake rate and are plotted in Fig. 2 (Confidence intervals and dilution factors are described in Additional file 1: Tables S3, S4, S5; Additional file 1: Figures S3, S4 show isotopomer fitting quality for individual metabolites). Additionally, labeling dynamics for citrate and malate are compared between *Synechococcus* 2973 and 7942 in Fig. 5. \(^1\)C-enrichment for key metabolites is presented in Fig. 5: Additional file 1: Figures S5, S6. The slower labeling pattern of certain metabolites may be indicative of the presence of metabolically inactive pools in bulk cytoplasm. To account for this problem, INST-MFA has employed dilution factors [3]. The isotopic dilutions in certain metabolites during non-stationary labeling suggest a non-homogeneous intracellular environment.

*Synechococcus* flux maps in PBR and SF cultures
Figure 2 provides a comparison of the flux distributions between *Synechococcus* 2973 PBR and SF cultures and *Synechococcus* 7942 PBR cultures. Although the cells have significantly different CO\(_2\) fixation rates and biosynthesis fluxes (i.e., *Synechococcus* 2973 PBR cultures contained more protein synthesis fluxes, while *Synechococcus* 2973 SF and *Synechococcus* 7942 PBR cultures had higher carbon allocation to glycogen), the flux ratios in central metabolism (the Calvin cycle and the TCA pathways) were similar after normalization to CO\(_2\) uptake rate. Regardless of the strain and growth rate, *Synechococcus*
did not exhibit flux through OPPP, and approximately 12–14% of the fixed carbon was directed to the TCA cycle similar to another recent study of *Synechococcus 7942* [14]. Here, *Synechococcus* had no flux from aKG to FUM, resulting in a linear TCA pathway that is consistent with those in leaves or other photosynthetic tissues [3, 4, 15]. The flux distribution explained the absence of a number of key genes in *Synechococcus 2973* genome,
including succinyl-CoA ligase, malate dehydrogenase, α-ketoglutarate dehydrogenase, and GABA shunt genes. Although cyanobacteria are thought to employ a cyclic TCA cycle via GABA shunt for respiratory energy production [7], a complete TCA cycle seems non-essential for the Synechococcus strains possibly due to evolutionary non-necessity or evolved energy balancing [16]. Based on genome annotation, Synechococcus synthesizes fumarate from aspartate via purine metabolism, which can then be converted into malate and succinate (dynamic labeling data shown in Additional file 1: Figure S5) providing a possible source of several TCA organic acids. Synechococcus 2973 had significant anaplerotic fluxes through PEP carboxylase for the synthesis of TCA cycle intermediates, but reduced flux through the malic enzyme (less than ~1% of total fixed CO₂ in both the PBR and SF conditions). Among tested culture conditions, there were noticeable differences in respiratory fluxes. Specifically, photosynthesis in Synechococcus 2973 was elevated in the PBR condition possibly to rebalance carbon and dissipate excessive reducing equivalents generated by high light. In contrast, Synechococcus 7942 demonstrated lower photosynthetic activity under PBR conditions.

**Isotopic metabolite profiling for relative pool size estimation**

We analyzed the change of metabolite concentrations in different cyanobacterial cultures under continuous light PBR conditions [17]. Metabolite leakage for gram negative cyanobacterial cells makes precise determination of cyanobacterial pool sizes difficult; therefore, we used isotopic ratio methods to benchmark the change of cyanobacterial metabolite pools against intracellular metabolites from a standard generated from *E. coli*. Figure 3 shows the relative ratio of metabolite pool sizes normalized by the amount of biomass. Compared to *E. coli*, cyanobacterial R5P/Ru5P and other sugar phosphates were abundant, whereas the pool sizes of TCA metabolites were significantly low (Fig. 3c). Synechococcus 2973 had generally smaller metabolite pools than *Synechocystis* 6803 (with the exception of R5P/Ru5P) and similar metabolite pools compared to *Synechococcus* 7942 (Fig. 3c) with the exception of UDP glucose (i.e., *Synechococcus* 7942 contained more UDP glucose for carbohydrate synthesis (Fig. 3b). Notably, *Synechococcus* 2973 had the highest NADPH concentration among the three cyanobacterial species (Fig. 3a). In addition, *Synechococcus* 2973 PBR cultivations had decreased central metabolite concentrations (mainly sugar phosphates) relative to SF cultures (Fig. 3d). Comparing two growth conditions of *Synechococcus* 2973 (Fig. 3d), the SF cells contained more ADP and AMP than PBR cells, which may suggest a change in energy charges (relative ATP concentration to overall adenosine phosphate concentrations) under sub-optimal light conditions.

**Effect of exogenous organic acids on Synechococcus 2973’s growth**

*Synechococcus* 2973 shows significant sugar phosphate interconversion, but the TCA cycle is incomplete with limited flux and reduced metabolite pools. We hypothesized that the addition of exogenous organic substrates may alleviate such biosynthesis bottlenecks; however, cell growth was not enhanced after supplying OAA, malate, αKG or organic acids (pyruvate and acetate) in either PBR or SF conditions (Fig. 4a), even though the cells demonstrated the ability to incorporate exogenous nutrients into proteinogenic amino acids (especially for OAA; Fig. 4b). Unlabeled malate from the culture medium could be significantly incorporated into aspartate, suggesting an uncharacterized malate dehydrogenase to synthesize oxaloacetate (the precursor of aspartate). In summary, *Synechococcus* 2973 shows limited capability to efficiently utilize organic acids and convert these nutrients into cyanobacteria biomass [18].

**Discussion**

**Flux comparisons among cyanobacterial species**

*Synechococcus* 2973 and 7942 have a number of phenotypic differences. The CO₂ uptake rate in *Synechococcus* 2973 was twofold greater than *Synechococcus* 7942. *Synechococcus* 2973 showed up to a 3.5-fold greater ¹³C-enrichment over *Synechococcus* 7942 for metabolites such as FBP and STP (Additional file 1: Figure S6) at 40 s, indicating greater Calvin cycle fluxes. From the normalized flux map, there are differences in the carbon partitioning as presented in Fig. 2 and Additional file 1: Figure S7. *Synechococcus* 2973 fluxes through fructose-bisphosphate aldolase and phosphatase were increased while the fluxes towards glycogen and the malic enzyme were reduced. Flux from F6P to G6P, which is the key step towards OPPP and sugar storage metabolism, was significantly decreased in *Synechococcus* 2973 under both conditions. Malate and citrate were labeled rapidly in *Synechococcus* 2973 due to the enhanced PEPC (phosphoenolpyruvate carboxylase) reaction. Unlike *Synechocystis* 6803 that maintains a functional GABA shunt [7], *Synechococcus elongatus* apparently does not contain the genes to operate a complete TCA cycle (as annotated by the KEGG Genome database), but based on labeling data presented here may utilize a linear pathway from OAA to make citrate and αKG. The GABA shunt contains one decarboxylating reaction and the absence of this shunt reduces CO₂ loss. A succinate dehydrogenase deletion mutant was created, and the mutant showed no growth defects under continuous light conditions, which
supports the absence of fluxes from aKGA to fumarate. Unlike Synechocystis 6803, Synechococcus 2973 pyruvate is predominately made from flux through pyruvate kinase with little flux through the malic enzyme (MAL → PYR + CO₂). Pyruvate kinase is a key step for producing pyruvate, but its activity in some cyanobacterial species can be inhibited by high ATP/ADP ratios during photosynthesis [19]. For example, the majority of pyruvate synthesis in Synechocystis 6803 [3] and Synechococcus 7942 [14] during slow growth with sub-optimal conditions was a result of malate degradation. Overexpression of pyruvate kinase for pyruvate synthesis improved cyanobacterial lactate and alcohol productions [20, 21]. Synechococcus 2973 has naturally evolved to overcome this pyruvate synthesis bottleneck, de-coupling pyruvate kinase inhibition from high photosynthesis rates. Furthermore, as shown in Fig. 2 and genome scale modeling [22], strong photorespiration is essential for high light photosynthesis and maximal Synechococcus 2973 growth despite its net carbon loss. Synechococcus 2973 exhibits twice the O₂ evolution rate of Synechococcus 7942 [11]. Water splitting results in heightened local oxygen concentration and increased linear electron flow, both of which may increase the need for photorespiration or cyclic electron flow [23].

The OPPP can generate NADPH and oxidize sugar phosphates as a means to derive more reduced compounds but its activation at the same time as the Calvin
Cycle would generate a futile cycle that reduces the efficient use of carbon and energy, which is why in higher plants the two are coordinated regulated by multiple mechanisms including redox state and pH [24, 25]. Due to lack of sufficient regulation or Calvin cycle bottlenecks, slow growing Synechococcus 6803 demonstrated ~12% CO₂ loss through OPPP under continuous light [3], while Synechococcus does not exhibit significant flux through this pathway. To further test this observation, we constructed Synechococcus 2973 zwf and pgl deletion mutants by disrupting the gene encoding G6P dehydrogenase and 6-phosphogluconolactonase, respectively, so that G6P could not be converted to 6PG (Additional file 1: Figure S2e, f). The mutants showed similar growth rate to the wild-type strain under continuous light and diurnal bioreactor conditions (Fig. 1), indicating the OPPP did not provide a benefit to the organism and was futile for photoautotrophic metabolism in Synechococcus 2973 [26]. These mutants and metabolic insights from the quantification of central carbon fluxes suggest new strategies for the redirection of carbon flux for the bioengineering of photosynthetic organisms. The observation that Synechococcus 2973 exhibits a similar growth phenotype to Synechococcus 7942 under diurnal conditions emphasizes the differences that can occur as a result of changes in light provided by optimal bioreactor conditions.

The optimal photoautotrophic metabolism (reduced pool sizes, enhanced energy levels, and repartitioning of biomass composition)

Metabolite concentrations are tied to input/output fluxes and can provide insight into biochemical level regulation [27, 28]. The lack of correlation between pool sizes and absolute fluxes has been shown in literature [15, 29], as well as here between Synechococcus 2973, Synechococcus 7942, and Synechocystis 6803. Compared to SF cultures and Synechocystis 6803, the
concentrations of a number of central metabolites from glycolysis and TCA in Synechococcus 2973 were reduced during the PBR conditions. Less metabolite accumulation in central metabolism under PBR conditions could indicate reduced feedback or higher rates of metabolite turnover. From literature, it has been suggested that as cell growth rates and carbon fluxes increase, the fraction of anabolic enzymes is increased to pull central metabolites towards biomass synthesis and thus decreases their pool sizes [30]. Conversely, the heterotroph E. coli had intracellular acetyl-CoA/organic acid concentrations much higher than cyanobacteria (Fig. 3c). Such observations, together with cultivation experiments using organic acids, indicated that the TCA cycle and its associated anabolic pathways in Synechococcus 2973 operate at lower rates than in E. coli, which would limit production of chemicals from the cyanobacterial TCA cycle [4]. In Synechococcus 2973, the photosynthetic capacity generates sugar phosphates readily (FBP phosphatase/aldolase, fructose-bisphosphatase, transketolase, and RuBisCO reactions) that can enhance its growth [31–33] and could potentially be applied for biotechnological productions [34].

Additionally, the higher energy charge in the PBR than SF for Synechococcus 2973 may benefit anabolic metabolism and cell tolerance to stress conditions [35]. Cyanobacterial biosynthesis has been closely tied to photosynthesis and changing bioreactor conditions [36]. In SF cultures, there was an increase in lower energy molecules (NADP+, ADP, and AMP) inhibiting anabolism, possibly resulting in higher levels of some central metabolites. The PBR grown Synechococcus 2973 showed the highest NADPH level that may reflect enhanced light-harvesting that can facilitate organic carbon assimilation and biomass growth [20, 37]. Synechococcus 2973 adjusts its PSII/PSI ratio throughout growth although it was found to have a slightly lower PSII/PSI ratio than Synechococcus 7942 while exhibiting greater O2 evolution [11]. Additionally, it may have a more optimal ATP/NADPH ratio to sustain metabolism while not exceeding ATP demand [22], which may be important to further unlocking cyanobacteria growth constraints. While CO2 fixation rate is a key factor for differences in phenotypes, CO2 fixation rate is influenced by optimal energy balance, biomass distribution/carbon fluxes, and compartmentalization of photosynthetic reactions.

Synechococcus 2973 PBR cells produce little glycogen (8–10 times less than PBR cells of Synechococcus 7942 and Synechocystis 6803 [38]). When less carbon is allocated to make glycogen, more can be partitioned toward synthesis of biomass, resulting in greater protein levels in Synechococcus 2973 than 7942. Consequently, Synechococcus 2973 may have more photosystems/RuBisCO proteins for effective photosynthesis. In contrast, Synechococcus 2973 produced significantly more glycogen under suboptimal SF conditions, leading to longer cell doubling times.

Metabolic features of Synechococcus 2973 for bio-production applications

Understanding and exploring native pathways with high metabolic strengths is a promising direction for future microbial cell factories [39]. Our flux results indicate that Synechococcus 2973 is advantageous for the synthesis of targeted products from sugar phosphate pathways under optimal bioreactor conditions. Recently, researchers have engineered Synechococcus 2973 to produce ~9 g L−1 of sucrose under potassium chloride stress [40]. This engineered strain also demonstrated potential to produce valuable polysaccharide products. It has been observed that while cyanobacteria have a flexible photosynthetic metabolism, there are a high number of essential metabolic genes involved in photautotrophic growth. This presents additional challenges for mutant strain generation or extensive genetic modification, and the knowledge gained from this study will help surmount the challenges facing cyanobacterial photo-biorefineries [41].

Subpopulations, metabolic inactive pools, and substrate channeling

Modeling results indicated the presence of unlabeled pools that required additional dilution factors for isotopic data fitting. For SF cultures, an increase in dilution factors was necessary to improve fitting, still resulting in an unaccepted SF fit. The dilution of the metabolite active pool by unlabeled metabolites infers phenotypic heterogeneity, i.e., the presence of inactive/non-growth cells or the presence of pre-existing sources of carbon that are more slowly turned over after 13C-bicarbonate pulses [42]. Inactive cells would not be involved in significant photosynthesis within the duration of the labeling period and may induce reflux of unlabeled carbon into the metabolic network [43]. Inactive subpopulations can be increased by poor light transmittance due to suboptimal culture conditions, cell self-shading and/or insufficient mixing conditions [44]. This may explain why Synechococcus 2973 does not exhibit significant growth advantages in SF conditions.

Metabolite channeling could also contribute to the observed phenotype. Channels pass metabolite intermediates between enzymes without intracellular diffusion. Channeling increases pathway efficiency [45] and is often associated with microcompartments (i.e., carboxysomes) or enzyme proximity [45–47]. For fast growing E. coli species, glycolysis channeling has been evident [48]. Substrate channeling can be inferred from transient labeling experiments [45]. When tracing 13C-enrichment
from 3PGA to downstream metabolites in *Synechocystis* 6803, it was previously found that certain downstream metabolites could be labeled faster than precursors (e.g., PEP was labeled faster than 3PGA) [3, 29]. In this study, we observed $^{13}$C-enrichment might not completely follow expected precursor-product relationships (Additional file 1: Figure S6). Comparing to *Synechococcus* 7942, *Synechococcus* 2973 showed rapid $^{13}$C-enrichment in 3TP over precursors. This can be explained by possible channels in the Calvin cycle that may influence the labeling of certain metabolites and provide advantages for CO$_2$ fixation. *Synechococcus* 2973 also shows evidence for channeling of glycolysis intermediates towards the TCA pathway with faster citrate M+4 distribution over time that exceeds the $^{13}$C-enrichment levels of 3PGA (Fig. 5). However, whether differences are a result of traditional channeling definitions or represent the presence of multiple pools separated through additional spatial compartments cannot be conclusively determined from our studies. These observations suggest a heterogenous distribution of intracellular metabolites as well as possible spatial organization of enzymes to confer growth benefits.

**Conclusion**

*Synechococcus* 2973 demonstrates high photosynthesis, efficient carbon fixation and small fluxes towards carbon loss and transient storage pathways under optimal growth conditions. These metabolic traits suggest this strain may be a promising platform to produce high-value compounds in well-controlled PBR conditions. Under suboptimal conditions, *Synechococcus* 2973 has inferior photosynthesis and decreased biomass synthesis, causing an accumulation of central carbon metabolites. Moreover, *Synechococcus* strains with different growth rates maintain similar flux distributions in central pathways. These observations offer insight into flux response to adaptive evolution. From a technological perspective, INST-MFA of cyanobacteria is not only important for genome-to-phenome mapping but also crucial for the rational application of platform photo-biorefineries. Currently, INST-MFA is still challenging, and flux results may be influenced by substrate channeling or subpopulations. Sorting out these factors will require advanced labeling experiments [40–52] to complement computer modeling and labeling experiments. Therefore, mutant creations and biomass composition
Methods
Cultivation conditions, and transient labeling experiments
*Synechococcus* 2973 was grown in BG-11 medium (pH 8.0–8.5) at 38 °C with 3% CO_2_ aeration (2000 mL min⁻¹) and continuous 500 µmol photons/m²s under photobioreactors (Additional file 1: photo 1). *Synechococcus* 2973 was also grown in shaking flasks (atmospheric CO_2_, 38 °C and continuous and diurnal (12 h intervals) 100 µmol photons/m²s light, 250 rpm). *Synechococcus* 7942 was grown in identical conditions, except in its optimal PBR conditions, where the light condition was 300 µmol photons/m²s. Higher light irradiance inhibits *Synechococcus* 7942 growth [2]. For the isotopic pool-size experiment, *Synechocystis* 6803 was grown in identical photobioreactors at 30 °C, 3% CO_2_ aeration and continuous 300 µmol photons/m²s light. For each labeling experiment, cultures were diluted to an OD_{730} 0.05 in BG-11 medium without citrate (so citrate could not be used as a carbon source or affect bulk citrate labeling dynamics) and grown to OD_{730} 0.6 (exponential phase) before the experiment was initiated. Prior to the experiment, the 3% CO_2_ aeration was replaced with air (0.04% CO_2_) and a 2-mL aliquot of saturated NaHCO_3_ (> 98% purity, Sigma Aldrich, St Louis) was injected into each PBR or shaking flask (SF) for a final concentration of 4 g L⁻¹ of NaHCO_3_ to saturate the cell with CO_2_. The use of 13C-bicarbonate rather than 1C-bicarbonate quickly equilibrates within the media without gas–liquid mass transfer limitations.

After the 13C-pulse, each identical culture in the PBR and SF was quenched at different time intervals (20, 30, 40, 60, 90, 150, 300, 600, 1200, 3600, 7200 s). These time-courses samples were analyzed to capture labeling dynamics of metabolites. Timed biological duplicates were used to generated standard deviations for experimentally measured MID values; however, a minimum standard deviation of 3% was used in INCA model.

Metabolite quenching, extraction, and analysis
Fast-cooling via cold solvents followed by centrifugation is reliable for recovering intracellular labeled metabolites [13]. After a 13C-pulse, samples were quenched by mixing with ice-cold minimal medium (≤ −5 °C, without a carbon source) in a falcón tube, which was immediately bathed in liquid N_2_ to sustain cold temperature (To avoid ice formation, the culture sample was rapidly swirled for several seconds). Then the quenched biomass samples (≤ 0 °C) were harvested via refrigerator centrifuge (3 min, 6000 rpm). For LC–MS analysis, the pelleted samples were extracted with a chloroform–methanol method [15]. Ion-pairing LC–MS/MS was performed at the Proteomics and Mass Spectrometry Facility, Donald Danforth Plant Science Center, St. Louis (details in Additional file 1: Part 1). The labeling of energy molecules (e.g., NADH and ATP), organic acids and acetyl-CoA were quantified using hydrophilic interaction liquid chromatography (HILIC) coupled to electrospray time-of-flight MS [53] at Joint Bio-Energy Institute, CA.

Estimation of relative pool sizes of metabolites
Changes of metabolite pool sizes in cyanobacteria were analyzed by an MS isotopomer ratio approach [54] (using fully labeled cell extracts as internal standards for semi-quantitative metabolomics [55]). Specifically, E. coli K-12 was cultured with uniformly labeled 13C-glucose and 13C-sodium bicarbonate in an M9 minimal media. The estimation of metabolite pool sizes was performed by mixing a known amount of labeled *E. coli* biomass with unlabeled cyanobacteria cultures (three biological replicates). Then, the mixtures were quenched by a liquid N_2_ bath and extracted for HILIC-MS analysis (Additional file 1: Figure S2c). The isotopic ratio of each metabolite (labeled vs. unlabeled) was normalized by the amount of *E. coli* and cyanobacterial biomass, respectively.

Biomass composition analysis
The cells were harvested during exponential growth by centrifugation, and cell pellets were washed with 0.9% NaCl and ddH_2_O and then freeze-dried. Protein and amino acid compositional analysis was performed by the Molecular Structure Facility, University of California (Davis, CA). Carbohydrates, lipids, chlorophyll a, and ash were measured by previously reported methods (details in Additional file 1: Part 1).

GC–MS analysis of proteinogenic amino acids
To confirm the capability of *Synechococcus* 2973 to uptake organic acids, unlabeled seed cultures were inoculated (4% inoculation ratio) into the medium with fully labeled NaHCO_3_ and unlabeled substrates (6 mM) including acetate, TCA cycle intermediates (e.g., malate and citrate), and pyruvate in shaking flasks or PBRs for cultivation of 48 h. The labeling in proteinogenic amino acids was then analyzed using a TBDMS method [56].

Isotopically nonstationary MFA
MFA is estimated based on isotope labeling dynamics of the free metabolites. The MFA model included the Calvin cycle, pentose phosphate pathway, TCA cycle, the glyoxylate shunt, anaplerotic pathways, and photorepospiration pathway. A list of reactions in the network model with their atom transitions is provided in Additional file 1: Table S1. A lumped biomass equation based on biomass
composition analysis and biochemical equations was used to fit the model to biomass production (Additional file 1: Table S2). The INCA platform [8] used a custom MATLAB ODE solver to fit over 890, 600, and 800 individual mass fragments for Synechococcus 2973 PBR, SF, and Synechococcus 7942 PBR to obtain Fluxes in the assumed cyanobacterial network, respectively. INCA provided a goodness of fit via a Chi square statistical test as well as confidence intervals of all estimated parameters (Model formulation details are included in the Additional file 1: Part 1).

Construction of the zwf, pgp, and the succinate dehydrogenase deletion mutants

Three mutants were built, which served to independently validate the related aspects of metabolism described by the modeling process. The Δzwf mutant was constructed by insertional mutagenesis of the first enzyme of the OPPP, glucose-6-phosphate 1 dehydrogenase. The Δzwf mutant was characterized via growth rate (Fig. 1). The mutant activity was confirmed via IP-LC−MS/MS by the absence of the downstream product of its disrupted gene, 6-phosphogluconate (Additional file 1: Figure S2a–d). The Δpgp and Δsdh mutants were constructed by a cngR-based CRISPR system used to delete succinate dehydrogenase from start codon to stop codon from the chromosome (Additional file 1: Part 1).

Additional file


Abbreviations

3PG: 2-phosphoglycerate; 2PG: 2-phosphoglycerate; 3PGA: 3-phosphoglycerate; dPGP: 4-phosphoglycerate; Ac: acetate; AcCoA: acetyl-CoA; ADRP: adenosine diphosphate; ADP: adenosine diphosphate; ADPG: adenosine diphosphoglucose; ADI: alpha-ketoglutarate; AMP: adenosine monophosphate; ATP: adenosine triphosphate; BPG: 1,3-bisphosphoglyceric acid; CIT: citrate; DHAP: dihydroxyacetone phosphate; EPF: enolpyruvyl phosphate; F6P: fructose 6-phosphate; FBP: fructose 1,6-diphosphate; FUM: fumarate; G6P: glucose 1-phosphate; GAP: glucose 6-phosphate; GABA: gamma-aminobutyric acid; GDC: glyceraldehyde 3-phosphate: GCL: glycine: GLY: glycine; GOG: oxoglutarate; HAP: malate: MEG: methanol: MGD: mass spectrometer distribution; NADP*: nicotinamide adenine dinucleotide phosphate; NADH: nicotinamide adenine dinucleotide: NADPH: nicotinamide adenine dinucleotide phosphate; reduced; OAA: oxaloacetate; OPPP: oxidative pentose phosphate pathway; PEP: phosphoenolpyruvate; PFR: pyruvate; R5P: ribose 5-phosphate; RuBP: ribulose 1,5-diphosphate; RuBP: ribulose 1,5-diphosphate; S7P: sedoheptulose 7-phosphate; S6P: sedoheptulose 1,7-diphosphate; SUCC: succinate; TCA: tricarboxylic acid cycle; UDP: uridine diphosphoglucose; XSP: xylulose 5-phosphate.

Authors’ contributions

HBP, CDM, and YIT initiated the project. MHA, FAM, YIT and ODKA designed and performed 13C-MFA and LC−MS analysis. YIT, HBP, MHA, and WH performed growth experiments and physiological analysis. HBP, ML, JL, MHA constructed mutants and helped with strain characteristics. CDM, SG and LH assisted the flux modeling and statistics. 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 Biology, Temple University, Philadelphia, PA 19122, USA. 3 Donald Danforth Plant Science Center, St. Louis, MO 63130, USA. 4 Department of Chemical Engineering, The Pennsylvania State University, University Park, PA 16802, USA. 5 United States Department of Agriculture, Agricultural Research Service, St. Louis, MO 63130, USA.

Acknowledgements

We are thankful for Dr. Edward Baitoo from Joint Bio-energy Institute for advice on LC−MS analysis and on Dr. Wei Lu at Michigan State University for biomass analysis.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

All datasets used and/or analyzed during the current study are available in this published article, its supplementary information files, or is available from the corresponding author on reasonable request.

Consent for publication

Not applicable.

Ethical approval and consent to participate

Not applicable.

Funding

This work was supported by DOE-BER (DESC0012722) to CDM, TST, and HBP, and NSF EF-1105249, DOE NE-000000220, and USDA-ARS funding to support FM and DKA. Analyses were in part enabled by major research instrumentation grants from the NSF (DBI-1427621 and DBI-0521350) that supported the acquisition of mass spectrometers within the Proteomics and Mass Spectrometry Core Facility at the Donald Danforth Plant Science Center. The core facility is additionally acknowledged for help with metabolite evaluations.

Publisher’s Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 8 June 2017 Accepted: 6 November 2017

Published online: 16 November 2017

References


11. Appendix 3: Supporting information for Chapter 3

Supporting Information: Carboxysome mutant reveals influence of enzyme compartmentalization on cellular metabolisms and metabolic network rigidity

Table S1. Measured succinic acid in the supernatant from cultures harvested during later exponential growth at 1% CO₂, 125 µmol photons m⁻²s⁻¹. A UV-method (Absorbance 340 nm) was used to measure succinate and A+ media served as a control.

<table>
<thead>
<tr>
<th>Strain</th>
<th>Succinic acid (mg/L)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Synechococcus</em> PCC 7002 control</td>
<td>1.7 ± 1.0</td>
</tr>
<tr>
<td><em>Synechococcus</em> PCC 7002 ΔccmKLMN</td>
<td>2.8 ± 0.3</td>
</tr>
</tbody>
</table>
Figure S1. **Labeling dynamics of key metabolites.** Percentages of total $^{13}$C enrichment are shown by red line. $^{13}$C enrichment was calculated at 15 minutes using the formula $\frac{1}{N} \sum_{i=1}^{N} M_i \times i$, where $M_i$ is the mass isotopomer distribution (e.g., M+0, M+1). Standard deviations are based on biological replicates (n=2).
Figure S2. Comparison of energy-molecule pool size between *Synechococcus* PCC 7002 control and ∆ccmKLMN strain. Relative pool size comparisons between the two strains. The relative pool size is normalized a $^{13}$C-labeled internal standard and optical density. A y-axis of 1 represents equal metabolite concentrations while a y-axis > 1 indicates the factor of pool size in ∆ccmKLMN greater than PCC 7002 ctrl.

Table S2. List of metabolic reactions and carbon transitions for the cyanobacterial network.

**Glycolysis and OPP**

G6P (abcdef) $\leftrightarrow$ F6P (abcdef)

G6P (abcdef) $\rightarrow$ RU5P (bcdef) + CO2 (a)

F6P (abcdef) $\leftrightarrow$ FBP (abcdef)

FBP (abcdef) $\leftrightarrow$ DHAP (cba) + G3P (def)

DHAP (abc) $\leftrightarrow$ G3P (abc)

G3P (abc) $\leftrightarrow$ 3PGA (abc)

3PGA (abc) $\leftrightarrow$ 2PGA (abc)

2PGA (abc) $\leftrightarrow$ PEP (abc)
PEP (abc) \rightarrow PYR (abc)

**Calvin cycle**

RU5P (abcde) \leftrightarrow X5P (abcde)

RU5P (abcde) \leftrightarrow R5P (abcde)

RU5P (abcde) \leftrightarrow RUBP (abcde)

RUBP (abcde) + CO2 (f) \rightarrow 3PGA (cde) + 3PGA (fba)

X5P (abcde) \leftrightarrow G3P (cde) + EC2 (ab)

F6P (abcdef) \leftrightarrow E4P (cdef) + EC2 (ab)

S7P (abcdefg) \leftrightarrow R5P (cdefg) + EC2 (ab)

F6P (abcdef) \leftrightarrow G3P (def) + EC3 (abc)

S7P (abcdefg) \leftrightarrow E4P (defg) + EC3 (abc)

DHAP (abc) + E4P (defg) \rightarrow SBP (cbadefg)

SBP (abcdefg) \rightarrow S7P (abcdefg)

**TCA**

PYR (abc) \rightarrow ACA (bc) + CO2 (a)

OAA (abcd) + ACA (ef) \leftrightarrow CIT (dcbfea)

CIT (abcdef) \leftrightarrow ICI (abcdef)

ICI (abcdef) \leftrightarrow AKG (abcde) + CO2 (f)

SUC (abcd) \leftrightarrow FUM (abcd)

MAL (abcd) \leftrightarrow OAA (abcd)

AKG (abcde) \rightarrow SUC (bcde) + CO2 (a)

**Amphibolic reactions**

MAL (abcd) \rightarrow PYR (abc) + CO2 (d)
PEP (abc) + CO2 (d) -> OAA (abcd)

**Photorespiration**
RUBP (abcde) -> 3PGA (cde) + 2PG (ba)
2PG (ab) <-> GOX (ab)
GOX (ab) + GOX (cd) -> GA (abd) + CO2 (c)
GA (abc) <-> 2PGA (abc)

**Carbohydrate synthesis**
G6P (abcdef) <-> G1P(abcdef)
G1P (abcdef) <-> ADPGLU (abcdef)
G1P (abcdef) <-> UDPGLU (abcdef)
ADPGLU (abcdef) -> CARB (abcdef)
UDPGLU (abcdef) -> CARB (abcdef)

**Biomass synthesis**

*ccmO*
0.304*R5P + 2.87*ACA + 0.35*E4P + 0.874*3PGA + 0.7*PEP + 3.178*PYR + 1.158*OAA + 0.86*AKG + 0.218*G3P + 1.247*CARB -> Biomass + 0.423*FUM + 1.851*CO2

*ctrl*
0.294*R5P + 2.32*ACA + 0.290*E4P + 0.724*3PGA + 0.58*PEP + 2.768*PYR + 1.028*OAA + 0.8*AKG + 0.218*G3P + 1.273*CARB -> Biomass + 0.388*FUM + 1.536*CO2

Table S3. Reported flux values for carboxysome-deficient mutant normalized to 100 mmol CO2 uptake with 95% confidence intervals.

<table>
<thead>
<tr>
<th>REACTION</th>
<th>FLUX</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1P &lt;-&gt; ADPG</td>
<td>2.9401</td>
<td>0</td>
<td>3.1172</td>
</tr>
</tbody>
</table>

199
<table>
<thead>
<tr>
<th>Reaction</th>
<th>Free Energy (kJ/mol)</th>
<th>Reaction Free Energy (kJ/mol)</th>
<th>Standard Free Energy (kJ/mol)</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1P &lt;-&gt; UDPG</td>
<td>0.1771</td>
<td>0</td>
<td>3.1172</td>
</tr>
<tr>
<td>ADPG -&gt; CARB</td>
<td>2.9401</td>
<td>0</td>
<td>3.1172</td>
</tr>
<tr>
<td>UDPG -&gt; CARB</td>
<td>0.1771</td>
<td>0</td>
<td>3.1172</td>
</tr>
<tr>
<td>G6P &lt;-&gt; G1P</td>
<td>3.1172</td>
<td>3.1172</td>
<td>3.1172</td>
</tr>
<tr>
<td>G6P &lt;-&gt; F6P</td>
<td>-3.1358</td>
<td>-3.8582</td>
<td>-3.1358</td>
</tr>
<tr>
<td>G6P -&gt; 6PG</td>
<td>0.0186</td>
<td>0.0186</td>
<td>0.741</td>
</tr>
<tr>
<td>6PG -&gt; RU5P + CO2</td>
<td>0.0186</td>
<td>0.0186</td>
<td>0.741</td>
</tr>
<tr>
<td>F6P &lt;-&gt; FBP</td>
<td>-81.5917</td>
<td>-82.4334</td>
<td>-78.8007</td>
</tr>
<tr>
<td>FBP &lt;-&gt; DHAP + G3P</td>
<td>-81.5917</td>
<td>-82.4334</td>
<td>-78.8007</td>
</tr>
<tr>
<td>DHAP &lt;-&gt; G3P</td>
<td>-82.5157</td>
<td>-83.6758</td>
<td>-80.1135</td>
</tr>
<tr>
<td>G3P &lt;-&gt; 3PGA</td>
<td>-205.704</td>
<td>-208.223</td>
<td>-202.561</td>
</tr>
<tr>
<td>3PGA &lt;-&gt; 2PGA</td>
<td>20.2565</td>
<td>19.9989</td>
<td>21.3306</td>
</tr>
<tr>
<td>2PGA &lt;-&gt; PEP</td>
<td>24.0751</td>
<td>24.0751</td>
<td>24.8792</td>
</tr>
<tr>
<td>PEP &lt;-&gt; PYR</td>
<td>17.2682</td>
<td>12.5943</td>
<td>17.9652</td>
</tr>
<tr>
<td>RU5P &lt;-&gt; X5P</td>
<td>-79.3798</td>
<td>-80.2052</td>
<td>-77.009</td>
</tr>
<tr>
<td>RU5P &lt;-&gt; R5P</td>
<td>-38.4925</td>
<td>-38.9052</td>
<td>-37.3071</td>
</tr>
<tr>
<td>RU5P -&gt; RUBP</td>
<td>117.891</td>
<td>114.2738</td>
<td>119.1046</td>
</tr>
<tr>
<td>RUBP + CO2 -&gt; 3PGA + 3PGA</td>
<td>110.2539</td>
<td>109.6525</td>
<td>112.2565</td>
</tr>
<tr>
<td>X5P &lt;-&gt; G3P + EC2</td>
<td>-79.3798</td>
<td>-80.2052</td>
<td>-77.009</td>
</tr>
<tr>
<td>F6P &lt;-&gt; E4P + EC2</td>
<td>40.1274</td>
<td>38.942</td>
<td>40.54</td>
</tr>
<tr>
<td>S7P &lt;-&gt; R5P + EC2</td>
<td>39.2525</td>
<td>38.0671</td>
<td>39.6651</td>
</tr>
<tr>
<td>F6P &lt;-&gt; G3P + EC3</td>
<td>38.3285</td>
<td>36.8091</td>
<td>38.7149</td>
</tr>
<tr>
<td>S7P &lt;-&gt; E4P + EC3</td>
<td>-38.3285</td>
<td>-38.7149</td>
<td>-36.8091</td>
</tr>
<tr>
<td>DHAP + E4P -&gt; SBP</td>
<td>0.924</td>
<td>0.6991</td>
<td>1.3489</td>
</tr>
<tr>
<td>Reaction</td>
<td>SBP -&gt; S7P</td>
<td>PYR -&gt; ACA + CO2</td>
<td>OAA + ACA -&gt; CIT</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------------</td>
<td>-------------------</td>
<td>------------------</td>
</tr>
<tr>
<td></td>
<td>0.924</td>
<td>12.454</td>
<td>2.1498</td>
</tr>
<tr>
<td></td>
<td>0.6991</td>
<td>9.324</td>
<td>2.1498</td>
</tr>
<tr>
<td></td>
<td>1.3489</td>
<td>12.454</td>
<td>2.9538</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CYCICLIC -&gt; ICI</td>
<td>2.1498</td>
<td>2.1498</td>
<td>2.9538</td>
</tr>
<tr>
<td></td>
<td>2.1498</td>
<td>2.1498</td>
<td>2.9538</td>
</tr>
<tr>
<td></td>
<td>2.1498</td>
<td>2.1498</td>
<td>2.9538</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AKG -&gt; SUC + CO2</td>
<td>1.00E-07</td>
<td>2.49E-14</td>
<td>0.8041</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SUC -&gt; FUM</td>
<td>-1.0700</td>
<td>-1.0700</td>
<td>-0.2659</td>
</tr>
<tr>
<td></td>
<td>2.49E-14</td>
<td>0.8041</td>
<td></td>
</tr>
<tr>
<td>MAL &lt;-&gt; OAA</td>
<td>-4.1732</td>
<td>-4.1732</td>
<td>2.78E-05</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OAA &lt;-&gt; FUM</td>
<td>0.0126</td>
<td>-0.7915</td>
<td>0.0126</td>
</tr>
<tr>
<td></td>
<td>-0.7915</td>
<td>2.78E-05</td>
<td></td>
</tr>
<tr>
<td>MAL -&gt; PYR + CO2</td>
<td>4.1732</td>
<td>2.78E-05</td>
<td>4.1732</td>
</tr>
<tr>
<td></td>
<td>4.1732</td>
<td>2.78E-05</td>
<td>4.1732</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEP + CO2 -&gt; OAA</td>
<td>5.0571</td>
<td>NaN</td>
<td>9.2303</td>
</tr>
<tr>
<td></td>
<td>2.78E-05</td>
<td>4.1732</td>
<td></td>
</tr>
<tr>
<td>RUBP -&gt; 3PGA + 2PG</td>
<td>7.6371</td>
<td>5.456</td>
<td>8.1179</td>
</tr>
<tr>
<td>2PG -&gt; GOX</td>
<td>7.6371</td>
<td>5.456</td>
<td>8.1179</td>
</tr>
<tr>
<td>GOX + GOX + -&gt; GA + CO2</td>
<td>3.8186</td>
<td>2.728</td>
<td>4.059</td>
</tr>
<tr>
<td>GA &lt;-&gt; 2PGA</td>
<td>3.8186</td>
<td>2.728</td>
<td>4.059</td>
</tr>
<tr>
<td>MAL ACTIVE POOL %</td>
<td>55.3376</td>
<td>49.4393</td>
<td>61.2363</td>
</tr>
<tr>
<td>G1P ACTIVE POOL %</td>
<td>96.0571</td>
<td>93.976</td>
<td>99.152</td>
</tr>
<tr>
<td>G3P ACTIVE POOL %</td>
<td>86.0372</td>
<td>83.7826</td>
<td>87.624</td>
</tr>
<tr>
<td>RUBP ACTIVE POOL %</td>
<td>73.3102</td>
<td>23.3224</td>
<td>29.9672</td>
</tr>
<tr>
<td>ADPG ACTIVE POOL %</td>
<td>80.8518</td>
<td>77.3005</td>
<td>84.5299</td>
</tr>
<tr>
<td>2PG ACTIVE POOL %</td>
<td>38.2508</td>
<td>33.3206</td>
<td>42.933</td>
</tr>
<tr>
<td>S7P ACTIVE POOL %</td>
<td>99.4983</td>
<td>98.227</td>
<td>100</td>
</tr>
<tr>
<td>CIT ACTIVE POOL %</td>
<td>49.8401</td>
<td>45.6888</td>
<td>52.906</td>
</tr>
<tr>
<td>------------------</td>
<td>---------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>AKG ACTIVE POOL %</td>
<td>37.4669</td>
<td>31.2021</td>
<td>45.7623</td>
</tr>
<tr>
<td>DHAP ACTIVE POOL %</td>
<td>96.8664</td>
<td>96.0636</td>
<td>97.3956</td>
</tr>
<tr>
<td>BIOMASS</td>
<td>2.4997</td>
<td>2.4997</td>
<td>2.4997</td>
</tr>
</tbody>
</table>

Table S4. Reported flux values for the PCC 7002 control strain normalized to 100 mmol CO₂ uptake with 95% confidence intervals.

<table>
<thead>
<tr>
<th>REACTION</th>
<th>FLUX</th>
<th>LB</th>
<th>UB</th>
</tr>
</thead>
<tbody>
<tr>
<td>G1P &lt;-&gt; ADPG</td>
<td>3.2956</td>
<td>0</td>
<td>3.2956</td>
</tr>
<tr>
<td>G1P &lt;-&gt; UDPG</td>
<td>0.0102</td>
<td>0</td>
<td>3.2956</td>
</tr>
<tr>
<td>ADPG -&gt; CARB</td>
<td>3.2956</td>
<td>0</td>
<td>3.2956</td>
</tr>
<tr>
<td>UDPG -&gt; CARB</td>
<td>0.0102</td>
<td>0</td>
<td>3.2956</td>
</tr>
<tr>
<td>G6P &lt;-&gt; G1P</td>
<td>3.3058</td>
<td>3.3058</td>
<td>3.3058</td>
</tr>
<tr>
<td>G6P &lt;-&gt; F6P</td>
<td>-3.3109</td>
<td>-4.4113</td>
<td>-3.3109</td>
</tr>
<tr>
<td>G6P -&gt; 6PG</td>
<td>0.0051</td>
<td>0.0000</td>
<td>0.8005</td>
</tr>
<tr>
<td>6PG -&gt; RU5P + CO2</td>
<td>0.0051</td>
<td>0.0000</td>
<td>0.8005</td>
</tr>
<tr>
<td>F6P &lt;-&gt; FBP</td>
<td>-80.3426</td>
<td>-96.0094</td>
<td>68.2221</td>
</tr>
<tr>
<td>FBP &lt;-&gt; DHAP + G3P</td>
<td>-80.3426</td>
<td>-96.0094</td>
<td>68.2221</td>
</tr>
<tr>
<td>DHAP &lt;-&gt; G3P</td>
<td>-80.9157</td>
<td>-81.9816</td>
<td>-80.8641</td>
</tr>
<tr>
<td>G3P &lt;-&gt; 3PGA</td>
<td>-210.805</td>
<td>-210.8050</td>
<td>-203.8840</td>
</tr>
<tr>
<td>3PGA &lt;-&gt; 2PGA</td>
<td>23.3265</td>
<td>19.8311</td>
<td>24.8898</td>
</tr>
<tr>
<td>2PGA &lt;-&gt; PEP</td>
<td>26.4518</td>
<td>23.0811</td>
<td>26.7398</td>
</tr>
<tr>
<td>PEP &lt;-&gt; PYR</td>
<td>17.1981</td>
<td>5.4974</td>
<td>18.4974</td>
</tr>
<tr>
<td>Reaction</td>
<td>Ka</td>
<td>Kb</td>
<td>Kc</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>------</td>
<td>------</td>
<td>------</td>
</tr>
<tr>
<td>RU5P ↔ X5P</td>
<td>-73.0284</td>
<td>-78.2532</td>
<td>-72.3706</td>
</tr>
<tr>
<td>RU5P ↔ R5P</td>
<td>-35.3217</td>
<td>-37.9401</td>
<td>-35.3217</td>
</tr>
<tr>
<td>RU5P → RUBP</td>
<td>119.525</td>
<td>119.5250</td>
<td>119.7250</td>
</tr>
<tr>
<td>RUBP + CO2 → 3PGA + 3PGA</td>
<td>113.0252</td>
<td>113.0252</td>
<td>113.2107</td>
</tr>
<tr>
<td>X5P ↔ G3P + EC2</td>
<td>-73.8284</td>
<td>-78.3706</td>
<td>-73.8284</td>
</tr>
<tr>
<td>F6P ↔ E4P + EC2</td>
<td>36.9499</td>
<td>36.5379</td>
<td>39.5966</td>
</tr>
<tr>
<td>S7P ↔ R5P + EC2</td>
<td>36.0785</td>
<td>36.0785</td>
<td>38.7740</td>
</tr>
<tr>
<td>F6P ↔ G3P + EC3</td>
<td>36.9499</td>
<td>33.0022</td>
<td>NaN</td>
</tr>
<tr>
<td>S7P ↔ E4P + EC3</td>
<td>-36.9499</td>
<td>NaN</td>
<td>33.0022</td>
</tr>
<tr>
<td>DHAP + E4P → SBP</td>
<td>3.0763</td>
<td>NaN</td>
<td>6.0763</td>
</tr>
<tr>
<td>SBP → S7P</td>
<td>3.0763</td>
<td>NaN</td>
<td>6.0763</td>
</tr>
<tr>
<td>PYR → ACA + CO2</td>
<td>12.396</td>
<td>8.8499</td>
<td>13.9086</td>
</tr>
<tr>
<td>OAA + ACA → CIT</td>
<td>2.1411</td>
<td>2.0693</td>
<td>2.3279</td>
</tr>
<tr>
<td>CIT ↔ ICI</td>
<td>2.1411</td>
<td>2.0693</td>
<td>2.3279</td>
</tr>
<tr>
<td>ICI ↔ AKG + CO2</td>
<td>2.1411</td>
<td>2.0693</td>
<td>2.3279</td>
</tr>
<tr>
<td>AKG → SUC + CO2</td>
<td>1.00E-07</td>
<td>0</td>
<td>0.0587</td>
</tr>
<tr>
<td>SUC ↔ FUM</td>
<td>-0.6500</td>
<td>-0.6500</td>
<td>-0.5913</td>
</tr>
<tr>
<td>MAL ↔ OAA</td>
<td>-3.0165</td>
<td>-3.4086</td>
<td>-2.7356</td>
</tr>
<tr>
<td>OAA ↔ FUM</td>
<td>-0.4506</td>
<td>-0.5093</td>
<td>-0.4506</td>
</tr>
<tr>
<td>MAL → PYR + CO2</td>
<td>3.0165</td>
<td>-3.4086</td>
<td>-2.7356</td>
</tr>
<tr>
<td>PEP + CO2 ↔ OAA</td>
<td>6.0891</td>
<td>6.0055</td>
<td>6.1636</td>
</tr>
<tr>
<td>RUBP → 3PGA + 2PG</td>
<td>6.5123</td>
<td>0</td>
<td>8.1206</td>
</tr>
<tr>
<td>2PG → GOX</td>
<td>6.5123</td>
<td>0</td>
<td>8.1206</td>
</tr>
<tr>
<td>Reaction</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>--------------------------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
</tr>
<tr>
<td>GOX + GOX + -&gt; GA + CO2</td>
<td>3.2562</td>
<td>0</td>
<td>4.0603</td>
</tr>
<tr>
<td>GA &lt;-&gt; 2PGA</td>
<td>3.2562</td>
<td>0</td>
<td>4.0603</td>
</tr>
<tr>
<td>MAL ACTIVE POOL %</td>
<td>73.9178</td>
<td>71.9027</td>
<td>74.6608</td>
</tr>
<tr>
<td>G1P ACTIVE POOL %</td>
<td>99.0018</td>
<td>98.9600</td>
<td>100</td>
</tr>
<tr>
<td>G3P ACTIVE POOL %</td>
<td>66.1728</td>
<td>64.9492</td>
<td>68.3012</td>
</tr>
<tr>
<td>RUBP ACTIVE POOL %</td>
<td>89.5316</td>
<td>89.2329</td>
<td>91.0861</td>
</tr>
<tr>
<td>ADPG ACTIVE POOL %</td>
<td>65.2371</td>
<td>63.9442</td>
<td>67.0693</td>
</tr>
<tr>
<td>2PG ACTIVE POOL %</td>
<td>40.6058</td>
<td>37.3491</td>
<td>NaN</td>
</tr>
<tr>
<td>S7P ACTIVE POOL %</td>
<td>97.6785</td>
<td>93.8407</td>
<td>96.558</td>
</tr>
<tr>
<td>CIT ACTIVE POOL %</td>
<td>47.2332</td>
<td>45.6201</td>
<td>48.3448</td>
</tr>
<tr>
<td>AKG ACTIVE POOL %</td>
<td>100.000</td>
<td>86.9059</td>
<td>100</td>
</tr>
<tr>
<td>DHAP ACTIVE POOL %</td>
<td>81.0967</td>
<td>79.8689</td>
<td>83.1862</td>
</tr>
<tr>
<td>BIOMASS</td>
<td>2.8365</td>
<td>2.8365</td>
<td>2.8365</td>
</tr>
</tbody>
</table>
12. Appendix 4: Supporting information for Chapter 4

Supporting Information: Metabolic engineering of cyanobacteria
For photoautotrophic production of heparosan,
a pharmaceutical precursor for heparin

UniProtKB - P74285 (CUGP_SYNY3)
Protein: UTP—glucose-1-phosphate uridylyltransferase
Gene: cugP
Organism: Synechocystis sp. (strain PCC 6803 / Kazusa)

>sp|P74285|CUGP_SYNY3 UTP--glucose-1-phosphate uridylyltransferase OS=Synechocystis sp. (strain PCC 6803 / Kazusa) GN=cugP PE=1 SV=1 MAFILIAAGKTVVRPITHTPKMIPILQKVMEFLLELQLQRGDFQIMVNVSHAEIIESYFRDGQRFQGVIASYFECNIVDSDLVGKALGSAAGGLKIQFNFVDFTVVMGCSDLIDDDTAVKLHREKGIA7IITKTVFCBLVSSYGVVTDDNGKILTFQFKEVSEALSTEINTGIYFEPVIDI1PSQQEYDLGLFLKVLDSGLPYAVNMDFEVD1GKVPDYWQAIRGVLREIKNVQ1PGIEVRPGYTTGINVAANWNEIEEGPVIYUGTMRIEDVGKIIIFSMIGFSCLCICQGAVVDNSVIFEVSRLPFGARLVKLVPGRYCVDKTGAAIDVQAAA
LDWLTDRHAAVQRQYPSQREISKLLQPE

**Figure S1:** cugP gene from PCC 6803 synthesizes enzyme UTP-glucose-1-phosphate uridylyltransferase possessing NTP_transferase domain. It displays 99% sequence similarity with mannose-1-phosphate guanylyltransferase from the same strain which is principally involved with synthesis of GDP-mannose (KEGG REACTION R00885).
**Figure S2:** Pair-wise sequence alignment of mannose-1-phosphate guanylyltransferases from PCC 6803 and PCC 7942 exhibited 79% sequence similarities with e-value of 0.0.
13. Appendix 5: Channeling in native microbial pathways: Implications and challenges for metabolic engineering

Doi: 10.1016/j.biotechadv.2017.06.004
Research review paper

Channeling in native microbial pathways: Implications and challenges for metabolic engineering

Mary H. Abernathy, Lian He, Yinjie J. Tang

Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO 63130, USA

ARTICLE INFO

Keywords: Compartmentalization Flux Glycolysis Isotopic labeling Thermodynamics Tunneling

ABSTRACT

Intercellular enzymes can be organized into a variety of assemblies, shuttling intermediates from one active site to the next. Eukaryotic compartmentalization within mitochondria and peroxisomes and substrate tunneling within multi-enzyme complexes have been well recognized. Intriguingly, the central pathways in prokaryotes may also form extensive channels, including the heavily branched glycolysis pathway. In vivo channeling through cascade enzymes is difficult to directly measure, but can be inferred from in vitro tests, reaction thermodynamics, transport/reaction modeling, analysis of molecular diffusion and protein interactions, or steady state/dynamic isotopic labeling. Channeling presents challenges but also opportunities for metabolic engineering applications. It rigidifies fluxes in native pathways by trapping or excluding metabolites for biocatalysis, causing substrate catabolite repression or inferior efficiencies in engineered pathways. Channeling is an overlooked regulatory mechanism used to control flux responses under environmental/genetic perturbations. The heterogeneous distribution of intracellular enzymes also confounds kinetic modeling and multiomics analyses. Understanding the scope and mechanisms of channeling in central pathways may improve our interpretation of robust fluxomic topology throughout metabolic networks and lead to better design and engineering of heterologous pathways.

1. Introduction

Metabolic flux is controlled by both enzyme kinetics and substrate diffusions. Under in vitro enzyme biochemistry, purified enzymes are often dissociated and metabolites are well-mixed; so enzymatic reactions have no spatial effect, and the homogenous system can be described by Michaelis–Menten models. However, macromolecular crowding in cells has been shown to promote enzyme associations and structural organizations within the cytoplasm (Miura et al., 2014). Meanwhile, fluidity in the cytoplasm has been found to be closely related to metabolic activity (Parry et al., 2014). Since in vivo environments may significantly impede enzyme kinetics (Minton, 2003), enzymes may organize to pass the product of one reaction site to an adjacent site without releasing into the bulk phase. Such a mechanism not only improves substrate diffusion and reaction-equilibrium, but also benefits other cellular functions, including protection of unstable intermediates, forestallment of substrate competition or inhibition among different pathways, and maintenance of a stable enzymatic microenvironment (Zhang, 2011).

Several mechanisms of enzyme complex channeling for cascade reactions have been recognized (Wheeldon et al., 2016). Intramolecular tunnels, as evidenced in tryptophan synthase, connect two active sites (Miles, 2001). Activated chemical swing arms can pass intermediates between active sites as seen in fatty acid synthase (type I) and pyruvate dehydrogenase complex. Moreover, electrostatic guidance uses complementary charges between protein residues spanning two active sites to direct intermediates as confirmed in malate dehydrogenase and citrate synthase (Wu and Minteer, 2015). Additionally, eukaryotic enzymes can compartmentalize enzymes in cellular organelles—e.g., the TCA cycle is carried out in the mitochondrion. On the other hand, evidence has shown that cascade enzymes in central and secondary pathways in prokaryotes, as well as eukaryotes, are also organized extensively (Table 1). For example, glycolytic enzymes may form glycolysosomes, which suggests intracellular enzymes are not randomly distributed in the cytosol. This hypothesis is not yet generally accepted because central metabolism enzymes often have high copy numbers and, metabolite diffusion is believed to be sufficiently fast (Wheeldon et al., 2016). In this review, we update evidences and scopes of natural channels in microbial metabolism as well as methods to measure metabolite channeling through cascade enzymes. We further discuss how channeling influences metabolic fluxes in central pathways. Validation and understanding of natural metabolite channeling through cascade
<table>
<thead>
<tr>
<th>Enzyme reactions</th>
<th>Detection method</th>
<th>System employed</th>
<th>Species studied</th>
<th>Citation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Polyfunctional enzyme forming molecular tunnels</td>
<td>Kinetic studies, enzyme cross-linking, LC-MS/MS, SDS-PAGE, and X-ray structural analysis</td>
<td>in vitro</td>
<td>Purified enzymes</td>
<td>(Chang et al., 2001; Alles, 2001; Gore, 2005)</td>
</tr>
<tr>
<td>Oxidative pentose phosphate pathway</td>
<td>Isotopic dilution of [3H]labeled GDP, intermediates by unlabeled intermediates</td>
<td>in vitro</td>
<td>Purified yeast enzymes, yeast extract, pea and soybean seed extracts</td>
<td>(Dobson et al., 1997)</td>
</tr>
<tr>
<td>Glycolysis to lactate</td>
<td>Isotopic dilution with [1-14C] glucose and [1-14C] maltose MPA</td>
<td>in vivo</td>
<td>Human Cell Line AGS-11</td>
<td>(Ullers et al., 2011)</td>
</tr>
<tr>
<td>TCA cycle</td>
<td>Isotopic dilution using [1-14C] malate and observed carbon orientation using [13-14C] MPA</td>
<td>in vivo</td>
<td>Saccharomyces cerevisiae</td>
<td>(Kang et al., 1990)</td>
</tr>
<tr>
<td>Calvin cycle</td>
<td>Isotopic dilution using [1-14C] CO2 and [14C] MPA</td>
<td>in vivo</td>
<td>Synchocystis sp. RCC 6803</td>
<td>(Ooi et al., 2011)</td>
</tr>
<tr>
<td>TCA cycle (from malate to citrate)</td>
<td>Isotopic dilution with challenging substrates</td>
<td>in vivo</td>
<td>Bacterial cell</td>
<td>(Vanderheide et al., 2000)</td>
</tr>
<tr>
<td>TCA cycle (from isocitrate to malate; malate to citrate)</td>
<td>Isotopic dilution, NMR (13C and 15N) and steady-state MPA</td>
<td>in vivo</td>
<td>Bacterial cell, enzymes from E. coli or Saccharomyces cerevisiae</td>
<td>(Kang et al., 2011)</td>
</tr>
<tr>
<td>TCA cycle (from isocitrate to oxaloacetate)</td>
<td>Isotopic labeling with [3-13C] glutamate and [3-13C] citrate; 13C NMR analysis</td>
<td>in vivo</td>
<td>Saccharomyces cerevisiae</td>
<td>(Williams et al., 2011)</td>
</tr>
<tr>
<td>Glycolysis (SFO to PEP)</td>
<td>Isotopic dilution using [4-13C] CO2 and [14C] MPA</td>
<td>in vivo</td>
<td>Saccharomyces cerevisiae</td>
<td>(Williams et al., 2011)</td>
</tr>
<tr>
<td>6PG to PEP to lactate</td>
<td>Isotopic dilution with [1,2-13C] glucose and [14C] MPA</td>
<td>in vivo</td>
<td>Bacterial cell</td>
<td>(Vanderheide et al., 2000)</td>
</tr>
<tr>
<td>Glycolysis</td>
<td>Protein analysis of isocitrate fractions and [1,2-13C] glucose studies</td>
<td>in vivo</td>
<td>Bacterial cell, enzymes from E. coli or Saccharomyces cerevisiae</td>
<td>(Williams et al., 2011)</td>
</tr>
<tr>
<td>Pyruvate degradation pathway</td>
<td>Spectral analysis compared between channeled complex and individual complexes</td>
<td>in vivo</td>
<td>Saccharomyces cerevisiae</td>
<td>(Williams et al., 2011)</td>
</tr>
<tr>
<td>GvA channeling through fatty acid synthase</td>
<td>Substrate pool quantification and competitive enzyme inhibition</td>
<td>in vivo</td>
<td>Purified chlorela from spinach and pea shoots</td>
<td>(Krey et al., 1997)</td>
</tr>
<tr>
<td>Fatty acid β-oxidation cycle</td>
<td>Protein crystallization, X-ray diffraction, and homology modeling</td>
<td>in vivo</td>
<td>Bacterial cell</td>
<td>(Kang et al., 2011)</td>
</tr>
<tr>
<td>Polyketide pathway</td>
<td>Kinetic analysis of modular enzymes</td>
<td>in vivo</td>
<td>Saccharomyces cerevisiae</td>
<td>(Williams et al., 2011)</td>
</tr>
<tr>
<td>Electron transport chain complexes</td>
<td>Analysis of enzyme activities in the presence of electron donors and inhibitors along with genetically modified complexes</td>
<td>in vivo</td>
<td>Mitochondria from mouse fibroblasts (Kang et al., 2011)</td>
<td></td>
</tr>
<tr>
<td>Photosynthesis light harvesting complexes</td>
<td>Spectroscopic analysis of isolated protein complexes and genetically modified complexes</td>
<td>in vivo</td>
<td>Chlamydomonas reinhardtii</td>
<td>(Nishi et al., 2015)</td>
</tr>
</tbody>
</table>
enzymes will offer new guidelines for genome-to-phenome mapping, metabolic modeling, and biosystems engineering.

2. Kinetic, thermodynamic and biochemistry evidences on channeling

2.1. Kinetic and thermodynamic implications

A cell's cytoplasm is very crowded with macromolecules constituting up to 40% of the matrix volume, which significantly increases viscosity and diffusional heterogeneity (Zhang, 2011). Consequently, the formation of enzyme associations for metabolite channeling has been hypothesized to overcome many barriers in metabolic pathways (Atkinson, 1977). First, low water concentration strongly inhibits a reaction's equilibrium and rate. If insufficient free water exists in the cell to support optimum metabolic reaction rates (Ellis, 2001), cells may link enzymes to channel metabolites directly through the pathway. Secondly, concentrations of key intracellular metabolites can be smaller than the Michaelis constant values of the catalyzing enzymes (Fendt et al., 2016), and these metabolite nodes may limit pathway fluxes (Bennett et al., 2009). Channeling improves enzyme conversion of low-concentration substrates while not inhibiting reaction rates. Lastly, many pathways have thermodynamic bottlenecks. For example, the oxidation of malate to oxaloacetate in the TCA cycle has a large positive ΔG°, but the reaction becomes thermodynamically favorable if malate → oxaloacetate → citrate can be treated as one step (Noor et al., 2014). Therefore, channeling is an important mechanism to overcome kinetic and thermodynamic obstacles in central metabolism.

2.2. Compartamentalization and enzyme tunneling

Compartamentalization of multiple enzymes and substrate tunneling within a multi-functional protein complex is well-studied. Cellular compartamentalization is a common characteristic of eukaryotes. Their organelles encapsulate enzymes or DNA by membranes. Compartments of enzyme complexes can be mitochondria, peroxisomes, chloroplasts, the endoplasmic reticulum, and the Golgi apparatus, which has been extensively explored (Bhatnagar and Feng, 2017). Moreover, polyfunctional enzyme subunits may contain caveolae for intermediate conversions. X-ray 3D structural analyses provide physical evidence of direct enzymatic conversion among active sites without diffusion. Tryptophan synthase was studied by X-ray structural analyses, which revealed the three-dimensional arrangement of the enzyme subunits to facilitate intramolecular tunneling of the intermediate indole (Huang et al., 2001). Transient-state kinetic approaches confirmed direct formation of L-tryptophan from indole-3-glycerol phosphate without releasing indole into the bulk phase. Similarly, RNA-dependent amidotransferase, 4-hydroxy-2-ketovalerate aldolase/acylating acetaldehyde dehydrogenase, cytidine triphosphate synthetase, pyruvate dehydrogenase, carbamoyl phosphate synthetase, glutamine phosphoribosylpyrophosphate amidotransferase, and asparagin synthetase contain multiple active sites (Huang et al., 2001; Weeks et al., 2006). These active sites are spatially close (ranging from −25 Å to −100 Å in length), which diminishes transit time and protects reactive intermediates from contacting with the external medium. Tunneling can also sequester reactive intermediates from hostile external environments. For example, acetyl-CoA synthase/carbox monooxide dehydrogenase from Maoredia thermoautotropha forms a molecular tunnel with a hydrophobic cavity of 138 Å, which controls the Wood-l-jungdahls pathway and reduces CO toxicity (Weeks et al., 2006). In addition to tunneling, other mechanisms may facilitate substrate transport between enzyme reaction sites. For instance, multiple subunits within fatty acid synthase (type I) pass intermediates for growing fatty acid chains to various catalytic sites via protein-protein interactions (Mashek et al., 2007). Similarly, pyruvate dehydrogenase complexes employ chemical swing arms to pass intermediates between dehydrogenase and acetyltransferase to transform pyruvate into acetyl-CoA (Pearson, 2002).

2.3. In vitro experiments on channeling of cascade enzymes

In vitro analyses of enzyme complexes and interactions have been performed by cross-linking/LC-MS, transient measurement of enzyme reaction rates, isotope labeling, kinetic modeling, SDS-PAGE, and detection of cascade resistance to a competing side reaction or a reaction inhibitor (Wheeldon et al., 2016). Recently, recombinant citrate synthase (CS), mitochondrial-malate dehydrogenase (mMDH), andaconitase were found to adopt a metabolon structure (Bulhogy et al., 2016). Site-directed mutagenesis of conserved arginine residues located in the positively charged electrostatic channel connecting mMDH and CS active sites led to impaired biocconversion of malate. This explications why high TCA cycle fluxes can operate in a viscous mitochondrial matrix. Additionally, a three-enzyme system (malate dehydrogenase, citrate synthase, and lactate dehydrogenase) was used to mimic the in vitro TCA pathway malate → oxaloacetate → citrate and found enzyme co-immobilization increased citrate production rates compared to the soluble system (Serre et al., 1973). Some purified enzymes can naturally aggregate to form stable channels under in vitro conditions (even in the presence of salt and detergent stresses). For example, the Kohl group studied tissue extracts and purified enzymes from the pentose phosphate pathway (PPP) (from yeast and plant cells), which channeled [1-14C] glucose into 14CO2 without influence by the addition of unlabelled intermediates, such as glucose 6-phosphate (G6P), 6-phosphogluconolactone, or 6-phosphogluconate (Dehse et al., 1997). They also found a lesser degree of channeling of G6P than 6-P-glucose. This may reflect the fact that G6P has more than one metabolic fate (sharing between PPP and glycolysis).

3. In vivo channeling studies

3.1. Approaches to reveal in vivo pathway channeling

Intracellular enzymes along central pathways may co-localize to form channels (Table 1). For example, the Embden-Meyerhof-Parnas pathway (EMP pathway) was thought to contain glycolysis enzymes (Shenoy et al., 2006). This opinion is still highly controversial because it is difficult to investigate enzyme interactions intracellularly. For in vivo studies, three common approaches are used. First, genetic engineering can be used to manipulate channelled pathways or disrupt enzyme interactions (Fig. 1c.). For example, to investigate channeling of TCA cycle enzymes in mitochondria, citrate synthase and malate dehydrogenase in Sacccharomyces cerevisiae were mutated (Vijot and Serre, 2000). The mutation in citrate synthase did not change its kinetic efficiency, but caused deficiency in the TCA cycle. Such an inhibitory phenotype could be overcome by malate dehydrogenase overexpression. This study inferred that enzyme interactions are essential for the TCA cycle functioning. Secondly, channeling can be characterized by the measurement of molecule movement through channels or the identification of co-localized enzymes. In vivo diffusivities of small solutes and macromolecules can be observed through photo bleaching of fluorescently labeled macromolecules, microscopy and time-resolved anisotropy (Milka et al., 2010; Verkman, 2002). The observation of metabolite turnover without aqueous-diffusive barriers in crowded cellular environments infers metabolite shuttling from one enzyme to an adjacent enzyme. Third, 13C or 15C experiments are used to detect elusive channeling phenomena because channeling occurrence may create a redistribution of C labeling (Niklas et al., 2011). In the TCA cycle of S. cerevisiae, the 13C fractions of the second and third carbon in oxaloacetate from [4-13C] 3-ketoglutarate were not randomized, which reveals possible channeling of succinate (a symmetric metabolite) in the enzyme reactions of α-ketoglutarate → succinate → fumarate (Buttigeg et al., 1993). In another example, when 13C position labeled glucose is
used as a substrate, the channeled EMP could bypass unlabeled tritiated phosphate input from the PPP and the Entner-Doudoroff pathway (EDP), altering the isotopomer composition of pyruvate and alanine (Tang et al., 2012). Therefore, $^{13}$C-metabolite flux analysis (MFA) can offer a noninvasive strategy for the detection of metabolite channeling. Based on the network structure and labeling strategies, $^{13}$C-MFA models may include the cytosolic pathway bypasses or subcellular compartmenations to characterize channeled flux (Williams et al., 2011). On the other hand, $^{13}$C-MFA often relies on the labeling of amino acids. Without extensive isotopomer information of pathway intermediates, $^{13}$C-MFA cannot provide conclusive evidence of channeling in central pathways.

3.2. Detection of leaky channels via dynamic labeling of intermediates

Intracellular pathways may form imperfect channels to release metabolites that connect to other pathways (Fig. 2). In pathways with leaky channels, intermediates appear in two separate compartments: a channeled pool (with high metabolite turnover rates) and a cytosolic pool (with much lower metabolite turnover rates). Under this assumption, dynamic labeling using $^{13}$C (or $^{14}$C) can identify these leaky nodes. For example, $^{13}$C analyses of phastauouric metabolism in Synchocystis 6803 found channeling of sugar phosphate pathways due to faster labeling in downstream products (FEP: phosphoenolpyruvic acid) than its upstream metabolite (3PG: 3-P-glycerate) right after the cells were fed with $^{13}$CO$_2$ (Young et al., 2011). Another dynamic labeling study saw possible metabolite channeling in Calvin cycle intermediates and introduced dilution factors into the flux analysis model to describe heterogeneous metabolite pools (Young et al., 2011). In other words, isotopically nonstationary experiments may reveal leaky metabolite nodes in a channel because its metabolically inactive cytosolic pool dilutes the bulk metabolite labeling rate. Although the observation of non-sequential labeling of cascade metabolites after the $^{13}$C-substrate pulse can be used as a “channeling signature,” such method has limitations. First, turnover rates of central metabolites could be so rapid in fast-growing microbes that labeling among cascade metabolites reaches steady state instantaneously after $^{13}$C-pulse (Link et al., 2013). Secondly, channeling signatures will not be observed if the pathway intermediates are perfectly channelled (the absence of leaked metabolite pools to change labeling dynamics).

Kohl et al. proposed an alternative approach based on the redistribution of isotopes in the presence of challenging metabolites (Shearer et al., 2005). In his study of EMP enzyme channeling, E. coli was engineered so that the mutants could uptake both $^{13}$C-glucose and unlabeled sugar phosphate intermediates (called challenging substrates). Based on different $^{13}$C enrichment rates of downstream metabolites with or without the presence of challenging substrates, they built a two-compartment model (channel pool vs. cytosol pool) to quantify the fraction of the total flux through the channel as seen in Fig. 1a. The authors found that the addition of unlabeled sugar phosphates in the labeled culture medium have different degrees of dilutions on the end product $^{14}$CO$_2$. They found that glycolysis intermediate fructose-bisphosphate (FBP) was perfectly channeled so that addition of unlabeled FBP did not affect $^{14}$CO$_2$ evolution from $^{14}$C-glucose. The challenging substrate G6P, on the other hand, diluted the $^{14}$CO$_2$ labeling by ~30%, while unlabeled fructose-6-P (F6P) diluted the $^{14}$CO$_2$ labeling by ~60%. Thus, G6P and F6P were both channeled to a lesser degree than FBP. This observation is consistent with the factors that G6P and
F6P are shared metabolites between the EMPP and PPP. Such leakage is essential for metabolic network connections. Moreover, fructose-bisphosphate aldolase (FBA) is a thermodynamic bottleneck in the EMPP (Flamholz et al., 2013), and the high degree of FBP channeling is essential to overcome this barrier.

3.3. “Squeeze and Pulse” to investigate pathways of high degree of channeling

The use of challenging substrates by Kohl group has two potential drawbacks. First, the cell cannot naturally uptake challenging substrates without engineering transporters, and secondly, the mutant may have catabolite repressions. As such, Hollinshead et al. (2016) designed a two-step approach to examine channeling without the use of challenging substrates. In the first step, genetic tools are used to downregulate a reaction step, creating a bottleneck in the pathway as seen in Fig. 1b. Such genetic modification aims to squeeze metabolites into the cytosol at the targeted node. This may decrease the pathway flux, making the measurements of metabolite turnover easier. The second step is to pulse a 14C-substrate into the cell culture, then time-course samples are collected to capture the labeling dynamics of the cascade metabolites. If a pathway is channelled, the labeling of the precursor metabolites will be slower than their product metabolites because a fraction of the precursor metabolites are squeezed outside of the channel to become less reactive. For example, a pF6 P. coli mutant was created to decrease the degree of channeling in the EMPP. Then 14C-pulse experiments found G6P/F6P had slower labeling rates than their product PEP (Hollinshead et al., 2016), confirming PEP was synthesized by the presumed glycolytic channel.

4. Modeling/theoretical investigation of pathway channeling

4.1. Channeling blocks flux towards thermodynamically favorable pathways

The EMPP is important for cell growth, thus many efforts have been proposed to bypass this pathway to increase catabolism. Fig. 3 indicates multiple thermodynamic bottlenecks in cascade enzymes (highlighted in Fig. 3a). Alternatively, the EDP requires less enzymatic proteins, and its cumulative $\Delta G^\circ$ is more favorable than the EMPP (Flamholz et al., 2013). However, E. coli prefers to use the EMPP rather than its native EDP even when the EDP is highly overexpressed (Hollinshead et al., 2015). This dilemma can be explained by the possible change of intrinsic Gibbs free energy by EMPP channeling. In a channelled pathway, the local concentrations of a reactant and a product can be altered, leading to a more negative $\Delta G^\circ$, which can drive a reaction forward. As illustrated by Fig. 3b, we present the $\Delta G$ of the pathway from glyceraldehyde-3-phosphate to 3-phosphoglycerate (enzymes ppd and ppg) as a function of concentrations of reactants and products. The $\Delta G$ of these two reactions in E. coli is close to zero (Park et al., 2016), represented as red circles in Fig. 3b. However, the value can be decreased exponentially with a higher concentration of reactant and a lower concentration of product. In addition, the $\Delta G$ of a reaction can be determined by forward and backward fluxes (Flamholz et al., 2012; Park et al., 2016). It is safe to assume that the forward flux is much higher than the backward flux in a channelled pathway, rendering $\Delta G$ a very negative number. Taken together, it is plausible that a channelled pathway is more thermodynamically favorable than a non-channelled pathway, altering pathway preference.

4.2. Channeling facilitates transport of metabolites between enzymes

Diffusion is assumed to be fast enough for reaction rates in small bacterial cells without the need for metabolite channeling (Woolfson et al., 2016). But others found that enzymes' active site orientation, size, distance (Baier et al., 2010), and electrostatic interactions (Tan et al., 2014) influence metabolite conversions. In this current study, a diffusion model examined the concentration gradient of a metabolite (c) between two cascade enzymes (A, B) with a distance $R$ (Fig. 4). If the enzymes are close to each other (within 20 nm), diffusion was fast enough to diminish concentration differences between A and B. As distance between enzymes A and B increases, the metabolite concentration at the location of enzyme B becomes more heavily influenced by the diffusion coefficient. Based on biochemistry experiments, most enzymes catalyze their reactions to a rate (average $k_{cat}/K_m \sim 10^9 M^{-1} s^{-1}$) that is 1000-10,000 times slower than the diffusion rate for collisions of proteins and metabolites (10^9 M^{-1} s^{-1}) (Bar-Even et al., 2011). However, such in vitro-measured kinetic parameters may not correlate with its in vivo enzyme activity. For example, in vitro reconstitution of the EMPP in a cell-free system observes protein aggregations and significant loss of kinetic activities at enzyme concentrations of a few micro-moles (Liu et al., 2017). In contrast, the bacterial cytoplasm is highly crowded with non-uniform macromolecules and “bound” water. At this high volume faction, intracellular suspensions exhibit properties that are characteristic of glass-forming liquids and provide a glauzy dynamic to the cytoplasm (Parr et al., 2014). Such a high viscosity and tortuosity causes transport barriers via the Stokes-Einstein law. If metabolite diffusion coefficients are reasonably reduced by 10 fold (Mika et al., 2016) through crowding effects, the metabolite concentration around the second
enzyme could be reduced by $85\%$, and thus limit flux towards enzyme B.

4.3. Diffusion and reaction rates along EMPF enzymes

An in silico study was conducted to relate enzyme copy number or concentration (enzyme concentration estimated as copy number $N_e \times$ cell volume (0.7 $\mu$m$^3$)) to diffusion and distance between entities. Relative concentration profiles (Fig. 4b) were determined based on relative time and positioning using the equation presented in Fig. 4a. The variables $x, c$, and $t$ are dimensionless after they are normalized to the distance between A and B (R), the initial concentration of A ($C_0$), and a characteristic time $R^2/D$, respectively. The resulting equation is irrelevant to $R$ and $D$, and here, we also assumed that the relative

$$
\Delta G = -RT \ln \left( \frac{J^f}{J} \right) = \Delta G^0 + RT \ln \left( \frac{\text{product}}{\text{reactant}} \right)
$$
concentration is one at $x = 0$ and zero elsewhere, and the concentration gradient is zero at $x = L$. To achieve 99% of the initial concentration near enzyme B, it was found to take roughly two relative time units. Fig. 4c shows the estimated time for one metabolite or protein to travel in a specific distance. A previous review suggests that 1 nM of molecules in cells corresponds to 1 µm average intermembrane distance, and that 100 nM corresponds to a distance of 0.25 µm (Grima and Schnell, 2008b). For enzymes involved in glycolysis, their copy numbers are generally high (between 10^6 and 10^7) (Voges et al., 2015), which gives a concentration of ~100 nM (between 24 nM and 240 nM) for each individual enzyme. Under such circumstances, the average time for a NBD-glucose molecule (fluorescently labeled deoxyglucose used to monitor glucose uptake) to travel from one enzyme to another is only 2.5 ms (Fig. 4c). As displayed in Fig. 4d with the dotted-line, 2.5 ms is smaller than reaction turnover time of most enzymes (except αm, δGL, ppm, pgf) in the glycolysis pathway (Flamholz et al., 2013), further suggesting diffusion poses few bottlenecks. However, if the distance between two enzymes increases 4 fold to 1 µm, the diffusion time increases to 40 ms, leading to increased diffusion barriers in the EMP (except αm). Further, if a pathway contains multiple steps, cumulative distance among all enzymes will be increased. For instance, the EMP contains 9 enzymatic steps from G6P to pyruvate. If the model assumes an average distance of 0.25 µm and a diffusion coefficient of 50 µm^2 s^-1, 156 ms is needed for conversion from G6P to pyruvate, and the time for diffusion accounts for 14.4% of the total time (the sum of diffusion time plus 1/∆ν of EMP enzymes). However, if the average distance among enzymes is shortened to 0.1 µm through compartmentalization, the time fraction for diffusion drops to 2.6%. This shows even a loose channel can promote fluxes through a long pathway.

4.4. Channeling alters enzyme-metabolite and enzyme-enzyme interactions

Proteins or macromolecules can be considered to be motionless when compared to free metabolites. For example, it may take 2 min for ribosome-loaded mRNAs to travel 1 µm, which is four orders of magnitude longer than a glucose molecule. This finding is consistent with the phenomenon of macromolecular crowding in cells (Ellis, 2001). If enzyme copy numbers are low, the average distance between cascade enzymes may be increased, and metabolite channeling is more likely to be beneficial to multiply the chances of enzyme-reactant collisions. For example, secondary metabolism in plants has been shown to form channelled pathways to improve low-concentration intermediates (Jandt et al., 2013; Jørgensen et al., 2005; Reze et al., 2011). In addition, metabolite channeling could also prevent promiscuous side reactions. Promiscuous enzymes (e.g., hydrolyase) are widely present in microbes as many cross-reactions are being discovered through gene knock outs of critical enzymes (Ólafsson and Tawfik, 2010). Metabolic channeling can help define the products of these promiscuous enzymes (Schub, 2003). While catalytic promiscuity may introduce unwanted side reactions, channeling may effectively repress side reactions by concentrating enzymes and their desired substrates within a pathway (Piedrafita et al., 2015). Therefore, another benefit for metabolite channeling is to improve specificity. Finally, intracellular proteins are constantly being broken down by diverse proteases (Oda, 2012). Proteolysis is important for metabolic regulation by deactivating enzymes, transcription factors, and receptors. In general, proteolysis can be highly promiscuous. To protect a wide range of essential enzymes, channeling may reduce the chance of these enzymes being freely exposed to non-specific intracellular proteases so that the cell only removes unnecessary or abnormal proteins. In eukaryotes, enzymes can protect themselves by concentrating in compartments; in prokaryotes, compartments are limited; consequently, the localization of enzymes could be a natural defense against enzyme proteolysis (Gottschalk, 1996; Meyer et al., 2014).

5. Implications of metabolic channeling

5.1. Challenges facing synthetic biology

In metabolic engineering applications, novel pathways with beneficial energy and carbon yields have been proposed to improve bio-productions. However, heterologous pathways typically have inferior efficiency relative to native pathways. Unlike native pathways, the engineered pathways are likely unable to overcome intracellular crowding effects, diffusion limitations, promiscuous side reactions or optimal regulation control. Channeling naturally increases the robustness of native pathways under environmental and genetic perturbations (Fischer et al., 2004; Fischer and Sauer, 2005) and may prevent native pathways from releasing intermediates as the precursors for engineered pathways. One of the challenges in metabolic engineering is glucose catabolite repression, which blocks E. coli utilization of ligandolose sugars. Disruption of the presumed glycolysis channel (such as knocking out the phosphotransferase system or a key gene in the EMP) decreases biomass growth but improves cell capability to co-utilize C5 sugars (Gonzalez and Antoniewicz, 2017; Hollinshead et al., 2016). This finding strongly implies that metabolite channeling may block EMP enzymes from co-utilizing intermediates produced from other pathways (e.g., xylose degradation pathway). Such catabolite repression may be reduced if the channelled flux is disrupted, allowing products of C5 sugars to diffuse from the cytosol into the pathway channel for metabolic conversion.

5.2. Applications of synthetic channel to benefit bio-productions

Metabolic engineers have attempted to increase the intrinsic inefficiency of metabolic reactions through enzyme localization and compartmentalization of metabolic pathways. Compartmentalization of pathways has been deployed in eukaryotes by tagging enzymes to specific cellular organelles (Avalos et al., 2013; Sheng et al., 2016). This approach offers benefits such as inhibition of competing reactions, pathway stabilization, substrate concentration, and reduction of toxic byproducts (Plattman and Feng, 2017). It also represents a starting point for innovative bacteria engineering tactics such as targeting bacterial microcompartment like carboxysomes or heterologously expressing micromcompartment (Chen and Silver, 2012; Parsons et al., 2010; Wundenhofer et al., 2011).

Another channeling strategy involves the engineering of protein interactions or direct protein fusion of interacting partners (i.e., protein or DNA/RNA scaffolds). Scaffolds reduce diffusion time, mitigate toxic metabolite inhibition and make heterologous enzymes more stable (Baek et al., 2013; Moon et al., 2010; Sachdeva et al., 2014; Wang and Yu, 2012). For example, co-localization of three mevalonate biosynthetic enzymes via a protein scaffold achieved improved product titers even with low enzyme expression (Gaeber et al., 2009; Wang and Yu, 2012). DNA scaffolds with distinct protein-docking sites have also been designed to offer precise enzyme spatial organization. Thus, in vivo RNA assembles can be used to engineer biological pathways with multidimensional architectures, such as hydrogen production (Delezene et al., 2011).

A particularly interesting area of study that benefits from channeling strategies is flavonoid production in engineered microorganisms. It is well established that flavonoid production in plants is facilitated by the presence of metabolons (Jørgensen et al., 2005). However, flavonoid isolation from naturally-producing plant material is difficult, making heterologous flavonoid production in recombinant organisms like E. coli attractive. Nonetheless, cofactor and substrate supply tend to be the limiting steps in flavonoid bio-production in E. coli. For heterologous flavonoid production in recombinant microorganisms, pathway
modulation and enzyme co-localization have been proved as valid strategies. Catechin flavonoid production in E. coli was enhanced by using combinatorial protein scaffolds with catechin biosynthesis (Zhan et al., 2015). To simulate the multi-enzyme complex involved in flavonoid production in plants, a translational protein fusion between two pathway enzymes was used to increase cyanidin 3-O-glucoside formation over tandem transcriptional expression of the two enzymes (Yao et al., 2009).

Heterologous expression of microcompartments and co-localization of enzymes are still difficult to employ widely in the metabolic engineering field. Synthetic channels place metabolic burdens on the host and may also affect protein folding. The formation of capsids can impair cell growth and productivity. Moreover, microcompartment and metabolon constructions are constrained by global cell regulations (Chen and Silver, 2012). Current synthetic channeling studies focus mainly on cell productivity without knowledge of the structure, mechanism, or kinetics of engineered channels to vary non-linearly with enzyme or model productivity is from metabolite shuffling, enzyme stability, or elimination of metabolite inhibition. To produce heterologous pathways as efficient as native pathways, we need more efforts to understand the delicate mechanisms of natural channels in central pathways.

5.3. The impact of channeling on genome-phenome mapping

There have been many studies on the mechanisms regulating metabolic fluxes in microbial systems. Most research focuses on transcriptional regulations under diverse conditions. Protein and metabolite interactions (allosteric regulation) were also proposed to enable instantaneous metabolic response and pathway modulation to maintain metabolic homeostasis (Link et al., 2013). Computational studies have been carried out to relate fluxes to data from the metabolome and proteome (Liebermeister and Klipp, 2006; Noor et al., 2016). Most models assume that metabolic fluxes are determined by metabolite concentrations and protein abundance. Intuitively, gene/protein expressions and metabolite concentrations should have strong influences over their functional outputs—i.e., fluxes. However, much evidence has shown that neither transcription/enzyme levels nor allosteric regulation are sufficient to explain the flux variations in response to various growth conditions (Blauk et al., 2005; Chubukov et al., 2013; Hackett et al., 2016). Even among different heterotrophic microbial species, the substrate catabolic fluxes in different bacterial families or their mutants show much smaller variations than their genome or transcriptional differences (Tang et al., 2009; Wu et al., 2016). This leaves unresolved the question of what are the hidden forces that maintain robustness and stability of flux ratios under genetic or environmental perturbations (Fischer and Sauer, 2005; He et al., 2017). Recent E. coli studies show surprisingly simple regulatory correlations, where only a few global transcription factors and intracellular metabolites (e.g., cyclic adenosine monophosphate, fructose-bisphosphate, and fructose-1-phosphate) appear to coordinate microbial central metabolism (Kochanowski et al., 2017). This leaves a strong hypothesis that native bacterial pathways form well-organized channels extensively to direct metabolic fluxes via several key regulation nodes. A wide distribution of pathway channels is beneficial for cell metabolism to minimize enzyme synthesis costs and to remember the optimal fluxome to organize nutrient influxes throughout metabolic network. From this point of view, channeling can cause flux divergence and absence of diffusion limitations. Additionally, 13C-MFA assumes that scrambling reactions net and exchange fluxes (such as succinate to fumarate) are symmetric (Wiechert et al., 2001). For a complex pathway, channeling can provide alternative isotopic labeling routes. For example, if [3-13C]propionate is used as a substrate, malate will be symmetrically labeled in the absence of metabolic channeling. In reality, malate is asymmetrically labeled in Saccharomyces cerevisiae, revealing the features of the channeling in TCA cycle enzymes (Suemegi et al., 1999). Therefore, precise 13C-MFA must consider the channeling influences on molecule orientation and pathway preferences (e.g., addition of alternate flux channels in the MFA model). Such 13C-MFA found that the assumption of channeling bypasses improved the MFA model fit and provided a more accurate description of the metabolic network (Williams et al., 2011). However, the use of 13C-MFA to investigate pathway channeling needs precise measurements of labeling positions of key metabolites using both MS and NMR as well as goodness of fit by realistic models. It may fail in highly connected networks or autotrophic metabolism if model or measurement resolution is not sufficient.

On the other hand, isotopically nonstationary flux analysis is a new approach to profile fluxes by tracing isotopic dynamics along cascade metabolites. Metabolite labeling patterns can be confounded by the presence of metabolically inactive pools or uncertain metabolic fluxes. To account for such phenomena, dilution parameters are used to describe a lack of isotopic equilibrium due to the existence of two or more metabolite pools (Antoniwiak et al., 2007; Young et al., 2011). Dilution factors improve the goodness of fit, but introduce additional degrees of freedom into model. This may increase undetermined parts of the metabolic network (van Winden et al., 2001). Dilution parameters, as channeling factors, provide quantifications of concentration heterogeneity of intracellular metabolites.

6. Future research on channeling

Channeling improves bioconversion rates, stabilizes low copy number enzymes, and reduces side products from promiscuous enzymes. It has been hypothesized to control fluxes through metabolic pathways and facilitate cascade reactions in a crowding environment. Channeling is a natural evolution outcome; yet despite experimental evidence, there are many unresolved questions. First, heterogeneous metabolite pools can change reaction thermodynamics and reaction paths. Kinetic models or flux analysis cannot precisely describe such process. Therefore, superior modeling and thermodynamic approaches need to be developed. Secondly, we still have little knowledge about how cells channel their native pathways. Recent research has found that heterologous gene expression is influenced by its location on the genome (Engelsander et al., 2017); however the explanation for how this occurs is still up for debate. During pathway engineering, over-expression of a gene cluster is in general more efficient than expressing multiple enzymes from different sources. Although it is common for a gene cluster to transcript/translate a biosynthesis pathway, genes of cascade enzymes in central metabolism are often not located next to each other but still show channeling phenomena. Thus, it is unclear whether gene locations, 3-D DNA structures, DNA intracellular locations, or other factors (e.g., intracellular pH, ion strength, or electrostatic interactions) coordinate locations of enzymes to form channels. To test this hypothesis, genetic engineering can disrupt the presumed channels via plasmid or chromosome relocations (i.e., after deletion of specific genes or at a non-native locus on the chromosome). Moreover, structure-guided enzyme mutations can be performed to weaken enzyme–enzyme interactions and cross-referencing to known protein–protein interactions. Chemical cross-linking combined with mass spectrometry may examine protein assemblies and reveal both the structure and the interactions of cascade enzymes (Liu et al., 2015; Tan et al., 2016).

In addition, genome scale modeling can assess channeling using experimental 13C-MFA and transcriptomic data. For example, predicted intracellular fluxes from genome scale modeling can be used as a
Synaptic tagging. Philos Trans 364, e16278.

Harriman, E.M., Feng, G., 2017. Compartimentalized metabolic engineering for bio-


14. Appendix 6: Supporting information for Chapter 6

Supporting Information: Comparative studies of glycolytic pathways and channeling under \textit{in vitro} and \textit{in vivo} modes

Figure S1. Total $^{13}$C Enrichment (fractional abundance) as a function of time following a $^{13}$C$_6$ glucose pulse in wild-type \textit{E. coli} K-12 and $\Delta$ptsG and $\Delta$ptsG + ED mutants. Total $^{13}$C Enrichment is calculated as $1/N \times \sum M_i \times i$, where \( N \) is the number of carbons in the metabolite and $M_i$ is the fractional abundance of the $i$th isotopomer.
Figure S2. Fractional abundance of mass isotopomers M+0 – M+5 for PEP, CIT, aKG, and MAL (30 seconds after U-\(^{13}\)C glucose pulse in wild-type E. coli K-12 and ΔptsG and ΔptsG+ED mutants). Predicted labeling patterns are predicted where L in AcCoA represents the labeled or active pool of AcCoA, U in AcCoA represents the unlabeled or inactive pool of AcCoA, and A represents the labeled OAA pool from anapleuortic pathways. Important enzymes are listed by gene names in the pathway in purple.
Figure S3. NAD(P)H formation as a measure of absorbance at 340 nm with EMP enzymes (100 nM) in various crowding agents.
Table S1. *E. coli* strains used in this study.

<table>
<thead>
<tr>
<th>Strains</th>
<th>Description</th>
<th>JBEI ICE code</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW25113 WT</td>
<td>Keio collection WT (ΔlacZ4787::rrnB3 hsdR514ΔaraD-B567ΔrhaD-B568 rph-1)</td>
<td>-</td>
<td>(Baba et al., 2006)</td>
</tr>
<tr>
<td>WH03</td>
<td>BW25113 + pED (derived from pBbE5c-YFP with gene replaced by <em>edd</em> and <em>eda</em>)</td>
<td>JBEI-11465</td>
<td>(Hollinshead et al., 2016)</td>
</tr>
<tr>
<td>ΔptsG</td>
<td>Keio collection strain JW1087-2 (ΔlacZ4787::rrnB3 hsdR514ΔaraD-B567ΔrhaD-B568 rph-1ΔptsG763::kan)</td>
<td>-</td>
<td>(Baba et al., 2006)</td>
</tr>
</tbody>
</table>

| WH05    | BW25113 ΔptsG + pED | -              | This study              |

Table S2. Purified proteins used in this study.

<table>
<thead>
<tr>
<th>Name</th>
<th>Description</th>
<th>Source organism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Glk</td>
<td>Glucokinase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>Pgi</td>
<td>Phosphoglucone isomerase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>PfkA</td>
<td>6-phosphofructokinase I</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>PfkB</td>
<td>6-phosphofructokinase II</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>FbaA</td>
<td>Fructose-bisphosphate aldolase class II</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>FbaB</td>
<td>Fructose-bisphosphate aldolase class I</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>TpiA</td>
<td>Triosephosphate isomerase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>Gene</td>
<td>Enzyme Description</td>
<td>Organism</td>
<td>Reference</td>
</tr>
<tr>
<td>------</td>
<td>--------------------</td>
<td>----------</td>
<td>-----------</td>
</tr>
<tr>
<td>GapA</td>
<td>NAD⁺-dependent glyceraldehyde 3-phosphate dehydrogenase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>Pgk</td>
<td>Phosphoglycerate kinase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>GpmA</td>
<td>Phosphoglycerate mutase I</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>GpmM</td>
<td>Phosphoglycerate mutase II</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>Eno</td>
<td>Enolase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>PykF</td>
<td>Pyruvate kinase I</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>PykA</td>
<td>Pyruvate kinase II</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>Zwf</td>
<td>Glucose-6-phosphate-1-dehydrogenase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>Pgl</td>
<td>6-phosphogluconolactonase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>(Liu et al., 2017)</td>
</tr>
<tr>
<td>Edd</td>
<td>6-phosphogluconate dehydratase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>This study</td>
</tr>
<tr>
<td>Eda</td>
<td>2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase</td>
<td><em>Escherichia coli</em> BL21</td>
<td>This study</td>
</tr>
</tbody>
</table>
15. Appendix 7: Comparing studies of glycolytic pathways and channeling under *in vitro* and *in vivo* modes

Doi: 10.1002/aic.16367
Comparative Studies of Glycolytic Pathways and Channeling Under
in Vitro and in Vivo Modes

Mary H. Ahearnathy
Dept. of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO, 63130

Yuchen Zhang
Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education and Wuhan University School of
Pharmaceutical Sciences, Wuhan, 430071, China

Whitney D. Hollinshead
Dept. of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO, 63130

George Wang and Edward E. K. Baidoo
Lawrence Berkeley National Laboratory, Emeryville, CA, 64608

Tiangang Liu†
Key Laboratory of Combinatorial Biosynthesis and Drug Discovery, Ministry of Education and Wuhan University School of
Pharmaceutical Sciences, Wuhan, 430071, China

Yinjie J. Tang†
Dept. of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO, 63130

Published online in Wiley Online Library (wileyonlinelibrary.com)

DOI 10.1002/aie.16367

Keywords: substrate channeling, ED pathway, EMP pathway, isotopic labeling, phosphotransferase system

Introduction

Synthetic biology has made great progress in establishing molecular tools for genome editing and controlling gene expression. However, redirecting metabolic fluxes for effective biosynthesis remains difficult because hosts often naturally suppress heterologous pathways and have undesired metabolic shifts. Metabolic fluxes do not always directly correlate with metabolite concentrations or gene/enzyme expressions. In reality, cell metabolism is highly regulated and coordinated through multiple interactions between genes, enzymes, and metabolites.

Additional Supporting Information may be found in the online version of this article.
M. H. A. and Y. Z. contributed equally to manuscript as co-authors

Correspondence concerning this article should be addressed to T. Liu, and Y. J. Tang at liu@whu.edu.cn and yinjie.tang@wustl.edu

© 2018 American Institute of Chemical Engineers

Figure 1. Substrate level regulations modulate the catalytic activity of enzymes via metabolite-protein interactions, namely allosteric inhibition. Additionally, metabolite-protein interactions are highly conserved across evolutionarily divergent species for the maintenance of metabolic homeostasis, and this regulation is responsible for fast metabolic dynamics under changing growth environments. Additionally, gene-protein interactions affect metabolic fluxes by regulating enzyme expressions. Transcriptional regulation is a commonly employed mechanism used to orchestrate metabolic responses. Less experimentally explored, enzymes may dynamically associate via pairwise interactions to form “pearl necklace like” channels or via nonpairwise interactions to form enzyme clusters. Independent of the specific channeling architecture, these channels function to direct metabolic fluxes in crowded cytosolic environments. Such protein-protein interactions (proximity channels) are linked via noncovalent forces rather than membrane encapsulation or covalent bonds. Chemotactic movement has been evidenced as the...
driving force for cascade enzymes to assemble in the presence of an initial substrate. Unlike transcription or protein expressions that often show significant variations, channeling might stabilize fluxes under environmental/genetic perturbation.\textsuperscript{10,11} Because small metabolite diffusion is considered sufficiently fast for enzyme kinetics,\textsuperscript{6} channeling may increase reaction rates, but it is more likely channeling functions as a regulatory role in controlling metabolic fluxes and protecting reaction intermediates from competing reactions or harsh cytosol environments.\textsuperscript{5} Although it has been recognized that enzymes can organize spatially and temporarily to shuttle intermediates without releasing them into the cytosol, it is difficult to validate weak transient interactions between enzymes under in vitro (cell free) conditions,\textsuperscript{3} thus proximity channeling in bacteria remains controversial.\textsuperscript{12}

In light of diverse regulatory theories, this study reexamined \textit{Escherichia coli} glycolytic pathways: the Embden-Meyerhof-Parnas (EMP) pathway, the pentose phosphate (PP) pathway and the Entner-Doudoroff (ED) pathway. Theoretically, the ED pathway would increase glycolytic flux in microbial hosts for desired products\textsuperscript{13-15}; thus, mutation efforts are employed to improve both native ED pathway usage\textsuperscript{14} and glucose catabolism in \textit{E.coli}.\textsuperscript{15,16} However, the ED pathway in \textit{E.coli} is always inferior to the EMP pathway despite its favorable thermodynamics, less enzymatic steps, and comparable enzyme activities.\textsuperscript{17,18} Studies on pathway regulations in bacteria often focus on reaction thermodynamics, enzyme kinetics, allosteric inhibitions, or enzyme expressions, while protein-protein interactions remain unfamiliar. To investigate flux controls among competing pathways for glucose catabolism, this study designed both in vitro and in vivo systems in conjunction with reported channeling theories and isotopic labeling analysis.\textsuperscript{19,20} Via cross-talk and comparisons among enzyme expressions and turnover of cascade metabolites under in vivo vs. well-mixed in vitro conditions, we aim to obtain new insights into glycolytic regulations. Moreover, the phosphotransferase system (PTS) is the main glucose transporter in \textit{E.coli}\textsuperscript{21} and plays a key role in phosphorylating glucose to glucose-6-phosphate (G6P), which is the branching point for the three glycolytic pathways\textsuperscript{22} (Figure 2). According to chemotactic assembly theory, the PTS would be a starting point of EMP channeling.\textsuperscript{21} Thus, by using PTS mutants, this study revisits the flux re-distributions at a major flux control point in \textit{E.coli}. The outcomes support recent views about in vivo flux regulations via channeling.\textsuperscript{5}

Materials and Methods

Strain construction

Wild type (WT) \textit{E.coli} (K-12 BW25113) and its mutant strains are listed in Supporting Information Table S1. The plasmid pED was derived from plasmid pBBR5c-YFP (50-70 copies per cell) with the YFP gene replaced with \textit{E. coli} \textit{edd} and \textit{eda} genes, as reported in our previous article.\textsuperscript{14} Briefly, the strain (\textit{Δ}ptsG + pED) was constructed as follows: competent cells were made from the \textit{Δ}ptsG mutant using chloramphenicol and storage solution. Heat shock transformation was used to transform the \textit{Δ}ptsG strain with the purified plasmid pED. The resulting strain was plated on kanamycin (20 mg/mL) and chloramphenicol (30 mg/mL) Luria-Bertani (LB) plates to isolate the transformed cells.

Media composition and growth experiments

All experiments were started with LB seed cultures with appropriate antibiotic. Next minimal M9 medium consisting of 11 g/L of 5x M9 salts (Sigma Aldrich, St. Louis, Missouri), 2 mM MgSO\textsubscript{4}, 7H\textsubscript{2}O, 0.1 mM CaCl\textsubscript{2}, 2H\textsubscript{2}O, 1 mg/L of thiamine hydrochloric acid, 3 mg/L of FeSO\textsubscript{4}, 7H\textsubscript{2}O, and trace minerals was inoculated (1% v/v) with 10 g/L glucose. All cultivations were aerobic 5 mL cultures on a rotary shaker at 250 rpm and 37°C. Mutant cultures were supplied with 20 μg/mL kanamycin and 30 mg/mL chloramphenicol as needed. To induce plasmid genes, 0.05 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) was added at early exponential phases.

![Figure 2. Glycolytic pathway map in E. coli.](image-url)
Figure 3. In vitro reaction rates of the EMPP and EDP.

(a). Reaction rates with protein concentrations at 1000 nM. Reactions were assumed to be first-order processes and the rate constants, k (1/h), were calculated based on absorbance curves and minimizing the sum of squared residuals. (b). Reaction rates with varied enzyme concentrations. [Color figure can be viewed at wileyonlinelibrary.com]

Steady-state labeling experiments and GC–MS analysis

E. coli cultures were inoculated in fresh M9 minimal medium (0.5% v/v) and were grown with 10 g/L 1-13C glucose (Sigma Aldrich) (n = 2) until exponential phase. Amino acid extraction and GC–MS analysis were performed as described previously during exponential growth phase.21 The fragments [M-15]7-9 or [M-57]7-9 and [M-159]7-9 or [M-85]7-9 were used for isotopic tracing. Using a previously reported algorithm, the natural isotopic abundance was corrected for the derivatized amino acids.20 Based on mass isotopomer distribution (MID) data from labeled amino acids, flux ratios among glycolytic pathways (Figure 4) were computed using software WUFlux.21 Standard deviations were calculated based on the measurement errors of MID data.

Isotopically nonstationary labeling experiments and pool size analysis

E. coli WT and mutant cultures were inoculated from overnight M9 cultures in fresh M9 minimal medium with 2 g/L glucose and allowed to grow for approximately 4 h. Cultures in exponential growth phase were pulsed with 13C6 glucose for a final concentration of 4 g/L. To measure the 13C-
Table 1. Glycolysis Genes and Reported Enzyme Kinetics. Gene location in E. coli (taxid:679895) was determined through NCBI nucleotide blast.

<table>
<thead>
<tr>
<th>Pathway</th>
<th>Reaction</th>
<th>Gene</th>
<th>Gene location</th>
<th>In vivo conc. (nM)</th>
<th>ΔG (standard condition) (kJ/mol)</th>
<th>ΔG (actual, kJ/mol)</th>
<th>kcat (1/s)</th>
<th>Km (μM)</th>
<th>Measured enzyme activity (U/mg)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper glycolysis</td>
<td>Glc &gt; G6P</td>
<td>G6P</td>
<td>2,501,904-2,502,905</td>
<td>169</td>
<td>2.5</td>
<td>-0.3</td>
<td>120</td>
<td>0.28</td>
<td>NM</td>
</tr>
<tr>
<td></td>
<td>G6P &gt; F6P</td>
<td>F6P</td>
<td>4,223,686-4,225,335</td>
<td>126</td>
<td>-17.2</td>
<td>-19.8</td>
<td>167</td>
<td>0.16</td>
<td>0.71</td>
</tr>
<tr>
<td></td>
<td>F6P &gt; FBP</td>
<td>FBP</td>
<td>4,075,600-4,099,442</td>
<td>37</td>
<td>1.80</td>
<td>1.81</td>
<td>1,054</td>
<td>0.86</td>
<td></td>
</tr>
<tr>
<td></td>
<td>FBP &gt; GAP + DHAP</td>
<td>GAP</td>
<td>3,063,524-3,064,603</td>
<td>767</td>
<td>21.5</td>
<td>-2.8</td>
<td>10,5</td>
<td>0.17</td>
<td>1.97</td>
</tr>
<tr>
<td></td>
<td></td>
<td>DHAP</td>
<td>4,100,668-4,101,435</td>
<td>768</td>
<td>5.5</td>
<td>-0.9</td>
<td>500</td>
<td>1</td>
<td>NM</td>
</tr>
<tr>
<td></td>
<td>F6P &gt; 6PG</td>
<td>6PG</td>
<td>1,929,096-1,930,571</td>
<td>113</td>
<td>-5</td>
<td>0.0</td>
<td>174</td>
<td>0.174</td>
<td>7.11</td>
</tr>
<tr>
<td></td>
<td>ED only</td>
<td>ED</td>
<td>794,062-795,037</td>
<td>738</td>
<td>-23.2</td>
<td>-31.2</td>
<td>190</td>
<td>0.02</td>
<td>NM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>KDPG</td>
<td>1,927,050-1,928,861</td>
<td>17</td>
<td>-43.1</td>
<td>-46.4</td>
<td>25</td>
<td>0.04</td>
<td>NM</td>
</tr>
<tr>
<td>Lower glycolysis</td>
<td>GAP &gt; BPG</td>
<td>GAP</td>
<td>1,926,372-1,927,013</td>
<td>ND</td>
<td>16.3</td>
<td>-0.8</td>
<td>83</td>
<td>0.1</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BPG</td>
<td>3,064,818-3,065,981</td>
<td>564</td>
<td>4.2</td>
<td>-0.4</td>
<td>654</td>
<td>0.0018</td>
<td>NM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3PGA</td>
<td>782,299-783,051</td>
<td>760</td>
<td>3.9</td>
<td>-0.4</td>
<td>530</td>
<td>0.65</td>
<td>NM</td>
</tr>
<tr>
<td></td>
<td>3PGA &gt; 2PGA</td>
<td>2PGA</td>
<td>3,778,620-3,780,164</td>
<td>567</td>
<td>4.2</td>
<td>-3.4</td>
<td>78</td>
<td>0.3</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td>2PGA &gt; PEP</td>
<td>PEP</td>
<td>2,900,002-2,901,300</td>
<td>119</td>
<td>-26.3</td>
<td>-11.8</td>
<td>232</td>
<td>0.3</td>
<td>0.17</td>
</tr>
</tbody>
</table>

The standard condition for ΔG: pH = 7.2, ionic strength = 0.25 M, metabolite concentrations = 1 M.

1Reference 44.
2The actual ΔG was calculated by eQuilibrate using reported metabolite concentrations.
3Reference 17.
4ND, no data available; NM, not measured.
5Source organism is E. coli unless noted.

incorporation into metabolites over time, cultures were quenched at different time points (from 5 s to 30 s) using ice-cold media and a liquid nitrogen bath. All time points were completed in duplicates. Then biomass was harvested through centrifugation, and the pelleted samples were extracted with a chloroform-methanol method before LC-MS analysis.20 The labeling of central metabolites were quantified at Joint BioEnergy Institute (Berkeley, CA), using hydrophilic interaction liquid chromatography (HILIC) coupled to electrospray time-of-flight MS, as described in our recent article.13 For metabolite pool-size analysis, WT E. coli was grown with 13C6 glucose and served as the internal standards. Mutant strains were grown with M9 minimal medium with 2 g/L unlabeled glucose and harvested at exponential phase. Fully labeled WT cultures and unlabeled culture were mixed and metabolites were extracted for LC-MS analysis. The relative pool sizes between WT and mutant strains were calculated based on isotope enrichment ratios.

**Protein purification**

Enzymes (Supporting Information Table S2) were individually overexpressed and purified from recombinant E. coli strains as described in our recent article.27 Freshly purified protein was concentrated using Amicon Ultra-15 centrifugal filter devices (Millipore, Darmstadt, Germany). Concentrated protein was applied to a PD-10 column (GE Healthcare, Chicago, IL) equilibrated with storage buffer (100 mM phosphate buffer, 10% glycerol, pH 7.6) for buffer exchange. Protein concentrations were measured with a Pierce BCA protein assay kit (Thermo Fisher Scientific, Massachusetts). Proteins were stored at -80°C after flash freezing in liquid nitrogen.

**In vitro assays of EMP and ED pathway enzymes**

In all in vitro assays, 0.1-1.0 μM of each protein component, 0.5 mM ATP, 2 mM ADP, 2.5 mM NAD+ were added to a reaction buffer containing 50 mM sodium phosphate buffer (pH 7.6), 1 mM potassium chloride, 1 mM Tris (2-carboxyethyl) phosphine (TCEP), 2 mM magnesium chloride, 0.2 mM 2,3-bisphosphoglyceric acid (2,3-BPG). For EDP related assays, 2.5 mM NADP+, and 0.2 mM FeSO4 were added as well. Reactions were initiated by adding 5 mM glucose and incubated at 37°C. Upon adding glucose to initiate the reactions, all contents were quickly mixed, and the increase in absorbance at 340 nm due to NADP+H formation was measured in a microplate reader (as a convenient way for tracking glycolytic activity). In brief, for individual enzyme activity, reactions were measured spectrophotometrically and the unit (U) or amount of purified protein that converted one μmol of substrate into product per minute was determined.

**Results and Discussion**

**Comparative studies on isozymes’ functions under in vitro and in vivo conditions**

E. coli phosphofructokinase, Pfk, has two isozymes that convert F6P to FBP. Under in vivo conditions, PfkA has higher copy numbers than PfkB (Table 1), and knocking out pfkA significantly impairs E. coli glucose metabolism (Table 2). In contrast, knocking out the isozyme gene pfkB has a negligible effect on E. coli growth.28,29 Under in vitro conditions here, purified PfkB exhibits twice the enzyme activity (U/mg) as PfkA and removal of PfkA has minimal impact on the in vitro glycolysis reaction rates (Table 2). Following Pfk, the isozymes, fructose-bisphosphate aldolase-A, and fructose-bisphosphate aldolase-B (FbaA and FbaB) reversibly cleave FBP (Figure 2). Under in vitro conditions, fbaA knockouts are lethal to E. coli, while ΔfbaB does not affect cell growth (Table 2). Under in vitro conditions, FbaA has a higher activity than FbaB and the removal of FbaA slowed down but did not stop the glycolytic reaction. As expected, all tested isozymes in Table 2 were interchangeable under in vitro conditions. For example, in the absence of PfkB or FbaA, in vitro EMP flux could be recovered if excessive PfkA or FbaB were added. We also tested the double deletion of Pfk or Fba and

4 DOI 10.1002/aic Published on behalf of the AIChE Month 2018 Vol. 00, No. 0 AIChE Journal
Table 2. In vitro Test of EMP Isoenzymes via a Factorial Experiment

<table>
<thead>
<tr>
<th>PLC (nM)</th>
<th>PLE (nM)</th>
<th>FbaA (nM)</th>
<th>PflB (nM)</th>
<th>In vitro rate change (95% CI L, U)</th>
<th>In vitro rate change</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>0% (-2%, +103%)</td>
<td>-75%</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>100</td>
<td>100</td>
<td>-42% (-34%, -51%)</td>
<td>-8%</td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>0</td>
<td>100</td>
<td>-71% (-61%, -83%)</td>
<td>-100% (lethal)</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>-5% (-2%, -9%)</td>
<td>-9%</td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>100</td>
<td>0</td>
<td>-71% (-55%, -73%)</td>
<td>No data</td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>0</td>
<td>100</td>
<td>-37% (-34%, -40%)</td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>-13% (-26%, -26%)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0</td>
<td>100</td>
<td>0</td>
<td>-42% (-37%, -49%)</td>
<td></td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>-26% (-21%, -35%)</td>
<td></td>
</tr>
<tr>
<td>1000</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>+144% (+131%, +165%)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>1000</td>
<td>+13% (+10%, +22%)</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>0</td>
<td>1000</td>
<td>+161% (+157%, +163%)</td>
<td></td>
</tr>
<tr>
<td>Control plus 225 g/L polyethylene glycol 3500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In vitro rate changes (%) based on the control condition after the removal or addition of Plk and Fba isoenzymes. The control condition is all EMP enzymes at a concentration of 100 nM. In vivo rate changes were based on a recent report. Confidence intervals (95% CI) were determined through repetitions of curve fitting calculations resulting in corresponding lower and upper bounds (L, U, B).

different crowding effects under in vitro conditions (Table 2). Removal of both Plk and FbaA isozymes caused the highest reduction of the EMP flux, while removal of PlkB and FbaB showed minimal impact. In summary, in vitro systems represent homogenous environments, and multiple systems level interactions are lost. Therefore, pathway function correlates well with activities and concentrations of individual enzymes. On the other hand, intracellular systems may not be 'well mixed' and enzyme–enzyme interactions or enzyme clustering may play a key role in directing EMP fluxes. For example, it has been reported that PlkB and FbaA enzymes form channels via electrostatic forces to confer catalytic advantages with minimal enzyme usage, and metabolic flux can be highly nonlinear to gene and enzyme expression levels.

The effect of ED pathway overexpression on in vitro and in vivo glycolytic rates

To explore the flux control of glycolytic pathways, the competing native ED pathway was overexpressed (ED pathway enzymes utilized; Figure 2). Reaction thermodynamics and kinetics of the ED reactions have advantages over the EMP pathway, and the ED pathway requires fewer reaction steps and much lower protein synthesis costs. In a cell-free system, we tested glycolytic enzymes with concentrations ranging from 100 to 1000 nM (enzymes included in in vitro reactions are listed in Supporting Information Table S2). Although the EMP pathway demonstrated catalytic rates lower than the ED pathway at the same enzyme concentrations (Figure 3a, Table 1), in vitro overexpression of ED enzymes could significantly increase glucose conversions (Figure 3b). As expected, in vitro introduction of the ED enzymes could also recover catalytic rates in the absence of PlkB or FbaA, two key enzymes for in vivo EMP pathway control.

E. coli metabolism is highly regulated, namely through transcriptional and allosteric regulation, which is closely tied to catalytic repressor/activator proteins and carbon substrates. Despite bypassing transcriptional regulation via plasmid engineering, overexpressing the native ED pathway does not increase flux through the ED pathway, and the engineered strain still prefers to utilize its EMP pathway rather than its ED pathway (Figure 4). One possible explanation is that EMP channels protect their intermediates from competing reactions such as the ED pathway. To test this hypothesis, presumed channels must be disrupted while still maintaining enzyme functions. In E. coli, the PTS uses PEP to phosphorylate glucose into G6P and is thought to be the first step of EMP channeling. By knocking out the PTS, the presumed channel could be disrupted, facilitating flux redistributions from the EMP towards the competing ED pathway. Thereby, we constructed PTS-knock-out mutants to examine their in vivo flux redistributions via metabolic flux analysis (Figure 4). The ΔptsG mutant showed less flux through the EMP pathway, while its PP pathway flux was significantly elevated. After overexpression of native edd and eda in the ΔptsG mutant, the ED pathway effectively hijacked PP pathway fluxes from the G6P node and became a dominating glycolysis route. Intracellular metabolite analyses further confirmed the flux redistributions (Figure 5). The ΔptsG mutant had much smaller pool sizes in upper EMP pathway metabolites (FBP, GAP, DHAP, and PGA) but larger PEP concentrations than WT strain. Furthermore, overexpression of the ED pathway drained the G6P and F6P pools, and the ΔptsG or ED pathway overexpression had slight impact on the concentrations of TCA cycle metabolites.

Figure 5. Metabolite pool-size comparison to WT E. coli K-12.

Pool size analysis of PTS mutant (ΔptsG) and a PTS mutant with an overexpressed ED pathway (ΔptsG + ED) compared to the WT. A ratio of 1 means both strains had the same pool size per biomass (n = 3). [Color figure can be viewed at wileyonlinelibrary.com]
Figure 6. Labeling dynamics following 13C6-glucose pulse in different E. coli strains.
(a) On the left axis, the fully labeled or M + 6/M + 3 mass isomeromer ratio is plotted for WT, ΔptsG, and ΔptsG + ED. Mass isomeromer ratios are the relative abundance of mass isomers where M + 0 is fully unlabeled, M + 3 is three labeled carbons (fully labeled for PGA and PEP) and M + 6 is six labeled carbons (fully labeled for FBP). Cultures were pulses with 13C6-glucose for 5 s. The yellow curve represents the intensity of fully labeled metabolites, and the unit is absolute counts per second (cpm, the right y-axis) (n = 2). (b) Mass isomeromer distribution of M + 3 in FBP and FBP (n = 2). [Color figure can be viewed at wileyonlinelibrary.com]

Transient labeling of E. coli EMP intermediates

To explore heterogeneous metabolite pools in the EMP pathway, we used isotopically nonstationary 13C-labeling, which can probe the bulk pool’s labeling dynamics and be used to investigate substrate channeling.

In our transient study, we observed faster FBP, DHAP, and PGA labeling (bulk M + 6 MID and M + 6 counts) than their upstream precursors G6P and F6P (Figure 6a). This reversal of the labeling order could be explained by the channeling theory: G6P is not uniformly mixed in the cytosol, and instead, the portion of G6P outside of the enzyme channel becomes less metabolically active, contributing largely to the M + 0 pool during dynamic labeling experiments. Such dilution of intracellular metabolite labeling has been reported by isotopic nonstationary flux analysis of bacterial metabolisms. Conversely, in the ΔptsG and ΔptsG + ED mutant, labeling between G6P, F6P, and FBP was strictly sequential (labeling enrichment: G6P > F6P > FBP), along with a decrease in the FBP pool (Figure 5). Moreover, ΔptsG and ΔptsG + ED mutants had much slower labeling in FBP than other cascade metabolites, PGA and DHAP, forming a "V" shape in Figure 6a. The higher labeling even downstream products was consistent with elevated bypass fluxes via PP pathway and ED pathway in these mutants. During dynamic 13C labeling of ΔptsG mutants (Figure 6b), FBP contained much more M + 3 isomeromer and other labeled carbons than its precursor G6P due to the increase of reversed Fba reactions from C3 sugar phosphates (PGA/DHAP→FBP). This result was consistent with the report of the gluconeogenesis in PTS mutants.66

Broader perspective on the glycolysisoxosomes

The common view of cell metabolism assumes enzymes and metabolites are homogeneously distributed throughout the cell, but amidst surmounting observations, this assumption falls and presents difficulties for metabolic modeling. In this study, both in vitro and in vivo glycolytic systems were compared. We demonstrated knocking out the PTS in E. coli loosens the metabolic flux network and changes reaction directions toward competing pathways. Although this work did not directly prove that intracellular enzymes self-organize to form subcellular compartments, dynamic labeling experiments provide reasonable in vitro channeling tests for systems perturbed at possible channeling flux control nodes. Intuitively, Malate→Oxaloacetate→Citrate in the TCA cycle of esukate cells have been proved to be highly channeled.37,38 In this study, labeling rates among PEP, citrate, malate, and succinate in the wild type E.coli were compared in Supporting Information Figure S2. After pulsed cells with 13C6-glucose, PEP carboxylase led to amount of 13C6-M + 3 labeled C4 TCA metabolites (e.g., malate and oxaloacetate), while pyruvate dehydrogenase produced fully labeled Acetyl units (M + 2). We expected that citrate synthase would generate 13C5 (M + 5) citrate. However, the citrate labeling rate was slower than expected with only M + 2 and M + 3 isomeromers detected by LC-MS at the initial stage. Such enzymatic observations were consistent with confounding labeling patterns due to metabolite channeling, and in turn could introduce errors into 13C-based flux analysis.

Metabolite channeling may play a key role in flux regulation. The EMP channel rigidifies glycolytic flux and influences catabolite repression as innate channeling would constrain fluxes and repress competing pathways.44 Understanding channeling-based regulation can help resolve inferior efficiencies in heterologous pathways by making proper co-localizations of engineered enzymes. Moreover, the heterogeneous distribution of intracellular enzymes may also confound kinetic modeling and multiple-omics analyses. Introducing channeling coefficients may improve the predictions of flux redistribution in branched pathway nodes. Another substanstel question is whether innate channeling is associated with gene locations (Table 1) or can be rebuilt via continuous strain evolution. Understanding the scope and the rules dictating the formation of metabolite channeling in native pathways may improve genome-to-phenome mapping, pathway engineering, and metabolic modeling.34,41,45

Greater cognizance of channeling will enable bio-systems engineers to add an extra level of control on cellular metabolic fluxes to current methods that rely mostly on gene copy numbers and transcriptional and translational regulations.

Conclusion

Proximity channeling through metabolic pathways cannot be measured intracellularly via typical tools (such as protein imaging and protein structural analysis) due to weak and possibly transient interactions. Therefore, the hypothesis of metabolite channeling through the EMP pathway is still controversial.30,33 This study compared in vitro and in vivo glycolytic activities, which elucidated different performances of
glycolysis enzymes under two modes and offered explanations of intracellular EMP functions using the channeling theory. The observed differences in metabolite isotopic dilutions may be due to disruption of metabolite channels and provides a general in vivo approach for probing organisms for substrate channeling. However, this study still cannot answer many remaining questions regarding the structural formation of glycolytic enzymes or potential mechanisms for metabolite channeling. Novel in vivo tools are needed to further understand and validate proximity channeling in native pathways.

Acknowledgments

This work was partially supported by DOE grant (DESC0018324) for understating flux regulation. It was also a part of the DOE Joint BioEnergy Institute (http://www.jbei.org) supported by the United States Department of Energy, Office of Science, Office of Biological and Environmental Research. Lawrence Berkeley National Laboratory and the United States Department of Energy. The authors declare no competing financial interests.

Notations

<table>
<thead>
<tr>
<th>AcCoA</th>
<th>acetyl-CoA</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>aKG</td>
<td>alpha-ketoglutarate</td>
</tr>
<tr>
<td>ADP</td>
<td>adenosine diphosphate</td>
</tr>
<tr>
<td>ASP</td>
<td>aspartate</td>
</tr>
<tr>
<td>Cit</td>
<td>citrate</td>
</tr>
<tr>
<td>DHP</td>
<td>dihydroxyacetone phosphate</td>
</tr>
<tr>
<td>EDP</td>
<td>Entner-Doudoroff pathway</td>
</tr>
<tr>
<td>EMP</td>
<td>Embden-Meyerhof-Parnas pathway</td>
</tr>
<tr>
<td>F6P</td>
<td>fructose-6-phosphate</td>
</tr>
<tr>
<td>F6P-1P</td>
<td>fructose-6-phosphatase phosphatase</td>
</tr>
<tr>
<td>KDPG</td>
<td>2-deoxy-3-deoxy-phosphogluconate</td>
</tr>
<tr>
<td>GAP</td>
<td>glyceraldehyde-3-phosphate</td>
</tr>
<tr>
<td>GRP</td>
<td>glucose-6-phosphate</td>
</tr>
<tr>
<td>GLU</td>
<td>glucose</td>
</tr>
<tr>
<td>MAL</td>
<td>malate</td>
</tr>
<tr>
<td>MDI</td>
<td>mass isotomer distribution</td>
</tr>
<tr>
<td>MFA</td>
<td>metabolic flux analysis</td>
</tr>
<tr>
<td>NADH</td>
<td>nicotinamide adenine dinucleotide</td>
</tr>
<tr>
<td>NADPH</td>
<td>nicotinamide adenine dinucleotide phosphate</td>
</tr>
<tr>
<td>PEP</td>
<td>phosphoenolpyruvate</td>
</tr>
<tr>
<td>PK</td>
<td>pyruvate kinase</td>
</tr>
<tr>
<td>PGAL</td>
<td>3-phosphoglycerate</td>
</tr>
<tr>
<td>PPP</td>
<td>pentose phosphate pathway</td>
</tr>
<tr>
<td>PYNR</td>
<td>pyruvate</td>
</tr>
<tr>
<td>OAA</td>
<td>oxaloacetate</td>
</tr>
<tr>
<td>SUCC</td>
<td>succinate</td>
</tr>
<tr>
<td>WT</td>
<td>wild-type</td>
</tr>
</tbody>
</table>

Literature Cited


Manuscript received Apr. 5, 2018, and revision received Jun. 28, 2018.