INTERROGATING CENTRAL CARBON METABOLISM IN *ESCHERICHIA COLI* VIA THE MAPPING OF FLUX RESPONSES TO GENE KNOCKOUTS AND ADAPTIVE EVOLUTION

by

Christopher P. Long

A dissertation submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Chemical Engineering

Spring 2018

INTERROGATING CENTRAL CARBON METABOLISM IN *ESCHERICHIA COLI* VIA THE MAPPING OF FLUX RESPONSES TO GENE KNOCKOUTS AND ADAPTIVE EVOLUTION

by

Christopher P. Long

Approved: _____
Eric M. Furst, Ph.D.
Chair of the Department of Chemical and Biomolecular Engineering

Approved: _____
Babatunde A. Ogunnaike, Ph.D.
Dean of the College of Engineering

Approved: _____
Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Maciek R. Antoniewicz, Ph.D.
Professor in charge of dissertation

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Eleftherios T. Papoutsakis, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Wilfred Chen, Ph.D.
Member of dissertation committee

I certify that I have read this dissertation and that in my opinion it meets the academic and professional standard required by the University as a dissertation for the degree of Doctor of Philosophy.

Signed: _____

Mark J. Nelson, Ph.D.
Member of dissertation committee

ACKNOWLEDGMENTS

I am indebted to many who have made my education, and this particular adventure, possible. While doctoral research can be a somewhat long and lonely road, the love of family and friends have never been far from my mind. I am forever grateful to my parents and family for their endless support, and for nurturing curiosity and wonder in my sisters and I from an early age. Hanna, you earned this with me. Your constant and present support throughout this journey was priceless, and I could always count on a full social calendar to keep life balanced. Thank you for everything; I love you, and can't wait to see what lies ahead.

My experience at UD was made by colleagues and friends. First and foremost I thank my adviser, Maciek Antoniewicz, for his mentorship, guidance, and contagious passion for science. I also thank my committee members for their insight and feedback, as well as the CBI community for exposing me to a wide breadth of scientific perspectives. The Antoniewicz Lab members, past and present, have been excellent colleagues and friends. A special thanks to Scott Crown, Jennifer Au, and Nikodimos Gebreselassie for stimulating discussions and productive collaborations, especially early on, that were educational and professional highlights of my time in the lab. Thanks to Jacqueline Gonzalez, Camil Diaz, Brian McConnell, and Ellie Oates for keeping these traditions alive and providing a fun environment to work in. I also thank the Papoutsakis Lab, especially Nich Sandoval for his help and teaching. Finally, my time at UD wouldn't have been nearly as much fun without my housemates and friends, especially Kay, Ed, Andrew, Megan, Chad, Cameron, and Niko.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Biological systems have enormous potential for chemical conversion, including otherwise inaccessible syntheses and utilizing renewable feedstocks. Designing production strains through metabolic engineering requires detailed systems knowledge, particularly in the mapping of genotype (i.e., where manipulations are most typically made) to the metabolic phenotype (the output, e.g., overproduction of a chemical product). In this work, advances were made in metabolic characterization methods, and these methods were then applied to map metabolic responses to gene knockouts and adaptive laboratory evolution. Specifically, novel methods for measuring biomass composition, useful new measurements for ^{13}C metabolic flux analysis (^{13}C-MFA), and strategies for optimal tracer design were developed. These optimized methods were used to comprehensively assess metabolic responses to 45 *E. coli* gene knockout strains of enzymes in central carbon metabolism. Analysis of flux rewiring in these strains revealed bottlenecks and areas of flexibility in metabolism, a novel reversibility of Enzyme I of the PTS system and a glucose secretion phenotype. These results constitute a significant new resource for systems biology, particularly for metabolic modeling where they will be directly applied to the development of ensemble kinetic models.

Genetic and metabolic responses to adaptive laboratory evolution in two strains were also characterized, providing new insights into the processes of microbial adaptation and fitness enhancement. Growth recovery (of up to 3.6-fold) in an *E. coli* knockout strain of a core glycolytic enzyme was enabled by a unique set of mutations

which alleviated rate limiting steps in metabolism. In evolved *E. coli* wild-type strains, growth rate enhancements of 50% did not correspond to intracellular flux rewiring, indicating broad and proportional regulatory change. Mutations in both experiments also suggest critical roles for global regulators in adaptation. Finally, the metabolism of *Vibrio natriegens*, a very fast-growing and potential next-generation host organism, was elucidated by ^{13}C-MFA. This provides an important baseline of knowledge to facilitate modeling and engineering of this organism. Further investigation into the mechanisms of fast-growth, both natural and evolved, will enable the development of hosts with superior productivity and economic potential.

Chapter 1

INTRODUCTION

Parts of this chapter are reprinted with permission from: Long CP, Antoniewicz MR (2014) Metabolic flux analysis of Escherichia coli knockouts: lessons from the Keio collection and future outlook. *Curr. Op. Biotech.* (28):127-133.

1.1 Systems-Level Analysis of Metabolism in Microbiology and Metabolic Engineering

Metabolism is the set of biochemical reactions that sustain cellular life. These reactions facilitate the consumption of nutrients, the production of useful forms of energy, and the synthesis of new cellular components. Metabolic reactions are typically catalyzed by specific proteins, or enzymes, of which there is an incredible natural diversity. The characterization of metabolic pathway reactions and associated genes/proteins, especially in model organisms such as the gram-negative bacterium *Escherichia coli*, was a major accomplishment of 20th century science (Caspi et al., 2012; Kanehisa and Goto, 2000). Subsequently, the challenge has been to develop quantitative, systems-level understanding of metabolism *in vivo*, particularly its regulation, kinetics, and adaptability. In the context of microorganisms, these questions have particular relevance to issues of the environment (Yao et al., 2015) and human health (Chandrasekaran and Price, 2010; Greenblum et al., 2013).

There has also been a long-recognized potential to leverage the diversity and specificity of biochemistry to develop novel chemical transformations, i.e. metabolic engineering (Bailey, 1991; Stephanopoulos, 1999). Advantages of biochemical processes include the ability to access the unique and highly specific chemistries, e.g.

for natural products (Paddon et al., 2013), as well as for utilizing renewable non-fossil fuel feedstocks (Langholtz et al., 2016). These efforts are enabled by genetic editing technologies, which allow for the manipulation of the system toward desired ends, such as optimal production of a chemical of interest. A fundamental challenge, however, is that the complexity of the cellular network (e.g., more than 4000 genes in *E. coli*) make identifying optimal engineering interventions difficult. Improved understanding of the quantitative, systems-level relationship between the genotype and phenotype is critical for such rational engineering efforts.

1.2 Quantitative Methods for Metabolic Analysis

The annotation of metabolic pathways provides a map of possible chemical conversions. To determine actual fluxes in the metabolic network, further assumptions or measurements are necessary. The constraint-based reconstruction and analysis (COBRA) family of *in silico* modeling tools utilizes the stoichiometry of the genome-scale model, and then applies various objective functions to identify a unique flux solution (Lewis et al., 2012; Schellenberger et al., 2011). For example, flux balance analysis (FBA) typically optimizes the growth rate, subject only to stoichiometry and substrate uptake limitations (Edwards and Covert, 2002). These methods are useful in identifying theoretical possibilities for performance, but do not provide direct insight into the actual *in vivo* metabolic state.

The predominant method for experimentally determining metabolic flux is via stable isotope tracers, particularly ^{13}C metabolic flux analysis (Antoniewicz et al., 2007a; Wiechert, 2001; Zamboni et al., 2009). An isotopically labeled substrate, for example glucose, is consumed by the organism of interest, and the resulting labeling of downstream metabolites in the network are measured. These labeling data contain

information about pathways usage, and provided sufficient measurements and an accurate network model of metabolism, fluxes and confidence intervals can be estimated (Antoniewicz et al., 2006). In such studies, the design of experiment, including single or parallel tracer experiments (Crown and Antoniewicz, 2013a), tracer selection, and choice of measurements are critical.

1.3 Toward a Comprehensive Study of Knockout Responses

The elucidation and quantification of complex regulatory and metabolic systems is of fundamental interest to biologists and engineers. A useful method in unraveling this complexity is to observe the biological system under perturbed conditions, for example, by removal of identified components (e.g. genes). As a model prokaryotic organism, *Escherichia coli* is ideally suited for gene knockout studies, and indeed, many studies on *E. coli* knockouts have been performed. The Keio library of all viable *E. coli* single-gene knockouts is now further facilitating these studies (Baba et al., 2006). In addition to having significant value in fundamental biological sciences, quantitative studies of cellular responses to gene knockouts provide relevant data for metabolic engineering and biotechnology applications.

Multiple omics techniques have been applied to gain insight into the systemic responses of *E. coli* to gene knockouts, including metabolomics (Ishii et al., 2007; M. Li et al., 2006a; Siddiquee et al., 2004a; Toya et al., 2010; Zhu and Shimizu, 2005), transcriptomics (Fong et al., 2006; Ishii et al., 2007; M. Li et al., 2006a; Siddiquee et al., 2004b; Waegeman et al., 2011), enzymatic activity measurements (M. Li et al., 2006a, 2006b; Siddiquee et al., 2004a; Zhao et al., 2004; Zhu and Shimizu, 2005), and metabolic flux analysis (fluxomics). Of all these omics levels, the fluxome provides the most relevant representation of the cellular phenotype for guiding metabolic

engineering efforts (He et al., 2013). Since early on in metabolic engineering, a lack of understanding and accurate predictability of complex perturbation responses has been identified as a major obstacle (Bailey, 1991; Stephanopoulos, 1999).

1.3.1 Predicting Flux Responses of *E. coli* Knockouts

There has been much work done in applying computational tools to predict responses of genetically perturbed biological systems. The performance limits of *E. coli* metabolic networks, subject to gene deletions and additions, have been traditionally assessed using FBA (Burgard and Maranas, 2001), where a linear optimization and an "objective function" are used to predict feasible flux distributions. The optimization serves as a proxy for evolutionary forces, e.g. a maximized biomass production flux represents evolutionary pressure under competitive growth conditions. The FBA approach and various objective functions have been evaluated for wild-type *E. coli*, and this approach been shown to generally work well (Schuetz et al., 2007). However, for unevolved genetically perturbed strains, the evolution-based objective function is questionable. Indeed, for many knockouts, FBA has been limited to evaluating the feasibility of growth (Edwards and Palsson, 2000), and predicting growth rates of evolved knockout strains (Fong and Palsson, 2004).

Several alternative approaches have been proposed to predict flux distributions of *E. coli* knockouts. Minimization of Metabolic Adjustment (MOMA) (Segre et al., 2002) postulates that the perturbed metabolic state will be as close as possible (by Euclidean distance) to the FBA optimum of the wild-type; this favors solutions with many small changes to fluxes rather than smaller number of large changes. Given that this can be inconsistent with the concepts of regulatory adaptation cost and linearity of flow, the alternative Regulatory On/Off Minimization (ROOM) approach has been

developed, which minimizes the number of large flux changes from the FBA solution (Shlomi et al., 2005). A third algorithm, RELATCH (RELATive CHange), suggests using experimental flux and expression data from a reference strain as the starting point. It then uses two parameters to describe the cell's efforts to first minimize regulatory and distribution pattern changes, and then after a short time activate previously latent pathways and increase the flux capacity of previously active pathways (Kim and Reed, 2012). Yet another approach uses FBA with additional regulatory constraints in the form of predicted gene co-occurrence (co-regulation), or "flux-coupled genes" (Kim et al., 2013; Park et al., 2010). Finally, a cybernetic modeling approach has been described that has the additional benefit of addressing microbial productivity as well as yields (Song and Ramkrishna, 2012). All these algorithms have reported increased accuracy in predicting flux responses to knockouts when compared to FBA, judged against existing experimental flux data from ^{13}C-MFA studies. However, a thorough assessment of the various methods has been difficult because of significant limitations in the existing knockout flux data (as described next).

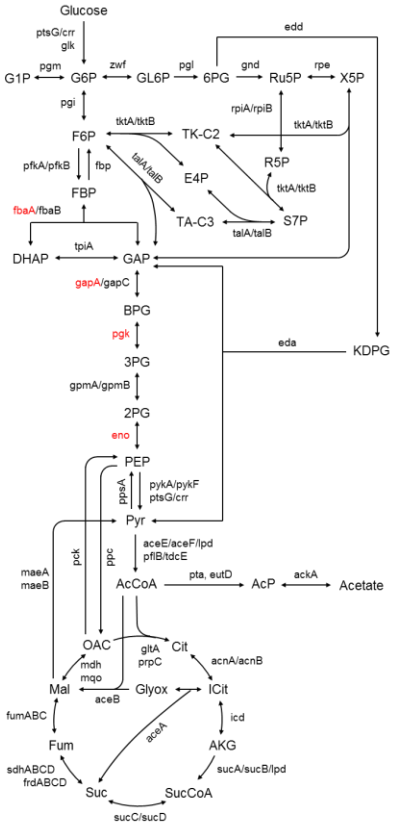
1.3.2 Previous Metabolic Flux Analysis Studies of *E. coli* Knockouts

In the past decade, numerous ^{13}C-MFA studies have been performed to measure fluxes in *E. coli* knockouts. The goals of these studies have been to explore specific parts of the metabolic and regulatory network, and to assess the general perturbation response. Examples of the former are studies which combine genetic and environmental perturbations to study oxygen sensing and aerobic regulatory response (Nizam et al., 2009; Nizam and Shimizu, 2008; Portnoy et al., 2010). Another example is a study by Nakahigashi et al., in which a novel hidden reaction in the pentose

phosphate pathway was discovered by studying sets of double knockouts (Nakahigashi et al., 2009). In addition to helping to resolve network structure, knockout studies can also help elucidate regulation and dynamics. For example, Siddiquee et al. studied flux distributions, enzyme activities, expression levels, and metabolite concentrations of a *pykF* knockout mutant to quantitatively describe the regulation of PEP carboxylase, malic enzyme, phosphofructokinase, acetate formation, and the oxidative pentose phosphate pathway (Siddiquee et al., 2004b).

There has also been interest in identifying challenges induced by knockouts, and the general and specific adaptive responses to them. For example, Canonaco and Sauer showed that due to overproduction of NADPH in the *pgi* knockout, transhydrogenase activity was kinetically limiting cell growth (Canonaco et al., 2001). Fong et al. showed that the initial response of *E. coli* to many gene knockouts was to activate latent pathways, such as the glyoxylate shunt and Entner-Doudoroff pathway, however, when the responses were monitored over many generations it was observed that the sub-optimal latent pathways were subsequently re-repressed and the capacities of more efficient pathways were expanded to reach new optima (Fong et al., 2006). Other examinations of the range of adaptive responses include the immediate flux response, based on a plasmid-based inducible knockout approach (Usui et al., 2012), and changes over the course of an extended batch culture (Toya et al., 2010). These studies inform our understanding of bacterial evolution, and have practical consequences for the industrial use of genetically modified strains. However, as a window to viewing the native metabolic and regulatory structure of *E. coli*, the studies of evolved strains are probably less informative than those studies immediately following a perturbation.

The most widely studied knockouts to date have been those of central carbon metabolism and global regulation. A summary of genes whose knockout mutant fluxes have been reported in literature are shown in Figure 1.1. While knockout flux studies are clearly an area of strong interest, bias in coverage and poor comparability of



E. coli knockout	Reference for ^{13}C-MFA study				
	(micro)aerobic		anaerobic		
Central carbon metabolism	**batch**	**/ continuous**	**batch**	**/ continuous**	**Other growth condition**
ptsG / crr	-	41	-	-	-
galM	-	4	-	-	-
glk	-	4	-	-	-
pgi	5,8,29,30,32,34	4,35,41	-	-	NH_4^+ limitation (35); growth on galactose (32)
pgm	-	4	-	-	-
pfkA / pfkB	34	4	-	-	-
fbp	-	4	-	-	-
fbaA / fbaB	-	4	-	-	-
tpiA	8	-	-	-	-
gapAC	-	4	-	-	-
pgk	-	-	-	-	-
gpmA / gpmB	-	4	-	-	-
eno	30	-	-	-	-
pykA / pykF	2,5,34	4,6,7,33	-	-	NH_4^+ limitation (33)
aceE / aceF	-	-	-	-	-
lpd	-	11	-	-	-
pflB / tdcE	-	-	-	-	-
zwf	31,32,34	4,10,35,38	-	-	NH_4^+ lim. (35); growth on pyr (10) and ac (38)
pgl	-	4	-	-	-
gnd	-	4,10,37	-	-	growth on pyruvate (10)
rpiA / rpiB	-	4	-	-	-
rpe	-	4	-	-	-
tktA / tktB	-	4	-	-	-
talA / talB	28[a]	4	-	-	-
edd	-	-	-	-	-
eda	-	-	-	-	-
gltA	-	-	-	-	-
prpC	-	-	-	-	-
acnA / acnB	-	-	-	-	-
icd	-	-	-	-	-
sucA / sucB	-	3	-	-	-
sucC / sucD	-	3	-	-	-
sdhABCD	34*,32	-	-	-	growth on galactose (32)
frdABCD	-	-	-	-	-
fumABC	34	-	-	-	-
mdh	34*	-	-	-	-
aceA	-	-	-	-	-
aceB	-	-	-	-	-
ppc	2,8	-	-	-	-
pck	34*	36	-	-	-
maeA / maeB	34*	-	-	-	-
ppsA	-	4	-	-	-
pta	2,8	-	-	-	-
ackA	-	-	-	-	-
mgsA	-	41	-	-	-
Regulatory genes					
arcA	32,39	9,26,40,42,43	39	-	Nitrate as electron acceptor (39, 43)
arcB	39	26	-	-	-
cra	39	-	-	-	-
crp	39	41	-	-	-
cya	39	-	-	-	-
fnr	39	40	-	-	-
mlc	39	41	-	-	-
iclR	-	9,42	-	-	-
fur	32	-	-	-	growth on galactose (32)
pdhR	32	-	-	-	growth on galactose (32)
ihfA	32	-	-	-	growth on galactose (32)
ihfB	32	-	-	-	growth on galactose (32)

Figure 1.1: A summary of the available literature for experimental flux studies on *E. coli* knockout strains in central carbon metabolism, as of 2014. The numbers refer to individual references as cited in *Long & Antoniewicz, Current Opinion in Biotechnology (2014) 28:127-133.*

results have prevented these results from being utilized to their full potential. Understandably, much attention has been paid to phosphoglucose isomerase (*pgi*),

glucose-6-phosphate dehydrogenase (*zwf*), 6-phosphogluconate dehydrogenase (*gnd*), and pyruvate kinase (*pykA* and *pykF*) genes, as these represent important nodes in metabolism: *pgi* initiates glycolysis, *zwf* and *gnd* initiate and propagate the oxidative pentose phosphate pathway, and *pykAF* provides the main route to pyruvate and acetyl-CoA, from which common fermentation products are produced. Regulatory gene knockouts, particularly of global regulators, have also been studied to some extent. The best studied is the ArcA/B system, which controls the aerobic metabolic response. Again, while it is understandable that these genes were targeted for initial studies, the overall limited coverage is limiting potential biological discoveries and the development of a more complete understanding of the general perturbation response.

Most studies to date have been performed under either substrate-rich conditions (e.g. batch), or substrate-limited conditions (e.g. continuous cultures). Cellular responses under these two conditions are likely to be significantly different. For example, remarkably robust flux profiles (i.e. relatively small flux changes) for 24 knockout strains grown under chemostat growth conditions were reported by Ishii et al. (Ishii et al., 2007). In contrast, much more pronounced metabolic responses were observed for similar strains grown under batch conditions. As an example of the difference, in batch culture the *zwf*-KO strain was reported to secrete acetate with a normalized flux of 44 and have a citrate synthase flux of 51 (Nicolas et al., 2007), while in a continuous culture the acetate flux was 0 and citrate synthase flux was 103 (Ishii et al., 2007). The fact that various continuous culture studies used different dilution rates further complicates a direct comparison of these studies. Thus, with current data it is often difficult to deconvolute the effects of the knockout mutation and environmental changes.

Even among studies using the same gene knockout and growth condition, significant variability in reported fluxes has been observed. This can result from differences in the genetic background of the wild-type, or differences in ^{13}C-MFA methodology employed. For example, for *pgi* knockouts grown under batch conditions, estimates of the normalized TCA flux range from 20 (Haverkorn van Rijsewijk et al., 2011) to 62 (Fong et al., 2006). In continuous studies of the *pykF* knockout at dilution rates 0.1 hr^{-1} and 0.2 hr^{-1} major differences were reported for *pgi*, *pyk* and *ppc* fluxes (Ishii et al., 2007; Siddiquee et al., 2004a).

1.3.3 A complete knockout flux data set would be valuable

Due to a lack of coverage and discrepancies in experimental conditions and methodology, knockout flux results are currently difficult to compare and generalize. To facilitate fundamental analysis of *E. coli* metabolic and regulatory processes, an unbiased and high-resolution data set consisting of methodologically self-consistent ^{13}C-flux results for a large number of knockout mutants would be ideal. This is now increasingly feasible given the significant improvements in ^{13}C-flux methodology in recent years (Antoniewicz, 2013a, 2013b, 2013c; Leighty and Antoniewicz, 2013), as well as the availability of a large number of viable *E. coli* knockout mutants from the Keio collection (Baba et al., 2006). A complete, systematic data set would provide an unbiased basis to assess the general *E. coli* perturbation response. Such a data set would also provide highly valuable inputs for systems biology approaches, e.g. the data could be integrated with other omics levels allowing discovery of novel network connectivities, regulatory relationships and metabolic activities (Cornelius et al., 2011; Fong et al., 2006; Nishikawa et al., 2008).

A well-developed quantitative model of metabolism would be an invaluable tool for testing current knowledge and a significant step forward as a predictive design tool for metabolic engineering. Several approaches to this kind of global metabolic modeling have been proposed and attempted. These include joint regulatory and metabolic models (Yeang and Vingron, 2006), models that incorporate enzyme activity (Kurata et al., 2007), cybernetic models (Young et al., 2008), and differential equation based models (Kadir et al., 2010). Ensemble modeling is one of these approaches where the goal is to obtain global metabolic mass-action kinetic parameters by iteratively constraining the parameter space with experimental metabolite and flux data, an application for which a large-scale ^{13}C-MFA knockout data set would be ideal (Tran et al., 2008). Many other modeling approaches have been equally limited by the available data for training and testing. A complete knockout flux data set would therefore be a valuable resource for the community and enable significant refinements of current models leading to improved accuracy and predictive power.

An improved quantitative understanding would then also facilitate improved interpretation of metabolism. Many existing theoretical frameworks could be applied to, tested, and enriched by comprehensive knockout flux data. Examples include graph theoretic assessments of metabolic robustness (Behre et al., 2008; Ghim et al., 2005; Wilhelm et al., 2004; Wunderlich and Mirny, 2006) and flux coupling (Burgard et al., 2004; Kim et al., 2013), quantifications of the influences of thermodynamics on network function (Henry et al., 2006), classifications of transcriptionally vs. metabolically limited fluxes (Shlomi et al., 2007), modular cost-benefit analysis (Carlson, 2007), Bayesian analysis of modularized networks (Kim et al., 2011),

evaluation of FBA optimality (Chen et al., 2011), and the logic of pathway latency/silencing (Fong et al., 2006; Nishikawa et al., 2008). Interpretation of knockout flux data through these frameworks would greatly improve our understanding of the evolved "logic" of cellular life.

1.3.4 Towards a large-scale standardized flux data set

Prioritization will be important in future efforts to produce a large set of flux data for *E. coli* knockouts. With almost four thousand viable single gene knockouts, and the relatively laborious nature of ^{13}C-MFA, a truly complete set would be a massive undertaking. A few key sets of metabolic genes would be of highest interest and practical value, and should therefore be prioritized. The first and most important knockouts are those of central carbon metabolism. This involves the high-traffic pathways that generate energy, cofactors, and precursors for all other cellular processes. Understanding bottlenecks and kinetics in central carbon metabolism will therefore be critical to many metabolic engineering efforts.

The next gene set of interest is that of regulators of central carbon metabolism. This regulatory network layer is essential to include in future comprehensive models, and quantitative understanding at this level is still not fully developed. For example, improved understanding of the regulatory system would allow for the replacement of stand-in modeling concepts such as "objective function" and other heuristics with more fundamental mechanistic models. Also essential to study are genes related to energy and redox metabolism, given that energetics and co-factor availability are fundamental to increasing metabolic yields. Finally, it would be valuable to investigate genes related to the aerobic/anaerobic response, stress response, and catabolism of carbon sources other than glucose, including xylose, glycerol and acetate.

An effort to produce a large-scale data set of *E. coli* knockout fluxes should utilize the best practices in ^{13}C-MFA. Within reasonable limits of effort and cost, the goal should be to produce the highest possible flux resolution. For example, optimal tracer selection and the use of parallel labeling experiments have been shown to dramatically improve flux measurement precision (Ahn and Antoniewicz, 2013; Crown et al., 2016b, 2012, Crown and Antoniewicz, 2013a, 2012; Leighty and Antoniewicz, 2012a). Consistent and proper reporting of flux data, error analysis, and transparency of methods, are also of utmost importance to enable reproducibility and comparability of results (Antoniewicz et al., 2006). A recent review has proposed minimum data standards for ^{13}C-MFA studies that will facilitate sharing of results (Crown and Antoniewicz, 2013b). With these guidelines and conditions, an experimental effort to obtain a large body of *E. coli* knockout flux data is warranted and promises to deliver significant value.

1.4 Objectives and Outline of this Thesis

With the above context, this thesis aims to advance four main areas of microbial metabolic research: (1) experimental methods for microbial physiological characterization and ^{13}C-MFA (Ch. 2-4); (2) further the accurate mapping of central carbon metabolism by identifying novel reactions (Ch. 5); (3) characterize the metabolic responses to a comprehensive set of 45 knockouts in *E. coli* central carbon metabolism (Ch. 6-9), and (4) dissect the metabolic and genetic mechanisms of adaptive evolution and fast growth (Ch. 10-12). The specific contributions described in each chapter are enumerated below.

- **Chapter 2** describes a set of novel GC-MS based methods for quantifying biomass composition. The composition of microbial biomass is an important

aspect of physiology, and directly informs the estimate of growth-related flux in ^{13}C-MFA. Here, a convenient and precise GC-MS based pipeline is presented for quantifying the four most abundant components of *E. coli* biomass (protein, RNA, lipids, and glycogen). These methods are also applicable to other organisms.

- **Chapter 3** demonstrates that the addition of labeling information from RNA (ribose moiety) and glycogen (glucose moiety) significantly aids the resolution and observability of certain fluxes in ^{13}C-MFA. This is demonstrated in wild-type *E. coli*, an engineered *E. coli* that co-consumes glucose and xylose, and CHO cells.
- **Chapter 4** presents a comprehensive analysis on tracer selection for ^{13}C-MFA. A novel scoring method for global flux resolution is presented, a large number of single, mixed, and parallel tracers are evaluated via simulation. Ultimately, the parallel use of [1,2] and [1,6-^{13}C]glucose are found to provide optimal precision, and this is validated experimentally.
- **Chapter 5** describes the identification of a novel reaction in *E. coli* central carbon metabolism, specifically the reversibility of Enzyme I of the PTS system. This flux is shown to be significantly reversible during glycolytic growth, and is responsible for a large net flux during gluconeogenic growth. Additionally, it is found to be sensitive to perturbations elsewhere in the PTS system.
- **Chapter 6** presents the physiological characterization of 20 *E. coli* knockouts from the 'upper' pathways of central carbon metabolism: glucokinase and phosphoglucomutase, glycolysis/EMP pathway (to

glyceraldehyde-3-phosphate), the pentose phosphate pathway, and Entner-Doudoroff pathway. Batch growth performance, including growth and uptake rates and yields, as well as biomass composition measurements, are included.

- **Chapter 7** presents the intracellular fluxes of the same 20 knockouts from upper metabolism. A novel glucose secretion reaction is observed in *ΔpfkA*, and the methylglyoxal pathway is resolved in *ΔtpiA*. Patterns of flux rewiring are analyzed, including via multivariate analysis, and compared to COBRA model predictions.
- **Chapter 8** presents the physiological characterization of 25 *E. coli* knockouts from the 'lower' pathways of central carbon metabolism: pyruvate kinase and PEP synthetase, pyruvate dehydrogenase and acetate pathways, TCA cycle, glyoxylate shunt, and amphibolic reactions. Batch growth performance, including growth and uptake rates and yields, as well as biomass composition measurements, are included.
- **Chapter 9** presents the intracellular fluxes of the same 25 knockouts from lower metabolism. Patterns of flux rewiring are analyzed, including via multivariate analysis, and compared to COBRA model predictions.
- **Chapter 10** explores how growth rate is recovered in *Δpgi* strains subjected to adaptive laboratory evolution (ALE). Unique sets of mutations are correlated to changes in metabolic flux, including in the cofactor transhydrogenase and PTS system. Phenomena previously described as latent pathway activation and repression are re-examined.

- **Chapter 11** describes the metabolic fluxes in six adaptively evolved *E. coli* MG1655 wild-type strains. The growth rates were increased by 50%, though the intracellular flux patterns were found not to change. This is consistent with the hypothesis that the effect of ALE in the wild-type is to broadly upregulate metabolic and growth functions.
- **Chapter 12** describes the elucidation of intracellular metabolism in the fast-growing organism *Vibrio natriegens*. This organism has been proposed as a next-generation host for biotechnology because of its fast growth. The metabolic network was reconstructed from the annotated genome, biomass composition measured, and fluxes elucidated by ^{13}C-MFA. Encouragingly for metabolic engineers, the normalized flux distribution was quite similar to *E. coli*.

Chapter 2

NOVEL GC-MS BASED METHODS FOR QUANTIFYING BIOMASS COMPOSITION

2.1 Introduction

Quantification of the various components of biomass is important for systems biology and bioengineering. The composition of an organism is a core feature of its phenotype, and provides insight into its underlying metabolic systems as well as differences between environmental conditions, genotypes, and species. In the field of fluxomics, the resolution and accuracy of metabolic flux models derived from approaches such as flux balance analysis (FBA) and ^{13}C metabolic flux analysis (^{13}C-MFA) are known to be sensitive to biomass composition (Pramanik and Keasling, 1998, 1997). Current methods for quantifying biomass are tedious and sometimes inaccurate, which can limit the performance of these fluxomic techniques.

The major components of microbial biomass are protein, RNA, lipids, and glycogen. For the model gram-negative microbe *Escherichia coli*, these four components have been reported to constitute 88% of the dry biomass (Neidhardt, 1987). Current methods to quantify these major biomass components rely on a variety of enzymatic and spectroscopic based methods. Colorimetric assays are often employed to measure total protein content (Bradford, 1976; Peterson, 1977; Smith et

al., 1985), while amino acid quantification requires hydrolysis followed by HPLC analysis (Rutherford and Gilani, 2009). The measurement of RNA typically requires purification followed by spectroscopic quantification (Benthin et al., 1991). While colorimetric methods are also available for total lipid quantification (Wawrik and Harriman, 2010), the use of GC/FID and GC/MS is common for profiling fatty acids (Dobson and Christie, 2002; Dodds et al., 2005; Y. Li et al., 2006). Glycogen is often quantified using enzymatic hydrolysis followed by glucose analysis via HPLC or a colorimetric method (McKinlay et al., 2007).

Here, we have developed a set of methods to quickly, accurately and precisely quantify 17 amino acids, all relevant fatty acids (5 demonstrated here, but easily extendable), RNA, and glycogen on a single, widely available analytical platform: gas chromatography-mass spectrometry (GC/MS). This approach offers a simplified and convenient workflow, summarized in Figure 2.1, and removes reliance on enzymatic and spectroscopic calibrations. All quantifications are based on isotope ratio analysis using analyte-specific standards which are isotopically unique from the sample, giving a high degree of confidence in the results. A similar approach has been previously reported for quantifying metabolite pools (Bennett et al., 2008). Here, "fully labeled" *E. coli* is used as the internal standard. To generate fully labeled *E. coli* in which all cellular carbon is ^{13}C, a large batch of *E. coli* is grown on [U-^{13}C]glucose. The components of this fully ^{13}C-labeled biomass are then quantified against known unlabeled standards. Once characterized, this fully labeled biomass can be used as an internal standard to quantify subsequent unlabeled biomass samples. This procedure adds flexibility and convenience to the workflow, and expands the potential application to organisms that are difficult to label fully (Swarup et al., 2014). The

methods described here are validated and then applied to three *E. coli* strains, confirming consistency, agreement with previously reported values, and demonstrating the practical importance of such measurements in systems biology.

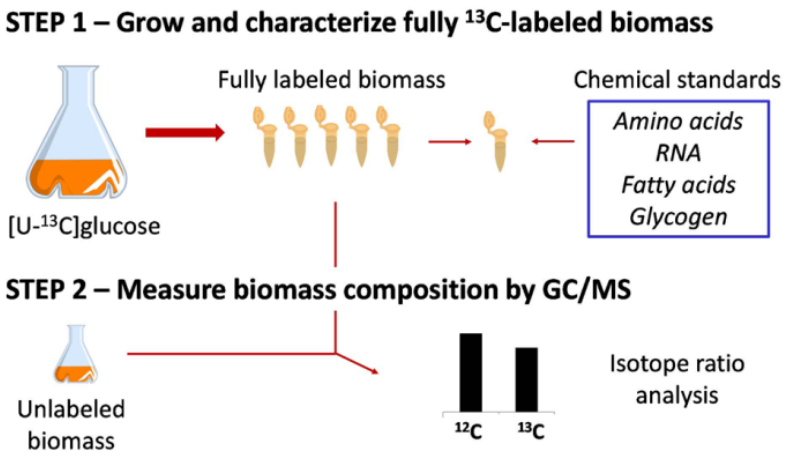


Figure 2.1: Schematic representation of workflow for GC-MS based methods for quantification of biomass composition.

2.2 Materials and Methods

2.2.1 Chemicals

All chemicals were purchased from Sigma-Aldrich (St. Louis, MO). [U-^{13}C]Glucose was purchased from Cambridge Isotope Laboratories (Andover, MA).

2.2.2 Strains and Cultures

E. coli BW25113 strains were obtained from the Keio Knockout Collection (Baba et al., 2006). The parent strain ("wild-type") was used in all validation studies and as the labeled biomass reference. Two knockouts, *Δpgi (phosphoglucose isomerase)* and *Δzwf (glucose-6-phosphate dehydrogenase)* were also analyzed. All

cultures were grown aerobically in M9 minimal medium with 2 g/L of glucose (Leighty and Antoniewicz, 2013), either naturally labeled (unlabeled) or [U-^{13}C] (fully labeled). Cells were harvested at mid-exponential phase ($OD_{600} \approx 0.7$). For all procedures, samples containing the equivalent of 1 mL of a culture at OD = 1.0 (roughly 0.3 mg of dry weight) were used. All biomass samples were washed twice with glucose-free M9 medium prior to analysis. Dry weights were measured for all strains by filtration of 50 mL of a culture at a density of OD = 1.0 using a 0.2 μm cellulose acetate filter (Sartorius 11107-47-N), followed by drying for several days at 80°C.

2.2.3 Chemical Standards

Unlabeled standards were prepared for all relevant analytes, and were added directly to the biomass pellet prior to the execution of the protocols described below. For amino acids, 40 μL of a 2.5 mM per amino acid solution (Pierce 20088) was used. For RNA, a 1 mg/mL solution (ribonucleic acid from torula yeast, Sigma R6625) in water was prepared, of which 80 μL was added to the sample. For glycogen, a 0.1 mg/mL solution (glycogen from bovine liver, Sigma G0885) in water was prepared, of which 100 μL was added to the sample. For fatty acids, a solution of 0.3 mg/mL of myristic acid (C14:0), palmitic acid (C16:0), palmitoleic acid (C16:1), stearic acid (C18:0), and elaidic acid (C18:1) in hexane was prepared, of which 20 μL was added.

2.2.4 GC/MS

GC/MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 μm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating

under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 200°C, the MS quad temperature at 150°C, the interface temperature at 280°C, and the inlet temperature at 250°C. For GC/MS analysis of amino acids, 1 μL was injected at 1:40 split ratio. The column was started at 80°C for 2 min, increased to 280°C at 7°C/min, and held for 20 min. For GC/MS analysis of fatty acid methyl esters (FAME) and sugar derivatives, 1 μL was injected splitless. The column was started at 80°C for 2 min, increased to 280°C at 10°C/min, and held for 12 min.

2.2.5 Amino Acid Analysis

The preparation and GC/MS analysis of biomass amino acids was performed as previously described by (Antoniewicz et al., 2007b). Briefly, biomass pellets were hydrolyzed with 500 μL of 6N HCl at 110°C for 24 h, then dried under air at 65°C. *Tert*-butyldimethylsilyl (TBDMS) derivatives of amino acids were prepared by adding 35 μL of pyridine and 50 μL of MTBSTFA + 1% TBDMCS (Sigma 375934) and incubating for 30 minutes at 60°C, and were then transferred to injection vials for GC/MS analysis. 17 of the 20 amino acids were detected and quantified by this method. The three not measured, arginine, cysteine, and tryptophan, are estimated to constitute 12% of total protein mass (Neidhardt, 1987). Since intracellular amino acid pools are known to be very small relative to proteinogenic amino acids (Bajad et al., 2006), the measured signal will be dominated by the latter.

2.2.6 RNA and Glycogen Analysis

These two biopolymers were analyzed simultaneously, as each required hydrolysis followed by quantification of a sugar monomer (ribose and glucose,

respectively). Biomass pellets were hydrolyzed by addition of 500 µL of 6N HCl. Samples were immediately placed under air at 65°C for drying. The aldonitrile propionate derivatives of glucose and ribose were prepared by the method described by (Antoniewicz et al., 2011). Briefly, 50 µL of 2 wt% hydroxylamine hydrochloride in pyridine was added to the dried sample, which was then incubated for 60 min at 90°C. Next, 100 µL of propionic anhydride was added followed by incubation at 60°C for 30 min. The samples were then transferred to injection vials for analysis. Ribose eluted approximately 2 minutes before glucose, and the *m/z* 173 fragment was used for quantification, which contains last two carbon atoms of each sugar. We found that other cellular sugars, such as fructose and deoxyribose, do not co-elute with either ribose or glucose and therefore do not interfere with these measurements. It is important to note that we used the biopolymers RNA and glycogen as standards. The use of the monomers ribose and glucose as standards yielded inconsistent and inaccurate results, due to the observed kinetics of hydrolysis and subsequent degradation of the sugars under the acidic conditions.

2.2.7 Fatty Acid Analysis

All glassware was first rinsed with chloroform and dried to remove contaminating lipid residues. The biomass pellets were re-suspended in approximately 500 µL of water, transferred to a glass Pyrex culture tube, and dried under air at 65°C. FAME derivatives were prepared by dissolving the dried biomass in 1 mL of methanol and 50 µL concentrated sulfuric acid, and incubating for 2 h at 100°C. The mixture was then cooled to room temperature, and the FAME's were extracted by the addition of 1.5 mL water and 3 mL hexane. The upper organic phase was isolated and dried under nitrogen flow at 40°C. The dried FAME's were then re-dissolved in 100 µL of

hexane and transferred to glass GC vials for analysis. The molecular ions of all species of interest were quantified: C14:0 (M0 *m/z* 242, fully ^{13}C-labeled M14 *m/z* 256), C16:1 (*m/z* 268, 284), C16:0 (*m/z* 270, 286), C18:1 (*m/z* 296, 314), C18:0 (*m/z* 298, 316).

2.2.8 GC/MS Data Analysis

For metabolite quantification using isotope ratio analysis, all measured mass isotopomer distributions were first corrected for natural abundances by the method of (Fernandez et al., 1996). Additionally, the unlabeled (M0) content of fully labeled biomass was assessed (typically ~1-2%, due to presence of unlabeled inoculum) and corrected for. The total ion counts of the labeled species were calculated as the sum of the fully labeled (M*N*) and one-less (M(*N-1)*) isotopes. This was necessary to account for the introduction of ^{12}C atoms due to isotopic impurities in the [U-^{13}C]glucose as well as the fixation of unlabeled CO_2 (Leighty and Antoniewicz, 2012a). The frequency of ^{12}C atoms from these sources were low enough such that significant amounts of M(*N-2*) isotopes were not observed, and thus this effect could be completely accounted for by the stated methods.

2.2.9 Validation of RNA and Glycogen Measurements

For RNA and glycogen quantification, care was taken to ensure that the target macromolecules were being measured, and not for example intracellular sugars such as ribose, glucose, or fructose phosphates. To validate this, a labeling switch experiment was performed. A culture of wild-type *E. coli* was grown on unlabeled glucose to OD 0.5, and then centrifuged, washed, and re-suspended in medium containing [U-^{13}C]glucose as the only carbon source. Time-course data of labeling

incorporation confirmed that low-turnover biomass components were indeed measured (Figure Appendix A.1).

2.2.10 Validation of Method Consistency

The methods described here depend on the quantification of isotope ratios of ^{13}C-labeled and unlabeled species in a sample. In practice, an unlabeled biomass sample can be measured against a labeled reference, or vice versa. Since most chemical standards are unlabeled, it is straightforward to quantify fully labeled biomass. To measure unlabeled biomass samples, we first grew a large batch of fully labeled *E. coli* biomass, and aliquoted a large number of identical samples, each containing an equivalent of 1 mL of a culture at OD = 1.0. The aliquoted biomass pellets were stored at -80°C. This stock of fully labeled biomass was characterized, and then used as a reference with which to measure subsequent unlabeled biomass samples. In this case, the reference and sample biomass pellets were combined directly at the beginning of the workup. To confirm that both methods yielded consistent results, *E. coli* was grown in two parallel cultures, one on unlabeled glucose and one on [U-^{13}C]glucose. The biomass compositions of both cultures were characterized by the two respective methods and we confirmed that both methods yielded consistent results (Appendix Figures A.2-A.4)

2.3 Results and Discussion

2.3.1 Method Validation

First, we validated that the methods for RNA and glycogen quantification were indeed measuring these low-turnover components of biomass. This was determined by measuring time-course labeling profiles in an experiment where *E. coli* was first

grown on unlabeled glucose followed by growth on fully ^{13}C-labeled glucose. The fractional labeling of both RNA and glycogen tracked well with the fraction of new biomass formed after the switch, over a period of three hours (Figure A.1). Intracellular metabolites, which turn over in minutes, would have become fully labeled in this time. Only a small fraction of the measured values, less than 15% for glucose and less than 5% for ribose, reflected fast-turnover metabolites. Therefore, we concluded that the methods we applied were almost entirely measuring the targeted biomass components RNA and glycogen.

Second, we validated that the presented methods gave consistent results regardless of whether a ^{13}C-labeled or unlabeled internal standard was used for quantification, i.e. 1) using unlabeled chemical standards for quantifying a labeled biomass sample; and 2) using fully labeled biomass as reference material for the analysis of unlabeled biomass sample. This was shown by culturing *E. coli* in parallel on unlabeled and fully labeled glucose and applying both approaches. As expected, both approaches were found to yield identical biomass composition values (Appendix Figures A.2- A.4). This result shows that the methods described here are consistent and flexible, and can be adapted for the convenience of the user. For example, the use of a fully labeled reference biomass stock may be preferable if it is found that frozen biomass is simpler to store or more stable over time than unlabeled standard solutions. More significantly, it may enable the convenient analysis of species which are difficult to fully label, such as organisms that require complex medium for growth (Ahn and Antoniewicz, 2013), for which labeled standards may otherwise be prohibitively expensive to generate.

2.3.2 Biomass Composition of Three *E. coli* Strains

Phenotypic differences between different genotypes, particularly gene knockouts, are of significant interest in systems biology (He et al., 2013; Long and Antoniewicz, 2014a). To demonstrate the utility of these methods, the biomass compositions of three *E. coli* strains were analyzed: wild-type (WT), and two knockout strains, *Δzwf* and *Δpgi*. These knockouts are missing the first reaction in the oxidative pentose phosphate pathway and glycolysis, respectively. As such, they have significantly altered metabolic states from the wild-type and each other. The overall compositional profiles of the three strains, shown in Figure 2.2, agree well with the established literature values from Neidhardt (Neidhardt, 1987). The total protein level was ~53% of dry weight for WT, which is comparable to Neidhardt's 55%. For total cell protein quantification, Neidhardt's values were assumed for the 3 unmeasured amino acids (other estimation approaches for these may be considered in the future, such as bioinformatic techniques to derive relative amino acid abundances from protein sequence data). There was slightly less protein in the knockouts, with 49% and 47% for *Δzwf* and *Δpgi*, respectively. There were a few differences in the amino acid profiles between the strains (Figure 2.3), such as elevated Glx (Glu+Gln) levels in WT compared to the knockouts. Neidhardt reported significantly higher levels of glycine, valine, isoleucine and lysine than what were measured in these *E. coli* strains. RNA was significantly reduced in the Δ*pgi* strain, at 14% of dry weight compared to 24% in the wild-type and 21% for *Δzwf*. The total fatty acids were a consistent 5% of dry weight for all three strains, slightly less than the 7% of Neidhardt. The distribution of fatty acids is shown in Figure 2.4. There was significant variability in C18:1 levels between the three strains, with the *Δzwf* strain showed elevated levels while Δ*pgi* showed decreased levels relative to the wild-type. Δ*pgi* also had less C16:1 than the

other two strains. The wild-type fatty acid levels, with the exception of C14:0 which is primarily associated with LPS, were less than the values reported by Neidhardt. This is likely due to differences in the *E. coli* strains characterized (K-12 in this study and B/r in Neidhardt).

Overall, the measured composition values for *E. coli* are in good agreement with those previously reported, while also demonstrating the importance of measuring the composition of novel phenotypes. This is often neglected in flux analysis and system biology studies (Crown and Antoniewicz, 2013b). Instead, Neidhardt's values, while summarized specifically for the *E. coli* B/r strain, are often assumed for other *E. coli* strains. Performing the biomass composition measurements presented here will therefore be useful for future systems microbiology applications, and in particular for advanced ^{13}C metabolic flux analysis studies (Crown and Antoniewicz, 2013a).

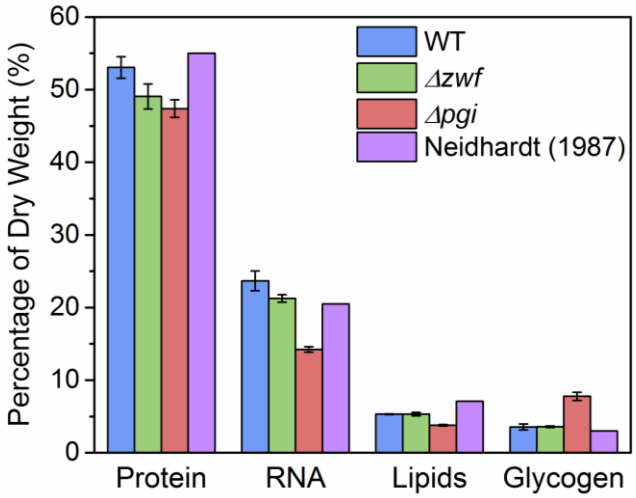


Figure 2.2: Biomass composition of three *E. coli* strains. Error bars indicate standard errors of the mean (n=4).

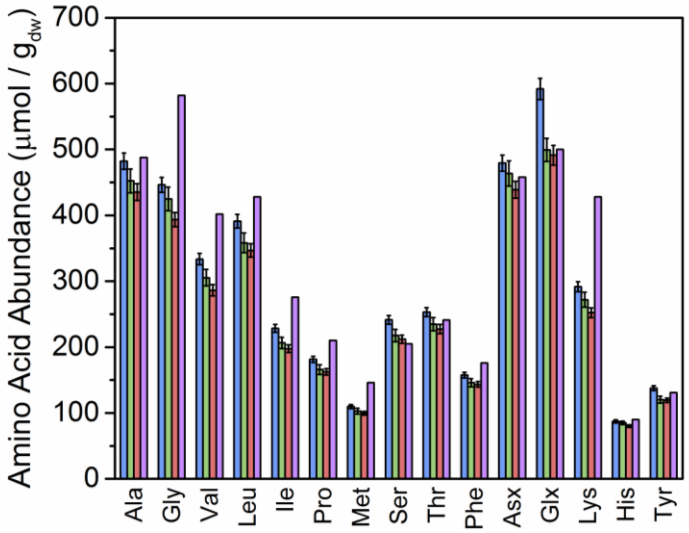


Figure 2.3: Amino acid distribution of three *E. coli* strains. The legend is the same as in Figure 2.2. Error bars indicate the standard error of the mean (n=4). "Asx" is the sum of aspartate and asparagine. "Glx" is the sum of glutamate and glutamine.

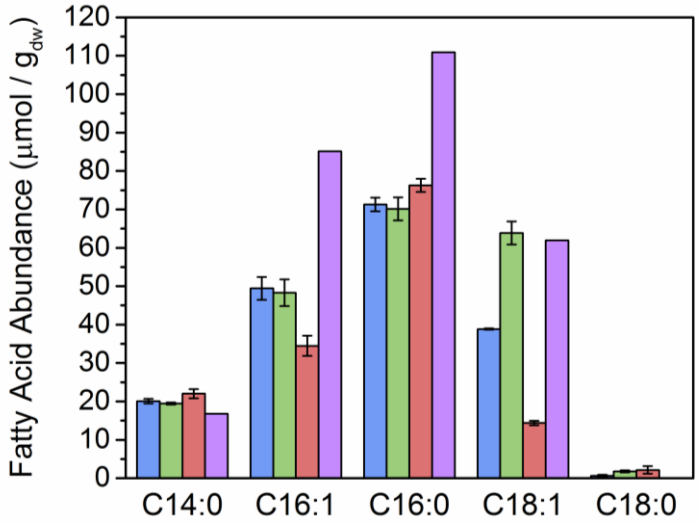


Figure 2.4 Fatty acid distribution of three E. coli strains. The key is the same as Fig. 2.2; error bars indicate standard error of the mean (n=4).

2.4 Conclusions

The methods presented here for biomass composition analysis are accurate and precise, as well as convenient and flexible. The use of a single analytical platform, GC/MS, and simple preparatory protocols means that a full biomass composition analysis (of the four major components) can be easily accomplished in many labs. The coverage of these techniques should be sufficient for most research endeavors in systems microbiology and engineering. The major components not explicitly measured here are DNA (3.1% of dry weight), lipopolysaccharide (3.4%), peptidoglycan (2.5%), and intracellular metabolites, cofactors, and ions (3.5%) (Neidhardt, 1987). In most circumstances, the amounts of these components are expected to be relatively constant (Pramanik and Keasling, 1997; Stephanopoulos et al., 1998).

The choice of fully labeled *E. coli* biomass as internal standard gives flexibility with respect to shelf-life and cell culturing concerns. For extension to other organisms, it is expected that the hydrolysis conditions used here will result in no significant differences in lysis or hydrolysis kinetics that might affect accuracy. Thus, we expect equally good performance with other classes of microbes, e.g. gram-positive bacteria and eukaryotic cells, although this should be validated prior to application. If verified, the convenience of using [U-^{13}C] labeled *E. coli* as internal standard, which is quickly and easily cultured, for analyzing more difficult organisms would offer a significant advantage.

2.4.1 Applications

Since the development of these methods, they have been applied in a number of studies in diverse organisms. These include numerous *E. coli* knockout strains

(Chapters 6 and 8) and in varied environmental conditions (Gonzalez et al., 2017), multiple species of extreme thermophiles (Cordova et al., 2017, 2016, 2015), the model algae species *Chlorella vulgaris* (Zuñiga et al., 2016), the very fast-growing bacterium *Vibrio natriegens* (Christopher P Long et al., 2017) (Chapter 12), and the model diazotroph *Azotobacter vinelandii* (unpublished). In these studies, the utility of biomass composition information was largely to define the coefficients of components and precursors in the biomass growth equation, e.g., for ^{13}C-MFA or genome-scale models. In one case, however, the composition measurements provided a key piece of fundamental insight (Yao et al., 2015). In environmental isolates from a phosphate-poor lake in Indonesia, strains were found to respond to phosphate starvation by significantly reducing their RNA content (typically ~75% of cellular phosphate). These results demonstrate the broad applicability and utility of the developed GC-MS based biomass composition measurements.

Chapter 3

MEASUREMENT OF SUGAR MONOMERS OF RNA AND GLYCOGEN INCREASE FLUX ESTMATE PRECISION IN ^{13}C-MFA

Reprinted with permission from: Long CP*, Au J*, Gonzalez JE, Antoniewicz MR (2016) ^{13}C metabolic flux analysis of microbial and mammalian systems is enhanced with GC-MS measurements of glycogen and RNA labeling. Metab. Eng. 38, 65-72. *Equal contribution

3.1 Introduction

^{13}C metabolic flux analysis (^{13}C-MFA) is a powerful tool for quantifying cellular metabolism in a wide range of metabolic engineering and biomedical applications (Maciek R. Antoniewicz, 2015a; Young, 2014). A major factor governing the accuracy and precision of flux estimates from ^{13}C-MFA is the selection of informative isotopic labeling measurements. In most ^{13}C-MFA studies to date, flux analysis has been conducted using measurements of isotopic labeling of proteinogenic amino acids and/or isotopic labeling of extracted intracellular metabolites measured with techniques such as mass spectrometry (MS), tandem mass spectrometry (MS/MS), and nuclear magnetic resonance (NMR) (Antoniewicz, 2013a; Antoniewicz et al., 2007b; Masakapalli et al., 2014; Truong et al., 2014). Recently, improvements in flux precision were also achieved by simultaneously fitting multiple sets of labeling measurements from parallel tracer experiments (Crown et al., 2015a).

In these data sets, however, metabolites derived from the upper half of central carbon metabolism (consisting of the upper portion of the glycolysis (EMP) pathway, pentose phosphate pathway (PPP), and Entner-Doudoroff (ED) pathway) are greatly

underrepresented. For example, observability of PPP fluxes in *E. coli* largely depends on accurate measurements of phenylalanine labeling, which is synthesized from erythrose 4-phosphate (E4P) and phosphoenolpyruvate (PEP). Although there are other metabolites derived from upper metabolism, e.g. histidine and serine, these measurements are less informative. In our experience, histidine is difficult to detect due to low abundance (Antoniewicz et al., 2007b), and both amino acids are related to one-carbon metabolism for which information may be incomplete. This leaves phenylalanine as the only relevant metabolite for flux analyses of PPP. In mammalian systems, flux analysis is even more challenging given that histidine and phenylalanine are not synthesized by mammalian cells, i.e. these are essential amino acids. Thus, ^{13}C-MFA depends on more distant labeling measurements, for example, of intermediates in the lower glycolytic pathway such as 3PG and PEP (Ahn and Antoniewicz, 2013). LC-MS and LC-MS/MS based approaches have provided direct labeling information on pentose phosphate pathway intermediates and fragments (Hanke et al., 2013; Rühl et al., 2012); however, these methods are often costly and laborious, as analysis of these intracellular metabolites requires rapid quenching and efficient extraction techniques, and in many cases large sample sizes due to low concentrations of these metabolites inside cells.

A recent effort to provide additional labeling data that are more directly related to upper metabolism was focused on fragments of nucleosides derived from DNA and RNA (Miranda-Santos et al., 2015). Specifically, 31 nucleoside fragments were identified and the applicability of these fragments was illustrated in yeast cultures. While this method employed GC-MS instead of LC-MS, the presented protocols required large sample sizes (1.5 to 100 mL of biomass at OD=3.5) and the labeling

data were not solely related to metabolism of PPP, but included other parts of metabolism, e.g. TCA cycle and one-carbon metabolism, which further complicated data analysis. Other research groups have measured ^{13}C-labeling of RNA, glycogen and glycan and demonstrated the value of these measurements for flux analysis (Badur et al., 2015; Guzmán et al., 2014; Murphy et al., 2013).

Our work builds on these previous studies. Specifically, we present here a convenient GC-MS based method for measuring isotopic labeling of multiple fragments of glucose and ribose derived from glycogen and RNA, respectively, for ^{13}C-MFA studies. Our approach is less sample-intensive and more informative. A key advantage of our approach over that of Miranda-Santos et al. is that the measured labeling data are directly related to PPP metabolism, which greatly simplifies data analysis and improves flux precision. We demonstrate the reliability of our approach in two biological systems: *E. coli* as a model microbial system and CHO cells as a model mammalian system. Specifically, we illustrate that isotopic labeling of glucose moiety from glycogen and ribose moiety from RNA permit precise quantification of net and exchange fluxes in PPP. We also use this approach to determine PPP fluxes during co-utilization of glucose and xylose. Overall, we demonstrate that incorporating labeling measurements of glycogen and RNA, which are stable and abundant in microbial and mammalian cells, greatly improves flux observability, thus paving the way for future applications of this approach in metabolic engineering and biomedical research.

3.2 Materials and Methods

3.2.1 Materials

Chemicals and M9 minimal medium were purchased from Sigma-Aldrich (St. Louis, MO). Isotopic tracers were purchased from Cambridge Isotope Laboratories (Tewksbury, MA): [1,6-^{13}C]glucose (99.2 % ^{13}C), [1,2-^{13}C]glucose (99.7 %), [1-^{13}C]glucose (99.5 %), [2-^{13}C]glucose (99.5 %), [3-^{13}C]glucose (99.5 %), [4,5,6-^{13}C]glucose (99.5 %), and [1,2-^{13}C]xylose (99.2 %). The isotopic enrichment of all tracers and the composition of tracer mixtures used in parallel labeling experiments were validated by GC-MS analysis as described in (Sandberg et al., 2016) and (Cordova and Antoniewicz, 2016). SFM4CHO medium (GE Healthcare Life Sciences SH3054901) and DMEM medium (Corning 17-207-CV, without glucose, glutamine, and sodium pyruvate) were purchased from Fisher Scientific (Pittsburgh, PA).

3.2.2 Strains and Culture Conditions

For *E. coli* tracer experiments, wild-type *E. coli* BW21135 and a *ΔptsG* (Keio collection) strain were used. All *E. coli* strains were purchased from GE Healthcare Dharmacon. *E. coli* was cultured aerobically in M9 minimal medium at 37°C in mini-bioreactors with 10 mL working volume as described previously (Crown et al., 2015a). For the wild-type experiments, *E. coli* cultures were inoculated at OD_{600} of 0.01. Tracers were added at the beginning of the culture. Cells were harvested (1 mL samples) for GC-MS analysis at mid-exponential growth when OD_{600} was about 0.6. For the *E. coli* Δ*ptsG* tracer experiment, M9 medium was supplemented with 50 ug/mL of kanamycin (selection marker for the knockout). The culture was inoculated at OD_{600} of 0.01, and [1,2-^{13}C]glucose and [1,2-^{13}C]xylose were added at the beginning of the culture, each at a concentration of 10 mM. Cells were harvested (1

mL samples) for GC-MS analysis at mid-exponential growth when OD_{600} was about 0.6. For CHO cell tracer experiments, CHO-K1 cells previously adapted to serum-free suspension culture were used (Valente et al., 2015). CHO cells were cultured in 15 mL of SFM4CHO/DMEM (1:1, v/v) medium supplemented with 4 mM glutamine and 1 mM sodium pyruvate, in vented 125-mL flasks (Corning 431143) in a humidified 5% CO_2 incubator at 37°C with slow shaking (~100 rpm). CHO cell cultures were inoculated at approximately 1.0×10^5 cells/mL. Tracers were added as a bolus when cell density reached 5.0×10^5 cells/mL. Cells were harvested (4 mL sample for hydrolysis of glycogen and RNA, and 6 mL sample for intracellular metabolite extraction) for GC-MS analysis after 23.25 h, when cell density was approximately 1.0×10^6 cells/mL.

3.2.3 Analytical Methods

Cell growth of E. coli cultures was monitored by measuring the optical density at 600nm (OD600) using a spectrophotometer (Eppendorf BioPhotometer). The OD600 values were converted to cell dry weight concentrations using a previously determined OD600-dry cell weight relationship for E. coli (1.0 OD600 = 0.31 gDW/L) (Long et al., 2016b). After centrifugation (5 min at 14,000 rpm), the supernatant was separated from the cell pellet, and the cell pellets were washed twice with glucose-free M9 medium. Acetate and xylose concentrations were determined using an Agilent 1200 Series HPLC (Au et al., 2014).

Cell growth of CHO cell cultures was monitored by measuring the cell concentration and cell viability using a Moxi Z Cell Counter and Moxi Z Cassettes Type S (ORFLO Technologies, Ketchum, ID). After centrifugation (2 min at 1,000 rpm), the supernatant was separated from the cell pellet, and the cell pellets were

washed twice with D-PBS (Mediatech, Inc., Manassas, VA). For intracellular metabolite extraction of CHO cells, the methanol/chloroform/water extraction method described in (Ahn and Antoniewicz, 2013) was followed. Glucose and lactate concentrations were determined using YSI 2700 biochemistry analyzer (YSI, Yellow Springs, OH).

3.2.4 Hydrolysis of Glycogen and RNA

For hydrolysis of glycogen and RNA, the following two-step hydrochloric acid hydrolysis procedure was used. First, 50 μL of 6N HCl was added to dry cell pellets and the samples were incubated for 30 min at 30 °C. Next, 250 μL of water was added (thus diluting the acid to 1 N) and samples were incubated for 60 min at 110 °C. The samples were then cooled to room temperature, neutralized with 50 μL of 5 N NaOH, and dried under air flow at 65°C. The development of this method was recently described in (McConnell and Antoniewicz, 2016).

3.2.5 Derivatization of Glucose and Ribose

Glucose and ribose released from hydrolysis of biomass (i.e. from glycogen and RNA, respectively) were derivatized using the aldonitrile propionate derivatization method described in (Antoniewicz et al., 2011). Briefly, 50 μL of 2 wt% hydroxylamine hydrochloride in pyridine was added to dried samples, which were then incubated for 60 min at 90°C. Next, 100 μL of propionic anhydride was added followed by incubation at 60°C for 30 min. The samples were then immediately transferred to injection vials for GC-MS analysis.

3.2.6 Gas Chromatography-Mass Spectrometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 µm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230°C, the MS quad temperature at 150°C, the interface temperature at 280°C, and the inlet temperature at 250°C. GC-MS analysis of *tert*-butyldimethylsilyl (TBDMS) derivatized proteinogenic amino acids was performed as described in (Leighty and Antoniewicz, 2013). GC-MS analysis of TBDMS derivatized intracellular metabolites was performed as described in (Ahn and Antoniewicz, 2011). For GC-MS analysis of glucose and ribose, 1 µL of a derivatized sample was injected at 1:2 or 1:10 split ratio. The column was started at 80°C and held for 2 min, increased to 280°C at 10°C/min, and held for 12 min. The *m/z* 173 and *m/z* 370 fragments of the glucose derivative (containing the last two and first five C-atoms of glucose, respectively (Antoniewicz et al., 2011)), and the *m/z* 173 and *m/z* 284 fragments of the ribose derivative (containing the last two and first four C-atoms of ribose, respectively (Long and Antoniewicz, 2014b)), were measured in single ion monitoring. Mass isotopomer distributions were obtained by integration (Antoniewicz et al., 2007b) and corrected for natural isotope abundances (Fernandez et al., 1996).

3.2.7 Metabolic Network Models and ^{13}C-Metabolic Flux Analysis

The metabolic network models used for ^{13}C-MFA in this study are provided in Appendix B. For *E. coli*, the full model described in (Crown et al., 2015a) was used, as well as a simplified model containing only the upper portion of central carbon

metabolism. For CHO cells, the model described in (Ahn and Antoniewicz, 2013) was used. All ^{13}C-MFA calculations were performed using the Metran software (Yoo et al., 2008), which is based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007a). Fluxes were estimated by minimizing the variance-weighted sum of squared residuals (SSR) between the experimentally measured and model predicted mass isotopomer distributions using non-linear least-squares regression (Antoniewicz et al., 2006). For integrated analysis of parallel labeling experiments, the data sets were fitted simultaneously to a single flux model as described in (Maciek R. Antoniewicz, 2015a; Leighty and Antoniewicz, 2012b). Flux estimation was repeated 10 times starting with random initial values for all fluxes to find a global solution. At convergence, accurate 68% and 95% confidence intervals were computed for all estimated fluxes by evaluating the sensitivity of the minimized SSR to flux variations (Antoniewicz et al., 2006).

To model fractional labeling of metabolites, G-value parameters were also included in ^{13}C-MFA (Maciek R. Antoniewicz, 2015b). The G-value represents the fraction of a metabolite pool that is produced during the labeling experiment, while 1-G represents the fraction that is naturally labeled, i.e. from the inoculum. By default, one G-value parameter was included for each measured metabolite in each data set. Reversible reactions were modeled as separate forward and backward fluxes. Net and exchange fluxes were determined as follows: $v_{net} = v_f - v_b$; $v_{exch} = \min(v_f, v_b)$. For visual representation of exchange fluxes, the exchange fluxes were rescaled as follows: exchange flux (%) = $100\% \times v_{exch} / (100 + v_{exch})$.

3.2.8 Goodness-of-Fit Analysis

To determine the goodness-of-fit, ^{13}C-MFA fitting results were subjected to a χ^2-statistical test. In short, assuming that the model is correct and data are without gross measurement errors, the minimized SSR is a stochastic variable with a χ^2-distribution (Antoniewicz et al., 2006). The number of degrees of freedom is equal to the number of fitted measurements *n* minus the number of estimated independent parameters *p*. The acceptable range of SSR values is between $\chi^2_{\alpha/2}(n\text{-}p)$ and $\chi^2_{1\text{-}\alpha/2}(n\text{-}p)$, where α is a certain chosen threshold value, for example 0.05 for 95% confidence interval.

3.3 Results and Discussion

3.3.1 Measuring Glycogen and RNA Labeling with GC-MS

Glycogen and RNA are abundant components of microbial and mammalian biomass. The glucose moiety of glycogen is derived from the glycolytic intermediate glucose-6-phosphate (G6P), and the ribose moiety of RNA is derived from the PPP intermediate ribose-5-phosphate (R5P). Here, we present a convenient procedure for measuring the labeling of glucose and ribose from glycogen and RNA using GC-MS for applications in ^{13}C-MFA. To observe the sugar monomers, the polymers are first hydrolyzed with hydrochloric acid and then derivatized as shown in Figure 3.1. Previously, we validated that this approach is selective for detecting the stable biomass components glycogen and RNA, rather than intracellular metabolites (Long and Antoniewicz, 2014b). For GC-MS analysis, we use the aldonitrile pentapropionate derivatization method that generates two reliable GC-MS fragments for each sugar. The first fragment, *m/z* 173, contains the last two carbons of each sugar, i.e. C5+C6 of glucose and C4+C5 of ribose. The second larger fragment contains the first five

carbons of glucose (C1-C5, *m/z* 370), and the first four carbons of ribose (C1-C4, *m/z* 284). The resulting positional labeling information provided by these fragments is critical for precise flux analysis, as is demonstrated in the next sections. The ribose and glucose peaks are clearly identifiable in the chromatograms as shown in Figures 1B and 1C for *E. coli* and CHO cells, respectively. For CHO cells, we also detected mannose and galactose peaks (Figure 3.1C), which could be useful for other applications such as glycan analysis (Badur et al., 2015).

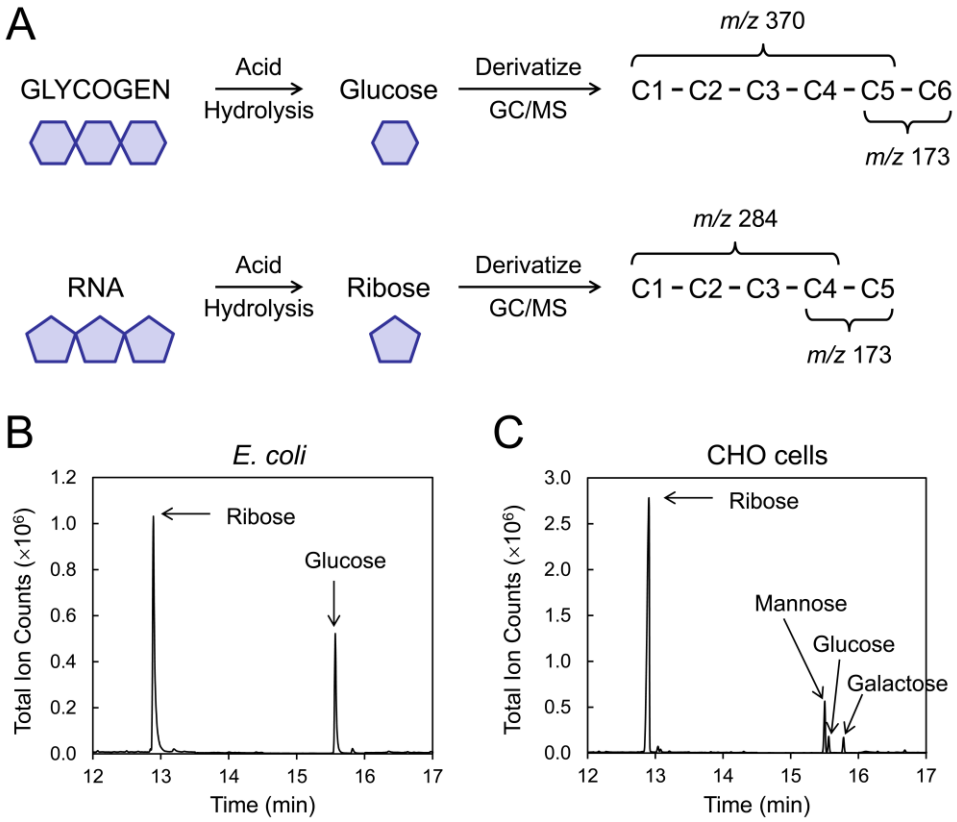


Figure 3.1: GC-MS analysis of glycogen and RNA labeling. A) The biopolymers glycogen and RNA are first broken down into the respective sugar monomers glucose and ribose by acid hydrolysis. The sugars are then subjected to aldonitrile propionate derivatization for subsequent GC-MS analysis. Two fragments of each species are measured to provide positional labeling information. B) Total ion chromatogram from GC-MS analysis of sugars from hydrolyzed *E. coli*, and C) CHO cells. Peaks corresponding to different sugar monomers are clearly resolved.

3.3.2 Glycogen and RNA Labeling Data Improve Resolution of PPP Fluxes in *E. coli*

To demonstrate that glycogen and RNA labeling data can improve flux resolution, we first applied this approach to *E. coli* as a model microbial system. ^{13}C-MFA of *E. coli* currently relies on measurements of protein-bound amino acids. To assess the complementarity of glycogen and RNA measurements with amino acid

measurements, a test case was performed using the tracer [1,6-^{13}C]glucose and exponentially growing wild-type *E. coli*. In preliminary studies, we identified [1,6-^{13}C]glucose as a promising tracer for ^{13}C-MFA with good performance throughout central carbon metabolism. After performing the tracer experiment, isotopic labeling of amino acids from hydrolyzed biomass, as well as glucose and ribose moieties of glycogen and RNA, were measured by GC-MS. Fluxes were then estimated, first using only the amino acid labeling data, and then using the amino acid data along with the glycogen and RNA data. The flux results are summarized in Figure 3.2. Acceptable fits were obtained in all cases, assuming a constant measurement error of 0.3 mol% for all GC-MS measurements, demonstrating that all data were in good agreement. Figure 3.2A shows the estimated net fluxes and Figure 3.2B shows the 68% and 95% confidence intervals of several key fluxes in central carbon metabolism. Importantly, the addition of glycogen and RNA labeling data significantly improved the precision of fluxes in upper glycolysis (e.g. PGI flux) and PPP (e.g. oxPPP, TKT, TAL fluxes). The confidence intervals of these fluxes were improved by 4-fold when glycogen and RNA measurements were included (Figure 3.2B).

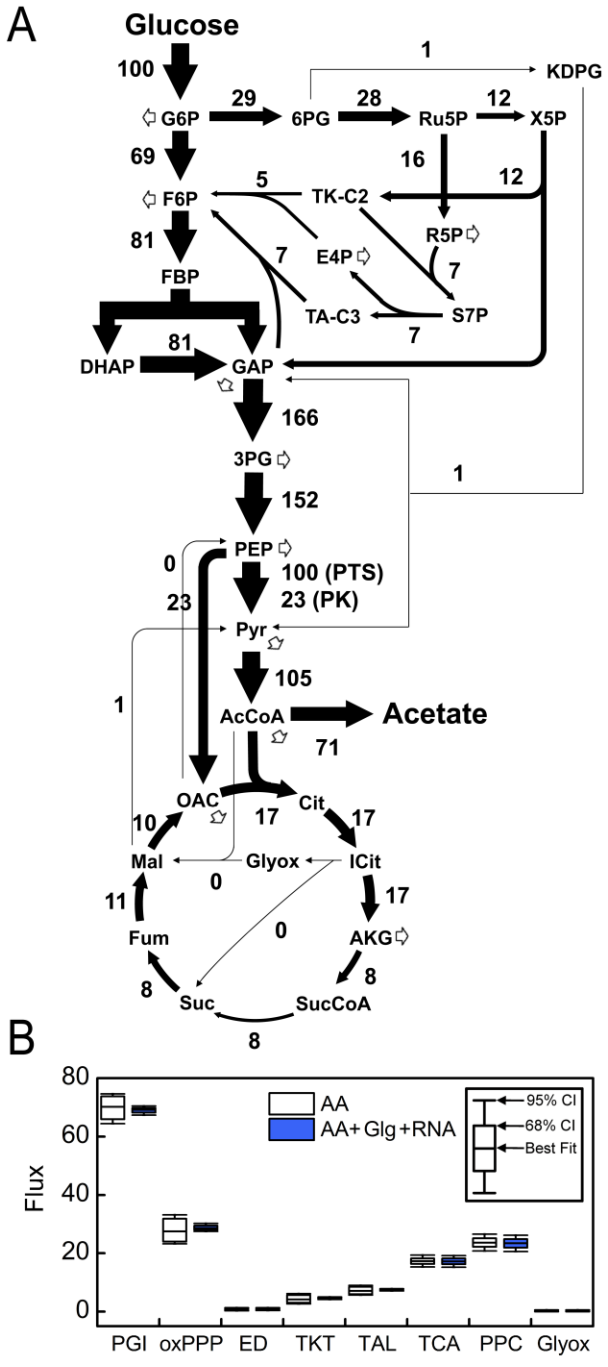

A
Glucose
100
G6P
29
6PG
28
Ru5P
12
X5P
1
KDPG
69
F6P
5
TK-C2
16
12
R5P
81
FBP
E4P
7
7
TA-C3
S7P
7
81
DHAP
GAP
166
3PG
1
152
PEP
0
23
100 (PTS)
23 (PK)
Pyr
105
1
AcCoA
Acetate
71
OAC
Cit
10
17
17
Mal
Glyox
ICit
0
11
17
Fum
0
AKG
8
8
Suc
SucCoA
8
B
Flux
80
60
40
20
0
AA
AA+ Glg +RNA
95% CI
68% CI
Best Fit
PGI
oxPPP
ED
TKT
TAL
TCA
PPC
Glyox

Figure 3.2: ^{13}C metabolic flux analysis of *E. coli* metabolism using amino acid, glycogen and RNA labeling data from a [1,6-^{13}C]glucose tracer experiment. A) Estimated flux map for *E. coli* central carbon metabolism. Fluxes were determined by simultaneously fitting amino acid (AA), glycogen (Glg), and RNA labeling data. White arrows represent outflux to biomass. B) Comparison of flux confidence intervals obtained from ^{13}C-MFA using only amino acid labeling data, and amino acid labeling data combined with glycogen and RNA data. The 68% and 95% flux confidence intervals are shown for eight representative metabolic fluxes in central carbon metabolism: phosphoglucose isomerase (v_2; PGI); oxidative pentose phosphate pathway (v_9; oxPPP); Entner-Doudoroff pathway (v_{18}; ED); transketolase (v_{14}; TKT); transaldolase (v_{16}; TAL); citrate synthase (v_{21}; TCA); PEP carboxylase (v_{33}; PPC); glyoxylate shunt (v_{29}; Glyox).

3.3.3 Estimation of Net and Exchange Fluxes in *E. coli* Upper Metabolism with Glycogen and RNA Data

Next, we evaluated if glycogen and RNA data alone (i.e. without amino acid data) could be used for reliable flux estimation in the upper half of metabolism, defined here to include the upper portion of the glycolysis (EMP) pathway, pentose phosphate pathway (PPP), and Entner-Doudoroff (ED) pathway. For this analysis we used a simplified network model shown in Figure 3.3A. Of the commonly measured amino acids, only phenylalanine could be used in this case (see Introduction section). To demonstrate the usefulness of glycogen and RNA measurements, a novel parallel tracer experiment scheme was employed here. Specifically, three mixtures of tracers were used (Ahn et al., 2016): [1-^{13}C]glucose + [4,5,6-^{13}C]glucose (1:1), [2-^{13}C]glucose + [4,5,6-^{13}C]glucose (1:1), and [3-^{13}C]glucose + [4,5,6-^{13}C]glucose (1:1). These tracers were selected based on the approaches for optimal tracer experiment design described in (Antoniewicz, 2013c; Maciek R. Antoniewicz, 2015a; Crown and Antoniewicz, 2012).

After performing the parallel labeling experiments, isotopic labeling of phenylalanine from hydrolyzed proteins, as well as the glucose and ribose moieties of glycogen and RNA, were measured by GC-MS. Flux analysis was then performed three times: first, using only phenylalanine labeling data (using fragments *m/z* 302, 308 and 336); second, using only glycogen and RNA labeling data; and third, using phenylalanine, glycogen and RNA labeling data together. All fits were statistically acceptable, assuming a constant measurement error of 0.3 mol% for all GC-MS measurements. The flux results are shown in Figure 3.3. The differences in flux precision for the different data sets are shown in Figure 3.3B, where for most reactions the glycogen and RNA measurements performed better than phenylalanine measurements alone, and the combined fits provided the most precise flux estimates. This was the case for all of the reactions except for the ED flux, where glycogen and RNA produced larger confidence intervals than phenylalanine. This is because while phenylalanine has some observability of the products of the ED reactions (i.e. PEP from GAP), glycogen and RNA are upstream and thus have little information about that flux. As shown in Figure 3.3C, the addition of glycogen and RNA measurements allowed good estimation of exchange fluxes. These exchange fluxes are notoriously difficult to estimate with amino acid labeling data alone across a wide spectrum of glucose tracers (Crown et al., 2015a). The PGI exchange flux was unobservable using only phenylalanine, but with the addition of glycogen and RNA measurements it was estimated with a narrow confidence interval. Dramatic improvements in precision were also observed for the exchange fluxes of transketolase (TKT) and transaldolase (TAL) half reactions leading to fructose-6-phosphate (F6P). Estimates were not improved in other half reactions.

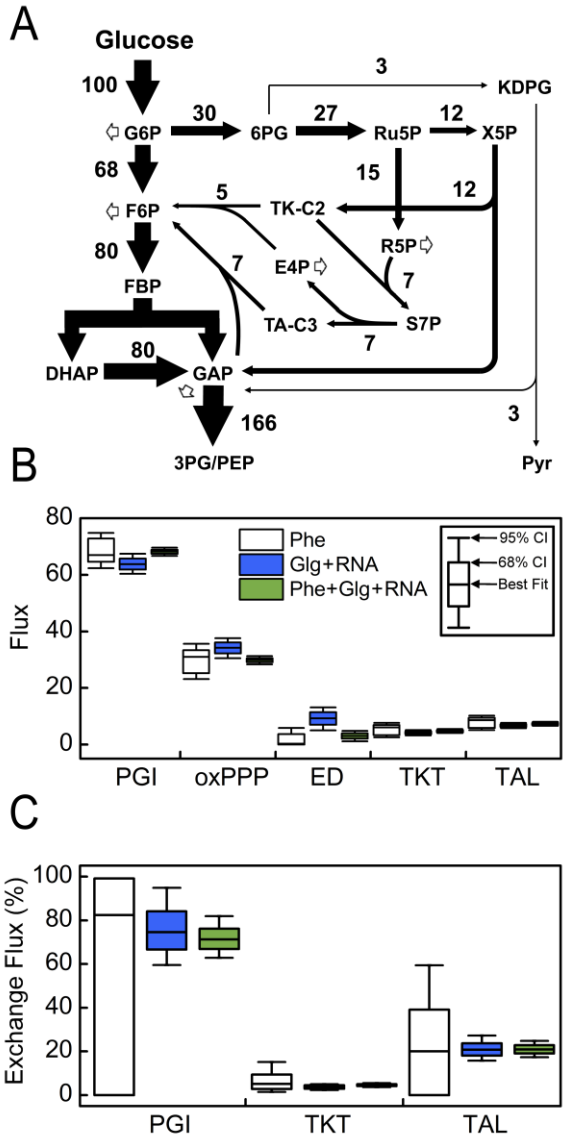

A
Glucose
100
G6P
30
6PG
3
KDPG
27
Ru5P
12
X5P
68
F6P
5
TK-C2
15
12
R5P
E4P
7
7
80
FBP
TA-C3
S7P
7
80
DHAP
GAP
166
3
3PG/PEP
Pyr
B
Flux
80
60
40
20
0
Phe
Glg+RNA
Phe+Glg+RNA
95% CI
68% CI
Best Fit
PGI
oxPPP
ED
TKT
TAL
C
Exchange Flux (%)
100
80
60
40
20
0
PGI
TKT
TAL

Figure 3.3: ^{13}C metabolic flux analysis of upper metabolism of *E. coli* using phenylalanine, glycogen and RNA labeling data from three parallel labeling experiments. A) Estimated flux map for E. coli determined by simultaneously fitting phenylalanine (Phe), glycogen (Glg), and RNA labeling data. White arrows represent outflux to biomass. B) Comparison of flux confidence intervals obtained from ^{13}C-MFA using only phenylalanine labeling data, only glycogen and RNA labeling data, and all three metabolites. The 68% and 95% flux confidence intervals are shown for five representative metabolic fluxes: phosphoglucose isomerase (v_2; PGI); oxidative pentose phosphate pathway (v_9; oxPPP); Entner-Doudoroff pathway (v_{18}; ED); transketolase (v_{14}; TKT); transaldolase (v_{16}; TAL). C) Comparison of confidence intervals for exchange fluxes estimated using different data sets. Note that in both (B) and (C), TKT and TAL refer to the terminal half reactions of transketolase and transaldolase involving fructose 6-phosphate (F6P).

3.3.4 Elucidation of Glucose and Xylose Co-Utilization Using Glycogen and RNA Data

To demonstrate a novel and practical application of our methodology, we applied it to elucidate the metabolism of *E. coli* Δ*ptsG* mutant, which has the ability to co-utilize glucose and xylose (Chiang et al., 2013; Li et al., 2007). A tracer experiment was performed using a 1:1 molar ratio of [1,2-^{13}C]glucose and [1,2-^{13}C]xylose. After performing the labeling experiment, isotopic labeling of phenylalanine, glycogen, and RNA were measured by GC-MS.

^{13}C-MFA was then performed three times as described above using different data sets. All fits were statistically acceptable, assuming a constant measurement error of 0.3 mol% for all GC-MS measurements. The flux results are shown in Figure 3.4, and the differences in flux precision for the different data sets are shown in Figure 3.4B. For ^{13}C-MFA, we did not include the measured glucose and xylose uptake rates as constraints; instead, xylose uptake rate was fixed at 100 and the relative glucose uptake rate was estimated by ^{13}C-MFA so that it could be compared to the measured

rate. Overall, including glycogen and RNA labeling data resulted in significantly more precise flux estimates, i.e. narrower confidence intervals, compared to using phenylalanine data alone (Figure 3.4B). Importantly, the addition of glycogen and RNA allowed the glucose uptake rate to be determined with high precision, which was not possible with phenylalanine data (Figure 4B). Based on ^{13}C-MFA results the glucose-to-xylose uptake ratio was 0.23 ± 0.02, which matches perfectly with the measured ratio of 0.22 (i.e. the measured glucose uptake rate was 1.93 mmol/g_{DW}/hr and the measured xylose uptake rate was 8.69 mmol/g_{DW}/hr). Previous studies on glucose and xylose co-utilization have relied on intracellular measurements and amino acid measurements (Aristilde et al., 2015; Cordova et al., 2016). This example clearly illustrates the value of adding glycogen and RNA labeling data for precise analysis of glucose and xylose co-utilization.

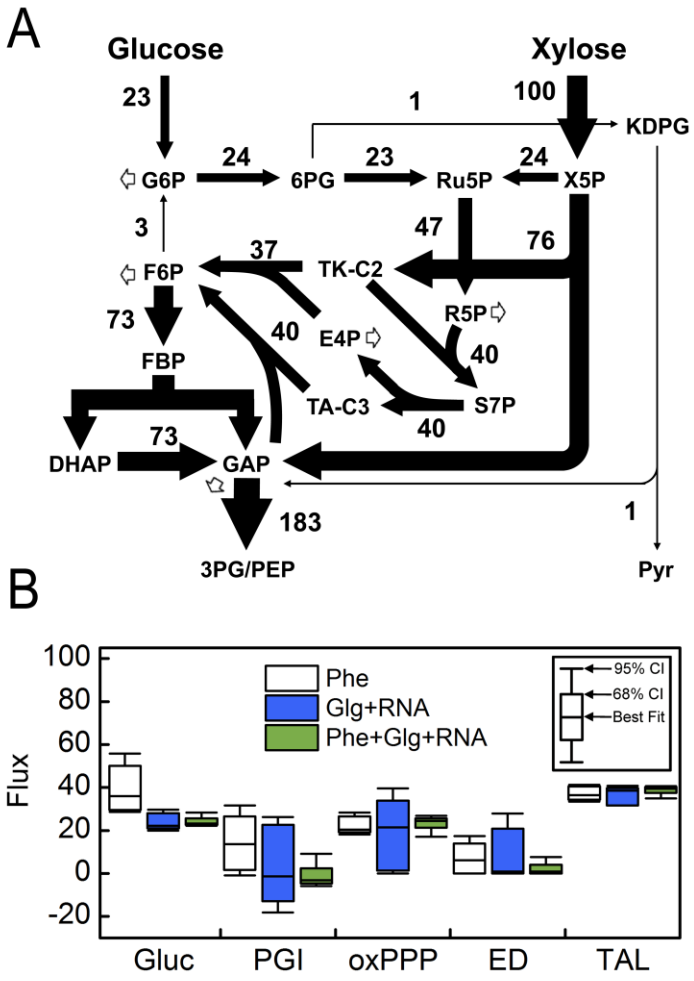


Figure 3.4: ^{13}C metabolic flux analysis of upper metabolism for *E. coli* strain *ΔptsG* that co-utilizes glucose and xylose. A) Estimated flux map determined by simultaneously fitting phenylalanine (Phe), glycogen (Glg), and RNA labeling data. White arrows represent outflux to biomass. B) Comparison of flux confidence intervals obtained from ^{13}C-MFA using only phenylalanine labeling data, only glycogen and RNA labeling data, and all three metabolites. The 68% and 95% flux confidence intervals are shown for six key metabolic fluxes: relative glucose uptake (v_1; Gluc); phosphoglucose isomerase (v_2; PGI); oxidative pentose phosphate pathway (v_9; oxPPP); Entner-Doudoroff pathway (v_{18}; ED); transketolase (v_{14}; TKT); transaldolase (v_{16}; TAL).

3.3.5 Estimation of Net and Exchange Fluxes in CHO Cell Upper Metabolism with Glycogen and RNA Data

Lastly, we applied the methodology to analyze the metabolism of CHO cells as a model mammalian system. To demonstrate the efficacy of glycogen and RNA labeling data for ^{13}C-MFA of mammalian cells, the same parallel tracer scheme was used as described in section 3.3.3. CHO cells were grown in serum-free suspension culture to the mid-exponential phase (~0.5×10^6 cells/mL), at which point tracers were introduced as a bolus. After ~24 hours, cells were harvested for analysis of 3PG and PEP labeling (intracellular metabolites), and glycogen and RNA labeling by GC-MS.

Flux analysis was then performed three times: first, using only 3PG and PEP labeling data, representing the current standard of using intracellular metabolites for ^{13}C-MFA; second, using only glycogen and RNA labeling data; and third, using 3PG, PEP, glycogen and RNA labeling data together. All fits were statistically acceptable, assuming a measurement error of 0.3 mol% for glycogen and RNA labeling measurements and a measurement error of 0.4 mol% for 3PG and PEP labeling measurements that were more noisy. The flux results are shown in Figure 3.5. As shown in Figure 3.5B, the confidence intervals of fluxes in upper metabolism were largest when fluxes were estimated with 3PG and PEP data alone, were reduced when glycogen and RNA data were used, and were greatly reduced when all measurements were used for ^{13}C-MFA. A similar trend was observed for confidence intervals of exchange fluxes (Figure 3.5C), where the precision of TKT and TAL exchange fluxes was greatly improved when all four metabolites were fitted at the same time. Together, these results demonstrate that glycogen and RNA labeling data are complementary to 3PG and PEP labeling measurements for estimating net and exchange fluxes in upper metabolism of CHO cells.

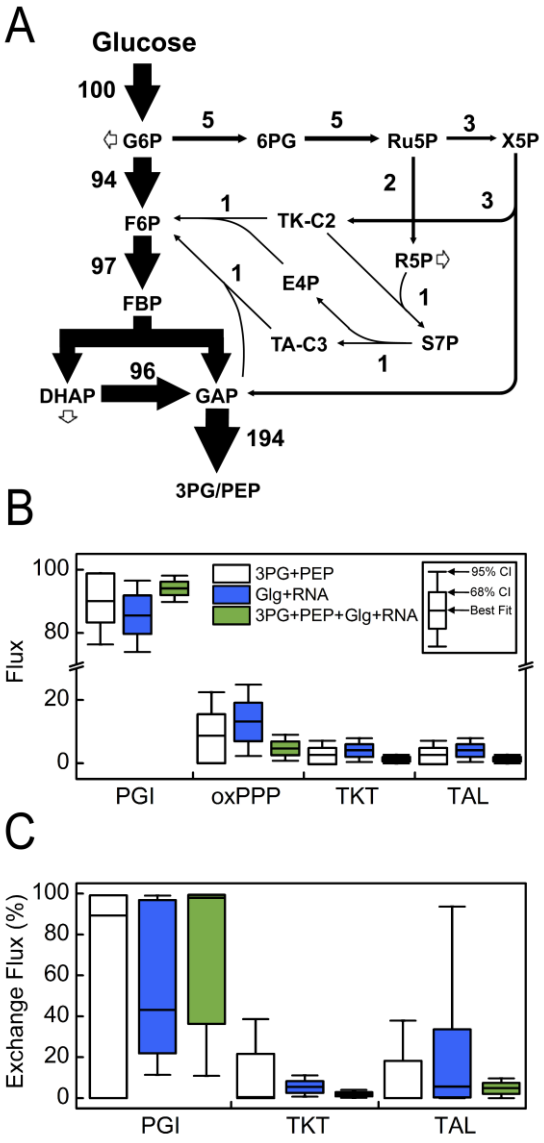

A
Glucose
100
G6P
5
6PG
5
Ru5P
3
X5P
94
2
3
F6P
1
TK-C2
R5P
97
1
E4P
1
FBP
TA-C3
S7P
1
96
DHAP
GAP
194
3PG/PEP
B
Flux
100
80
20
0
3PG+PEP
Glg+RNA
3PG+PEP+Glg+RNA
95% CI
68% CI
Best Fit
PGI
oxPPP
TKT
TAL
C
Exchange Flux (%)
100
80
60
40
20
0
PGI
TKT
TAL

Figure 3.5: ^{13}C metabolic flux analysis of upper metabolism of CHO cells using 3PG, PEP, glycogen and RNA labeling data from three parallel labeling experiments. A) Estimated flux map determined by simultaneously fitting 3PG, PEP, glycogen (Glg), and RNA labeling data. White arrows represent outflux to biomass. B) Comparison of flux confidence intervals obtained from ^{13}C-MFA using only 3PG and PEP labeling data, only glycogen and RNA labeling data, and all four metabolites. The 68% and 95% flux confidence intervals are shown for four representative metabolic fluxes: phosphoglucose isomerase (v_2; PGI); oxidative pentose phosphate pathway (v_8; oxPPP); transketolase (v_{13}; TKT); and transaldolase (v_{15}; TAL). C) Comparison of confidence intervals for exchange fluxes estimated using different data sets. Note that in both (B) and (C), TKT and TAL refer to the terminal half reactions of transketolase and transaldolase involving fructose 6-phosphate (F6P).

3.3.6 Determining Turnover Rates of Glycogen and RNA

In addition to determining intracellular fluxes, glycogen and RNA labeling data can also be used to determine turnover rates of glycogen and RNA(Murphy et al., 2013). For this purpose, the estimated G-values for glycogen and RNA are first used to calculate an apparent labeling rate for each macromolecule. This apparent labeling rate reflects the generation of new biomass (i.e. growth) as well as turnover (i.e. breakdown and regeneration) of the macromolecules. This relationship is described by:

$$\textit{Apparent labeling rate} = -\ln(1-G)/t = \textit{growth rate} + \textit{turnover rate}$$

Here, G is the estimated G-value of glycogen (or RNA) from ^{13}C-MFA, and t is the length of the labeling experiment (in this study, t = 23.25 h). To illustrate this approach for determining turnover rates of glycogen and RNA, the growth rate of CHO cells was determined directly by cell counting (Figure 6A), and the apparent labeling rates were determined from the estimated G-values (Figure 6B). Based on cell

counting, a specific cell growth rate of 0.035 ± 0.001 h^{-1} was determined, and from on the estimated G-values for glycogen (G = 0.65±0.01) and RNA (G = 0.61±0.01), apparent labeling rates of 0.045±0.001 and 0.040±0.001 h^{-1} were determined for glycogen and RNA, respectively. Thus, the turnover rate of glycogen was 1.0% per hour (i.e. 0.045 – 0.035 = 0.010 h^{-1}), and the turnover rate of RNA was 0.5% per hour (i.e. 0.040 – 0.035 = 0.005 h^{-1}).

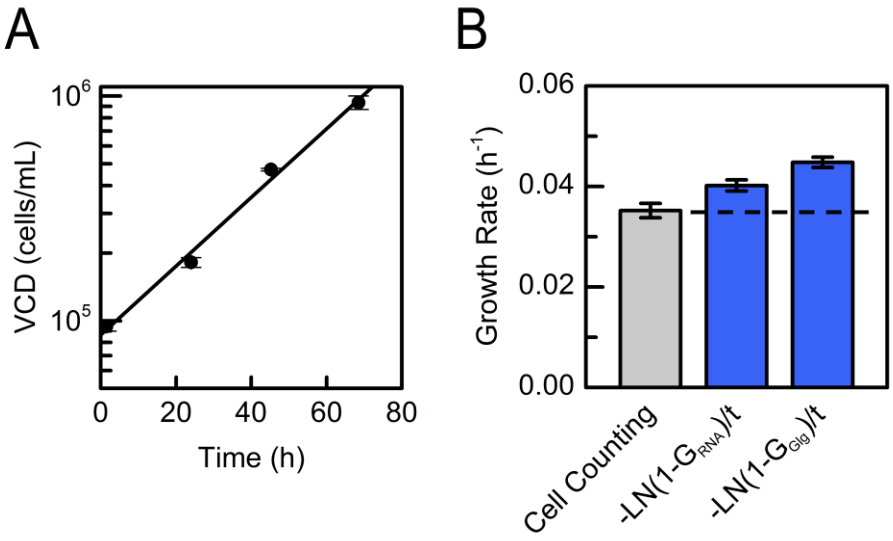


Figure 3.6: Determining turnover rates of glycogen and RNA in CHO cells. A) Time profile of viable cell density of CHO cells in suspension culture. B) Comparison of the growth rate of CHO cells (determined by cell counting) and the apparent labeling rates of glycogen (Glg) and RNA (determined from estimated G-values). The differences between the growth rate of CHO cells and the apparent labeling rates are attributable to turnover of glycogen and RNA. Error bars represent mean ± SD (*n*=3, biological replicates from parallel labeling experiments).

3.4 Conclusions

In this work, we have presented an approach for GC-MS based analysis of isotopic labeling of glycogen and RNA, and demonstrated the usefulness of these measurements for ^{13}C-MFA. Compared to alternative approaches, our method requires relatively little biomass (<0.2 mg of dry biomass for *E. coli*, and $<4\times10^6$ CHO cells), and provides multiple fragments of glucose and ribose moieties with valuable information regarding metabolic fluxes in upper metabolism, including glycolysis and pentose phosphate pathway. Additionally, we demonstrate that these measurements are complementary to other commonly used measurements for ^{13}C-MFA, including amino acids in microbial systems and intracellular metabolites in mammalian systems. We also demonstrate that glycogen and RNA labeling data are valuable for estimating precise glucose and xylose uptake rates when both substrates are co-utilized.

Beyond facilitating more precise flux estimates in combination with other common measurements, glycogen and RNA measurements provide unique observability of net and exchange fluxes in upper metabolism. In *E. coli*, these measurements are sufficient for precise quantification of fluxes in a simplified model of upper central carbon metabolism, performing better and with excellent complementarity to current measurements. In CHO cells, these measurements perform similarly to the commonly measured intracellular metabolites, which require much more laborious sample preparations and larger sample sizes. Similarly, we found strong synergy between these measurements and other commonly used measurements for ^{13}C-MFA. It is important to note that glycogen and RNA measurements allow significant improvements in the precision of exchange fluxes that are often unobservable with other commonly used measurements (Crown et al., 2015a). Given the convenience of measuring glycogen and RNA labeling and the high information

content of these measurements for resolving metabolic fluxes, we believe that the approach presented in this study will be broadly applied in future ^{13}C-MFA studies.

3.5 Author Contributions

Research was conceived of by CPL, JA, and MRA, and overseen by MRA. CPL, JEG, and JA performed the experiments and data analysis for wild-type *E. coli*, *E. coli ΔptsG*, and CHO cells, respectively. All authors contributed to writing the published article.

Chapter 4

OPTIMAL TRACERS FOR PARALLEL LABELING EXPERIMENTS AND ^{13}C METABOLIC FLUX ANALYSIS: A NEW PRECISION AND SYNERGY SCORING SYSTEM

Reprinted with permission from: Crown SB*, Long CP*, Antoniewicz MR (2016) Optimal tracers for parallel labeling experiments and ^{13}C metabolic flux analysis: A new precision and synergy scoring system. Metab. Eng. 38, 10-18.
*Equal contribution

4.1 Introduction

The significance of a judicious selection of isotopic tracers for ^{13}C metabolic flux analysis (^{13}C-MFA) has been known since the early years of ^{13}C-MFA (Follstad and Stephanopoulos, 1998; Mollney et al., 1999; Wittmann and Heinzle, 2001). In recent years, in the quest for ever increasing accuracy and precision in ^{13}C-MFA, parallel labeling experiments have emerged as the new state-of-the-art technique (Antoniewicz, 2015a; Antoniewicz, 2015b; Crown and Antoniewicz, 2013a; Crown and Antoniewicz, 2013b). This powerful flux analysis approach presents new opportunities for metabolic engineering studies; however, it also brings with it new challenges in the identification of optimal tracers. Significantly, parallel labeling experiments require careful selection of complementary tracers that take full advantage of the additional experimental effort that is required (Antoniewicz, 2013a). A bottleneck in the selection of optimal tracers is the lack of a systematic approach to evaluate the results of *in silico* simulations and *in vivo* labeling experiments to identify complementary tracers. For example, while it is relatively straightforward to determine the optimal tracer for a single flux of interest in a model, the question is

more challenging when multiple fluxes must be estimated with high precision (Crown and Antoniewicz, 2012). Often, selecting a particular tracer that improves the precision of one flux in the model results in decreased precision of another flux (Crown et al., 2012). Parallel labeling experiments offer a potential solution to this problem by tailoring specific isotopic tracers for different parts of metabolism. However, it is still not clear how optimal tracers should be selected for parallel labeling experiments.

One of the predominant methods for tracer selection and experiment design is based on a grid-search strategy combined with the use of linearized statistics. The process involves calculating the parameter covariance matrix for various isotopic tracers of interest. For this analysis the following information is needed: an assumed network model, a set of fluxes (measured or assumed), and an assumed measurement set. To compare between different tracers and determine which tracer is optimal for a given system, the D-optimality criterion is commonly applied (Mollney et al., 1999). The D-optimality criterion is related to the covariance matrix of the free fluxes and provides a measure of single parameter confidence intervals and correlations between estimated parameters. A relative information score for each tracer is then determined from the D-optimality criterion, given an assumed reference tracer experiment. The tracer scheme that produces the highest information score is then selected as the optimal tracer (Arauzo-Bravo and Shimizu, 2003; Noh and Wiechert, 2006; Yang et al., 2006). A drawback of this approach is that it inherently relies on the assumption that the underlying non-linear ^{13}C-isotopomer balances can be approximated by linearization near the optimal solution, which is not always valid (Antoniewicz et al., 2006).

Metallo et al. (Metallo et al., 2009) introduced another approach for tracer selection using a precision scoring system that captures the nonlinear behavior of ^{13}C-MFA. The suggested precision scoring method relies on calculating accurate nonlinear confidence intervals for free fluxes (Antoniewicz et al., 2006). The method computes a score for each flux, based on the upper and lower bounds for the confidence intervals and using a flux weighting parameter. If a flux has a score of zero, the flux is unidentifiable; if the score is one, the flux is optimally identifiable. The precision score is then calculated as the sum of the scores for each flux. Similar to the grid-search approach, precision scores are compared for various tracers of interest and the tracer that has the highest score is selected as optimal. More recently, Walther et al. (Walther et al., 2012) proposed a genetic algorithm for tracer selection. Despite addressing the nonlinearities of flux confidence intervals, the proposed approach potentially introduces biases due to normalization of flux confidence intervals with respect to flux values (i.e. pathways with small fluxes values such as ED pathway and glyoxylate shunt were weighted more heavily than pathways with large fluxes such as glycolysis, PP pathway and TCA cycle), as well as the method's reliance on empirically derived parameters to determine scores.

In this work, we propose a new precision scoring metric that captures the nonlinear behavior of flux confidence intervals, i.e. similar to the methods by Metallo et al. and Walther et al., but does not rely on empirically derived parameters and avoids potential biases due to flux normalization. We also propose a new synergy scoring metric that allows, for the first time, optimal tracers to be selected for parallel labeling experiments. Through the use of these two new scoring metrics we have

identified new optimal tracers for high-resolution ^{13}C-MFA studies and have validated the performance of these tracers experimentally.

4.2 Theory

4.2.1 The Precision Score

Here, we propose the following precision scoring metric to evaluate the precision of estimated fluxes:

$$P = \frac{1}{n}\sum_{i=1}^{n} p_i$$

with

$$p_i = \left(\frac{\left(UB_{95,i} - LB_{95,i}\right)_{ref}}{\left(UB_{95,i} - LB_{95,i}\right)_{\exp}}\right)^2$$

The precision score (P) for a given tracer experiment is calculated as the average of individual flux precision scores (p_i) for n number of fluxes of interest. An individual flux precision score is calculated as the squared ratio of the 95% flux confidence interval obtained for a reference tracer experiment ("ref") relative to the tracer experiment that is being evaluated ("exp"). An individual flux precision score is thus roughly equivalent to the fold-improvement in flux variance relative to the reference tracer experiment, i.e. in linear statistics: variance = (standard deviation)^2. An individual flux precision score of 1.0 indicates that the flux precision for a particular tracer experiment is the same as for the reference tracer experiment. A precision score greater than one is desirable, as this means that the tracer experiment results in a narrower confidence interval compared to the reference experiment. Since individual flux precision scores can vary from flux to flux, i.e. a tracer experiment

may perform better than the reference for some fluxes and worse for other fluxes, the ultimate gauge of tracer performance is the total precision score for the fluxes of interest. A precision score greater than one indicates that the tracer experiment on a whole outperforms the reference tracer experiment. The larger the precision score, the more substantial the increase in flux precision is compared to the reference tracer experiment. To make sure that a handful precision scores don't dominate the total precision score, we set a maximum value of 9 for any individual precision score when evaluating single tracer experiments (and 30 for parallel labeling experiments).

If deemed necessary, the precision score can be tailored further by applying different weighting factors, w_i, for different fluxes of interest:

$$P = \left(\sum_{i=1}^{n} w_i * p_i\right) \Big/ \left(\sum_{i=1}^{n} w_i\right)$$

4.2.2 The Synergy Score

In addition to the new precision scoring metric described above, we also propose a new synergy scoring metric that quantifies the increase in flux information obtained as a result of conducting multiple parallel labeling experiments and simultaneously fitting the data for ^{13}C-MFA:

$$S = \frac{1}{n}\sum_{i=1}^{n} s_i$$

with

$$s_i = \frac{p_{i,1+2}}{p_{i,1} + p_{i,2}}$$

The synergy score (S) is calculated as the average of individual flux synergy scores (s_i). An individual flux synergy score is calculated by dividing the precision

score for the parallel labeling experiment, denoted by $p_{i,1+2}$, by the sum of the precision scores for the respective individual experiments, $p_{i,1}$ and $p_{i,2}$. Note that this definition of the synergy score can be expanded for more than two parallel labeling experiments by adding additional terms in the denominator.

An individual flux synergy score of 1.0 indicates that no additional information is gained by fitting multiple parallel labeling experiments simultaneously. Intuitively, by performing *N* number of parallel labeling experiments the precision score is expected to increase by about *N*-fold. For non-linear problems such as ^{13}C-MFA, the synergy score can be smaller than one, or greater than one. A synergy score greater than 1.0 indicates a greater than expected gain in flux information, while a synergy score of 1.0 or less indicates a smaller than expected improvement in flux precision. As with the precision scores, synergy scores can vary from flux to flux. For parallel labeling experiments it is desirable to have a total synergy greater than one, as this indicates that the global flux resolution is improved synergistically through the use of complementary tracers.

Note that the synergy score can also be expressed as:

$$S = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(UB_{95,i} - LB_{95,i}\right)_{1+2}^{-2}}{\left(UB_{95,i} - LB_{95,i}\right)_{1}^{-2} + \left(UB_{95,i} - LB_{95,i}\right)_{2}^{-2}}$$

From the above equation it is clear that the synergy score is independent of the reference tracer experiment. Similar to the precision score, the synergy score can be tailored further by applying different weighting factors, w_i, for different fluxes of interest:

$$S = \left(\sum_{i=1}^{n} w_i * s_i\right) \Big/ \left(\sum_{i=1}^{n} w_i\right)$$

4.3 Materials and Methods

4.3.1 Materials

Media and chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Tracers were purchased from Cambridge Isotope Laboratories: [1-^{13}C]glucose (99.6 atom% ^{13}C), [1,2-^{13}C]glucose (99.8%), [1,6-^{13}C]glucose (99.2%), and [U-^{13}C]glucose (99.3%). M9 minimal medium was used for all tracer experiments. All solutions were sterilized by filtration.

4.3.2 Strain and Growth Conditions

E. coli BW25113 (GE Healthcare Dharmacon OEC5042) was used in this study. Four parallel labeling experiments were performed with 2 g/L of [1,2-^{13}C]glucose; [1,6-^{13}C]glucose; 51.5% [1,2-^{13}C]glucose + 48.5% [1,6-^{13}C]glucose; and 81% [1-^{13}C]glucose + 19% [U-^{13}C]glucose, as described previously (Crown et al., 2015a). The isotopic purity of glucose tracers and the composition of glucose tracer mixtures was validated by GC-MS. Cells were grown in aerated mini-bioreactors with a working volume of 10 mL at 37°C (Leighty and Antoniewicz, 2013). The cultures were inoculated from the same pre-culture that was grown overnight in a shaker flask at 37°C.

4.3.3 Analytical Methods

Samples were collected during the exponential growth phase to monitor cell growth and glucose consumption. Cell growth was monitored by measuring the optical density at 600nm (OD600) using a spectrophotometer (Eppendorf BioPhotometer). The OD600 values were converted to cell dry weight concentrations using a pre-determined OD600-dry cell weight relationship for E. coli (1.0 OD600 = 0.32 gDW/L)

(Long et al., 2016b). After centrifugation, the supernatant was separated from the biomass pellet and glucose concentration was measured with a YSI 2700 biochemistry analyzer (YSI, Yellow Springs, OH). Acetate was measured by HPLC (Au et al., 2014).

4.3.4 Gas Chromatography-Mass Spectrometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 µm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230°C, the MS quad temperature at 150°C, the interface temperature at 280°C, and the inlet temperature at 250°C. GC-MS analysis of *tert*-butyldimethylsilyl (TBDMS) derivatized proteinogenic amino acids was performed as described in (Long and Antoniewicz, 2014b). Labeling of glucose was determined using the aldonitrile propionate derivatization method described in (Antoniewicz et al., 2011).

4.3.5 Metabolic Network Model and ^{13}C-Metabolic Flux Analysis

The metabolic network model used for ^{13}C-MFA was described previously (Crown et al., 2015a), and is shown in Appendix Table B.1 (i.e., same as full model used in Chapter 3). The model includes all major metabolic pathways of central carbon metabolism, lumped amino acid biosynthesis reactions, and a lumped biomass formation reaction. The model also accounts for dilution of intracellular labeling from incorporation of unlabeled CO_2 (Leighty and Antoniewicz, 2012). All simulations and ^{13}C-MFA calculations were performed using the Metran software (Yoo et al., 2008)

which is based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007a). Fluxes were estimated by minimizing the variance-weighted sum of squared residuals (SSR) between the measured and model predicted mass isotopomer distributions using non-linear least-squares regression (Antoniewicz et al., 2006). For integrated analysis of parallel labeling experiments, the data sets were fitted simultaneously to a single flux model as described in (Leighty and Antoniewicz, 2013). Flux estimation was repeated 10 times starting with random initial values for all fluxes to find a global solution.

Three methods were used to calculate 95% confidence intervals of fluxes. The first method, described in (Antoniewicz et al., 2006), calculates accurate nonlinear 95% confidence intervals by evaluating the sensitivity of the minimized SSR to flux variations. The second method is based on Monte Carlo simulations, where random errors from a normal distribution (here we assumed 0.4 mol% measurement errors for all GC-MS measurements) are introduced and flux estimation is repeated with the corrupted data sets. In this work, we have performed 1,000 Monte Carlo simulations to determine 95% confidence intervals of fluxes. The third method is based on linearized statistics, where 95% confidence intervals of fluxes are obtained from the parameter covariance matrix (Mollney et al., 1999).

To model fractional labeling of biomass amino acids G-value parameters were also included in ^{13}C-MFA. As described previously (Antoniewicz et al., 2007b), the G-value represents the fraction of a metabolite pool that is produced during the labeling experiment, while 1-G represents the fraction that is naturally labeled (e.g. from inoculum). By default, one G-value parameter was included for each measured amino acid in each data set. Reversible reactions were modeled as separate forward

and backward fluxes. Net and exchange fluxes were determined as follows: $v_{net} = v_f - v_b$; $v_{exch} = \min(v_f, v_b)$. To determine the goodness-of-fit, ^{13}C-MFA fitting results were subjected to a χ^2-statistical test. In short, assuming that the model is correct and data are without gross measurement errors, the minimized SSR is a stochastic variable with a χ^2-distribution (Antoniewicz et al., 2006). The number of degrees of freedom is equal to the number of fitted measurements *n* minus the number of estimated independent parameters *p*. The acceptable range of SSR values is between $\chi^2_{\alpha/2}(n-p)$ and $\chi^2_{1-\alpha/2}(n-p)$, where α is a certain chosen threshold value, e.g. 0.05 for 95% confidence interval.

4.3.6 Assumption of No Kinetic Isotope Effect

A common assumption in ^{13}C-MFA is that there is no ^{13}C kinetic isotope effect; essentially, that transporters and enzymes don't discriminate between ^{12}C and ^{13}C (Feng and Tang, 2011). The assumption of no kinetic isotope effect was also applied in this study for all tracers. Support for this assumption comes from several studies. Sandberg et al. (Sandberg et al., 2016), for example, demonstrated that there was no measurable difference in the uptake of ^{12}C glucose and ^{13}C glucose by wild-type *E. coli* and several evolved *E. coli* strains. Additionally, massively parallel labeling experiments have been used to test this assumption. In a recent study by Crown et al. (Crown et al., 2015a), 14 parallel labeling experiments were successfully combined into one global flux solution. This would not be possible if the different tracers had caused a significant change in metabolism. Thus, based on best available methods, the assumption of no kinetic isotope effect appears to be valid for ^{13}C-MFA.

4.4 Results and Discussion

4.4.1 Evaluation of Single Glucose Tracers

First, we performed *in silico* simulations to evaluate the performance of 19 commercially available glucose tracers (Table 4.1). We also evaluated the performance of two commonly used glucose tracer mixtures: 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose (which was selected as the reference tracer experiment in this study), and 20% [U-^{13}C]glucose + 80% natural glucose. These two tracer schemes are widely used because of the relatively low cost of the tracers involved, i.e. [1-^{13}C]glucose (~$100/g) and [U-^{13}C]glucose (~$200/g), compared to the cost of other glucose tracers (Table 4.1).

Flux precision for each glucose tracer was determined as follows: 1) for each tracer, GC-MS measurements of proteinogenic amino acids were simulated using a previously determined flux map for wild-type *E. coli* (Crown et al., 2015a); 2) ^{13}C-MFA was performed on the simulated data. Glucose influx was fixed at 100 and acetate yield was 70 ± 5 mol/mol (Long et al., 2016b). No other external constrains were imposed. A constant measurement error of 0.4 mol% was assumed for all GC-MS measurements; 3) 95% confidence intervals of fluxes were determined using three different methods: i) using the method described in (Antoniewicz et al., 2006), which produces accurate nonlinear confidence intervals by evaluating the sensitivity of SSR to flux variations; ii) using 1000 Monte Carlo simulations; and iii) using linearized statistics that approximate 95% confidence intervals at the optimal solution; 4) Precision scores were calculated as the average of individual precision scores for the following eight key fluxes in central carbon metabolism: upper glycolysis (v_2, G6P $\rightarrow$ F6P), oxidative pentose phosphate pathway (v_{10}, 6PG $\rightarrow$ Ru5P + CO_2), non-oxidative

pentose phosphate pathway (v_{14}, F6P $\rightarrow$ E4P + E-C2), Entner–Doudoroff pathway (v_{18}, 6PG $\rightarrow$ KDPG), TCA cycle (v_{21}, AcCoA + OAC $\rightarrow$ Cit), glyoxylate shunt (v_{29}, ICit $\rightarrow$ Glyox + Suc), cataplerosis (v_{31}, Mal $\rightarrow$ Pyr + CO_2), and gluconeogenesis (v_{34}, OAC $\rightarrow$ PEP + CO_2).

Table 4.1**:** 19 commercially available ^{13}C-glucose tracers.

Glucose tracers	Abbreviation	List Price ($/g)*
[1-^{13}C]Glucose	[1]Gluc	$87 (Isotec)
[2-^{13}C]Glucose	[2]Gluc	$200 (Omicron)
[3-^{13}C]Glucose	[3]Gluc	$1200 (Omicron)
[4-^{13}C]Glucose	[4]Gluc	$1600 (Omicron)
[5-^{13}C]Glucose	[5]Gluc	$1700 (Omicron)
[6-^{13}C]Glucose	[6]Gluc	$700 (Omicron)
[1,2-^{13}C]Glucose	[12]Gulc	$650 (Omicron)
[1,3-^{13}C]Glucose	[13]Gluc	$1700 (Omicron)
[1,6-^{13}C]Glucose	[16]Gluc	$1500 (Omicron)
[2,3-^{13}C]Glucose	[23]Gluc	$1800 (Omicron)
[2,5-^{13}C]Glucose	[25]Gluc	$2600 (Omicron)
[3,4-^{13}C]Glucose	[34]Gluc	$3200 (Omicron)
[4,5-^{13}C]Glucose	[45]Gluc	$2550 (Omicron)
[4,6-^{13}C]Glucose	[46]Gluc	$9760 (Omicron)
[5,6-^{13}C]Glucose	[56]Gluc	$2600 (Omicron)
[1,2,3-^{13}C]Glucose	[123]Gluc	$1700 (Omicron)
[4,5,6-^{13}C]Glucose	[456]Gluc	$3200 (Omicron)
[2,3,4,5,6-^{13}C]Glucose	[23456]Gluc	$9400 (Omicron)
[U-^{13}C]Glucose	[U]Gluc	$195 (Isotec)

* Shown is the lowest listed price per gram of tracer on 3/27/2016.

* Omicron, http://www.omicronbio.com/index.html

* Isotec, http://www.sigmaaldrich.com/chemistry/stable-isotopes-isotec.html

* Cambridge Isotope Laboratories, http://www.isotope.com

Figure 4.1 shows the calculated standard deviations of key fluxes in central carbon metabolism for the different glucose tracers, and Figure 4.2 shows the calculated precision scores, sorted from the best to the worst performing glucose tracer. Tracers with a precision score greater than 1.0 performed better than the reference tracer experiment, 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose; tracers with a precision score smaller than 1.0 performed worse than the reference tracer experiment. Figures 4.2A, 4.2B and 4.2C compare the precision scores determined using the three different methods for determining 95% confidence intervals of fluxes. Overall, there was excellent agreement between the accurate nonlinear confidence intervals method (Figure 4.2A) and the Monte Carlo simulations method (Figure 4.2B). In contrast, linearized statistics produced significantly different precision scores (Figure 4.2C). For example, the first two methods determined that the reference tracer experiment was one of the worst performing tracers. Both methods determined that the mixture 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose performed worse than 100% [1-^{13}C]glucose, which was recently validated experimentally (Crown et al., 2015a). In contrast, the linearized statistics method predicted the opposite result. Overall, this method overestimated the performance of tracer mixtures. It is important to note that nearly all studies thus far have relied on linearized statistics to identify optimal tracers. Consistent with our simulation results, in many cases mixtures of tracers have been predicted to perform better than pure glucose tracers (Arauzo-Bravo and Shimizu, 2003; Mollney et al., 1999). Here, we demonstrate that using linearized statistics may not be appropriate for tracer selection. In the remainder of this study, we used the accurate nonlinear confidence intervals method by Antoniewicz et al. (2006) to determine 95% confidence intervals of fluxes, since this method produced the same

results as the Monte Carlo simulation method, but was computationally much more efficient.

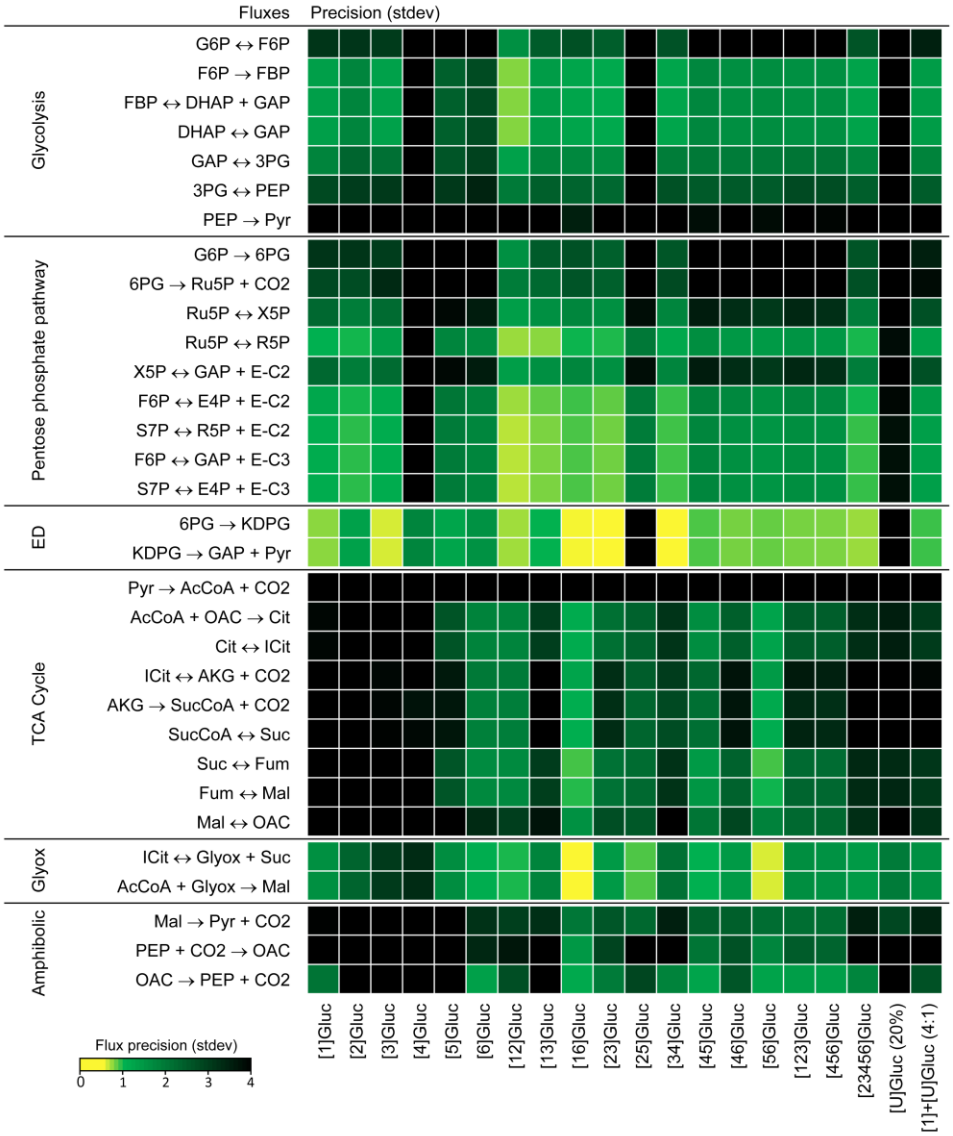


Figure 4.1: Precision of estimated fluxes in central carbon metabolism obtained with ^{13}C-MFA for 19 commercially available glucose tracers and two common glucose tracer mixtures, 20% [U-13C]glucose; and 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose. ^{13}C-MFA was performed using simulated GC-MS data.

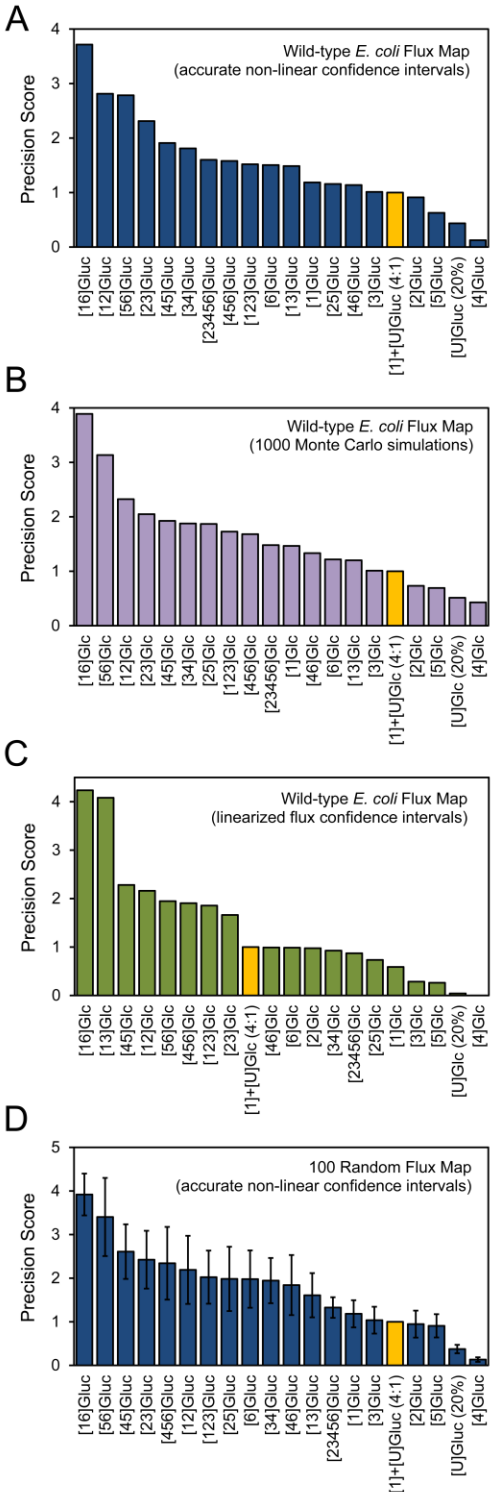

A
Wild-type E. coli Flux Map
(accurate non-linear confidence intervals)
Precision Score
[16]Gluc
[12]Gluc
[56]Gluc
[23]Gluc
[45]Gluc
[34]Gluc
[23456]Gluc
[456]Gluc
[123]Gluc
[6]Gluc
[13]Gluc
[1]Gluc
[25]Gluc
[46]Gluc
[3]Gluc
[1]+[U]Gluc (4:1)
[2]Gluc
[5]Gluc
[U]Gluc (20%)
[4]Gluc
B
Wild-type E. coli Flux Map
(1000 Monte Carlo simulations)
Precision Score
[16]Glc
[56]Glc
[12]Glc
[23]Glc
[45]Glc
[34]Glc
[25]Glc
[123]Glc
[456]Glc
[23456]Glc
[1]Glc
[46]Glc
[6]Glc
[13]Glc
[3]Glc
[1]+[U]Glc (4:1)
[2]Glc
[5]Glc
[U]Glc (20%)
[4]Glc
C
Wild-type E. coli Flux Map
(linearized flux confidence intervals)
Precision Score
[16]Glc
[13]Glc
[45]Glc
[12]Glc
[56]Glc
[456]Glc
[123]Glc
[23]Glc
[1]+[U]Glc (4:1)
[46]Glc
[6]Glc
[2]Glc
[34]Glc
[23456]Glc
[25]Glc
[1]Glc
[3]Glc
[5]Glc
[U]Glc (20%)
[4]Glc
D
100 Random Flux Map
(accurate non-linear confidence intervals)
Precision Score
[16]Gluc
[56]Gluc
[45]Gluc
[23]Gluc
[456]Gluc
[12]Gluc
[123]Gluc
[25]Gluc
[6]Gluc
[34]Gluc
[46]Gluc
[13]Gluc
[23456]Gluc
[1]Gluc
[3]Gluc
[1]+[U]Gluc (4:1)
[2]Gluc
[5]Gluc
[U]Gluc (20%)
[4]Gluc

Figure 4.2: Precision scores for 19 commercially available glucose tracers and two common glucose tracer mixtures, 20% [U-^{13}C]glucose; and 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose. The reference tracer experiment, 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose, is highlighted in yellow. Three different methods were used to calculate 95% confidence intervals of fluxes. (A) Accurate nonlinear 95% confidence intervals of fluxes were determined using the method described in (Antoniewicz et al., 2006). ^{13}C-MFA was performed using fluxes for wild-type *E. coli* to simulate isotopic labeling. (B) 1000 Monte Carlo simulations were used to determine 95% confidence intervals of fluxes. (C) Linearized statistics were used to determine 95% confidence intervals of fluxes. (D) Average precision scores for 100 random flux maps. The random flux maps captured a wide range of possible flux scenarios. Accurate nonlinear 95% confidence intervals of fluxes were determined using the method described in (Antoniewicz et al., 2006).

Of the 19 commercially available glucose tracers, 14 performed better than the reference tracer experiment (Figure 4.2A). The highest precision score of 3.7 was obtained for [1,6-^{13}C]glucose, which was particularly good at determining fluxes of Entner–Doudoroff pathway and glyoxylate shunt (Figure 4.1). Interestingly, the top six best performing tracers were all doubly labeled glucose tracers: [1,6-^{13}C]glucose (precision score = 3.7), [1,2-^{13}C]glucose (precision score = 2.8), [5,6-^{13}C]glucose (precision score = 2.8), [2,3-^{13}C]glucose (precision score = 2.3), [4,5-^{13}C]glucose (precision score = 1.9), and [3,4-^{13}C]glucose (precision score = 1.8). Doubly labeled tracers have not been widely used for ^{13}C-MFA, with the exception of [1,2-^{13}C]glucose (Ahn and Antoniewicz, 2013; Crown and Antoniewicz, 2013b; Murphy et al., 2013; Walther et al., 2012). The tracers that performed the worst were 20% [U-^{13}C]glucose (precision score = 0.4) and [4-^{13}C]glucose (precision score = 0.1).

4.4.2 Optimal Tracers are not Sensitive to Flux Values

In theory, different tracers could be optimal for different flux maps. To evaluate the sensitivity of optimal tracer selection with respect to flux values, we generated 100 random flux maps and repeated the analysis described above. The results are summarized in Figure 4.2D. The 100 random flux maps captured a wide range of possible flux scenarios, with the glycolysis flux ranging from 7 to 98 (normalized to glucose uptake of 100), the oxidative pentose phosphate flux ranging from 0 to 92, the Entner–Doudoroff pathway flux ranging from 0 to 38, the TCA cycle flux ranging from 0 to 64, the glyoxylate shunt flux ranging from 0 to 43, and the acetate secretion flux ranging from 0 to 99. We found that the precision scores calculated for wild-type *E. coli* flux map (Figure 4.2A) and 100 random flux maps (Figure 4.2D) were very similar, thus suggesting that optimal tracer selection doesn't depend strongly on the actual flux values used for tracer selection. For example, in both cases: 1) [1,6-^{13}C]glucose was determined to be the best tracer; 2) doubly labeled glucose tracers produced the highest precision scores; 3) the reference tracer experiment with 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose performed very poorly overall; and 4) 20% [U-^{13}C]glucose and [4-^{13}C]glucose were the two worst performing tracers.

4.4.3 Evaluation of Mixtures of Glucose Tracers

Next, we evaluated if mixtures of glucose tracers would perform better than pure glucose tracers. Note that the reference tracer experiment is an example of a mixture of glucose tracers: 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose. In a previous study, we experimentally evaluated four different mixtures of glucose tracers: [1-^{13}C]glucose and [U-^{13}C]glucose (1:1 mixture), [1-^{13}C]glucose and [U-^{13}C]glucose (4:1

mixture), [1-^{13}C]glucose and [4,5,6-^{13}C]glucose (1:1 mixture), and [U-^{13}C]glucose and unlabeled glucose (1:4 mixture), in addition to a large number of pure glucose tracers (Crown et al., 2015a). Experimentally, we found that pure glucose tracers performed better than mixtures of glucose tracers. Here, we wanted to determine if this observation could be generalized. Thus, we evaluated all possible dual mixtures of glucose tracers, as well as mixtures of all individual glucose tracers and unlabeled glucose. For all mixtures, nine different ratios were evaluated: 10/90, 20/80, 30/70, 40/60, 50/50, 60/40, 70/30, 80/20, and 90/10.

The results of this extensive analysis are summarized in Figure 4.3. Overall, we found that in the vast majority of cases pure tracers performed better than mixtures of tracers. In fact, mixing different tracers often resulted in a significantly lower precision score. For example, a 50/50 mixture of [1,2-^{13}C]glucose and [1,6-^{13}C]glucose produced a precision score that was less than half of the precision scores of the respective pure tracers. There were only 6 cases where a mixture of tracers performed slightly better than the respective pure tracers (highlighted in red in Figure 3); however, in none of these cases did the precision score approach that of the best pure glucose tracers identified in the previous section. Thus, based on this exhaustive analysis we can conclude that pure glucose tracers in general perform better than mixtures of tracers for ^{13}C-MFA in *E. coli*.

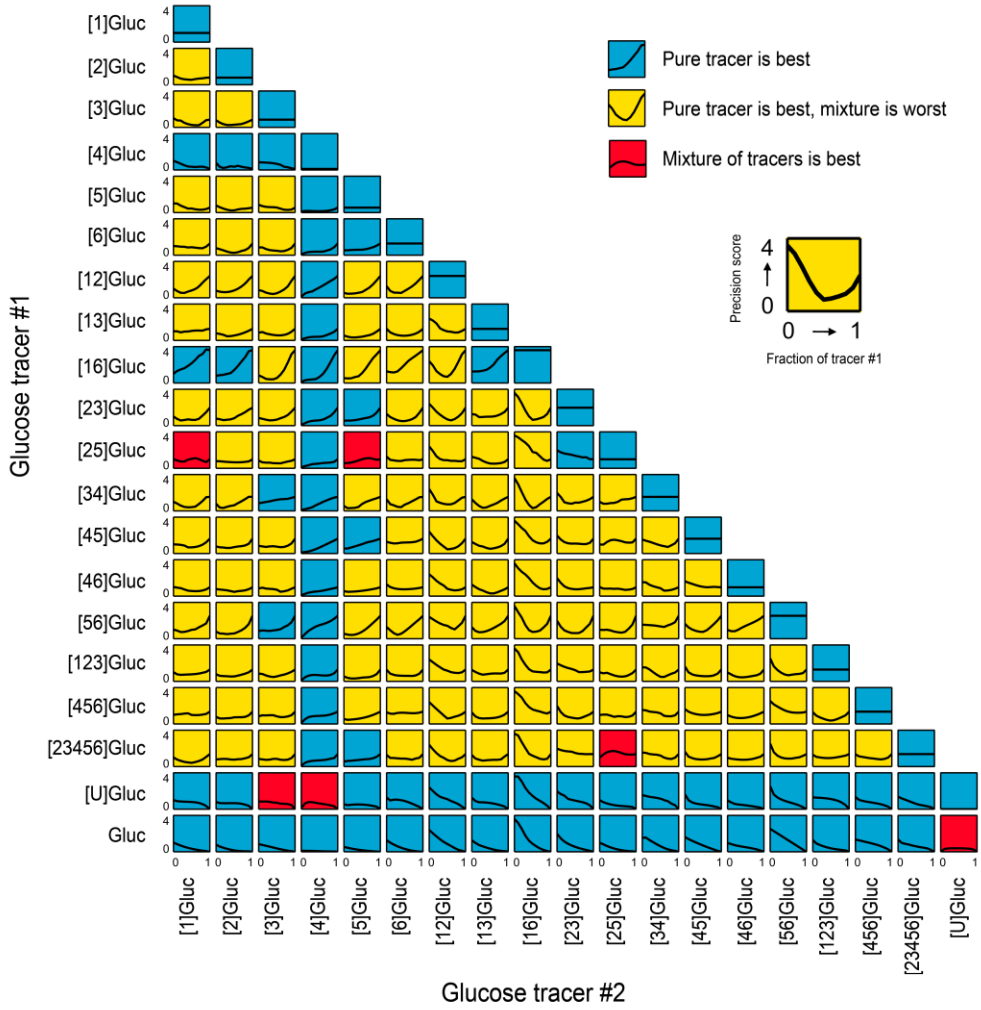


Figure 4.3: Precision scores for mixtures of glucose tracers. For each combination of two glucose tracers, nine mixing ratios were evaluated, ranging from 10%/90% to 90%/10%. Blue squares correspond to cases were the precision score monotonically increased (or decreased) with respect to the mixing ratio. Yellow squares correspond to cases where mixing pure tracers resulted in a significantly reduced precision score. Red squares correspond to cases where mixing of tracers resulted in an improved precision score compared to pure tracers. ^{13}C-MFA was performed using simulated GC-MS data and assuming wild-type *E. coli* fluxes.

4.4.4 Identification of Optimal Tracers for Parallel Labeling Experiments

Next, we set out to identify optimal tracers for parallel labeling experiments. We examined all possible combinations of two parallel labeling experiments with the 19 commercially available glucose tracers. Mixtures of tracers were not included in this analysis based on the results in the previous section. For consistency, the tracer experiment with 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose was used as the reference experiment. The calculated precision scores and synergy scores are shown in Figure 4.4. The highest precision score was obtained for parallel labeling experiments with [1,6-^{13}C]glucose and [1,2-^{13}C]glucose, which had a precision score of 7.8. This precision score was higher than the sum of the precision scores for the respective single pure tracers, i.e. 3.7 for [1,6-^{13}C]glucose and 2.8 for [1,2-^{13}C]glucose. Thus, there was significant synergy in the use of these two tracers. In this case, the synergy score was 1.2 (=7.8/(3.7+2.8)). There were several other combinations of parallel labeling experiments that produced a high precision score and synergy score. Interestingly, all of the best performing parallel labeling experiments (with precision scores above 7.0) included at least one doubly labeled glucose tracer.

The precision and synergy scores on the diagonal in Figure 4.4 correspond to performing the same tracer experiment twice. We found that synergy scores on the diagonal were all less than 1, thus indicating that no synergistic flux information is gained by performing the same tracer experiment twice, as might be expected. Of course, there could be other good reasons for performing the same tracer experiment multiple times, e.g. to evaluate biological variability (Au et al., 2014); however, from the perspective of improving flux precision, it is always better to use two different tracers in parallel. Several tracers had particularly high synergy scores, e.g. [2,5-^{13}C]glucose, suggesting that this tracer brings significant complementarity in parallel

labeling experiments. Overall, the synergy scores varied from values less than 1 (no synergy) to values much greater than 1 (high synergy). This result indicates that judicious selection of tracers for parallel labeling experiments is important, since some (but not all) parallel labeling experiments produce synergistic improvements in flux precision. To our knowledge, this is the first time that synergies resulting from the use of parallel labeling experiments have been rigorously quantified.

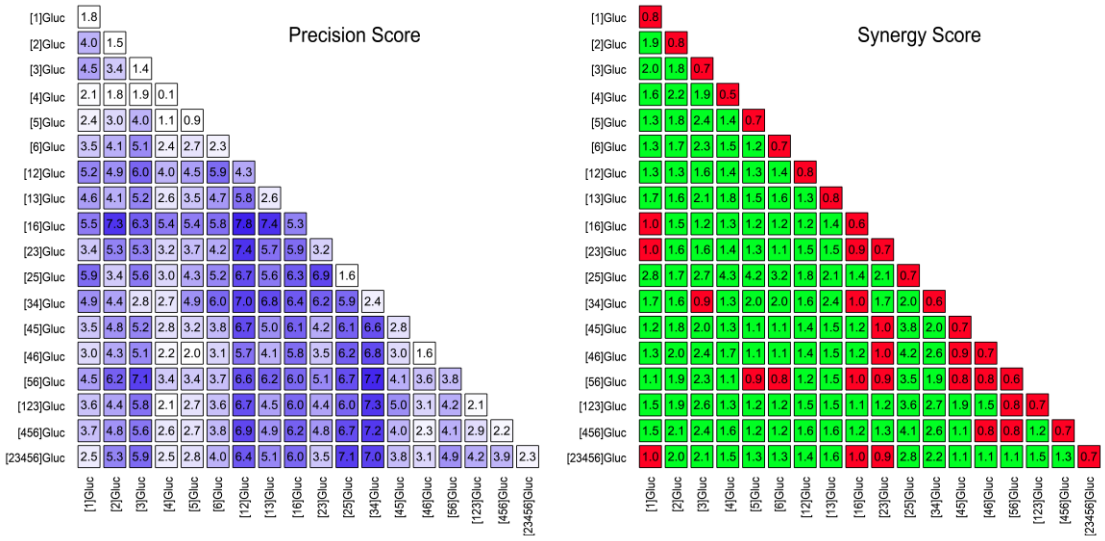


Figure 4.4: Precision and synergy scores for parallel labeling experiments with pure glucose tracers. Synergy scores above 1 (positive synergy) are highlighted in green, and synergy scores below 1 (no synergy) are highlighted in red. ^{13}C-MFA was performed using simulated GC-MS data and assuming wild-type *E. coli* fluxes.

4.4.5 Experimental Validation of Optimal Tracers

To validate the predictions described in the previous sections, we performed four parallel labeling experiments with wild-type *E. coli*. Specifically, labeling

experiments were performed with the following tracers: 100% [1,2-^{13}C]glucose; 100% [1,6-^{13}C]glucose; 51.5% [1,2-^{13}C]glucose + 48.5% [1,6-^{13}C]glucose; and 81% [1-^{13}C]glucose + 19% [U-^{13}C]glucose. The first two tracers were used to validate the prediction that [1,6-^{13}C]glucose and [1,2-^{13}C]glucose are optimal tracers for ^{13}C-MFA when single tracer experiments are used (see section 4.4.1), and to test the synergy of these two tracers in parallel labeling experiments (see section 4.4.4). The third tracer was chosen to validate the prediction that mixing tracers will result in a dramatically poorer performance compared to using the respective pure tracers (see section 4.4.3). The fourth tracer experiment was used as the reference tracer experiment.

E. coli was grown aerobically in parallel batch cultures and mass isotopomer distributions of proteinogenic amino acids were measured by GC-MS. Data from the four tracer experiments were then first analyzed separately by ^{13}C-MFA. In all cases a statistically acceptable fit was obtained. The estimated fluxes agreed well with those previously reported for the closely related *E. coli* K-12 MG1655 strain (Crown et al., 2015a; Leighty and Antoniewicz, 2013). Next, we fitted the experiments with [1,2-^{13}C]glucose and [1,6-^{13}C]glucose in parallel, which also resulted in a statistically acceptable fit. Figure 4.5 shows the calculated precision scores. As predicted, the tracer experiments with [1,2-^{13}C]glucose and [1,6-^{13}C]glucose performed significantly better (precision scores of 6.1 and 5.4, respectively) than the reference tracer experiment. Moreover, as predicted, when the two tracers were mixed (51.5% [1,2-^{13}C]glucose + 48.5% [1,6-^{13}C]glucose), a significantly lower precision score of 2.1 was obtained. Finally, parallel fitting of [1,2-^{13}C]glucose and [1,6-^{13}C]glucose tracer experiments resulted in the highest precision score (18.3) with a synergy score of 1.6 (=18.3/(6.1+5.4)). The higher than expected precision score obtained here (i.e.

compared to simulation results) was mainly due to a poorer than expected performance of the reference tracer experiment. Taken together, the experimental results described in this section confirmed all important predictions of our *in silico* simulations.

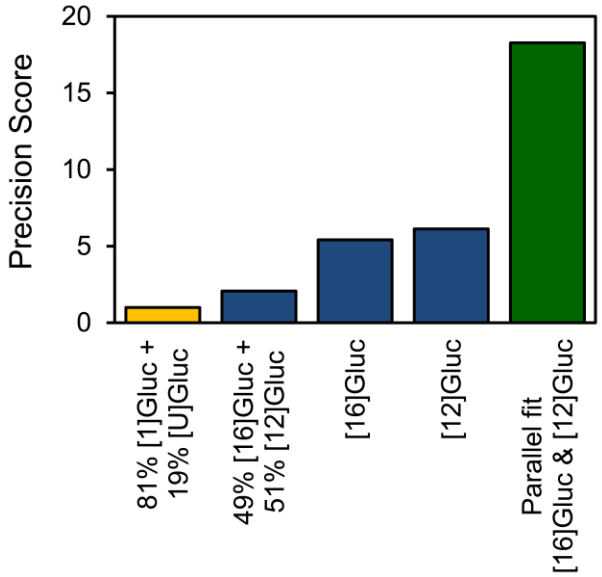


Figure 4.5: Experimentally determined precision scores for four different tracers, and for the parallel fit of tracer experiments with [1,2-^{13}C]glucose and [1,6-^{13}C]glucose. The precision score for the reference tracer experiment, 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose, is by definition 1 (highlighted in yellow).

4.5 Conclusions

In this contribution, we have introduced a new scoring system for identifying optimal tracers for ^{13}C-MFA. Unlike previous efforts (Arauzo-Bravo and Shimizu, 2003; Metallo et al., 2009; Walther et al., 2012), the proposed precision scoring metric accounts for nonlinear flux intervals and is not biased due to normalization by flux

values. The new synergy score introduced here provides information whether it is beneficial to conduct tracer experiments in parallel. Through extensive *in silico* simulations, 19 commercially available glucose tracers were evaluated for ^{13}C-based flux analysis. We demonstrated that the current standard tracer, 80% [1-^{13}C]glucose + 20% [U-^{13}C]glucose, performs poorly in general. A large number of pure glucose tracers, especially doubly ^{13}C-labeled tracers such as [1,6-^{13}C]glucose and [1,2-^{13}C]glucose, performed significantly better. Flux precision was dramatically improved through the use of parallel labeling experiments. For example, we demonstrated an 18-fold improvement in the precision score (compared to the reference tracer experiment) by using parallel experiments with [1,6-^{13}C]glucose and [1,2-^{13}C]glucose.

In this work, we have focused on identifying optimal tracers for ^{13}C-flux analysis in *E. coli* because of the significant importance of this organism in both academia and industry (Long and Antoniewicz, 2014a). The optimal tracers identified here may not be optimal for other organisms, especially if the structure of the metabolic pathways are dramatically different from *E. coli*. However, if the metabolic pathways are similar then likely the same tracers will be optimal, given the fact that we have evaluated a very wide range of possible flux scenarios and consistently found the same tracers to perform optimally.

It is also important to note that flux results not only depend on the selected tracers, but also on the labeling measurements used for ^{13}C-MFA. In this work, we have assumed that proteogenic amino acids (measured by GC-MS) were used for ^{13}C-MFA. GC-MS is a wide used technology in the ^{13}C-MFA field (Antoniewicz, 2015a). However, several alternative measurement techniques are available, and in some cases, may be preferred: tandem mass spectrometry (Antoniewicz, 2013b; Choi and

Antoniewicz, 2011; Choi et al., 2012), LC-MS/MS (McCloskey et al., 2016), and NMR (Masakapalli et al., 2014; Szyperski, 1995; Tang et al., 2007). Moreover, in addition to measuring labeling of protein-bound amino acids, it may be advantageous to measure ^{13}C-labeling of carbohydrates (McConnell and Antoniewicz, 2016), fatty acids (Crown et al., 2015b), nucleosides (Miranda-Santos et al., 2015), glycogen and RNA (Guzman et al., 2014; Long et al., 2016a), and intracellular metabolites (Ahn and Antoniewicz, 2013; Millard et al., 2014). Finally, a practical consideration is the cost associated with performing ^{13}C-MFA studies: the cost of tracers (Table 4.1), analytical instruments, sample preparation, analysis time, and the cost of performing parallel labeling experiments vs. single tracer experiments (Hollinshead et al., 2016). All of these factors should be considered collectively to determine the best or most convenient strategy for completing a ^{13}C-MFA study.

4.6 Author Contributions

SBC developed the theory and performed simulations. CPL performed the experiments and analysis thereof. MRA oversaw the research. All authors contributed to writing the published article.

Chapter 5

ENZYME I FACILITATES REVERSE FLUX FROM PHOSPHOENOLPYRUVATE TO PYRUVATE IN *ESCHERICHIA COLI*

Reprinted with permission from: Long CP*, Au J*, Sandoval NR, Gebreselassie NA, Antoniewicz MR (2017) Enzyme I facilitates reverse flux from pyruvate to phosphoenolpyruvate in *Escherichia coli*. Nature Communications. 8(14316)
*Equal contribution

5.1 Introduction

The phosphoenolpyruvate-carbohydrate phosphotransferase system (PTS) is used by many bacteria and some archaea for the uptake and phosphorylation of sugar substrates (Deutscher et al., 2014). It is the main mechanism of glucose uptake and utilization in the model organism *Escherichia coli*, where it also has an important role in carbon catabolite repression and regulating central carbon metabolism (Deutscher et al., 2014, 2006; Escalante et al., 2012). The PTS consists of four proteins carrying out successive phosphotransferase reactions, coupling glucose transport and phosphorylation to the lower glycolytic reaction of phosphoenolpyruvate (PEP) to pyruvate (PYR). This allows for the coupled regulation of substrate uptake and glycolytic flux, as the PEP/PYR ratio has been shown to act as part of a flux sensor (Kotte et al., 2010; Kremling et al., 2007) and controller of phosphofructokinase (encoded by *pfkA*) activity via allosteric inhibition by PEP (Fenton and Reinhart, 2009). Due to its central metabolic function and complex regulatory role, the PTS is a

frequent target of metabolic engineering interventions (De Anda et al., 2006; Flores et al., 1996; Gosset, 2005; Meza et al., 2012).

Although individual steps of the PTS are known to be reversible (Deutscher et al., 2014, 2006; Postma et al., 1993), current understanding allows only for a net forward flux during the uptake of a PTS sugar (e.g. glucose). Indeed, the conversion from PEP to PYR, which is also facilitated by pyruvate kinases (encoded by *pykA* and *pykF* in *E. coli*), is often assumed to be a committed step in lower glycolysis. This assumption has practical implications, for example in the analysis of stable isotope labeling data through ^{13}C metabolic flux analysis and in flux balance analysis studies. The reverse reaction, PYR to PEP, is carried out by the gluconeogenic enzyme PEP synthetase (PpsA, encoded by *ppsA* in *E. coli*). This enzyme is minimally expressed during growth on glycolytic substrates (Trauchessec et al., 2014), as significant activity would cause a wasteful futile cycle. However, PpsA is actively expressed under gluconeogenic conditions via transcriptional regulation by Cra (Ramseier, 1996).

In this work, we show that Enzyme I (EI), the terminal phosphotransferase in the PTS responsible for the conversion of PEP to PYR, is responsible for a significant *in vivo* flux in the reverse direction (i.e. PYR to PEP) during both gluconeogenic and glycolytic growth. We use knockout strains and ^{13}C alanine tracer experiments to directly quantify this reverse flux and determine gene-reaction relationships. We demonstrate that PpsA and EI are able to interchangeably and exclusively support the major gluconeogenic flux from PYR to PEP during growth on acetate and pyruvate. Similar experiments under growth on glycolytic substrates glucose and xylose demonstrate that this reverse flux is mainly attributable to EI, indicating an unexpected

role for this enzyme in the context of central carbon metabolism. Furthermore, we show that this reverse flux is modulated by genetic perturbation of other PTS components.

5.2 Methods

5.2.1 Materials

Chemicals and culture media were purchased from Sigma-Aldrich (St. Louis, MO). [1-^{13}C]Alanine (99 atom% ^{13}C) and [U-^{13}C]alanine (98+ atom% ^{13}C) were purchased from Cambridge Isotope Laboratories (Andover, MA). M9 minimal medium was used for all labeling experiments. All solutions were sterilized by filtration.

5.2.2 Strains

For wild-type *E. coli* grown on acetate, *E. coli* K-12 MG1655 (ATCC Cat. No. 700925, Manassas, VA) was used. For all other cultures, *E. coli* strains were obtained from the Keio collection (GE Dharmacon), which was generated by one-step inactivation of all non-essential genes in *E. coli* K-12 BW25113 (Δ(*araD-araB*)567, Δ(*lacZ*4787::*rrnB*-3), lambda−, rph-1, Δ(*rhaD*−*rhaB*)568, *hsdR*514) (Baba et al., 2006). The strains used in this study, with identifying information from the collection, are listed in Table 5.1. Double deletion strains were constructed following the method of Datsenko and Wanner on existing Keio collection strains (Baba et al., 2006; Datsenko and Wanner, 2000). Kanamycin resistance cassettes were cured by transformation of pCP20 (Cherepanov and Wackernagel, 1995), which carries the FLP recombinase gene; the pCP20 plasmid was subsequently cured by growth at 42°C overnight and confirmed via replica plating and PCR amplification. Kanamycin

resistance cassettes for second-gene knockouts were amplified from Keio collection single deletion strains using the original Keio collection primers with homologous regions corresponding to the desired deletion. The purified amplicon was electroporated into the cured single deletion host *E. coli* strain expressing 1 mM arabinose-induced λ-Red recombinase genes from the pKD46 plasmid (Datsenko and Wanner, 2000) and grown on solid LB with kanamycin at 37°C. Successful recombination was confirmed via PCR of both the mutated loci. The recombination plasmid pKD46 was subsequently cured by growth overnight at 42°C and confirmed via replica plating. All strains carrying pCP20 and pKD46 were grown at 30°C.

Knockout Gene	Plate-Row-Col ID	JW_id-Strain
ppsA	3-A-3	JW1692-1
ptsG	55-G-3	JW1087-2
crr	57-H-8	JW2410-1
ptsH	57-F-8	JW2408-2
ptsI	57-G-8	JW2409-1

Table 5.1: List of strains used in this study.

5.2.3 Culture Conditions

E. coli strains were cultured aerobically in M9 minimal medium at 37°C in aerated mini-bioreactors with 10 mL working volume (Crown et al., 2015a). Cultures were inoculated at OD_{600} of 0.01, and biomass concentration and growth rates were determined by periodic measurements of OD_{600} using a spectrophotometer (Eppendorf BioPhotometer). The medium contained, for the respective experiments, 1.2 g/L

acetate, 2 g/L pyruvate, 2 g/L glucose, or 4 g/L xylose. In the acetate and pyruvate experiments, a bolus of 1 mM [1-^{13}C]alanine was added when the cultures reached an OD_{600} of ~0.1, and the cells were harvested for analysis at an OD_{600} of 0.5. In the glucose and xylose experiments, a bolus of 10 mM [U-^{13}C]alanine was added when the culture reached an OD_{600} of 0.5, and cells were harvested at an OD_{600} of 1.5. In all experiments, the non-tracer substrate (i.e. acetate, glucose, or xylose) was not limiting throughout, and exponential growth was maintained (i.e. culture performance was not affected by the presence of the tracer). For glucose experiments with strains containing the Δ*ptsI* mutation, pre-cultures were grown in medium with 4 g/L galactose. The cells were centrifuged, washed, and inoculated into the standard glucose medium described above at an OD_{600} of 0.05-0.1, and harvested at an OD_{600} of 0.3.

5.2.4 Gas Chromatography Mass Spectrometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 μm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230°C, the MS quad temperature at 150°C, the interface temperature at 280°C, and the inlet temperature at 250 °C. GC-MS analysis of tert-butyldimethylsilyl (TBDMS) derivatized proteinogenic acids was performed to measure isotopic labeling (Leighty and Antoniewicz, 2013). Mass isotopomer distributions were obtained by integration (Antoniewicz et al., 2007b) and corrected for natural isotope abundances (Fernandez et al., 1996).

5.2.5 Calculations

For the [1-^{13}C]alanine tracer experiments, the fraction of PEP derived from PYR was determined from the labeling of PEP (determined from phenylalanine *m/z* 302 fragment labeling, C1-C2) and PYR (determined from valine *m/z* 288 fragment labeling, C1-C5):

$$\%PEP\ from\ Pyr = \frac{PEP_{M1}}{Pyr_{M1}} = \frac{Phe302_{M1}}{Val288_{M1}}$$

For the [U-^{13}C]alanine tracer experiments, the fraction of PEP derived from PYR was determined by least-squares regression using the measured mass isotopomer distributions (MID) of PEP (determined from phenylalanine *m/z* 302 fragment, C1-C2), OAC (determined from phenylalanine *m/z* 302 fragment, C1-C2), and PYR (determined from phenylalanine *m/z* 260 fragment, C2-C5), after correction for unlabeled biomass

$$MID_{PEP} = (\%PEP\ from\ Pyr) * MID_{Pyr} + (\%PEP\ from\ OAC) * MID_{OAC} + (\%PEP\ from\ gluc\ or\ xyl) * MID_{unlabeled}$$

5.2.5.1 Correction of labeling data for unlabeled biomass (after natural abundance correction)

Since we measure labeling of amino acids from hydrolyzed biomass proteins, the mass isotopomer data must be corrected for unlabeled biomass that was present prior to the introduction of ^{13}C-alanine. The fraction of *old unlabeled* biomass and fraction of *new* biomass (i.e. generated after the introduction of ^{13}C-alanine) is calculated as follows:

Fraction of *old unlabeled* biomass = Ala_{M0}

Fraction of *new* biomass = $1 - Ala_{M0}$

The mass isotopomer data of other amino acids, e.g. valine, phenylalanine, aspartate, can then be corrected for the presence of *old unlabeled* biomass as follows:

$$Val_{M0}^{corr} = \frac{Val_{M0} - Ala_{M0}}{1 - Ala_{M0}} \quad \text{(for M0)}$$

$$Val_{Mi}^{corr} = \frac{Val_{Mi}}{1 - Ala_{M0}} \quad \text{(for Mi, i> 0)}$$

5.2.5.2 Tracer experiments with [1-^{13}C]alanine (growth on acetate and pyruvate)

For [1-^{13}C]alanine tracer experiments, the fraction of PEP derived from pyruvate is calculated from the M1 labeling of pyruvate and M1 labeling of PEP. Labeling of pyruvate is inferred from that of valine. GC-MS analysis of valine produces *m/z* 288 fragment which contains carbons C1-C5 of valine. C1 of valine is derived from C1 of pyruvate, and the remaining carbons are derived from C2-C3 of pyruvate. Since carbons C2-C3 of pyruvate are unlabeled, the M1 labeling of *m/z* 288 fragment reflects the M1 labeling of pyruvate:

$$Pyr_{M1} = Val288_{M1}^{corr}$$

The M1 labeling of PEP is inferred from the M1 labeling of *m/z* 302 fragment of phenylalanine, which contains C1-C2 of phenylalanine that are derived from C1-C2 of PEP:

$$PEP_{M1} = Phe302_{M1}^{corr}$$

Thus, the fraction of PEP derived from pyruvate is calculated as follows:

$$\%PEP\ from\ Pyr = \frac{PEP_{M1}}{Pyr_{M1}} = \frac{Phe302_{M1}^{corr}}{Val288_{M1}^{corr}} = \frac{Phe302_{M1}/(1 - Ala260_{M0})}{Val288_{M1}/(1 - Ala260_{M0})} = \frac{Phe302_{M1}}{Val288_{M1}}$$

In the above equation, we have assumed that the contribution of oxaloacetate to PEPM1 is minimal, which we have confirmed by measuring aspartate M1 labeling directly and showing that it was negligible in all cases.

5.2.5.3 Tracer experiments with [U-^{13}C]alanine (growth on glucose and xylose)

For [U-^{13}C]alanine tracer experiments, the fraction of PEP derived from pyruvate is calculated by least squares regression using the measured mass isotopomer distributions (MID) of PEP (determined from phenylalanine *m/z* 302 fragment, C1-C2), oxaloacetate (OAC) (determined from phenylalanine *m/z* 302 fragment, C1-C2), and PYR (determined from valine *m/z* 260 fragment, C2-C5).

$$MID_{PEP,C1-C2} = (\%PEP\ from\ Pyr) * MID_{Pyr,C1-C2} + (\%PEP\ from\ OAC) * MID_{OAC,C1-C2} + (\%PEP\ from\ gluc\ or\ xyl) * MID_{unlabeled}$$

The mass isotopomer distribution of pyruvate (carbon atoms C1-C2) is determined as follows:

$$MID_{Pyr,C1-C2} = \begin{bmatrix} Pyr_{C1-C2,M0} \\ Pyr_{C1-C2,M1} \\ Pyr_{C1-C2,M2} \end{bmatrix} = \begin{bmatrix} \sqrt{Val260_{M0}^{corr}} \\ 0 \\ 1 - \sqrt{Val260_{M0}^{corr}} \end{bmatrix}$$

The mass isotopomer distributions of PEP and OAC (carbon atoms C1-C2) are determined as follows:

$$MID_{PEP,C1-C2} = \begin{bmatrix} PEP_{C1-C2,M0} \\ PEP_{C1-C2,M1} \\ PEP_{C1-C2,M2} \end{bmatrix} = \begin{bmatrix} Phe302_{M0}^{corr} \\ Phe302_{M1}^{corr} \\ Phe302_{M2}^{corr} \end{bmatrix}$$

$$MID_{OAC,C1-C2} = \begin{bmatrix} OAC_{C1-C2,M0} \\ OAC_{C1-C2,M1} \\ OAC_{C1-C2,M2} \end{bmatrix} = \begin{bmatrix} Asp302_{M0}^{corr} \\ Asp302_{M1}^{corr} \\ Asp302_{M2}^{corr} \end{bmatrix}$$

Finally:

$$MID_{unlabeled} = \begin{bmatrix} M0 \\ M1 \\ M2 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$$

5.3 Results

5.3.1 Enzyme I Supports a Significant Gluconeogenic Flux

There are two possible gluconeogenic routes for acetate metabolism (Fig. 5.1a). Acetate enters central carbon metabolism as acetyl-CoA (AcCoA) and can either be metabolized to PEP via PEP carboxykinase (*pck*) (shown in green in Fig. 5.1a) or via malic enzyme (*maeAB*) followed by conversion of PYR to PEP (shown in purple). As discussed above, this latter reaction is known to be carried out under gluconeogenic conditions by PEP synthetase (*ppsA)*.

To resolve the relative contribution of these two gluconeogenic routes, a tracer experiment using [1-^{13}C]alanine was applied (Fig. 5.1). The tracer was added during growth on excess acetate (growth rates shown in Fig. 5.1b). Alanine equilibrates with intracellular PYR (Fig. 5.1f), which results in a PYR pool (observed via valine labeling) that is a mixture of unlabeled PYR (M0) produced from unlabeled sources in central carbon metabolism, and [1-^{13}C]PYR (M1) produced from the tracer (Fig. 5.1c). As oxaloacetate (OAC, observed via aspartate) is almost entirely unlabeled (Fig. 5.1d) (the labeled C-1 of PYR is lost in the pyruvate dehydrogenase reaction before entering the TCA cycle), the relative contribution of each route to PEP production (as measured by phenylalanine labeling, Fig. 5.1e) is easily calculated.

In the wild-type, a significant amount (~60%) of PEP was generated from PYR (Fig. 5.1g). In order to confirm that PpsA was responsible for this flux, the tracer

experiment was repeated with a Δ*ppsA* knockout strain. Surprisingly, the contribution of PYR to PEP (~65%) was similar to the wild-type. Following a database search for enzymes able to interconvert PYR and PEP(Kanehisa et al., 2016), we hypothesized that Enzyme I (EI, encoded by the gene *ptsI*) may be involved. EI is known to react reversibly(Weigel et al., 1982), but is not known to have a role in gluconeogenesis. In the knockout strain Δ*ptsI*, the contribution of PYR to PEP (~65%) was still similarly high to the wild-type and Δ*ppsA* strains. To determine if any other enzymes were involved with this flux, a double knockout, Δ*ppsA*Δ*ptsI*, was constructed and the tracer experiment was repeated. In this double knockout, PEP labeling was entirely eliminated, indicating that the flux from PYR to PEP was zero (Figs. 5.1e and 5.1g). These results suggest that PpsA and EI interchangeably and exclusively support the large gluconeogenic flux from PYR to PEP observed in the wild-type during growth on acetate. This result was also observed during growth on pyruvate, which is shown in Figure 5.2. With pyruvate, the WT, Δ*ppsA*, and Δ*ptsI* strains grew similarly well (~0.25 hr^{-1}), while Δ*ppsA*Δ*ptsI* was unable to grow (Fig. 5.2a). For the three viable strains, nearly 100% of PEP was generated directly from PYR (Fig. 5.2b), indicating that alternative routes via the glyoxylate shunt and PCK or MAE reactions were not utilized. The unaffected growth rates and PYR to PEP fluxes in both Δ*ppsA* and Δ*ptsI* during growth on acetate and pyruvate reveal a high degree of flexibility in the system, which requires either rapid transcriptional compensation or large excess capacity for each enzyme.

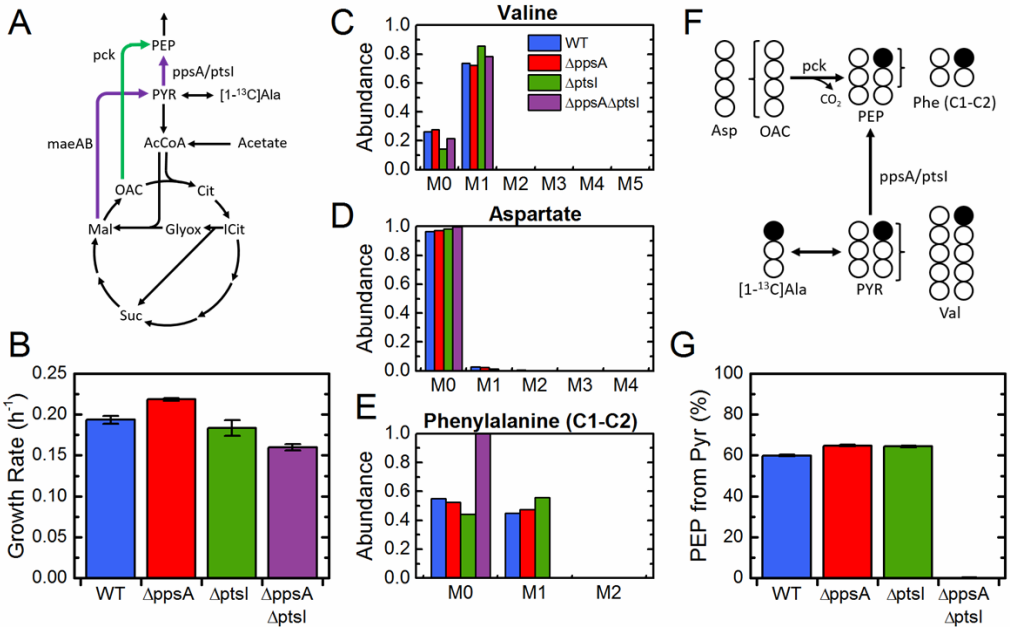


Figure 5.1: Quantification of alternative routes of PEP generation during growth on acetate. **(A)** Schematic showing two routes of PEP synthesis during growth on acetate. After malate is produced via glyoxylate shunt, malic enzyme (*maeAB*) can convert malate to PYR, from which PEP can be formed by the activity of *ppsA* or *ptsI*. Alternatively, *pck* can convert oxaloacetate (OAC) to PEP directly. **(B)** Growth rates of four strains during growth on acetate, wild-type (WT), Δ*ppsA*, Δ*ptsI*, and the double-knockout Δ*ppsA*Δ*ptsI*. **(C)** Labeling of valine from [1-^{13}C]alanine, reflecting PYR labeling. Labeling is M1 (from tracer) and M0 (from unlabeled precursors in central metabolism). **(D)** Labeling of aspartate from [1-^{13}C]alanine, reflecting OAC labeling. Aspartate is almost entirely unlabeled (M0). **(E)** Labeling of the first two (C1-C2) carbons of phenylalanine, reflecting the labeling of the first two carbons of PEP. **(F)** Schematic depicting the conversion of [1-^{13}C]alanine to PEP and the measured amino acids. Opened and filled circles represent unlabeled (^{12}C) and labeled (^{13}C) carbons, respectively. **(G)** Percentage of PEP generated from PYR. Approximately 60% of PEP is generated from PYR in the WT and each single knockout strain; however, the flux is completely eliminated in the double knockout, indicating dual responsibility of *ppsA* and *ptsI* for the conversion of PYR to PEP. Data presented in (B) are mean ± s.e.m. of two biological replicates. Labeling data in (C), (D), and (E) have been corrected for natural abundances and unlabeled biomass present prior to tracer introduction. The error presented in (G) reflects the propagation of GC-MS measurement error through the calculation.

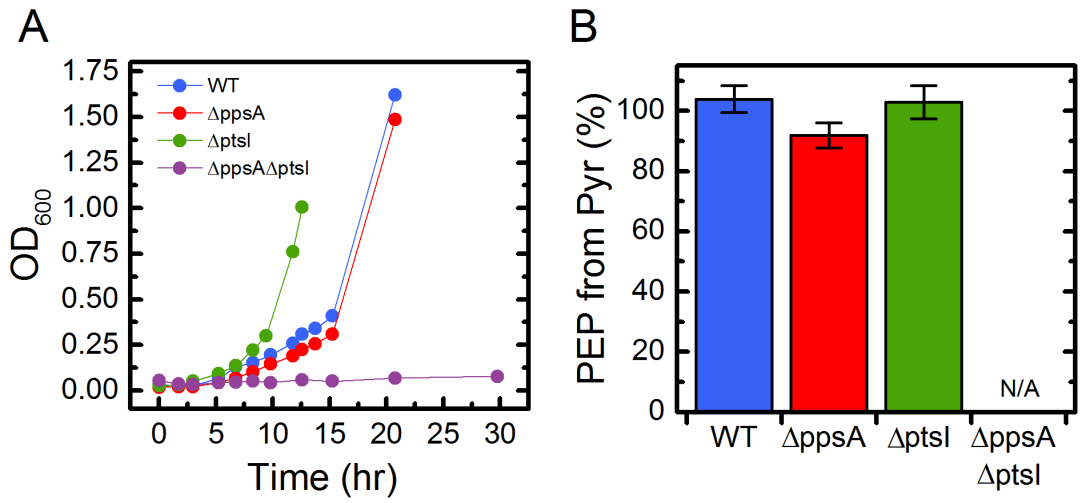


Figure 5.2: Enzyme I supports gluconeogenic growth on pyruvate. **(A)** Growth profiles of four *E. coli* strains during growth on pyruvate: wild-type (WT), Δ*ppsA*, Δ*ptsI*, and the double knockout Δ*ppsA*Δ*ptsI*. The double knockout strain Δ*ppsA*Δ*ptsI* did not grow on pyruvate. **(B)** [1-^{13}C]alanine experiments were performed and the percentage of PEP derived from pyruvate was determined. For the three strains able to grow on pyruvate, effectively all PEP was derived from pyruvate.

5.3.2 A Significant Back-Flux is Measured During Growth on Glucose

Given the surprising activity of EI under gluconeogenic growth conditions, we next sought to determine whether there was any measurable flux from PYR to PEP during growth on glucose. This flux was expected to be minimal or nonexistent, as this reaction is traditionally understood to have a large forward thermodynamic driving force, and gluconeogenic flux via PpsA would create a futile cycle. Glucose is a PTS sugar, meaning that during its consumption EI actively participates in the conversion of PEP to PYR. The experimental approach was modified slightly from the acetate and pyruvate cases by using [U-^{13}C]alanine as tracer instead of [1-^{13}C]alanine (Fig. 5.3). Again, PYR labeling was observed via valine labeling, OAC via aspartate, and PEP labeling via phenylalanine labeling. The contributions to PEP from OAC and PYR

were distinguishable due to M1 labeling in OAC, generated from scrambling in the TCA cycle (Fig. 5.3f). During growth on glucose, the wild-type was determined to have a statistically significant back-flux through which 10% of PEP was generated from PYR. When Δ*ppsA* strain was analyzed, it was found to have the same back-flux as the wild-type (Fig. 5.3g), a result consistent with the reports of minimal PpsA expression during growth on glucose.(Trauchessec et al., 2014) Given the dual contribution of PpsA and EI to gluconeogenic flux under growth on acetate and pyruvate, it was suspected that EI may also be responsible for this back-flux on glucose.

However, the EI mutant (Δ*ptsI*) is known to grow minimally on glucose, and only after a long lag phase.(Flores et al., 1996; Liang et al., 2015) After pre-growth on LB medium and transferring to minimal medium with glucose, little or no growth was observed over 60 hours (Fig. 5.4a). Previous studies have found that growth on glucose can be facilitated by the induction of the *gal* operon, natively used for galactose transport and metabolism.(De Anda et al., 2006) Upon induction, the GalP proton-symport transporter is able to non-specifically transport glucose(Flores et al., 2005, 1996; Hernández-Montalvo et al., 2003; Liang et al., 2015), which can be subsequently phosphorylated by glucokinase (*glk*). To take advantage of this phenomenon, Δ*ptsI* and Δ*ppsA*Δ*ptsI* strains were pre-grown in minimal medium with galactose. Upon transferring to minimal medium with glucose, growth immediately commenced, albeit at a relatively slow growth rate (~0.1 hr^{-1}) (Fig. 5.3b, Fig. 5.4a). The absence of PTS transport was confirmed by the lack of growth under the same conditions of a Δ*ptsI*Δ*glk* double knockout strain (Fig. 5.4b).

In both the Δ*ptsI* and Δ*ppsA*Δ*ptsI* strains first pre-grown in this way, a significant amount of PEP labeling was observed (Fig. 5.3e). In Δ*ptsI*, 13% of PEP came from PYR, indicating significant PpsA activity (Fig. 5.3g), which is likely a result of active expression of *ppsA* in the altered regulatory state of this strain. In the Δ*ppsA*Δ*ptsI* strain, almost all of the PEP labeling was generated from OAC via the PCK reaction (note the significant M1 labeling in Fig. 5.3e). As a result, only 2.4% of PEP was generated from PYR, demonstrating again responsibility of EI and PpsA for this flux.

It is important to note that in these experiments WT and Δ*ppsA* strains take up glucose via the PTS system, while Δ*ptsI* and Δ*ppsA*Δ*ptsI* strains are using non-PTS transporters. In order to get a more direct comparison of these four strains during glycolytic growth, the non-PTS sugar xylose was chosen for subsequent experiments.

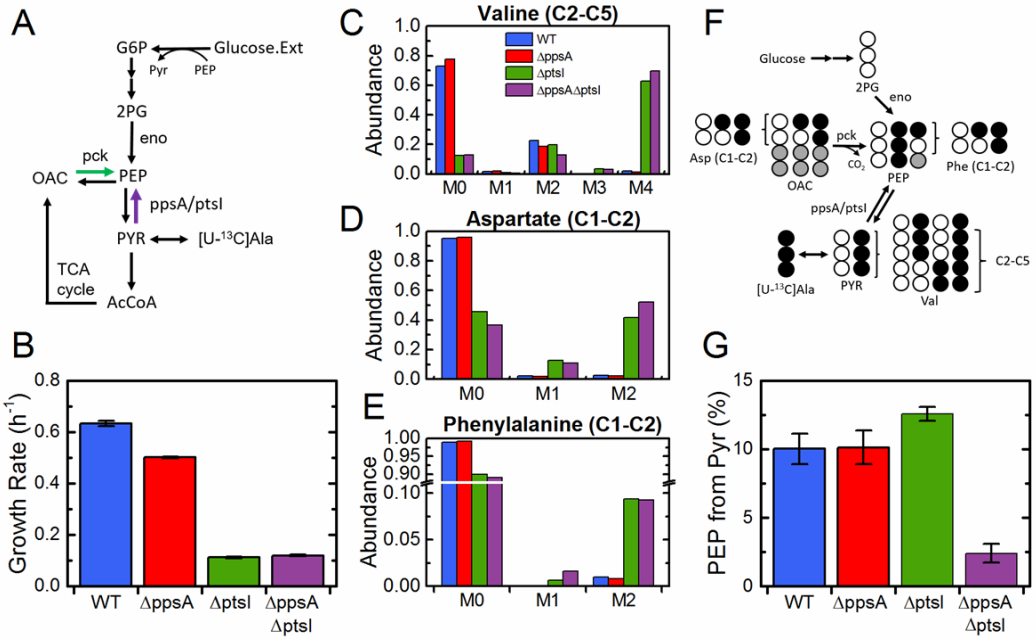

A
G6P
Glucose.Ext
Pyr
PEP
2PG
eno
pck
OAC
PEP
ppsA/ptsI
PYR
[U-13C]Ala
TCA cycle
AcCoA
B
Growth Rate (h-1)
WT
ΔppsA
ΔptsI
ΔppsA ΔptsI
C
Valine (C2-C5)
Abundance
WT
ΔppsA
ΔptsI
ΔppsAΔptsI
M0
M1
M2
M3
M4
D
Aspartate (C1-C2)
M0
M1
M2
E
Phenylalanine (C1-C2)
M0
M1
M2
F
Glucose
2PG
eno
Asp (C1-C2)
pck
CO2
OAC
PEP
Phe (C1-C2)
ppsA/ptsI
[U-13C]Ala
PYR
Val
C2-C5
G
PEP from Pyr (%)
WT
ΔppsA
ΔptsI
ΔppsA ΔptsI

Figure 5.3: During growth on glucose, there is a significant back-flux from PYR to PEP not carried out by PEP synthetase (*ppsA*). **(A)** Schematic of glucose consumption and metabolism related to PEP and PYR interconversion. Glucose is transported and phosphorylated by the PTS, simultaneously converting PEP to PYR via Enzyme I (*ptsI*). **(B)** Growth rates of four strains during growth on glucose; Δ*ptsI* and Δ*ppsA*Δ*ptsI* were pre-grown on galactose to facilitate growth on glucose. **(C)** Labeling of last four carbons (C2-C5) of valine, representing the condensation of the last two carbons (C2-C3) of two PYR molecules. Labeling is mainly M0 (condensation of two unlabeled PYR's), M2 (condensation of one fully labeled PYR and one unlabeled PYR) and M4 (condensation of two labeled PYRs). **(D)** Labeling of the first two (C1-C2) carbons of aspartate, reflecting the labeling of the same carbons in OAC. M1 labeling is generated through scrambling in the TCA cycle. **(E)** Labeling of the first two (C1-C2) carbons of phenylalanine, reflecting the labeling of the first two carbons of PEP. **(F)** Schematic depicting the conversion of [U-^{13}C]alanine to PEP and the measured amino acids. The relative contributions of the three sources of PEP were quantified via regression. Opened and filled circles represent unlabeled (^{12}C) and labeled (^{13}C) carbons, respectively. **(G)** Percentage of PEP generated from PYR. Approximately 10% is generated from PYR in both WT and Δ*ppsA* strains. The contribution is slightly elevated in Δ*ptsI*, likely due to activity of *ppsA*. There is minimal back-flux in Δ*ppsA*Δ*ptsI*. Data presented in (B) are mean ± s.e.m. of two biological replicates. Labeling data in (C), (D), and (E) have been corrected for natural abundances and unlabeled biomass present prior to tracer introduction. The error presented in (G) reflects the propagation of GC-MS measurement error through the calculation.

5.3.3 Enzyme I Accounts for Back-Flux During Growth on Xylose

The [U-^{13}C]alanine tracer experiments were repeated for the same four strains (WT, Δ*ppsA*, Δ*ptsI* and Δ*ppsA*Δ*ptsI*) using the non-PTS sugar xylose as substrate (Fig. 5.5). Xylose is transported into the cell via an ABC transporter (*xylGHF*) or a proton symporter (*xylE*) (Linton and Higgins, 1998) (Fig. 5.5a), which renders EI non-essential for growth (Fig. 5.5b). During growth on xylose, the back-flux observed in

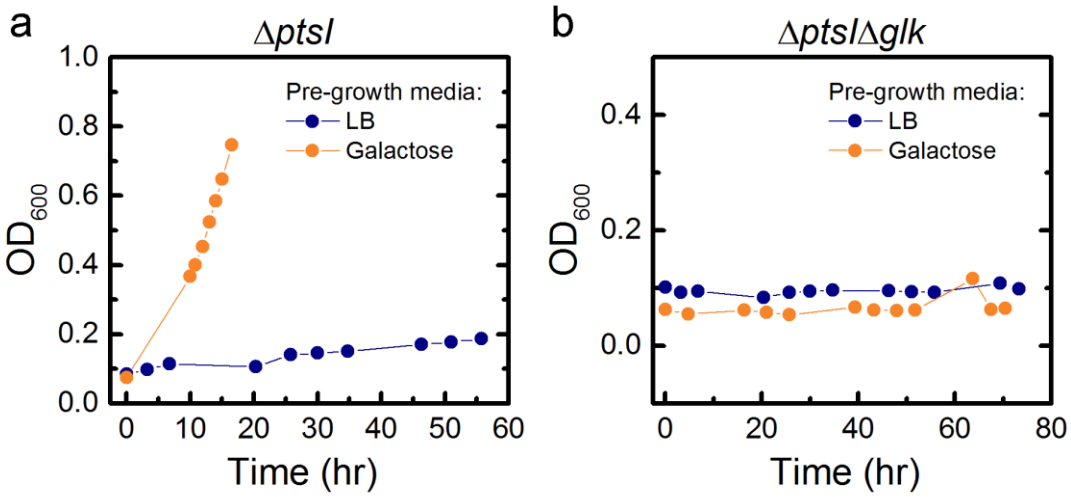


Figure 5.4 **(A)** Growth profiles of Δ*ptsI* strain with glucose as substrate. When Δ*ptsI* was pre-grown on LB medium, little or no growth on glucose was observed. When Δ*ptsI* was pre-grown on M9 medium with galactose, the cells grew without a lag phase on glucose. **(B)** The double knockout Δ*ptsI*Δ*glk* did not grow on glucose, regardless if the cells were pre-grown on LB medium or M9 medium with galactose. These results demonstrate that PTS transport is inactive in the Δ*ptsI* strain.

the wild-type was similar to that observed during growth on glucose, with 11% of PEP formed from PYR (Fig. 5.5g). Once again, this flux was not significantly reduced in the Δ*ppsA* strain. However, this flux was almost completely eliminated in the Δ*ptsI* and Δ*ppsA*Δ*ptsI* strains (Fig. 5.5g), providing strong evidence that EI was exclusively responsible for the conversion of PYR to PEP under these conditions. Thus, during growth on xylose PpsA was not active, as may be expected under normal glycolytic growth conditions (Trauchessec et al., 2014).

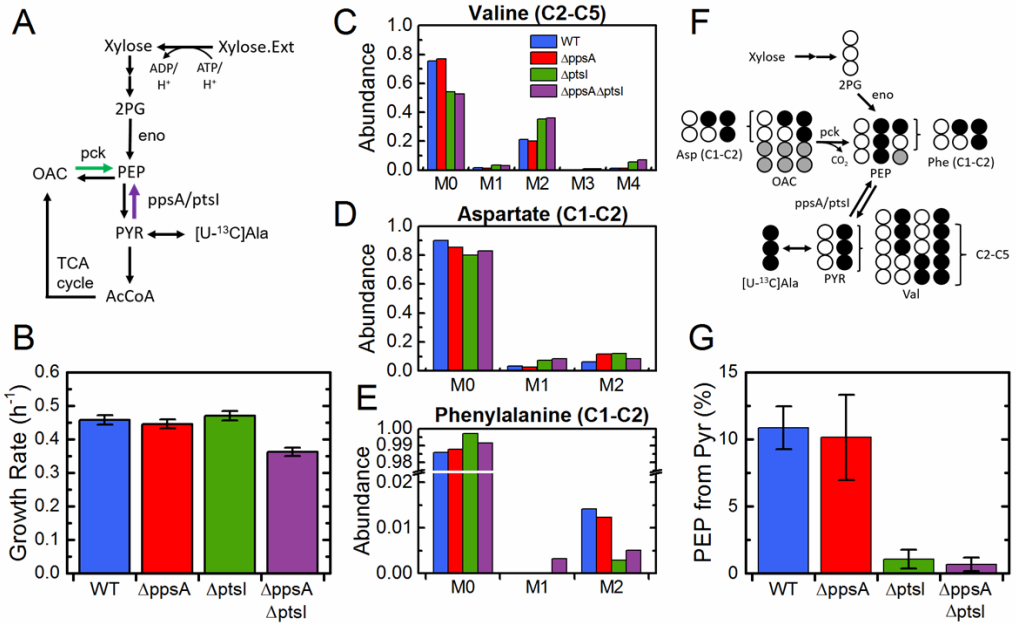

A
Xylose
Xylose.Ext
ADP/ H+
ATP/ H+
2PG
eno
pck
OAC
PEP
ppsA/ptsI
PYR
[U-13C]Ala
TCA cycle
AcCoA
B
Growth Rate (h-1)
WT
ΔppsA
ΔptsI
ΔppsA ΔptsI
C
Valine (C2-C5)
Abundance
WT
ΔppsA
ΔptsI
ΔppsAΔptsI
M0
M1
M2
M3
M4
D
Aspartate (C1-C2)
E
Phenylalanine (C1-C2)
F
Xylose
2PG
eno
Asp (C1-C2)
pck
CO2
OAC
PEP
Phe (C1-C2)
ppsA/ptsI
[U-13C]Ala
PYR
Val
C2-C5
G
PEP from Pyr (%)

Figure 5.5 Enzyme I (*ptsI*) is responsible for the back-flux from PYR to PEP during growth on xylose. **(A)** Schematic of xylose consumption and metabolism related to PEP and PYR interconversion. Xylose is transported via non-PTS transporters. **(B)** Growth rates of four strains during growth on xylose, wild-type (WT), Δ*ppsA*, Δ*ptsI*, and the double knockout Δ*ppsA*Δ*ptsI*. **(C)** Labeling of last four carbons (C2-C5) of valine, representing the condensation of the last two carbons (C2-C3) of two PYR molecules. Labeling is mainly M0 (condensation of two unlabeled PYR's), M2 (condensation of one fully labeled PYR and one unlabeled PYR) and M4 (condensation of two labeled PYR's). **(D)** Labeling of the first two (C1-C2) carbons of aspartate, reflecting the labeling of the same carbons in OAC. M1 labeling is generated through scrambling in the TCA cycle. **(E)** Labeling of the first two (C1-C2) carbons of phenylalanine, reflecting the labeling of the first two carbons of PEP. **(F)** Schematic depicting the conversion of [U-^{13}C]alanine to PEP and the measured amino acids. The relative contributions of the three sources of PEP were quantified via regression. Opened and filled circles represent unlabeled (^{12}C) and labeled (^{13}C) carbons, respectively. **(G)** Percentage of PEP generated from PYR. Approximately 10% is generated from PYR in both the WT and Δ*ppsA* strains. The flux is nearly completely eliminated in the Δ*ptsI* and Δ*ppsA*Δ*ptsI* strains, indicating a major role for Enzyme I (*ptsI*) in facilitating the back-flux. Data presented in (B) are mean ± s.e.m. of two biological replicates. Labeling data in (C), (D), and (E) have been corrected for natural abundances and unlabeled biomass present prior to tracer introduction. The error presented in (G) reflects the propagation of GC-MS measurement error through the calculation.

5.3.4 Back-Flux is Affected by Genetic Knockouts of PTS Components

Given the strong evidence for EI involvement in the back-flux from PYR to PEP under glycolytic conditions, it was further hypothesized that this activity would be perturbed in knockout mutants of other PTS components. The PTS and its components are shown in Fig. 5.6a. The phosphotransferase partner of EI is HPr (encoded by the gene *ptsH*), which then interacts with the soluble (*crr*) and membrane-bound (*ptsG*) components of the glucose-specific Enzyme II complex (EIIABC^{Glc}). The growth rates on glucose and xylose for the mutant strains Δ*ptsG*,

Δ*crr*, and Δ*ptsH* are shown in Fig. 5.6b. The mannose Enzyme II complex (EII^{Man}) is also known to transport glucose (Curtis and Epstein, 1975), allowing for the Δ*ptsG* and Δ*crr* strains to grow on glucose with no lag phase, presumably still utilizing the PTS system. The lack of significant lag phase or growth defect for Δ*ptsH* on glucose is less clear, but it has been suggested that this strain may be able to recruit non-PTS transporters such as *galP* more quickly than Δ*ptsI* (Liang et al., 2015). Alternatively, it has been suggested that the HPr-like protein FPr from the fructose PTS may be able to substitute its activity for HPr (Bettenbrock et al., 2007).

The [U-^{13}C]alanine tracer experiments described above for glucose and xylose were performed for all knockouts of PTS components. There was a striking increase in the back-flux for several knockout strains (Fig. 5.6c). For example, in the Δ*ptsG* strain grown on glucose, 21% of PEP was formed from PYR. Similarly high back-fluxes were also observed for Δ*crr* on both glucose (26%) and xylose (24%). However, the back-flux was nearly eliminated in Δ*ptsH* on glucose, and significantly reduced on xylose. This indicates that HPr is likely the primary, if not sole, phosphotransferase partner of EI.

While *ppsA* is generally not expected to be expressed during normal glycolytic conditions, it could be expressed and active in the altered regulatory environments of these mutant strains, as was seen in the case of Δ*ptsI* on glucose. The phosphorylation states of PTS proteins are known effectors in regulatory circuits, particularly in the signaling of glucose availability for global metabolic regulation (Deutscher et al., 2006; Reddy and Kamireddi, 1998). Perturbations in the PTS system could conceivably result in expression of gluconeogenic genes, e.g. through the activation of the global regulator Cra. To help to assign responsibility for the increased back-flux in

these strains, several double knockouts of *ptsG*, *crr*, *ppsA*, and *ptsI* were constructed (Fig. 5.6b,c). The strain Δ*ptsG*Δ*ppsA* had a significant reduction in back-flux, down to 10% from 21% in the single knockout during growth on glucose (Fig. 5.6c). Thus the excess back-flux in Δ*ptsG*, relative to the wild-type, was attributable to PpsA. In contrast, there was only a modest decrease in back-flux for Δ*crr*Δ*ppsA* on glucose, and there was no decrease on xylose. In the Δ*crr*Δ*ptsI* strain, pre-grown on galactose as previously described, there was a reduction in the back-flux during growth on glucose, down to 14% from 26% in Δ*crr*. On xylose, the flux for the double knockout was nearly zero, compared with 24% in the single knockout. The fact that Δ*crr*Δ*ptsI* on glucose still had some back flux likely indicates some degree of PpsA activity. Therefore, both enzymes play a role in the elevated back-flux of Δ*crr* on glucose. During growth on xylose, however, the elevated back-flux is entirely the result of EI.

Overall, the double knockout studies are helpful in deconvoluting the relative contributions of PpsA and EI. These results suggest that in certain PTS mutant strains, particularly Δ*ptsI*, Δ*crr*, and Δ*ptsG*, the perturbed regulatory environments during growth on glucose results in *ppsA* being expressed and active. PpsA does not appear to play a role during growth on xylose, allowing for a more direct analysis of the interactions of PTS component perturbations with the back-flux activity.

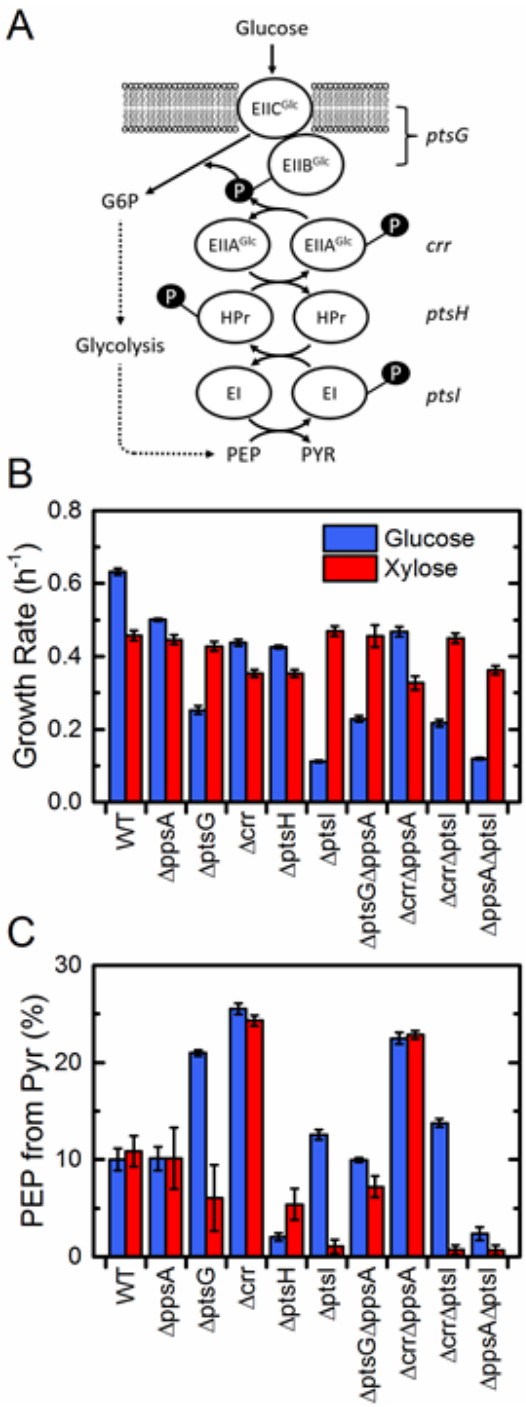

A
Glucose
EIIC^Glc
EIIB^Glc
ptsG
G6P
EIIA^Glc
EIIA^Glc
crr
HPr
HPr
ptsH
Glycolysis
EI
EI
ptsI
PEP
PYR
B
Growth Rate (h^-1)
0.0
0.2
0.4
0.6
0.8
Glucose
Xylose
WT
ΔppsA
ΔptsG
Δcrr
ΔptsH
ΔptsI
ΔptsGΔppsA
ΔcrrΔppsA
ΔcrrΔptsI
ΔppsAΔptsI
C
PEP from Pyr (%)
0
10
20
30
WT
ΔppsA
ΔptsG
Δcrr
ΔptsH
ΔptsI
ΔptsGΔppsA
ΔcrrΔppsA
ΔcrrΔptsI
ΔppsAΔptsI

Figure 5.6 Genetic perturbations of PTS components significantly impact the back-flux from PYR to PEP. **(A)** Schematic of the PTS sugar transport system, which couples the transport and phosphorylation of glucose at $EIIBC^{Glc}$ (*ptsG*) to conversion of PEP to PYR (*ptsI*), via the phosphotransferases ptsH (*ptsH*) and $EIIA^{Glc}$ (crr). **(B)** Growth rates of wild-type (WT) *E. coli* and 9 knockout strains, including all single knockouts of PTS components and selected double knockouts, grown on glucose and xylose. All *ptsI* knockout strains grown on glucose were pre-grown on galactose. **(C)** [U-^{13}C]alanine experiments were performed for all strains, and the percentage of PEP derived from PYR was determined. Several strains had significantly higher percentages of PEP derived from PYR, particularly Δ*ptsG* on glucose and Δ*crr* on glucose and xylose, indicating that PTS component perturbation impacts back-flux. Double knockouts of these strains and Δ*ppsA* or Δ*ptsI* showed that *ppsA* plays a significant role during growth on glucose, accounting for all of the elevated back-flux in Δ*ptsG*, and for some in Δ*crr*, as evidenced by the residual flux in ΔcrrΔptsI. On xylose, the elevated back-flux in Δ*crr* is caused exclusively by EI. Data presented in (B) are mean ± s.e.m. of two biological replicates. The error presented in (C) reflects the propagation of GC-MS measurement error through the calculation.

5.4 Discussion

The results presented here show that Enzyme I facilitates reverse flux from PYR to PEP in *E. coli*. This function is active both under conditions in which the PTS is the primary means of transporting substrate (growth on glucose) and is not (growth on acetate, pyruvate or xylose). Knowledge of this gene-reaction relationship will improve our understanding and annotation of *E. coli* central carbon metabolism, which is of central importance in metabolic modeling and engineering efforts such as ^{13}C metabolic flux analysis (Long and Antoniewicz, 2014a) and the development of production strains (De Anda et al., 2006; Flores et al., 1996; Gosset, 2005; Meza et al., 2012).

This metabolic activity also raises biological questions about the PTS, its regulation, and whether there are additional unannotated connectivities in the network.

For example, are there other kinases that interact with HPr or EI to provide the phosphoryl groups needed to sustain this large flux from PYR to PEP? In the gluconeogenic conditions studied here, there is clearly an unknown phosphate donor that allows EI to drive a large net flux from PYR to PEP. In the glycolytic conditions it is less clear whether the observed labeling is a result of simple reversibility in the EI/HPr system, or whether there is another donor specifically responsible for the PYR to PEP flux. The large increase in the back-flux in Δ*crr*, along with its elimination in Δ*ptsH* during growth on glucose, is consistent with a model in which HPr is the primary or sole phosphotransferase partner of EI, and the Δ*crr* mutation perturbs the equilibria of the PTS chain as the less abundant or efficient EII^{Man} is substituted for EII^{Glc}. This would cause the phosphorylated form of HPr to accumulate, driving the partial reversal of the EI and PEP/PYR reactions. The fact that this also occurs during growth on xylose is surprising, and indicates either a robust dynamic equilibrium in the PTS system even when not being actively used, or the activity of other unknown factors.

The complexity of the enzymatic and regulatory interactions in the PTS system require caution when interpreting these results. It is possible that EI could phosphorylate or activate another, yet unknown enzyme which phosphorylates PYR, or that the altered phosphorylation states of PTS components in the mutant strains studied result in the activation of such an enzyme. For example, $P\sim EIIA^{Glc}$ activates adenylate cyclase, generating cAMP which activates the global regulator Crp(Deutscher et al., 2006), which controls transcription of over 100 genes(Zheng et al., 2004). Other regulatory functions spanning from carbon and nitrogen metabolism to chemotaxis are directly influenced by the phosphorylation state of PTS components

(Deutscher et al., 2014). In this work we were able to deconvolute the role of *ppsA*, which was active in some knockout strains. Further work is needed to clarify the network, including potential candidates implicated in this work (e.g. a phosphotransferase donor during gluconeogenic growth) as well as to rule out the presence of any more unknown intermediary enzymes.

The reversibility of the PEP to PYR step in glycolysis is also surprising in the global sense, particularly in the context of textbook understanding of this reaction as a "committed step" with a large Gibbs free energy drop. In fact, the results presented here are consistent with a recent study that estimated *in vivo* ΔG values in central carbon metabolism using measured metabolite concentrations and observed cellular ΔG values (Park et al., 2016). For this reaction, the estimated ΔG was significantly lower than historically assumed. Similar labeling studies to those presented here demonstrated the reversibility of pyruvate kinase in human iBMK cells (Park et al., 2016). Taken together these results are cause for reconsideration of our understanding of the thermodynamics, control, and engineering targets of central carbon metabolism.

5.5 Author Contributions

CPL and JA performed the experiments in the glucose, xylose, and pyruvate conditions. NAG performed the experiments in the acetate condition. NRS constructed all double knockout strains. CPL, JA and MRA designed the project and wrote the paper with help from all authors.

Chapter 6

CHARACTERIZATION OF PHYSIOLOGICAL RESPONSES TO 22 GENE KNOCKOUTS IN *ESCHERICHIA COLI* UPPER CENTRAL CARBON METABOLISM

Reprinted with permission from: Long CP, Gonzalez JE, Sandoval NR, Antoniewicz MR (2016) Characterization of physiological responses to 22 gene knockouts in *Escherichia coli* central carbon metabolism. Metab. Eng. 37, 102-113.

6.1 Introduction

Gene knockouts are widely used in biology to identify specific functions of corresponding gene products, as well as to study their roles in broader systems contexts. The utility of this approach has long been appreciated in the study of microbial metabolism, particularly by observing responses to knockouts of metabolic enzymes and global regulators. For example, in the model organism *Escherichia coli*, multiple '-omics' techniques have been applied to gain insights into metabolic robustness (Ishii et al., 2007), transcriptional regulatory control (Haverkorn van Rijsewijk et al., 2011), hidden reactions (Nakahigashi et al., 2009), as well as responses following adaptive evolution (Charusanti et al., 2010; Fong et al., 2006). Metabolic fluxes are often the output of interest, and numerous studies have applied ^{13}C-metabolic flux analysis (^{13}C-MFA) to knockouts of genes in *E. coli* central carbon metabolism and its global regulators (Long and Antoniewicz, 2014a). Many of these studies have been facilitated by the Keio collection of all viable single-gene *E. coli* knockouts (Baba et al., 2006). Knockout studies such as these are highly useful in improving our understanding of the structure and dynamics of metabolic networks,

and provide data for assessing and improving our ability to predict complex genetic perturbation responses. This ability is critical to rational strain design in metabolic engineering, and the difficulty of such predictions has long been identified as a key obstacle in the field (Bailey, 1991; Stephanopoulos, 1999).

Due to the centrality of gene perturbation response prediction to metabolic engineering, a variety of modeling approaches have been developed and applied. The most prominent set are the constraint-based reconstruction and analysis (COBRA) methods, which rely on metabolic network stoichiometry and a defined 'objective function'. These include flux balance analysis (FBA) (Edwards and Covert, 2002), minimization of metabolic adjustment (MOMA) (Segre et al., 2002), regulatory on/off minimization of metabolic flux changes (ROOM) (Shlomi et al., 2005), and relative optimality in metabolic networks (RELATCH), which also includes gene expression data from the reference state (Kim and Reed, 2012). These predictive models are then implemented as tools for rational strain design, such as in OptKnock (Burgard et al., 2003; Chowdhury et al., 2014; Ranganathan et al., 2010).

In a recent review (Long and Antoniewicz, 2014a), we highlighted important gaps in studies of *E. coli* knockout metabolism. For example, knockouts of many genes in central carbon metabolism have not been studied at all, and flux results from those that have been studied are often inconsistent or difficult to compare due to differences in culturing conditions and analytical approaches. This has impeded broad systems biology analyses that would be possible with a large, self-consistent study, as well as by integrating flux data with complimentary phenotypic observations. Illustrating the broad interest of such data, Mackie et al. (Mackie et al., 2014) recently

suggested that knockout phenotypic data would be an important addition to the Ecocyc database.

Limitations in the available experimental data have inhibited thorough evaluation of the various competing COBRA models and the development more mechanistic kinetic and regulatory models. Demonstrating the potential of the latter, Khodayari et al. (Khodayari et al., 2014) recently used flux data from seven *E. coli* knockout mutants to estimate the parameters of a kinetic model of central carbon metabolism. In efforts such as these, additional high quality physiological data would undoubtedly be of great use.

In this study, we present a comprehensive physiological characterization of wild-type *E. coli* and 22 knockouts of central carbon metabolism enzymes. The scope of this work consists of the following physiological data: growth rates, substrate uptake rates, product secretion rates, biomass and acetate yields, and biomass composition. These data were collected for the knockouts of upper central carbon metabolism, defined here to include glucose transporters, the upper portion of the glycolysis (EMP) pathway, pentose phosphate pathway (PPP), and Entner-Doudoroff (ED) pathway (Figure 6.1). In this work, we chose to study cellular responses under aerobic, batch (glucose-rich) conditions during exponential growth in M9 minimal medium. Under these conditions, no external growth limitations are imposed on the cells. We therefore expect to observe metabolic responses that inherently reflect the altered metabolic state of the knockouts, i.e. more so than would be observed with imposed growth limitations. Previous continuous culture studies reported surprisingly little metabolic variation when *E. coli* knockouts were grown at an arbitrary low growth rate (Ishii et al., 2007).

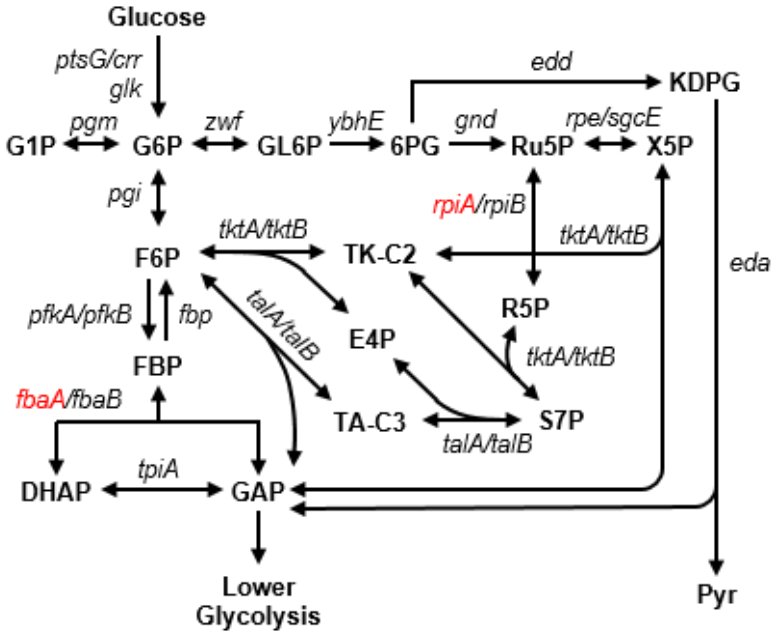


Figure 6.1: Upper central carbon metabolism, with all genes studied here shown in their metabolic contexts (*fbaA* and *rpiA* were not included in this study, shown in red, see Methods section).

6.2 Methods

6.2.1 Chemicals

All chemicals were purchased from Sigma-Aldrich (St. Louis, MO). [U-^{13}C]Glucose was purchased from Cambridge Isotope Laboratories (Andover, MA). M9 minimal medium was used for all experiments. All media and solutions were sterilized by filtration.

6.2.2 Strains and Culture Conditions

E. coli strains were obtained from the Keio collection (GE Healthcare Dharmacon), which were generated by one-step inactivation of all non-essential genes in *E. coli* K-12 BW25113 (Baba et al., 2006). The specific strains used in this study

are listed in Table 6.1. The Keio collection contains two strains per gene knockout, and the reader should note which specific strain was used here, as differences resulting from adaptive evolution or contamination of stock cultures are possible. Two knockouts were excluded from this study: *ΔfbaA* and *ΔrpiA* (Figure 6.1). *ΔfbaA* is not available from the Keio collection, indicating it is likely essential; the purchased *ΔrpiA* was validated to not be the correct strain. We were unable to create a *rpiA* knockout *de novo*, suggesting that this gene may be essential.

Pathway	Knockout Gene	Plate-Row-Col	Strain ID
Wild-type	none		
PTS glucose transport	ptsG	55-G-3	JW1087-2
	crr	57-H-8	JW2410-1
Glucose kinase	glk	3-C-5	JW2385-1
Glycogen biosynthesis	pgm	5-E-5	JW0675-1
Oxidative pentose phosphate pathway	zwf	3-C-3	JW1841-1
	ybhE	19-G-6	JW0750-3
	gnd	4-E-3	JW2011-3
Non-oxidative pentose phosphate pathway	rpe	3-C-6	JW3349-2
	sgcE	73-D-6	JW4263-1
	rpiB	4-G-7	JW4051-2
	tktA	5-A-6	JW5478-1
	tktB	3-F-5	JW2449-3
	talA	3-E-5	JW2448-1
	talB	1-H-10	JW0007-1
Entner-Doudoroff pathway	edd	51-D-3	JW1840-1
	eda	51-C-3	JW1839-1
Upper glycolysis	pgi	3-F-7	JW3985-1
	pfkA	3-F-6	JW3887-1
	pfkB	77-B-4	JW5280-1
	fbp	5-H-9	JW4191-1
	fbaB	77-C-8	JW5344-1
	tpiA	4-G-6	JW3890-2

Table 6.1: *E. coli* strains from the Keio collection (GE Healthcare Dharmacon) used in this study.

For assessment of biomass and excreted metabolite yields, each strain was grown in aerated mini-bioreactors with 10 mL working volume (Leighty and Antoniewicz, 2013) in M9 minimal medium with 2 g/L glucose. Cultures were grown until glucose depletion. Biomass yields were calculated by regression of glucose concentrations and optical density (OD_{600}) measurements (Eppendorf BioPhotometer). Supernatant was collected at the time of glucose depletion for analysis of excreted metabolites by HPLC (Au et al., 2014). For assessment of dry weight and biomass composition, two biological replicate 100 mL cultures were grown aerobically in shaker flasks, in M9 minimal medium with 2 g/L glucose. Cells were harvested at mid-exponential phase ($OD_{600} \approx 0.7$). At that point, samples for dry weight analysis were taken by filtration of 70 mL of culture using a 0.2 μm cellulose acetate filter (Sartorius 11107-47-N), followed by drying for several days at 80 °C until constant weight. Additionally, samples containing the equivalent of 1 mL of culture at OD_{600} = 1.0 (roughly 0.3 mg of dry weight) were washed twice with glucose-free M9 medium and used for biomass composition analysis.

6.2.3 Biomass Composition Analysis

The methods used for quantifying biomass composition were previously described in detail (Long and Antoniewicz, 2014b). Briefly, samples were prepared by three respective methods: hydrolysis of protein and subsequent TBDMS derivatization of amino acids; hydrolysis of RNA and glycogen and subsequent aldonitrile propionate derivatization of sugars (ribose and glucose, respectively); and fatty acid methyl ester derivatization. In total, 17 amino acids were quantified. The amino acids

arginine, cysteine and tryptophan are degraded during hydrolysis and were thus not detected. For total protein quantification, we assumed the values previously reported (Neidhardt, 1987) for these three amino acids. Glutamine and asparagine were deaminated to glutamate and aspartate, respectively, during hydrolysis; thus, we report the combined pools of each. Quantification of all species was achieved by isotope ratio analysis using an isotopically labeled standard and a naturally labeled sample. In this study, the standard was generated by growing wild-type *E. coli* on [U-^{13}C]glucose and aliquoting identical (1 mL of an OD_{600} = 1.0) samples of this "fully labeled" biomass. These were centrifuged and washed twice with M9 medium. The composition of the labeled biomass was characterized using unlabeled chemical standards, and subsequently these were used as standards by co-dissolving with the unlabeled samples at the beginning of each respective analytical method.

6.2.4 Gas Chromatography-Mass Spectrometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 μm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230 °C, the MS quad temperature at 150 °C, the interface temperature at 280 °C, and the inlet temperature at 250 °C. For GC-MS analysis of amino acids, 1 μL was injected at 1:40 split ratio. The column was started at 80 °C for 2 min, increased to 280 °C at 7 °C/ min, and held for 20 min. For GC-MS analysis of fatty acid methyl esters (FAME) and sugar derivatives, 1 μL was injected splitless (Crown et al., 2015b). The column was started at 80 °C for 2 min, increased to 280 °C at 10 °C/min, and held for 12 min.

6.2.5 Calculation of Oxygen Uptake and Carbon Dioxide Production Rate

The carbon dioxide production rate was calculated from an overall carbon balance, accounting for glucose uptake, acetate secretion, and biomass formation. The oxygen uptake rate was calculated from an overall redox balance using the degrees of reduction and production/consumption rates of glucose, acetate and biomass. The degree of reduction of biomass of each strain was calculated from the measured biomass composition. The degree of reduction (γ) was calculated as follows (Antoniewicz et al., 2007c; Crown and Antoniewicz, 2013a):

$$\gamma = 4C + H - 2O - 3N + 6S + 5P - charge$$

6.2.6 COBRA Modeling

Flux balance analysis (FBA), minimization of metabolic adjustment (MOMA), and the relative change (RELATCH) algorithms were implemented through the COBRA Toolbox 2.0 in Matlab 2012b (Kim and Reed, 2012; Schellenberger et al., 2011) using the *E. coli* iAF1260 genome scale model (Feist et al., 2007). CPLEX from Tomlab (http://tomopt.com) was used as the LP and QP solvers. RELATCH was downloaded from the Reed Laboratory website (http://reedlab.che.wisc.edu/codes.php). For FBA calculations, the upper bounds for glucose and oxygen uptake rates were set at wild-type values of 8.5 mmol/g_{dw}/hr and 12 mmol/g_{dw}/hr respectively (see Results), and growth rate was optimized. For MOMA and RELATCH, the reference state was generated using fluxes previously reported from a ^{13}C-MFA study (Leighty and Antoniewicz, 2013) as follows. First, the reported fluxes were re-normalized to the reported glucose uptake rate of 8.5 mmol/g_{dw}/hr. Then, in the iAF1260 model, the upper and lower bounds for key branch point reactions in central carbon metabolism were set as the measured value plus or

minus 0.05 respectively, corresponding approximately to the reported 95% confidence intervals. These included the glucose-6-phosphate isomerase ('PGI'), glucose-6-phosphate dehydrogenase ('G6PDH2r', otherwise referred to as "zwf" here), 6-phosphogluconate dehydratase ('EDD'), triose phosphate isomerase ('TPI'), glyceraldehyde-3-phosphate dehydrogenase ('GAPD'), pyruvate kinase ('PYK'), phosphotransacetylase ('PTAr', toward acetate production), citrate synthase ('CS'), isocitrate lyase ('ICL'), malic enzyme (ME1, ME2), phosphoenolpyruvate carboxylase ('PPC'), and phosphoenolpyruvate carboxykinase ('PPCK') reactions. The upper limit of the glucose uptake rate was set to 8.5 mmol/g_{dw}/hr, at least 95% of which was required to be transported via the PTS system ('Glcptspp'). An FBA simulation was then run to obtain a genome-scale reference state flux solution. For RELATCH, gene expression data for the wild-type reference state was included (Covert et al., 2004). All reaction bounds, including all transport bounds, were lifted for the calculation of the knockout phenotype using MOMA and RELATCH.

6.3 Results

6.3.1 Characterization of Physiological Responses to Gene Knockouts

6.3.1.1 Biomass Dry Weights

The experimentally determined OD_{600} to cell-dry-weight conversion factors are summarized in Figure 6.2 (top right). Out of the 23 strains studied here, 21 strains had a conversion factor that conformed to a normal distribution (as confirmed by the Lilliefors normality test) with a mean of 0.31 g/L/OD_{600} and a standard deviation of 0.02. The two notable outliers were *ΔptsG* with a dry weight conversion factor of 0.42 g/L/OD_{600} and *ΔtpiA* with a dry weight conversion factor of 0.41 g/L/OD_{600}. The high

conversion factors for these two strains were consistently observed in multiple biological replicates. Interestingly, we also observed significant foaming in cultures of these two strains, while little or no foaming was observed with the other 21 strains. For all subsequent calculations we used the individual dry weight conversion factors for *ΔptsG* (0.42 g/L/OD_{600}) and *ΔtpiA* (0.41 g/L/OD_{600}), and used the average value of 0.31 g/L/OD_{600} for the remaining 21 strains.

6.3.1.2 Growth Rates

The measured growth rates are shown in Figure 6.2 (top left). The growth rate of wild-type strain was 0.63 hr^{-1}. Only one knockout strain had a significantly higher growth rate, *Δpgm*, which grew with a specific growth rate of 0.68 hr^{-1}. The slowest growing strains were *ΔpfkA* (0.16 hr^{-1}), *Δtpi* (0.17 hr^{-1}), *Δpgi* (0.19 hr^{-1}), *ΔptsG* (0.25 hr^{-1}), *Δrpe* (0.30 hr^{-1}) and *Δcrr* (0.44 hr^{-1}). Other strains had growth rates similar to or slightly less than the wild-type.

6.3.1.3 Biomass Yields

We observed relatively little variability in the measured biomass yields (Figure 6.2, bottom left). Most knockouts had similar values to the wild-type (0.41 g_{dw}/g). The *ΔptsG* strain had a higher yield of 0.57 g_{dw}/g. Other knockouts had yields similar to or less than the wild-type, with *ΔybhE* having the lowest biomass yields (~0.31 g_{dw}/g).

6.3.1.4 Acetate Yields

Relatively large variability was observed for acetate yields (Figure 6.2, bottom right), and correspondingly, the acetate secretion rates. The wild-type strain produced 0.71 mol of acetate per mol of glucose consumed. The knockouts *ΔptsG* and *Δpgi* produced little or no acetate, while others produced anywhere from ~0.4 mol/mol

(*ΔpfkA* and *ΔtpiA*) to as much as 0.95 mol/mol (*ΔybhE* and *Δgnd*). The variations in acetate yields likely reflect significant differences in underlying internal fluxes, particularly with respect to the acetyl-CoA node and the TCA cycle.

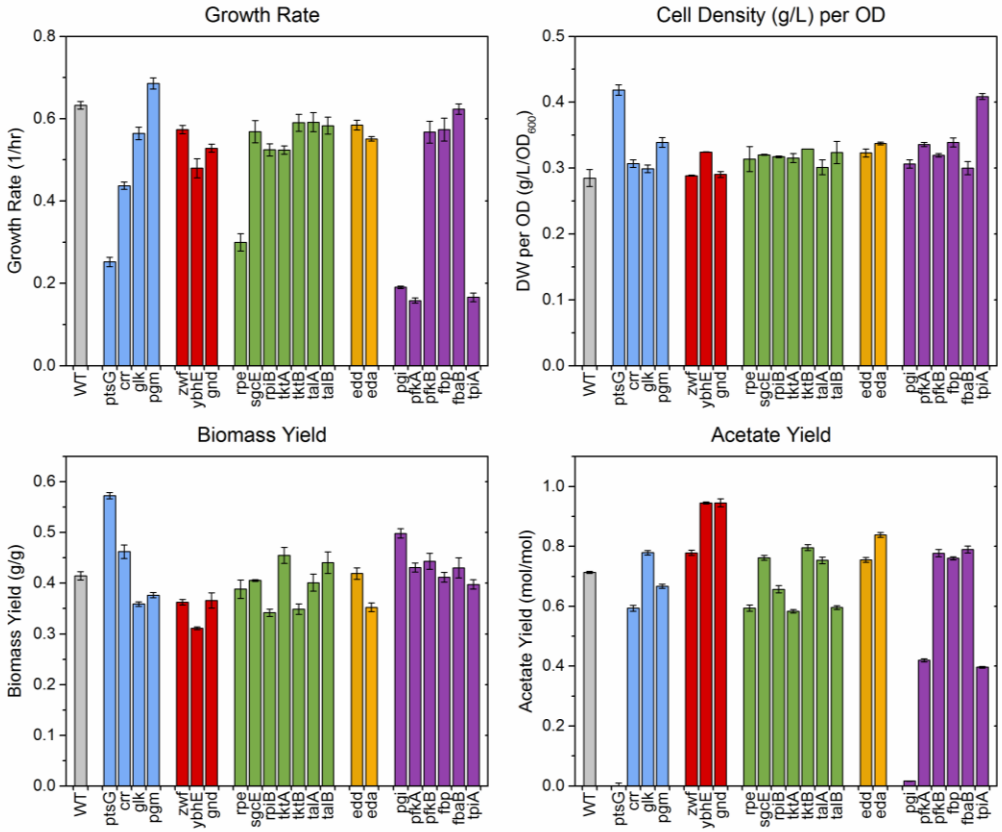


Figure 6.2: Measured physiological parameters. Bar colors reflect pathway assignment (wild-type: gray, transporters and phosphoglucomutase: blue, oxidative pentose phosphate pathway: red, non-oxidative pentose phosphate pathway: green, Entner-Doudoroff pathway: orange, upper EMP pathway: purple). Error bars indicate standard errors of the mean for growth rate (n=3) and cell density per OD (n=2), and standard errors attributable to regression and measurement error for biomass and acetate yield.

6.3.1.5 Glucose Uptake Rates

The calculated glucose uptake rates are shown in Figure 6.3. The uptake rate for wild-type strain was 8.5 mmol/g_{dw}/hr, consistent with previous reports (Crown et al., 2015a; Haverkorn van Rijsewijk et al., 2011; Leighty and Antoniewicz, 2013). Large reductions in glucose uptake rate were observed in some of the slow growing strains, with *ΔptsG*, *Δpgi*, *ΔpfkA*, and *ΔtpiA* all at approximately 2 mmol/g_{dw}/hr. Additionally, *Δcrr* and *Δrpe* had lower glucose uptake rates at approximately 5 mmol/g_{dw}/hr. Interestingly, two knockouts consumed glucose at a higher rate than the wild-type, the fastest being *Δpgm* at 10.1 mmol/g_{dw}/hr. Other knockouts had uptake rates similar to or slightly less than the wild-type.

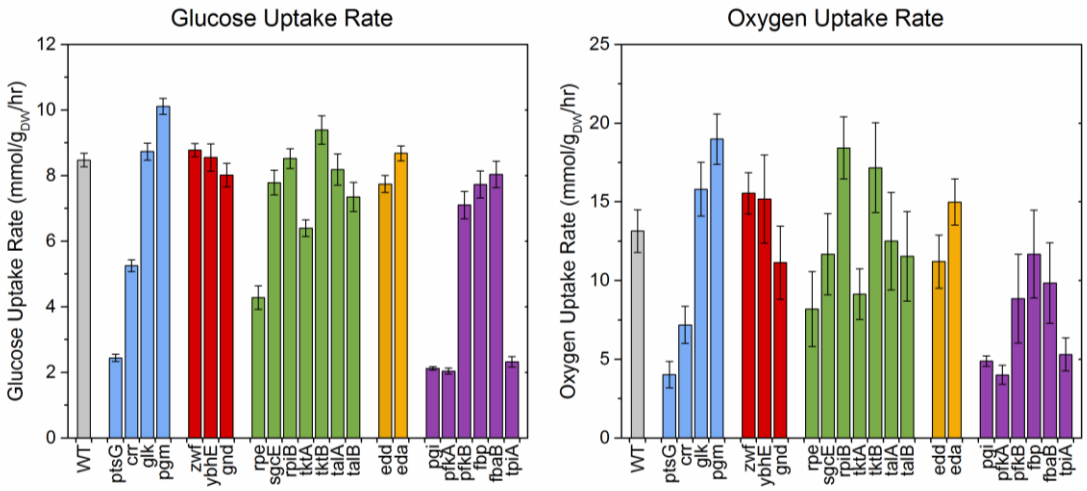


Figure 6.3: Calculated uptake rates of glucose and oxygen. Bar colors reflect pathway assignment (wild-type: gray, transporters and phosphoglucomutase: blue, oxidative pentose phosphate pathway: red, non-oxidative pentose phosphate pathway: green, Entner-Doudoroff pathway: orange, upper EMP pathway: purple). Error bars reflect the propagation of the directly measured standard errors.

6.3.1.6 Gas Exchange Rates

Specific oxygen uptake (Figure 63, right) and carbon dioxide evolution rates were calculated from redox and carbon balances, respectively (see also Figure S1). Despite relatively larger uncertainty in these estimated rates, resulting from propagation of measurement errors, some significant variations were still apparent. The specific oxygen uptake rate of wild-type was estimated to be ~13 mmol/g_{dw}/hr, consistent with prior literature (Chen et al., 2011; Leighty and Antoniewicz, 2013). The strains *ΔptsG*, *Δpgi*, *ΔpfkA*, and *ΔtpiA* consumed oxygen at a much lower rate, ~4 to 5 mmol/g_{dw}/hr; while other strains consumed oxygen at a higher rate than wild-type, particularly *Δpgm* and *ΔrpiB* (~19 mmol/g_{dw}/hr). The carbon dioxide secretion rates and respiratory coefficients (RQ = CO_2 produced per O_2 consumed) are shown in Figure 6.4. The RQ values for all strains were between 1.1 and 1.3, with no significant outliers.

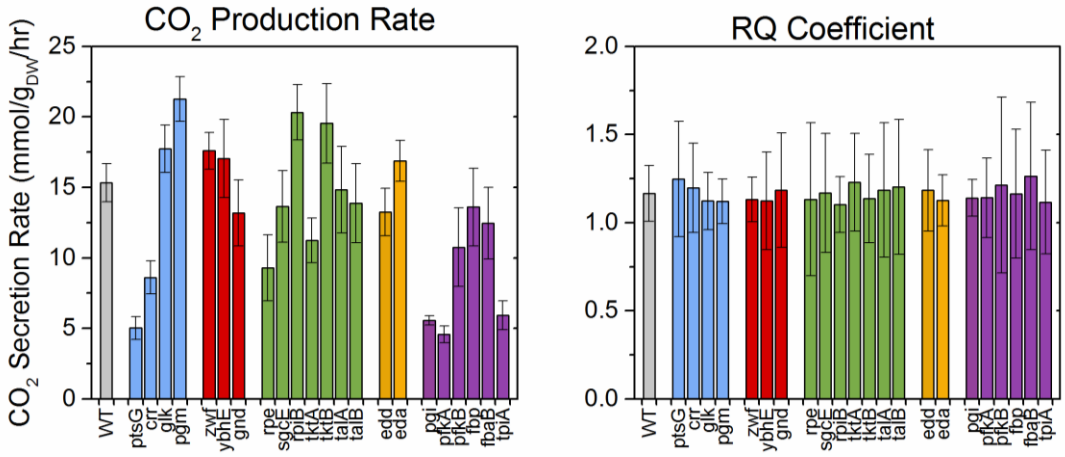


Figure 6.4: Calculated carbon dioxide secretion rates and RG coefficient. Bar colors reflect pathway assignment (wild-type: gray, transporters and phosphoglucomutase: blue, oxidative pentose phosphate pathway: red, non-oxidative pentose phosphate pathway: green, Entner-Doudoroff pathway: orange, upper EMP pathway: purple). Error bars reflect the propagation of the directly measured standard errors.

6.3.1.7 Biomass Composition

The compositions of biomass for all 23 strains are shown in Figure 6.5. We observed several striking differences in the relative distribution of the four major biomass components, i.e. proteins, RNA, lipids and glycogen. The total protein content varied between 45 wt% and 59 wt% (wild-type strain 51 wt%). The RNA content ranged from 14 wt% to 23 wt%, with most knockouts having lower RNA content than wild-type (21 wt%). The lipid content varied significantly between 3.7 wt% and 6.4 wt% (wild-type strain 4.7 wt%). The glycogen content also varied widely, ranging from 0.6 wt% to 8.9 wt%, with most knockout strains accumulating more glycogen than the wild-type (3.2 wt%). A notable exception was *Δpgm*, which had almost no glycogen, consistent with the gene's key role in generating the precursor glucose 1-phosphate for glycogen synthesis. The relatively large variations

in biomass composition highlight the relevance of these measurements for interpreting the metabolic and physiological states of these strains.

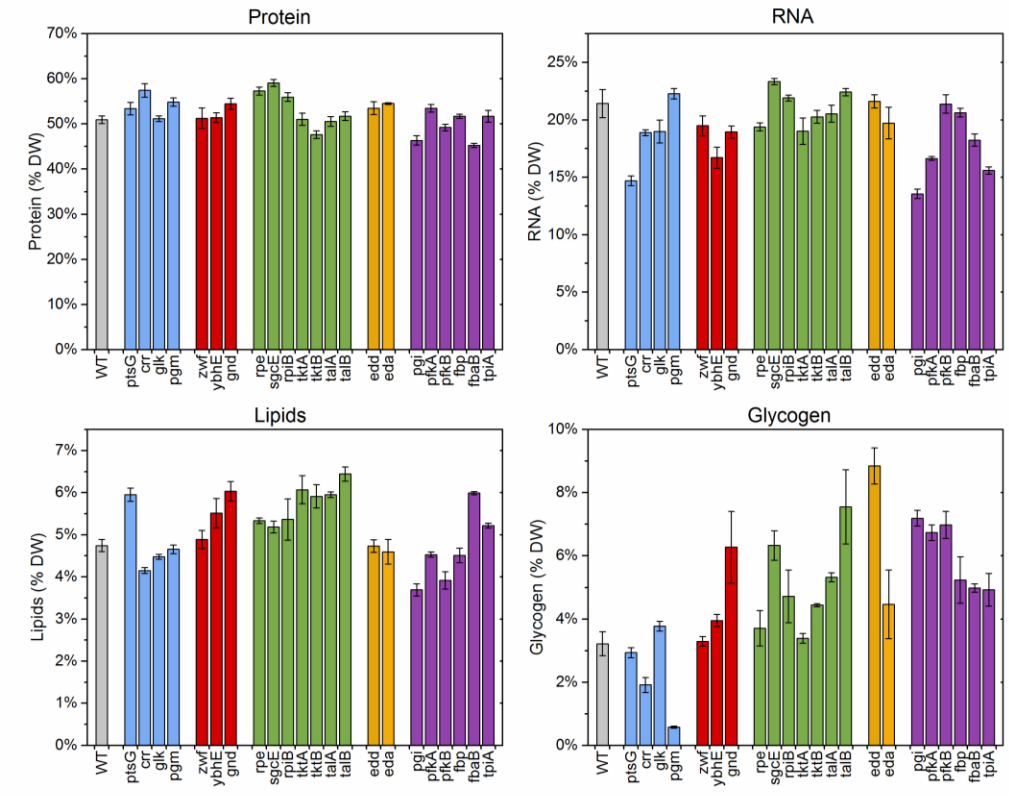


Figure 6.5: Dry weights for major biomass components. Bar colors reflect pathway assignment (wild-type: gray, transporters and phosphoglucomutase: blue, oxidative pentose phosphate pathway: red, non-oxidative pentose phosphate pathway: green, Entner-Doudoroff pathway: orange, upper EMP pathway: purple). Error bars represent standard errors of the mean (n=4, 2 biological replicates with 2 technical replicates each).

The normalized distributions of amino acids (per gram of protein) were constant for all strains (Figure 6.6). In contrast, there was large variability in the distribution of fatty acids, particularly amongst C16:0, C16:1 and C18:1 fatty acids

(Figure 6.7). Interestingly, all slow-growing strains (*Δrpe*, *Δpgi*, *ΔpfkA*, and *ΔtpiA*) had elevated C16:0 levels and decreased C18:1 levels. Another strain that stood out was *Δfbp*, which had more C16:1 than C16:0.

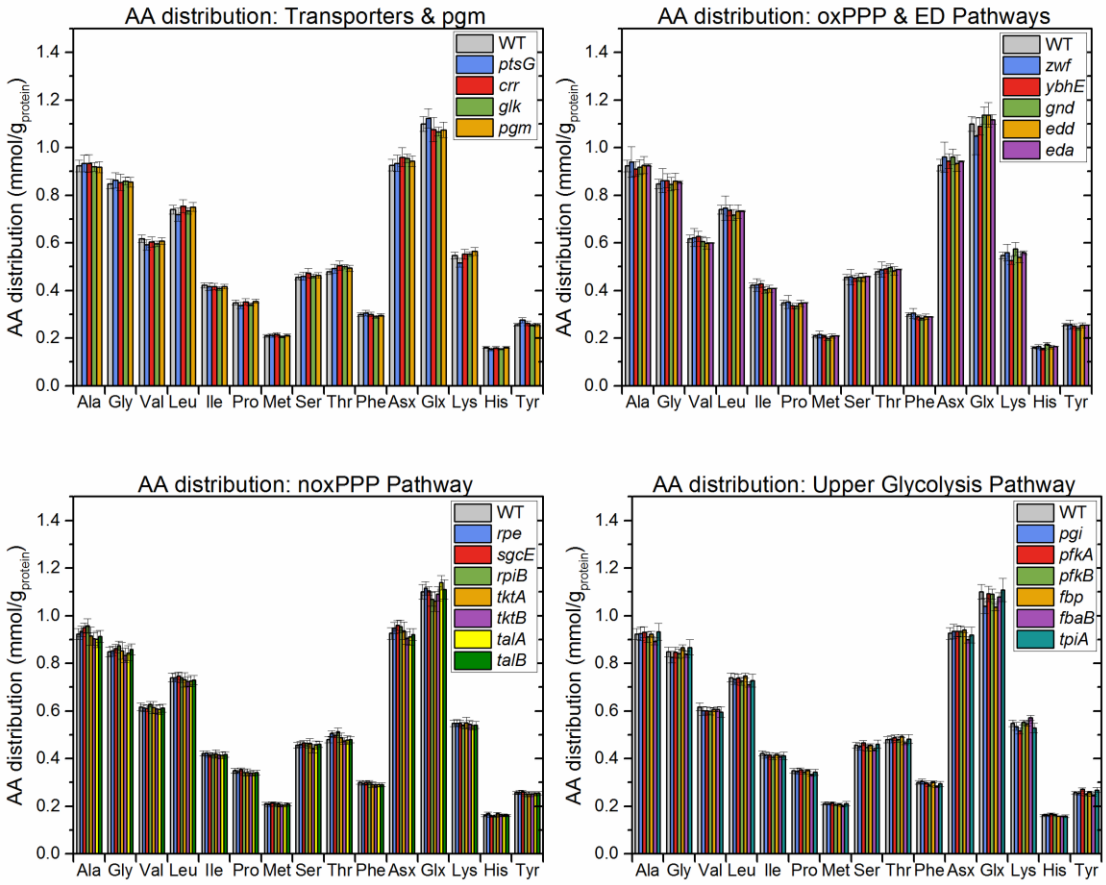


Figure 6.6: Amino acid distribution of each strain. Error bars represent standard errors of the mean (n=4, 2 biological replicates with 2 technical replicates each).

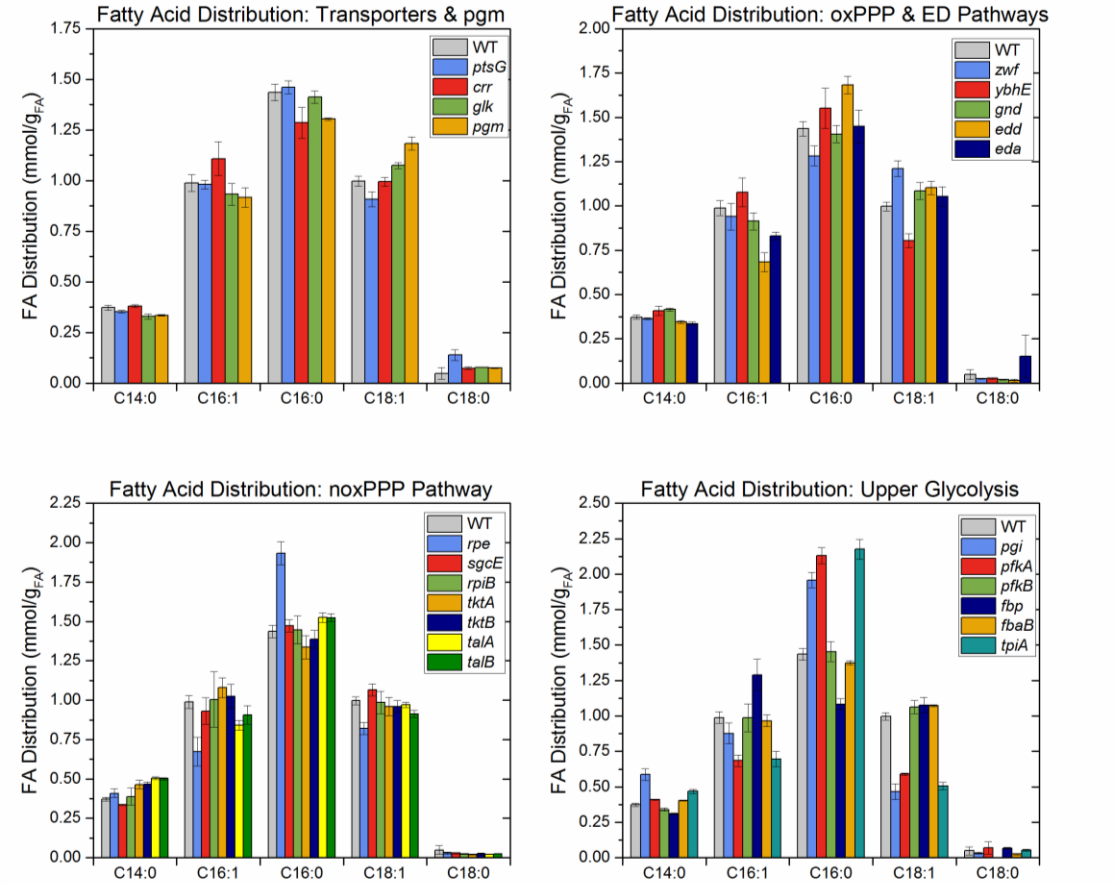


Figure 6.7: Amino acid distribution of each strain. Error bars represent standard errors of the mean (n=4, 2 biological replicates with 2 technical replicates each).

The calculated degrees of reduction and molecular weights of biomass are listed in Table 6.2. The degrees of reduction had minimal variation, ranging from 4.34 to 4.45 electrons per Cmol of biomass. The molecular weight of biomass also varied minimally, ranging from 26 to 27 g_{dw} per Cmol of biomass.

Strain	**Degree of Reduction**	**MW(g/mmol-C)**
WT	4.371	0.0270

ptsG	4.414	0.0261
crr	4.340	0.0262
glk	4.368	0.0268
pgm	4.353	0.0265
zwf	4.382	0.0268
ybhE	4.414	0.0265
gnd	4.414	0.0267
rpe	4.383	0.0264
sgcE	4.373	0.0268
rpiB	4.385	0.0267
tktA	4.426	0.0266
tktB	4.432	0.0270
talA	4.422	0.0269
talB	4.434	0.0272
edd	4.381	0.0272
eda	4.366	0.0267
pgi	4.384	0.0268
pfkA	4.380	0.0266
pfkB	4.363	0.0272
fbp	4.363	0.0272
fbaB	4.451	0.0270
tpiA	4.398	0.0269

Table 6.2: Calculated degree of reduction and molecular weight of each strain.

6.3.2 Correlations in Physiological Data and PCA Analysis

Next, we analyzed the measured growth data, yield data, and biomass composition data for pairwise correlations. The results are summarized in Figure 6.8, and scatter plots of several highly correlated physiological parameters are shown in

Figure 6.9. Several correlations that we identified were anticipated based on prior literature reports (see below), while other correlations were more surprising. The growth rate correlated with several physiological parameters, including acetate yield, RNA content, and the content of the fatty acids C16:0 and C18:1. There were also clear trade-offs between biomass and acetate yields, and between fatty acids, particularly C16:1 vs. C16:0, and C16:0 vs. C18:1. The strong correlations between biomass composition and growth rate are also highlighted in Figure 6.9. The RNA dependence on growth rate has been reported previously (Neidhardt, 1987; Pramanik and Keasling, 1997), and is thought to reflect the need for more ribosomes to support fast growth rates. The trade-off in fatty acid composition, on the other hand, in which the unsaturated C18:1 is preferred to the saturated C16:0 at faster growth rates, has not been previously reported to our knowledge.

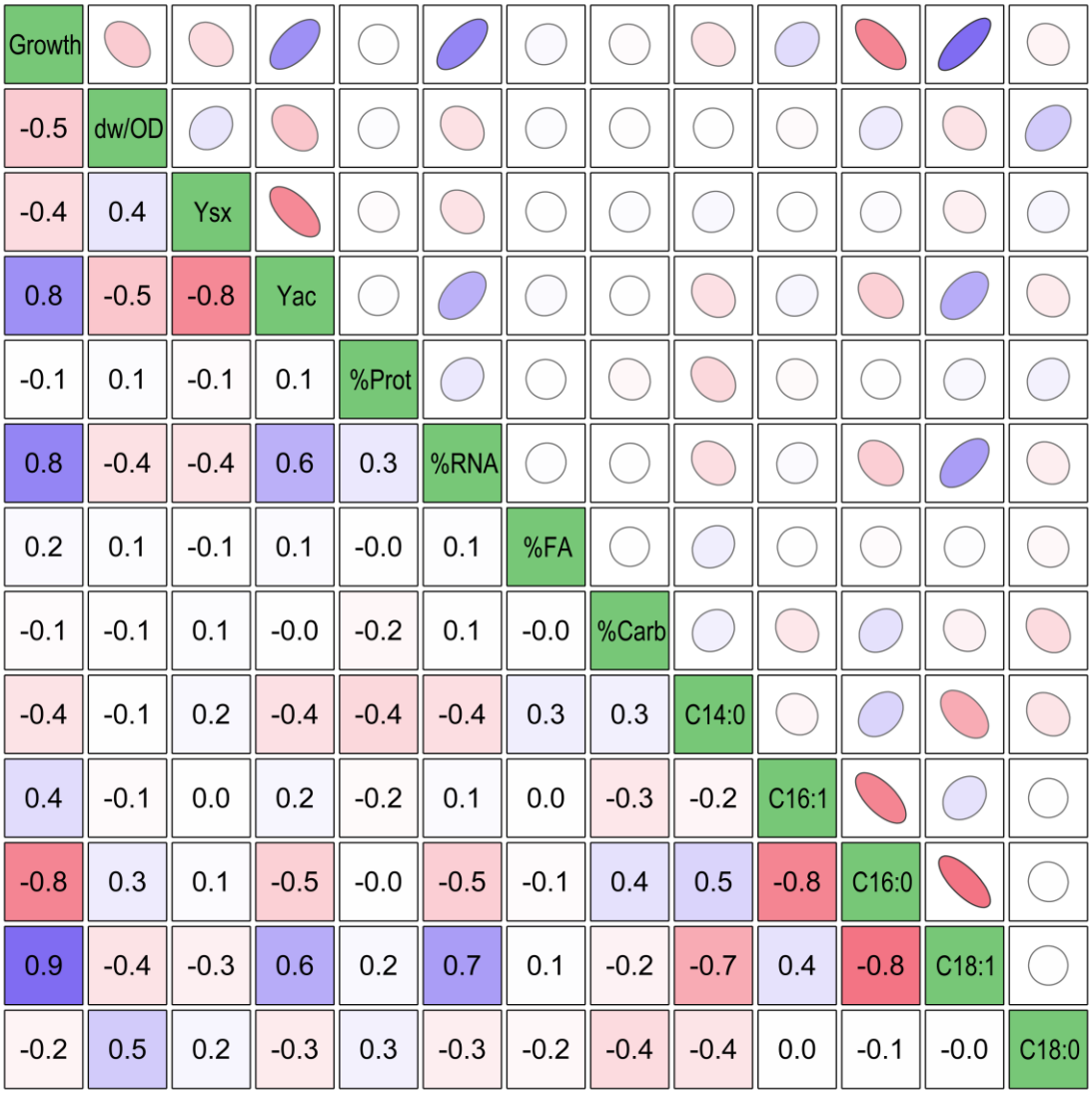


Figure 6.8: Exhaustive pairwise correlation coefficients for all measured data. The coefficients are given in the lower triangle, and the quality and direction of the correlation is represented visually by ellipsoids in the upper triangle (more elongated ellipsoid = higher quality correlation). The coloring is scaled to reflect value from -1.0 (red) to 0 (white) to 1.0 (blue). The included data sets (left to right) are growth rate (h^{-1}) , dry weight per OD ($g/L/OD_{600}$), biomass yield (g/g), acetate yield (mol/mol), percentages of the four major biomass components, and the relative fatty acid contents (mmol/g(lipid)). All coefficients greater than 0.4 indicate a significant nonzero correlation at 95% confidence.

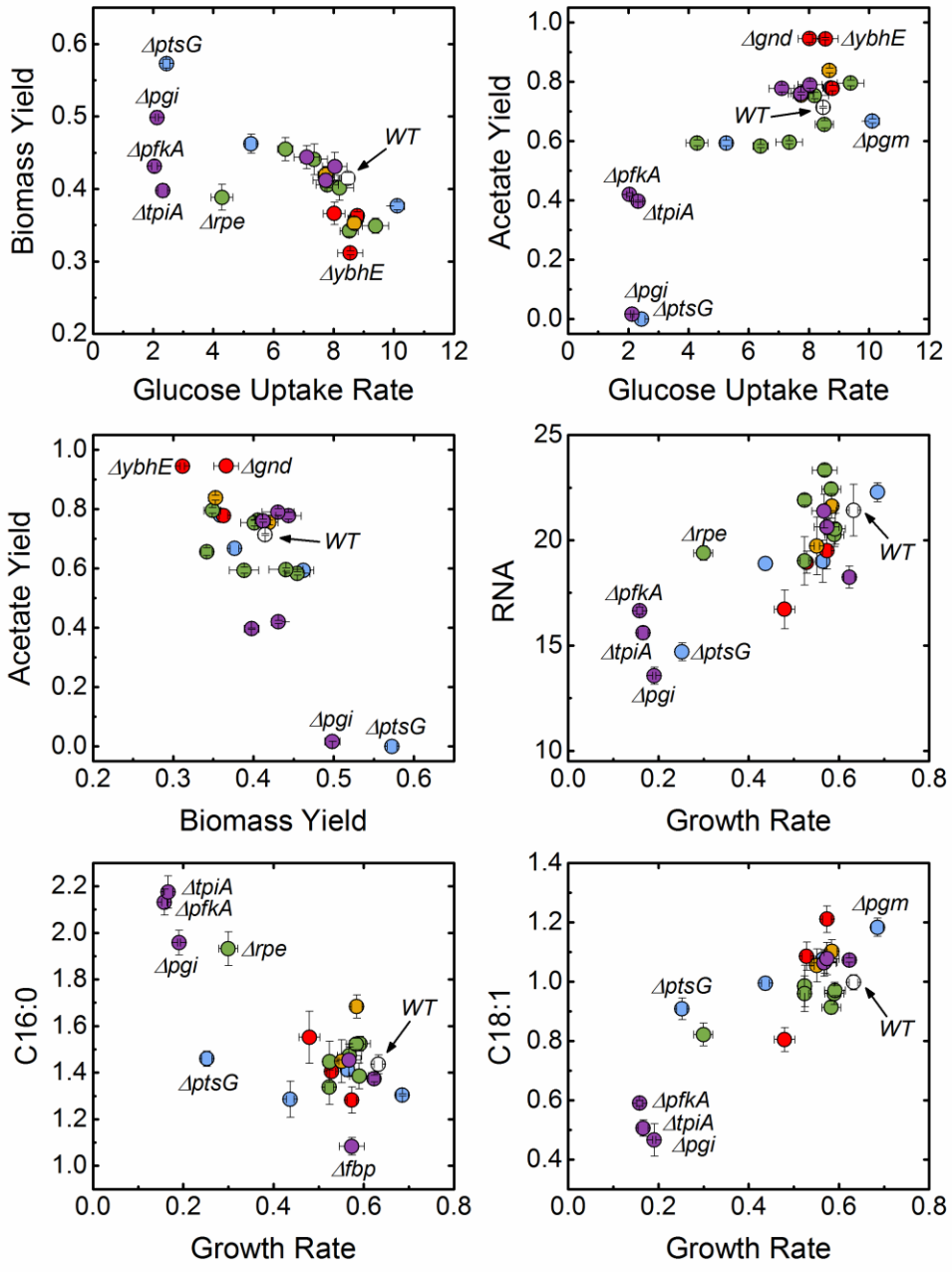
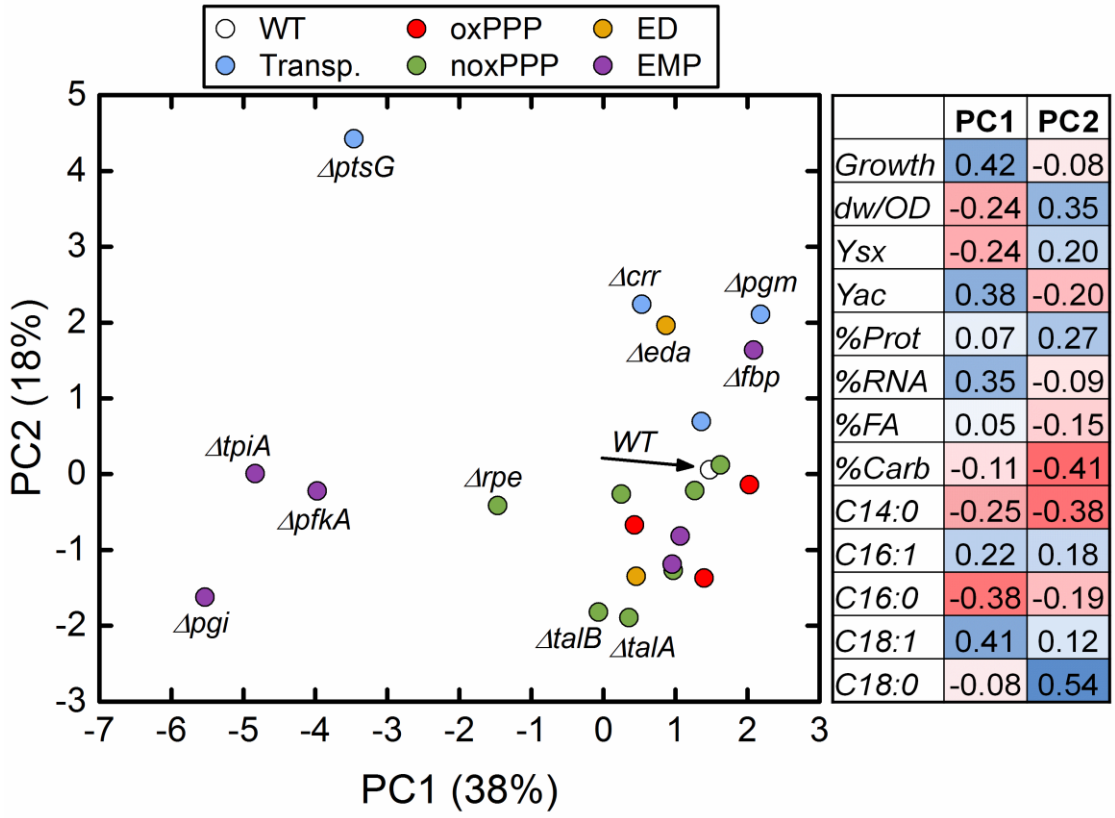


Figure 6.9: Scatter plots of correlated data. Marker colors reflect pathway assignment (wild-type: gray, transporters and phosphoglucomutase: blue, oxidative pentose phosphate pathway: red, non-oxidative pentose phosphate pathway: green, Entner-Doudoroff pathway: orange, upper EMP pathway: purple).

Also shown in Figure 6.9 are the correlation between glucose uptake rate and both biomass and acetate yields, which has implications in terms of the acetate "overflow" metabolism interpretation (Basan et al., 2015; Majewski and Domach, 1990). *E. coli* is believed to produce acetate under aerobic conditions in order to maximize glucose uptake and growth rate under constrained oxidative phosphorylation capacity. Therefore, at a lower glucose uptake rate, a relatively higher oxidative phosphorylation flux would result in higher biomass yields and lower acetate flux. For example, Basan et. al (Basan et al., 2015) showed a highly linear dependence of acetate flux on growth rate. In our data, the expected trends are present: biomass yield correlates negatively with glucose uptake rate and acetate yield correlates positively with glucose uptake rate. However, it is interesting to note the imperfect nature of this relationship in our data, especially relative to the results of Basan et al. These results indicate that the gene knockouts are likely perturbing the system in ways that exceed this relatively well-understood phenomenon.

Principal component analysis (PCA) was also applied to explore correlations in our data set. PCA is a data reduction technique that can be used to identify redundant and linearly dependent measurements, allowing for the identification and interpretation of nontrivial variation in large multidimensional data sets such as is presented here. PCA defines a new lower-dimensional space spanned by new variables, or "principal components", that are linear combinations of the original variables and that capture the maximum amount of original variation in the data. For PCA analysis, all data were normalized and standardized prior to analysis. The results of PCA analysis are shown in Figure 6.10. Principal components 1 and 2, accounted for 56% of the total data variability. PC1 mostly captured the growth rate

dependencies discussed above (i.e. covariation of growth rate with acetate yield, %RNA, C16:0, and C18:1), while PC2 captured subtler correlations mostly in biomass composition data, particularly between %Carb, C14:0, and C18:0 (see coefficients in Figure 6.10). The various *E. coli* knockouts clustered in informative ways. For example, knockouts of the dominant enzymes in the EMP pathway (*Δpgi*, *ΔpfkA*, *ΔtpiA*) clustered in a region with low PC1 and low PC2 values. These knockouts are likely to disrupt the highly active glycolytic flux of wild-type cells. *ΔptsG* was an outlier with low PC1 and high PC2 values. Both isozymes of *ΔtalA* and *ΔtalB* also clustered together with low PC2 values, and several strains clustered with high PC1 and moderately high PC2 values. Interestingly, the two knockouts of the ED pathway (*Δedd* and *Δeda*) did not cluster together. Hierarchical clustering was also applied as an alternative clustering approach, and the resulting dendrogram is included in Figure 6.11. Overall, similar general relationships were observed with hierarchical clustering and PCA analysis, for example, the clustering of *Δpgi*, *ΔpfkA*, *ΔtpiA* and *Δrpe*, and the largest distances being between wild-type and this cluster, and between wild-type and *ΔptsG*.



	PC1	PC2
Growth	0.42	-0.08
dw/OD	-0.24	0.35
Ysx	-0.24	0.20
Yac	0.38	-0.20
%Prot	0.07	0.27
%RNA	0.35	-0.09
%FA	0.05	-0.15
%Carb	-0.11	-0.41
C14:0	-0.25	-0.38
C16:1	0.22	0.18
C16:0	-0.38	-0.19
C18:1	0.41	0.12
C18:0	-0.08	0.54

Figure 6.10: Principal component analysis (PCA) plot showing the first two components, which together account for more than half of the total variation in the data. The coefficients mapping these components to the original (normalized and standardized) data are shown in the table to the right.

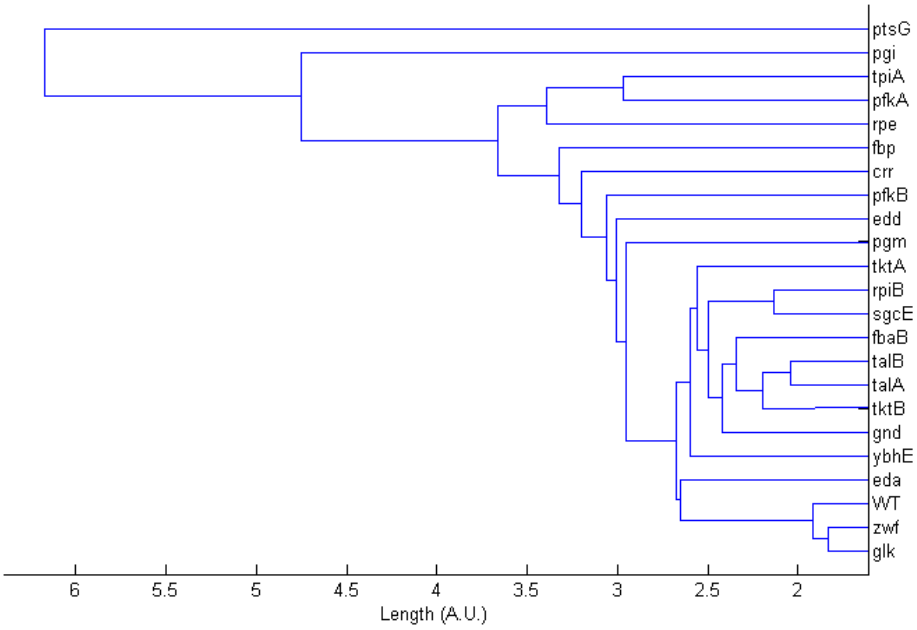


Figure 6.11: Hierarchical clustering of the knockout strains.

6.3.3 Evaluating COBRA Modeling Predictions

Next, we applied several widely used COBRA modeling approaches to predict several of the directly measured physiological characteristics for the 23 strains. For FBA, the upper bounds on glucose and oxygen uptakes were taken to be those of the wild-type *E. coli*. The wild-type glucose uptake and oxygen uptake rates applied were 8.5 mmol/g_{dw}/hr and 12 mmol/g_{dw}/hr, respectively. The former was as measured in this study, and the latter value was chosen because it is within the experimental error of measurements in this study, is consistent with previous studies (Chen et al., 2011; Leighty and Antoniewicz, 2013), and results in FBA predictions of wild-type growth rate, biomass yield, and acetate yield that are consistent with experiment (Figure 6.12). These constraints are a necessary and common assumption of FBA in which the wild-

type is assumed to be optimal. No uptake rate constraints were applied for MOMA or RELATCH. The reference state, generated as described in the Methods section, also had good agreement with the experimental wild-type growth rate, biomass and acetate yields (Figure 8). Of the 22 knockout strains studied here, 13 were knockouts of isozymes catalyzing the same reaction, which do not impose new stoichiometric constraints, thus allowing FBA and MOMA to find the trivial wild-type solution. The same is true for *Δglk*, *Δedd*, *Δeda*, and *Δfbp*, which do not carry significant flux in the wild-type (both in the FBA solution and according to (Leighty and Antoniewicz, 2013)). In contrast, RELATCH utilizes gene expression data to calculate the metabolic costs of adjustment associated with each individual gene knockout, including isozymes.

The performance of the three models in terms of predicting growth rate, biomass yield, and acetate yield are summarized in Figure 6.12. Correlation coefficients were calculated to quantify the agreement between prediction and measurement. In general, FBA predicted very few deviations from the wild-type physiology for the 22 knockouts. This resulted in poor agreement with the data, with correlation coefficients of 0.43, 0.17, and -0.12 for growth, biomass yield, and acetate yield respectively. As assuming growth rate optimization has long been considered invalid for un-evolved knockouts, this poor performance is not surprising. MOMA performed better in some cases, for example in predicting growth rates for the oxidative pentose phosphate pathway knockouts (predicted: 0.51-0.52 hr^{-1}, observed: 0.58, 0.48, 0.53 hr^{-1}), *Δpgi* (predicted: 0.46, observed: 0.19 hr^{-1}) and *ΔtpiA* (predicted: 0.31, observed: 0.16 hr^{-1}). The overall growth rate correlation coefficient was 0.51. MOMA had a very poor overall agreement with the biomass yield data (ρ=0.07), and

in particular overestimated the biomass yield penalties for *Δpgi* (predicted: 0.31 g_{dw}/g, observed: 0.46 g_{dw}/g) and *ΔtpiA* (predicted: 0.15 g_{dw}/g, observed: 0.39 g_{dw}/g). MOMA predicted almost no changes to acetate yields across all strains except *ΔtpiA* (0.55 mol/mol), also resulting in poor overall agreement (ρ=0.24). Of the three models tested here, RELATCH predictions agreed best with measurement for growth rate (ρ=0.54), biomass yield (ρ=0.52), and acetate yield (ρ=0.40). This is partially a result of the fact, as mentioned above, that RELATCH was the only one of the three without a large number of trivial (wild-type) solutions for knockouts. While performing best of this set of COBRA methods, RELATCH did not fully capture the range of growth rates observed. The growth rates of all of the slowest growing strains were over-estimated. While performing by far the best in terms of biomass yield, it did not predict the high yield for *ΔptsG*. The acetate yield predictions of RELATCH were generally low relative to the measured values, but high for the non-producing strains *ΔptsG* and Δ*pgi*.

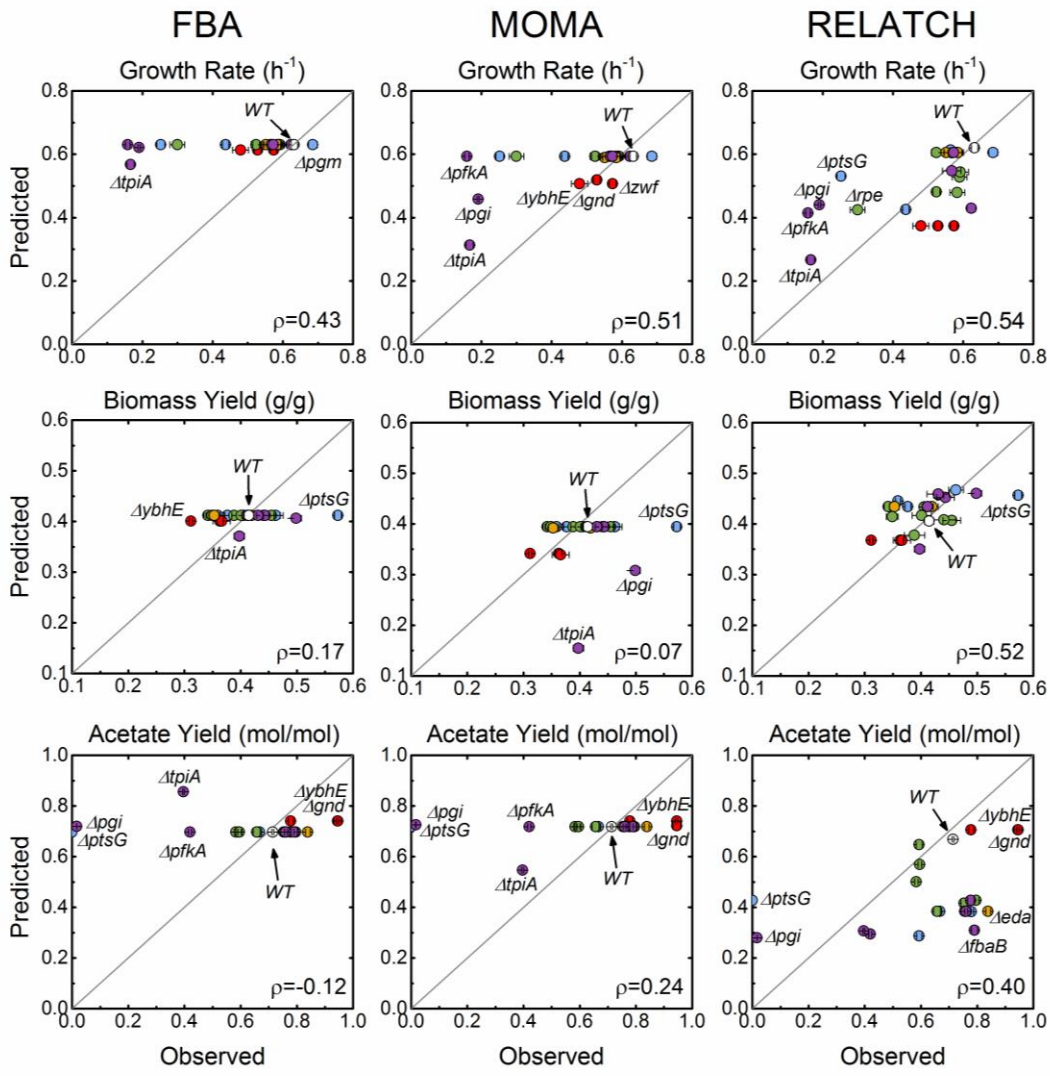


Figure 6.12: Comparison of experimental growth rates, biomass yields, and acetate yields to those predicted by three constraint-based modeling approaches (FBA: flux balance analysis, MOMA: minimization of metabolic adjustment, RELATCH: relative optimality in metabolic networks). Marker colors reflect pathway assignment (wild-type: gray, transporters and phosphoglucomutase: blue, oxidative pentose phosphate pathway: red, non-oxidative pentose phosphate pathway: green, Entner-Doudoroff pathway: orange, upper EMP pathway: purple). Correlation coefficients (ρ) describe the agreement between prediction and measurement. Wild-type data were excluded from this correlation.

6.4 Discussion

The phenotypic data collected and reported here demonstrate that there are significant metabolic and physiological changes associated with the knockout of central carbon metabolism enzymes. The "rewired" metabolic phenotypes are direct responses to the altered metabolic network, with effects of stoichiometry, kinetics, and regulation manifested. The study of this rewiring can offer new insights into these areas. In the knockouts studied here, we observed a 4.3 fold range of growth rates, a 1.8 fold range in biomass yields, and a wide range of acetate producing phenotypes from non-producing (~0 mol/mol for *Δpgi*) to high-producing (~1 mol/mol for *ΔybhE* and *Δgnd*). The corresponding metabolic rates (glucose uptake, oxygen uptake, carbon dioxide secretion, and acetate secretion) had correspondingly high variations. These results demonstrate that under substrate-rich conditions, the impact of many central carbon metabolism knockouts on the metabolic phenotype is dramatic. This contrasts with previous studies in which, for example, at low dilution rates of 0.2 hr^{-1} no acetate was produced by any knockout strains, including many of the same as are presented here (Ishii et al., 2007).

There are several interesting cases to which more specific analyses can be applied. For example, it was unexpected that two knockouts to complimentary components of the PTS glucose transport (*ΔptsG* is the glucose-specific transporter, and *Δcrr* is its first partner phosphotransferase) displayed different growth and glucose uptake rates (*Δcrr* took up glucose approximately twice as fast). This could indicate the presence of regulatory effects or kinetic differences as alternative transporters or phosphotransferases are utilized. Also of note are the phenotypes of the many isozymes studied. Some of these knockouts were severely impacted in terms of growth, indicating a heavy reliance on one dominant isozyme. This was the case for

Δrpe and *ΔpfkA*, which had been previously reported to be the dominant isozymes (Kotlarz et al., 1975; Lyngstadaas et al., 1998). This was even more severely the case for fructose bisphosphate aldolase, where the primary *fbaA* is essential for *E. coli* growth even with rich media (Baba et al., 2006). In contrast, the growth rate was not nearly as affected for *tktA/B* and *talA/B*, indicating perhaps a more equitable and flexible distribution of the flux load. However, significant differences in biomass and acetate yields and biomass composition indicate that measurable phenotypic changes do occur in these cases. The one knockout studied here that was not completely integrated into a central carbon metabolic pathway was *Δpgm*. Phosphoglucomutase interconverts G6P and G1P, and under glucose rich conditions generates G1P for glycogen biosynthesis (Eydallin et al., 2007). In this knockout, we observed an elevated growth rate, higher glucose and oxygen uptake rates, and negligible glycogen content, in agreement with results reported by Eydallin et al. These results show quantitatively the metabolic and growth "opportunity cost" that *E. coli* pays for security against future glucose starvation.

Looking at the growth and yield data in aggregate, we can extract general trends and identify outliers of interest. For example, we observed a positive correlation between acetate yield and glucose uptake rate (Figure 6.9). This is consistent with the general interpretation of acetate overflow metabolism (Majewski and Domach, 1990). However, we can see deviations from this trend, including very slowly growing strains that still produce significant amounts of acetate such as *ΔpfkA* and *ΔtpiA*. Similarly, acetate yield correlates negatively with biomass yield, as would be expected given the trade-off of carbon fates; and biomass yield correlates negatively with glucose uptake rate. Along with the two strains mentioned above, *Δrpe* also clearly falls into a

category of knockout strains that grows both slowly and inefficiently. It is notable that while the severity of the EMP disruptions (also including *Δpgi*) may be intuitive due to the large wild-type glycolytic flux, the RPE flux is much less (Crown et al., 2015a; Leighty and Antoniewicz, 2013), showing that disruptions even in nominally small fluxes can have large effects.

The presented biomass composition information also provided valuable perspective and opportunity for insight. It is notable that the total protein content varied only slightly, and the relative amino acid composition did not vary at all. However RNA, fatty acids, and glycogen contents exhibited much more variation. We demonstrated that RNA content correlated positively with growth rate, which had been previously reported (Neidhardt, 1987; Pramanik and Keasling, 1997). However, we did not observe the also previously reported negative correlations between growth rate and both protein and glycogen content (Neidhardt, 1987; Pramanik and Keasling, 1997). Many knockouts exhibited an increased level of glycogen storage, which could indicate that it is an effect of a general stress response. Most surprising were strains that did not have impaired growth or other significant differences from the wild-type, but had much higher glycogen levels, such as *Δedd*, *ΔtalB*, and *ΔpfkB*. The total fatty acid content ranged moderately both higher and lower than the wild-type, but we observed striking correlations in fatty acid content. Faster growing strains had a more equitable distribution of C16:1, C16:0, and C18:1 fatty acids. Slower growing strains, however, contained much more C16:0, which came primarily at the expense of C18:1. These fatty acids are associated with the phospholipid membrane, so the differing compositions likely correspond to different membrane properties, e.g. membrane fluidity and stability (Zhang and Rock, 2008). On the other hand, C14:0, which did not

change much between the strains, is primarily associated with lipopolysaccharide (LPS) (Neidhardt, 1987). Some growth rate effects on membrane lipid have been identified previously (Shokri et al., 2002), but the identified trends were not as monotonic or clear as is reported here. The exact mechanism behind this correlation is worthy of future investigation.

These correlations and others were captured also in the presented PCA analysis. Clustering was observed among knockout strains with similar roles in a pathway, particularly the dominant EMP enzymes *Δpgi*, *ΔpfkA*, and *ΔtpiA*. These knockouts were likely to have similar physiological effects such as sugar-phosphate stress from accumulation of glycolytic intermediates (Richards et al., 2013). The *ΔptsG* strain was also significantly unique from all other strains. Surprisingly, a significant distance was observed between the two ED pathway knockouts, *Δedd* and *Δeda*. This was due to differences in biomass composition, particularly higher glycogen levels and less C18:0 in *Δedd*. It is not known by what mechanism these two strains would be affected in this way, as this pathway carries negligible flux in the wild-type. Both the PCA analysis and hierarchical clustering provided a quantitative sense of the similarity between all studied strains, and could identify the knockout strains with the most severe impacts relative to the wild-type strain.

Metabolic networks have been most thoroughly studied and understood thus far at the level of stoichiometry. This is reflected in the prevalence of the COBRA modeling tools. Using the external flux data collected here, we have applied three COBRA methods, FBA, MOMA, and RELATCH, to determine whether the observed phenotypes could be predicted on the basis of stoichiometry and wild-type phenotypic data alone (fluxes for MOMA; and fluxes and gene expression for RELATCH). While

RELATCH performed the best of the three, all fell short of broad quantitative accuracy. For the isozymes, which make up a large fraction of the included knockout set, any deviation from the wild-type could not be predicted on the basis of stoichiometry alone by FBA or MOMA. These results demonstrate the difficulty of predicting knockout flux behavior, particularly in nutrient-rich environments where glucose uptake rates can vary significantly. In this regime, there are likely to be significant kinetic and/or regulatory effects, which further study of knockout metabolism will help to elucidate.

6.5 Conclusions

Knockouts of central carbon metabolism enzymes in *E. coli* show significant physiological and metabolic changes, or "metabolic rewiring". This is particularly true in the case of unconstrained growth conditions used in this study, in contrast to previous reports of robustness seen in slow dilution rate chemostat cultures (Ishii et al., 2007). Large variations in growth rate, biomass yield, acetate yield, uptake rates, and biomass composition were observed. The external metabolic fluxes of glucose uptake, biomass generation, and acetate generation provide a glimpse into the kinetic limitations and regulatory adjustments occurring in these knockouts. These also provide an important data set for assessing and improving metabolic modeling, as we have demonstrated that common COBRA modeling approaches do not accurately predict observed behavior. Significantly more detailed understanding of these rewired metabolisms, and commensurate modeling opportunities, will be possible with more detailed ^{13}C-MFA studies. The results presented also demonstrate the importance and complementarity of biomass composition measurements to the other physiological measurements. Relationships between general composition changes and common

physiological changes like altered growth rate were observed in RNA and fatty acid composition. Large differences in glycogen composition, uncorrelated with growth or other measured parameters, could offer insights into other types of stress response in these strains.

6.6 Author Contributions

CPL performed the experiments and analysis except the following. JEG performed the analysis of excreted metabolites, and NRS did the genetic work necessary to show that *ΔrpiA* was likely essential in these conditions. CPL and MRA designed the project and wrote the paper with help from all authors.

Chapter 7

METABOLIC FLEXIBILITY AND LIMITATIONS REVEALED IN FLUXOMIC RESPONSES TO DELETION OF 20 GENES IN UPPER CENTRAL CARBON METABOLISM

Christopher P. Long, Maciek R. Antoniewicz

7.1 Introduction

Quantitative and predictive understanding of cell metabolism is a fundamental objective in cell biology and bioengineering. Since the mapping of most major metabolic enzymes and associated genes over the past century (Caspi et al., 2012; Kanehisa and Goto, 2000), focus has shifted to studying the systems-level properties of metabolism, including kinetics and regulation. A commonly applied and fruitful approach to generating systems information is to study perturbation responses, such as to environmental or genetic changes (Long and Antoniewicz, 2014a). Knockouts of genes encoding metabolic enzymes and regulators (Haverkorn van Rijsewijk et al., 2011) have been used to study regulatory responses, kinetics (e.g, identify potential rate limiting steps (Toya et al., 2010)), and even to identify novel reactions (Nakahigashi et al., 2009). Beyond any single discovery or analysis, such data also inform the development of metabolic models, for which predicting the genotype-phenotype relationship is a fundamental objective (Kim and Reed, 2012; Lewis et al., 2012; Segre et al., 2002; Shlomi et al., 2005). These models represent the expression of current systems-level understanding, and their utility is essential for the realization of rational, forward engineering of strains in biotechnology. Rigorous assessment of

existing models such as the constraint based reconstruction and analysis (COBRA) models, and development of next-generation models such as ensemble kinetic approaches (Khodayari and Maranas, 2016; Tran et al., 2008), have been limited by available experimental measurements (Long and Antoniewicz, 2014a).

In the model Gram-negative microbe *Escherichia coli*, gene perturbation studies have been significantly enabled by the Keio knockout strain collection (Baba et al., 2006). In published fluxomic studies, however, incomplete coverage in genes and differences in experimental conditions across studies have limited the generalized interpretation of results (Long and Antoniewicz, 2014a). Furthermore, recent advances in ^{13}C-metabolic flux analysis (^{13}C-MFA) methodology have significantly improved the precision and identifiability of fluxes (Maciek R. Antoniewicz, 2015a; Crown et al., 2016b; Long et al., 2016a). In this study, 20 gene knockout strains, which comprehensively span the upper portion of central carbon metabolism, were characterized under identical growth conditions. The chosen condition of aerobic batch growth on glucose minimal medium compliments a previous study which reported robust metabolic responses to 14 knockouts under substrate-limited conditions (Ishii et al., 2007), is relevant to many biotechnological applications, and offers novel insights into intracellular kinetic limitations as external limits to growth are removed. The results will provide a basis for better understanding metabolic limitations to growth, and enable the application of next-generation modeling to this important regime.

7.2 Materials and Methods

7.2.1 Materials

Chemicals and M9 minimal medium were purchased from Sigma-Aldrich (St. Louis, MO). Isotopic tracers were purchased from Cambridge Isotope Laboratories (Tewksbury, MA): [1,6-13C]glucose (99.2 % 13C), and [1,2-13C]glucose (99.7 %). The isotopic purity and enrichment of all tracers were validated by GC-MS analysis (Cordova and Antoniewicz, 2016; Sandberg et al., 2016). All solutions were sterilized by filtration.

7.2.2 Strains and Culture Conditions

E. coli strains were obtained from the Keio collection (GE Healthcare Dharmacon), which were generated by one-step inactivation of all non-essential genes in *E. coli* K-12 BW25113 (Baba et al., 2006). The specific strains used in this study are listed in Table 6.1 (except *ΔptsG* and *Δcrr*, excluded here), and were verified using PCR amplification of the knockout gene region (Long et al., 2016b). For ^{13}C-tracer experiments, strains were cultured aerobically in glucose M9 minimal medium at 37°C in mini-bioreactors with 10 mL working volume (Long et al., 2016b). Pre-cultures were grown overnight and then used to inoculate the experimental culture at an OD_{600} of 0.01, in which 2 g/L of the glucose tracers were present. Cells were harvested for GC-MS analysis at mid-exponential growth when OD_{600} was approximately 0.7. For the glucose secretion experiments, experimental cultures were inoculated as above, with 2 g/L of either [$^{2}H_{7}$]glucose (i.e., d_7 glucose) or [1,2-^{13}C]glucose. The experiment continued until glucose was entirely depleted.

7.2.3 Analytical Methods

Cell growth was monitored by measuring the optical density at 600nm (OD_{600}) using a spectrophotometer (Eppendorf BioPhotometer). The OD_{600} values were converted to cell dry weight concentrations using a previously determined OD_{600}-dry cell weight relationship for *E. coli* (1.0 OD_{600} = 0.32 g_{DW}/L) (Long et al., 2016b). After centrifugation, the supernatant was separated from the cell pellet. The cell pellets were washed with glucose-free M9 medium prior to subsequent analysis. Glucose concentrations in the media were determined using YSI 2700 biochemistry analyzer (YSI, Yellow Springs, OH). Acetate concentrations were determined using an Agilent 1200 Series HPLC (Gonzalez et al., 2017). Growth rate was calculated using linear regression of the natural logarithm of the OD_{600} and time.

7.2.4 Gas Chromatography-Mass Spectrometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 µm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230°C, the MS quad temperature at 150°C, the interface temperature at 280°C, and the inlet temperature at 250°C. GC-MS analysis of *tert*-butyldimethylsilyl (TBDMS) derivatized proteinogenic amino acids was performed as described (Long and Antoniewicz, 2014b). Labeling of glucose (derived from glycogen) and ribose (from RNA) were determined as described (Long et al., 2016a; McConnell and Antoniewicz, 2016). For the additional labeling measurement in *ΔtpiA*, the intracellular metabolite glycerol-3-phosphate was derivatized as described (Ahn et al., 2016). The labeling of medium glucose in the

glucose excretion experiments was determined as described (Antoniewicz et al., 2011). In all cases, mass isotopomer distributions were obtained by integration (Antoniewicz et al., 2007b) and corrected for natural isotope abundances (Fernandez et al., 1996).

7.2.5 Metabolic network model and ^{13}C-metabolic flux analysis

The metabolic network model used for ^{13}C-MFA is provided in Appendix Table C.1, and the full flux results are in the Supplemental File. The model (Crown et al., 2015a; Gonzalez et al., 2017) includes all major metabolic pathways of central carbon metabolism, lumped amino acid biosynthesis reactions, and a lumped biomass formation reaction. The coefficients for precursors in the biomass formation reaction were determined for each strain, based on the measured biomass compositions, as previously described (Antoniewicz et al., 2007c). Modifications to the model were made for two cases. For *ΔpfkA*, secretion of G6P was added, such that the net uptake of glucose was still 100. For *ΔtpiA*, the methylglyoxal pathway was added. To resolve this flux, an additional tracer experiment was performed using 50% [1-^{13}C]glucose and 50% [4,5,6-^{13}C]glucose. The labeling of the intracellular metabolite glycerol-3-phosphate, which is produced from DHAP, was measured and used in flux estimation. ^{13}C-MFA calculations were performed using the Metran software (Yoo et al., 2004), which is based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007a). Fluxes were estimated by minimizing the variance-weighted sum of squared residuals (SSR) between the measured and model predicted mass isotopomer distributions and acetate yield using non-linear least-squares regression. For integrated analysis of parallel labeling experiments, the data sets were fitted simultaneously to a single flux model (Leighty and Antoniewicz, 2013). Flux estimation was repeated 10

times starting with random initial values for all fluxes to find a global solution. At convergence, accurate 95% confidence intervals were computed for all estimated fluxes by evaluating the sensitivity of the minimized SSR to flux variations. Precision of estimated fluxes was determined as follows (Antoniewicz et al., 2006):

Flux precision (stdev) = [(flux $_{\text{upper bound 95\%}}$) − (flux $_{\text{lower bound 95\%}}$)] / 4

To describe fractional labeling of biomass amino acids G-value parameters were included in ^{13}C-MFA. As described previously (Antoniewicz et al., 2007c), the G-value represents the fraction of a metabolite pool that is produced during the labeling experiment, while 1-G represents the fraction that is naturally labeled (e.g., from inoculum). By default, one G-value parameter was included for each measured amino acid in each data set. Reversible reactions were modeled as separate forward and backward fluxes. Net and exchange fluxes were determined as follows: $v_{net} = v_f - v_b$; $v_{exch} = \min(v_f, v_b)$. To determine the goodness-of-fit, ^{13}C-MFA fitting results were subjected to a χ^2-statistical test (Antoniewicz et al., 2006).

7.2.6 Quantification of Glucose Secretion

For each strain, experiments were performed independently with [$^2H_7$]glucose (d_7-glucose) and [1,2-^{13}C]glucose. The glucose in the medium was sampled throughout the course of a batch culture, for both total concentration and isotopic labeling. Biomass samples were taken at multiple time points from OD~0.7 to the end of the experiment at glucose depletion (OD~2), and analyzed for glycogen labeling. In all cases the glycogen labeling was found to not change significantly in time. The amount of glucose secretion was quantified by fitting to a differential equation-based model (See Appendix C).

7.2.7 Constraint-Based Reconstruction and Analysis (COBRA) Modeling

Flux balance analysis (FBA), minimization of metabolic adjustment (MOMA), and the relative change (RELATCH) algorithms were implemented through the COBRA Toolbox 2.0 in Matlab 2012b (Kim and Reed, 2012; Schellenberger et al., 2011) using the *E. coli* iAF1260 genome scale model (Feist et al., 2007) (specifically "Ec_iAF1260_flux1.xml"). CPLEX from Tomlab (http://tomopt.com) was used for the LP and QP solvers. RELATCH was downloaded from the Reed Laboratory website (http://reedlab.che.wisc.edu/codes.php). For FBA calculations, the upper bounds for glucose and oxygen uptake rates were set at wild-type values of 8.5 mmol/gdw/hr and 12 mmol/gdw/hr respectively and growth rate was optimized. For MOMA, the reference state was generated using the measured (from 13C-MFA) wild-type fluxes as follows. Forty reactions were mapped from the network model applied here for ^{13}C-MFA to the iAF1260 model, spanning glucose uptake, central carbon metabolic pathways, and acetate secretion. In iAF1260, the upper and lower allowable bounds for each reaction was set to the values obtained for the 95% confidence intervals from ^{13}C-MFA. FBA with growth optimization was then run to obtain the 'reference' state fluxes. Good agreement with measured growth rate (0.60 h-1), acetate yield (0.66 mol/mol) and biomass yield (0.39 g/g) were observed. For RELATCH, the same 40 fluxes and standard errors were inputted, as well as gene expression data previously reported for the wild-type (Covert et al., 2004). All bounds for substrate uptake and central carbon reactions were lifted for the calculation of knockout phenotypes using MOMA and RELATCH.

For the assessment of correlations between measured and estimated normalized intracellular fluxes (Fig. 7.7E-F, Fig. C.3), the following specific fluxes were used (reaction names correspond to iAF1260 designation and numbers correspond to

network model in Table S2): PGI (v_2), PYK (v_9), GND (v_{11}), EDD (v_{19}), PDH (v_{21}), CS (v_{22}), SUCOAS (v_{28}), ICL (v_{32}), PPC (v_{36}), PCK (v_{37}), and EX_ac(e) (v_{70}).

7.3 Results

7.3.1 Detailed Characterization of the Metabolic Phenotype

For each of the 21 studied strains, a detailed description of carbon metabolism was acquired through three complimentary analyses (Fig. 7.1A). First, direct measurements of cell density, substrates, and products during batch culture (aerobic, glucose minimal medium) allowed for the quantification of growth, uptake, and secretion rates. Second, the biomass composition of each strain was quantified using previously described GC-MS based methods (Long and Antoniewicz, 2014b). These measurements provide important constraints in ^{13}C-MFA by quantifying the amounts of each precursor and co-factor utilized in biomass formation (e.g., ribose phosphates for RNA synthesis, or acetyl-CoA for fatty acids). Accuracy of flux estimates in secondary metabolism (i.e., biosynthetic pathways) is being increasingly recognized as important for developing large-scale models (Khodayari and Maranas, 2016; McCloskey et al., 2016). The biomass composition and physiological parameters of these strains were previously reported (Long et al., 2016b), and are updated here (see Materials and Methods). Lastly, ^{13}C tracer experiments were performed with a parallel experimental design, employing [1,2-^{13}C] and [1,6-^{13}C]glucose, previously identified as optimal for flux estimate precision (Crown et al., 2016b). The resulting labeling of proteinogenic amino acids and the sugar moieties of RNA (ribose) and glycogen (glucose) (Long et al., 2016a) were fit to a comprehensive network model of *E. coli*

metabolism, subject to the measured external fluxes and biomass compositions, to obtain the global metabolic phenotype.

The scope of this study includes the knockouts, from the Keio collection (Baba et al., 2006), of twenty enzymes spanning the 'upper' part of central carbon metabolism (Fig. 7.1B), defined here as including glucose kinase (*glk*), phosphoglucomutase (*pgm*, the first step in glycogen biosynthesis), the first half of glycolysis (i.e., EMP pathway), the oxidative (oxPPP) and non-oxidative pentose phosphate pathways (noxPPP), and the ED pathway. The effects of knocking out these genes on the growth rate, biomass yield, glucose uptake rate, and acetate yield are shown in Fig. 7.1C (absolute values in Appendix Fig. C.1, and growth rate reproducibility in Appendix Fig. C.2). The most severe growth defects (>50%) were observed in *Δpgi*, *ΔpfkA*, *ΔtpiA*, and *Δrpe*. Interestingly, *Δpgm* grew 10% faster than the wild-type, likely reflecting the cost the wild-type incurs to sacrifice growth for glycogen storage, hedging against future scarcity or stress. Variations in biomass yield of up to 0.1 g/g from the wild-type (0.4 g_{DW}/g_{gluc}) were observed, with a low of 0.3 g/g (*ΔtpiA*), and a high of 0.5 (*Δpgi*). Corresponding to the growth rates and biomass yields, the glucose uptake rates also varied widely. The acetate secretion phenotype ranged from none produced (*Δpgi*) to nearly 1 mol per mol glucose (*ΔrpiB*), showing both significant increases and decreases from the wild-type's 0.7 mol/mol. These phenotypic differences reflect large metabolic perturbations in these knockouts during batch growth.

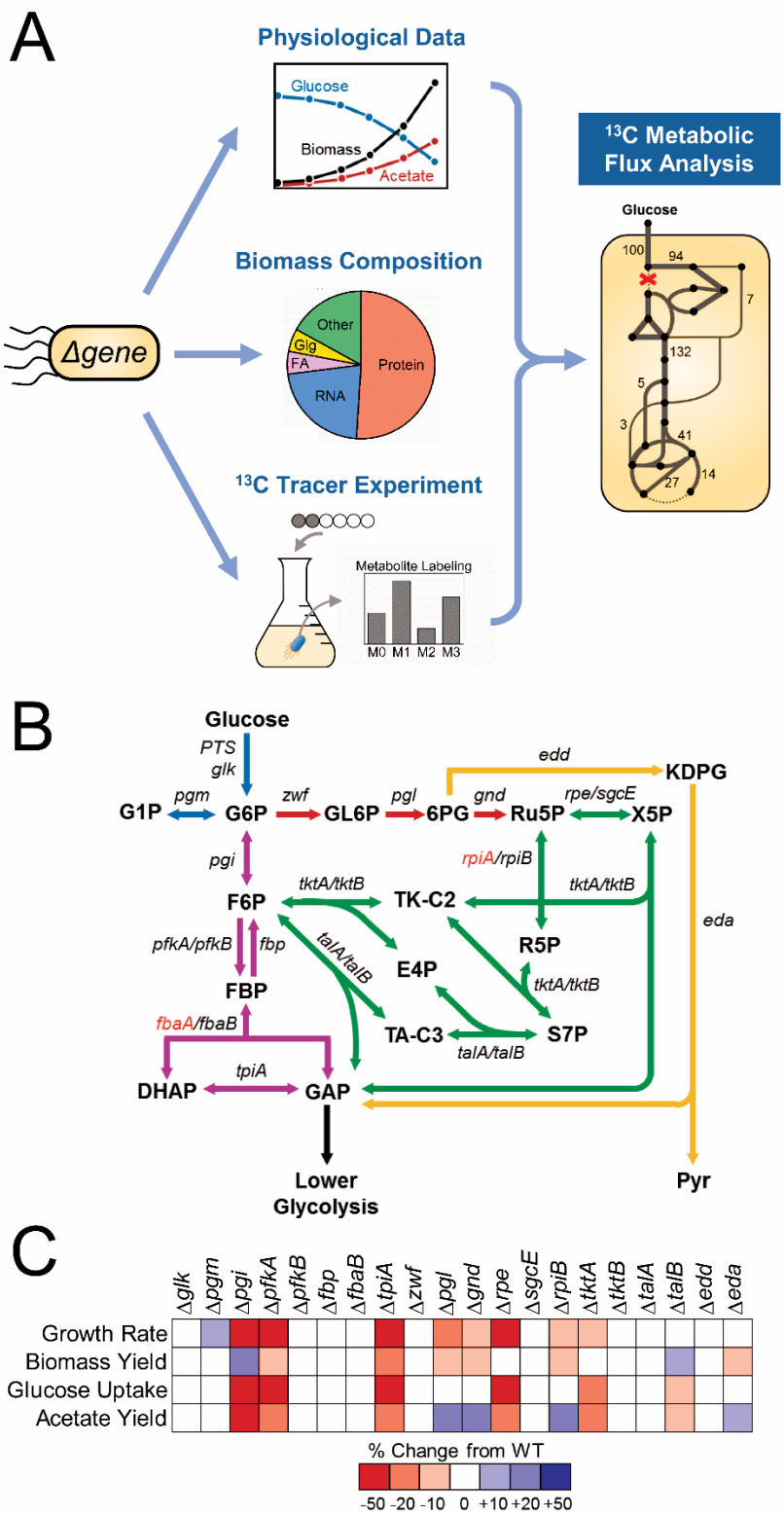

A
Physiological Data
Glucose
Biomass
Acetate
Biomass Composition
Other
Glg
FA
Protein
RNA
Δgene
13C Tracer Experiment
Metabolite Labeling
M0 M1 M2 M3
13C Metabolic Flux Analysis
Glucose
100
94
7
132
5
3
41
27
14
B
Glucose
PTS
glk
G1P
pgm
G6P
zwf
GL6P
pgl
6PG
gnd
Ru5P
rpe/sgcE
X5P
edd
KDPG
pgi
rpiA/rpiB
tktA/tktB
F6P
TK-C2
tktA/tktB
eda
pfkA/pfkB
fbp
talA/talB
R5P
E4P
tktA/tktB
FBP
fbaA/fbaB
TA-C3
S7P
talA/talB
tpiA
DHAP
GAP
Lower Glycolysis
Pyr
C
Δglk
Δpgm
Δpgi
ΔpfkA
ΔpfkB
Δfbp
ΔfbaB
ΔtpiA
Δzwf
Δpgl
Δgnd
Δrpe
ΔsgcE
ΔrpiB
ΔtktA
ΔtktB
ΔtalA
ΔtalB
Δedd
Δeda
Growth Rate
Biomass Yield
Glucose Uptake
Acetate Yield
% Change from WT
-50 -20 -10 0 +10 +20 +50

Figure 7.1: Characterization of 20 E. coli knockouts of upper central carbon metabolism enzymes. (A) The methods applied include 1) Physiological data from aerobic batch growth on glucose minimal medium, 2) The measured biomass composition of each strain, and 3) The metabolite isotopic labeling resulting from parallel ^{13}C isotopic tracer experiments using [1,2-^{13}C]glucose and [1,6-^{13}C]glucose. Together, these measurements are used to precisely estimate fluxes via ^{13}C metabolic flux analysis (^{13}C-MFA). (B) The scope of this study includes twenty single gene knockout strains from the Keio collection spanning the pathways of upper central carbon metabolism. These include *glk* (glucokinase) and *pgm* (to glycogen production) (blue), the upper portion of the glycolysis (EMP) pathway (purple), the oxidative (red) and non-oxidative (green) pentose phosphate pathway, and the ED pathway (orange). The genes denoted in red were excluded because of unavailability (*fbaA* is unconditionally essential and its knockout is not in the Keio collection; *rpiA* was previously determined to be essential in these conditions (20) (C) Physiological changes in knockout strains are summarized, expressed as percentage change from the wild-type. Biomass yields were estimated by ^{13}C-MFA and used to calculate glucose uptake rates.

7.3.2 Identification of Novel Reactions Active in Knockout Strains

Precise intracellular fluxes were determined for each strain by fitting isotopic labeling data and product yields to a previously validated model of *E. coli* metabolism. Statistically acceptable fits were obtained for 19 of the 21 strains (Fig. 7.2A). The two exceptions were *ΔpfkA* and *ΔtpiA*, for which the sum of squared residuals was significantly outside the acceptable range. Such poor model fits in ^{13}C-MFA often indicate an incomplete metabolic network model. Indeed, for *ΔtpiA* the methylglyoxal (MG) pathway has been previously reported to carry significant flux (Fong et al., 2006). Although tightly repressed in the wild-type due to the toxicity of the intermediate methylglyoxal (Tötemeyer et al., 1998), the MG pathway relieves the dead-end at DHAP caused by the knockout, and its addition to the model resulted in an acceptable fit. To resolve the unacceptable fit for *ΔpfkA*, a wide range of alternative

metabolic pathways and reactions was evaluated. The only change to the model that resulted in an acceptable fit of the data was the introduction of an outflux of glucose 6-phoshate (G6P). After validating that *ΔpfkA* did not accumulate additional carbohydrates internally, we hypothesized that *ΔpfkA* may secrete glucose into the medium. This was verified with a targeted tracer experiment (Fig. 7.2B-E), in which fully-deuterated d_7-glucose was introduced as the substrate. Exchange of hydrogens in the PPP and PGI reactions resulted in intracellular G6P labeling that was distinct from extracellular glucose, and was measurable via glycogen. In Fig. 7.2C-D, the extracellular glucose labeling in the batch culture is compared to glycogen labeling. During the course of the tracer experiment, glucose labeling in the medium increasingly resembled the labeling of G6P, clear evidence of glucose secretion by *ΔpfkA*. The glucose secretion rate was determined to be 18% of the uptake rate (Fig. 7.2E; these results were also replicated using a ^{13}C tracer, excluding any ^{2}H-specific kinetic isotope effect). No evidence of glucose secretion was found for wild-type *E. coli* and *ΔtpiA* (Fig 7.2E). These results support and extend recent reports that a *ΔpfkA* strain, modified to prevent glucose consumption, accumulated glucose in the medium when grown on pentoses (Niyas and Eiteman, 2017; Xia et al., 2015). Mechanistic investigations concluded that several phosphatases were involved in dephosphorylating G6P, allowing for the diffusion of free glucose out of the cell (Niyas and Eiteman, 2017).

The intracellular fluxes of wild-type *E. coli*, *ΔpfkA* and *ΔtpiA* are shown in Fig. 7.2F-H. The wild-type flux distribution (Fig. 7.2F) is in good agreement with previous reports (Chen et al., 2011; Crown et al., 2015a; Leighty and Antoniewicz, 2013; Perrenoud and Sauer, 2005), and large flux changes in *ΔpfkA* (Fig. 7.2G) and *ΔtpiA*

(Fig. 7.2H) are apparent (changes greater than 30% from wild-type are highlighted), including the unique reactions identified for these two strains. The PFK I (*pfkA*) and TpiA (*tpiA*) enzymes carry high flux in the wild-type, and their loss results in large reductions in growth rate (Fig. 7.1C) and glucose uptake rate (Fig. 7.1C and Fig. 7.2F-H). In addition to the glucose secretion phenotype, in *ΔpfkA* there is increased relative utilization of the oxPPP, ED pathway, serine deaminase, and the TCA cycle, while the PGI reaction and acetate production fluxes are decreased. The co-occurrence of the unique phosphatase and serine deaminase (no deaminase flux was measured in any other strain) activities may indicate a common regulatory mechanism. The *ΔtpiA* fluxes are quite different, as the MG pathway takes a large flux (86 per 100 glucose) from DHAP to pyruvate, and the TCA cycle, glyoxylate shunt, and PPCK reaction (OAC to PEP) are highly elevated. The oxPPP, lower glycolysis, and acetate production fluxes are decreased. These significant differences demonstrate that there is no common 'slow growth' metabolic phenotype, but that these strains encounter unique limitations and challenges.

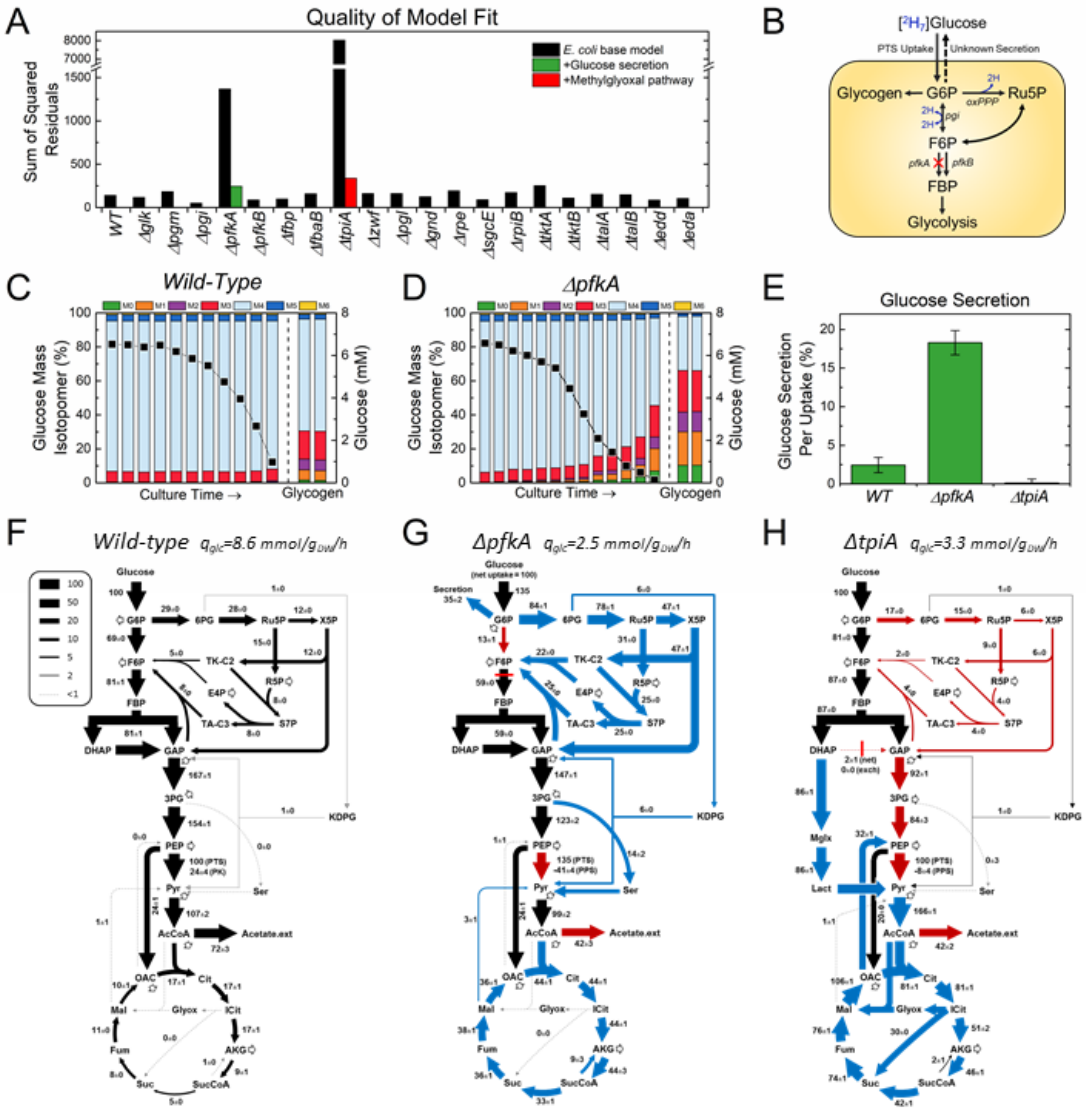

A
Quality of Model Fit
Sum of Squared Residuals
E. coli base model
+Glucose secretion
+Methylglyoxal pathway
WT
Δglk
Δpgm
Δpgi
ΔpfkA
ΔpfkB
Δfbp
ΔfbaB
ΔtpiA
Δzwf
Δpgl
Δgnd
Δrpe
ΔsgcE
ΔrpiB
ΔtktA
ΔtktB
ΔtalA
ΔtalB
Δedd
Δeda
B
[2H7]Glucose
PTS Uptake
Unknown Secretion
Glycogen
G6P
Ru5P
oxPPP
pgi
F6P
pfkA
pfkB
FBP
Glycolysis
C
Wild-Type
Glucose Mass Isotopomer (%)
Glucose (mM)
Culture Time →
Glycogen
D
ΔpfkA
E
Glucose Secretion
Glucose Secretion Per Uptake (%)
F
Wild-type qglc=8.6 mmol/gDW/h
G
ΔpfkA qglc=2.5 mmol/gDW/h
H
ΔtpiA qglc=3.3 mmol/gDW/h
Glucose
G6P
6PG
Ru5P
X5P
F6P
TK-C2
R5P
E4P
FBP
TA-C3
S7P
DHAP
GAP
3PG
KDPG
PEP
Pyr
Ser
AcCoA
Acetate.ext
OAC
Cit
Mal
Glyox
ICit
Fum
AKG
Suc
SucCoA
Mglx
Lact

Figure 7.2: Model validation and glucose secretion determined by ^{13}C-MFA. (A) Acceptable fits (SSR less than ~250) for all but two strains: *ΔpfkA* and *ΔtpiA*. Acceptable fits for these were acquired by adding glucose secretion, and the methylglyoxal pathway, respectively, to the models. (B) To confirm the glucose secretion phenotype, a tracer experiment using [$^{2}H_{7}$]glucose was implemented. Deuterated hydrogens are lost in the oxPPP and PGI reactions, resulting in an intracellular G6P labeling pattern (measured by glycogen) distinct from the extracellular glucose. (C) For the wild-type, extracellular glucose labeling (left, bars) does not change appreciably during batch culture as concentration (squares) diminishes. Glycogen labeling is compared on the right. (D) In *ΔpfkA*, the extracellular glucose labeling changes significantly, and trends toward the intracellular glycogen labeling, reflecting secretion. (E) The secretion was quantified using a differential-equation based model, showing that 18% of glucose taken up is secreted in *ΔpfkA*. (F) Intracellular flux maps of the wild-type, *ΔpfkA* (G), and *ΔtpiA* (H). Fluxes are normalized to 100 units of glucose uptake, and the absolute uptake rates are noted. Red hash marks indicate the reaction affected by knockout, and colors reflect relative flux changes of greater than 30% compared to the wild-type (blue increased, red decreased).

7.3.3 Carbon Metabolism Rewiring in 20 Knockout Strains

The normalized (per 100 glucose) distributions of fluxes through the central carbon metabolic pathways are summarized for all strains in Fig. 7.3A. As might be expected, the strains with severe growth defects (i.e., *Δpgi*, *ΔpfkA*, *ΔtpiA*, *Δrpe*) also had highly altered intracellular flux distributions. However, there were also several strains with large flux changes despite much smaller external physiological effects, particularly *Δzwf*, *Δgnd*, *ΔrpiB*, and *ΔtktA*. This indicates that some compensating pathways are more flexible than others, resulting in a more robust physiology. Other knockout strains had flux distributions close to the wild-type. These include *Δpgm*, as well as knockouts of enzymes that carry minimal flux in the wild-type such as *glk* (most glucose is transported and phosphorylated via PTS), the ED pathway genes *edd* and *eda*, and the gluconeogenic *fbp*. Additionally, minimal changes were observed in

knockouts of minor isozymes *pfkB, fbaB, sgcE, tktB,* and *talA*. The normalized fluxes are also represented in Fig. 7.4, focusing on the pathway utilization through three key branch points in metabolism.

The local effects of enzyme loss on the directly involved reactions can also be evaluated in absolute flux units (mmol/g_{DW}/h, Fig. 7.3B). In the cases where a single gene is responsible for the flux, complete elimination of flux was observed in most cases (e.g., *Δpgm*, *Δpgi*, *ΔtpiA*, *Δzwf,* and *Δgnd*). A notable exception was *Δpgl*, where a significant PGL (6-phosphogluconolactonase) flux was maintained in the knockout strains. This reaction has been shown to proceed spontaneously *in vivo* (Kupor and Fraenkel, 1972; Miclet et al., 2001; Thomason et al., 2004), and a significant oxPPP flux has also been measured in *E. coli* strain BL21, which is lacking the *pgl* gene (Waegeman et al., 2011). In the cases where multiple enzymes are known to catalyze the same reactions (i.e., isozymes), it is apparent in some cases that one isozyme is dominant. For example, the PTS system (glucose uptake) is more utilized than *glk*; *pfkA* is preferred over *pfkB, fbaA* over *fbaB, rpiA* over *rpiB,* rpe over *sgcE,* and *tktA* over *tktB.*

The nature of flux rewiring can also be assessed more globally, for example to determine whether compensatory flux increases are common (i.e., flexible redistribution) or if rate limitations occur (Fig. 7.3C-D). When changes in fluxes from the wild-type are expressed in relative flux units (per 100 glucose), a symmetrical, unimodal distribution of flux changes is observed, suggesting that many fluxes are both increased and decreased in the knockout strains (as in Fig. 7.3A). However, when expressed in absolute units (mmol/g_{DW}/h), the pattern of flux changes is highly skewed towards negative flux changes, i.e., most flux changes result in lower rates in the

knockouts and very few result in higher absolute flux. These changes are expanded upon in Fig. 7.4, which shows for example that the elevated normalized fluxes through the oxPPP and TCA cycle in *Δpgi* and *ΔpfkA* do not represent increases in absolute terms. Instead, as the rates of glucose uptake decrease in these strains, similar absolute rates of flux through these pathways play relatively larger roles. This lack of expansion of absolute flux capacity indicates a rate limitation in these slow growing strains. Limitations can also occur in cofactor metabolism, which is affected directly by carbon pathway usage (Fig. 7.6). For example, it has been shown that reversal of the pyridine nucleotide cofactor transhydrogenase is at least partially rate-limiting in *Δpgi* (Canonaco et al., 2001).

The distributions of absolute fluxes throughout central carbon metabolism across strains is visualized in Fig. 7.3D. The rates of glycolysis varied most widely, with several glycolysis knockouts and *Δrpe* much slower than the wild-type and other strains. There were less drastic changes (in absolute flux units) elsewhere in metabolism, as these fluxes are much smaller in magnitude in the wild-type. Notably, only *Δgnd* had a significantly increased absolute ED pathway flux, and only *ΔtpiA* had significantly elevated absolute TCA, glyoxylate shunt, and PPCK fluxes. In all other strains these pathways were not significantly increased relative to the wild-type, in contrast to previous interpretations of 'latent pathway activation' (for example in *Δpgi*) based on relative flux changes (Fischer and Sauer, 2003; Hua et al., 2003). Notably, the higher variability pathways (e.g., glycolysis) have been shown to be the most highly regulated at the metabolite level, while the pentose phosphate pathway and TCA cycle are comparatively less regulated (Reznik et al., 2017). This may help explain the lack of absolute flux increases in these pathways in most strains, and why

the oxPPP is not able to sufficiently compensate for the decreased glycolysis flux in *Δpgi* and *ΔpfkA*.

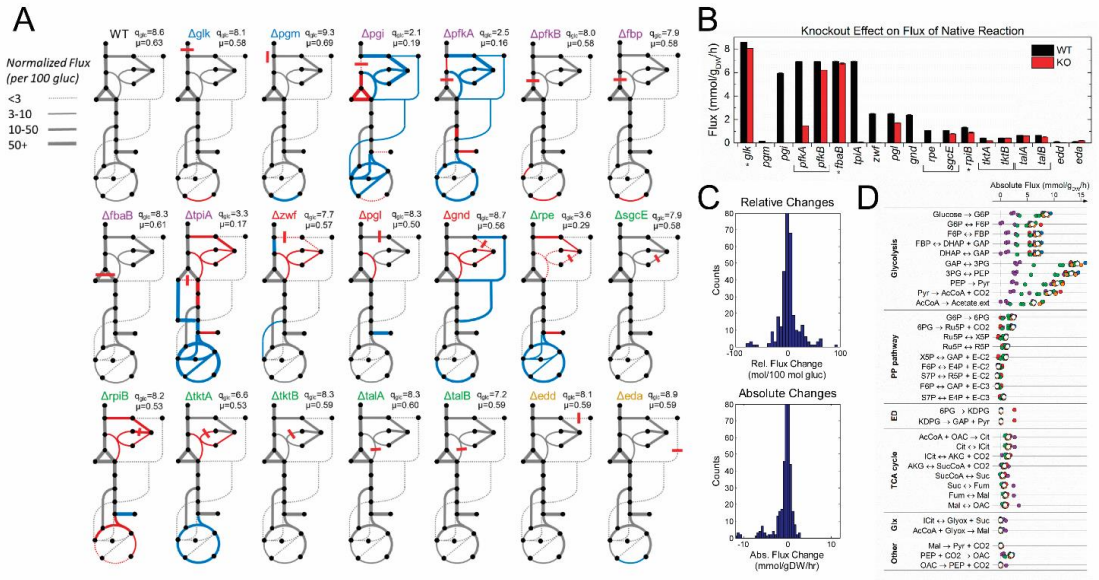


Figure 7.3: Intracellular fluxes in the wild-type and 20 knockout strains. (A) Fluxes for all strains are normalized to 100 units of glucose uptake, with growth and glucose uptake rates noted. The red hash marks denote the reaction affected by the knockout, and the line thickness corresponds to flux value according the legend (upper, left). Colored reactions indicate changes of greater than 30% from the wild-type (blue increased, red decreased). (B) The absolute flux for each reaction directly affected by a knockout, in both the wild-type and corresponding mutant strain. For example, the PGI flux is 6 mmol/g_{DW}/h in the wild-type, but 0 in *Δpgi*. Brackets indicate isozymes corresponding to the same reaction, while asterisks indicate the presence of an isozyme outside the scope of this study. (C) The intracellular flux changes are distributed in a symmetrical way in relative (per 100 glucose) terms, reflecting both flux increases and decreases. In absolute terms the changes are highly skewed to the negative, indicating that most fluxes are maintained or decreased in knockouts, and very few are increased. (D) Absolute fluxes in central carbon metabolism, including the wild-type (white), and mutants in *glk* and *pgm* (blue), glycolysis (purple), oxPPP (red), noxPPP (green), and the ED pathway (orange).

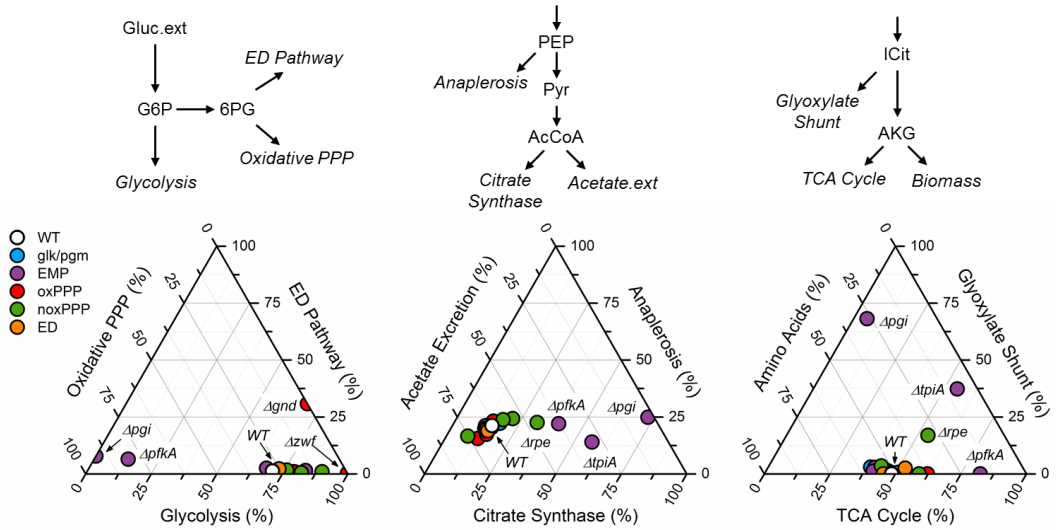


Figure 7.4: Distribution of flux through three key branch points in central carbon metabolism. Three branch points in central carbon metabolism are summarized, each with three possible metabolite fates. These relative fluxes are shown in ternary diagrams. First, the fate of G6P into glycolysis, oxPPP, or ED pathway. The wild-type is ~71% glycolysis, ~28% oxPPP. Glycolytic (EMP) knockouts *Δpgi* and *ΔpfkA*, oxPPP knockout *Δzwf*, and noxPPP knockout *Δgnd* are notably altered. Second, in lower glycolysis PEP can be used for anaplerosis (to OAC) into the TCA cycle and glyoxylate shunt (citrate synthase), or secreted as acetate. EMP knockouts are most distinct in this branch point. Lastly, isocitate can be used for glyoxylate shunt, TCA cycle, or biomass (primarily amino acid biosynthesis) fluxes. *ΔpfkA* has highly elevated TCA cycle usage, and *Δpgi* and *ΔtpiA* have significantly utilized glyoxylate shunt fluxes.

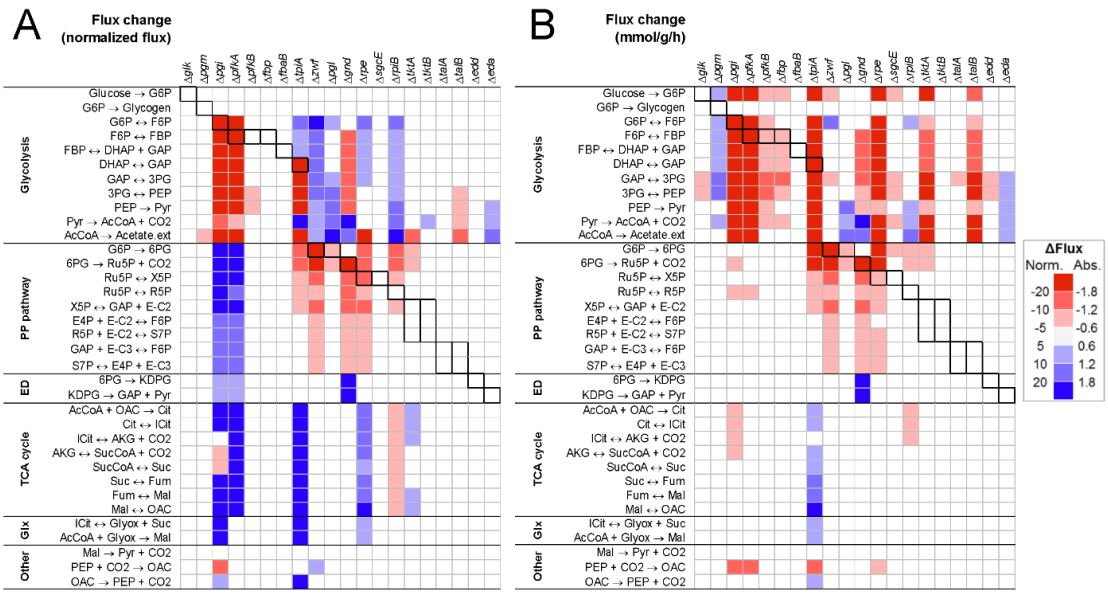


Figure 7.5: Flux changes from the wild-type, in normalized (per 100 glucose) (**A**) and absolute (mmol/g_{DW}/h) (**B**). Colors denote the sign and magnitude of the change, with blue indicating flux increase and red flux decrease, per the key to the right. The flux directly affected by each knockout is bolded. Comparing the two reveals that some increased normalized usage in knockout strains, e.g., pentose phosphate and TCA cycle fluxes in *Δpgi* and *ΔpfkA*, do not reflect increases in absolute fluxes in those pathways. Furthermore, absolute flux increases following knockouts are rare. These cases include glycolysis in *Δpgm*, the TCA cycle, glyoxylate shunt and PPCK fluxes in *ΔtpiA*, and ED pathway in *Δgnd*.

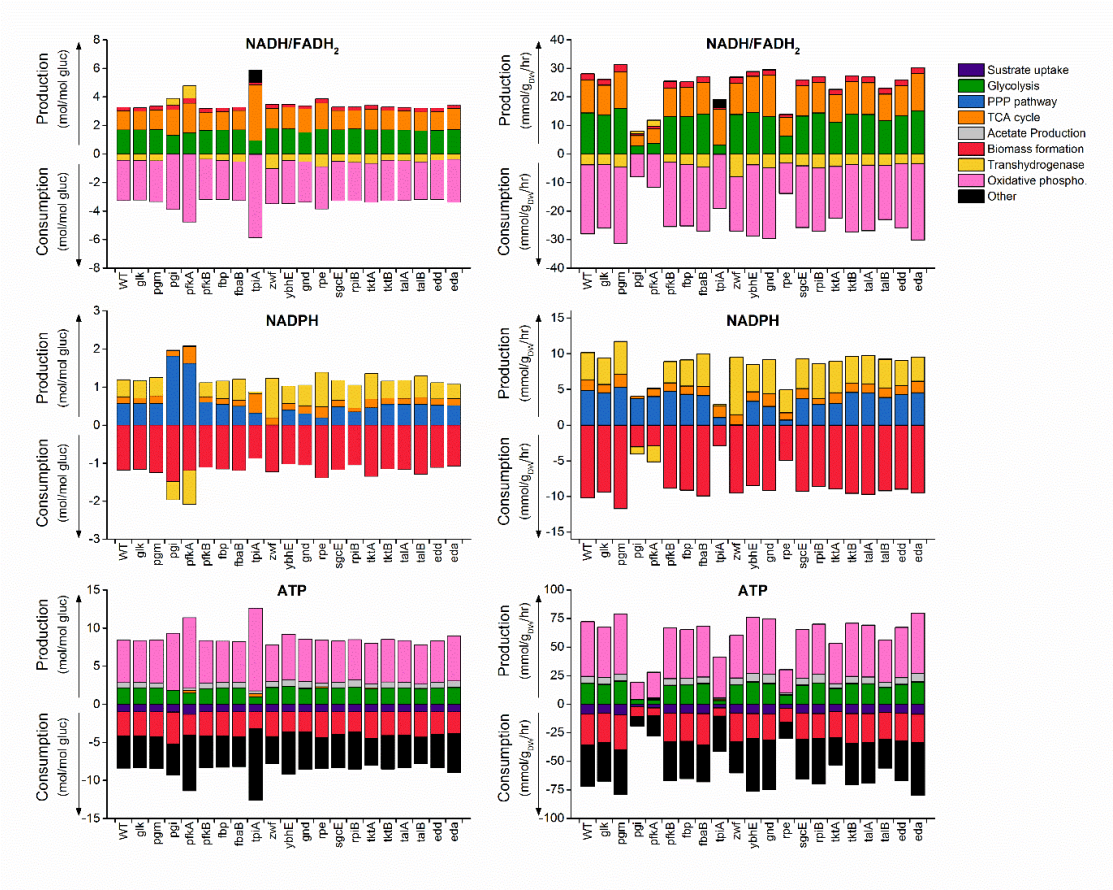


Figure 7.6: Cofactor balances, in normalized (left) and absolute (right) units. The contribution of the major central carbon metabolic pathways (key, top right) to the production and consumption of $NADH/FADH_2$ (lumped), NADPH, and ATP are shown. Bars representing positive values, show cofactor production, and negative values show consumption. In the wild-type and most knockout strains, excess NADH is converted by transhydrogenase to produce NADPH. In *Δpgi* and *ΔpfkA*, elevated oxPPP usage results in a reversal of the transhydrogenase flux, where excess NADPH is converted to NADH. The estimate of ATP generation from oxidative phosphorylation assumes a constant P/O ratio of 2.0 across strains, and the cell maintenance ATP cost ("Other") is calculated as the balance of production and consumption. If no changes to oxidative phosphorylation occur, the calculated elevation in cell maintenance cost in *ΔtpiA* may arise from the toxicity of methylglyoxal.

7.3.4 Multivariate Analysis and Model Prediction Assessment

For additional insights into the key underlying drivers of the observed metabolic rewiring, multivariate statistical analyses were applied. Principal component analysis (PCA) (Fig. 7.7A) shows that the highly dimensional flux data across the 21 strains can be largely explained by a few key principal components (70% of variance explained by first two, 83% by first three). PC1 shows that absolute glycolytic flux and growth rate are coupled, and inversely relate to glyoxylate shunt and PPCK flux rates. PC2 captures an inverse relationship between pentose phosphate pathway and TCA cycle fluxes. The distribution of strains across these two principal components is reflective of the degree of perturbation from the wild-type, and of the highly-perturbed strains there is clustering of knockouts by pathway perturbation (e.g., *Δzwf* and *Δgnd*; *ΔpfkA* and *Δpgi*). Pairwise correlations between fluxes were also examined (Fig. 7.7B-C), with random flux maps generated (Supplementary Text, Appendix C) to control for trivial, stoichiometry-driven relationships (Fig. 7.7B). There were interesting differences between the random and measured flux correlations. The random fluxes showed a highly negative relationship between the fluxes at the branch point of glycolysis and the oxPPP (as well as the ED pathway), whereas no such relationships were observed in the measured flux pattern. This could indicate that this split is regulated, or is kinetically limited (i.e., reduced glycolytic flux did not result in increased absolute oxPPP flux in *Δpgi* or *ΔpfkA*). Additionally, the PPC (PEP to OAC) flux correlated strongly with glycolysis. This flux replenishes the TCA cycle to allow for supply of precursor for biomass synthesis, and its correlation reflects a strong driving force to maintain a relatively constant biomass yield. The strongest non-trivial flux correlation was between the PDH flux (pyruvate to AcCoA) and acetate secretion (ρ=0.99). Above a PDH flux of 2 mmol/g_{DW}/h, the acetate secretion rate increased

linearly by 0.85 mmol/g_{DW}/h per flux unit of PDH reaction. Below a PDH flux of 2 mmol/g_{DW}/h, no acetate secretion was observed. This result provides evidence for the hypothesis of acetate overflow metabolism that is likely caused by limitations in the electron transport chain capacity (Basan et al., 2015; Liu et al., 2014; Majewski and Domach, 1990).

Finally, the measured flux changes were compared to predictions from three commonly used constraint-based reconstruction and analysis (COBRA) models. These models consider the genome scale stoichiometry (here, the *E. coli* iAF1260 genome scale model (Feist et al., 2007)), and represent increasing levels of complexity from FBA to MOMA to RELATCH. Significantly here, only RELATCH can account for the respective contributions of isozymes. The comparison of model predictions to measurements are summarized in Fig. 7.7D-F. FBA and MOMA performed quite poorly at predicting extracellular rates (Fig. 7.7D), often predicting minimal or no phenotypic impact. While RELATCH performed better, it still lacked broad accuracy. For the intracellular flux distributions (normalized), the most severely rewired strains were also the most difficult to predict (Fig. 7.7E). For *Δpgi* and *ΔpfkA*, fluxes were most accurately predicted with RELATCH, however *ΔtpiA* was poorly predicted. For many strains, the accuracy of COBRA predictions was only slightly better than the "assume-no-flux-change" scenario, and these methods were most significantly challenged by the highly-perturbed strains discussed.

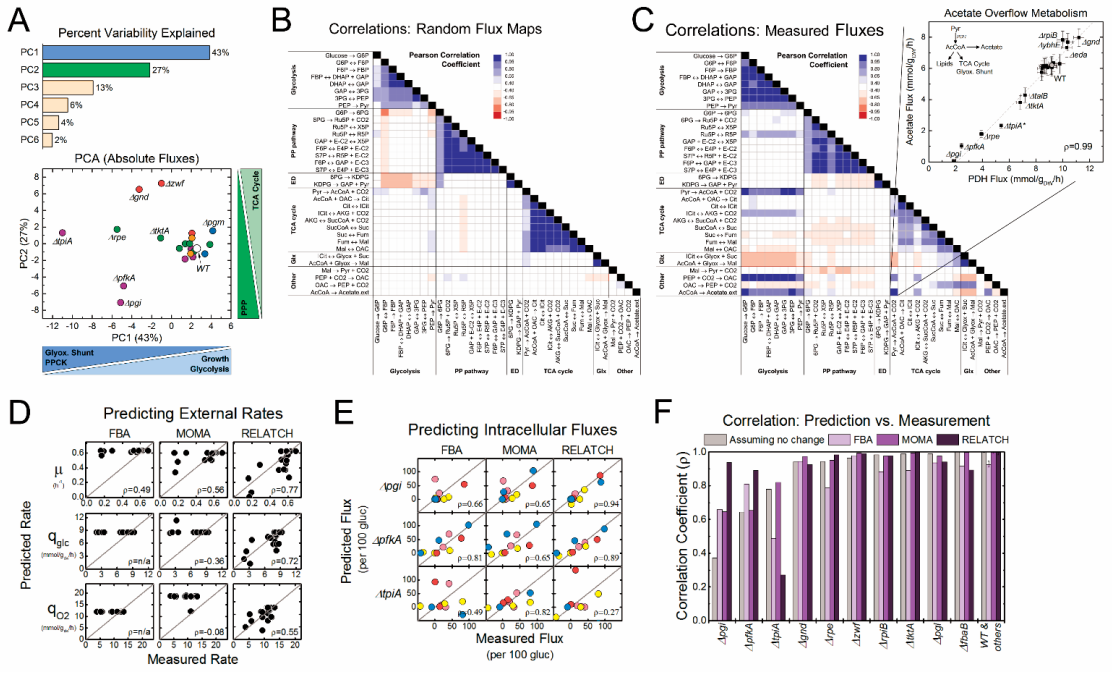


Figure 7.7: Multivariate analysis and model assessment. (A) PCA was performed on all measured absolute fluxes. A large amount (70%) of the variance was captured by the first two principal components, and 93% by the top five. PC1 and PC2 capture pathway fluxes shown by the axis labels. (B-C) Nontrivial correlations between fluxes were identified by comparing those from randomly generated flux maps (B) to those of the measured fluxes (C) (Pearson correlation coefficient represented by color per the legend). The pyruvate dehydrogenase (PDH) and acetate secretion (expanded) fluxes were highly correlated (ρ=0.99), consistent with the acetate overflow model. (D-F) Measured flux values were compared to predictions from common COBRA models FBA, MOMA and RELATCH. (D) External rates of growth, glucose uptake, and oxygen uptake were compared to predictions, with quality of agreement scored by Pearson correlation coefficients. (E) Predictions and measured values of eleven key normalized intracellular fluxes are compared for the three most perturbed strains. The fluxes correspond to intracellular pathways as follows: glycolysis (light blue), oxPPP and ED pathways (red), TCA cycle and glyoxylate shunt (yellow), acetate and ana/cataplerotic reactions (pink). (F) The correlations of these fluxes are shown for all strains. The gray bars reflect the correlation between the knockout and wild-type fluxes (i.e., assuming no change from the wild-type).

7.4 Discussion

In this study, we have leveraged the Keio collection and advances in ^{13}C-MFA to comprehensively interrogate system-level responses to genetic perturbations in the upper pathways of central carbon metabolism. In contrast to previous studies of substrate limited growth (Ishii et al., 2007), the substrate rich conditions presented here revealed and allowed for the quantification of rate limitations in central carbon metabolism. In some cases, remarkable robustness and flexibility was observed, as in the redistribution of oxPPP flux in *Δzwf* and *Δgnd* without significantly reduced glucose uptake. In other cases, however, loss of certain enzymes created massive bottlenecks as flux was redirected (*Δpgi*, *ΔpfkA*), reactions had to be reversed (*Δrpe*), or new pathways had to be activated (*ΔtpiA*, *ΔpfkA*). In cataloguing these various responses, these results will provide a valuable resource for strain engineering and metabolic modeling. For example, the rarity of absolute flux increases in these strains suggests that metabolism is perhaps more kinetically controlled than currently appreciated. As such, the stoichiometry-based COBRA models struggled to predict absolute rates. The development of increasingly sophisticated COBRA models, such as the 'ME' models (Lerman et al., 2012) which account for macromolecular constraints, and ensemble kinetic models (Khodayari et al., 2014; Khodayari and Maranas, 2016), offer promising pathways to better accuracy and predictability. The latter is dependent on high-quality knockout flux data for kinetic parameter identification.

The application of advanced ^{13}C-MFA techniques has again been demonstrated as a useful tool in biological discovery. Here, glucose secretion in *ΔpfkA* was identified and confirmed, building upon previous observations which have already shown potential for biotechnological application (Niyas and Eiteman, 2017; Xia et al.,

2015). Recently, targeted ^{13}C tracer studies confirmed the *in vivo* reversibility of Enzyme I of the PTS system (Christopher P. Long et al., 2017a). These examples highlight that methods designed to improve precision in ^{13}C-flux analysis, e.g., optimized tracer selection and parallel labeling experiments, also provide more resolving power to identify non-conventional pathway usage, even in the extensively studied *E. coli*. Importantly, the directly measured mass isotopomer data can be easily re-analyzed by the community under different network model assumptions or model sizes (García Martín et al., 2015; Gopalakrishnan and Maranas, 2015). Widening the scope of the available flux data to knockouts of other metabolic pathways, and complementing fluxomic data with metabolomic, transcriptomic and/or proteomic data, will strengthen these efforts. Further collection of ^{13}C-MFA results in online public databases, of *E. coli* (Ishii et al., 2007; Long and Antoniewicz, 2014a; Mackie et al., 2014) and other species (Blank et al., 2005; Fischer and Sauer, 2005), would ease access and encourage comparative studies.

Chapter 8

CHARACTERIZATION OF PHYSIOLOGICAL RESPONSES TO 25 GENE KNOCKOUTS IN *ESCHERICHIA COLI* LOWER CENTRAL CARBON METABOLISM

Christopher P. Long, Jacqueline E. Gonzalez, Maciek R. Antoniewicz

8.1 Introduction

Mapping and understanding the genotype-phenotype relationship has fundamental value across systems biology and metabolic engineering practice. In the model organism *Escherichia coli*, the availability of the Keio collection has facilitated large-scale interrogation of gene knockout responses, including growth screens (Baba et al., 2006) and various omics-based investigations into pathways of interest (Haverkorn van Rijsewijk et al., 2011; Ishii et al., 2007). Such studies are critical in the development of systems-level analysis of metabolic networks, including in modeling efforts with direct utility in strain design (Burgard et al., 2003; Khodayari and Maranas, 2016; Kim and Reed, 2010). In a recent review, we highlighted important gaps in the study of *E. coli* knockout metabolism, particularly in central carbon metabolism (Long and Antoniewicz, 2014a). We found that quantitative phenotypic data were not available for numerous gene knockouts, and those that were available from disparate sources often used different conditions and were therefore difficult to compare and assemble for broad analysis.

The pathways of central carbon metabolism under aerobic conditions were identified as the highest priority target for a detailed and comprehensive study of

knockout metabolism. These pathways serve as the essential backbone for substrate assimilation, energy generation, and biomass precursor production, and are necessarily the targets of many engineering interventions. Due to alternative pathways and enzyme redundancy, however, predicting specific knockout responses is far from trivial and has been the subject of extensive study (Fischer and Sauer, 2003; Jahan et al., 2016; Lewis et al., 2012; Long and Antoniewicz, 2014a). Novel reaction activities have even recently been identified in the canonical pathways (Christopher P. Long et al., 2017a; Nakahigashi et al., 2009). Previously, we reported physiological measurements, including growth rates and yields of biomass and products, as well as biomass composition, for the wild-type and 21 knockouts spanning the 'upper' pathways of central carbon metabolism (i.e., glucose transport, glucokinase, the EMP pathway to glyceraldehyde-3-phosphate, the pentose phosphate pathway, and the ED pathway) (Chapter 6) (Long et al., 2016b). Here, we extend this to the 'lower' central carbon metabolic pathways, shown in Fig. 8.1 to include lower glycolysis (downstream of PEP), pyruvate dehydrogenase and acetate production pathways, the TCA cycle, glyoxylate shunt, and amphibolic reactions. These pathways include key metabolic nodes of high interest to metabolic engineering, including pyruvate, acetyl-CoA, citrate, and oxaloacetate (Cordova and Alper, 2016; Krivoruchko et al., 2015). Detailed physiological characterizations are presented, followed by correlation and PCA analysis, as well as assessment of predictions made by constraint-based reconstruction and analysis (COBRA) modeling approaches.

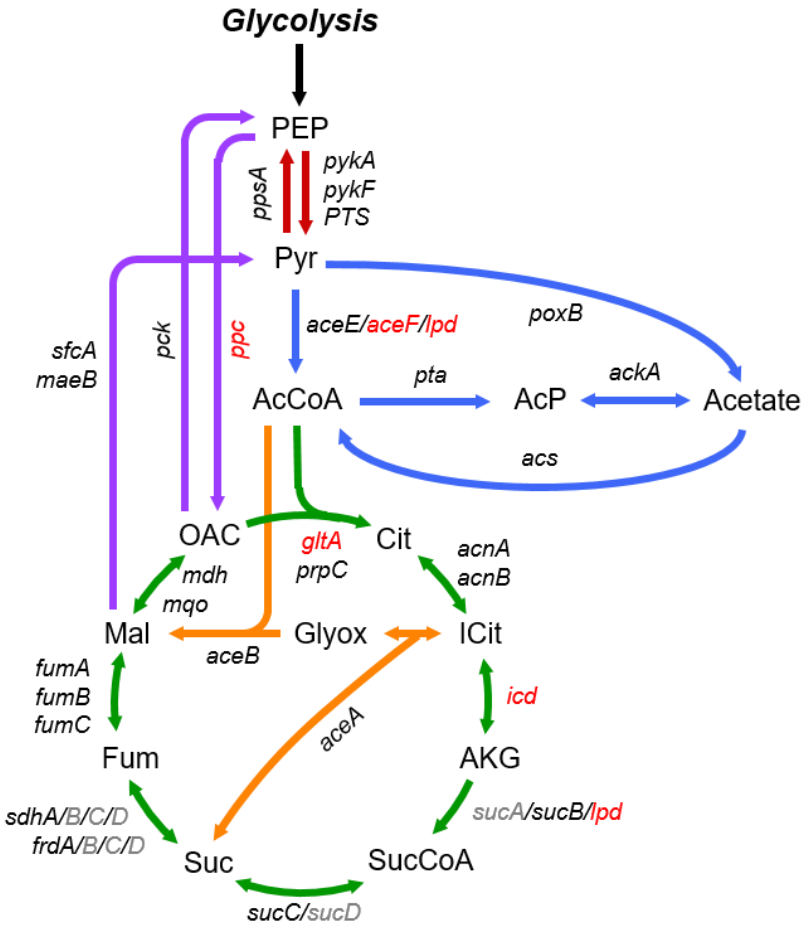


Figure 8.1: The genes included in this knockout study, shown in context of the central carbon metabolic pathways. These include genes in lower glycolysis (red), pyruvate dehydrogenase and acetate metabolism (blue), TCA cycle (green), glyoxylate shunt (orange), and amphibolic reactions (purple). Multiple genes listed for a given reaction indicates isozymes for that reaction, except where slash marks indicate an enzyme complex (e.g. PDH complex denoted *aceE/aceF/lpd*). Genes listed in black are included in the study, those in red did not grow in the studied conditions, and gray were not studied (one knockout per complex was included).

8.2 Methods

8.2.1 Chemicals

All chemicals were purchased from Sigma-Aldrich (St. Louis, MO). [U-^{13}C]Glucose was purchased from Cambridge Isotope Laboratories (Andover, MA). M9 minimal medium was used for all experiments. All media and solutions were sterilized by filtration.

8.2.2 Strains and Culture Conditions

E. coli strains were obtained from the Keio collection (GE Healthcare Dharmacon), which were generated by one-step inactivation of all non-essential genes in *E. coli* K-12 BW25113 (Baba et al., 2006). The strains used in this study are listed in Table 8.1, with the wild-type having been previously described (Long et al., 2016b). The Keio collection contains two independently generated copies of each gene knockout, and the reader should note which was used here as differences resulting from adaptive evolution or contamination of stock cultures are possible.

For assessment of biomass and excreted metabolite yields, each strain was grown in aerated mini-bioreactors with 10 mL working volume (Leighty and Antoniewicz, 2013) in M9 minimal medium with 2 g/L glucose. Cultures were grown until glucose depletion. Biomass yields were calculated by regression of glucose concentrations and optical density (OD_{600}) measurements (Eppendorf BioPhotometer). Supernatant was collected prior to glucose depletion for analysis of excreted metabolites. Acetate was quantified by HPLC (Au et al., 2014), and organic acids in the medium were quantified using isotopically labeled standards (pyruvate, lactate, citrate, succinate, malate, and fumarate) and methoxylamine-TBDMS derivatization (Crown et al., 2016a). For assessment of dry weight and biomass composition, two

biological replicate 100 mL cultures were grown aerobically in shaker flasks, in M9 minimal medium with 2 g/L glucose. Cells were harvested at mid-exponential phase ($OD_{600} \approx 0.7$). At that point, samples for dry weight analysis were taken by filtration of 70 mL of culture using a 0.2 μm cellulose acetate filter (Sartorius 11107-47-N), followed by drying for several days at 80 °C until constant weight. Additionally, samples containing the equivalent of 1 mL of culture at $OD_{600} = 1.0$ (roughly 0.3 mg of dry weight) were washed twice with glucose-free M9 medium and used for biomass composition analysis.

Table 8.1: *E. coli* strains from the Keio collection (GE Healthcare Dharmacon) used in this study.

Pathway	Knockout Gene	Plate-Row-Col	Strain ID
Wild-type	none		
Lower glycolysis	pykA	3-D-3	JW1843-1
	pykF	4-H-2	JW1666-3
	ppsA	4-A-3	JW1692-2
PDH & Acetate	aceE	5-F-9	JW0110-2
	poxB	5-A-5	JW0855-1
	pta	5-H-4	JW2294-1
	ackA	5-E-10	JW2293-1
	acs	49-B-4	JW4030-1
TCA Cycle	prpC	1-C-11	JW0324-1
	acnA	4-B-2	JW1268-2
	acnB	1-B-11	JW0114-3
	sucB	1-F-12	JW0716-1
	sucC	1-G-12	JW0717-1
	frdA	3-D-8	JW4115-1
	sdhA	2-C-12	JW0713-2
	fumA	3-G-2	JW1604-1
	fumB	3-H-7	JW4083-1
	fumC	3-F-2	JW1603-1

	mdh	3-A-6	JW3205-1
	mqo	3-F-3	JW2198-1
Glyoxylate shunt	aceA	3-E-7	JW3975-2
	aceB	3-D-7	JW3974-1
Amphibolic reactions	pck	3-D-6	JW3366-1
	sfcA	77-B-1	JW5238-1
	maeB	3-D-5	JW2447-3

8.2.3 Biomass Composition Analysis

The methods used for quantifying biomass composition were previously described in detail (Long and Antoniewicz, 2014b). Briefly, samples were prepared by three respective methods: hydrolysis of protein and subsequent TBDMS derivatization of amino acids; hydrolysis of RNA and glycogen and subsequent aldonitrile propionate derivatization of sugars (ribose and glucose, respectively); and fatty acid methyl ester derivatization. In total, 17 amino acids were quantified. The amino acids arginine, cysteine and tryptophan are degraded during hydrolysis and were thus not detected. For total protein quantification, we assumed the previously reported ratios of these amino acids relative to alanine (Neidhardt, 1987). Glutamine and asparagine were deaminated to glutamate and aspartate, respectively, during hydrolysis; thus, we report the combined pools of each. Quantification of all species was achieved by isotope ratio analysis using an isotopically labeled standard and a naturally labeled sample. In this study, the standard was generated by growing wild-type *E. coli* on [U-^{13}C]glucose and aliquoting identical (1 mL of an OD_{600} = 1.0) samples of this "fully labeled" biomass. These were centrifuged and washed twice with M9 medium. The composition of the labeled biomass was characterized using unlabeled chemical

standards, and subsequently these were used as standards by co-dissolving with the unlabeled samples at the beginning of each respective analytical method.

8.2.4 Gas Chromatography-Mass Spectrometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 μm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230 °C, the MS quad temperature at 150 °C, the interface temperature at 280 °C, and the inlet temperature at 250 °C. For GC-MS analysis of amino acids, 1 μL was injected at 1:40 split ratio. The column was started at 80 °C for 2 min, increased to 280 °C at 7 °C/ min, and held for 20 min. For GC-MS analysis of fatty acid methyl esters (FAME) and sugar derivatives, 1 μL was injected splitless (Crown et al., 2015b). For GC-MS analysis of organic acids in the medium, 1 μL was injected at 1:10 split ratio. The column was started at 80 °C for 2 min, increased to 280 °C at 10 °C/min, and held for 12 min.

8.2.5 Calculation of Oxygen Uptake and Carbon Dioxide Production Rate

The carbon dioxide production rate was calculated from an overall carbon balance, accounting for glucose uptake, acetate secretion, and biomass formation. The oxygen uptake rate was calculated from an overall redox balance using the degrees of reduction and production/consumption rates of glucose, acetate and biomass. The degree of reduction of biomass of each strain was calculated from the measured biomass composition. The degree of reduction (γ) was calculated as follows (Antoniewicz et al., 2007c; Crown and Antoniewicz, 2013a):

$$\gamma = 4C + H - 2O - 3N + 6S + 5P - charge$$

8.2.6 COBRA Modeling

Flux balance analysis (FBA), minimization of metabolic adjustment (MOMA), and the relative change (RELATCH) algorithms were implemented through the COBRA Toolbox 2.0 in Matlab 2012b (Kim and Reed, 2012; Schellenberger et al., 2011) using the *E. coli* iAF1260 genome scale model (Feist et al., 2007) (specifically "Ec_iAF1260_flux1.xml"). CPLEX from Tomlab (http://tomopt.com) was used for the LP and QP solvers. RELATCH was downloaded from the Reed Laboratory website (http://reedlab.che.wisc.edu/codes.php). For FBA calculations, the upper bounds for glucose and oxygen uptake rates were set at wild-type values of 8.5 mmol/g_{dw}/hr and 12 mmol/g_{dw}/hr respectively and growth rate was optimized. For MOMA, the reference state was generated using measured wild-type fluxes from ^{13}C-MFA (Chapter 7), as follows: Forty reactions were mapped from the network model applied here for ^{13}C-MFA to the iAF1260 model, spanning glucose uptake, central carbon metabolic pathways, and acetate secretion. In iAF1260, the upper and lower allowable bounds for each reaction was set to the values obtained for the 95% confidence intervals from ^{13}C-MFA. FBA with growth optimization was then run to obtain the 'reference' state fluxes. Good agreement with measured growth rate (0.60 h-1), acetate yield (0.66 mol/mol) and biomass yield (0.39 g/g) were observed. For RELATCH, the same 40 fluxes and standard errors were inputted, as well as gene expression data previously reported for the wild-type (Covert et al., 2004). All bounds for substrate uptake and central carbon reactions were lifted for the calculation of knockout phenotypes using MOMA and RELATCH.

8.3 Results

8.3.1 Characterization of Physiological Responses to Gene Knockouts

8.3.1.1 Biomass Dry Weights

The experimentally determined conversion factors of OD_{600} to cell-dry-weight are shown in the upper right of Fig. 8.2. There was minimal variance between strains, with an average conversion of 0.31 g/L/OD_{600} that was consistent with those from the 22 previous measurements reported previously (Long et al., 2016b). The overall average (n=48 strains) of 0.31 g/L/OD_{600} was used for subsequent calculations.

8.3.1.2 Growth Rate

The growth rates of each knockout strain are shown in the upper left panel of Fig. 8.2. Pyruvate kinase I (*pykF*) shows a slight growth defect (0.53 h^{-1}) compared to the wild-type (0.63 h^{-1}) and pyruvate kinase II (*pykA*). The large role of the PTS system in pyruvate to PEP conversion likely contributes to this robustness. More surprisingly, there was also slightly impeded growth in *ΔppsA* (0.50 h^{-1}), which encodes the gluconeogenic PEP synthetase. This is believed to be minimally expressed during aerobic growth on glucose (Trauchessec et al., 2014) as its activity creates a futile cycle. By far the most severe growth defect was observed in *ΔaceE* (0.06 h^{-1}), the E1 component of the pyruvate dehydrogenase (PDH) complex. Interestingly, knockouts of the other components E2 and E3, *aceF* and *lpd*, respectively, did not grow in the studied conditions. A minor defect was observed in *Δpta* (0.51 h^{-1}), the first step in acetate production. In the TCA cycle, the most significant growth defects were in *ΔacnB* (0.21 h^{-1}) and *ΔsucB* (0.23 h^{-1}). The former is the dominant aconitase isozyme, with the presence of *acnA* alleviating the glutamate auxotrophy observed in

the double knockout mutant (Gruer et al., 1997), which also precludes growth in *Δicd*. The latter is a component of the 2-oxoglutarate dehydrogenase multi-enzyme complex (OGDHC) that converts AKG to SuccCoA. No growth reductions were observed in knockouts in the glyoxylate shunt or amphibolic reactions, consistent with the known low flux through these pathways in the wild-type (Leighty and Antoniewicz, 2013). The knockout of the active anaplerotic reaction (PEP to OAC), *Δppc*, did not grow under these conditions.

8.3.1.3 Biomass and Product Yields

The measured biomass yields are also shown in Fig. 8.2, in the center-left panel. Most knockout strains had similar yields to the wild-type's 0.41 g_{DW}/g_{gluc}. The notable exceptions to this were *ΔaceE*, with a very low biomass yield of 0.09 g/g, and *ΔsucB* (0.21 g/g). The highest measured biomass yield was in *Δpta*, at 0.47 g/g. The low biomass yield strains were characterized by 3-carbon products (Fig. 8.2, lower panels), with *ΔaceE* producing high amounts of lactate (0.54 mol/mol) and pyruvate (1.05 mol/mol) and *ΔsucB* producing lactate (0.69 mol/mol). There was a wide spectrum of acetate phenotypes, the primary aerobic product in wild-type *E. coli* (Long et al., 2016b). Most notably, knockouts in PDH (*ΔaceE)* and the dominant *pta-ackA* pathway produced very low amounts of acetate. Interestingly, *Δpta* produced some lactate and pyruvate, while *ΔackA* (the next reaction in the acetate production pathway) did not. Some other mutants in the TCA cycle caused the acetate yield to significantly increase, particularly *ΔacnB* (1.05 mol/mol) and *ΔsdhA* (1.13 mol/mol). Additionally, a small amount of citrate (0.03 mol/mol) were detected from *ΔacnB*, and succinate (0.08 mol/mol) from *ΔsdhA*. In both cases, the excreted product is the substrate of the affected reaction, likely reflecting intracellular accumulation.

A carbon balance, reflecting the measured yields, was performed for each strain and is shown in Fig. 8.3. Remarkably, in *ΔaceE* 79% of the carbon is converted to lactate and pyruvate. The unique biomass and product profiles of *Δpta* and *ΔsucB* are apparent as well. In *ΔackA*, which produces very little acetate or other product, an estimated 40% of all carbon is fully oxidized to CO_2.

8.3.1.4 Substrate Uptake Rates

The uptake rates of glucose and oxygen were calculated and are shown in Fig. 8.4. The lowest glucose uptake rates were observed in *ΔaceE* (3.9 mmol/g_{DW}/h) and *ΔacnB* (3.3 mmol/g_{DW}/h). Corresponding to their low biomass yields, the reductions in glucose uptake rate for *ΔaceE* and *ΔsucB* (uptake rate of 6.0 mmol/g_{DW}/h) were relatively less severe than their reductions in growth rate. A modest decrease was also observed in *Δpta* (6.1 mmol/g_{DW}/h), compared to the wild-type's 8.5 mmol/g_{DW}/h. The oxygen uptake rates were calculated based on electron balances, and revealed significant decreases (from the wild-type's 12.9 mmol/g_{DW}/h) in respiration rates in *ΔaceE*, *Δpta*, *ΔacnB*, and *ΔsucB* (4-6 mmol/g_{DW}/h), as well as less severe reductions in *ΔsucC* and *ΔsdhA* (~8 mmol/g_{DW}/h). Inhibited flux through the TCA cycle may be expected to limit the available reduced cofactors for the electron transport chain. The highest estimated oxygen uptake rate was 17.9 mmol/g_{DW}/h in *ΔackA*, presumably corresponding to an elevated TCA flux and lack of secreted products.

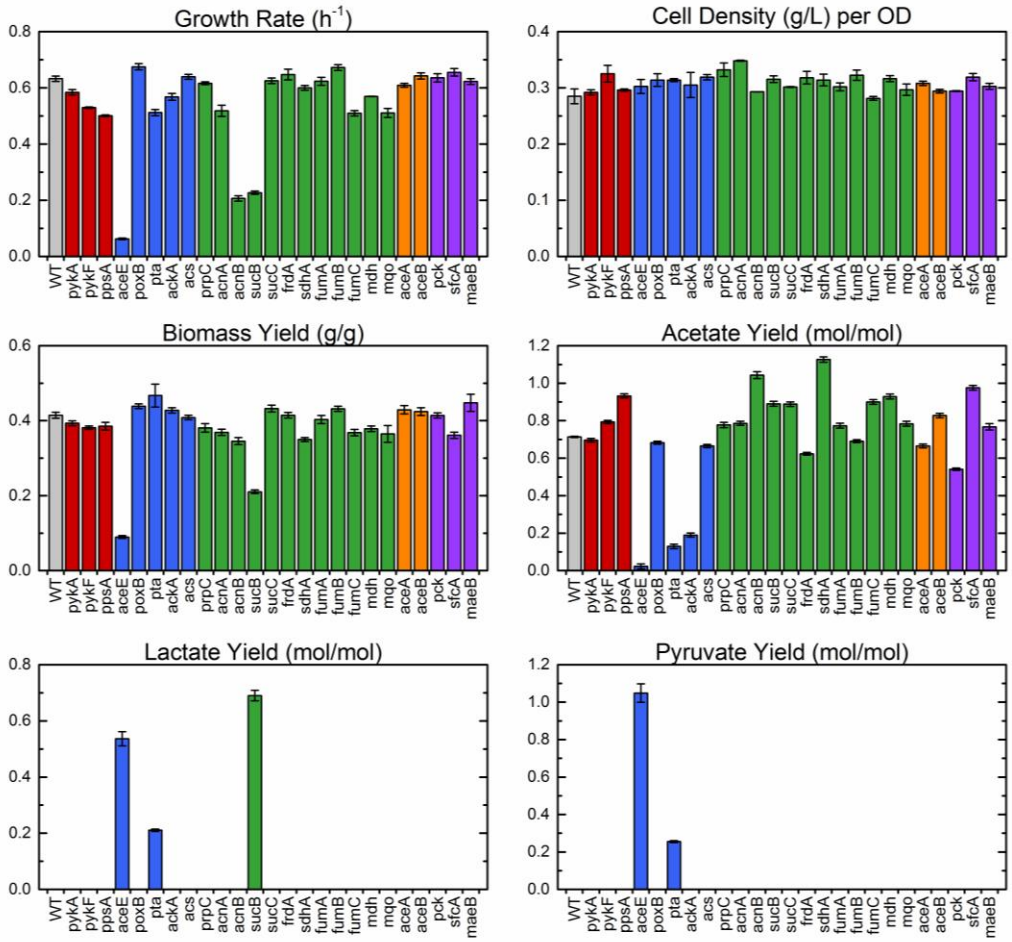


Figure 8.2: Physiological measurements for the wild-type and 25 knockouts. The growth rate, cell density to OD600 conversion, and biomass, acetate, lactate, and pyruvate yields are shown. Colors denote gene pathways: lower glycolysis (red), pyruvate dehydrogenase and acetate metabolism (blue), TCA cycle (green), glyoxylate shunt (orange), and amphibolic reactions (purple). Errors in growth rate and cell density per OD600 reflect standard errors of the mean (n=3, n=2, respectively). Biomass yield errors reflect the standard error from regression of biomass and glucose during culture. The product yield errors reflect standard errors based on the respective assumed errors of measurement (HPLC and GC-MS).

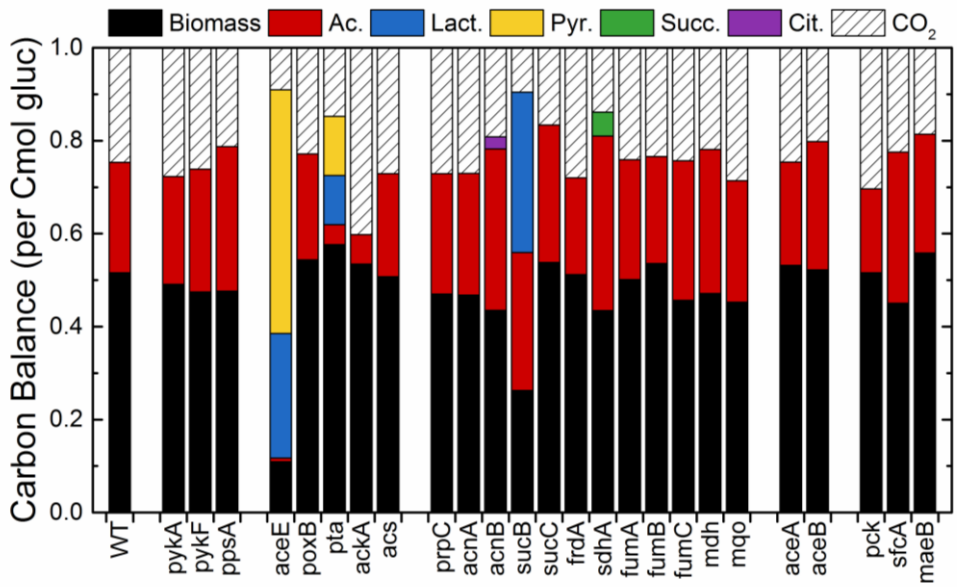


Figure 8.3: Carbon balances for wild-type and knockout strains. The fates of glucose carbon into biomass, products, and CO_2 were calculated on a Cmol per Cmol basis. Strains are grouped according to their metabolic pathways, as in Figure 2.

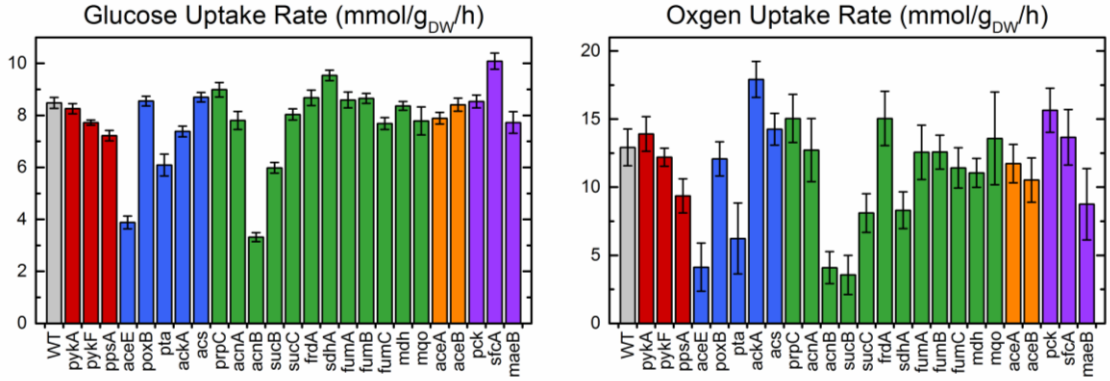


Figure 8.4: Glucose and oxygen uptake rates for the wild-type and knockout strains. Colors denote gene pathways: lower glycolysis (red), pyruvate dehydrogenase and acetate metabolism (blue), TCA cycle (green), glyoxylate shunt (orange), and amphibolic reactions (purple). Error bars reflect standard error bars as propagated from direct measurements through the respective calculations.

8.3.1.5 Biomass Composition

To further elucidate the physiological effects of these knockouts, as well as to facilitate further metabolic network analysis, the biomass composition of each strain was measured. The compositions are summarized in Fig. 8.5. The total protein content varied minimally in most strains from the 51% of cell dry weight seen in the wild-type, consistent with the previous knockout report (Long et al., 2016b). The biggest change was a decrease in *ΔaceE* (39%), and the highest value observed was 56% in *ΔackA*. The RNA content varied more widely, with low amounts (11-14%) observed in *ΔaceE* and *ΔsucB*. Other strains were closer to the wild-type's 21%. Total lipid content ranged from 4% (*ΔsucB*) to 7% (*ΔacnA*), representing both decreases and increases from the wild-type (5%). With respect to glycogen content, most strains were in a range of 2-4%, similar to the wild-type (3%). Several strains had moderately increased glycogen content of ~6%, e.g., *Δacs*, *ΔfrdA*, and *ΔaceB*. By far most notably, however, was *ΔaceE* with a very highly elevated glycogen content of 11% of dry weight. The relative abundances of amino acids were constant across strains, but did vary significantly for fatty acids, particularly C16:0, C16:1, and C18:1.

The compositions were used to calculate the degree of reduction and molecular weights of biomass, which were used here in calculating oxygen and CO_2 exchange rates. The degrees of reduction varied minimally, from 4.36 to 4.46 electrons per Cmol of biomass. The molecular weights varied from 26 to 27 g_{DW} per Cmol of biomass.

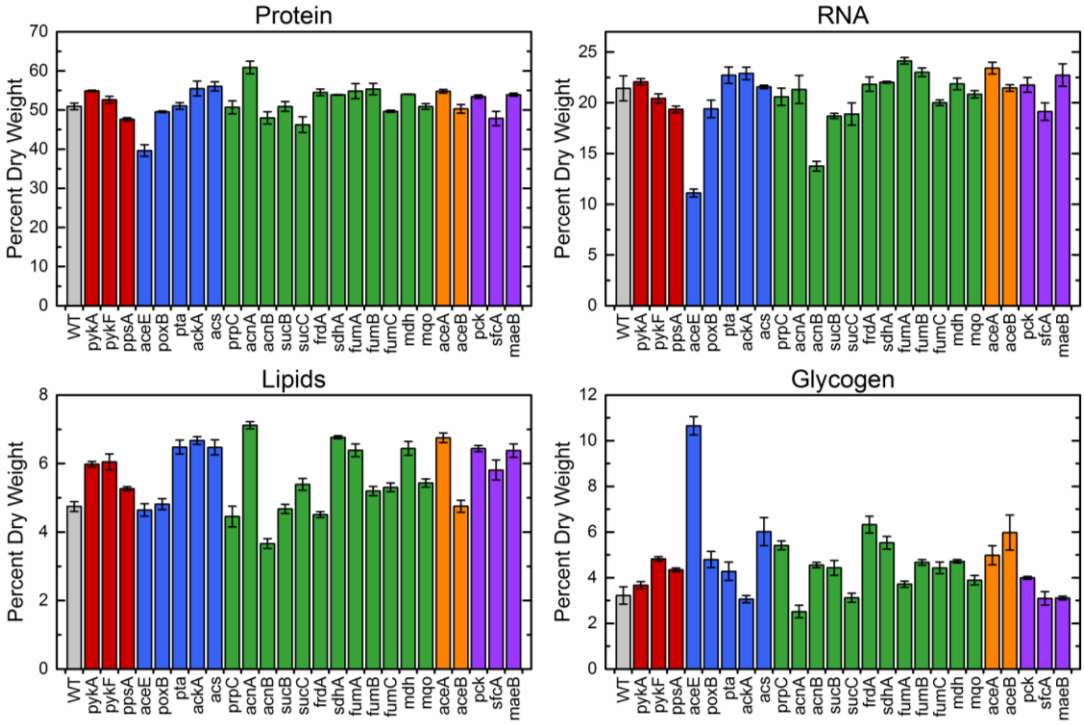


Figure 8.5: Biomass compositions of wild-type and knockout strains. The four most abundant components of *E. coli* biomass are quantified, including protein, RNA, lipids, and glycogen. Colors denote gene pathways: lower glycolysis (red), pyruvate dehydrogenase and acetate metabolism (blue), TCA cycle (green), glyoxylate shunt (orange), and amphibolic reactions (purple). Error bars reflect standard errors of the mean based on 4 measurements (2 biological by 2 technical replicates).

8.3.2 Correlations and Principal Component Analysis of Physiological Data

Pairwise correlation analysis was performed using the physiological data, and a relevant subset are shown in Fig. 8.6. Here they are also compared to 21 previously reported knockouts from upper central carbon metabolism (Long et al., 2016b), which provides interesting context and insight into the effects of perturbing the different central carbon pathways. For example, in the upper central carbon pathway knockouts, higher glucose uptake rate was correlated with reduced biomass yield and increased

product (i.e., acetate) yield. This was consistent with the concept of acetate overflow metabolism. In the lower central carbon metabolic knockouts however, these relationships do not hold. In fact, some of the strains with the lowest glucose uptake rates have the highest product yields and lowest biomass yields (*ΔaceE* and *ΔsucB*). *ΔacnB* is also notable with its very low uptake rate, moderately high product yield, and a biomass yield only slightly lower than the wild-type. As the acetate overflow model is predicated on there being a static capacity for TCA cycle flux, it is perhaps not surprising that perturbations directly to it and product formation pathways would disrupt this phenomenon. There is consistency between the sets of mutant data with respect to the expected trade-off in carbon fates between product and biomass yield (middle-left panel). Here, the addition of *ΔaceE* and *ΔsucB* significantly extend the observed range of this correlation.

The correlations previously reported (Long et al., 2016b) in growth rate and biomass composition were also recapitulated here. The strong positive correlation between growth rate and RNA composition is well-established (Neidhardt, 1987; Pramanik and Keasling, 1997), and is thought to reflect the increased numbers of ribosomes needed for faster growth. The growth rate dependence of fatty acid composition, first reported for the upper central carbon pathway knockouts, is also observed here. At higher growth rates, the C16:0 content decreases and C18:1 content increases. The mechanism of this is still unknown. For each of these relationships, the very low growth rate of *ΔaceE* extends the range of observations beyond what was reported previously.

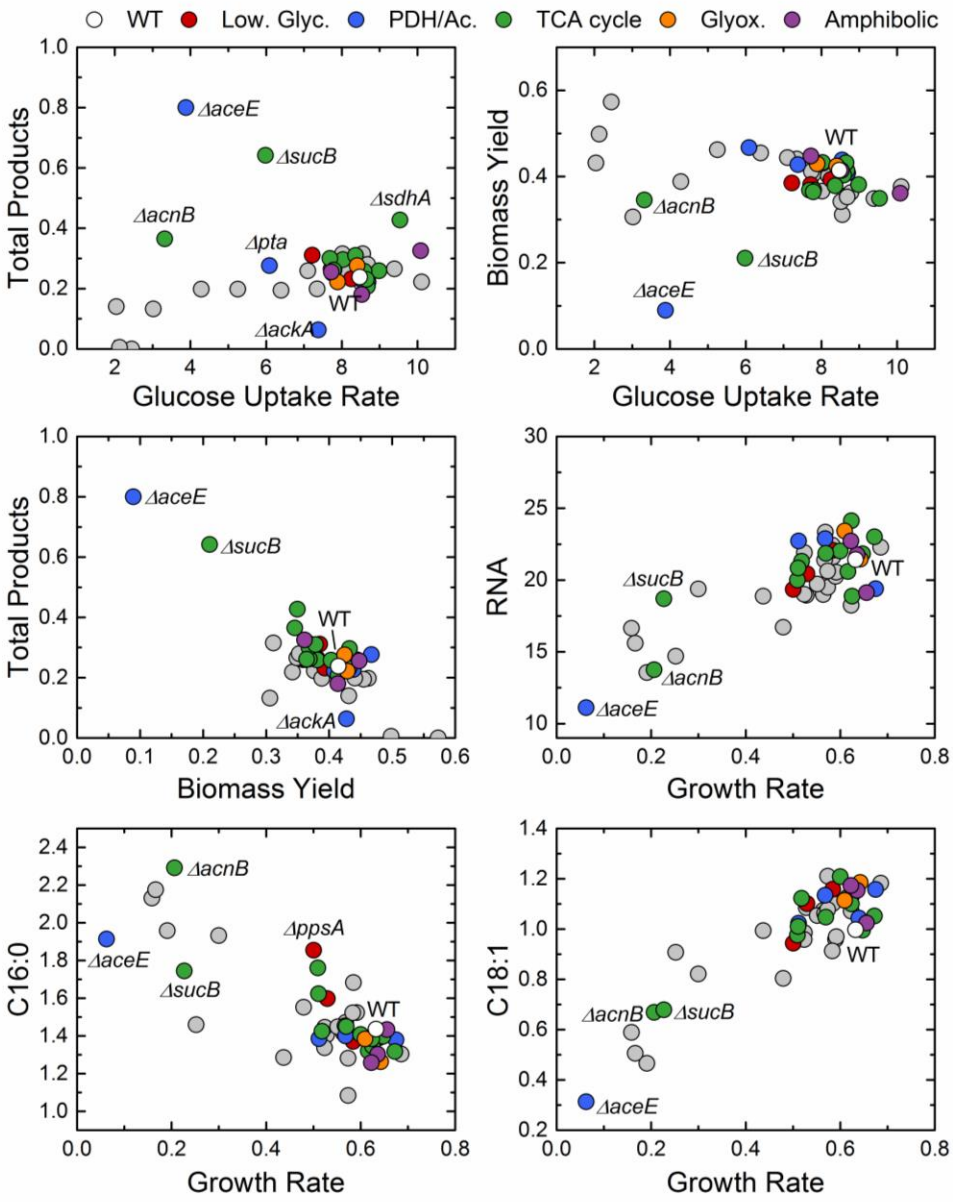


Figure 8.6: Correlations between physiological parameters. For each, previously reported values for 22 knockouts from upper central carbon metabolism are shown in gray, and strains from this study are shown in color, according to the pathway: lower glycolysis (red), pyruvate dehydrogenase and acetate metabolism (blue), TCA cycle (green), glyoxylate shunt (orange), and amphibolic reactions (purple). Six relevant correlations are shown, based on notable differences between upper and lower metabolism knockouts (see text).

Principal component analysis (PCA) was performed to further deconvolute the relationships between physiological parameters and between strains. PCA is a multivariate data reduction technique that minimizes that combines redundant and linearly dependent variables into ‘principal components’, which are able to capture the maximum amount of total variance across the data set. The result of this analysis is shown in Fig. 8.7. The first two principal components explain a large portion of the overall data variance: 42% and 17% respectively. PC1 captures the growth rate and associated biomass composition effects. High PC1 values reflect high growth rate, high RNA and C18:1 content, and low C16:0 content. PC2 captures the carbon fate, of product formation (high PC2 values) and biomass (low values). Again, the upper central carbon pathway knockouts are included in this analysis for further context. Many strains cluster around the wild-type, as expected for knockouts of small or no impact. Interestingly, *ΔacnB* clusters with two slow-growing knockouts in glycolysis, *Δpgi* and *ΔtpiA*. As in Fig. 7, *ΔaceE* and *ΔsucB* are significant outliers compared to the upper central carbon pathway knockouts, due to their very high product yields.

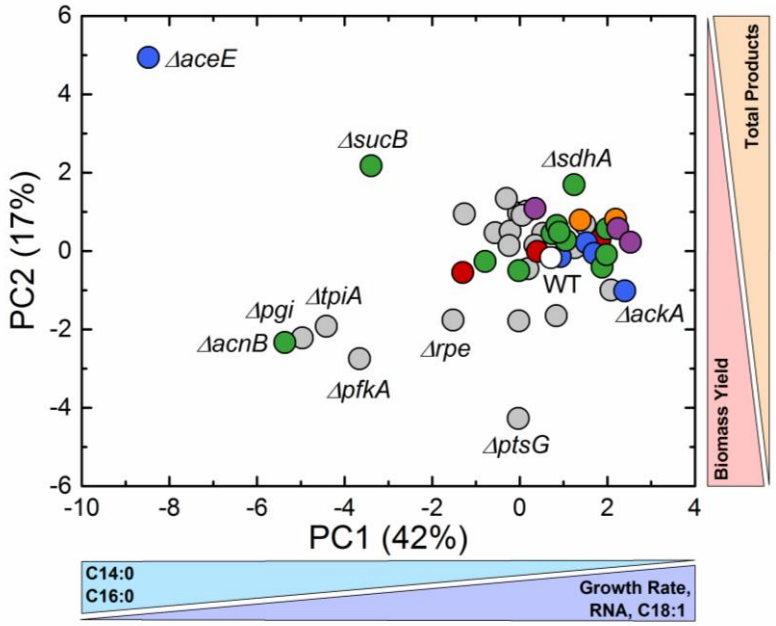


Figure 8.7: Principal component analysis (PCA) of directly measured physiological parameters, including the wild-type (white), 22 knockouts of upper central carbon metabolism (previously reported, gray), and the 25 knockouts in lower central carbon metabolism: lower glycolysis (red), pyruvate dehydrogenase and acetate metabolism (blue), TCA cycle (green), glyoxylate shunt (orange), and amphibolic reactions (purple). PC1 and PC2 capture 59% of the total measurement variability, with PC1 grouping together the growth rate related variance, and PC2 the trade-off of biomass and product yields. The mapping of physiology to PC values is indicated by the triangles on the axes, e.g., the purple triangle shows that high growth rate values correspond to high PC1 values, and vice versa.

8.3.3 Evaluating COBRA Model Predictions

The application of constraint-based reconstruction and analysis (COBRA) modeling approaches to metabolic network analysis and engineering is widespread (Lewis et al., 2012). The acquisition of large phenotypic data sets such as this one presents an opportunity to evaluate the various 'objective functions' used in these approaches. Thus, we compared the measured growth rate, biomass yield, and product

yields to the predictions made by three commonly used techniques: FBA with growth rate optimization (Edwards and Covert, 2002), MOMA (Segre et al., 2002), and RELATCH (Kim and Reed, 2012) (Fig. 8.8). MOMA minimizes the sum of squares difference from the wild-type to the newly constrained knockout network, and RELATCH accounts for wild-type gene expression to further constrain pathway capacities. Details of the implementations of these methods are described above (Methods).

The FBA predictions of growth rates are characterized by very little impact relative to the wild-type, although the no-growth phenotypes are accurately predicted for *ΔgltA* and *Δicd*. MOMA is more accurate in predicting a significant growth rate impact for *ΔaceE*, but the effect is underestimated at 0.2 h^{-1}. RELATCH is actually less accurate for *ΔaceE* (0.4 h^{-1}), but does accurately predict modest decreases in growth rate for *ΔpykF*, *Δpta*, and *ΔackA*. It has a false-negative growth prediction for *ΔacnB*, but does uniquely and accurately predict no growth for *Δppc*. With respect to biomass yield, the only notable predictions for MOMA are *ΔaceE* (accurately low), and for RELATCH again the false-negative for *ΔacnB*. The acetate yields are shown in the bottom row of Fig. 8.8, as the predictions for other products were near zero for all strains. The exceptions were an FBA prediction of pyruvate yield for *ΔaceE* (measured: 1.05 mol/mol, predicted: 0.10 mol/mol) and succinate for *ΔsdhA* (measured: 0.08 mol/mol, MOMA: 0.05 mol/mol, RELATCH: 0.03 mol/mol). The acetate yield predictions for all models were quite poor, with only *ΔaceE* registering differences in FBA and MOMA. RELATCH did predict modest decreases for *Δpta* and *ΔackA*, but these were still quite overestimated relative to the measurements.

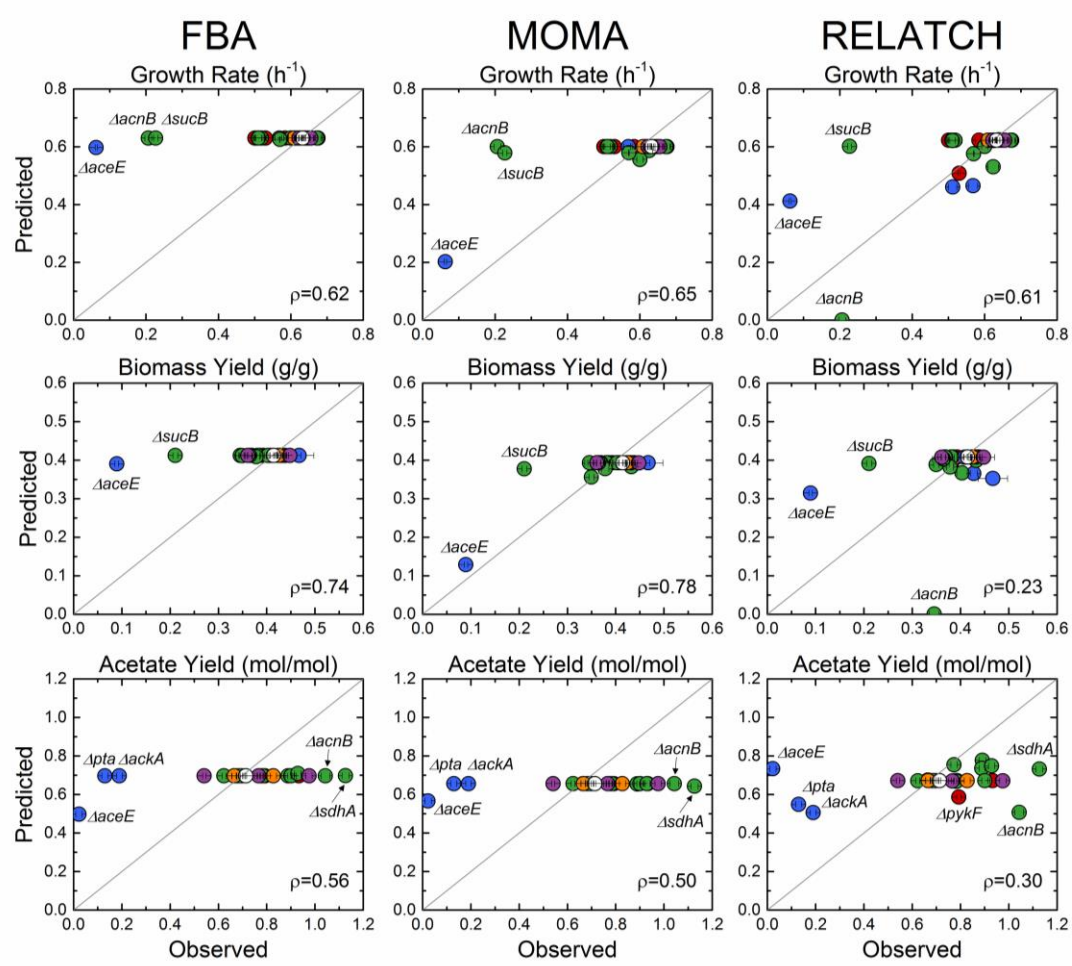


Figure 8.8: Comparison of measured physiological parameters to predictions made by three commonly-used COBRA modeling approaches: FBA, MOMA, and RELATCH. The growth rates, biomass yields, and acetate yields are compared, with model predictions (y-axis) plotted against observed value (x-axis), with the gray (y=x) line representing perfect agreement. The overall quality of the agreements were quantified using Pearson correlation coefficients (ρ, shown).

8.4 Conclusions

In this study, we identify key genes in lower glycolysis, acetate formation, and the TCA cycle that cause significant metabolic bottlenecks and rewiring of

fermentation pathways. The PDH mutant *ΔaceE* was particularly notable here given its severely impeded growth rate and biomass yield, and its 79% carbon conversion to pyruvate-derived products make it an attractive engineering target. While it is unclear why this strain is able to grow while mutants of E2 and E3 (*aceF* and *lpd*) cannot, it should be noted that Nakashima et al. observed a similar high pyruvate production phenotype following RNA silencing of *aceE* (Nakashima et al., 2014). In subsequent reactions which consume the key node metabolite AcCoA, an interesting activation of pyruvate and lactate formation in *Δpta*, not observed in the next step *ΔackA*, could reflect a mechanism to effectively divert glycolytic overflow to the upstream pyruvate node. In the TCA cycle, *ΔacnB* and *ΔsucB* uniquely represent significant growth limiting, but not preventing, perturbations. Continued ^{13}C-MFA studies will confirm whether these and other TCA cycle mutants result in incomplete, truncated cycles.

Taken together with the previous report of wild-type and 21 knockouts in upper central carbon metabolic pathways (Long et al., 2016b), this study completes a comprehensive mapping of physiological effects of knockouts in aerobic central carbon metabolism. This should be a resource to the modeling and engineering communities, which will grow in scope and utility when ^{13}C flux and other omics measurements become available. As constraint-based approaches become increasingly sophisticated, e.g. the ME-models (O'Brien et al., 2013), such data can be re-applied for assessing prediction accuracy. More detailed fluxomic, metabolomic, and/or transcriptomic data will facilitate further development of ensemble kinetic models as well (Chowdhury et al., 2015; Khodayari et al., 2014; Khodayari and Maranas, 2016).

8.5 Author Contributions

CPL performed the experiments and analysis except the analysis of excreted acetate, performed by JEG. CPL and MRA designed the project and wrote the paper with help from all authors.

Chapter 9

FLUXOMIC RESPONSES TO 25 GENE KNOCKOUTS IN LOWER CENTRAL CARBON METABOLISM

Christopher P. Long, Maciek R. Antoniewicz

9.1 Introduction

Here, the methods and analysis introduced in Chapter 7 are extended to the 25 knockout strains in lower central carbon metabolism introduced in Chapter 8. These include knockouts of genes involved in lower glycolysis, pyruvate dehydrogenase and acetate pathways, the TCA cycle, glyoxylate shunt, and amphibolic reactions. The intracellular fluxes of these strains provide important information about the flow of carbon around the key nodes pyruvate and acetyl-CoA, including the usage of the TCA cycle versus product formation. In the knockouts of upper central carbon metabolism (Chapter 7), several strains utilized the glyoxylate shunt and PCK (cataplerotic) reaction. Additionally, a very strong correlation was observed between pyruvate dehydrogenase flux and acetate production, i.e., the acetate overflow effect. Now that genes directly involved in these pathways are perturbed, will the glyoxylate shunt, PCK, or malic enzyme reactions be induced in any cases? Will the acetate overflow effect hold, even when the PDH and acetate pathways are directly affected by knockout? In the upper central carbon metabolic knockouts, significant changes far from the site of knockout, e.g., in the lower pathways, were observed. In these strains, will we see ‘upstream’ impacts of the knockouts, such as redistribution of flux between glycolysis, oxidative pentose phosphate, and Entner-Doudoroff pathways? In

this work, insights into these questions are developed as intracellular fluxes are measured by ^{13}C-MFA, analyzed, and compared to predictions made by COBRA models. The presented results complete the comprehensive metabolic study of knockout responses of 45 central carbon metabolism genes.

9.2 Materials and Methods

9.2.1 Materials

Chemicals and M9 minimal medium were purchased from Sigma-Aldrich (St. Louis, MO). Isotopic tracers were purchased from Cambridge Isotope Laboratories (Tewksbury, MA): [1,6-13C]glucose (99.2 % 13C), and [1,2-13C]glucose (99.7 %). The isotopic purity and enrichment of all tracers were validated by GC-MS analysis (Cordova and Antoniewicz, 2016; Sandberg et al., 2016). All solutions were sterilized by filtration.

9.2.2 Strains and Culture Conditions

E. coli strains were obtained from the Keio collection (GE Healthcare Dharmacon), which were generated by one-step inactivation of all non-essential genes in *E. coli* K-12 BW25113 (Baba et al., 2006). The specific strains used in this study are listed in Table 8.1. For ^{13}C-tracer experiments, strains were cultured aerobically in glucose M9 minimal medium at 37°C in mini-bioreactors with 10 mL working volume (Long et al., 2016b). Pre-cultures were grown overnight and then used to inoculate the experimental culture at an OD_{600} of 0.01, in which 2 g/L of the glucose tracers were present. Cells were harvested for GC-MS analysis at mid-exponential growth when OD_{600} was approximately 0.7.

9.2.3 Analytical Methods

Cell growth was monitored by measuring the optical density at 600nm (OD_{600}) using a spectrophotometer (Eppendorf BioPhotometer). The OD_{600} values were converted to cell dry weight concentrations using a previously determined OD_{600}-dry cell weight relationship for *E. coli* (1.0 OD_{600} = 0.32 g_{DW}/L) (Long et al., 2016b). After centrifugation, the supernatant was separated from the cell pellet. The cell pellets were washed with glucose-free M9 medium prior to subsequent analysis.

9.2.4 Gas Chromatography-Mass Spectrometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 µm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230°C, the MS quad temperature at 150°C, the interface temperature at 280°C, and the inlet temperature at 250°C. GC-MS analysis of *tert*-butyldimethylsilyl (TBDMS) derivatized proteinogenic amino acids was performed as described (Long and Antoniewicz, 2014b). Labeling of glucose (derived from glycogen) and ribose (from RNA) were determined as described (Long et al., 2016a; McConnell and Antoniewicz, 2016). In all cases, mass isotopomer distributions were obtained by integration (Antoniewicz et al., 2007b) and corrected for natural isotope abundances (Fernandez et al., 1996).

9.2.5 Metabolic network model and ^{13}C-metabolic flux analysis

The metabolic network model used for ^{13}C-MFA is provided in Appendix Table C.1, and the full flux results are in the Supplemental File. The model (Crown et al., 2015a; Gonzalez et al., 2017) includes all major metabolic pathways of central

carbon metabolism, lumped amino acid biosynthesis reactions, and a lumped biomass formation reaction. The coefficients for precursors in the biomass formation reaction were determined for each strain, based on the measured biomass compositions, as previously described (Antoniewicz et al., 2007c). A modification was made to the model for *ΔaceE*, where it was found that the reversible reaction between aspartate and fumarate (E.C. 4.3.1.1) significantly improved the quality of fit. ^{13}C-MFA calculations were performed using the Metran software (Yoo et al., 2004), which is based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007a). Fluxes were estimated by minimizing the variance-weighted sum of squared residuals (SSR) between the measured and model predicted mass isotopomer distributions and acetate yield using non-linear least-squares regression. For integrated analysis of parallel labeling experiments, the data sets were fitted simultaneously to a single flux model (Leighty and Antoniewicz, 2013). Flux estimation was repeated 10 times starting with random initial values for all fluxes to find a global solution. At convergence, accurate 95% confidence intervals were computed for all estimated fluxes by evaluating the sensitivity of the minimized SSR to flux variations. Precision of estimated fluxes was determined as follows (Antoniewicz et al., 2006):

$$\text{Flux precision (stdev)} = [(\text{flux}_{\text{upper bound 95\%}}) - (\text{flux}_{\text{lower bound 95\%}})] / 4$$

To describe fractional labeling of biomass amino acids G-value parameters were included in ^{13}C-MFA. As described previously (Antoniewicz et al., 2007c), the G-value represents the fraction of a metabolite pool that is produced during the labeling experiment, while 1-G represents the fraction that is naturally labeled (e.g., from inoculum). By default, one G-value parameter was included for each measured amino acid in each data set. Reversible reactions were modeled as separate forward

and backward fluxes. Net and exchange fluxes were determined as follows: $v_{net} = v_f - v_b$; $v_{exch} = \min(v_f, v_b)$. To determine the goodness-of-fit, ^{13}C-MFA fitting results were subjected to a χ^2-statistical test (Antoniewicz et al., 2006).

9.2.6 Constraint-Based Reconstruction and Analysis (COBRA) Modeling

Flux balance analysis (FBA), minimization of metabolic adjustment (MOMA), and the relative change (RELATCH) algorithms were implemented through the COBRA Toolbox 2.0 in Matlab 2012b (Kim and Reed, 2012; Schellenberger et al., 2011) using the *E. coli* iAF1260 genome scale model (Feist et al., 2007) (specifically "Ec_iAF1260_flux1.xml"). CPLEX from Tomlab (http://tomopt.com) was used for the LP and QP solvers. RELATCH was downloaded from the Reed Laboratory website (http://reedlab.che.wisc.edu/codes.php). For FBA calculations, the upper bounds for glucose and oxygen uptake rates were set at wild-type values of 8.5 mmol/gdw/hr and 12 mmol/gdw/hr respectively and growth rate was optimized. For MOMA, the reference state was generated using the measured (from ^{13}C-MFA) wild-type fluxes as follows. Forty reactions were mapped from the network model applied here for ^{13}C-MFA to the iAF1260 model, spanning glucose uptake, central carbon metabolic pathways, and acetate secretion. In iAF1260, the upper and lower allowable bounds for each reaction was set to the values obtained for the 95% confidence intervals from ^{13}C-MFA. FBA with growth optimization was then run to obtain the 'reference' state fluxes. Good agreement with measured growth rate (0.60 h-1), acetate yield (0.66 mol/mol) and biomass yield (0.39 g/g) were observed. For RELATCH, the same 40 fluxes and standard errors were inputted, as well as gene expression data previously reported for the wild-type (Covert et al., 2004). All bounds for substrate

uptake and central carbon reactions were lifted for the calculation of knockout phenotypes using MOMA and RELATCH.

For the assessment of correlations between measured and estimated normalized intracellular fluxes (Fig. 9.9B-C), the following specific fluxes were used (reaction names correspond to iAF1260 designation and numbers correspond to network model in Table S2): PGI (v_2), PYK (v_8), GND (v_{10}), EDD (v_{18}), PDH (v_{20}), CS (v_{21}), SUCOAS (v_{27}), ICL (v_{31}), PPC (v_{35}), PCK (v_{36}), and EX_ac(e) (v_{74}), EX_lac_D(e) (v_{75}), and EX_pyr(e) (v_{76}).

9.3 Results and Discussion

9.3.1 Physiology

As presented in Chapter 8, the scope of this study was 25 knockout strains spanning the pathways of lower central carbon metabolism (Fig. 9.1A). The basic physiological characteristics, including growth rate, biomass yield, glucose uptake rate, and acetate yield, are expressed as a percentage change from the wild-type in Fig. 9.1B (complete data in Appendix Fig. D.1-2). The most severely growth-limited strains were *ΔaceE*, *ΔacnB*, and *ΔsucB*. In addition to the changes highlighted here, some strains also had significant production of pyruvate (*ΔaceE* and *Δpta*) and lactate (*ΔaceE*, *Δpta*, and *ΔsucB*).

9.3.2 Intracellular flux distributions

Simplified flux maps, representing the normalized intracellular flux distributions, are shown in Fig. 9.2. The large changes in pathway usage are highlighted (blue for increases, red for decreases greater than 30%). Growth rates and glucose uptake rates are also noted. Notably, the relative usages of the upper pathways, i.e., glycolysis,

pentose phosphate pathway, and ED pathway, do not undergo large rewiring in these strains. This may reflect a regulatory mechanism that maintains the splits in these branch points across a wide range of glucose uptake rates. The knockouts have significant impacts across the pathways they are more proximal to, in product formation and the TCA cycle. The three strains *ΔaceE*, *Δpta*, and *ΔsucB* begin producing pyruvate derived (pyruvate and lactate) products, while the knockouts in the *pta-ackA* acetate pathway cause significant reductions in acetate excretion and increased usage of the TCA cycle. Several knockouts restrict flux in the TCA cycle, including *ΔaceE*, *ΔacnB*, *ΔsucB*, *ΔsucC*, and *Δmdh*. These include some of the most severely perturbed overall phenotypes. Interestingly, despite direct perturbations in the TCA cycle and amphibolic reactions, none of the knockouts utilized the glyoxylate shunt or PCK reactions. Only *Δmdh* had an active malic enzyme flux.

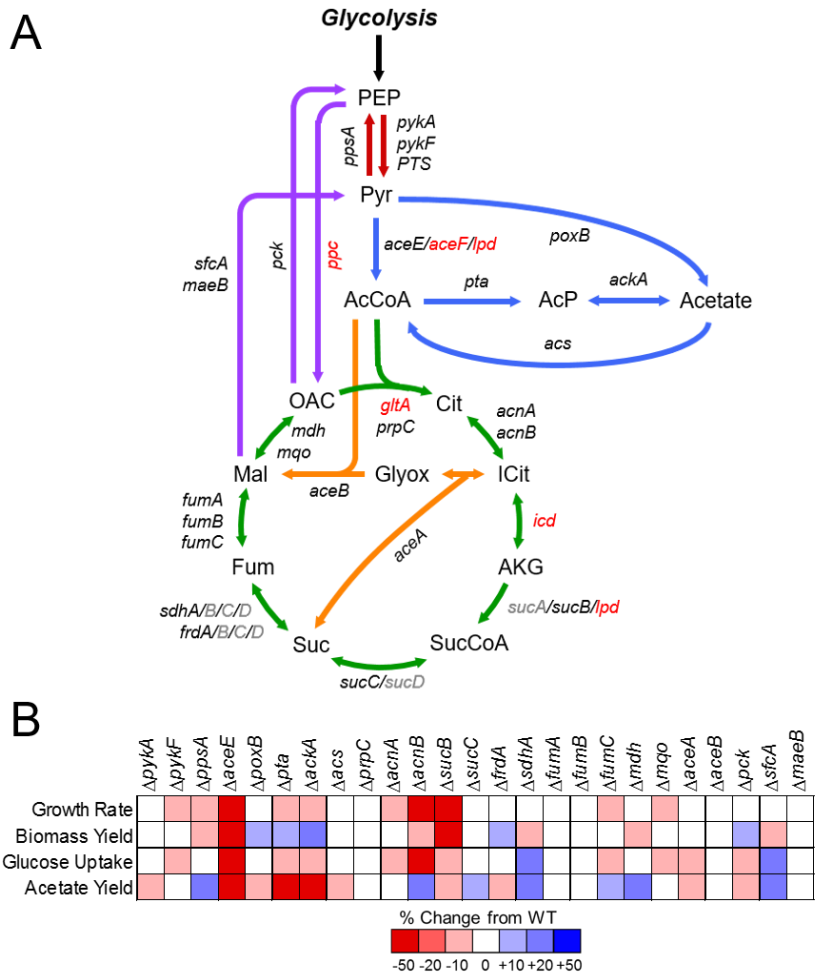


Figure 9.1 (A) The genes included in this knockout study, shown in context of the central carbon metabolic pathways. These include genes in lower glycolysis (red), pyruvate dehydrogenase and acetate metabolism (blue), TCA cycle (green), glyoxylate shunt (orange), and amphibolic reactions (purple). Multiple genes listed for a given reaction indicates isozymes for that reaction, except where slash marks indicate an enzyme complex (e.g. PDH complex denoted *aceE/aceF/lpd*). Genes listed in black are included in the study, those in red did not grow in the studied conditions, and gray were not studied (one knockout per complex was included). (B) Physiological changes in knockout strains are summarized, expressed as percentage change from the wild-type. Biomass yields were estimated by ^{13}C-MFA and used to calculate glucose uptake rates.

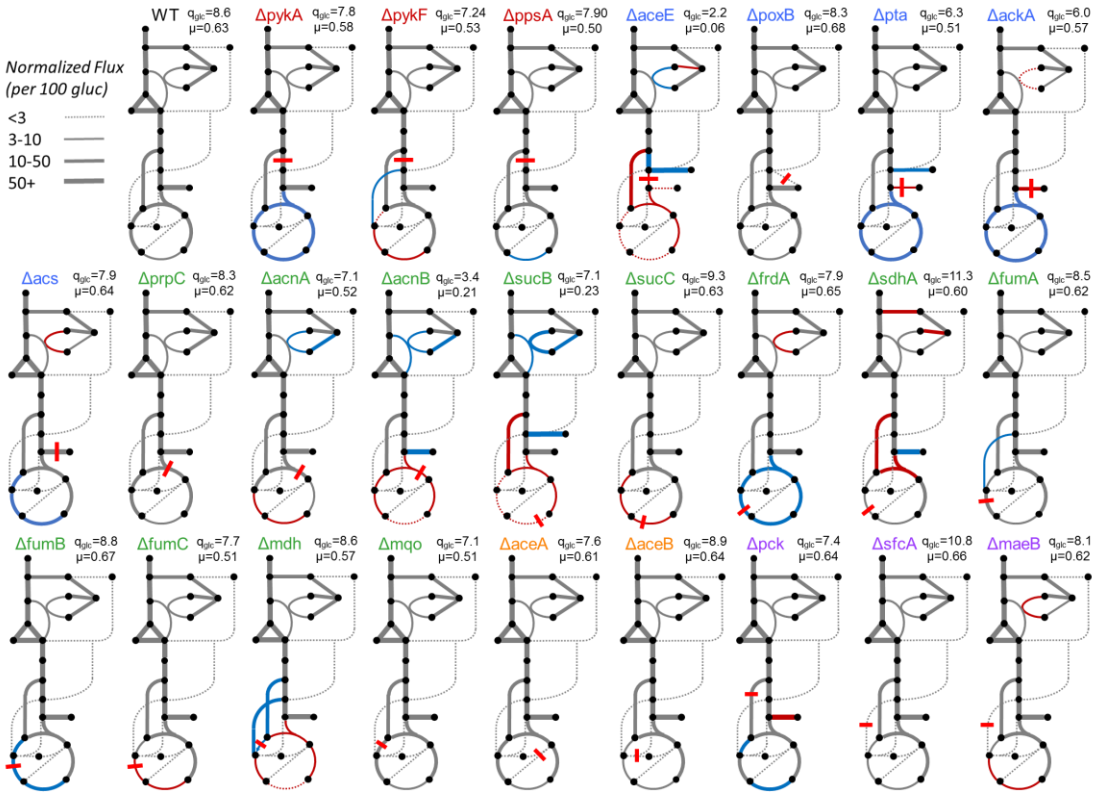


Figure 9.2: Intracellular flux distributions for the wild-type and 25 knockouts. Glucose uptake rate and growth rate are noted for each strain, and the location of the knockout is shown by the red hash mark. The magnitude of the normalized intracellular fluxes are represented by the line thickness, as noted in the key to the upper left, and changes greater than 30% are reflected by the line color (blue for increase, red for decrease).

Three interesting knockouts affecting the key pyruvate and acetyl-CoA nodes are *ΔaceE*, *Δpta*, and *ΔackA*, the fluxes of which are shown in detail in Fig. 9.3. The E1 subunit of the pyruvate dehydrogenase (PDH) complex is knocked out in *ΔaceE*, which severely restricts this flux (Pyr→AcCoA). Due to identical atom transitions, it is not possible to distinguish flux through the PDH mechanism or others such as pyruvate formate-lyase (EC 2.3.1.54). The result of the knockout was a massive out-flux from pyruvate, excreted as pyruvate (101 per 100 mol gluc) and lactate (50).

Here, no acetate was produced and the TCA cycle is incomplete, only generating the minimum necessary precursors, e.g. AKG, for biomass production. Additionally, the large pathway changes resulted in exactly zero cofactor transhydrogenase flux, i.e., NADH and NADPH were generated at exactly the needed proportion. The *pta-ackA* pathway is the primary route of acetate production in aerobically growing *E. coli*, and has been shown to be largely thermodynamically controlled (Enjalbert et al., 2017; Wolfe, 2005). The knockout of *pta*, the first step in this pathway, significantly reduced acetate outflux, and interestingly resulted in excretion of pyruvate products (22 lactate and 17 pyruvate per 100 gluc). The impact of *ackA* knockoutwas quite different; although acetate production was similarly knocked down, no pyruvate-derived products were produced and instead the TCA cycle flux was significantly increased. This resulted in a larger transhydrogenase flux producing NADPH. The differences between these two strains are interesting with respect to the regulation of acetate overflow metabolism. One potential explanation for these differences is the role of the intermediate acetyl-phosphate in protein regulation, where it is well known to have global control of lysine acetylation (Weinert et al., 2013).

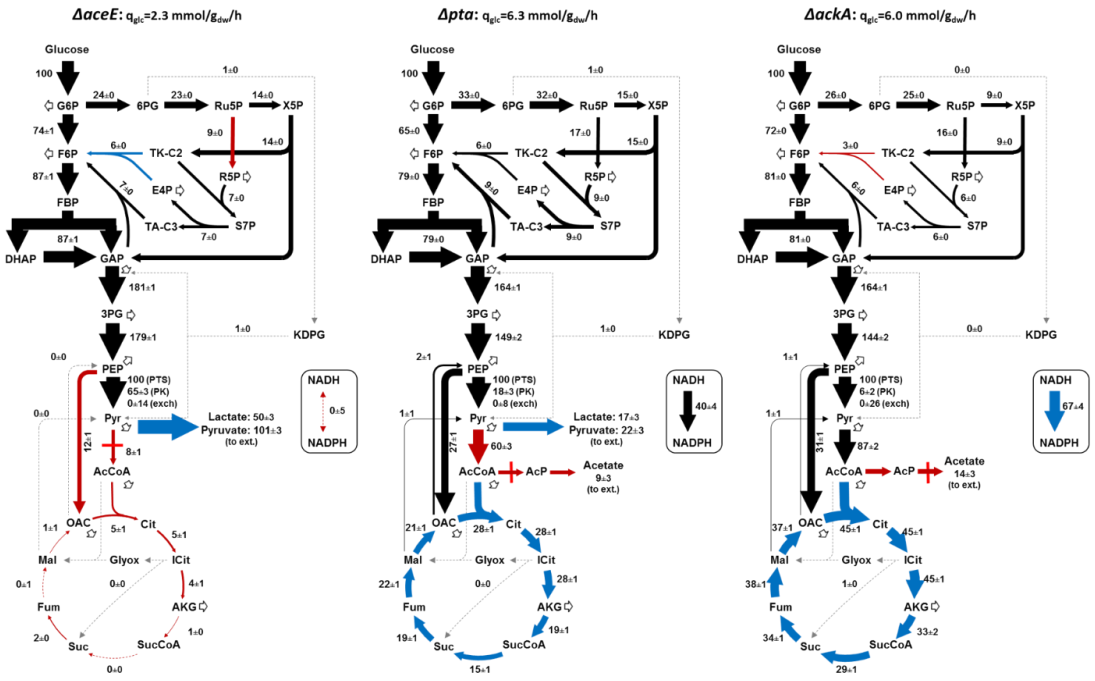


Figure 9.3: Detailed intracellular flux maps for three strains in the PDH and acetate production pathways. Glucose uptake rates are noted, and the knockouts are represented by red hash mark. The line thickness reflects the magnitude of the normalized intracellular flux (per 100 glucose uptake). Changes in normalized flux of greater than 30% are reflected by color (blue for increase, red for decrease).

In the TCA cycle, *ΔacnB*, *ΔsucB*, and *Δmdh* had the most severely perturbed fluxes (detailed flux maps shown in Fig. 9.4). The knockout of *acnB*, the dominant aconitate hydratase, caused a large reduction in glucose uptake rate (3.4 mmol/g_{DW}/h) and an incomplete TCA cycle providing only biomass precursors. The isozyme *acnA* is presumably responsible for the aconitase flux of 9 per 100 glucose in this strain. As the TCA cycle is limited, acetate overflow is increased. In *ΔsucB*, the flux through the 2-oxoglutarate dehydrogenase is blocked, resulting in an incomplete TCA cycle. The limited flux into the TCA cycle provides AKG for biosynthesis (*icd*, which produces

AKG, is essential in these conditions). *ΔsucB* maintains fast glucose uptake (7.1 mmol/g_{DW}/h), but has a much lower growth rate (0.23 h^{-1}) and biomass yield (0.18 g_{DW}/g_{gluc}). Somewhat surprisingly, a significant portion of this missing carbon goes into lactate production (81 per 100 gluc). In this strain, the transhydrogenase flux is reversed (forming NADH). Deletion of the major malate dehydrogenase *mdh* results in a unique rewiring of the TCA cycle and malic enzyme. The TCA cycle is incomplete, and flux is routed from PEP to OAC (PPC reaction) to Mal to Pyr (via malic enzyme). Malate dehydrogenase is known to be reversible, but the malate:quinone oxidoreductase *mqo* is thought to be irreversible, so the mechanism of this oxidative (OAC→Mal) flux is unclear and worthy of future investigation.

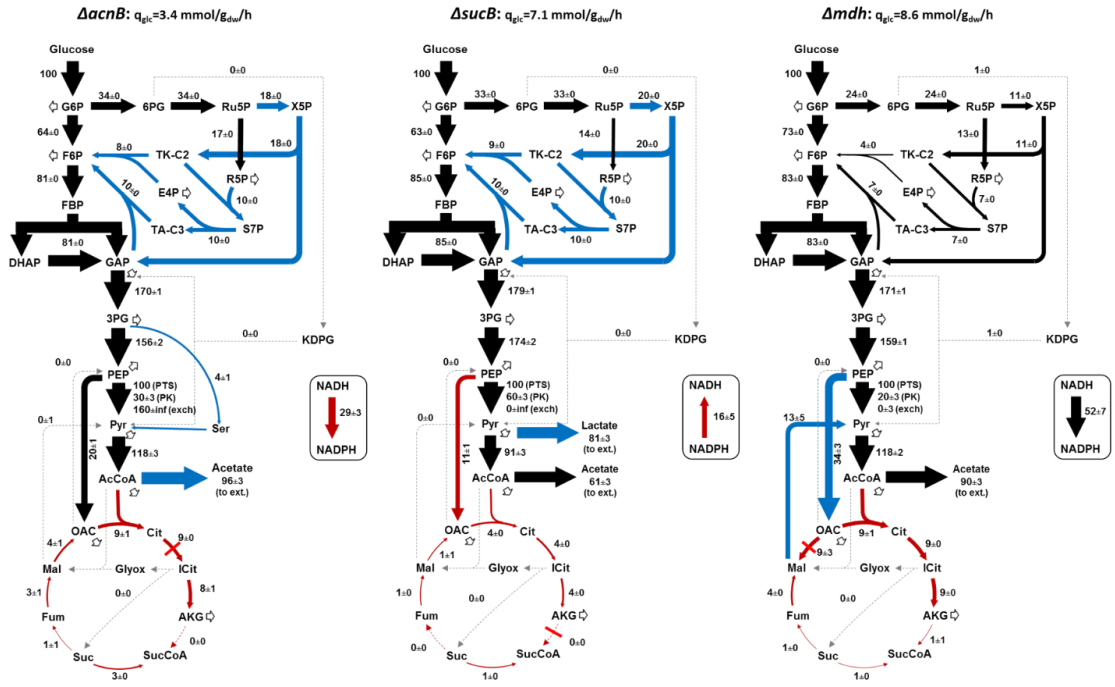


Figure 9.4: Detailed flux maps of three knockouts in the TCA cycle. Line thickness and color have the same meaning as in Fig. 9.3.

9.3.3 Metabolic flux changes

The nature of flux rewiring can also be analyzed more holistically, as in Figs. 9.5 and 9.6. In Fig. 9.5A, the change in directly affected flux (i.e., in the reaction associated with each knocked-out gene) for each mutant is shown. For example, of the genes involved in pyruvate and PEP interconversion, *ΔpykF* causes a slight decrease in absolute flux, while knock out of *ppsA* (gluconeogenic reaction) causes a slight increase in flux in the glycolytic direction. *ΔaceE, Δpta*, and *ΔackA* cause severe, near-complete flux decreases, while the alternative acetate production pathway *poxB* and acetate consumption pathway *acs* do not affect the acetate flux. As noted above, *ΔacnB* has a more significant impact than *ΔacnA*, while *ΔsucB* and *ΔsucC* eliminate flux as expected. Except for the reversal of the malate dehydrogenase in *Δmdh*, the rest of the genes either have robust isozymes or carry very little flux in the wild-type.

The global network flux changes are visualized in Fig. 9.5B-C and Fig. 9.6. In absolute units, by far the most variance is observed in glycolysis and product formation pathways (Fig. 9.5B). Histograms of flux changes, both in relative and absolute terms, are shown in Fig. 9.5C. While most changes are near zero, in both relative and absolute terms fluxes are seen to increase and decrease. This is in contrast with the upper metabolism knockouts (Ch. 7), where absolute flux increases were rare.

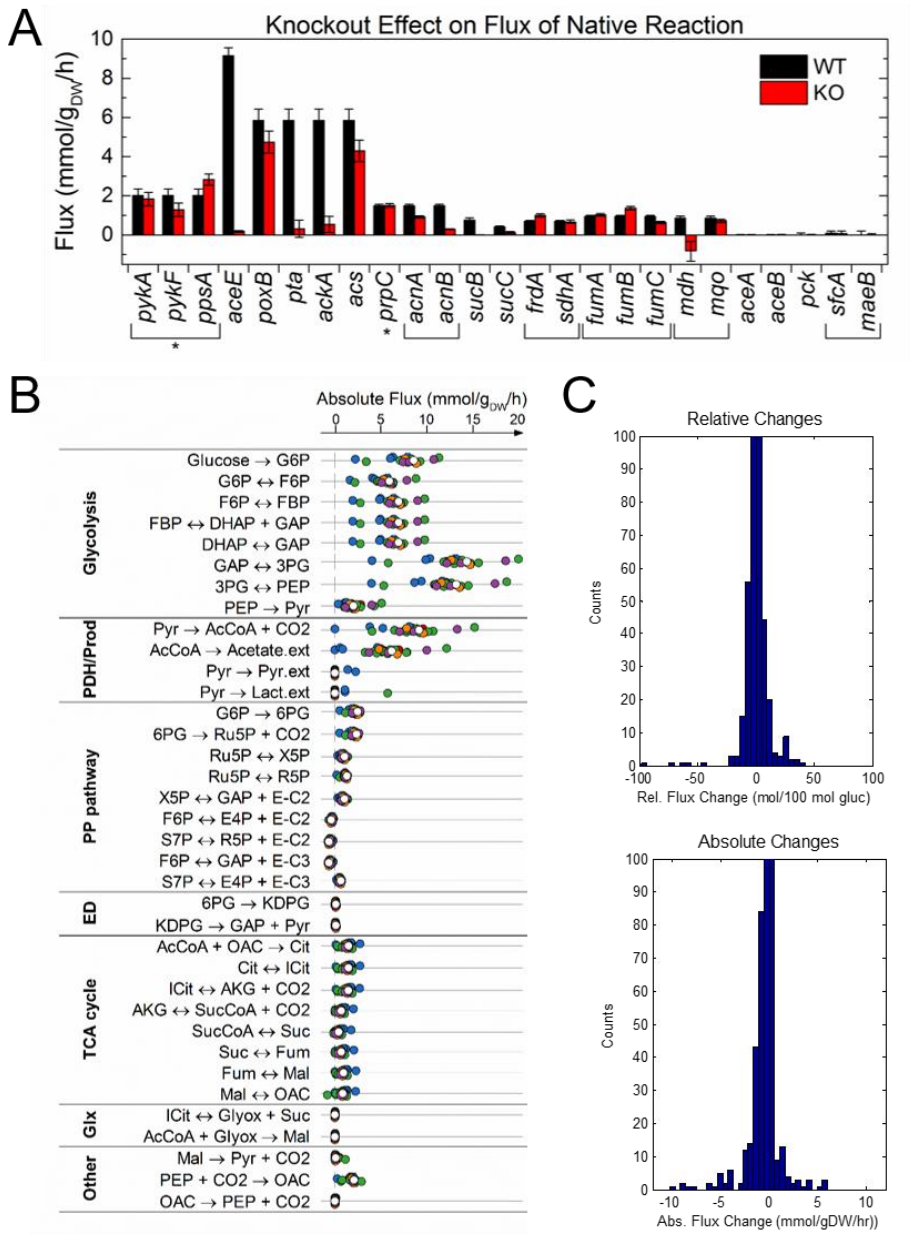


Figure 9.5: (A) The change in absolute flux in the reaction directly affected by the knockout. Asterisks reflect the presence of an isozyme not included in the study, and brackets group isozymes together. (B) Absolute fluxes of all strains, throughout central carbon metabolism. Each point is a single strain, with the color reflecting the pathway as in Fig. 9.1 (wild-type is white). (C) Histograms of relative and absolute flux changes from the wild-type in the knockout strains. Values below zero represent reductions in flux, whereas those greater than zero represent increases.

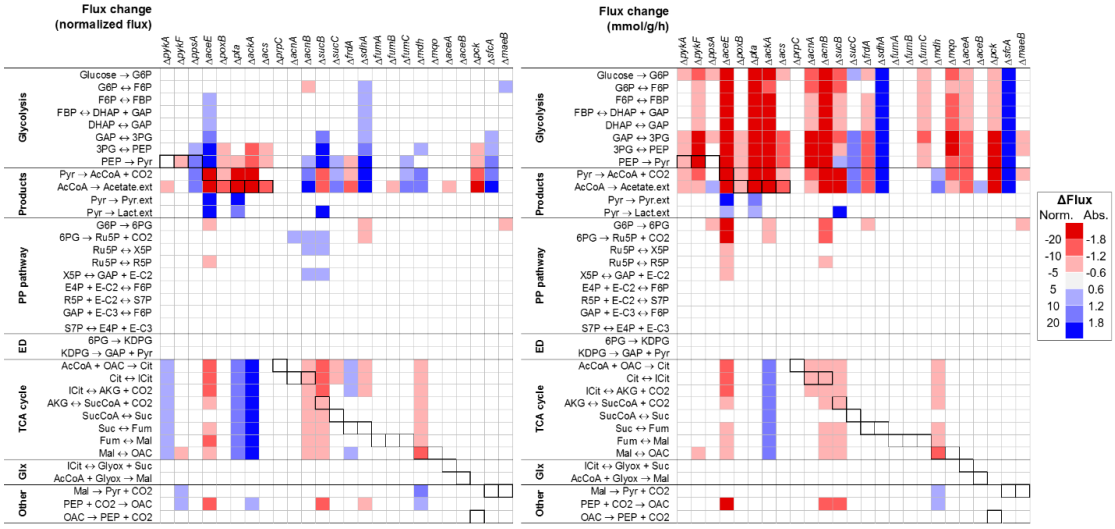


Figure 9.6: Flux changes from the wild-type, in normalized (per 100 glucose) (**A**) and absolute (mmol/g_{DW}/h) (**B**). Colors denote the sign and magnitude of the change, with blue indicating flux increase and red flux decrease, per the key to the right. The flux directly affected by each knockout is bolded.

In Fig. 9.6, the specific occurrences of absolute increases are apparent, and include pyruvate and lactate production, TCA cycle flux in *ΔackA*, and glycolytic flux in *ΔsdhA* and *ΔsfcA*. The latter two have elevated glucose uptake, glycolysis, and acetate secretion rates. *ΔsdhA* knocks out the succinate dehydrogenase, which does perturb the TCA cycle. *ΔsfcA* encodes the NAD-dependent malic enzyme, which does not carry flux in the wild-type. The mechanisms of these phenotypes should be investigated further, as regulatory changes of potential interest are possible.

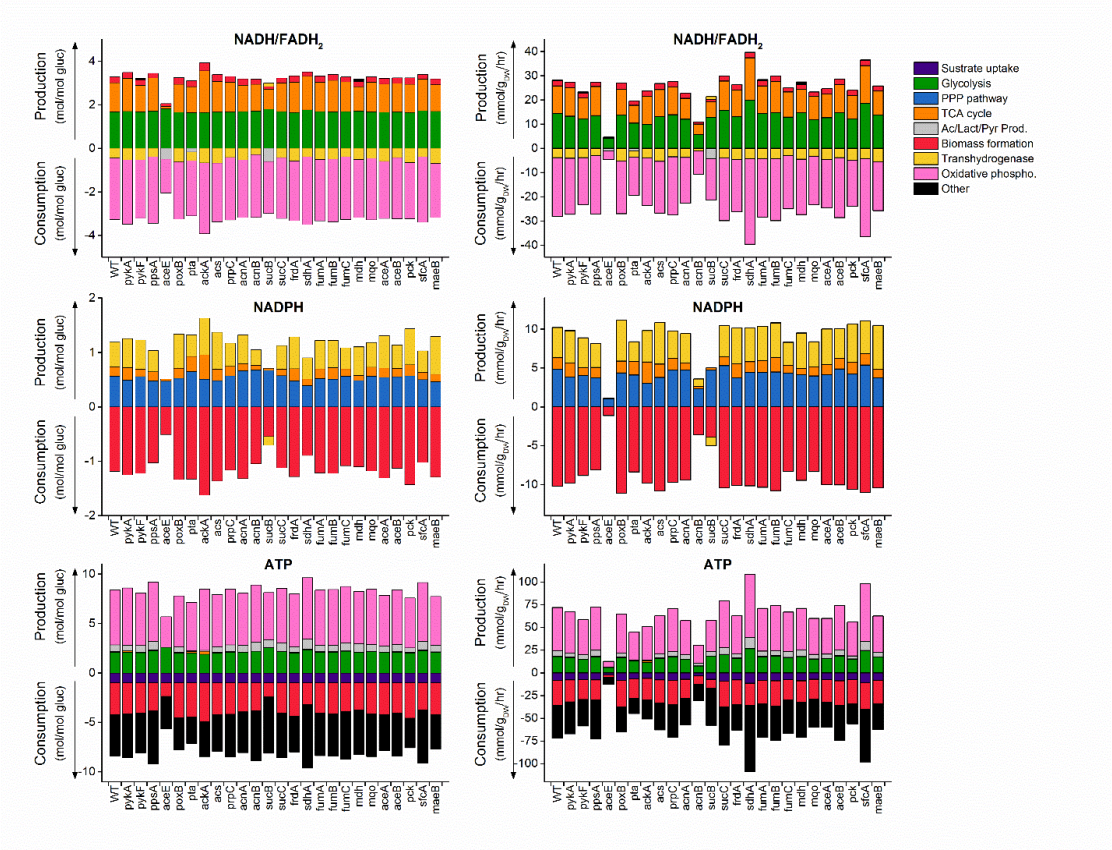


Figure 9.7: Cofactor balances, in normalized (left) and absolute (right) units. The contribution of the major central carbon metabolic pathways (key, top right) to the production and consumption of NADH/$FADH_2$ (lumped), NADPH, and ATP are shown. Bars representing positive values, show cofactor production, and negative values show consumption. The estimate of ATP generation from oxidative phosphorylation assumes a constant P/O ratio of 2.0 across strains, and the cell maintenance ATP cost ("Other") is calculated as the balance of production and consumption.

Changing fluxes result in changes to the sources and sinks of metabolic cofactors, including NADH, $FADH_2$, NADPH, and ATP. The pathway contributions to the production and consumption of these are shown in Fig. 9.7, in normalized and absolute terms. As discussed above, the changes in central carbon metabolism change

the role of the pyridine cofactor transhydrogenase in some strains. In the wild-type, it is responsible for converting a small fraction of cellular NADH to approximately 40% of the needed NADPH. In *ΔaceE* and *ΔsucB*, lactate production contributes to consuming some NADH, and severely reduced biomass yields reduce the demand for NADPH. As a result, the transhydrogenase flux is zero in *ΔaceE*, and reversed in *ΔsucB*. Interestingly, despite the many perturbations to the TCA cycle in these strains, the only knockout with severely reduced oxidative phosphorylation (in normalized terms) was *ΔaceE*.

9.3.4 Multivariate Analysis

As was done in Chapter 7, the absolute fluxes across all strains can be analyzed for pairwise correlations (Fig. 9.8). In this set, the correlation between the PYK (PEP→Pyr) and PDH (Pyr→AcCoA) fluxes is weaker, due to the strains that secrete products from pyruvate. An important, and nontrivial, correlation is the acetate overflow relationship between the PDH flux and acetate secretion. This reflects the proposed mechanism of glycolytic flux exceeding the capacity of the TCA cycle, and the excess being excreted as acetate. As these knockouts directly impact the product formation and TCA cycle pathways, it was unclear if the relationship would hold. As is clear in the figure, however, the correlation held and was still strong (ρ=0.96), with the same line of best fit as for the upper knockouts. Notably, the acetate knockouts, particularly *ΔackA*, which did not secrete pyruvate products and increased its TCA flux in absolute terms, were the most significant deviants (less acetate than expected per PDH flux). The two strains with elevated absolute glycolytic fluxes, *ΔsdhA* and *ΔsfcA*, laid clearly on the same trendline.

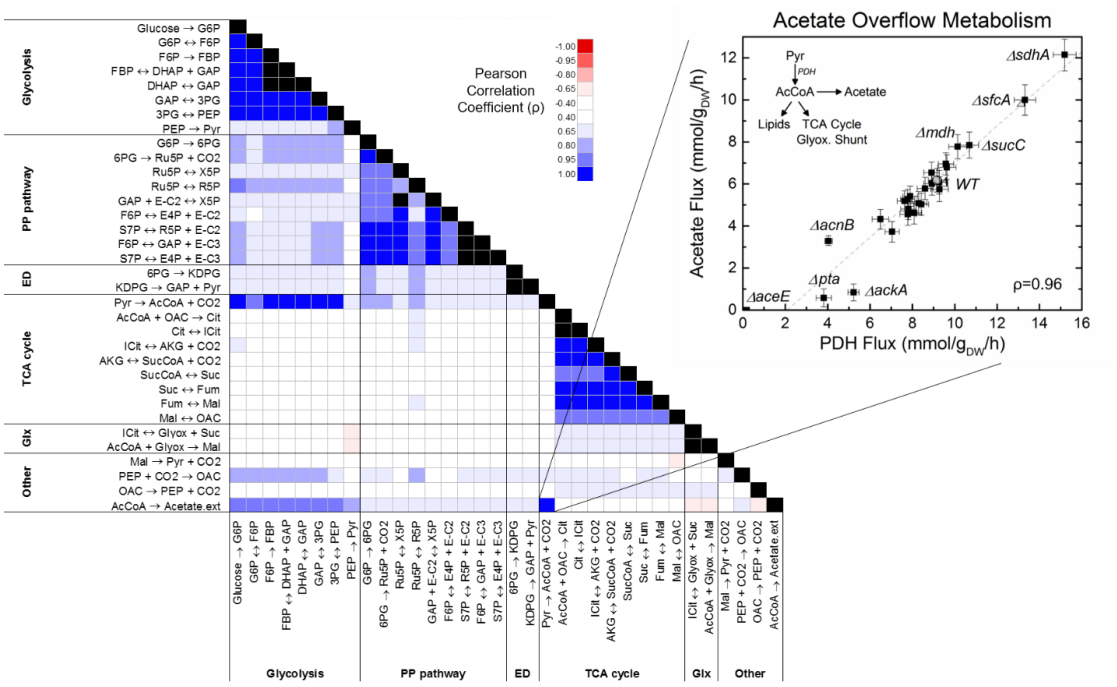


Figure 9.8: Pairwise correlations of fluxes in central carbon metabolism. Stoichiometrically trivial correlations are controlled for by comparison to correlations seen in 100 randomly generated flux maps (Fig. 7.7B). The heat map reflects the Pearson correlation coefficient. A key, non-trivial relationship is the acetate overflow relationship between pyruvate dehydrogenase flux and acetate production. The scatter plot of this strong correlation is shown.

The principal component analysis, introduced in Chapter 7 and utilizing absolute fluxes from throughout the network, can also be extended to include the additional strains presented here. This is shown in Fig. 9.9, where the top two principal components account for 59% of the total variance across the 45 strains, and the top six account for 91%. The generalized identities of PC1 and PC2 are the same as in Chapter 7, where the growth rate related effects are in PC1 (i.e., growth rate, glycolytic rate, and acetate production rate positively related, with the rarely active glyoxylate shunt and PPCK pathways negatively related), and PC2 capturing the TCA

cycle (positive) and the pentose phosphate pathway and lactate production (negative). As expected, many strains cluster around the wild-type, but other interesting clusters are also observed: *Δgnd*, *ΔackA*, and *Δzwf*; and *ΔaceE*, *Δpgi*, *ΔpfkA*, and *ΔacnB*.

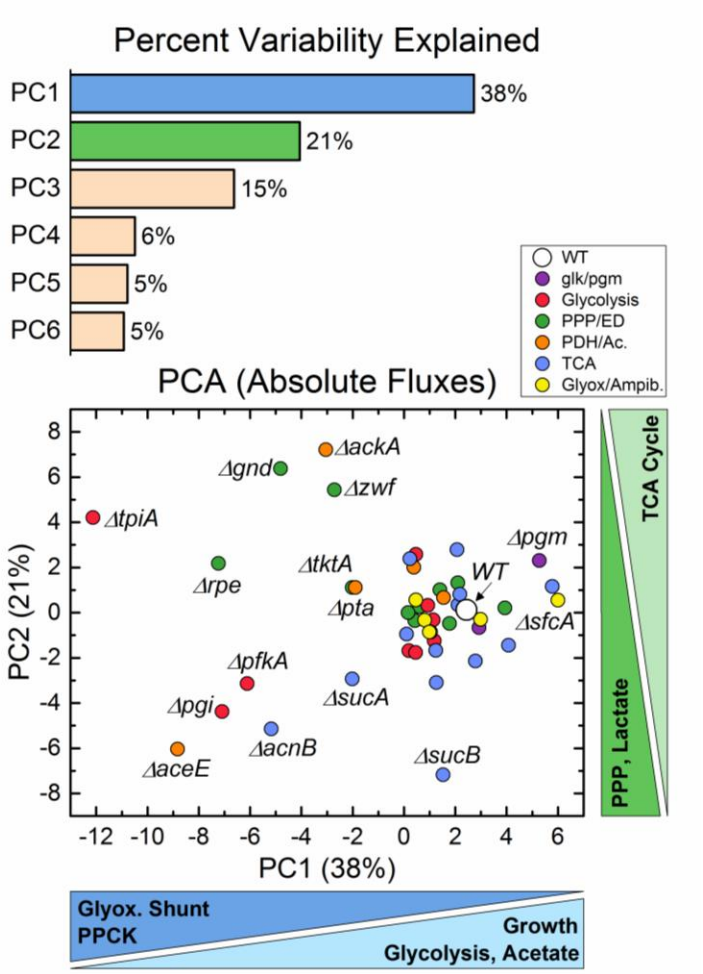


Figure 9.9: Principal component analysis of absolute fluxes in 45 knockout strains. The percentage of total variability explained by the top six PC's are shown above, and the distribution of the strains across the top two PC's are shown below. The colors reflect pathway association per the key, and the triangles to the bottom and right of the axis reflect the primary determinants of PC1 and PC2, respectively.

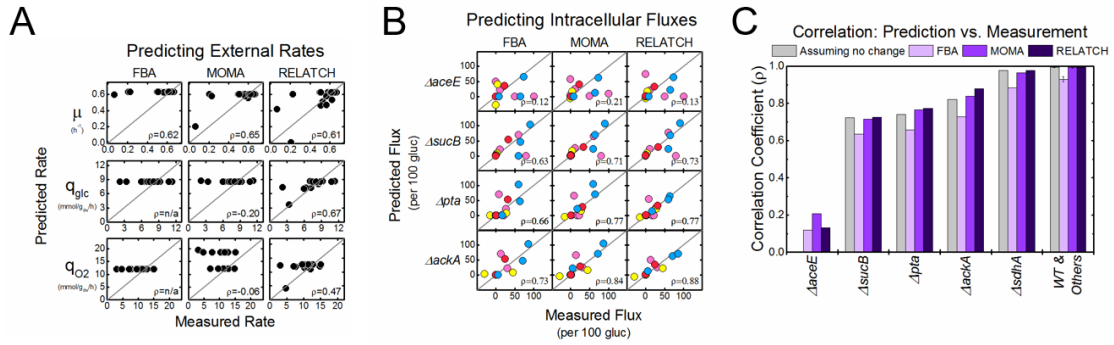


Figure 9.10: Comparison of measurements to predictions made by three COBRA models: FBA, MOMA and RELATCH. (A) External rates: growth, glucose uptake, and oxygen uptake. Predictions (y-axis) vs. measurements (x-axis) are plotted, with correlation coefficients shown. (B) Normalized intracellular fluxes for the four significantly perturbed strains. The fluxes correspond to intracellular pathways as follows: glycolysis (light blue), oxPPP and ED pathways (red), TCA cycle and glyoxylate shunt (yellow), product and ana/cataplerotic reactions (pink). (C) Correlations for the normalized intracellular fluxes for all strains. The gray bar reflects a trivial "no change from wild-type" model, i.e., strains with a good fit by this model are quite unperturbed and are therefore trivial to predict.

Finally, the metabolic flux phenotypes of these strains were compared to COBRA predictions, made by FBA, MOMA, and RELATCH. The results are summarized in Fig. 9.10. As established in Chapter 8, these methods perform quite poorly at predicting external rates (Figs. 8.8, 9.10A). For assessing the accuracy of predictions for normalized intracellular flux patterns, 13 key fluxes (see Methods) spanning the central carbon and excretion pathways were examined. The correlation between prediction and measurement was used to describe the agreement between the two. As seen in Fig. 9.10C, many strains had insignificant changes from the wild-type, i.e., a trivial model that assumes no flux changes at all performed very well for 20 of the strains. Four strains that had non-trivial changes are shown in Fig. 9.9B as well:

ΔaceE, *ΔsucB*, *Δpta*, and *ΔackA*. None of the COBRA methods predicted significant pyruvate or lactate excretion, which harmed the agreement in three of these four. All methods performed very poorly with *ΔaceE*, with MOMA performing best with a correlation coefficient of 0.21. Predictions were altogether more accurate for *ΔsucB*, but still struggled due to the fluxes around pyruvate (elevated PYK, active lactate and pyruvate secretion). For both *Δpta* and *ΔackA*, the models all predict that high acetate secretion should be maintained through other pathways.

9.4 Conclusions

The results presented here significantly further our quantitative understanding of the pathways of lower central carbon metabolism. Particularly, understanding the dynamics of pyruvate and AcCoA production and consumption is central to efforts to describe and model metabolism. As key node metabolites, these species are also of high interest as precursors for products of interest in biotechnology. The insights that blocking the PDH with *ΔaceE* allow for the efficient conversion to pyruvate-derived products could be quite useful in strain engineering. Additionally, the differing effects of blocking acetate production at *pta* and *ackA* could be taken advantage of, in the case where either pyruvate availability or TCA cycle flux was desired. In the TCA cycle, it is useful to know which reactions are essential (i.e., upstream of AKG) and yet can be modulated to significant effect (*acnB*), and which are not but still cause significant rate limitations when perturbed (*sucB*). These results also inform additional lines of scientific inquiry, particularly in the differing regulatory effects of *Δpta* and *ΔackA*, the mechanism of elevated glycolytic rates in *ΔsdhA* and *ΔsfcA*, and the oxaloacetate reduction to malate in *Δmdh*.

Taken together with the results from upper central carbon metabolism (Chapter 7), we have presented a significant new resource of 45 characterized knockout strains spanning central carbon metabolism. This resource will be useful to metabolic engineers who have specific interests in genes and pathways, as well as to the broader systems biology community. Although we have presented some analysis of these flux results, many more analyses and potential insights are possible and will hopefully be explored. More specifically, these fluxes will be highly applicable to metabolic modeling, both in assessing existing models as was done here, but also in developing new approaches. Mutant flux data can be used for the parameterization of kinetic models by the ensemble method, and the breadth and precision of this study will be a significant advantage in that application (Khodayari et al., 2014; Khodayari and Maranas, 2016).

Chapter 10

DISSECTING THE GENETIC AND METABOLIC MECHANISMS OF ADAPTATION TO THE KNOCKOUT OF PGI, A MAJOR METABOLIC ENZYME IN *ESCHERICHIA COLI*

Reprinted with permission from: Long CP, Gonzalez JE, Feist AM, Palsson BO, Antoniewicz MR (2018) Dissecting the genetic and metabolic mechanisms of adaptation to the knockout of a major metabolic enzyme in *Escherichia coli*. Proc. Natl. Acad. Sci. USA. 115(1) 222-227

10.1 Introduction

In the study of microbial metabolism, understanding responses to genetic perturbation and adaptive evolution is fundamental. Mutations in metabolic enzymes force a rewiring of flux in the cell, the nature of which can inform our understanding of alternative pathways, kinetics and regulation (Ishii et al., 2007; Long and Antoniewicz, 2014a). Adaptive laboratory evolution (ALE) is a powerful approach by which a microbe is cultured continuously for many generations, typically achieving improved fitness (e.g., faster growth rate) through natural selection. The final mutants are then sequenced and phenotypically characterized (Dragosits and Mattanovich, 2013; Herring et al., 2006), with the identification of causal genetic mutations and mechanistic insights enabled by replicate experiments and detailed 'omics' analysis (LaCroix et al., 2015). Often used to study adaptations to environmental conditions like varied substrates (Conrad et al., 2009; Herring et al., 2006; LaCroix et al., 2015; Lee and Palsson, 2010; Sandberg et al., 2016) or the presence of toxic chemicals (Atsumi et al., 2010; Horinouchi et al., 2010; Mundhada et al., 2017; Reyes et al.,

2012), ALE has also been previously applied to study the adaptive responses to genetic perturbations such as the loss of major metabolic enzymes (Charusanti et al., 2010; Fong et al., 2006; Fong and Palsson, 2004). These studies provide a valuable dimension for both evolutionary and metabolic research, as new metabolic phenotypes are evolved subject to significant and unnatural constraints. The metabolic response to knockouts before and after adaptive evolution has been an area of significant theorizing and *in silico* model development (Cornelius et al., 2011; Fong and Palsson, 2004; Kim and Reed, 2012).

Phosphoglucose isomerase (*pgi*) knockouts of *E. coli* are of significant interest in metabolic engineering and have been the subject of many investigations (Long and Antoniewicz, 2014a). Phosphoglucose isomerase catalyzes the first reaction in glycolysis, the conversion of glucose 6-phosphate (G6P) to fructose 6-phosphate (F6P), which in the wild-type during aerobic growth on glucose catabolizes approximately 70% of glucose (Crown et al., 2015a; Leighty and Antoniewicz, 2013). Its loss results in a correspondingly severe growth impairment (70-80% lower growth rate) (Fong et al., 2006; Long et al., 2016b) as the oxidative pentose phosphate pathway (oxPPP) and Entner-Doudoroff pathway must compensate. Several studies have used ^{13}C-metabolic flux analysis (^{13}C-MFA) to characterize *Δpgi*, frequently describing the activation of normally latent (i.e., non-utilized) pathways and a redox imbalance caused by over-production of NADPH in the pentose phosphate pathway (Canonaco et al., 2001; Fischer and Sauer, 2003; Ishii et al., 2007; Toya et al., 2010; Usui et al., 2012). The major flux, redox, and growth rate changes caused by loss of *pgi* make it a rich target for ALE experiments (Charusanti et al., 2010; Fong et al., 2006). Previously, Charusanti et al. adaptively evolved ten strains in replicate

experiments over 50 days of continuous culture in glucose minimal media, reporting significant growth recovery of 3.6-fold (Charusanti et al., 2010). However, no underlying intracellular fluxes have been reported for these strains or any similarly large scale ALE study of genetic mutants.

To gain fundamental insight into the mechanisms and outcomes of adaptive evolution, both the mutations and the selected-upon phenotype (here, metabolism) must be measured. In this study, we applied high-resolution ^{13}C-MFA and next-generation sequencing to the ten evolved *Δpgi* strains and the parental strain reported previously (Charusanti et al., 2010). Novel mutations were identified, and comparisons to recently reported wild-type ALE studies (LaCroix et al., 2015; Sandberg et al., 2016) helped to identify mutations unique to *Δpgi*. Together with complimentary fluxomic information, a detailed picture of how *Δpgi* metabolically adapts to achieve faster growth is attained. Areas of convergence and divergence on the genetic and fluxomic levels highlight the large number of genetic solutions possible for achieving similar metabolic phenotypes, as well as some differences in metabolic optima. In several cases, specific causal mutation-flux relationships were identified.

10.2 Methods

10.2.1 DNA Sequencing

Clonal genomic DNA was extracted using NucleoSpin Tissue kit from Macherey-Nagel (Bethlehem, PA). The quality of DNA was assessed with ultraviolet absorbance ratios using a Nano drop. DNA was quantified using Qubit dsDNA High Sensitivity assay. Paired-end resequencing libraries were generated using a Nextera

XT kit from Illumina (San Diego, CA) with 1 ng of input DNA total. Sequences were obtained using an Illumina MiSeq MS-102-3003 600 cycle kit. The breseq pipeline (Barrick et al., 2009) version 0.27.1 was used to map sequenced reads and identify mutations relative to the *E. coli* K-12 MG1655 genome (NCBI accession number NC_000913.3). All samples had a mean coverage of at least 80.

10.2.2 Materials

Chemicals and M9 minimal medium were purchased from Sigma-Aldrich (St. Louis, MO). Isotopic tracers were purchased from Cambridge Isotope Laboratories (Tewksbury, MA): [1,6-^{13}C]glucose (99.2 % ^{13}C), [1,2-^{13}C]glucose (99.7 %), and [U-^{13}C]alanine (98+% ^{13}C). The isotopic purity and enrichment of all tracers were validated by GC-MS analysis (Cordova and Antoniewicz, 2016; Sandberg et al., 2016). All solutions were sterilized by filtration.

10.2.3 Strains and Culture Conditions

The strains used in this study were previously described (Charusanti et al., 2010). Briefly, the starting strain for adaptive evolution was an *E. coli Δpgi* strain constructed from wild-type *E. coli* K12 MG1655 (ATCC, Manassas, VA) (Fong et al., 2006). Adaptive laboratory evolution had been carried out with ten *Δpgi* replicates using serial passaging for 50 days in 250 mL flasks containing glucose M9 minimal medium. For the wild-type strain data presented here, physiological culture measurements were used as reported (Crown et al., 2015a).

For ^{13}C-tracer experiments, strains were cultured aerobically in glucose M9 minimal medium at 37°C in mini-bioreactors with 10 mL working volume (Long et al., 2016b). Pre-cultures were grown overnight and then used to inoculate the

experimental culture at an OD_{600} of 0.01. For ^{13}C-MFA, glucose tracers were added at the beginning of the culture. Cells were harvested for GC-MS analysis at mid-exponential growth when OD_{600} was approximately 0.7. For the quantification of the pyruvate to PEP flux, [U-^{13}C]alanine tracer experiments were performed and the data were analyzed as described (Christopher P. Long et al., 2017a). Briefly, the cultures were inoculated as above in naturally labeled glucose M9 medium. At an OD_{600} of 0.5, a bolus of 10 mM [U-^{13}C]alanine was added and the cells were collected at an OD_{600} of 1.5 for GC-MS analysis.

10.2.4 Analytical Methods

Cell growth was monitored by measuring the optical density at 600nm (OD600) using a spectrophotometer (Eppendorf BioPhotometer). The OD600 values were converted to cell dry weight concentrations using a previously determined OD600-dry cell weight relationship for E. coli (1.0 OD600 = 0.32 gDW/L) (Long et al., 2016b). After centrifugation, the supernatant was separated from the cell pellet. The cell pellets were washed with glucose-free M9 medium prior to subsequent analysis. Glucose concentrations in the media were determined using YSI 2700 biochemistry analyzer (YSI, Yellow Springs, OH). Acetate concentrations were determined using an Agilent 1200 Series HPLC (Gonzalez et al., 2017). Growth rate was calculated using linear regression of the natural logarithm of the OD600 and time. Biomass yields and oxygen uptake were estimated as part of 13C-MFA, and used to calculate absolute uptake rates.

10.2.5 Gas Chromatography-Mass Spectrometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 µm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230°C, the MS quad temperature at 150°C, the interface temperature at 280°C, and the inlet temperature at 250°C. GC-MS analysis of *tert*-butyldimethylsilyl (TBDMS) derivatized proteinogenic amino acids was performed as described (Long and Antoniewicz, 2014b). Labeling of glucose (derived from glycogen) and ribose (from RNA) were determined as described (Long et al., 2016a; McConnell and Antoniewicz, 2016). In all cases, mass isotopomer distributions were obtained by integration (Antoniewicz et al., 2007b) and corrected for natural isotope abundances (Fernandez et al., 1996).

10.2.6 Metabolic Network Model and ^{13}C-Metabolic Flux Analysis

The metabolic network model used for ^{13}C-MFA is provided in Appendix Table E.1. The model (Crown et al., 2015a; Gonzalez et al., 2017) includes all major metabolic pathways of central carbon metabolism, lumped amino acid biosynthesis reactions, and a lumped biomass formation reaction. For the wild-type, labeling data reported for 14 parallel tracer experiments (Crown et al., 2015a) were refitted. ^{13}C-MFA calculations were performed using the Metran software (Yoo et al., 2004), which is based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007a). Fluxes were estimated by minimizing the variance-weighted sum of squared residuals (SSR) between the measured and model predicted mass isotopomer distributions and acetate yield using non-linear least-squares regression. For integrated

analysis of parallel labeling experiments, the data sets were fitted simultaneously to a single flux model (Leighty and Antoniewicz, 2013). Flux estimation was repeated 10 times starting with random initial values for all fluxes to find a global solution. At convergence, accurate 95% confidence intervals were computed for all estimated fluxes by evaluating the sensitivity of the minimized SSR to flux variations. Precision of estimated fluxes was determined as follows (Antoniewicz et al., 2006):

Flux precision (stdev) = [(flux $_{\text{upper bound 95\%}}$) − (flux $_{\text{lower bound 95\%}}$)] / 4

To describe fractional labeling of biomass amino acids G-value parameters were included in ^{13}C-MFA. As described previously (Antoniewicz et al., 2007c), the G-value represents the fraction of a metabolite pool that is produced during the labeling experiment, while 1-G represents the fraction that is naturally labeled (e.g., from inoculum). By default, one G-value parameter was included for each measured amino acid in each data set. Reversible reactions were modeled as separate forward and backward fluxes. Net and exchange fluxes were determined as follows: $v_{net} = v_f - v_b$; $v_{exch} = \min(v_f, v_b)$. To determine the goodness-of-fit, ^{13}C-MFA fitting results were subjected to a χ^2-statistical test (Antoniewicz et al., 2006).

For the [U-^{13}C]alanine tracer experiments, the fraction of PEP derived from pyruvate was determined by least-squares regression using the measured mass isotopomer distributions (MID) of PEP (determined from phenylalanine *m/z* 302 fragment, C1-C2), oxaloacetate (determined from aspartate *m/z* 302 fragment, C1-C2), and pyruvate (determined from valine *m/z* 260 fragment, C2-C5), after correction for unlabeled biomass (see Chapter 5 for more information).

$$MID_{PEP} = (\%PEP\ from\ Pyr) * MID_{Pyr} + (\%PEP\ from\ OAC) * MID_{OAC} + (\%PEP\ from\ gluc) * MID_{unlabeled}$$

The absolute PPS/EI flux was calculated from this value using the estimated fluxes of enolase (3PG→PEP) and PCK (OAC→PEP) from ^{13}C-MFA.

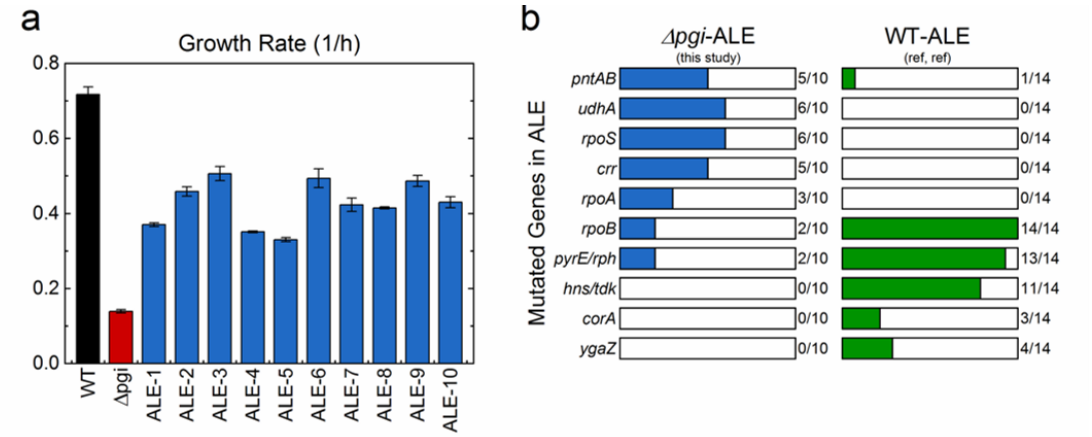
10.3 Results

10.3.1 Recovery of Growth Fitness in *Δpgi* ALE Strains is Associated with Unique Mutations

In *E. coli* K-12 MG1655, the knockout of phosphoglucose isomerase (*pgi*) results in a reduction in growth rate of approximately 80% compared to the wild-type, from 0.72 h^{-1} to 0.14 h^{-1}. Following ALE, a significant fraction of this growth rate can be recovered (46-71% of the wild-type growth rate, Fig 10.1A). These 2.4 to 3.6-fold increases in growth rate are quite large when compared with, for example, similar ALE experiments with wild-type *E. coli* which reported 1.6-fold increases in growth rate (LaCroix et al., 2015).

To assess the genetic basis of the large increases in growth rate, whole-genome sequencing was performed and sequences were mapped to the *E. coli* reference to identify mutations in ten independent ALE experiments. Recent advances in sequencing allowed for improved determination of mutations in clones isolated from final populations, particularly of insertion sequences (IS elements), as compared with a previous effort which utilized both a microarray hybridization-based method and an earlier Illumina short-read technology protocol (Charusanti et al., 2010). In all, 52 unique mutations were identified across the ten ALE strains, spanning 34 different genetic regions (29 structural mutations, i.e., within in ORF, and 5 intergenic regions). The complete mutation table is in the Supplementary File. A key advantage and reason for using replicate ALE experiments is to use mutation frequency to help differentiate causal mutations from genetic ‘hitch-hikers’ that do not affect fitness (LaCroix et al.,

2015). Fig. 10.1B lists the genes which were mutated in ≥ 2 of the 10 *Δpgi* ALE strains, as well as the genes frequently mutated in reported wild-type K-12 MG1655 ALE studies performed in similar conditions (LaCroix et al., 2015; Sandberg et al., 2016). The two sets are striking in their lack of overlap; mutations that occur in almost every reported wild-type ALE experiment, in *rpoB*, *pyrE/rph*, and *hns/tdk* occur rarely or not at all in *Δpgi* strains. Instead, the *Δpgi* ALE strains have a high frequency of mutations in the pyridine nucleotide transhydrogenases *pntAB* and *sthA*, the transcription factor *rpoS*, and the PTS sugar transport system component *crr*. The distribution of these mutations across the ten *Δpgi* ALE strains is also in Fig. 1A, showing that while some strains had many of these common mutations (ALE-2 had 6 out of the top 7), others had fewer (ALE-8 had only 1/7). This likely reflects less common but equally effective adaptive routes. It is worthwhile to mention that some genes or genetic regions had many unique mutations in parallel evolutions (e.g., 6 for *rpoS* and 5 for *pntAB*). In contrast, the IS element insertion in *crr* was identical in 5 different strains. Both patterns clearly demonstrate evidence for causality. It was previously demonstrated that the combination of *rpoS* and *sthA* mutations are causal for increased growth in *Δpgi*, and exhibit positive epistasis (Charusanti et al., 2010). The difference in mutation profiles demonstrate that there are unique selective pressures in *Δpgi*, which result in unique adaptive responses. To further investigate how these mutations enabled the large increases in growth rates from the initial perturbed *Δpgi* state, we next characterized the carbon metabolism of each strain using ^{13}C-MFA.



	Δpgi ALE-									
Gene	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
pntAB (5)	✓	✓		✓			✓			✓
sthA (4)	✓	✓	✓				✓	✓	✓	
rpoS (6)		✓	✓	✓	✓	✓	✓			
crr (1)		✓		✓	✓		✓			✓
rpoA (2)		✓								✓
rpoB (2)	✓								✓	
pyrE/rph (1)		✓				✓				

Figure 10.1: Growth rate recovery in evolved *Δpgi* strains is supported by unique mutations. (a) The growth rate is severely reduced in the *Δpgi* knockout strain relative to wild-type (WT). This is significantly, but not completely, recovered through adaptive laboratory evolution (ALE), as seen in ten independently evolved *Δpgi*-ALE strains (growth rate, mean ± sem, n≥3). (b) Whole-genome sequencing of ALE strains allows for analysis of frequently occurring mutations. Here, the frequency of mutations in 10 *Δpgi*-ALE strains is compared to those previously reported for 14 WT-ALE strains of *E. coli* (LaCroix et al., 2015; Sandberg et al., 2016). The profile of mutations is quite distinct for *Δpgi*-ALE and WT-ALE strains, highlighting the unique evolutionary pressures and mechanisms of growth rate recovery in *Δpgi*. The distribution of mutations from the seven most frequently affected genes in the 10 *Δpgi*-ALE is also shown in (**a**, bottom), with the number of unique mutations per gene noted.

10.3.2 Activation of Latent Pathways, or Not?

To characterize the rewiring of central carbon metabolism in the parental and ALE *Δpgi* strains, high-resolution ^{13}C-MFA was performed. The analysis consisted of two parallel labeling experiments with [1,2-^{13}C] and [1,6-^{13}C]glucose (an experimental design previously identified as providing optimal flux estimate precision (Crown et al., 2016b)), and the simultaneous fitting of labeling from proteinogenic amino acids, the ribose moiety of RNA, and glucose moiety of glycogen (Long et al., 2016a) to estimate fluxes. For the wild-type, data from parallel labeling experiments previously reported were refitted (Crown et al., 2015a). The full network model is in Appendix E.1, and the estimated metabolic fluxes are in the Supplementary File.

In Fig. 10.2A-C, the estimated intracellular fluxes of the wild-type, *Δpgi* parental strain, and *Δpgi* ALE-3 are summarized. ALE-3 was the fastest-growing *Δpgi* strain (0.51 h^{-1}) and had an intracellular flux distribution typical of most of the ALE strains. The growth rates and glucose uptake rates for each strain are noted, and the fluxes shown are normalized to 100 units of glucose uptake. In all *Δpgi* strains, the forward and reverse fluxes of the PGI reaction were estimated to be zero, thus confirming the *pgi* knockout. The unevolved *Δpgi* (Fig. 10.2B) was found to utilize reactions and pathways that carry minimal flux in the wild-type (Fig. 10.2A), including the ED pathway, glyoxylate shunt, and PCK reaction (oxaloacetate to PEP). These flux changes have been noted in previous studies, and have been described as 'latent pathway activation' (Fischer and Sauer, 2003; Hua et al., 2003; Ishii et al., 2007; Toya et al., 2010; Usui et al., 2012). After adaptive evolution, ALE-3 (Fig. 10.2C) and the other ALE strains significantly reduced the usage of these pathways, for example the glyoxylate shunt flux was reduced from 25 to 6, and the PCK reaction from 22 to 3 in ALE-3. This 're-repression' following adaptive evolution has also

been previously observed and been the focus of various speculations and computational analyses (Cornelius et al., 2011; Fong et al., 2006). The previous terminology implies the presence of a transient regulatory response, activated in response to the stress caused by the knockout and then repressed during evolution to facilitate faster growth.

When interpreting metabolic fluxes, it is important to consider both normalized fluxes (e.g., relative to glucose uptake) and absolute fluxes (mmol/g_{DW}/h), as each provide complimentary information. Several key fluxes are shown in both units in Fig. 10.2D, corresponding to the pathway map shown in Fig. 10.2E. The glucose uptake rates used to calculate absolute fluxes are in Appendix Table E.2. In Fig. 10.2B and 10.2D, we can see that although the oxPPP is the dominant route of glucose catabolism in *Δpgi*, the absolute flux is reduced by roughly half in the unevolved strain relative to the wild-type. It has been previously reported that G6P accumulates in *Δpgi* and that G6PDH (encoded *zwf*), the first step in the oxPPP, is likely rate limiting for growth due to allosteric inhibition caused by an elevated NADPH/$NADP^+$ ratio (Toya et al., 2010). This limitation is overcome in the evolved strains, where oxPPP fluxes are increased by 3 to 4-fold, to rates higher even than in the WT (by up to 2-fold). Intriguingly, the highest absolute flux was observed in ALE-9, which was the only evolved strain with a mutation in *zwf*.

In cases of 'activated latent' pathways, the absolute fluxes provide an especially illuminating perspective. There was no statistically significant increase in the absolute ED, glyoxylate shunt, or PCK fluxes in *Δpgi* compared to the wild-type. Instead, a similar low level of absolute flux was maintained, which only appeared much larger in relative terms due to the dramatic reductions in absolute glucose uptake

caused by the *pgi* knockout. Very little change was observed in the ALE strains for these latent pathway fluxes, with the lone exception of an elevated PCK flux in ALE-9. These results challenge the notion that these latent pathways are 'activated' in a regulatory sense that increases their absolute flux capacity. Perhaps more likely is that these enzymes are expressed at low levels in the wild-type, and this is maintained in the *Δpgi* strain, where due to the perturbation in glycolysis the same small rates of flux play a larger relative role. The 're-repression' in the ALE strains, then, is instead simply the recovery of faster glucose uptake rate (Appendix Fig. E.1).

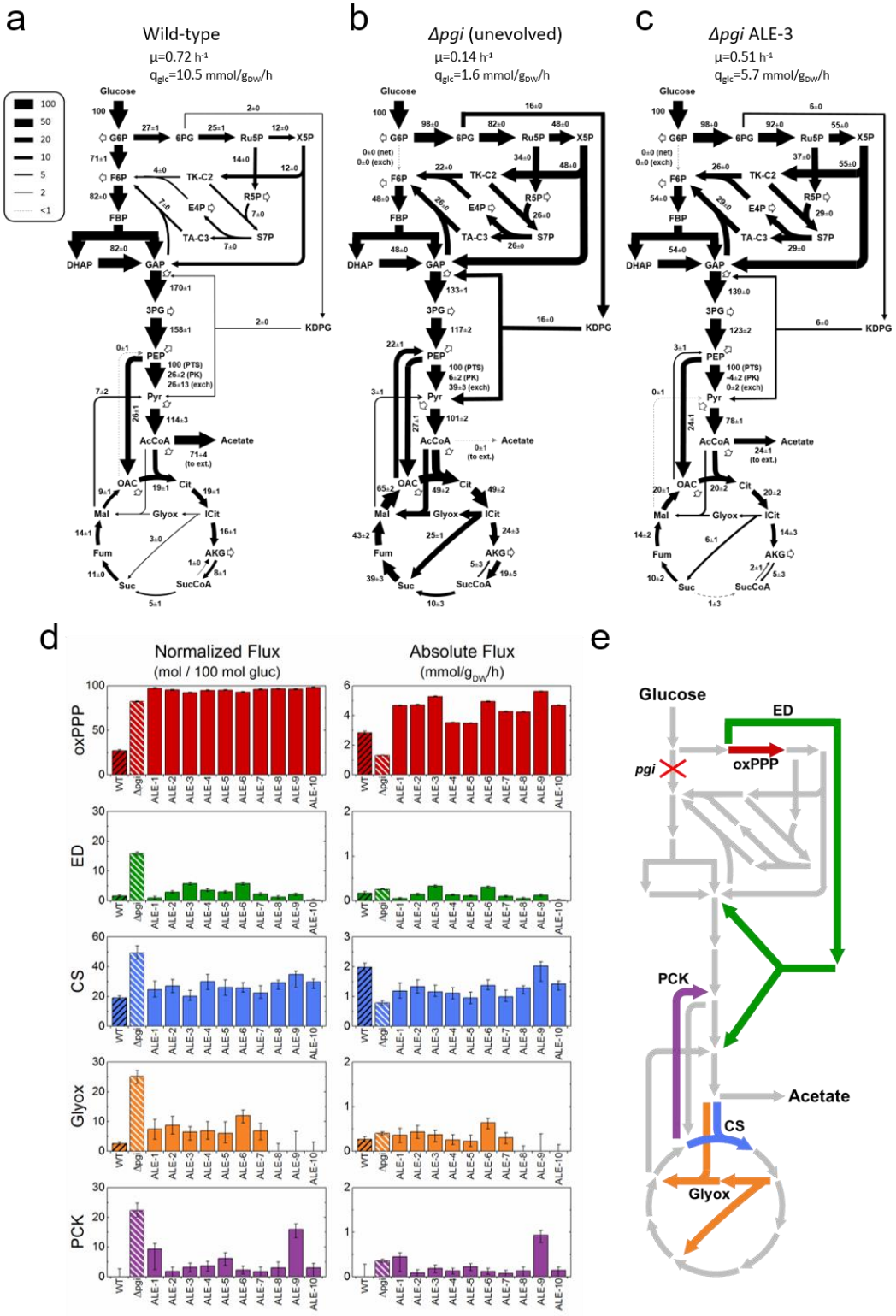

a
Wild-type
μ=0.72 h-1
qglc=10.5 mmol/gDW/h
b
Δpgi (unevolved)
μ=0.14 h-1
qglc=1.6 mmol/gDW/h
c
Δpgi ALE-3
μ=0.51 h-1
qglc=5.7 mmol/gDW/h
Glucose
G6P
6PG
Ru5P
X5P
F6P
TK-C2
R5P
E4P
FBP
TA-C3
S7P
DHAP
GAP
3PG
KDPG
PEP
Pyr
AcCoA
Acetate
OAC
Cit
Mal
Glyox
ICit
Fum
AKG
Suc
SucCoA
d
Normalized Flux
(mol / 100 mol gluc)
Absolute Flux
(mmol/gDW/h)
oxPPP
ED
CS
Glyox
PCK
WT
Δpgi
ALE-1
ALE-2
ALE-3
ALE-4
ALE-5
ALE-6
ALE-7
ALE-8
ALE-9
ALE-10
e
Glucose
ED
pgi
oxPPP
PCK
Acetate
CS
Glyox

Figure 10.2: ^{13}C-Metabolic flux analysis reveals large flux redistributions, but ambiguous activation of latent pathways. Intracellular flux distributions, normalized to 100 units of glucose uptake, are shown for the wild-type (a), unevolved *Δpgi* (b), and a representative evolved *Δpgi* strain ALE-3 (c). The net and exchange fluxes through the PGI reaction were determined to be zero for all knockout strains, thus confirming the *pgi* knockout. Significant flux rewiring is observed in *Δpgi*, including increased usage of oxPPP and TCA, as well as activation of the ED pathway, glyoxylate shunt, and PCK reaction. Many of these apparently activated latent pathways are re-repressed in the ALE-3 strain, and other ALE strains (d). Selected fluxes are shown in normalized units (as in a-c) as well as absolute units (mmol/g_{DW}/h), in the context of central metabolic pathways (e). The error bars in (d) reflect the 95% confidence intervals of flux estimates. The absolute fluxes reveal that the apparent activation of the ED pathway, glyoxylate shunt, and PCK reactions, and increased TCA cycle fluxes in *Δpgi* do not correspond to increases in absolute flux, and that the relative increases are therefore results of lower glucose uptake. Thus, what may be interpreted as pathway activation when expressed as relative pathway usage may not represent a real increase in expression or pathway capacity.

10.3.3 Energy Metabolism is not Significantly Affected by Adaptive Evolution in *Δpgi*

The measured metabolic fluxes can also inform a broader analysis of energy metabolism in these strains. In Fig. 10.3A-B, oxygen uptake rates and acetate yields are shown. In the unevolved *Δpgi*, the oxygen uptake is reduced to 4.3 mmol/g_{DW}/h, down from 15 mmol/g_{DW}/h in the wild-type, corresponding to the overall slowed metabolism and growth rate. The unevolved *Δpgi* does not produce acetate, as the citrate synthase (CS) flux can easily accommodate all the flux from acetyl-CoA at less than half of its wild-type rate (Fig. 10.2D). In the ALE strains, oxygen uptake recovers to 62-91% of the wild-type flux and some strains produce acetate. This may represent a limit in TCA cycle or oxidative phosphorylation capacity that these strains encounter, above which excess glycolytic flux is diverted to acetate production. This

acetate overflow effect is shown in Fig. 10.3C, which shows that the absolute flux through the pyruvate dehydrogenase (acetyl-CoA generation) strongly correlates with the acetate secretion flux in all strains. Fig. 10.3D-E show the normalized cofactor balances for ATP and $NADH/FADH_2$ (the electron carriers used in oxidative phosphorylation for ATP production), with contributions to production and consumption by the various pathways and cell functions. One noticeable difference in the unevolved *Δpgi* strain is an increased contribution of the TCA cycle and oxidative phosphorylation to energy metabolism, leading to a slightly higher overall ATP yield. The ALE strains mostly reverted to normalized levels of total cofactor production and consumption that were very similar to the wild-type. Overall, the profile of energy metabolism is remarkably conserved in these strains. Perhaps the most significant difference between the wild-type and *Δpgi* strains in cofactor metabolism is in the reversed role of transhydrogenase (Fig. 10.3D).

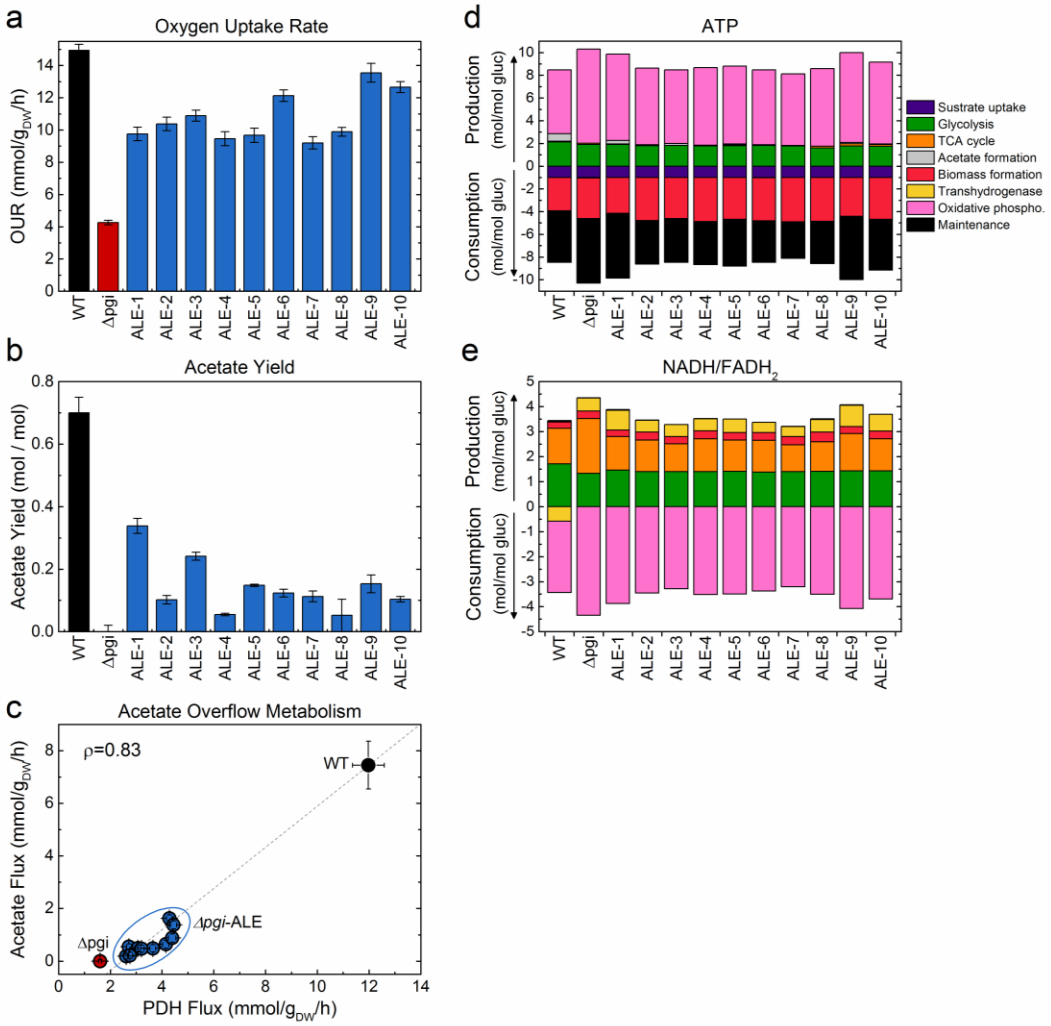

a
Oxygen Uptake Rate
OUR (mmol/g$_{DW}$/h)
b
Acetate Yield
Acetate Yield (mol / mol)
c
Acetate Overflow Metabolism
ρ=0.83
WT
Δpgi
Δpgi-ALE
Acetate Flux (mmol/g$_{DW}$/h)
PDH Flux (mmol/g$_{DW}$/h)
d
ATP
Production (mol/mol gluc)
Consumption (mol/mol gluc)
Sustrate uptake
Glycolysis
TCA cycle
Acetate formation
Biomass formation
Transhydrogenase
Oxidative phospho.
Maintenance
e
NADH/FADH$_2$
WT
Δpgi
ALE-1
ALE-2
ALE-3
ALE-4
ALE-5
ALE-6
ALE-7
ALE-8
ALE-9
ALE-10

Figure 10.3: Energy metabolism is mostly unaffected by adaptive evolution in *Δpgi*. Oxygen uptake rates (OUR) (a) and acetate yields (b) are shown for the wild-type, unevolved *Δpgi*, and ten *Δpgi*-ALE strains. The OUR recovers in some cases to near wild-type levels in ALE strains after being severely reduced in *Δpgi*. OUR was estimated by ^{13}C-MFA, and the error bars indicate standard deviations of the estimates. The acetate secretion phenotype also varies widely in these strains (b). The unevolved *Δpgi* secretes no acetate, compared to 0.7 mol/mol glucose in the wild-type. The ALE strains produce varying amounts of acetate, but all significantly less than the wild-type. The acetate overflow effect is represented in (c), which shows the absolute acetate production flux plotted against the absolute pyruvate dehydrogenase (PDH) flux in absolute units. Error bars represent 95% confidence intervals. The two fluxes strongly correlate (Pearson correlation coefficient of 0.83). Both the oxygen uptake and acetate secretion phenotypes directly impact cellular energy metabolism, summarized in more detail through ATP (d) and NADH/$FADH_2$ (e) balances. These both show the normalized (per glucose) contributions of various metabolic pathways to the production and consumption of these cofactors. Oxidative phosphorylation supports the majority of ATP generation while consuming NADH or $FADH_2$ and oxygen. Glycolysis contributes to ATP production through substrate-level phosphorylation. ATP is utilized in glucose uptake, biomass synthesis, and other cellular maintenance costs. NADH/$FADH_2$ are produced in glycolysis, the TCA cycle, and in the *Δpgi* strains, through the transhydrogenase. Overall, energy metabolism is not significantly altered following adaptive evolution.

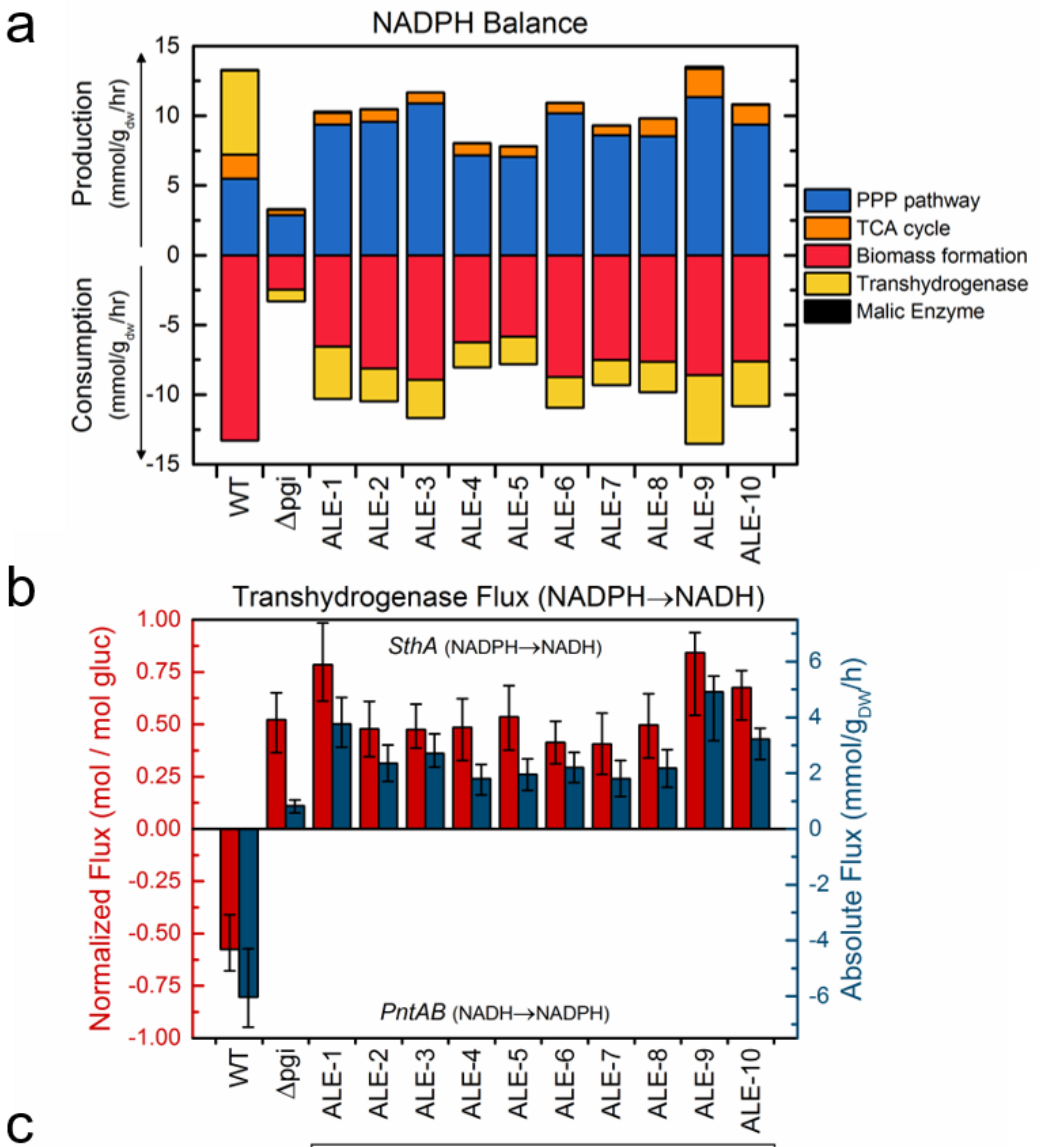
10.3.4 Transhydrogenase Genes are Mutated in Many, but Not All *Δpgi* ALE Strains

E. coli is able to interconvert reduced cofactors with two pyridine nucleotide transhydrogenases, the membrane-bound PntAB which primarily converts NADH to NADPH, and the soluble form SthA (also referred to as UdhA) which primarily converts NADPH to NADH (Sauer et al., 2004). In the wild-type, excess NADH produced in glycolysis and the TCA cycle is used to produce approximately half of the needed NADPH through the transhydrogenase (Figs. 10.3D, 10.4A, E.2). In *Δpgi*, the redirection of glucose flux in upper central carbon metabolism, away from glycolysis

and into the oxPPP, results in a significant excess of NADPH generation. Some of this is utilized by elevated biosynthesis flux (i.e., biomass yield) (Fig. E.1), but most of the imbalance must be corrected by a reversal of the transhydrogenase to convert NADPH to NADH.

The absolute rates of NADPH production and consumption are shown in Fig. 10.4A. Here again is shown the impact of the forcing of flux through the oxPPP, which generates a large excess of NADPH and necessitates the reversal of the transhydrogenase. This transhydrogenase flux is shown in Fig. 10.4B in both absolute and relative flux units, with the 95% confidence intervals calculated from ^{13}C-MFA. In the *Δpgi* strains (both unevolved and evolved), the normalized flux does not change significantly (also Fig. E.2), reflecting that other parts of central carbon metabolism were not rewired to relieve the cofactor imbalance. Instead, the absolute transhydrogenase fluxes increase significantly in the ALE strains. To gain insight into how these flux increases were achieved, the mutations directly related to the regions of the transhydrogenase genes in the ALE strains were compared to the flux changes (Fig. 10.4C). As noted above (Fig. 10.1), *pntA*, *pntB*, and *sthA* were some of the most frequently mutated genes in this study. Eight of the ten ALE strains had at least one transhydrogenase mutation, with five having two. Based on the nature of the mutations, they presumably increase SthA activity, reduce PntAB activity, or both. The *pntAB* mutations were diverse, including deletions and a duplication, as well as an upstream IS element insertion. Three of the four mutations in the coding regions likely result in truncated, nonfunctional proteins: ALE-2 has a nonsense mutation in *pntA*, and ALE-4 and ALE-7 have major truncations in *pntB* and *pntA*, respectively (Charusanti et al., 2010). The *sthA* mutations were SNP's, including a commonly

mutated site (5 strains) 64 bp upstream. The exact effect of the upstream mutations on transcriptional regulation is uncertain, but we hypothesize that they increase *sthA* and reduce *pntAB* expression. Despite the high frequency of these mutations, two strains (ALE-5 and ALE-6) achieved the increased flux with no observed mutations directly in the transhydrogenase genes, raising questions about other possible mechanisms for cofactor rebalancing. Both ALE-5 and ALE-6 possess mutations in genes which directly affect transcription levels (e.g., *lrp* and *rpoA*). In fact, *pntAB* has been identified as a regulatory target of Lrp (Cho et al., 2008).



c

		ALE-									
Gene	**Mutation**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
pntA ←	Δ11 bp (677-687/1533 nt)							✓			
pntA ←	S66* (TCA→TAA)		✓								
pntA ← / → ydgH	IS5 (+) +4 bp, intergenic (-75/-446)	✓									
pntB ←	6 bp duplication (682/1389 nt)										✓
pntB ←	Δ1 bp (430/1389 nt)				✓						
sthA ←	S317G (AGC→GGC)	✓									
sthA ←	G20S (GGC→AGC)								✓	✓	
sthA ← / → fabR	intergenic C→A (-64/-270)			✓							
sthA ← / → fabR	intergenic C→T (-64/-270)		✓					✓	✓	✓	

Figure 10.4: Reversal of transhydrogenase flux corresponds to genetic mutations in many, but not all *Δpgi* ALE strains. The function of the pyridine nucleotide transhydrogenases changes dramatically in *Δpgi* strains. The cofactors NADH and NADPH are produced and consumed in various metabolic pathways, and transhydrogenases can interconvert the reduced forms of each. In (a), the pathway contributions to the NADPH production and consumption are shown in absolute units. In the wild-type, excess NADH is used to produce approximately half of NADPH needed for biomass synthesis. In *Δpgi* strains, elevated oxPPP pathway flux creates an excess of NADPH, leading to a reversal of net transhydrogenase flux. This flux is shown in normalized and absolute units in (b). There are two transhydrogenases in *E. coli*: the soluble SthA primarily converts NADPH to NADH, and the membrane-bound PntAB primarily converts NADH to NADPH. The absolute, but not the normalized transhydrogenase flux increases significantly following adaptive evolution. This suggests that metabolism is not rewired elsewhere to relieve the imbalance, but instead the transhydrogenase activity in the needed direction increases. The transhydrogenase enzymes *pntA*, *pntB*, and *sthA* were frequently mutated in the ALE strains (c). Check marks reflect the presence of the described mutations in specific ALE strain. At least one transhydrogenase mutation occurred in 8 out of 10 strains.

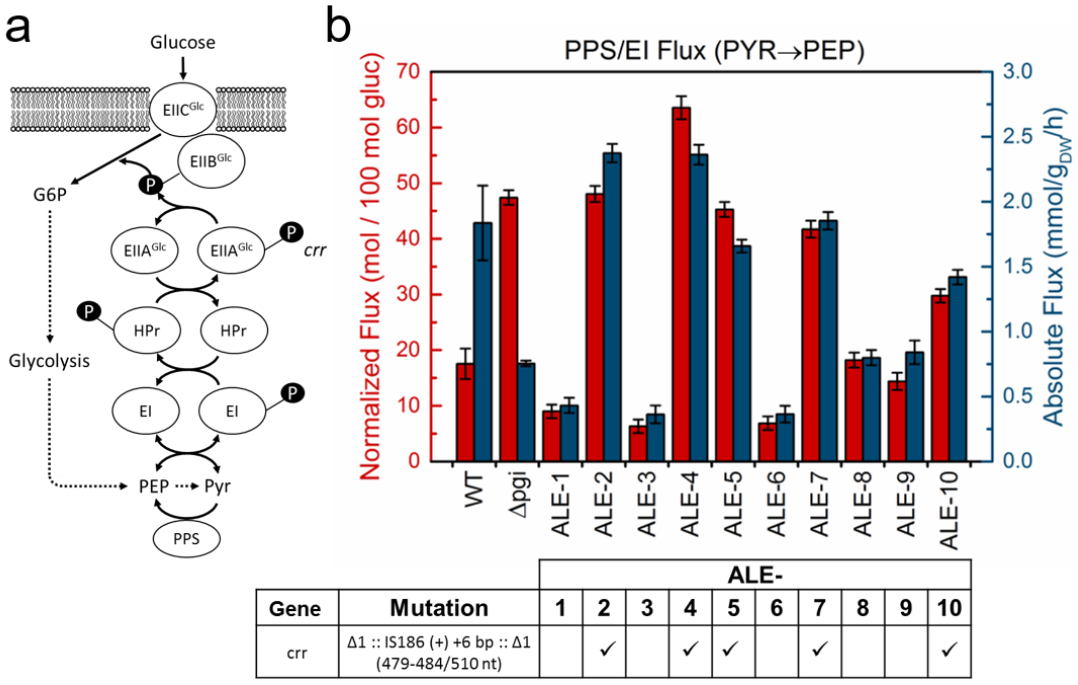
10.3.5 Mutations in PTS Component *crr* are Associated with Increased Back-Flux from Pyr to PEP

Another frequently occurring, and more unexpected, mutation was an IS element insertion in *crr* in 5 of 10 ALE strains (Fig. 5). *Crr* encodes $EIIA^{Glc}$, the cytosolic sub-unit of Enzyme II in the PTS glucose transport system. The PTS system is the primary mode of glucose uptake in *E. coli*, and links the uptake and phosphorylation of glucose (by EII) to the glycolytic conversion of PEP to pyruvate (Pyr) (by EI, linked by the intermediary phosphotransferase HPr) (Deutscher et al., 2014; Escalante et al., 2012) (Fig. 10.5A). Previous work has shown that Enzyme I (EI) of the PTS sugar transport system is reversible *in vivo* (Christopher P. Long et al.,

2017a), and that in the wild-type 10% of PEP is produced from Pyr in the reverse (i.e., gluconeogenic) direction by this mechanism. Previous work has also shown that in a *Δcrr* strain, this flux is increased by more than 2-fold due to perturbation of the PTS equilibrium as well as some possible activation of PPS, the gluconeogenic enzyme also able to facilitate the conversion of Pyr to PEP. Given the prevalence of the *crr* mutation and global metabolic perturbations in *Δpgi*, we hypothesized that the PPS/EI flux (Pyr to PEP) would be altered in these ALE strains.

The PPS/EI flux was measured using an [U-^{13}C]alanine tracer approach developed recently (Christopher P. Long et al., 2017a) (see Methods, Appendix Fig. E.3). As shown in Fig. 10.5B, the extent of this flux varied widely among the ALE strains, and its magnitude corresponded strongly with the presence of the *crr* mutation. Expressed as normalized flux (relative to 100 units of glucose uptake), the PPS/EI flux was significantly elevated in the unevolved *Δpgi* (from 18 in the wild-type to 47), and was reduced subsequently in strains lacking the *crr* mutation, but maintained at a high level in the strains with the mutated *crr*. In absolute terms, the flux was significantly reduced in all *Δpgi* strains except for those with the *crr* mutation, where the flux was more similar to the wild-type. Without further analysis of the activity of PPS in these strains, it is difficult to deduce the exact mechanism of these changes, but they are consistent with the previous observations in *Δcrr* (Christopher P. Long et al., 2017a). This result strongly supports a genetic-metabolic flux relationship between the *crr* IS element mutation and elevated PPS/EI flux. The high frequency of this identical mutation indicates a strong selective pressure for this mutation, but the exact mechanism of the fitness benefit is unclear. In addition to its direct role in glucose

uptake, *crr* is also involved in signaling the global regulator Crp (Deutscher et al., 2006), which controls the transcription of over 100 genes (Zheng et al., 2004).



		ALE-									
Gene	**Mutation**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
crr	Δ1 :: IS186 (+) +6 bp :: Δ1 (479-484/510 nt)		✓		✓	✓		✓			✓

Figure 10.5: Elevated back-flux from PYR to PEP corresponds with a frequently occurring mutation in *crr*. In (a), the PTS glucose transport system is shown. This cascade of phosphotransferases couples the uptake and phosphorylation of glucose to the glycolytic conversion of phosphoenolpyruvate (PEP) to pyruvate (Pyr). The terminal phosphotransferase Enzyme I (EI) has been previously shown to be reversible, and capable of converting Pyr to PEP. This reaction is also carried out by the gluconeogenic enzyme PPS. Another intermediate phosphotransferase, $EIIA^{Glc}$, is encoded by *crr* as shown. The flux from Pyr to phosphoenolpyruvate (i.e., in the gluconeogenic direction) was quantified using [U-^{13}C]alanine tracer experiments. The estimated flux is shown in (b) in both normalized and absolute flux units, with error bars reflecting standard deviations of the estimate. This flux was significantly elevated in *Δpgi* relative to the wild-type in normalized units, but decreased in absolute units. There was wide variability in the ALE strains, which strongly correlated with the presence of a specific and frequent insertion element mutation in *crr* in 5 of 10 strains. Strains containing this mutation, indicated above with check marks, had highly elevated PPS/EI fluxes. *Crr* is functionally linked to EI through their mutual involvement in the PTS sugar transport system, and its knockout has been previously shown to modulate EI reversibility.

10.4 Discussion

In this work, we have explored how recovery of growth rate in *Δpgi* is enabled by unique genetic mutations and significant metabolic rewiring. In this system, fitness recovery was driven by global transcriptional regulation (i.e., sigma factors and other RNA polymerase components) and relief of a rate-limiting step at the cofactor transhydrogenase. This led to increased absolute flux through the oxidative pentose phosphate pathway, and a corresponding recovery of faster glucose uptake and growth rates. The usage of latent pathways including the ED pathway, glyoxylate shunt, and PCK reaction, were shown to represent no increase in absolute flux relative to the wild-type, and in absolute terms did not appreciably change after adaptive evolution. The availability of these pathways, expressed at low levels in the wild-type, may offer flexibility when facing changing conditions. However, they do not appear to be 'activated' as part of a general stress response in this case. These insights and the data set presented here should help advance predictive metabolic modeling (Chowdhury et al., 2015; Khodayari et al., 2014; Lerman et al., 2012). Overall, these results add to our understanding of adaptive evolution by elucidating how challenges to specific cellular sub-systems, i.e., central carbon metabolism and glycolysis, are overcome. Future application of this approach to other significant, growth-limiting metabolic perturbations may similarly illuminate associated kinetic and regulatory limitations as well as reveal useful solutions to ameliorate them. Paired with the appropriate measurements (as shown here with fluxomics), such studies are likely to uncover detailed mechanisms of adaptation.

Comparing the mutations and phenotypes of wild-type and *Δpgi* derived ALE strains deepens our mechanistic understanding of adaptive evolution by providing insight into the context-dependent selective forces and impacts of various mutations.

For example, RNA polymerase components are commonly mutated in ALE, and in experiments started with wild-type, mutations to *rpoB* and *rpoC* dominate (Fig. 1) (Herring et al., 2006; LaCroix et al., 2015; Sandberg et al., 2017, 2016, 2014). These have been shown to contribute to a rebalancing of the proteome to promote growth (Utrilla et al., 2016), corresponding to broad and proportional intracellular metabolic flux increases (i.e., no changes in normalized flux distributions) (Christopher P. Long et al., 2017b). In contrast, ALE experiments started with *Δpgi* acquired frequent mutations in the stress-response associated sigma factor *rpoS*. This would seem to indicate the presence of a unique maladaptive *rpoS*-mediated stress response in *Δpgi*. However, *rpoB* mutations were also seen in 2 of 10 *Δpgi* ALE experiments here, but never co-occurring with *rpoS* mutations. Further work is needed to deconvolute the role of each in adapting the global transcriptome and whether the two mutations would have additive benefits in *Δpgi*. Across both the wild-type and *Δpgi* studies, mutations to more than two RNA polymerase subunits (*rpo*) genes are rarely observed, possibly pointing to overlapping mechanisms where the selected-for mutations depend on the state of the cell (in this case, perturbed (*Δpgi*) versus a wild-type state). Other mutations seen more when starting with a wild-type strain, such as in *pyrE/rph* are likely relevant based on the overall growth state (the evolved strains started with wild-type are significantly faster when evolved under the same conditions) thus they may never be selected for in a 'crippled' starting strain such as the *Δpgi* starting strain (Jensen, 1993). However, such mutations and a convergence of mutations may be seen if evolved for more extended lengths of time under the same conditions (Wiser et al., 2013).

Other unique mutations in *Δpgi* were in the transhydrogenases and in *crr*. With regards to the former, a recent report of an ALE study of an oxPPP mutant (i.e., reduced NADPH production) found a high frequency of mutations in *pntAB* which led to increased activity (Chou et al., 2015). Given the importance of cofactors to the formation of metabolic products, the reported mutations from these two studies may provide valuable new candidates for rationally manipulating transhydrogenase activity in metabolic engineering (Jan et al., 2013). Lastly, the mechanism of fitness enhancement of the *crr* IS element mutation identified here requires further study. One possibility is that the reduced glycolytic flux in *Δpgi* could be sensed (Kochanowski et al., 2013) by the PTS system, e.g., via perturbation of the PEP/Pyr concentration ratio, and result in feedback inhibition of glucose uptake. Another is that the accumulated G6P (Toya et al., 2010) could induce a maladaptive regulatory response via CRP activation by P~EIIB^{Glc}. In these scenarios, a *crr* mutation may help to decouple feedback inhibition or limit the harmful regulatory effect. Ultimately, it will be desirable to confirm the identities and interactions of causal mutations by reproducing them synthetically. Previous work (Canonaco et al., 2001; Charusanti et al., 2010) confirmed causality and epistasis for *sthA* and *rpoS* mutations, but did not fully recapitulate the observed growth phenotypes of the evolved clones themselves. This likely points to a complex landscape that will become more feasible to explore as high throughput genome engineering methods mature and many strains can be tested efficiently.

10.5 Author Contributions

CPL performed all cell culture and ^{13}C tracer experiments; CPL and JEG analyzed metabolic data; AMF performed whole-genome sequencing; AMF and BOP

analyzed genomic data; CPL and MRA designed the project and wrote the paper with help from all authors.

Chapter 11

FAST GROWTH PHENOTYPE OF *E. COLI* K-12 FROM ADAPTIVE LABORATORY EVOLUTION DOES NOT REQUIRE INTRACELLULAR FLUX REWIRING

Reprinted with permission from: Long CP, Gonzalez JE, Feist AM, Palsson BO, Antoniewicz MR (2017) Fast growth phenotype of *E. coli* K-12 from adaptive laboratory evolution does not require intracellular flux rewiring. Metab. Eng. 44, 100-107.

11.1 Introduction

Adaptive laboratory evolution (ALE) is a method in which microorganisms are continuously cultured in a controlled environment over many generations, allowing for fitness improvement through the accumulation of beneficial mutations. ALE has been applied to increasing chemical tolerance (Atsumi et al., 2010; Horinouchi et al., 2010; Mundhada et al., 2017; Reyes et al., 2012), rates of growth on diverse substrates (Cordova et al., 2016; Herring et al., 2006; Lee and Palsson, 2010; Sandberg et al., 2017), and gaining general insight into microbial responses to environmental or genetic perturbations (Charusanti et al., 2010; Fong and Palsson, 2004; Tenaillon et al., 2012). Following an ALE experiment, the resulting strains are sequenced to identify genetic mutations (Herring et al., 2006). The difficulty inherent in identifying causal mutations has led to the practice of performing multiple independent ALE experiments and using the frequency of mutations to guide analysis (LaCroix et al., 2015). Phenotypic characterization is then necessary to quantitatively describe the extent of the fitness improvement (e.g. increase in growth rate) and the associated

physiology. Detailed cellular characterizations involving omics techniques such as transcriptomics, proteomics, metabolomics, and fluxomics can then enable systems and pathway-level analysis of the phenotype and its enabling mechanisms. Ideally, such approaches would result in genotype-phenotype insights that improve our general scientific understanding of the cell system and inform future rational engineering efforts (Long and Antoniewicz, 2014a).

Exponential aerobic growth of *E. coli* K-12 MG1655 on glucose minimal media is arguably the most widely used combination of organism and condition in basic science and biotechnology (Janssen et al., 2005). ALE applied in this context, particularly serial passaging of batch cultures such that the exponential phase is maintained, selects for increased growth rate. Such efforts probe the limits of *E. coli* growth performance and allow for the study of fast growing phenotypes that might be useful in biotechnology. Previously, LaCroix et al. reported the phenotypes and transcriptional analysis of ten independent ALE experiments of *E. coli* MG1655 (LaCroix et al., 2015). They reported up to 1.6-fold increases in growth rate, and identified frequent causal mutations in *rpoB* and intergenic regions of *hns/tdk* and *pyrE/rph*. The *pyrE/rph* mutation ameliorates a well-characterized strain-specific defect in pyrimidine biosynthesis (Jensen, 1993), and the other two likely result in broad transcriptional changes as *rpoB* and *hns* are global regulators. Transcriptomic analyses revealed increases in expression of genes associated with protein production (amino acid metabolism, transcription, translation, folding), glucose transport, and glycolysis, and reductions in enzymes involved in the TCA cycle and glyoxylate shunt. Similarly, Sandberg et al. reported the results of six independent ALE experiments, also using *E. coli* MG1655 and aerobic exponential growth on glucose

minimal media (Sandberg et al., 2016). Here, ALE was carried out over 40 days, or approximately 1000 generations, through serial passaging such that stationary phase was avoided. The glucose was ^{13}C labeled to test the hypothesis that a subtle kinetic isotope effect may influence metabolism and the trajectory of ALE, but this was disproven in isotopic preference studies described previously (Sandberg et al., 2016). Furthermore, the final growth rates and most frequent mutations were very similar to those in (LaCroix et al., 2015), and the same key mutations (*rpoB*, *pyrE/rph*, and *hsns/tdk*) occurred with high frequency in the six ALE experiments.

The genetic and transcription level changes in these studies suggest the possibility of broad metabolic shifts in the adaptively evolved strains. However, to our knowledge, intracellular metabolic fluxes of such strains have not yet been reported. In this study, we applied high-resolution ^{13}C-metabolic flux analysis (^{13}C-MFA) (Maciek R. Antoniewicz, 2015a) to the six ALE strains previously described in (Sandberg et al., 2016) to determine whether their high growth rate is enabled by or associated with rewiring of central carbon metabolism. For additional context, the fluxes of the ALE strains were compared to the parent MG1655 strain, a related K-12 strain (BW25113), and a more distantly related *E. coli* strain BL21. Finally, flux balance analysis (FBA) was performed to compare the calculated optimal stoichiometric solution to the measured *in vivo* fluxes.

11.2 Materials and Methods

11.2.1 Materials

Chemicals and M9 minimal medium were purchased from Sigma-Aldrich (St. Louis, MO). Isotopic tracers were purchased from Cambridge Isotope Laboratories

(Tewksbury, MA): [1,2-^{13}C]glucose (99.7 %) and [1,6-^{13}C]glucose (99.2 % ^{13}C). The isotopic purity and enrichment of the tracers were validated by GC-MS analysis as described in (Sandberg et al., 2016) and (Cordova and Antoniewicz, 2016). All solutions were sterilized by filtration.

11.2.2 Strains and Growth Conditions

E. coli BL21(DE3) was obtained from Invitrogen (Cat. No. C600003). *E. coli* BW25113 was obtained from the Keio collection (GE Healthcare Dharmacon, Cat. No. OEC5042). The MG1655 wild-type and adaptively evolved (ALE) strains were previously described in (Sandberg et al., 2016). The wild-type was K-12 MG1655 (ATCC 700926). Six independent cultures were adaptively evolved in M9 minimal glucose medium for 40 days, corresponding to an average of 963 generations, or 2.82 x 10^{12} cumulative cell divisions (CCD). Passaging was frequent enough to avoid glucose depletion and initiation of stationary phase. In this study, all strains were cultured aerobically in glucose M9 minimal medium at 37°C in mini-bioreactors with 10 mL working volume as previously described (Long et al., 2016b). Pre-cultures were grown overnight and then used to inoculate the experimental culture at an OD_{600} of 0.01. For ^{13}C-MFA, glucose tracers were added at the beginning of the culture. Cells were harvested (1 mL samples) for GC-MS analysis at mid-exponential growth when OD_{600} was approximately 0.7. In all cases, parallel tracer experiments were performed using [1,2-^{13}C]glucose and [1,6-^{13}C]glucose. These tracers were previously determined to be optimal for high-resolution ^{13}C-MFA of *E. coli* (Crown et al., 2016b).

11.2.3 Analytical Methods

Cell growth was monitored by measuring the optical density at 600nm (OD600) using a spectrophotometer (Eppendorf BioPhotometer). The OD600 values were converted to cell dry weight concentrations using previously determined OD600-dry cell weight relationship for E. coli (1.0 OD600 = 0.32 gDW/L; molecular weight of dry biomass = 24.6 gDW/C-mol (Long et al., 2016b). After centrifugation of the samples, the supernatant was separated from the biomass pellet. Acetate concentrations in the supernatant were determined using an Agilent 1200 Series HPLC (Gonzalez et al., 2017). Glucose concentrations were determined using a YSI 2700 biochemistry analyzer (YSI, Yellow Springs, OH). Growth rate was calculated using linear regression of the natural logarithm of the OD600 and time, and biomass yield via regression of biomass dry weight and glucose concentration in the medium.

11.2.4 Gas Chromatography-Mass Spectrometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 µm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230°C, the MS quad temperature at 150°C, the interface temperature at 280°C, and the inlet temperature at 250°C. GC-MS analysis of *tert*-butyldimethylsilyl (TBDMS) derivatized proteinogenic amino acids was performed as described in (Long and Antoniewicz, 2014b). Labeling of glucose (derived from glycogen) and ribose (derived from RNA) were determined as described in (Long et al., 2016a; McConnell and Antoniewicz, 2016). In all cases, mass isotopomer distributions were obtained by integration (Antoniewicz et al.,

2007b) and corrected for natural isotope abundances (Fernandez et al., 1996). Measurement errors of 0.3% were assumed for all mass isotopomers (Antoniewicz et al., 2007b).

11.2.5 Metabolic Network Model and ^{13}C-Metabolic Flux Analysis

The metabolic network model used for 13C-MFA is provided in Appendix Table F.1. The model is based on the E. coli model described previously (Crown et al., 2015a; Gonzalez et al., 2017), which includes all major metabolic pathways of central carbon metabolism, lumped amino acid biosynthesis reactions, and a lumped biomass formation reaction. Updates to the model include: i) making the reactions between PEP and pyruvate (Christopher P. Long et al., 2017a), and between α-ketoglutarate and succinyl-CoA reversible; ii) allowing for deamination of serine to pyruvate; and iii) modeling atmospheric CO2 dilution of each parallel experiment independently (Leighty and Antoniewicz, 2012a).

^{13}C-MFA calculations were performed using the Metran software (Yoo et al., 2004), which is based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007a). Fluxes were estimated by minimizing the variance-weighted sum of squared residuals (SSR) between the measured and model predicted mass isotopomer distributions and acetate yield using non-linear least-squares regression. For integrated analysis of parallel labeling experiments, the data sets were fitted simultaneously to a single flux model as described previously (Leighty and Antoniewicz, 2013). Flux estimation was repeated 10 times starting with random initial values for all fluxes to find a global solution. At convergence, accurate 95% confidence intervals were computed for all estimated fluxes by evaluating the

sensitivity of the minimized SSR to flux variations. Precision of estimated fluxes was determined as follows :

Flux precision (stdev) = [(flux $_{\text{upper bound 95\%}}$) − (flux $_{\text{lower bound 95\%}}$)] / 4

To describe fractional labeling of metabolites, G-value parameters were included in ^{13}C-MFA. As described previously (Antoniewicz et al., 2007c), the G-value represents the fraction of a metabolite pool that is produced during the labeling experiment, while 1-G represents the fraction that is naturally labeled, i.e. from the inoculum. By default, one G-value parameter was included for each measured metabolite in each data set. Reversible reactions were modeled as separate forward and backward fluxes. Net and exchange fluxes were determined as follows: $v_{net} = v_f - v_b$; $v_{exch} = \min(v_f, v_b)$.

11.2.6 Goodness-of-Fit Analysis

To determine the goodness-of-fit, ^{13}C-MFA fitting results were subjected to a χ^2-statistical test. In short, assuming that the model is correct and data are without gross measurement errors, the minimized SSR is a stochastic variable with a χ^2-distribution (Antoniewicz et al., 2006). The number of degrees of freedom is equal to the number of fitted measurements *n* minus the number of estimated independent parameters *p*. The acceptable range of SSR values is between $\chi^2_{\alpha/2}(n-p)$ and $\chi^2_{1-\alpha/2}(n-p)$, where α is a certain chosen threshold value, for example 0.05 for 95% confidence interval.

11.2.7 Flux Balance Analysis (FBA) and Flux Variability Analysis (FVA)

For FBA and FVA calculations, the COBRA Toolbox 2.0 implemented in Matlab was used (Schellenberger et al., 2011). Gurobi was used for the linear solver

(http://www.gurobi.com/). The *E. coli* iAF1260 genome scale model was used (Feist et al., 2007) for all calculations. The measured glucose and oxygen uptake rate were used as constraints as described in the text. All additional import and export fluxes, as well as internal constraints, were identical to those contained in the iAF1260 model file "Ec_iAF1260_flux1.xml".

11.3 Results and Discussion

11.3.1 Growth and Physiology

In this study, three unevolved *E. coli* strains, i.e. BL21(DE3), BW25113, and MG1655, and six adaptively evolved MG1655 strains were investigated. The complete list of specific mutations observed in the six evolved strains, labeled ALE-1 through ALE-6, had been reported in (Sandberg et al., 2016). Briefly, the same key mutations previously identified as causal (LaCroix et al., 2015) were recapitulated in the evolved strains, particularly various insertion sequence mutations in the *hns*/*tdk* intergenic region (ALE-1, ALE-3, ALE-5, ALE-6), deletions in the *pyrE*/*rph* intergenic region of either 1 bp (ALE-2, ALE-5) or 82 bp (ALE-1, ALE-3, ALE-4) and SNP's in rpoB (ALE-1, ALE-2, ALE-3, ALE-4, ALE-5). ALE-6 was noticeably distinct from the other five strains, as it was lacking *pyrE*/*rph* and *rpoB* mutations, but instead had a unique *rpoC* mutation.

The growth rates, biomass and acetate yields, and glucose uptake rates for all nine strains investigated here (i.e. three unevolved and six evolved strains) are summarized in Figs. 11.1 and 11.2. The growth physiology of the BW25113 strain was previously described (Long et al., 2016b). Like MG1655, this strain is a K-12 derivative and thus is closely related, whereas BL21 is a more distantly related *E. coli*

strain. All wild-type strains had similar growth rates (0.63 to 0.68 h^{-1}) (Fig. 1A), and the ALE strains grew significantly faster (approximately 0.9 h^{-1}) as expected and previously reported (Sandberg et al., 2016). This represents a 28-38% increase in fitness (i.e. growth rate) under our experimental conditions. Most of the strains had similar acetate production phenotypes (Fig. 11.1B). The K-12 strains produced approximately 0.7 mol acetate per mol glucose, in good agreement with previous reports (Chen et al., 2011; Leighty and Antoniewicz, 2013; Rahman and Shimizu, 2008). The BL21 strain produced significantly less acetate (0.39 mol/mol), a phenotype which has also been well characterized previously (Monk et al., 2016; Waegeman et al., 2012, 2011). In fact, previous studies have reported even lower acetate yields of 0.2 mol/mol and less, which may indicate a relatively larger variability in BL21 strains compared to K-12 strains. Interestingly, the acetate phenotypes following adaptive evolution were mostly unchanged, with the most significant change being an increase in acetate yield in ALE-6 from 0.66 to 0.83 mol/mol.

The biomass yields were relatively consistent across all strains (0.41 to 0.44 g_{DW}/g_{glc}) (Fig. 11.1C). The directly measured yields are compared to those estimated by ^{13}C-MFA in Appendix Fig. F.2. While there is strong overall consistency, a larger difference is suggested between BL21 (0.45 g/g) and ALE-6 (0.38 g/g), as would be expected given the divergent acetate yield phenotypes. This suggests that increases in growth rate that were attained during adaptive evolution came not from increased carbon efficiency, but rather from increased overall metabolic rate. This was also reflected in the calculated glucose uptake rates (Fig. 11.2A), which increased from 8.5 $mmol/g_{DW}/h$ in the wild-type to up to 12.5 $mmol/g_{DW}/h$ in ALE-6. The oxygen uptake

rates (Fig. 11.2B) also increased significantly in the ALE strains relative to the wild-type, but reached levels similar or only slightly higher (17 mmol/g_{DW}/h) than the oxygen uptake rate of BL21 (15 mmol/g_{DW}/h). Notably, the strain with the unique genetic mutations, i.e. ALE-6, had the highest rates of glucose uptake and acetate excretion.

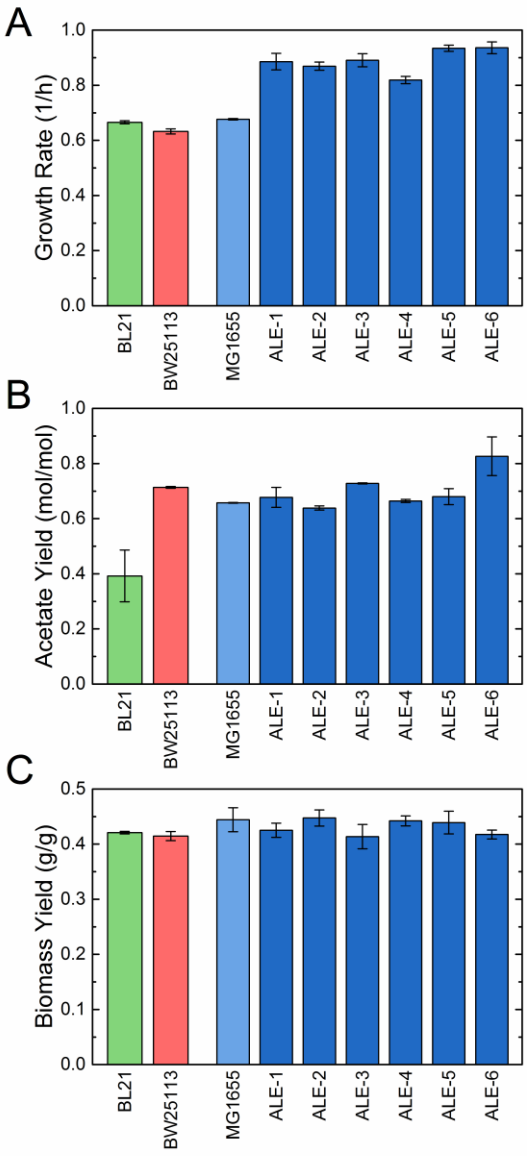


Figure 11.1: Physiology of all strains during exponential growth on glucose minimal medium. The three *E. coli* wild-type strains, BL21 (green), BW25113 (red), and MG1655 (blue) are shown along with the adaptively evolved MG1655 strains (ALE, dark blue). Growth rates (A) were measured in triplicate cultures, and acetate yields (B) on media HPLC measurements of duplicate cultures. Error bars indicate standard errors of the mean. Biomass yield (C) was based on regression of substrate and biomass measurements in a culture, with error bars reflecting the uncertainty in the parameter fitting.

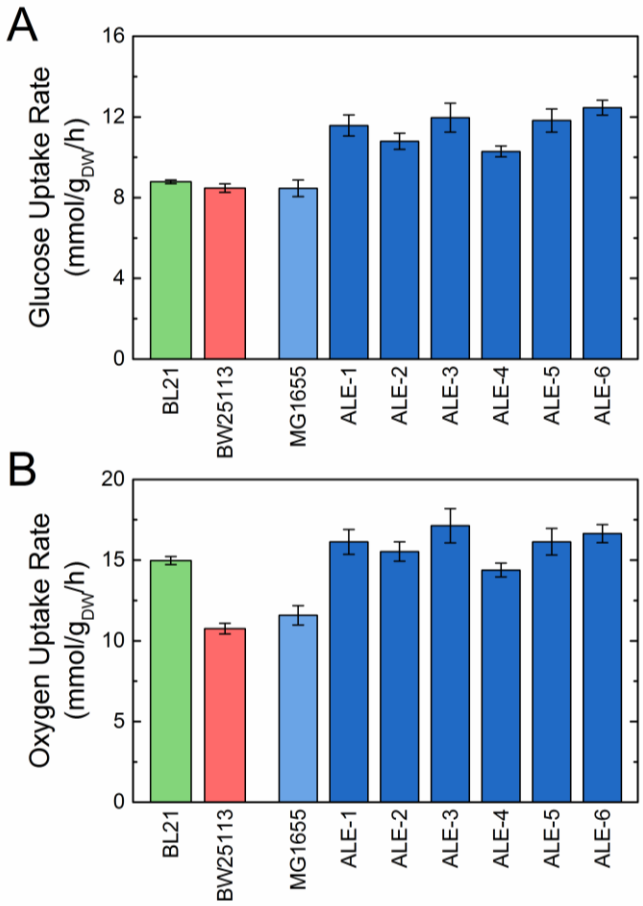


Figure 11.2: Substrate uptake rates of all strains during exponential growth on glucose minimal medium. The three *E. coli* wild-type strains, BL21 (green), BW25113 (red), and MG1655 (blue) are shown along with the adaptively evolved MG1655 strains (ALE, dark blue). Glucose uptake rates (A) were calculated from the growth rate and biomass yield (Fig. 11.1), and oxygen uptake rate (B) was estimated by ^{13}C-MFA. Error bars reflect standard errors.

11.3.2 ^{13}C-Metabolic Flux Analysis

To quantify the intracellular metabolic fluxes supporting the observed increases in growth and glucose uptake rates, high-resolution ^{13}C-MFA was performed. For each strain, two parallel labeling experiments were performed with

[1,2-^{13}C]glucose and [1,6-^{13}C]glucose, as this was previously identified to provide optimal precision in flux estimates throughout *E. coli* central carbon metabolism (Crown et al., 2016b). Labeling of proteinogenic amino acids, labeling of ribose moiety of RNA, and glucose moiety of glycogen (Long et al., 2016a) from each parallel experiment were fitted simultaneously, along with the measured acetate yield, to estimate fluxes. The estimated metabolic fluxes are provided in the Supplemental File. Statistically acceptable fits were achieved in all cases, assuming GC-MS measurement errors of 0.3% (Antoniewicz et al., 2007b).

The results of ^{13}C-MFA are summarized in Figs. 11.3 and 11.4. In Fig. 11.3, the distributions of fluxes through two key branch points in central carbon metabolism are shown for all strains. The first branch point (Fig. 11.3A) describes the split in upper central carbon metabolism between glycolysis (EMP pathway), the oxidative pentose phosphate pathway (oxPPP), and the Entner Doudoroff (ED) pathway. The relative usage of these three pathways was remarkably consistent among the nine strains studied here. The main route of glucose catabolism being the EMP pathway (74 to 78%), with almost all the rest going to the oxPPP. ED pathway usage was minimal, not exceeding 2% in any of these strains. There was a small increase in EMP usage in four of the six ALE strains relative to the parental strain. More variations between strains were observed in lower metabolism (Fig. 11.3B). The branch point here describes the fate of the lower glycolytic intermediate phosphoenolpyruvate (PEP), into anaplerosis (conversion to oxaloacetate via PPC), the TCA cycle via citrate synthase, or acetate production. The differences in the pathway usage here also reflect the differing acetate secretion phenotypes (Fig. 11.1B). Particularly, the BL21 strain has a much lower acetate flux (35% of PEP) and a correspondingly higher TCA cycle

flux (40%) than all K-12 strains (15% for BW25113 and 18% for MG1655). In the adaptively evolved strains, the relative pathway usage is once again remarkably unchanged, with the only significant difference being a higher acetate secretion and reduced TCA cycle flux in ALE-6. The constant relative intracellular pathway usage (i.e., fluxes normalized to these branch points or to glucose uptake rate) in the ALE strains corresponds to substantial and proportional increases in absolute flux throughout central carbon metabolism, as the glucose uptake rates are significantly elevated (Fig. 11.2A).

Detailed flux maps of central carbon metabolism are shown for three selected strains in Fig. 11.4. The wild-type BL21 and MG1655 are compared along with ALE-6, which was the fastest growing strain and exhibited the most unique phenotype of the evolved strains. The fluxes shown were normalized to 100 units of glucose uptake, with the growth and glucose uptake rates for each strain noted. As discussed above, there was a slight increase in normalized EMP flux from the wild-type to ALE-6, from 72% to 76% of glucose, at the expense of the oxPPP. The flux differences in lower metabolism, particularly the relative rate of TCA cycle and acetate secretion in these three strains, can also be seen here. No significant fluxes were observed in the ED pathway, glyoxylate shunt, malic enzyme, or PCK reactions in any of the studied strains. Given this, the two branch points described in Fig. 11.3 captured the main variations in the normalized intracellular fluxes of the strains. As noted above, ALE-6 was the most different from the wild-type, with elevated acetate yield and reduced TCA cycle usage. In absolute terms, however, the citrate synthase flux of ALE-6 (1.74 mmol/g_{DW}/h) was quite similar to that of the wild-type (1.67). The normalized fluxes of the other ALE strains were highly conserved from the parental strain.

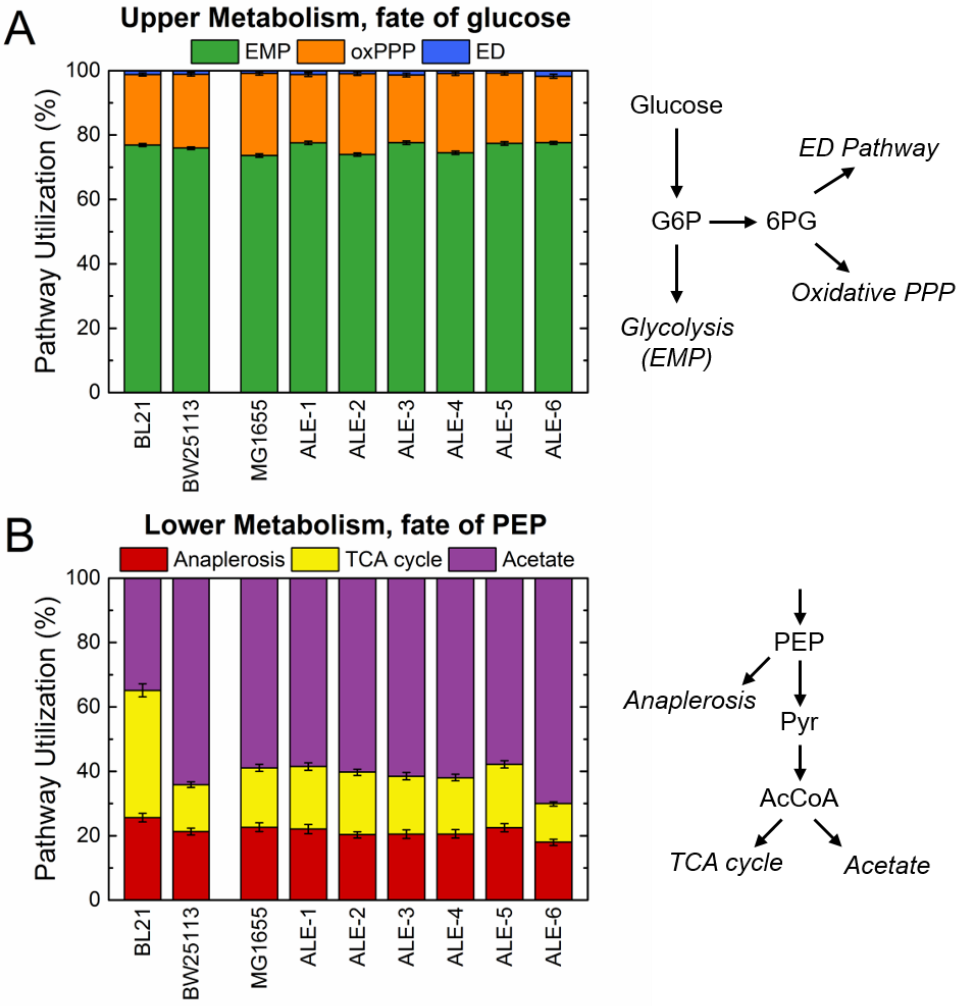


Figure 11.3: Key flux branch points for all strains as measured by ^{13}C-MFA. The upper flux branch point (A) reflects the fate of glucose into one of the EMP pathway (glycolysis), the oxidative pentose phosphate pathway (oxPPP) or the ED pathway. The lower branch point shown (B) reflects the fate of phosphoenolpyruvate (PEP) into anaplerosis, the TCA cycle via citrate synthase, or to acetate production. Error bars reflect the 95% confidence interval of the flux estimates.

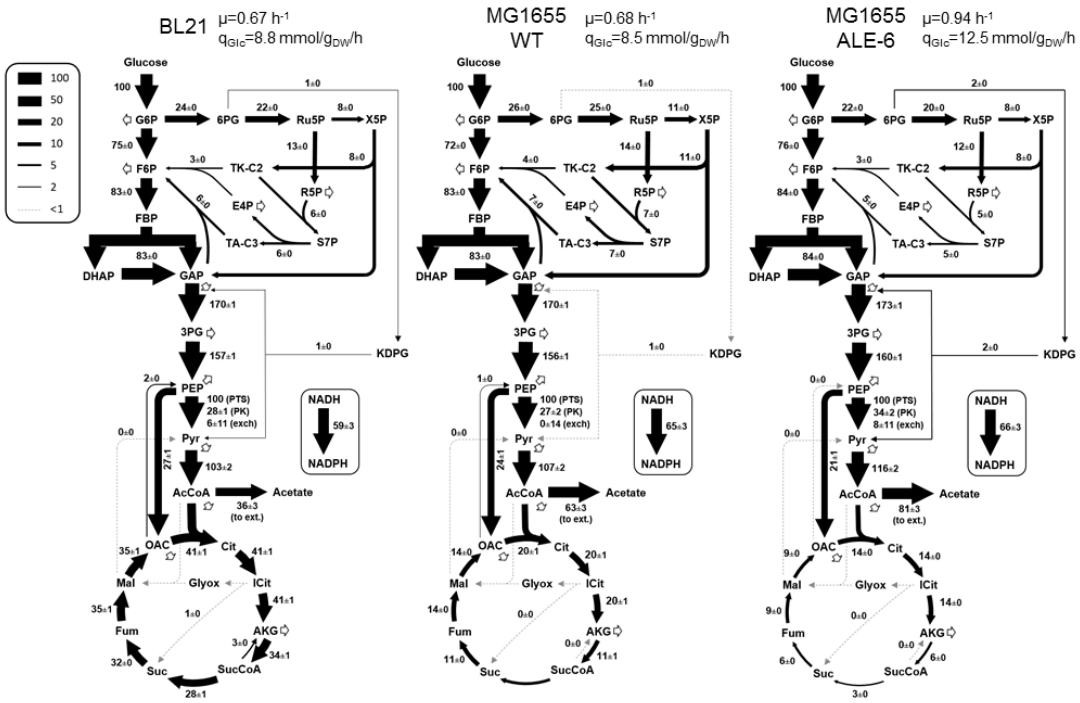


Figure 11.4: Complete flux maps of selected strains. All fluxes in central carbon metabolism are shown for the BL21 strain, as well as the wild-type MG1655 and one of its evolved descendants (ALE-6). Growth and glucose uptake rates for each strain are listed above, and all intracellular fluxes shown are then normalized to 100 units of glucose uptake. The indicated uncertainties for fluxes represent standard errors of the estimates.

11.3.3 Cofactor Metabolism

The measured intracellular fluxes can be used to calculate the contributions of individual pathways to the production or consumption of key cofactors in metabolism, including NADH and $FADH_2$ (Fig. 5A), NADPH (Fig. 5B), and ATP (Fig. 5C) (shown here in absolute units). This analysis highlights that NADH is roughly evenly produced in glycolysis and the TCA cycle (approximately 40% and 50%, respectively, for BL21, and the reverse for the K-12 strains), and mostly consumed by the electron transport chain as part of oxidative phosphorylation in all strains. 15-23% of NADH is

converted by transhydrogenase to NADPH, accounting for 41-55% of NADPH produced and supplementing the oxPPP and TCA cycle. The large TCA cycle flux in BL21 makes it a particularly significant source of NADPH in that strain (28%). All NADPH is utilized for biomass synthesis. Fig. 5C illustrates that ATP is mainly produced by oxidative phosphorylation (64-70%), with a smaller contribution from glycolysis (23-27%), and is consumed for biosynthesis and maintenance costs (under the category of "Other" in Fig. 5C). A constant P/O ratio of 2 was assumed for all strains to calculate ATP production, and the maintenance costs were estimated as the difference between total production and consumption for substrate uptake and growth. As a more conservative P/O ratio of 1.5 (Noguchi et al., 2004; Taymaz-Nikerel et al., 2010) would reduce the estimated ATP production and maintenance consumption rates somewhat, these results should be interpreted with caution. The inter-strain differences remain relevant though, as no mutations were observed in oxidative phosphorylation genes that would indicate a changing P/O ratio in the ALE strains. Given the conservation of normalized fluxes in the MG1655 strains, much of the variation in absolute cofactor rates is due to differences in glucose uptake and overall metabolic rate. In Appendix Fig. F.1, the cofactor balances are shown normalized to glucose uptake, where it is apparent that there are only very subtle differences in relative cofactor metabolism across the K-12 wild-types and ALE strains. The elevated TCA cycle of BL21 does contribute significantly more to cofactor production, and the increased use of oxidative phosphorylation results in a higher overall ATP yield.

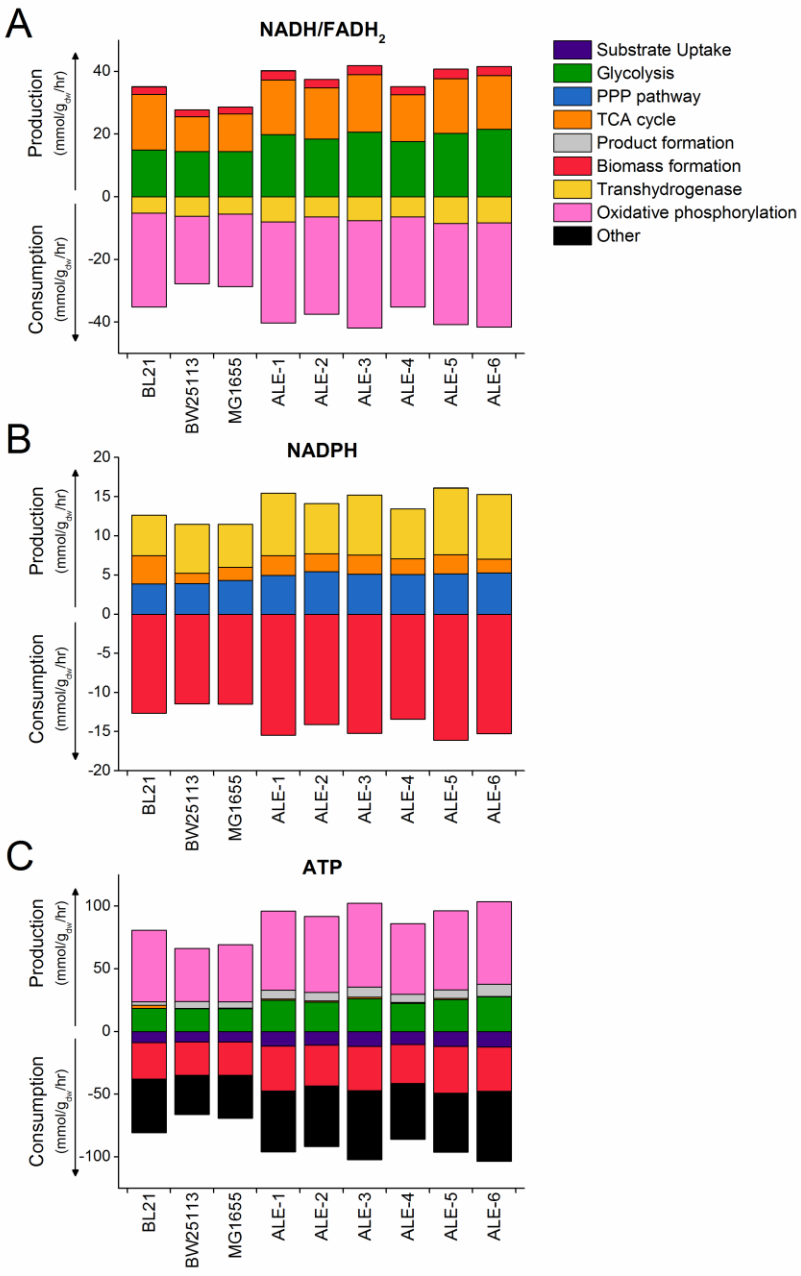


Figure 11.5: Quantitative cofactor balances. For each strain, the contributions of metabolic pathways to the production and consumption of cofactors are calculated in absolute units. Positive values reflect production of cofactor, and negative values reflect consumption. Shown are balances for NADH/$FADH_2$ (lumped) (A), NADPH (B), and ATP (C). "Other" in the ATP panel represents the estimated ATP maintenance cost (here, assuming P/O ratio=2.0).

11.3.4 Principal Component Analysis and Flux Balance Analysis

To further assess the degree of similarity between the metabolic flux profiles of the investigated strains, principal component analysis (PCA) was performed using nine key normalized fluxes in central carbon metabolism. PCA is a data reduction technique in which large multivariate data sets can be described in a new lower dimensional space in terms of principal components, which are linear combinations of the original variables. These principal components capture the maximum amount of original variation in the data. The results of PCA are shown in Fig. 11.6. The fluxes used for the analysis were selected from various representative intracellular and extracellular pathways, and excluded those with minimal flux (e.g. the ED pathway or glyoxylate shunt) where the inter-strain variance was not meaningful. The first two principal components capture nearly all (95%) of the flux variation, with PC1 (52%) reflecting the lower metabolism split between acetate production and the TCA cycle, and PC2 (43%) the upper split between glycolysis and oxPPP. This confirms our analysis above in Fig. 11.3, suggesting that these were the major areas of flux variance among the strains. All the K-12 strains, including BW25113, the MG1655 wild-type and ALE strains, clustered together in the PCA plot. ALE-6 was positioned at the extreme end of the group with a high PC1 value, reflecting its particularly low TCA cycle and high acetate flux, while BL21 was positioned at the opposite end with a low PC1 value.

We were also interested to compare the flux phenotypes of the adaptively evolved strains to an ‘optimal’ flux distribution as predicted by flux balance analysis (FBA), a widely-used tool in metabolic engineering. FBA finds a set of fluxes that optimizes the maximal growth rate given substrate uptake constraints, the genome scale network stoichiometry, and a biomass growth equation. This is commonly

justified as reflecting the selection pressure and result of evolution in laboratory strains (Edwards and Palsson, 2000; García Sánchez and Torres Sáez, 2014; Segre et al., 2002). FBA calculations were performed using the *E. coli* iAF1260 genome scale model (Feist et al., 2007) with two sets of glucose and oxygen uptake constraints (all others were set to simulate the glucose minimal media environment): one corresponding to the MG1655 wild-type (q_{glc}=8.5 mmol/g_{DW}/h, q_{O2}=12 mmol/g_{DW}/h), and one corresponding to the ALE strains (q_{glc}=12, q_{O2}=17). The predicted growth rates from FBA agreed well with the measured growth rates (i.e., unevolved predicted 0.63 h^{-1}, and evolved predicted 0.92 h^{-1}) and acetate yields (unevolved 0.7 mol/mol, evolved 0.63 mol/mol). The corresponding normalized flux predictions from FBA were included in the PCA plot (Fig. 11.6). There were some notable disagreements between the predicted and measured metabolic fluxes, especially in upper metabolism, where FBA predicted in both cases (unevolved and evolved) that approximately half of glucose flux was catabolized through oxPPP (54-55%). As discussed previously and shown in Figs. 3 and 4, this flux was measured by ^{13}C-MFA as 21-25% of glucose flux in all strains studied here. Since the measured and FBA-predicted growth rates agreed well, it appears that the optimal growth rate is not strongly affected by upper pathway usage and alternate optimal solutions may enable the observed growth rates.

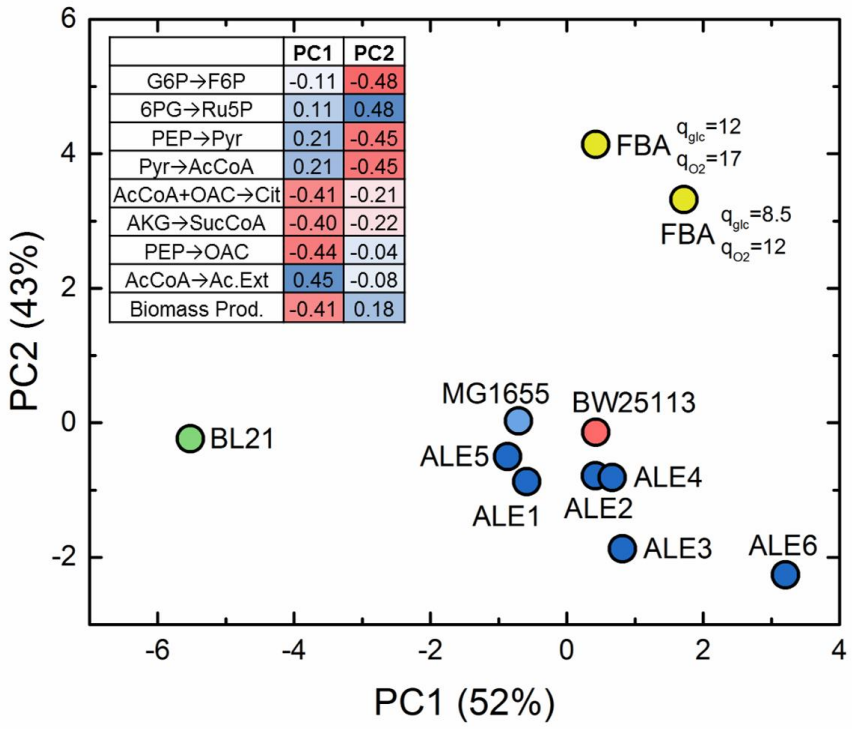


Figure 11.6: Principal component analysis of key normalized intracellular metabolic fluxes in measured strains and two flux balance analysis (FBA) simulations. The simulations were based on two different sets of substrate uptake constraints as noted. The coefficients of the top two principal components are shown in the table.

This hypothesis of alternate optimal solutions was confirmed with flux variability analysis (FVA) (Mahadevan and Schilling, 2003), which calculates a range of possible flux values that can support a given rate of growth. FVA was applied to both sets of glucose and oxygen uptake rate constraints, yielding very similar normalized flux variabilities in each case. The results described below are from the higher uptake rate case, corresponding to the ALE phenotype. Focusing again on the oxPPP flux, stepping down slightly to 99% of the optimal growth rate, the oxPPP flux varied from 34%-67% of glucose uptake. Stepping further, this range increased to 14-

74% of glucose uptake for 98% of optimal growth rate. Interestingly, in looking at another major central metabolism pathway, the branch point between the TCA cycle and acetate production, was more constrained with citrate synthase (TCA) fluxes varying only from 8-19% at 99% optimal growth, and 8-27% at 98% optimal growth. Thus, this analysis demonstrates that alternate optimal flux distributions can support rapid growth in the *E. coli* network and some pathways have more flexibility than others.

11.4 Conclusions

In this work, intracellular fluxes of *E. coli* subject to adaptive laboratory evolution were analyzed for the first time using ^{13}C-MFA. Given the numerous genetic mutations (Sandberg et al., 2016) and previously reported transcriptional changes in evolved strains (LaCroix et al., 2015), it was expected that significant intracellular metabolic rewiring would be occurring in these strains. Instead, we show here that normalized intracellular metabolic fluxes change very little in six independently evolved MG1655 strains. In absolute terms, intracellular fluxes increased proportionally and substantially, along with the glucose uptake rate, to support faster growth. The one significant change, a 26% increase in acetate yield in ALE-6, corresponded to a unique set of mutations. Interestingly, it was previously reported that in similarly evolved strains, enzymes involved in the TCA cycle were broadly transcriptionally repressed (LaCroix et al., 2015), but this did not correspond to reductions in normalized (5 out of 6 strains) or absolute (6 out of 6) TCA flux in the strains analyzed here. Future studies may explore whether some TCA cycle enzymes (or others in central metabolism) have excess flux capacity in the wild-type, allowing for increases in absolute flux under certain conditions without commensurate increases

in expression.. Overall, it was found that the magnitude of the differences between wild-type *E. coli* strains, particularly between BL21 and the K-12 strains (MG1655 and BW25113), exceeded the variation in unevolved and adaptively evolved MG1655 strains. The broad similarities, but notable differences between *E. coli* strains, should further inform analyses of cell metabolism rigidity across species (Tang et al., 2009; Wu et al., 2016).

Principal component analysis of the differences in normalized intracellular fluxes highlighted the similarity between all K-12 strains and the uniqueness of BL21 strain. It also showed that the particular solution found using FBA optimization of the growth rate function predicted a high oxPPP flux, differing significantly from that measured here using ^{13}C-MFA. This apparently reflects alternate optima or near-optima, which was further supported through an FVA approach. While growth rate optimization is a commonly used objective function, and reflects the selective pressure in the ALE experiment, alternative objective functions could also be explored (García Sánchez and Torres Sáez, 2014; Schuetz et al., 2007). Overall, in the case of K-12 MG1655, ^{13}C-MFA demonstrated there was no strong selective pressure to change fluxes from the starting flux distribution. As the research community accumulates more results for adaptively evolved strains with different initial metabolic phenotypes (e.g. different wild-type or gene knockout strains (Fong et al., 2006)), it may become possible to elucidate the path dependency of the evolved 'optima' achieved through ALE. For example, it would be interesting to determine whether the high TCA flux of BL21 strain is reduced upon evolution or is maintained during fast growth.

The knowledge that faster growth of adaptively evolved strains was not enabled by any particular change in metabolic pathway usage adds some clarity to the

picture previously presented via genetic and transcriptomic analysis (LaCroix et al., 2015). Along with increases in protein producing machinery, i.e. transcription and translation, changes to expression levels of central carbon metabolic enzymes were also reported. These included increases in PTS glucose transporters, enzymes in glycolysis and acetate production, and decreases in TCA cycle and glyoxylate shunt enzymes. The fact that the overall state of the metabolic network remained the same, despite adjustment in expression of metabolic enzymes, may indicate that the adaptive evolution responses are a matter of proteomic allocation rather than optimization of cellular processes. For example, the *rpoB* mutation has been shown to affect the balance between growth and stress functions (Utrilla et al., 2016). Moving forward, combined multi-omics analysis of ALE strains will be useful in advancing cellular modeling of kinetics (Khodayari et al., 2014) and physical and macromolecular constraints on phenotype (O'Brien et al., 2013). Identifying a growth-optimal proteome and regulatory mechanisms by which it can be achieved will be useful in engineering efficient strains. Understanding the hard constraints of *E. coli* performance, which in addition to the proteome can also include membrane space limitations (Liu et al., 2014), may motivate the development of alternative high-performance organisms for future applications (Cordova et al., 2015; Lee et al., 2016).

11.5 Author Contributions

CPL performed all cell culture and ^{13}C tracer experiments; CPL and JEG analyzed metabolic data; AMF performed whole-genome sequencing; AMF and BOP analyzed genomic data; CPL and MRA designed the project and wrote the paper with help from all authors.

Chapter 12

METABOLISM OF THE FAST-GROWING BACTERIUM *VIBRIO NATRIEGENS* ELUCIDATED BY ^{13}C METABOLIC FLUX ANALYSIS

Reprinted with permission from: Long CP, Gonzalez JE, Cipolla RM, Antoniewicz MR (2017) Metabolism of the fast growing bacterium *Vibrio natriegens* elucidated by ^{13}C metabolic flux analysis. Metab. Eng. 44, 191-197.

12.1 Introduction

Escherichia coli is the most widely studied microorganism in academia (Janssen et al., 2005; Long and Antoniewicz, 2014a). A wealth of knowledge has been generated over the past century for this model microbe, and many molecular tools have been developed for genetic engineering (Datsenko and Wanner, 2000; Gibson et al., 2009; Li et al., 2015; Wang et al., 2009). As a result, *E. coli* is often the go-to organism for metabolic engineering efforts. However, researchers are increasingly interested in selecting alternative hosts for various applications in biotechnology. One of the key physiological characteristics that impacts industrial performance is growth rate, or perhaps more importantly, biomass specific substrate uptake rate. This rate determines the maximum productivity that can be achieved for a given size bioreactor. As such, faster growing organisms offer a clear advantage over slower growing organisms. In this respect, wild-type *E. coli* has a relatively high growth rate of about 0.7 h^{-1} (60 min doubling time) when grown in minimal medium with glucose as the carbon source, and ~30 min doubling time when grown in rich medium. Faster growing *E. coli* strains have also been generated through adaptive laboratory evolution with a maximum growth rate of about 1.0 h^{-1} (40 min doubling time) in glucose

minimal medium (LaCroix et al., 2015; Sandberg et al., 2016); however, for yet unknown reasons this appears to be the upper limit for *E. coli* growth rate.

To achieve even faster conversion rates, scientists are now interested in identifying alternative fast-growing organisms to replace *E. coli* as the workhorse host (Lee et al., 2016; Weinstock et al., 2016). One such organism is the Gram-negative, non-pathogenic marine bacterium *Vibrio natriegens*, which is commonly found in marine and coastal waters and sediments, and has a reported doubling time of 10 minutes or less when cultured under ideal conditions in rich medium (Eagon, 1962). *V. natriegens* (initially *Pseudomonas natriegens*) was described for the first time by Payne et al. in the early 1960s (Payne et al., 1961). It is a moderate halophile, requiring about 1.5% NaCl for optimal growth, and grows well under laboratory conditions with glucose as the only carbon source, with an optimal growth temperature of 37 °C (Eagon, 1962; Lee et al., 2016). Two annotated genomes are available for *V. natriegens*, for strains ATCC 12048 (Wang et al., 2013) and DSMZ 759 (Maida et al., 2013); and recently, a wide range of genetic tools were developed and described to engineer *V. natriegens* (Weinstock et al., 2016). Additionally, the fast growth of this species was recently shown to reduce the time needed to execute common cloning pipelines, which has clear advantages in highly iterative strain building efforts (Weinstock et al., 2016).

In this work, we have investigated the metabolism of *V. natriegens* using state-of-the-art tools for ^{13}C metabolic flux analysis (^{13}C-MFA) (Gonzalez and Antoniewicz, 2017). Currently, no information is available about intracellular metabolism of *V. natriegens* and this lack of knowledge significantly impacts our ability to apply rational strategies to engineer this organism. For example, in the

absence of an experimentally validated flux map, constraint-based modeling and analysis (COBRA) approaches cannot be applied to guide new metabolic engineering designs (Becker et al., 2007; Schellenberger et al., 2011). To address this critical gap in knowledge, here we have characterized the growth physiology, constructed a detailed model core metabolism, measured biomass composition, and performed high-resolution ^{13}C-MFA to estimate intracellular fluxes in *V. natriegens.* Analysis of co-factor balances provide additional insights into its metabolism. Taken together, the results presented in this study provide an important quantitative description of the physiology and metabolism of *V. natriegens* that can serve as the basis for informed host selection, future model building, and strain design efforts.

12.2 Materials and Methods

12.2.1 Materials

Media and chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Tracers were purchased from Cambridge Isotope Laboratories, [1,2-^{13}C]glucose (99.5 atom% ^{13}C) and [1,6-^{13}C]glucose (99.5%). The isotopic purity and enrichment of glucose tracers was determined by GC-MS analysis (Cordova et al., 2016). Wolfe's minerals (Cat. No. MD-TMS) and Wolfe's vitamins (Cat. No. MD-VS) were purchased from ATCC (Manassas, VA). The growth medium for *V. natriegens* was M9 minimal medium supplemented with (per liter of medium): 15 g of NaCl, 10 mL of Wolfe's minerals, 10 mL of Wolfe's vitamins, and 0.05 g of yeast extract. Glucose was added as indicated in the text. All media and stock solutions were sterilized by filtration.

12.2.2 Strains and Growth Conditions

V. natriegens (ATCC Cat. No. 14048, Manassas, VA) was used in this study. For tracer experiments, cells from frozen stock were first pre-cultured overnight at 37 °C in a shaker flask with 20 mM initial glucose. Next, 1 mL of this culture was washed with fresh medium and used to inoculate a new shaker flask with 10 mL of medium (20 mM initial glucose). After 1 hr, 100 uL of this culture was used to inoculate two mini-bioreactors, one containing 10 mM of [1,2-^{13}C]glucose and one containing 10 mL of [1,6-^{13}C]glucose. The optical density (OD_{600}) of the inoculated cultures was about 0.015. We estimated that about 0.2 mM of unlabeled glucose was carried over from the inoculum to the cultures. Cells were then grown at 37°C in the mini-bioreactors as described before (Gonzalez et al., 2017). Air was sparged into the liquid at a rate of 12 mL/min to provide oxygen and to ensure sufficient mixing of the culture by the rising gas bubbles. Cell pellets were collected for GC-MS analysis of isotopic labeling during the mid-exponential growth phase when biomass concentration (OD_{600}) was between 0.7 and 1.0.

12.2.3 Analytical Methods

Samples were collected at multiple times during the growth phase to monitor cell growth, substrate uptake and acetate accumulation. Cell growth was monitored by measuring the optical density at 600nm (OD600) using a spectrophotometer (Eppendorf BioPhotometer). The OD600 values were converted to cell dry weight concentrations using the following relationship for V. natriegens: 1.0 OD600 = 0.27 gDW/L, which was experimentally determined as described in (Long et al., 2016b). Acetate concentrations were determined using an Agilent 1200 Series HPLC

(Whitaker et al., 2017), and glucose and lactate concentrations were determined using a YSI 2700 biochemistry analyzer (YSI, Yellow Springs, OH).

12.2.4 Biomass Composition Analysis

The methods used for quantifying biomass composition were described in (Long and Antoniewicz, 2014b). Briefly, samples were prepared by three respective methods: hydrolysis of protein and subsequent TBDMS derivatization of amino acids; hydrolysis of RNA and glycogen and subsequent aldonitrile propionate derivatization of sugars (ribose and glucose, respectively); and fatty acid methyl ester derivatization for analysis of fatty acid. In total, 17 amino acids were quantified. Glutamine and asparagine are deaminated during hydrolysis to glutamate and aspartate, respectively; thus, we report the combined pools of each. The amino acids arginine, cysteine and tryptophan are degraded during hydrolysis and thus were not quantified. Quantification of all components was achieved by isotope ratio analysis using an isotopically labeled standard and a naturally labeled biomass sample. In this study, the standard was generated by growing wild-type *E. coli* on [U-^{13}C]glucose and aliquoting identical (1 mL of an OD_{600} = 1.0) samples of this "fully labeled" biomass. These were centrifuged and washed twice with M9 medium. The composition of the fully labeled biomass was determined using unlabeled chemical standards as described in (Long and Antoniewicz, 2014b).

12.2.5 Gas Chromatography-Mass Spectometry

GC-MS analysis was performed on an Agilent 7890B GC system equipped with a DB-5MS capillary column (30 m, 0.25 mm i.d., 0.25 µm-phase thickness; Agilent J&W Scientific), connected to an Agilent 5977A Mass Spectrometer operating

under ionization by electron impact (EI) at 70 eV. Helium flow was maintained at 1 mL/min. The source temperature was maintained at 230°C, the MS quad temperature at 150°C, the interface temperature at 280°C, and the inlet temperature at 250°C. GC-MS analysis of *tert*-butyldimethylsilyl (TBDMS) derivatized proteinogenic amino acids was performed as described in (Antoniewicz et al., 2007b). Labeling of glucose in the medium was determined after aldonitrile propionate derivatization as described in (Antoniewicz et al., 2011; Sandberg et al., 2016). Labeling of fatty acids was determined after derivatization to fatty acid methyl esters (FAME) (Crown et al., 2015b). Labeling of glucose (derived from glycogen) and ribose (derived from RNA) were determined as described in (Long et al., 2016a; McConnell and Antoniewicz, 2016). In all cases, mass isotopomer distributions were obtained by integration (Antoniewicz et al., 2007b) and corrected for natural isotope abundances (Fernandez et al., 1996).

12.2.6 Metabolic Network Analysis and ^{13}C-Metabolic Flux Analysis

A metabolic network model of *V. natriegens* core metabolism was constructed for ^{13}C-MFA based on the reactions annotated in KEGG and BioCyc databases (Caspi et al., 2012; Kanehisa et al., 2012; Kanehisa and Goto, 2000). The complete ^{13}C-MFA model is provided in Appendix Table G.1. The model includes all major metabolic pathways of central carbon metabolism, including glycolysis, pentose phosphate pathway, Entner–Doudoroff pathway, TCA cycle, glyoxylate shunt, and various anaplerotic and cataplerotic reactions, lumped amino acid biosynthesis reactions, and a lumped biomass formation reaction, which was derived using the measured biomass composition for *V. natriegens*. Since it is not possible to distinguish between the fluxes of NAD-dependent malic enzyme (EC 1.1.1.38), NADP-dependent malic

enzyme (EC 1.1.1.40), and oxaloacetate decarboxylase (EC 4.1.1.3) using ^{13}C-MFA (Kappelmann et al., 2015), we included only one reaction in the model (the NADP-dependent malic enzyme) to describe the combined flux of all three reactions. The model also accounts for the exchange of intracellular and atmospheric unlabeled CO_2 (Leighty and Antoniewicz, 2012a), and G-value parameters to describe fractional labeling of amino acids. As described previously (Antoniewicz et al., 2007c), the G-value represents the fraction of a metabolite produced from labeled glucose, while 1-G represents the fraction that is naturally labeled, e.g. from to the inoculum. By default, one G-value parameter was included for each measured amino acid in each data set. Reversible reactions were modeled as separate forward and backward fluxes. Net and exchange fluxes were calculated as follows: $v_{net} = v_f - v_b$; $v_{exch} = \min(v_f, v_b)$.

All ^{13}C-MFA calculations were performed using the Metran software (Yoo et al., 2008) which is based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007a). Fluxes were estimated by minimizing the variance-weighted sum of squared residuals (SSR) between the experimentally measured and model predicted mass isotopomer distributions of biomass amino acids, glucose derived from glycogen, ribose derived from RNA, and the measured acetate yield, using non-linear least-squares regression (Antoniewicz et al., 2006). For integrated analysis of parallel labeling experiments, the data sets were fitted simultaneously to a single flux model (Maciek R Antoniewicz, 2015). Flux estimation was repeated at least 10 times starting with random initial values for all fluxes to find a global solution. At convergence, accurate 95% confidence intervals were computed for all estimated fluxes by evaluating the sensitivity of the minimized SSR to flux variations. Precision of estimated fluxes was determined as follows (Antoniewicz et al., 2006):

Flux precision (stdev) = [(flux $_{\text{upper bound 95\%}}$) − (flux $_{\text{lower bound 95\%}}$)] / 4

12.2.7 Goodness-of-Fit Analysis

To determine the goodness-of-fit, ^{13}C-MFA fitting results were subjected to a χ^2-statistical test. In short, assuming that the model is correct and data are without gross measurement errors, the minimized SSR is a stochastic variable with a χ^2-distribution (Antoniewicz et al., 2006). The number of degrees of freedom is equal to the number of fitted measurements *n* minus the number of estimated independent parameters *p*. The acceptable range of SSR values is between $\chi^2_{\alpha/2}(n\text{-}p)$ and $\chi^2_{1\text{-}\alpha/2}(n\text{-}p)$, where α is a certain chosen threshold value, for example 0.05 for 95% confidence interval.

12.3 Results and Discussion

12.3.1 Growth Physiology

Growth characteristics of *V. natriegens* were determined in aerobic batch culture at 37 °C in medium containing 1.8 g/L of glucose as the main carbon source. The growth medium was M9 minimal medium supplemented with 1.5% NaCl, vitamins, minerals, and 0.05 g/L of yeast extract. The addition of small amount of yeast extract eliminated a short lag phase that was observed when cells were subcultured without it. The yeast extract did not impact glucose uptake rate or the specific growth rate. During the exponential growth phase, the specific growth rate of *V. natriegens* was 1.70 ± 0.02 h^{-1}, which corresponds to a doubling time of about 24 min (Fig 12.1). The biomass yield was 0.44 g_{DW}/g and the biomass-specific glucose uptake rate was 21.4 ± 1.3 mmol/g_{DW}/h. *V. natriegens* produced acetate as a byproduct

during exponential growth, with about 0.8 mol of acetate produced per mol of glucose consumed (Table 12.1). No other byproducts were detected by HPLC analysis.

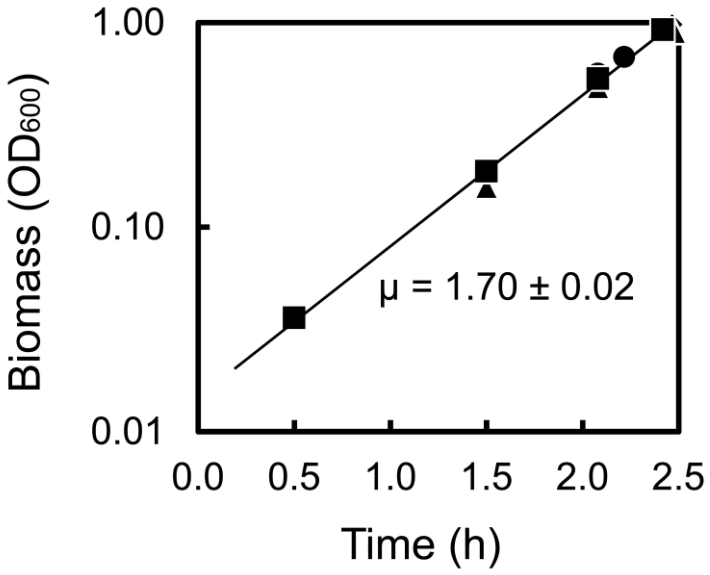


Figure 12.1: The growth rate of *V. natriegens*, as measured in three replicate cultures.

Table 12.1: Physiological characteristics of *V. natriegens* grown in aerobic batch culture on glucose minimal medium at 37°C (mean ± sem, $n = 4$ biological replicates).

Growth rate	1.70 ± 0.02	1/h
Biomass yield	0.44 ± 0.03	g_{DW}/g
Acetate yield	0.8 ± 0.1	mol/mol
Glucose uptake rate	21.4 ± 1.3	$mmol/g_{DW}/h$

12.3.2 Metabolic Model Construction and Biomass Composition Analysis

To facilitate quantitative studies of *V. natriegens*, a detailed network model of core metabolism was constructed for ^{13}C-MFA based on reactions annotated in the KEGG and BioCyc databases. As illustrated in Fig 12.2, central carbon metabolism of

V. natriegens is similar to that of *E. coli* and includes the following metabolic pathways: glycolysis (EMP pathway), pentose phosphate pathway, Entner-Doudoroff pathway, TCA cycle, glyoxylate shunt, and various anaplerotic and cataplerotic reactions. Compared to *E. coli*, *V. natriegens* has one additional cataplerotic reaction, oxaloacetate decarboxylase (EC 4.1.1.3), which is not present in *E. coli*. Further analysis of annotated genes revealed that *V. natriegens* engages the same canonical amino acid biosynthesis pathways as *E. coli*.

An important reaction in the ^{13}C-MFA model is the lumped biomass reaction that captures the drain of precursor metabolites and cofactors needed for cell growth. To determine the coefficients for this biomass formation reaction, the biomass composition of *V. natriegens* was determined experimentally in this study. The results of this analysis are provided in Appendix Table G.2 and shown in Fig. 12.3, where the biomass composition of *V. natriegens* is also compared to that of *E. coli* (Long and Antoniewicz, 2014b) and the fast-growing thermophile *Geobacillus* LC300 (Cordova et al., 2015; Cordova and Antoniewicz, 2016). Proteins were the most abundant component of *V. natriegens* biomass (47% of dry weight), followed by RNA (29%), lipids (7%) and glycogen (3%). The RNA content was higher for *V. natriegens* (29%) compared to *E. coli* (21%), but similar to that of *Geobacillus* LC300 (28%). It has been observed previously that RNA content is often higher for fast growing strains (Long et al., 2016b; Pramanik and Keasling, 1997) (which is thought to reflect the need for more ribosomes to support the higher growth rates. The ratio of RNA to protein of 0.6 g/g is consistent with a previous report at a similar growth rate (Aiyar et al., 2002). The relative distribution of fatty acids in *V. natriegens* and *E. coli* were similar, with the most abundant fatty acids being C16:1 and C16:0 followed by C18:1

and C14:0. No odd-chain fatty acids (e.g., C15 or C17) were detected in *V. natriegens*. The relative distribution of amino acids in biomass was similar for the three microbes, with the notable exception of glutamate/glutamine (Glx), which was significantly elevated in *V. natriegens* compared to *E. coli* and *Geobacillus* LC300 (Fig. 12.3). The relative abundance of aspartate/asparagine (Asx) was lower for *V. natriegens* and *Geobacillus* LC300 compared to *E. coli*.

12.3.3 ^{13}C-Metabolic Flux Analysis

Next, we quantified intracellular metabolic fluxes for *V. natriegens* during exponential growth on glucose using high-resolution ^{13}C-MFA. The analysis consisted of first performing two parallel labeling experiments with [1,2-^{13}C]glucose and [1,6-^{13}C]glucose (an experimental design previously identified as providing optimal flux precision (Crown et al., 2016b)). Subsequently, fluxes were estimated by simultaneously fitting the measured acetate yield and the labeling data from proteinogenic amino acids, the ribose moiety of RNA, and the glucose moiety of glycogen to the model described in the previous section. A statistically acceptable fit was obtained. The minimized SSR value of 188 was lower than the maximum acceptable SSR value of 224 at 95% confidence level, assuming a constant measurement error of 0.3 mol% for all GC-MS measurements (Antoniewicz et al., 2007b). The estimated metabolic fluxes and 95% flux confidence intervals are provided in Appendix Table G.3.

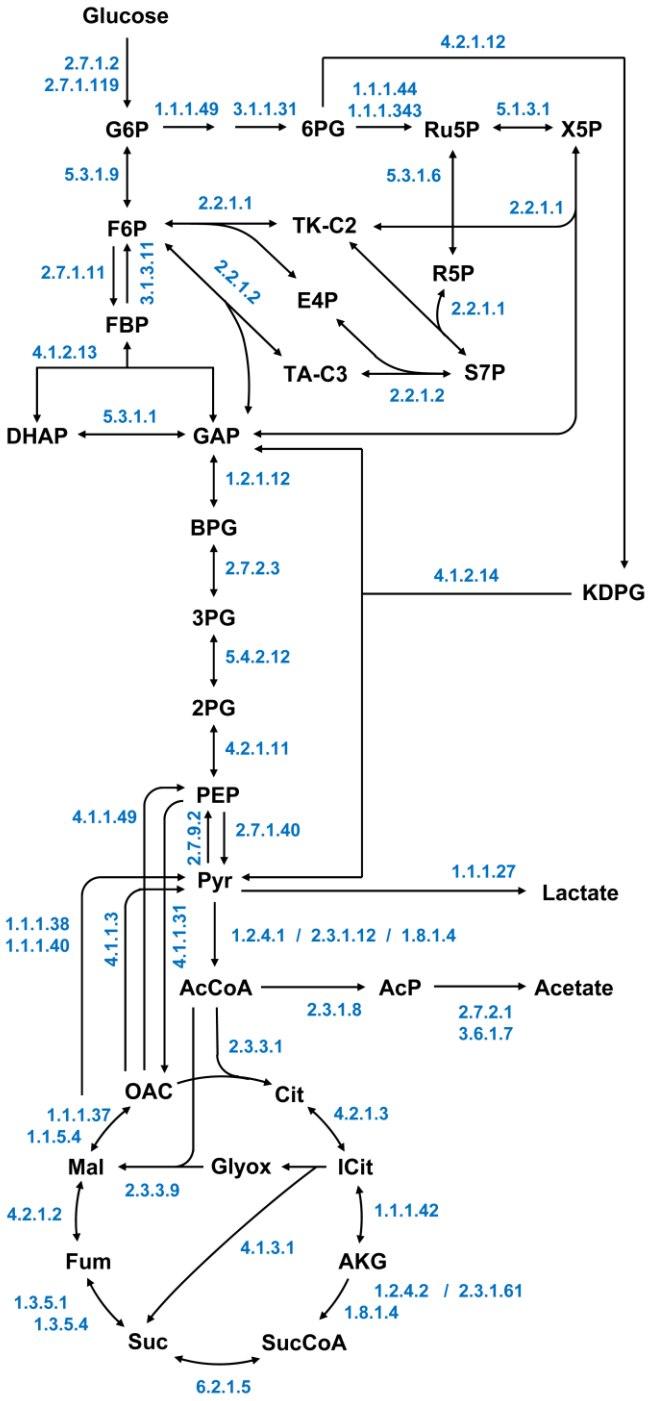


Figure 12.2: The central carbon metabolic network, as reconstructed from available genome annotations. EC numbers corresponding to annotated enzymes are shown for each reaction.

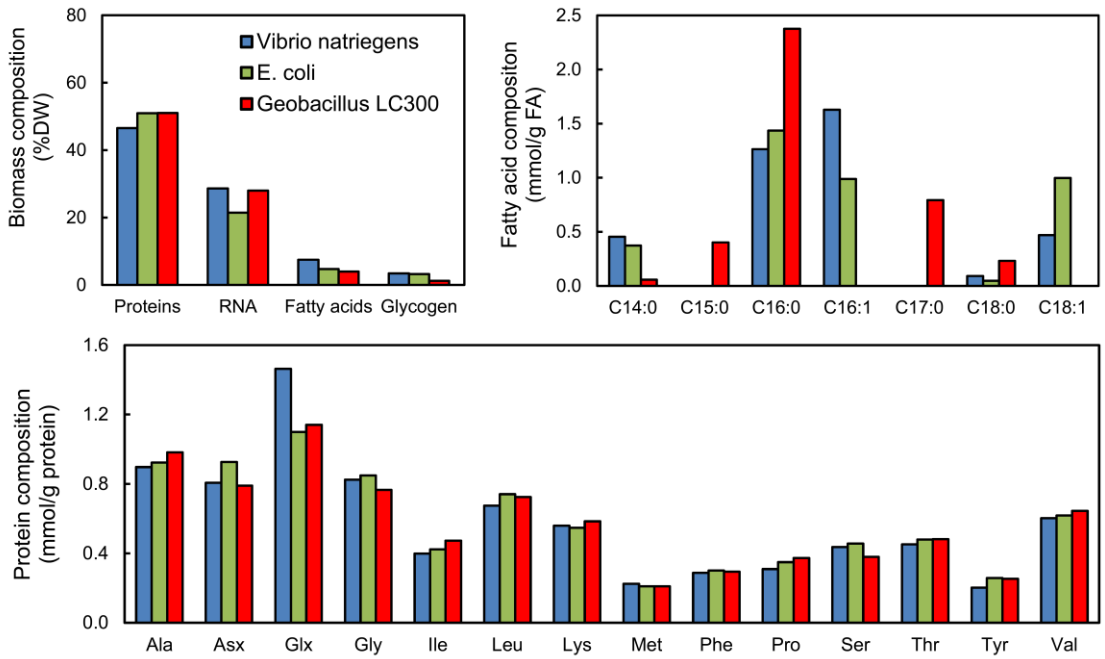


Figure 12.3: The biomass composition of *V. natriegens* and comparisons to *E. coli* and the fast-growing thermophile *Geobacillus* LC300. The fractional dry weights of the major biomass components are shown, as well as the distributions of fatty acids and amino acids.

Fig. 12.4 shows the estimated fluxes in central carbon metabolism for *V. natriegens* during aerobic growth on glucose (fluxes were normalized to glucose uptake rate of 100). The flux map of *V. natriegens* was characterized by a high glycolytic flux (80 ± 0.5 for phosphoglucose isomerase, and 169 ± 1 for glyceraldehyde 3-phosphate dehydrogenase), relatively high anaplerosis flux via phosphoenolpyruvate carboxylase (27 ± 1), and low oxidative PPP flux (18 ± 0.4) and TCA cycle fluxes (e.g. 17 ± 1 for citrate synthase; and 7 ± 0.4 for α-ketoglutarate dehydrogenase). Several pathways were found to be inactive (or nearly inactive) during aerobic growth on glucose, including the Entner-Doudoroff pathway (0.6 ± 0.1), the glyoxylate shunt (0.0 ± 0.1), and gluconeogenesis via phosphoenolpyruvate

carboxykinase (0.0 ± 0.3). In our model, the NAD-dependent malic enzyme (EC 1.1.1.38), the NADP-dependent malic enzyme (EC 1.1.1.40), and oxaloacetate decarboxylase (EC 4.1.1.3) were lumped into a single reaction, since it was not possible to independently estimate these three fluxes. The combined flux of these three reactions was estimated to be low (3.5 ± 0.4). Taken together, we determined that the *V. natriegens* glucose-normalized flux map was similar to that of *E. coli* during aerobic growth on glucose (Crown et al., 2015a; Leighty and Antoniewicz, 2013), with the main difference being a 33% lower flux of glucose through oxidative PPP, i.e. 27% of glucose for *E. coli* compared to 18% for *V. natriegens*. Given approximately 2.5-fold higher glucose uptake rate in *V. natriegens* compard to *E. coli*, however, the absolute carbon fluxes (i.e., mmol/g_{DW}/h) through the central carbon metabolic pathways are much higher in this organism.

12.3.4 Quantitative Analysis of Co-Factor Balances

To provide additional insights into the physiology of *V. natriegens*, we analyzed the production and consumption rates of key co-factors in metabolism, based on the ^{13}C-MFA estimated fluxes, and calculated the overall carbon balance. In Fig. 12.5, the results are summarized (see Appendix Table G.3 for additional details) and are compared to *E. coli* based on previous ^{13}C-MFA results (Crown et al., 2015a), on a glucose-normalized basis. The overall carbon balance for *V. natriegens* was similar to that of *E. coli*, with about 50% of glucose being converted to biomass, 25% to acetate, and 25% to CO_2. Fig. 12.5B shows the normalized production and consumption rates of NADH/$FADH_2$, NADPH and ATP, with contributions by the various pathways (absolute contributions in units of mmol/g_{DW}/h are shown in

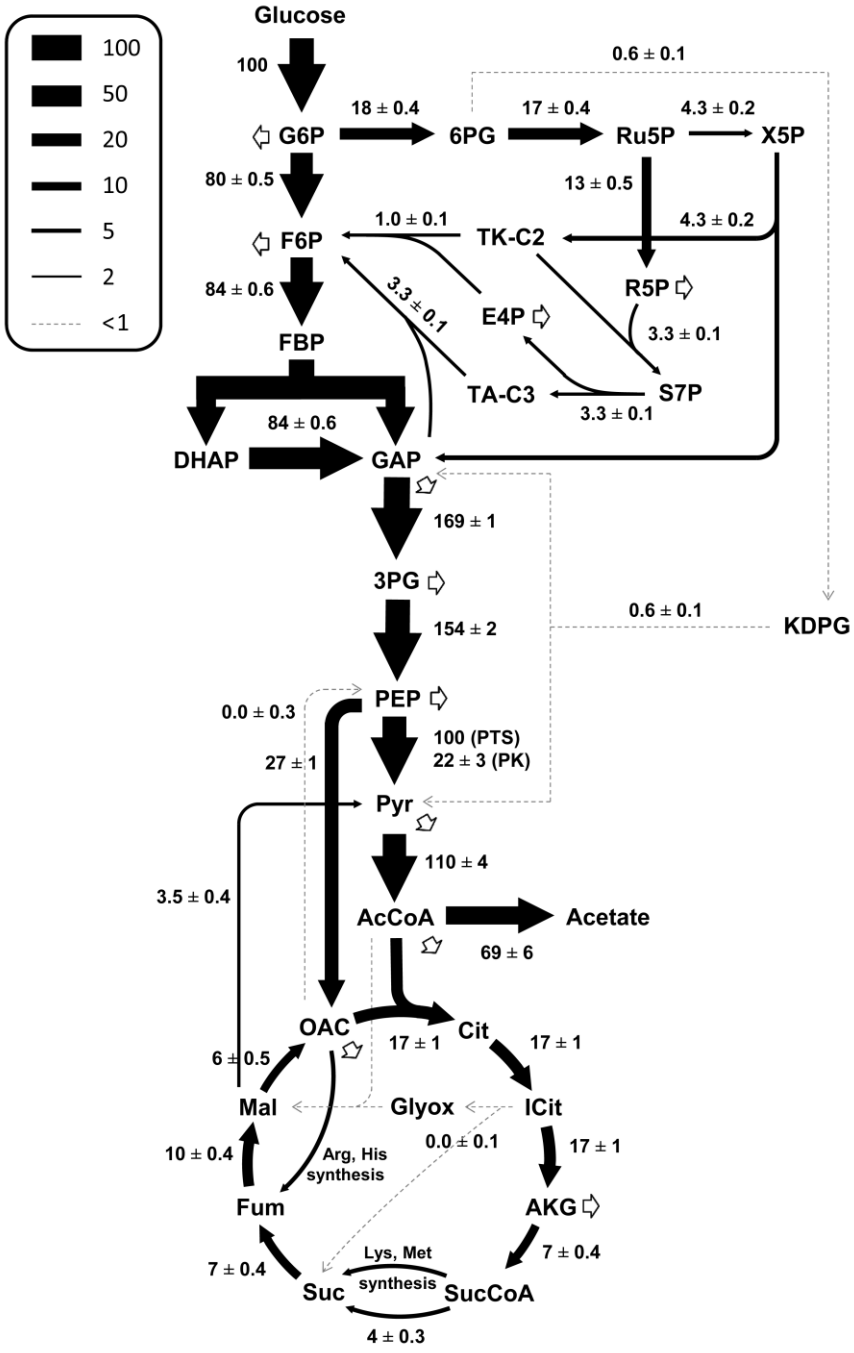


Figure 12.4: The central carbon metabolic fluxes of *V. natriegens*. All fluxes are normalized to 100 units of glucose uptake, assumed to be assimilated via the PTS transport system. Indicated errors reflect the standard errors of the flux estimate. The malic enzymes and oxaloacetate are lumped into a single reaction, shown here as malate to pyruvate.

Appendix Fig. G.1). The uncertainty introduced by the lumping of the NAD and NADP-dependent cataplerotic reaction (malic enzymes and oxaoloacetate decarboxylase) is minimal given the small measured flux. The amount of $NADH/FADH_2$ produced (per mol of glucose) was similar for *V. natriegens* and *E. coli*, with glycolysis (~50% contribution) and the TCA cycle (~40% contribution) being the main contributing pathways. The vast majority of $NADH/FADH_2$ was oxidized to generate ATP via oxidative phosphorylation (~80%), and 20% of NADH was converted to NADPH by transhydrogenases.

There were relatively larger differences in the production of NADPH between *V. natriegens* and *E. coli*. For *V. natriegens*, the majority of NADPH was produced from NADH by transhydrogenases (56%), with oxPPP contributing 25% and TCA cycle 14% to NADPH production, and negligible contribution from malic enzyme (<3%). For comparison, in *E. coli*, transhydrogenases and oxPPP contributed about equally to NADPH production (45% and 41%, respectively), with the TCA cycle contributing the remaining 13%. Based on the genome annotations, both *V. natriegens* and *E. coli* are believed to contain two pyridine nucleotide transhydrogenases, the membrane-bound PntAB (EC 1.6.1.2) which primarily converts NADH to NADPH, and the soluble form SthA (EC 1.6.1.1, also referred to as UdhA) which primarily converts NADPH to NADH (Sauer et al., 2004). The absolute net transhydrogenase flux from NADH to NADPH was 3-fold higher for *V. natriegens* (14.7 $mmol/g_{DW}/h$,) compared to *E. coli* (5.0 $mmol/g_{DW}/h$), thus suggesting a much more active PntAB enzyme.

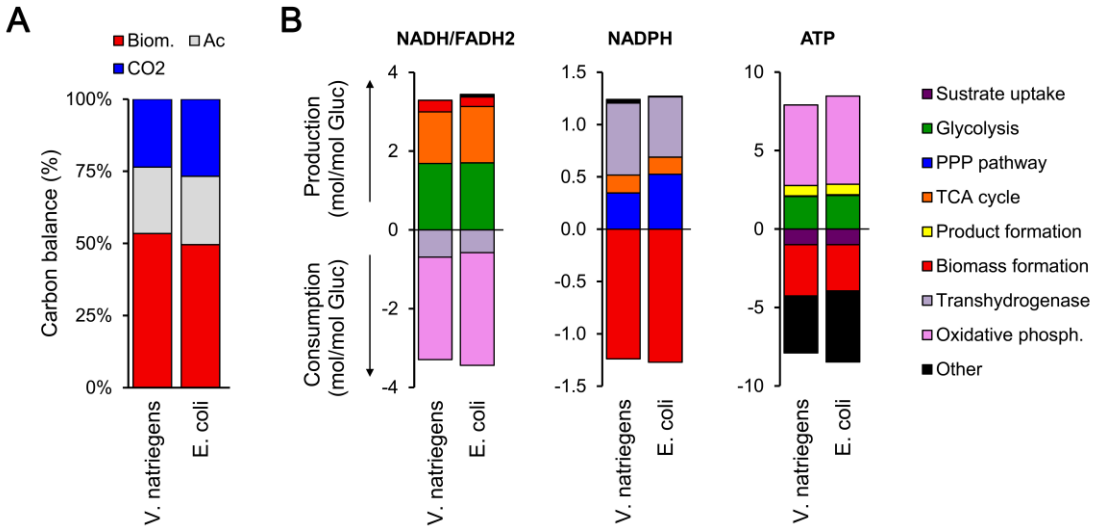


Figure 12.5: Carbon and cofactor balances of *V. natriegens* and *E. coli*. (A) The overall carbon balances reflect the fates of glucose on a C-mol basis. (B) The metabolic pathways responsible for the production and consumption of the cofactors NADH/$FADH_2$, NADPH, and ATP are summarized, on a normalized (per unit glucose) basis. For the ATP balance, the production rate from oxidative phosphorylation and the maintenance cost ('Other') are based on an assumed effective P/O ratio of 2. As this has not been reliably measured in *V. natriegens*, this result should be interpreted with caution.

Biological energy in the form of ATP is needed for three key cellular processes: 1) transport of substrates and nutrients into the cells; 2) anabolism (i.e. cell growth); and 3) maintenance. The overall ATP balances were similar for *V. natriegens* and *E. coli* (Fig 5), with the oxidative phosphorylation contributing the most towards ATP production, approximately 65% for both organisms, followed by substrate level phosphorylation in glycolysis (~25%), and acetate production (~10%). The contribution of the TCA cycle to ATP production was negligible for both organisms. It should be noted that the absolute ATP production rates calculated here should be viewed with caution, since little information is available about the effective P/O ratio

for *V. natriegens*. For simplicity, we assumed the same P/O ratio of 2.0 for both organisms. Once a reliable estimate of P/O ratio for *V. natriegens* is available, the ATP analysis results reported here should be re-evaluated.

12.3.5 Comparison of *V. natriegens* Physiology and Other Fast Growing Bacteria

Finally, we compared key physiological characteristics (i.e. growth rate, glucose uptake rate, and oxygen uptake rate) of *V. natriegens* to three other strains (Fig. 12.6): wild-type *E. coli* (Leighty and Antoniewicz, 2013), adaptively evolved *E. coli* (Sandberg et al., 2016), and the fast-growing extreme thermophile *Geobacillus* LC300 (Cordova et al., 2015; Cordova and Antoniewicz, 2016). Like *V. natriegens*, the *E. coli* characteristics reflect growth at 37 °C, while those of *Geobacillus* LC300 are from growth at 72 °C. The growth rate of *V. natriegens* on minimal glucose medium of 1.70 h^{-1} (doubling time 24 min) was significantly higher than the growth rate of both wild-type *E. coli* (~0.7 h^{-1}, doubling time 60 min) and evolved *E. coli* (~1.0 h^{-1}, doubling time 40 min), but significantly lower than the growth rate of the thermophile *Geobacillus* LC300, which has a growth rate on minimal glucose medium of 2.15 h^{-1} (doubling time of 19 min). The biomass-specific glucose uptake rates followed the same pattern as growth rates (Fig 12.6B), since the biomass yields are similar for all four strains, with *Geobacillus* LC300 having the highest glucose uptake rate of 31 mmol/g_{DW}/h, followed by *V. natriegens* (21 mmol/g_{DW}/h), the evolved *E. coli* (12 mmol/g_{DW}/h), and wild-type *E. coli* (8.5 mmol/g_{DW}/h). Even more striking were the differences between the three mesophilic strains and *Geobacillus* LC300 when the biomass-specific oxygen uptake rates were compared. The oxygen uptake rate of *Geobacillus* LC300 (~60 mmol/g_{DW}/h) was more than 2-fold higher than the oxygen uptake rate of *V. natriegens* (28 mmol/g_{DW}/h), more than 3-fold higher

compared to evolved *E. coli* (19 mmol/g_{DW}/h), and about 5-fold higher compared to wild-type *E. coli* (12 mmol/g_{DW}/h). These values suggest that the gram-positive thermophile *Geobacillus* LC300 has a significantly more active electron transport chain compared to the gram-negative mesophiles *V. natriegens* and *E. coli*. Unfortunately, metabolism of *Geobacillus* LC300 and evolved *E. coli* strains have not been elucidated by ^{13}C-MFA during growth on glucose, so a more comprehensive comparison of metabolic fluxes between the four strains is not possible at this time.

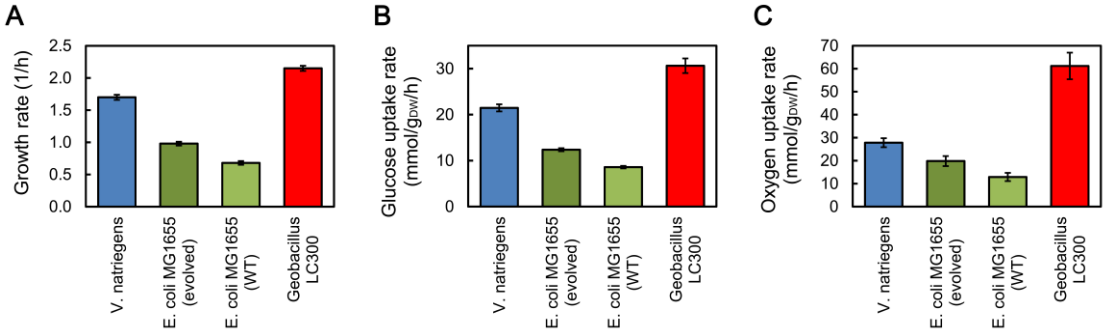


Figure 12.6: Overall physiological assessment of four potential host strains for biotechnology applications: *V. natriegens*, *E. coli* (following adaptive evolution and the un-evolved wild-type), and the thermophilic *Geobacillus* LC300. The values reflect growth at 37 °C for the mesophilic strains and 72 °C for *Geobacillus* LC300. The growth rates (A), and biomass specific glucose (B) and oxygen (C) uptake rates are shown.

12.4 Conclusions

The physiological and fluxomic analysis presented here are important foundational components of a developing knowledge base from which the community will be able to study and engineer *V. natriegens*. The verified intracellular metabolic pathways and wild-type reference flux map will enable the use of several COBRA

model-based analyses and designs. Additional systems measurements, particularly of the proteome and transcriptome, will complement these results and help begin to answer the fundamental question of how this organism achieves such fast growth. For example, these measurements may determine whether *V. natriegens* has a more efficiently allocated proteome for growth. In adaptively evolved *E. coli*, mutations in global regulation are frequently observed which shift gene expression toward catabolic processes at the expense of stress response functions (LaCroix et al., 2015; Utrilla et al., 2016). There is some limited evidence of less robust stress responses already, with the observation of substantially lower intrinsic catalase activity in *V. natriegens* than in *E. coli* (Weinstock et al., 2016). Other molecular mechanisms may also be investigated, such as the DNA replication and protein synthesis machinery (the high RNA content measured is believed to be related to high rRNA levels and ribosome number (Aiyar et al., 2002)). As more biological knowledge is developed, advanced *in silico* approaches including ME models (Liu et al., 2014; O'Brien et al., 2013), kinetic models (Khodayari and Maranas, 2016; Tran et al., 2008), and even whole cell models (Karr et al., 2012) may be applied to develop systems-level understanding.

Insights into fast growing strains like *V. natriegens* (as well as *Geobacillus* LC300, adaptively evolved *E. coli*, and others) may enable rational engineering approaches of conventional organisms for improved performance. Given its natural advantages, however, it has been proposed that *V. natriegens* may be a superior next-generation host for biotechnological applications. Its similar flux distribution to *E. coli* in glucose minimal medium is encouraging in that raises the prospect that some metabolic engineering strategies previously developed in *E. coli* may also be successful in *V. natriegens*. Further study of metabolic kinetics and regulation in *V.*

natriegens, and the nature of responses to genetic and environmental perturbations, will be critical. Finally, we have shown that although *V. natriegens* has large advantages in rate over its mesophilic counterpart *E. coli*, it lags behind the thermophilic *Geobacillus* LC300. Similar development would be needed in systems knowledge and genetic tools to develop this (and any other non-model organism) into a viable host, after which an informed choice based on desired process temperature and other characteristics would be possible.

12.5 Author Contributions

All authors contributed to the execution and analysis of experimental data. CPL and MRA wrote the manuscript with contributions from all authors.

Chapter 13

CONCLUSIONS AND FUTURE DIRECTIONS

13.1 Methods in ^{13}C Metabolic Flux Analysis

Novel contributions to methods in microbial physiological and fluxomic characterization were described in Chapters 2-4. These developments shed light on critical aspects of designing and executing ^{13}C-MFA, specifically tracer selection, measurement choice, and biomass composition. As was discussed in the respective chapters, each of these can determine flux observability and flux estimate precision. The detailed mapping of optimal tracer selection (Chapter 4) is a particularly significant development toward enabling rational and informed experimental design. Given the significant nontrivial challenges in predicting the relationship between tracer selection and flux estimate quality, many previous studies defaulted to tracers which had been previously used (e.g. mixtures of [U-^{13}C] and [1-^{13}C]glucose) or were least expensive. The use of RNA and glycogen in ^{13}C-MFA (Chapter 3) is directly useful in microbial flux analysis, and the presented approach for identifying value in additional measurements is more broadly applicable. For example, with other analytical techniques such as LC-MS analysis of intracellular metabolites, additional exchange fluxes in metabolism may become observable.

It has also become increasingly clear that accurate determination of the biomass composition is critical for accurate determination of fluxes. Many previous flux studies assumed a constant biomass composition, for example among *E. coli* knockout strains in a previous large-scale study (Ishii et al., 2007). Here (Chapters 6

and 8), we have shown that this assumption is weak, especially across strains with diverse growth rates. Significant differences in RNA or glycogen content were observed, for example, which affect the mass balance and flux analysis of upper metabolism when the drain of the precursors R5P and G6P are properly considered. The accurate quantification of fluxes through secondary biosynthetic pathways rely entirely on these measurements, which is of high interest as ^{13}C-MFA and kinetic models of metabolism grow in scope to incorporate these pathways (García Martín et al., 2015; Gopalakrishnan and Maranas, 2015). Furthermore, the demonstrated and facile extension of the composition analysis to diverse organisms (Section 2.4.1), including algae (Zuñiga et al., 2016) and several genera of bacteria, provides an important enabling tool for applying ^{13}C-MFA to non-model (and even unidentified (Yao et al., 2015)) organisms.

These and related improvements to ^{13}C-MFA methodologies in recent years have contributed to the technology's maturation, to the point that highly precise flux estimates can be obtained at reasonable experimental effort and throughput. For example, the large-scale knockout study presented here (Chapters 7 and 9) utilized the parallel labeling experiment identified as optimal ([1,2] and [1,6-^{13}C]glucose), with RNA, glycogen, and amino acid labeling measurements and biomass composition enabling the highly precise characterization of 45 strains. Altogether in this thesis, ^{13}C-MFA was applied to 66 *E. coli* strains (Chapters 7, 9, 10, 11) and the wild-type *Vibrio natriegens* (Chapter 12). Ongoing work is also underway to determined fluxes for an additional 26 wild-type and evolved *E. coli* strains (see Section 13.3.2, below). Continued experience and standardization will make ^{13}C-MFA an increasingly accessible tool for systems biology research.

13.2 Characterization of Metabolic Responses to Gene Knockouts

Detailed physiological and fluxomic characterizations were performed for 45 *E. coli* strains with knockouts of central carbon metabolism enzymes (Chapters 6-9). This study revealed physiological relationships such as the dependence of biomass components on growth rate, and the fluxomic results provide a detailed picture of metabolic flexibility and robustness. Novel reactions were identified, including the reversibility of Enzyme I of the PTS system (Chapter 5) and glucose excretion in *ΔpfkA* (Chapter 7). Systems-level analysis revealed some bottlenecks in metabolism, including seven knockouts with a growth rate decrease of 50% or greater (in addition to the conditionally essential genes not included). In upper metabolism, the oxidative phosphate pathway was observed to be inflexible in compensating for blockages in glycolysis (e.g., *Δpgi* and *ΔpfkA*), an issue which was resolved after adaptive evolution (Chapter 10). In lower metabolism, activation of lactate and pyruvate secretion was observed in several strains, including in *ΔaceE* where these were produced with 80% carbon yield. Surprisingly, the glyoxylate shunt and PCK reactions were not activated in any lower metabolism knockout, and the malic enzyme only activated in one (*Δmdh*).

13.2.1 Utility in Systems Biology and Engineering

This dataset of knockout metabolic responses will be a valuable resource for the systems biology and metabolic engineering communities. Although we have presented an initial analysis and interpretation of the results, additional insights will undoubtedly emerge as other researchers apply their own analyses and perspectives. Given the complexity of the systems biology literature and even this dataset, it is likely that additional inferences can be made based on information not considered

here. For example, specialists in specific regulatory mechanisms may be able to recognize or explain various phenomena such as the different metabolic responses to the two *pta/ackA* pathway knockouts, or the activated ED pathway in *Δgnd*. Assembly of multi-omics data including these knockout fluxes in appropriate databases will facilitate these types of insights. More broadly, the many analytical approaches that have been applied to metabolism previously, described in Sections 1.3.1 and 1.3.3, can be re-applied to these results. Perhaps most notably, existing models of metabolism that aim to predict knockout responses can be assessed, as was done here (Chapter 6-9) for the COBRA models FBA, MOMA, and RELATCH. The poor agreement that was found motivates the need for more approaches that incorporate kinetics and regulation.

13.2.2 Development of Ensemble Kinetic Models

A direct application of this knockout flux data that is being currently pursued is ensemble kinetic modeling. Detailed descriptions of this approach can be found here (Khodayari et al., 2014; Khodayari and Maranas, 2016; Tan and Liao, 2012; Tran et al., 2008). Briefly, a detailed model is developed based on mass-action kinetics and substrate-level regulation. An "ensemble" of candidate models is then generated which spans the very large parameter space (e.g., an ensemble size of 131,072 was used in (Khodayari and Maranas, 2016) of models that are consistent with wild-type fluxes. Subsequently, the models can be challenged to predict knockout fluxes, and those inconsistent with experimental observations are discarded. Provided sufficient experimental knockout data, the ensemble can be pruned to a unique or nearly unique "valid" model. To this end, the comprehensive, self-consistent, and detailed approach taken in this study make the collected set of knockout flux measurements ideal for

ensemble modeling. In a collaborative effort, transcriptomics measurements for these same knockout strains are being obtained, and together will provide an opportunity for developing a kinetic model with unprecedented precision. Such a model will be a significant advance in descriptive and predictive understanding of metabolism, and will be hopefully open new levels of insight and opportunity in systems biology and strain engineering.

13.2.3 Extension to Additional Knockouts

As discussed in Section 1.3.4, central carbon metabolism under the glucose and aerobic growth conditions was identified as the highest priority for knockout studies. There are many potentially valuable directions to investigate next. First, the global regulation of metabolism is of high interest. There are hundreds of transcription factors in *E. coli*, a much smaller number (~10-20) are known to dominate control of metabolic processes (Fang et al., 2017). Other pathways of interest might include secondary metabolism, for example in amino acid, nucleic acid, or lipid biosynthesis, or metabolite membrane transporters. With regard to glucose transport, further investigation of the PTS system mutants (fluxes, transcriptional regulatory impact, carbon catabolite repression) would be of particular interest.

13.3 Metabolic Responses to Adaptive Laboratory Evolution

The value of adaptive laboratory evolution in developing systems metabolic knowledge was demonstrated in Chapters 10-11. In Chapter 10, mechanisms of growth recovery in *Δpgi* were explored. As discussed in Chapters 6-7, this knockout strain has a highly perturbed phenotype, with very slow growth and massive flux redistribution. The coupling of genetic analysis of mutations with metabolic

characterization via ^{13}C-MFA revealed that growth recovery was enabled by the alleviation of key rate limitations in cofactor metabolism and the oxidative pentose phosphate pathway. What were previously described as important processes of latent pathway activation and re-repression were found to represent instead low constant levels of flux. In Chapter 11, the metabolic fluxes of fast-growing adaptively evolved strains were characterized for the first time, and surprisingly little change was observed in relative intracellular pathway usage. This is consistent with the hypothesis that the main mechanism of fast growth is a broad shift in the expressed proteome, to reduce functions that hedge against environmental change or challenge, and increase catabolism and growth functions. Out of these two studies arise new questions about the impact of global regulators, for example the RNA polymerase subunits, on metabolic control and fitness. In both of these instances, additional systems measurements such as transcriptomics or proteomics, would be useful in more fully elucidating the regulated responses.

13.3.1 ALE for Growth Recovery in Gene Knockout Strains

Just as the ALE of *Δpgi* revealed mechanisms to overcome bottlenecks in the pentose phosphate pathway and cofactor metabolism, so too might similar studies provide rich insight into other areas of metabolism. The study of central carbon metabolic knockouts (Chapters 6-9) provides a rich set of potential targets. For example, the seven knockouts with growth rates more than 50% slower than the wild-type are likely to have significant fitness enhancements realized in ALE, and each could provide unique and specific insights. For example, how the *Δrpe* mutant corrects imbalances in the pentose phosphate pathway is likely to be very different from how *ΔtpiA* modulates its methyglyoxal, glyoxylate, and PCK pathways, which will

undoubtedly be different still from how *ΔaceE* improves the supply of carbon to AcCoA and glutamate. If and how *ΔpfkA* reduces the glucose secretion flux would likely shed light onto the mechanism and regulation of that reaction. ALE would also illuminate the metabolic potential of various modes of pathway usage, for example what growth rates are possible when the pentose phosphate pathway or TCA cycle is unable to be used (e.g., *Δzwf* and *ΔsucB*, respectively).

13.3.2 ALE of Wild-Type *E. coli* for Fast Growth

The results of Chapter 11 show that intracellular flux redistribution is not required for fast growth, and does not occur in *E. coli* MG1655. A key finding was that inter-strain differences, e.g., between MG1655 and BL21 wild-type strains, were more pronounced than those caused by ALE. Therefore, a key question is whether these metabolic distinct wild-type strains would converge to a single 'optimum' phenotype through ALE, or if they would find alternative optima dependent on the initial condition. To address this and gain additional systems insight into fast growth in *E. coli*, our collaborators Bernhard Palsson and Adam Feist have led an effort to evolve six different wild-type strains: BL21, C, Crooks, MG1655, W, and W3110. A multi-omics approach including genetic resequencing, transcriptomics, and fluxomics is being applied to the parent strains and 20 total evolved strains representing multiple time points during evolution. These strains have significant metabolic differences, particularly the high TCA cycle flux and lack of a *pgl* gene in BL21, as well as the significant ED pathway usage (~10%) in C and Crooks strains (unpublished data). A particularly interesting analysis will be to compare transcript level changes to flux changes, which will possibly illuminate key areas of kinetic limitation.

13.4 Fast Growth and Future Host Strain Development

The multi-omics approach in this effort, and future similar works in other fast-growing strains such as *Vibrio natriegens*, will provide an important picture of the limits of metabolic rates. It appears, in *E. coli* at least, that faster growth is achieved through a proteomic "re-prioritization". Whereas the wild-type allocates some fraction of its proteome to preparedness for environmental stress or perturbation, for example substrate starvation or switch, temperature or pH change etc., the faster growing evolved strain reduces these to allow for increased expression of metabolic and biosynthetic genes (Utrilla et al., 2016). Strains with such properties would be desirable for bioprocesses, where the controlled environment of a bioreactor would require less robustness and absolute rates of productivity are crucial. So far, all the evolved *E. coli* strains have not surpassed a maximum growth rate of ~1.1 h^{-1}. Systems biology measurements and modeling will be needed to identify the reasons for this limitation, and whether it can be surpassed. Some possibilities, for example, are spatial limitations in the membrane (Liu et al., 2014) or kinetic limitations in biosynthesis or DNA replication. Comparison to *Vibrio natriegens* (1.7 h^{-1}) will surely be a relevant approach in this effort. This organism should be further characterized in terms of its robustness, shape, and enzyme variants. Provided continued development of genetic tools, it may become a preferred host in some biotechnology applications. Lessons learned in probing these organisms and phenotypes may also inform efforts to create genetically re-factored or minimized chassis microbes (Hutchison et al., 2016; Kuznetsov et al., 2017).

REFERENCES

Ahn, W.S., Antoniewicz, M.R., 2013. Parallel labeling experiments with [1,2-^{13}C]glucose and [U-^{13}C]glutamine provide new insights into CHO cell metabolism. Metab. Eng. 15, 34–47.

Ahn, W.S., Antoniewicz, M.R., 2011. Metabolic flux analysis of CHO cells at growth and non-growth phases using isotopic tracers and mass spectrometry. Metab. Eng. 13, 598–609.

Ahn, W.S., Crown, S.B., Antoniewicz, M.R., 2016. Evidence for transketolase-like TKTL1 flux in CHO cells based on parallel labeling experiments and ^{13}C-metabolic flux analysis. Metab. Eng. 37, 72–78.

Aiyar, S.E., Gaal, T., Gourse, R.L., 2002. rRNA Promoter Activity in the Fast-Growing Bacterium Vibrio natriegens. J. Bacteriol. 184, 1349–1358.

Antoniewicz, M.R., 2015a. Parallel labeling experiments for pathway elucidation and 13C metabolic flux analysis. Curr. Opin. Biotechnol. 36, 91–97.

Antoniewicz, M.R., 2015b. Methods and advances in metabolic flux analysis: a mini-review. J. Ind. Microbiol. Biotechnol. 42, 317–325.

Antoniewicz, M.R., 2015. Parallel labeling experiments for pathway elucidation and 13C metabolic flux analysis. Curr. Opin. Biotechnol. 36, 91–97.

Antoniewicz, M.R., 2013a. Tandem mass spectrometry for measuring stable-isotope labeling. Curr. Opin. Biotechnol. 24, 48–53.

Antoniewicz, M.R., 2013b. Dynamic metabolic flux analysis-tools for probing transient states of metabolic networks. Curr. Opin. Biotechnol. 24.

Antoniewicz, M.R., 2013c. 13C metabolic flux analysis: optimal design of isotopic labeling experiments. Curr. Opin. Biotechnol. 24.

Antoniewicz, M.R., Kelleher, J.K., Stephanopoulos, G., 2011. Measuring deuterium enrichment of glucose hydrogen atoms by gas chromatography/mass spectrometry. Anal. Chem. 83, 3211–6.

Antoniewicz, M.R., Kelleher, J.K., Stephanopoulos, G., 2007a. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. Metab. Eng. 9, 68–86.

Antoniewicz, M.R., Kelleher, J.K., Stephanopoulos, G., 2007b. Accurate Assessment of Amino Acid Mass Isotopomer Distributions for Metabolic Flux Analysis. Anal. Chem. 79, 7554–7559.

Antoniewicz, M.R., Kelleher, J.K., Stephanopoulos, G., 2006. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. Metab. Eng. 8, 324–337.

Antoniewicz, M.R., Kraynie, D.F., Laffend, L. a, González-Lergier, J., Kelleher, J.K., Stephanopoulos, G., 2007c. Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of E. coli producing 1,3-propanediol. Metab. Eng. 9, 277–92.

Aristilde, L., Lewis, I.A., Park, J.O., Rabinowitz, J.D., 2015. Hierarchy in pentose sugar metabolism in Clostridium acetobutylicum. Appl. Environ. Microbiol. 81, 1452–1462.

Atsumi, S., Wu, T.-Y., Machado, I.M.P., Huang, W.-C., Chen, P.-Y., Pellegrini, M., Liao, J.C., 2010. Evolution, genomic analysis, and reconstruction of isobutanol tolerance in Escherichia coli. Mol. Syst. Biol. 6, 1–11.

Au, J., Choi, J., Jones, S.W., Venkataramanan, K.P., Antoniewicz, M.R., 2014. Parallel labeling experiments validate Clostridium acetobutylicum metabolic network model for (13)C metabolic flux analysis. Metab. Eng. 26C, 23–33.

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H., 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol. 2, 1–11.

Badur, M.G., Zhang, H., Metallo, C.M., 2015. Enzymatic passaging of human embryonic stem cells alters central carbon metabolism and glycan abundance. Biotechnol. J. 10, 1600–1611.

Bailey, J.E., 1991. Toward a Science of Metabolic Engineering. Science 252, 1668–1675.

Bajad, S.U., Lu, W., Kimball, E.H., Yuan, J., Peterson, C., Rabinowitz, J.D., 2006. Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. J. Chromatogr. A 1125,

76–88.

Barrick, J., Yu, D., Yoon, S., Jeong, H., Oh, T., Schneider, D., Lenski, R., Kim, J., 2009. Genome evolution and adaptation in a long-term experiment with Escherichia coli. Nature 461, 1243–1247.

Basan, M., Hui, S., Zhang, Z., Shen, Y., Williamson, J.R., Hwa, T., 2015. Overflow metabolism in bacteria results from efficient proteome allocation for energy biogenesis. Nature 528, 99–106.

Becker, S.A., Feist, A.M., Mo, M.L., Hannum, G., Palsson, B.Ø., Herrgard, M.J., 2007. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat. Protoc. 2, 727–738.

Behre, J., Wilhelm, T., von Kamp, A., Ruppin, E., Schuster, S., 2008. Structural robustness of metabolic networks with respect to multiple knockouts. J. Theor. Biol. 252, 433–441.

Bennett, B.D., Yuan, J., Kimball, E.H., Rabinowitz, J.D., 2008. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nat. Protoc. 3, 1299–311.

Benthin, S., Nielsen, J., Villadsen, J., 1991. A simple and reliable method for the determination of cellular RNA content. Biotechnol. Tech. 5, 39–42.

Bettenbrock, K., Sauter, T., Jahreis, K., Kremling, A., Lengeler, J.W., Gilles, E.D., 2007. Correlation between growth rates, EIIACrr phosphorylation, and intracellular cyclic AMP levels in Escherichia coli K-12. J. Bacteriol. 189, 6891–6900.

Blank, L.M., Kuepfer, L., Sauer, U., 2005. Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol. 6, R49.

Bradford, M., 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. 72, 248–254.

Burgard, A.P., Maranas, C.D., 2001. Probing the Performance Limits of the Escherichia coli Metabolic Network Subject to Gene Additions or Deletions. Biotechnol. Bioeng. 74, 364–375.

Burgard, A.P., Nikolaev, E. V, Schilling, C.H., Maranas, C.D., 2004. Flux coupling analysis of genome-scale metabolic network reconstructions. Genome Res. 14,

301–312.

Burgard, A.P., Pharkya, P., Maranas, C.D., 2003. Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol. Bioeng. 84, 647–57.

Canonaco, F., Hess, T. a, Heri, S., Wang, T., Szyperski, T., Sauer, U., 2001. Metabolic flux response to phosphoglucose isomerase knock-out in Escherichia coli and impact of overexpression of the soluble transhydrogenase UdhA. FEMS Microbiol. Lett. 204, 247–252.

Carlson, R.P., 2007. Metabolic systems cost-benefit analysis for interpreting network structure and regulation. Bioinformatics 23, 1258–1264.

Caspi, R., Altman, T., Dreher, K., Fulcher, C.A., Pallavi, S., Keseler, I.M., Kothari, A., Krummenacker, M., Latendresse, M., Mueller, L.A., Ong, Q., Paley, S., Pujar, A., Shearer, A.G., Travers, M., Weerasinghe, D., Zhang, P., Karp, P.D., 2012. The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. Nucleic Acids Res. 40, D472–D753.

Chandrasekaran, S., Price, N.D., 2010. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in Escherichia coli and Mycobacterium tuberculosis. Proc. Natl. Acad. Sci. U. S. A. 107, 17845–50.

Charusanti, P., Conrad, T.M., Knight, E.M., Venkataraman, K., Fong, N.L., Xie, B., Gao, Y., Palsson, B.Ø., 2010. Genetic basis of growth adaptation of Escherichia coli after deletion of pgi, a major metabolic gene. PLoS Genet. 6, e1001186.

Chen, X., Alonso, A.P., Allen, D.K., Reed, J.L., Shachar-Hill, Y., 2011. Synergy between 13C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in E. coli. Metab. Eng. 13, 38–48.

Cherepanov, P.P., Wackernagel, W., 1995. Gene disruption in Escherichia coli: TcR and KmR cassettes with the option of Flp-catalyzed excision of the antibiotic-resistance determinant. Gene 158, 9–14.

Chiang, C., Lee, H.M., Guo, H.J., Wang, Z.W., Lin, L., 2013. Systematic Approach To Engineer Escherichia coli Pathways for Co-Utilization of a Glucose-Xylose Mixture. J. Agric. Food Chem. 61, 7583–7590.

Cho, B.-K., Barrett, C.L., Knight, E.M., Park, Y.S., Palsson, B.O., 2008. Genome-scale reconstruction of the Lrp regulatory network in Escherichia coli. Proc. Natl. Acad. Sci. 105, 19462–19467.

Chou, H.H., Marx, C.J., Sauer, U., 2015. Transhydrogenase promotes the robustness and evolvability of E. coli deficient in NADPH production. PLoS Genet. 11, e1005007.

Chowdhury, A., Khodayari, A., Maranas, C.D., 2015. Improving prediction fidelity of cellular metabolism with kinetic descriptions. Curr. Opin. Biotechnol. 36, 57–64.

Chowdhury, A., Zomorrodi, A.R., Maranas, C.D., 2014. k-OptForce: Integrating Kinetics with Flux Balance Analysis for Strain Design. PLoS Comput. Biol. 10, e1003487.

Conrad, T.M., Joyce, A.R., Applebee, M.K., Barrett, C.L., Xie, B., Gao, Y., Palsson, B.Ø., 2009. Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. Genome Biol. 10, R118.

Cordova, L.T., Alper, H.S., 2016. Central metabolic nodes for diverse biochemical production. Curr. Opin. Chem. Biol. 35, 37–42.

Cordova, L.T., Antoniewicz, M.R., 2016. ^{13}C metabolic flux analysis of the extremely thermophilic, fast growing, xylose-utilizing Geobacillus strain LC300. Metab. Eng. 33, 148–157.

Cordova, L.T., Cipolla, R.M., Swarup, A., Long, C.P., Antoniewicz, M.R., 2017. 13C metabolic flux analysis of three divergent extremely thermophilic bacteria: Geobacillus sp. LC300, Thermus thermophilus HB8, and Rhodothermus marinus DSM 4252. Metab. Eng. 44, 182–190.

Cordova, L.T., Long, C.P., Venkataramanan, P., Antoniewicz, M.R., 2015. Complete genome sequence, metabolic model construction and phenotypic characterization of Geobacillus LC300, an extremely thermophilic, fast growing, xylose-utilizing bacterium. Metab. Eng. 32, 74–81.

Cordova, L.T., Lu, J., Cipolla, R.M., Sandoval, N.R., Long, C.P., Antoniewicz, M.R., 2016. Co-utilization of glucose and xylose by evolved Thermus thermophilus LC113 strain elucidated by 13C metabolic flux analysis and whole genome sequencing. Metab. Eng. 37, 63–71.

Cornelius, S.P., Sang, J., Motter, A.E., 2011. Dispensability of Escherichia coli's latent pathways. Proc. Natl. Acad. Sci. U. S. A. 108, 3124–3129.

Covert, M.W., Knight, E.M., Reed, J.L., Herrgard, M.J., Palsson, B.Ø., 2004. Integrating high-throughput and computational data. Nature 429, 92–96.

Crown, S.B., Ahn, W.S., Antoniewicz, M.R., 2012. Rational design of ^{13}C-labeling experiments for metabolic flux analysis in mammalian cells. BMC Syst. Biol. 6, 43.

Crown, S.B., Antoniewicz, M.R., 2013a. Parallel labeling experiments and metabolic flux analysis: Past, present and future methodologies. Metab. Eng. 16, 21–32.

Crown, S.B., Antoniewicz, M.R., 2013b. Publishing 13C metabolic flux analysis studies: A review and future perspectives. Metab. Eng. 20, 42–48.

Crown, S.B., Antoniewicz, M.R., 2012. Selection of tracers for 13C-Metabolic Flux Analysis using Elementary Metabolite Units (EMU) basis vector methodology. Metab. Eng. 14, 150–161.

Crown, S.B., Kelleher, J.K., Rouf, R., Muoio, D.M., Antoniewicz, M., 2016a. Comprehensive metabolic modeling of multiple 13C-isotopomer data sets to study metabolism in perfused working hearts. Am. J. Physiol. Heart Circ. Physiol. ajpheart.00428.2016.

Crown, S.B., Long, C.P., Antoniewicz, M.R., 2016b. Optimal tracers for parallel labeling experiments and ^{13}C metabolic flux analysis: A new precision and synergy scoring system. Metab. Eng. 38, 10–18.

Crown, S.B., Long, C.P., Antoniewicz, M.R., 2015a. Integrated 13C-metabolic flux analysis of 14 parallel labeling experiments in Escherichia coli. Metab. Eng. 28, 151–158.

Crown, S.B., Marze, N., Antoniewicz, M.R., 2015b. Catabolism of Branched Chain Amino Acids Contributes Significantly to Synthesis of Odd-Chain and Even-Chain Fatty Acids in 3T3-L1 Adipocytes. PLoS One 10, e0145850.

Curtis, S.J., Epstein, W., 1975. Phosphorylation of D-glucose in Escherichia coli mutants defective in glucosephosphotransferase, mannosephosphotransferase, and glucokinase. J. Bacteriol. 122, 1189–1199.

Datsenko, K. a, Wanner, B.L., 2000. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc. Natl. Acad. Sci. U. S. A. 97, 6640–5.

De Anda, R., Lara, A.R., Hernández, V., Hernández-Montalvo, V., Gosset, G., Bolívar, F., Ramírez, O.T., 2006. Replacement of the glucose phosphotransferase transport system by galactose permease reduces acetate accumulation and improves process performance of Escherichia coli for recombinant protein production without impairment of growth rate. Metab. Eng. 8, 281–290.

Deutscher, J., Aké, F.M.D., Derkaoui, M., Zébré, A.C., Cao, T.N., Bouraoui, H., Kentache, T., Mokhtari, A., Milohanic, E., Joyet, P., 2014. The Bacterial Phosphoenolpyruvate:Carbohydrate Phosphotransferase System: Regulation by Protein Phosphorylation and Phosphorylation-Dependent Protein-Protein Interactions. Microbiol. Mol. Biol. Rev. 78, 231–256.

Deutscher, J., Francke, C., Postma, P.W., 2006. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. Microbiol. Mol. Biol. Rev. 70, 939–1031.

Dobson, G., Christie, W.W., 2002. Mass spectrometry of fatty acid derivatives. Eur. J. Lipid Sci. Technol. 104, 36–43.

Dodds, E.D., McCoy, M.R., Rea, L.D., Kennish, J.M., 2005. Gas chromatographic quantification of fatty acid methyl esters: flame ionization detection vs. electron impact mass spectrometry. Lipids 40, 419–28.

Dragosits, M., Mattanovich, D., 2013. Adaptive laboratory evolution – principles and applications for biotechnology. Microb. Cell Fact. 12, 64.

Eagon, R.G., 1962. Pseudomonas natriegens, a marine bacterium with a generation time of less than 10 minutes. J. Bacteriol. 83, 736–737.

Edwards, J.S., Covert, M., 2002. Minireview Metabolic modelling of microbes : the flux-balance approach 4, 133–140.

Edwards, J.S., Palsson, B.O., 2000. The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc. Natl. Acad. Sci. U. S. A. 97, 5528–5533.

Enjalbert, B., Millard, P., Dinclaux, M., Portais, J.-C., Létisse, F., 2017. Acetate fluxes in Escherichia coli are determined by the thermodynamic control of the Pta-AckA pathway. Sci. Rep. 7, 42135.

Escalante, A., Cervantes, A.S., Gosset, G., Bolívar, F., 2012. Current knowledge of the Escherichia coli phosphoenolpyruvate-carbohydrate phosphotransferase system: Peculiarities of regulation and impact on growth and product formation. Appl. Microbiol. Biotechnol. 94, 1483–1494.

Eydallin, G., Viale, A.M., Morán-Zorzano, M.T., Muñoz, F.J., Montero, M., Baroja-Fernández, E., Pozueta-Romero, J., 2007. Genome-wide screening of genes affecting glycogen metabolism in Escherichia coli K-12. FEBS Lett. 581, 2947–53.

Fang, X., Sastry, A., Mih, N., Kim, D., Tan, J., Yurkovich, J.T., Lloyd, C.J., Gao, Y., Yang, L., Palsson, B.O., 2017. Global transcriptional regulatory network for Escherichia coli robustly connects gene expression to transcription factor activities. Proc. Natl. Acad. Sci. 201702581.

Feist, A.M., Henry, C.S., Reed, J.L., Krummenacker, M., Joyce, A.R., Karp, P.D., Broadbelt, L.J., Hatzimanikatis, V., Palsson, B.Ø., 2007. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol. Syst. Biol. 3, 121.

Fenton, A.W., Reinhart, G.D., 2009. Disentangling the Web of Allosteric Communication in a Homotetramer: Heterotropic Inhibition in Phosphofructokinase from Eschericia coli. Biochemistry 48, 12323–12328.

Fernandez, C. a, Des Rosiers, C., Previs, S.F., David, F., Brunengraber, H., 1996. Correction of 13C mass isotopomer distributions for natural stable isotope abundance. J. Mass Spectrom. 31, 255–62.

Fischer, E., Sauer, U., 2005. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. Nat. Genet. 37, 636–640.

Fischer, E., Sauer, U., 2003. Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS. Eur. J. Biochem. 270, 880–891.

Flores, N., Flores, S., Escalante, A., De Anda, R., Leal, L., Malpica, R., Georgellis, D., Gosset, G., Bolívar, F., 2005. Adaptation for fast growth on glucose by differential expression of central carbon metabolism and gal regulon genes in an Escherichia coli strain lacking the phosphoenolpyruvate:carbohydrate phosphotransferase system. Metab. Eng. 7, 70–87.

Flores, N., Xiao, J., Berry, A., Bolivar, F., Valle, F., 1996. Pathway engineering for the production of aromatic compounds in Escherichia coli. Nat. Biotechnol. 14, 620–623.

Fong, S.S., Nanchen, A., Palsson, B.O., Sauer, U., 2006. Latent pathway activation and increased pathway capacity enable Escherichia coli adaptation to loss of key metabolic enzymes. J. Biol. Chem. 281, 8024–8033.

Fong, S.S., Palsson, B.O., 2004. Metabolic gene-deletion strains of Escherichia coli evolve to computationally predicted growth phenotypes. Nat. Genet. 36, 1056–1058.

García Martín, H., Kumar, V.S., Weaver, D., Ghosh, A., Chubukov, V.,

Mukhopadhyay, A., Arkin, A., Keasling, J.D., 2015. A Method to Constrain Genome-Scale Models with 13C Labeling Data. PLOS Comput. Biol. 11, e1004363.

García Sánchez, C.E., Torres Sáez, R.G., 2014. Comparison and analysis of objective functions in flux balance analysis. Biotechnol. Prog. 30, 985–991.

Ghim, C.-M., Goh, K.-I., Kahng, B., 2005. Lethality and synthetic lethality in the genome-wide metabolic network of Escherichia coli. J. Theor. Biol. 237, 401–411.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C. a, Smith, H.O., Iii, C.A.H., America, N., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–5.

Gonzalez, J.E., Antoniewicz, M.R., 2017. Tracing metabolism from lignocellulosic biomass and gaseous substrates to products with stable-isotopes. Curr. Opin. Biotechnol. 43, 86–95.

Gonzalez, J.E., Long, C.P., Antoniewicz, M.R., 2017. Comprehensive analysis of glucose and xylose metabolism in Escherichia coli under aerobic and anaerobic conditions by 13C metabolic flux analysis. Metab. Eng. 39, 9–18.

Gopalakrishnan, S., Maranas, C.D., 2015. 13C Metabolic Flux Analysis at a Genome-Scale. Metab. Eng. 32, 12–22.

Gosset, G., 2005. Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate:sugar phosphotransferase system. Microb. Cell Fact. 4, 14.

Greenblum, S., Chiu, H.-C., Levy, R., Carr, R., Borenstein, E., 2013. Towards a predictive systems-level model of the human microbiome: progress, challenges, and opportunities. Curr. Opin. Biotechnol. 24, 810–820.

Gruer, M.J., Bradbury, A.J., Guest, J.R., 1997. Construction and properties of aconitase mutants of Escherichia coli. Microbiology 143, 1837–1846.

Guzmán, S., Marin, S., Miranda, A., Selivanov, V.A., Centelles, J.J., Harmancey, R., Smih, F., Turkieh, A., Durocher, Y., Zorzano, A., Rouet, P., Cascante, M., 2014. 13C metabolic flux analysis shows that resistin impairs the metabolic response to insulin in L6E9 myotubes. BMC Syst. Biol. 8, 109.

Hanke, T., Nöh, K., Noack, S., Polen, T., Bringer, S., Sahm, H., Wiechert, W., Bott, M., 2013. Combined fluxomics and transcriptomics analysis of glucose

catabolism via a partially cyclic pentose phosphate pathway in Gluconobacter oxydans 621H. Appl. Environ. Microbiol. 79, 2336–48.

Haverkorn van Rijsewijk, B.R.B., Nanchen, A., Nallet, S., Kleijn, R.J., Sauer, U., 2011. Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in Escherichia coli. Mol. Syst. Biol. 7, 477.

He, L., Xiao, Y., Gebreselassie, N., Zhang, F., Antoniewicz, M.R., Tang, Y.J., Peng, L., 2013. Central metabolic responses to the overproduction of fatty acids in Escherichia coli based on 13C-metabolic flux analysis. Biotechnol. Bioeng. 1–11.

Henry, C.S., Jankowski, M.D., Broadbelt, L.J., Hatzimanikatis, V., 2006. Genome-scale thermodynamic analysis of Escherichia coli metabolism. Biophys. J. 90, 1453–1461.

Hernández-Montalvo, V., Martínez, A., Hernández-Chavez, G., Bolivar, F., Valle, F., Gosset, G., 2003. Expression of galP and glk in a Escherichia coli PTS mutant restores glucose transport and increases glycolytic flux to fermentation products. Biotechnol. Bioeng. 83, 687–694.

Herring, C.D., Raghunathan, A., Honisch, C., Patel, T., Applebee, M.K., Joyce, A.R., Albert, T.J., Blattner, F.R., van den Boom, D., Cantor, C.R., Palsson, B.O., 2006. Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale. Nat Genet 38, 1406–1412.

Horinouchi, T., Tamaoka, K., Furusawa, C., Ono, N., Suzuki, S., Hirasawa, T., Yomo, T., Shimizu, H., 2010. Transcriptome analysis of parallel-evolved Escherichia coli strains under ethanol stress. BMC Genomics 11, 579.

Hua, Q., Yang, C., Baba, T., Mori, H., Shimizu, K., 2003. Responses of the Central Metabolism in Escherichia coli to Phosphoglucose Isomerase and Glucose-6-Phosphate Dehydrogenase Knockouts. J. Bacteriol. 185, 7053–7067.

Hutchison, C.A., Chuang, R.-Y., Noskov, V.N., Assad-Garcia, N., Deerinck, T.J., Ellisman, M.H., Gill, J., Kannan, K., Karas, B.J., Ma, L., Pelletier, J.F., Qi, Z.-Q., Richter, R.A., Strychalski, E.A., Sun, L., Suzuki, Y., Tsvetanova, B., Wise, K.S., Smith, H.O., Glass, J.I., Merryman, C., Gibson, D.G., Venter, J.C., 2016. Design and synthesis of a minimal bacterial genome. Science 351, aad6253.

Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P.Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T.,

Shimizu, K., Mori, H., Tomita, M., 2007. Multiple High-Throughput Analyses Monitor the Response of E. coli to Perturbations. Science 316, 593–597.

Jahan, N., Maeda, K., Matsuoka, Y., Sugimoto, Y., Kurata, H., 2016. Development of an accurate kinetic model for the central carbon metabolism of Escherichia coli. Microb. Cell Fact. 15, 112.

Jan, J., Martinez, I., Wang, Y., Bennett, G.N., San, K.Y., 2013. Metabolic engineering and transhydrogenase effects on NADPH availability in escherichia coli. Biotechnol. Prog. 29, 1124–1130.

Janssen, P., Goldovsky, L., Kunin, V., Darzentas, N., Ouzounis, C. a., 2005. Genome coverage, literally speaking. EMBO Rep. 6, 397–399.

Jensen, K.F., 1993. The Escherichia coli K-12 “wild types” W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. J. Bacteriol. 175, 3401–3407.

Kadir, T.A.A., Mannan, A.A., Kierzek, A.M., McFadden, J., Shimizu, K., 2010. Modeling and simulation of the main metabolism in Escherichia coli and its several single-gene knockout mutants with experimental verification. Microb. Cell Fact. 9, 88.

Kanehisa, M., Goto, S., 2000. KEGG: Kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 28, 27–30.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M., 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res. 40, 109–114.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., Tanabe, M., 2016. KEGG as a reference resource for gene and protein annotation. Nucleic Acids Res. 44, D457–D462.

Kappelmann, J., Wiechert, W., Noack, S., 2015. Cutting the Gordian Knot: Identifiability of anaplerotic reactions in Corynebacterium glutamicum by means of 13C-metabolic flux analysis. Biotechnol. Bioeng. 113, 661–674.

Karr, J.R., Sanghvi, J.C., Macklin, D.N., Gutschow, M. V, Jacobs, J.M., Bolival, B., Assad-garcia, N., Glass, J.I., Covert, M.W., 2012. A Whole-Cell Computational Model Predicts Phenotype from Genotype 389–401.

Khodayari, A., Maranas, C.D., 2016. A genome-scale Escherichia coli kinetic metabolic model satisfying flux data for multiple mutant strains. Nat. Commun.

7, 1–12.

Khodayari, A., Zomorrodi, A.R., Liao, J.C., Maranas, C.D., 2014. A kinetic model of Escherichia coli core metabolism satisfying multiple sets of mutant flux data. Metab. Eng. 25, 50–62.

Kim, H.U., Kim, T.Y., Lee, S.Y., 2011. Framework for network modularization and Bayesian network analysis to investigate the perturbed metabolic network. BMC Syst. Biol. 5 Suppl 2, S14.

Kim, H.U., Kim, W.J., Lee, S.Y., 2013. Flux-coupled genes and their use in metabolic flux analysis. Biotechnol. J. 8, 1035–42.

Kim, J., Reed, J.L., 2012. RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations. Genome Biol. 13, R78.

Kim, J., Reed, J.L., 2010. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. BMC Syst. Biol. 4, 53.

Kochanowski, K., Volkmer, B., Gerosa, L., Haverkorn van Rijsewijk, B.R., Schmidt, A., Heinemann, M., 2013. Functioning of a metabolic flux sensor in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 110, 1130–5.

Kotlarz, D., Garreau, H., Buc, H., 1975. Regulation of the amount and of the activity of phosphofructokinases and pyruvate kinases in Escherichia coli. Biochim. Biophys. Acta 381, 257–268.

Kotte, O., Zaugg, J.B., Heinemann, M., 2010. Bacterial adaptation through distributed sensing of metabolic fluxes. Mol. Syst. Biol. 6, 355.

Kremling, A., Bettenbrock, K., Gilles, E.D., 2007. Analysis of global control of Escherichia coli carbohydrate uptake. BMC Syst. Biol. 1, 42.

Krivoruchko, A., Zhang, Y., Siewers, V., Chen, Y., Nielsen, J., 2015. Microbial acetyl-CoA metabolism and metabolic engineering. Metab. Eng. 28, 28–42.

Kupor, S.R., Fraenkel, D.G., 1972. Glucose Metabolism in 6-Phosphogluconolactonase Mutants of Escherichia coli. J. Biol. Chem. 40, 1904–1911.

Kurata, H., Zhao, Q., Okuda, R., Shimizu, K., 2007. Integration of enzyme activities into metabolic flux distributions by elementary mode analysis. BMC Syst. Biol. 1, 31.

Kuznetsov, G., Goodman, D.B., Filsinger, G.T., Landon, M., Rohland, N., Aach, J., Lajoie, M.J., Church, G.M., 2017. Optimizing complex phenotypes through model-guided multiplex genome engineering. Genome Biol. 18, 100.

LaCroix, R.A., Sandberg, T.E., O'Brien, E.J., Utrilla, J., Ebrahim, A., Guzman, G.I., Szubin, R., Palsson, B.O., Feist, A.M., 2015. Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium. Appl. Environ. Microbiol. 81, 17–30.

Langholtz, M.H., Stokes, B.J., Eaton, L.M., 2016. 2016 Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy. Oak Ridge Natl. Lab.

Lee, D.H., Palsson, B.O., 2010. Adaptive evolution of escherichia coli K-12 MG1655 during growth on a nonnative carbon source, L-l,2-propanediol. Appl. Environ. Microbiol. 76, 4158–4168.

Lee, H.H., Ostrov, N., Wong, B.G., Gold, M.A., Khalil, A.S., Church, G.M., 2016. Vibrio natriegens, a new genomic powerhouse. bioRxiv.

Leighty, R.W., Antoniewicz, M.R., 2013. COMPLETE-MFA: Complementary parallel labeling experiments technique for metabolic flux analysis. Metab. Eng. 20, 49–55.

Leighty, R.W., Antoniewicz, M.R., 2012a. Parallel labeling experiments with [U-13C]glucose validate E. coli metabolic network model for 13C metabolic flux analysis. Metab. Eng. 14, 533–541.

Leighty, R.W., Antoniewicz, M.R., 2012b. Parallel labeling experiments with [U-^{13}C]glucose validate E. coli metabolic network model for ^{13}C metabolic flux analysis. Metab. Eng. 14, 533–41. doi:10.1016/j.ymben.2012.06.003

Lerman, J.A., Hyduke, D.R., Latif, H., Portnoy, V.A., Lewis, N.E., Orth, J.D., Schrimpe-Rutledge, A.C., Smith, R.D., Adkins, J.N., Zengler, K., Palsson, B.Ø., 2012. In silico method for modelling metabolism and gene product expression at genome scale. Nat. Commun. 3, 929.

Lewis, N.E., Nagarajan, H., Palsson, B.O., 2012. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. Nat. Rev. Microbiol. 10, 291–305.

Li, M., Ho, P.Y., Yao, S., Shimizu, K., 2006a. Effect of sucA or sucC gene knockout on the metabolism in Escherichia coli based on gene expressions, enzyme activities, intracellular metabolite concentrations and metabolic fluxes by 13C-

labeling experiments. Biochem. Eng. J. 30, 286–296.

Li, M., Ho, P.Y., Yao, S., Shimizu, K., 2006b. Effect of lpdA gene knockout on the metabolism in Escherichia coli based on enzyme activities, intracellular metabolite concentrations and metabolic flux analysis by 13C-labeling experiments. J. Biotechnol. 122, 254–266.

Li, R., Chen, Q., Wang, P.G., Qi, Q., 2007. A novel-designed Escherichia coli for the production of various polyhydroxyalkanoates from inexpensive substrate mixture. Appl. Microbiol. Biotechnol. 75, 1103–1109.

Li, Y., Beisson, F., Pollard, M., Ohlrogge, J., 2006. Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. Phytochemistry 67, 904–15.

Li, Y., Lin, Z., Huang, C., Zhang, Y., Wang, Z., Tang, Y. jie, Chen, T., Zhao, X., 2015. Metabolic engineering of Escherichia coli using CRISPR-Cas9 meditated genome editing. Metab. Eng. 31, 13–21.

Liang, Q., Zhang, F., Li, Y., Zhang, X., Li, J., Yang, P., Qi, Q., 2015. Comparison of individual component deletions in a glucose-specific phosphotransferase system revealed their different applications. Sci. Rep. 5, 13200.

Linton, K.J., Higgins, C.F., 1998. The Escherichia coli ATP-binding cassette (ABC) proteins. Mol. Microbiol. 28, 5–13.

Liu, J.K., O Brien, E.J., Lerman, J. a, Zengler, K., Palsson, B.O., Feist, A.M., 2014. Reconstruction and modeling protein translocation and compartmentalization in Escherichia coli at the genome-scale. BMC Syst. Biol. 8, 110.

Long, C.P., Antoniewicz, M.R., 2014a. Metabolic flux analysis of Escherichia coli knockouts: lessons from the Keio collection and future outlook. Curr. Opin. Biotechnol. 28, 127–133.

Long, C.P., Antoniewicz, M.R., 2014b. Quantifying Biomass Composition by Gas Chromatography/Mass Spectrometry. Anal. Chem. 86, 9423–7.

Long, C.P., Au, J., Gonzalez, J.E., Antoniewicz, M.R., 2016a. 13C metabolic flux analysis of microbial and mammalian systems is enhanced with GC–MS measurements of glycogen and RNA labeling. Metab. Eng. 38, 65–72.

Long, C.P., Au, J., Sandoval, N.R., Gebreselassie, N.A., Antoniewicz, M.R., 2017a. Enzyme I facilitates reverse flux from pyruvate to phosphoenolpyruvate in Escherichia coli. Nat. Commun. 8, 14316.

Long, C.P., Gonzalez, J.E., Cipolla, R.M., Antoniewicz, M.R., 2017. Metabolism of the fast-growing bacterium Vibrio natriegens elucidated by 13C metabolic flux analysis. Metab. Eng. 44, 191–197.

Long, C.P., Gonzalez, J.E., Feist, A.M., Palsson, B.O., Antoniewicz, M.R., 2017b. Fast growth phenotype of E. coli K-12 from adaptive laboratory evolution does not require intracellular flux rewiring. Metab. Eng. 44, 100–107. doi:10.1016/j.ymben.2017.09.012

Long, C.P., Gonzalez, J.E., Sandoval, N.R., Antoniewicz, M.R., 2016b. Characterization of physiological responses to 22 gene knockouts in Escherichia coli central carbon metabolism. Metab. Eng. 37, 102–113.

Lyngstadaas, A., Sprenger, G. a., Boye, E., 1998. Impaired growth of an Escherichia coli rpe mutant lacking ribulose-5-phosphate epimerase activity. Biochim. Biophys. Acta 1381, 319–330.

Mackie, A., Paley, S., Keseler, I.M., Shearer, A., Paulsen, I.T., Karp, P.D., 2014. Addition of Escherichia coli K-12 growth observation and gene essentiality data to the ecocyc database. J. Bacteriol. 196, 982–988.

Mahadevan, R., Schilling, C.H., 2003. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. Metab. Eng. 5, 264–276.

Maida, I., Bosi, E., Perrin, E., Papaleo, M.C., Orlandini, V., Fondi, M., Fani, R., Wiegel, J., Bianconi, G., Canganella, F., 2013. Complete Genome Sequence of the Fast-Growing Bacterium Vibrio natriegens Strain DSMZ 759. Genome Announc. 1, e00648-13.

Majewski, R., Domach, M., 1990. Simple constrained optimization view of acetate overflow in E.coli. Biotechnol. Bioeng. 35, 732–738.

Masakapalli, S.K., Ratcliffe, R.G., Williams, T.C.R., 2014. Quantification of ^{13}C enrichments and isotopomer abundances for metabolic flux analysis using 1D NMR spectroscopy. Methods Mol. Biol. 1090, 73–86.

McCloskey, D., Young, J.D., Xu, S., Palsson, B.O., Feist, A.M., 2016. Modeling Method for Increased Precision and Scope of Directly Measurable Fluxes at a Genome-Scale. Anal. Chem. 88, 3844–3852.

McConnell, B.O., Antoniewicz, M.R., 2016. Measuring the Composition and Stable-Isotope Labeling of Algal Biomass Carbohydrates via Gas Chromatography/Mass Spectrometry. Anal. Chem. 88, 4624–4628.

McKinlay, J.B., Shachar-Hill, Y., Zeikus, J.G., Vieille, C., 2007. Determining Actinobacillus succinogenes metabolic pathways and fluxes by NMR and GC-MS analyses of 13C-labeled metabolic product isotopomers. Metab. Eng. 9, 177–92.

Meza, E., Becker, J., Bolivar, F., Gosset, G., Wittmann, C., 2012. Consequences of phosphoenolpyruvate:sugar phosphotranferase system and pyruvate kinase isozymes inactivation in central carbon metabolism flux distribution in Escherichia coli. Microb. Cell Fact. 11, 127.

Miclet, E., Stoven, V., Michels, P.A.M., Opperdoes, F.R., Lallemand, J.Y., Duffieux, F., 2001. NMR Spectroscopic Analysis of the First Two Steps of the Pentose-Phosphate Pathway Elucidates the Role of 6-Phosphogluconolactonase. J. Biol. Chem. 276, 34840–34846.

Miranda-Santos, I., Gramacho, S., Pineiro, M., Martinez-Gomez, K., Fritz, M., Hollemeyer, K., Salvador, A., Heinzle, E., 2015. Mass isotopomer analysis of nucleosides isolated from RNA and DNA using GC/MS. Anal. Chem. 87, 617–623.

Monk, J.M., Koza, A., Campodonico, M.A., Machado, D., Seoane, J.M., Palsson, B.O., Herrgård, M.J., Feist, A.M., 2016. Multi-omics Quantification of Species Variation of Escherichia coli Links Molecular Features with Strain Phenotypes. Cell Syst. 238–251.

Mundhada, H., Seoane, J.M., Schneider, K., Koza, A., Christensen, H.B., Klein, T., Phaneuf, P. V., Herrgard, M., Feist, A.M., Nielsen, A.T., 2017. Increased production of L-serine in Escherichia coli through Adaptive Laboratory Evolution. Metab. Eng. 39, 141–150.

Murphy, T.A., Dang, C. V., Young, J.D., 2013. Isotopically nonstationary 13C flux analysis of Myc-induced metabolic reprogramming in B-cells. Metab. Eng. 15, 206–217.

Nakahigashi, K., Toya, Y., Ishii, N., Soga, T., Hasegawa, M., Watanabe, H., Takai, Y., Honma, M., Mori, H., Tomita, M., 2009. Systematic phenome analysis of Escherichia coli multiple-knockout mutants reveals hidden reactions in central carbon metabolism. Mol. Syst. Biol. 5, 306.

Nakashima, N., Ohno, S., Yoshikawa, K., Shimizu, H., Tamura, T., 2014. A vector library for silencing central carbon metabolism genes with antisense RNAs in escherichia coli. Appl. Environ. Microbiol. 80, 564–573.

Neidhardt, F.C., 1987. Escherichia coli and Salmonella Typhimurium. ASM Press.

Nicolas, C., Kiefer, P., Letisse, F., Krömer, J., Massou, S., Soucaille, P., Wittmann, C., Lindley, N.D., Portais, J.-C., 2007. Response of the central metabolism of Escherichia coli to modified expression of the gene encoding the glucose-6-phosphate dehydrogenase. FEBS Lett. 581, 3771–3776.

Nishikawa, T., Gulbahce, N., Motter, A.E., 2008. Spontaneous reaction silencing in metabolic optimization. PLoS Comput. Biol. 4, e1000236.

Niyas, A.M.M., Eiteman, M.A., 2017. Phosphatases and phosphate affect the formation of glucose from pentoses in Escherichia coli. Eng. Life Sci. 17, 579–584.

Nizam, S.A., Shimizu, K., 2008. Effects of arcA and arcB genes knockout on the metabolism in Escherichia coli under anaerobic and microaerobic conditions. Biochem. Eng. J. 42, 229–236.

Nizam, S.A., Zhu, J., Ho, P.Y., Shimizu, K., 2009. Effects of arcA and arcB genes knockout on the metabolism in Escherichia coli under aerobic condition. Biochem. Eng. J. 44, 240–250.

Noguchi, Y., Nakai, Y., Shimba, N., Toyosaki, H., Kawahara, Y., Sugimoto, S., Suzuki, E.I., 2004. The energetic conversion competence of Escherichia coli during aerobic respiration studied by 31P NMR using a circulating fermentation system. J. Biochem. 136, 509–515.

O'Brien, E.J., Lerman, J. a, Chang, R.L., Hyduke, D.R., Palsson, B.Ø., 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. 9, 693.

Paddon, C.J., Westfall, P.J., Pitera, D.J., Benjamin, K., Fisher, K., McPhee, D., Leavell, M.D., Tai, A., Main, A., Eng, D., Polichuk, D.R., Teoh, K.H., Reed, D.W., Treynor, T., Lenihan, J., Fleck, M., Bajad, S., Dang, G., Dengrove, D., Diola, D., Dorin, G., Ellens, K.W., Fickes, S., Galazzo, J., Gaucher, S.P., Geistlinger, T., Henry, R., Hepp, M., Horning, T., Iqbal, T., Jiang, H., Kizer, L., Lieu, B., Melis, D., Moss, N., Regentin, R., Secrest, S., Tsuruta, H., Vazquez, R., Westblade, L.F., Xu, L., Yu, M., Zhang, Y., Zhao, L., Lievense, J., Covello, P.S., Keasling, J.D., Reiling, K.K., Renninger, N.S., Newman, J.D., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. Nature 496, 528–32.

Park, J.M., Kim, T.Y., Lee, S.Y., 2010. Prediction of metabolic fluxes by incorporating genomic context and flux-converging pattern analyses. Proc. Natl. Acad. Sci. U. S. A. 107, 14931–14936.

Park, J.O., Rubin, S.A., Xu, Y., Amador-noguez, D., Fan, J., Shlomi, T., Rabinowitz, J.D., 2016. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat. Chem. Biol. 12, 482–489.

Payne, W.J., Eagon, R.G., Williams, A.K., 1961. Some observations on the physiology of Pseudomonas natriegens nov. spec. Antonie Van Leeuwenhoek 27, 121–128.

Perrenoud, A., Sauer, U., 2005. Impact of Global Transcriptional Regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on Glucose Catabolism in Escherichia coli. J. Bacteriol. 187, 3171–3179.

Peterson, G.L., 1977. A simplification of the protein assay method of Lowry et al. which is more generally applicable. Anal. Biochem. 83, 346–356.

Portnoy, V.A., Scott, D.A., Lewis, N.E., Tarasova, Y., Osterman, A.L., Palsson, B.O., 2010. Deletion of genes encoding cytochrome oxidases and quinol monooxygenase blocks the aerobic-anaerobic shift in Escherichia coli K-12 MG1655. Appl. Environ. Microbiol. 76, 6529–6540.

Postma, P.W., Lengeler, J.W., Jacobson, G.R., 1993. Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria. Microbiol. Rev. 57, 543–594.

Pramanik, J., Keasling, J.D., 1998. Effect of Escherichia coli biomass composition on central metabolic fluxes predicted by a stoichiometric model. Biotechnol. Bioeng. 60, 230–8.

Pramanik, J., Keasling, J.D., 1997. Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. Biotechnol. Bioeng. 56, 398–421.

Rahman, M., Shimizu, K., 2008. Altered acetate metabolism and biomass production in several Escherichia coli mutants lacking rpoS-dependent metabolic pathway genes. Mol. Biosyst. 4, 160–169.

Ramseier, T.M., 1996. Cra and the control of carbon flux via metabolic pathways. Res. Microbiol. 147, 489–493.

Ranganathan, S., Suthers, P.F., Maranas, C.D., 2010. OptForce: an optimization procedure for identifying all genetic manipulations leading to targeted overproductions. PLoS Comput. Biol. 6, e1000744.

Reddy, P., Kamireddi, M., 1998. Modulation of Escherichia coli adenylyl cyclase

activity by catalytic-site mutants of protein IIA(Glc) of the phosphoenolpyruvate: sugar phosphotransferase system. J Bacteriol 180, 732–736.

Reyes, L.H., Almario, M.P., Winkler, J., Orozco, M.M., Kao, K.C., 2012. Visualizing evolution in real time to determine the molecular mechanisms of n-butanol tolerance in Escherichia coli. Metab. Eng. 14, 579–590.

Reznik, E., Christodoulou, D., Goldford, J.E., Briars, E., Sauer, U., Segrè, D., Noor, E., 2017. Genome-Scale Architecture of Small Molecule Regulatory Networks and the Fundamental Trade-Off between Regulation and Enzymatic Activity. Cell Rep. 20, 2666–2677.

Richards, G.R., Patel, M. V., Lloyd, C.R., Vanderpool, C.K., 2013. Depletion of glycolytic intermediates plays a key role in glucose-phosphate stress in escherichia coli. J. Bacteriol. 195, 4816–4825.

Rühl, M., Rupp, B., Nöh, K., Wiechert, W., Sauer, U., Zamboni, N., 2012. Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of 13C metabolic flux analysis. Biotechnol. Bioeng. 109, 763–771.

Rutherford, S.M., Gilani, G.S., 2009. Amino Acid Analysis. Curr. Protoc. Protein Sci. 58, 11.9.1-11.9.37.

Sandberg, T.E., Lloyd, C.J., Palsson, B.O., Feist, A.M., 2017. Laboratory Evolution to Alternating Substrate Environments Yields Distinct Phenotypic and Genetic Adaptive Strategies. Appl. Environ. Microbiol. AEM.00410-17.

Sandberg, T.E., Long, C.P., Gonzalez, J.E., Feist, A.M., Antoniewicz, M.R., Palsson, B.O., 2016. Evolution of E. coli on [U-13C]Glucose Reveals a Negligible Isotopic Influence on Metabolism and Physiology. PLoS One 11, e0151130.

Sandberg, T.E., Pedersen, M., Lacroix, R.A., Ebrahim, A., Bonde, M., Herrgard, M.J., Palsson, B.O., Sommer, M., Feist, A.M., 2014. Evolution of Escherichia coli to 42 °C and subsequent genetic engineering reveals adaptive mechanisms and novel mutations. Mol. Biol. Evol. 31, 2647–2662.

Sauer, U., Canonaco, F., Heri, S., Perrenoud, A., Fischer, E., 2004. The Soluble and Membrane-bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of Escherichia coli. J. Biol. Chem. 279, 6613–6619.

Schellenberger, J., Que, R., Fleming, R.M.T., Thiele, I., Orth, J.D., Feist, A.M., Zielinski, D.C., Bordbar, A., Lewis, N.E., Rahmanian, S., Kang, J., Hyduke, D.R., Palsson, B.Ø., 2011. Quantitative prediction of cellular metabolism with

constraint-based models: the COBRA Toolbox v2.0. Nat. Protoc. 6, 1290–307.

Schuetz, R., Kuepfer, L., Sauer, U., 2007. Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli. Mol. Syst. Biol. 3, 119.

Segre, D., Vitkup, D., Church, G.M., 2002. Analysis of optimality in natural and perturbed metabolic networks. Proc. Natl. Acad. Sci. U. S. A. 99, 15112–15117.

Shlomi, T., Berkman, O., Ruppin, E., 2005. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. Proc. Natl. Acad. Sci. U. S. A. 102, 7695–7700.

Shlomi, T., Eisenberg, Y., Sharan, R., Ruppin, E., 2007. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. Mol. Syst. Biol. 3, 101.

Shokri, A., Sandén, A.M., Larsson, G., 2002. Growth rate-dependent changes in Escherichia coli membrane structure and protein leakage. Appl. Microbiol. Biotechnol. 58, 386–392.

Siddiquee, K.A.Z., Arauzo-Bravo, M.J., Shimizu, K., 2004a. Metabolic flux analysis of pykF gene knockout Escherichia coli based on 13C-labeling experiments together with measurements of enzyme activities and intracellular metabolite concentrations. Appl. Microbiol. Biotechnol. 63, 407–417.

Siddiquee, K.A.Z., Arauzo-Bravo, M.J., Shimizu, K., 2004b. Effect of a pyruvate kinase (pykF-gene) knockout mutation on the control of gene expression and metabolic fluxes in Escherichia coli. FEMS Microbiol. Lett. 235, 25–33.

Smith, P., Krohn, R., Hermanson, G., Mallia, A., Gartner, F., Provenzano, M., Fujimoto, E., Goeke, N., Olson, B., Klenk, D., 1985. Measurement of protein using bicinchoninic acid. Anal. Biochem. 150, 76–85.

Song, H.-S., Ramkrishna, D., 2012. Prediction of dynamic behavior of mutant strains from limited wild-type data. Metab. Eng. 14, 69–80.

Stephanopoulos, G., 1999. Metabolic fluxes and metabolic engineering. Metab. Eng. 1, 1–11.

Stephanopoulos, G., Aristidou, A., Nielsen, J., 1998. Metabolic Engineering: Principles and Methodologies.

Swarup, A., Lu, J., DeWoody, K.C., Antoniewicz, M.R., 2014. Metabolic network reconstruction, growth characterization and 13C-metabolic flux analysis of the

extremophile Thermus thermophilus HB8. Metab. Eng. 24, 173–180.

Tan, Y., Liao, J.C., 2012. Metabolic ensemble modeling for strain engineers. Biotechnol. J. 7, 343–53.

Tang, Y.J., Martin, H.G., Dehal, P.S., Deutschbauer, A., Llora, X., Meadows, A., Arkin, A., Keasling, J.D., 2009. Metabolic flux analysis of shewanella spp. Reveals evolutionary robustness in central carbon metabolism. Biotechnol. Bioeng. 102, 1161–1169.

Taymaz-Nikerel, H., Borujeni, A.E., Verheijen, P.J.T., Heijnen, J.J., van Gulik, W.M., 2010. Genome-derived minimal metabolic models for Escherichia coli MG1655 with estimated in vivo respiratory ATP stoichiometry. Biotechnol. Bioeng. 107, 369–381.

Tenaillon, O., Rodriguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D., Gaut, B.S., 2012. The Molecular Diversity of Adaptive Convergence. Science (80-.). 335, 457–461.

Thomason, L.C., Court, D.L., Datta, A.R., Khanna, R., Rosner, J.L., 2004. Identification of the Escherichia coli K-12 ybhE Gene as pgl, Encoding 6-Phosphogluconolactonase. J. Bacteriol. 186, 8248–8253.

Tötemeyer, S., Booth, N.A., Nichols, W.W., Dunbar, B., Booth, I.R., 1998. From famine to feast: The role of methylglyoxal production in Escherichia coli. Mol. Microbiol. 27, 553–562.

Toya, Y., Ishii, N., Nakahigashi, K., Hirasawa, T., Soga, T., Tomita, M., Shimizu, K., 2010. 13C-metabolic flux analysis for batch culture of Escherichia coli and its Pyk and Pgi gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. Biotechnol. Prog. 26, 975–992.

Tran, L.M., Rizk, M.L., Liao, J.C., 2008. Ensemble modeling of metabolic networks. Biophys. J. 95, 5606–5617.

Trauchessec, M., Jaquinod, M., Bonvalot, A., Brun, V., Bruley, C., Ropers, D., de Jong, H., Garin, J., Bestel-Corre, G., Ferro, M., 2014. Mass Spectrometry-based Workflow for Accurate Quantification of Escherichia coli Enzymes: How Proteomics Can Play a Key Role in Metabolic Engineering. Mol. Cell. Proteomics 13, 954–968.

Truong, Q.X., Yoon, J.M., Shanks, J. V, 2014. Isotopomer measurement techniques in metabolic flux analysis I: nuclear magnetic resonance. Methods Mol. Biol. 1083, 65–83.

Usui, Y., Hirasawa, T., Furusawa, C., Shirai, T., Yamamoto, N., Mori, H., Shimizu, H., 2012. Investigating the effects of perturbations to pgi and eno gene expression on central carbon metabolism in Escherichia coli using 13C metabolic flux analysis. Microb. Cell Fact. 11, 87.

Utrilla, J., O'Brien, E.J., Chen, K., McCloskey, D., Cheung, J., Wang, H., Armenta-Medina, D., Feist, A.M., Palsson, B.O., 2016. Global Rebalancing of Cellular Resources by Pleiotropic Point Mutations Illustrates a Multi-scale Mechanism of Adaptive Evolution. Cell Syst. 2, 260–271.

Valente, K.N., Lenhoff, A.M., Lee, K.H., 2015. Expression of difficult-to-remove host cell protein impurities during extended Chinese hamster ovary cell culture and their impact on continuous bioprocessing. Biotechnol. Bioeng. 112, 1232–1242.

Waegeman, H., Beauprez, J., Moens, H., Maertens, J., De Mey, M., Foulquié-Moreno, M.R., Heijnen, J.J., Charlier, D., Soetaert, W., 2011. Effect of iclR and arcA knockouts on biomass formation and metabolic fluxes in Escherichia coli K12 and its implications on understanding the metabolism of Escherichia coli BL21 (DE3). BMC Microbiol. 11.

Waegeman, H., Maertens, J., Beauprez, J., De Mey, M., Soetaert, W., 2012. Effect of iclR and arcA deletions on physiology and metabolic fluxes in Escherichia coli BL21 (DE3). Biotechnol. Lett. 34, 329–337.

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., Church, G.M., 2009. Programming cells by multiplex genome engineering and accelerated evolution. Nature 460, 894–8.

Wang, Z., Lin, B., Iv, W.J.H., Vora, G.J., 2013. Draft Genome Sequence of the Fast-Growing Marine Bacterium Vibrio natriegens Strain ATCC 14048. Genome Announc. 1, e00589-13.

Wawrik, B., Harriman, B.H., 2010. Rapid, colorimetric quantification of lipid from algal cultures. J. Microbiol. Methods 80, 262–6.

Weigel, N., Waygood, E.B., Kakuruzinska, M.A., Nakazawa, A., Roseman, S., 1982. Sugar Transport by the Bacterial Phosphotransferase System. J. Biol. Chem. 257, 14526–14537.

Weinert, B., Iesmantavicius, V., Wagner, S., Schölz, C., Gummesson, B., Beli, P., Nyström, T., Choudhary, C., 2013. Acetyl-Phosphate is a critical determinant of Lysine Acetylation in E.coli. Mol. Cell 51, 265–272.

Weinstock, M.T., Hesek, E.D., Wilson, C.M., Gibson, D.G., 2016. Vibrio natriegens

as a fast-growing host for molecular biology. Nat. Methods 13, 1–39.

Whitaker, W.B., Jones, J.A., Bennett, R.K., Gonzalez, J.E., Vernacchio, V.R., Collins, S.M., Palmer, M.A., Schmidt, S., Antoniewicz, M.R., Koffas, M.A., Papoutsakis, E.T., 2017. Engineering the biological conversion of methanol to specialty chemicals in Escherichia coli. Metab. Eng. 39, 49–59.

Wiechert, W., 2001. 13C metabolic flux analysis. Metab. Eng. 3, 195–206.

Wilhelm, T., Behre, J., Schuster, S., 2004. Analysis of structural robustness of metabolic networks. Syst. Biol. (Stevenage). 1, 114–120.

Wiser, M.J., Ribeck, N., Lenski, R.E., 2013. Long-term dynamics of adaptation in asexual populations. Science (80-.). 342, 1364–1367.

Wolfe, A.J., 2005. The acetate switch. Microbiol. Mol. Biol. Rev. 69, 12–50.

Wu, S.G., Wang, Y., Jiang, W., Oyetunde, T., Yao, R., Zhang, X., Shimizu, K., Tang, Y.J., Bao, F.S., 2016. Rapid Prediction of Bacterial Heterotrophic Fluxomics Using Machine Learning and Constraint Programming. PLoS Comput. Biol. 12.

Wunderlich, Z., Mirny, L., 2006. Using the topology of metabolic networks to predict viability of mutant strains. Biophys. J. 91, 2304–2311.

Xia, T., Han, Q., Costanzo, W. V., Zhu, Y., Urbauer, J.L., Eiteman, M. a., 2015. Accumulation of d-Glucose from Pentoses by Metabolically Engineered Escherichia coli. Appl. Environ. Microbiol. 81, 3387–3394.

Yao, M., Elling, F.J., Jones, C., Nomosatryo, S., Long, C.P., Crowe, S.A., Antoniewicz, M.R., Hinrichs, K.-U., Maresca, J.A., 2015. Heterotrophic bacteria from an extremely phosphate-poor lake have conditionally reduced phosphorus demand and utilize diverse sources of phosphorus. Environ. Microbiol. 1, n/a-n/a.

Yeang, C.-H., Vingron, M., 2006. A joint model of regulatory and metabolic networks. BMC Bioinformatics 7, 332.

Yoo, H., Antoniewicz, M.R., Stephanopoulos, G., Kelleher, J.K., 2008. Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line. J. Biol. Chem. 283, 20621–7.

Yoo, H., Stephanopoulos, G., Kelleher, J.K., 2004. Quantifying carbon sources for de novo lipogenesis in wild-type and IRS-1 knockout brown adipocytes. J. Lipid Res. 45, 1324–1332.

Young, J.D., 2014. 13C metabolic flux analysis of recombinant expression hosts. Curr.

Opin. Biotechnol. 30, 238–245.

Young, J.D., Henne, K.L., Morgan, J.A., Konopka, A.E., Ramkrishna, D., 2008. Integrating cybernetic modeling with pathway analysis provides a dynamic, systems-level description of metabolic control. Biotechnol. Bioeng. 100, 542–559.

Zamboni, N., Fendt, S.-M., Rühl, M., Sauer, U., 2009. 13C-based metabolic flux analysis. Nat. Protoc. 4, 878–892.

Zhang, Y.-M., Rock, C.O., 2008. Membrane lipid homeostasis in bacteria. Nat. Rev. Microbiol. 6, 222–233.

Zhao, J., Baba, T., Mori, H., Shimizu, K., 2004. Global metabolic response of Escherichia coli to gnd or zwf gene-knockout, based on 13C-labeling experiments and the measurement of enzyme activities. Appl. Microbiol. Biotechnol. 64, 91–98.

Zheng, D., Constantinidou, C., Hobman, J.L., Minchin, S.D., 2004. Identification of the CRP regulon using in vitro and in vivo transcriptional profiling. Nucleic Acids Res. 32, 5874–5893.

Zhu, J., Shimizu, K., 2005. Effect of a single-gene knockout on the metabolic regulation in Escherichia coli for D-lactate production under microaerobic condition. Metab. Eng. 7, 104–115.

Zuñiga, C., Li, C.-T., Huelsman, T., Levering, J., Zielinski, D.C., McConnell, B.O., Long, C.P., Knoshaug, E.P., Guarnieri, M.T., Antoniewicz, M.R., Betenbaugh, M.J., Zengler, K., 2016. Genome-Scale Metabolic Model for the Green Alga Chlorella vulgaris UTEX 395 Accurately Predicts Phenotypes under Autotrophic, Heterotrophic, and Mixotrophic Growth Conditions. Plant Physiol. 172, 589–602.

Appendix A

SUPPLEMENTARY MATERIAL FOR CHAPTER 2

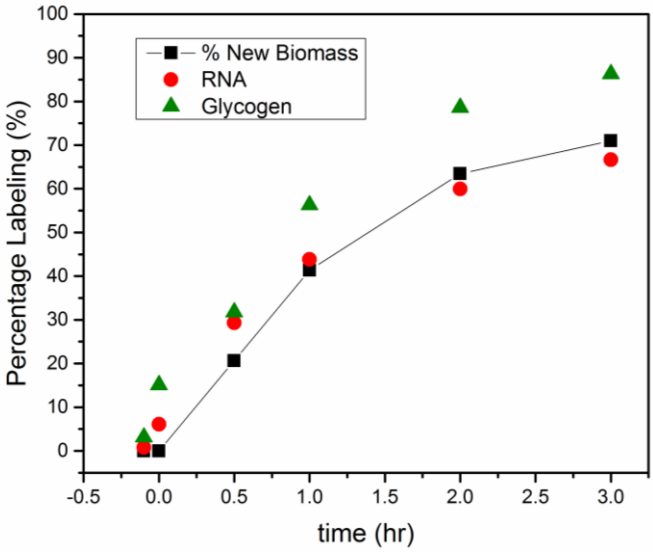


Figure A.1 Following a switch from unlabeled to labeled substrate, the measured labeling of RNA and glycogen tracked with the percentage of new biomass from growth, demonstrating that low-turnover biopolymers (i.e., RNA and glycogen) were measured.

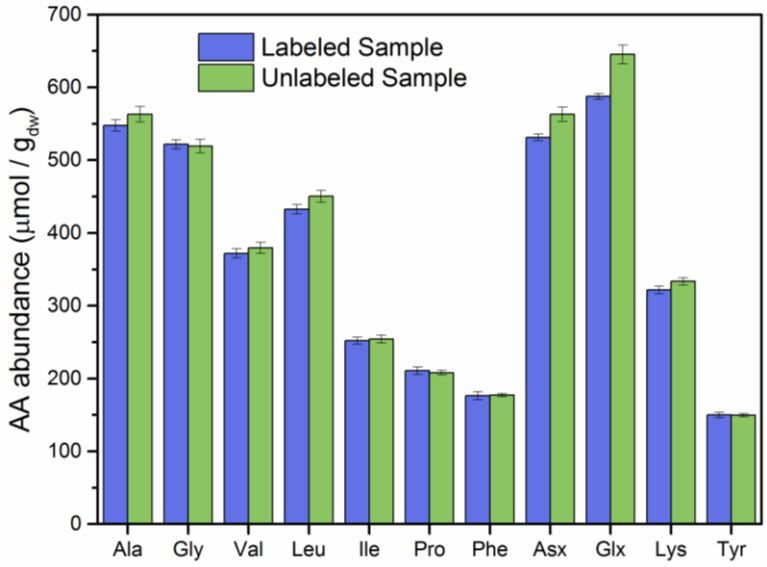


Figure A.2 The abundance of amino acids in E. coli biomass as measured by both approaches (labeled sample and unlabeled standard, and vice versa). Error bars indicate typical measurement uncertainty for both approaches.

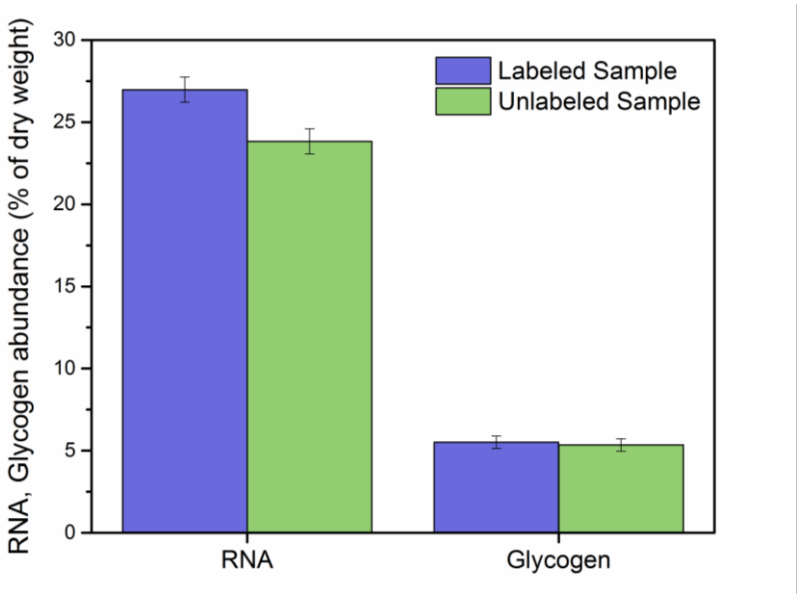


Figure A.3 The abundance of RNA and glycogen in E. coli biomass as measured by both approaches. Error bars indicate typical measurement uncertainty.

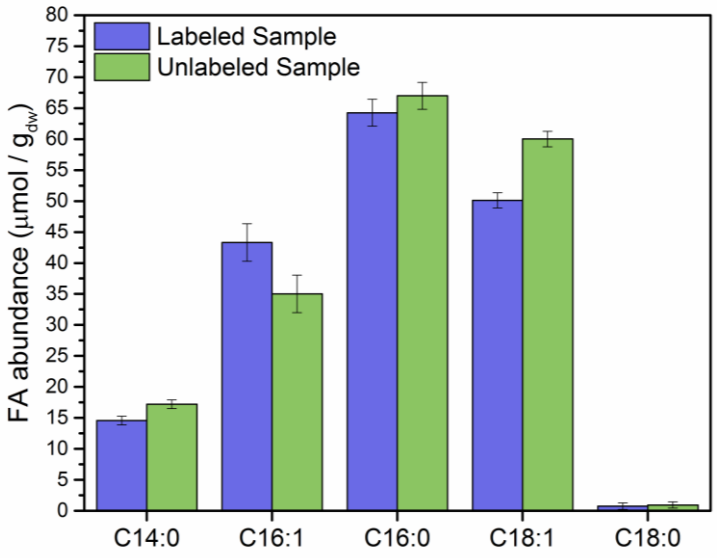


Figure A.4 The abundance of fatty acids in E. coli biomass as measured by both approaches. Error bars indicate typical measurement uncertainty for both approaches.

Appendix B

SUPPLEMENTARY MATERIAL FOR CHAPTER 3

Table B.1: Metabolic Network Model for ^{13}C-MFA of *E. coli* (full model)

Glycolysis

v_1 Gluc.ext (abcdef) + PEP (ghi) → G6P (abcdef) + Pyr (ghi)
v_2 G6P (abcdef) ↔ F6P (abcdef)
v_3 F6P (abcdef) + ATP → FBP (abcdef)
v_4 FBP (abcdef) ↔ DHAP (cba) + GAP (def)
v_5 DHAP (abc) ↔ GAP (abc)
v_6 GAP (abc) ↔ 3PG (abc) + ATP + NADH
v_7 3PG (abc) ↔ PEP (abc)
v_8 PEP (abc) → Pyr (abc) + ATP

Pentose Phosphate Pathway

v_9 G6P (abcdef) → 6PG (abcdef) + NADPH
v_{10} 6PG (abcdef) → Ru5P (bcdef) + CO2 (a) + NADPH
v_{11} Ru5P (abcde) ↔ X5P (abcde)
v_{12} Ru5P (abcde) ↔ R5P (abcde)
v_{13} X5P (abcde) ↔ TK-C2 (ab) + GAP (cde)
v_{14} F6P (abcdef) ↔ TK-C2 (ab) + E4P (cdef)
v_{15} S7P (abcdefg) ↔ TK-C2 (ab) + R5P (cdefg)
v_{16} F6P (abcdef) ↔ TA-C3 (abc) + GAP (def)
v_{17} S7P (abcdefg) ↔ TA-C3 (abc) + E4P (defg)

Entner-Doudoroff Pathway

v_{18} 6PG (abcdef) → KDPG (abcdef)
v_{19} KDPG (abcdef) → Pyr (abc) + GAP (def)

TCA Cycle

v_{20} Pyr (abc) → AcCoA (bc) + CO2 (a) + NADH
v_{21} OAC (abcd) + AcCoA (ef) → Cit (dcbfea)

v_{22} Cit (abcdef) ↔ ICit (abcdef)
v_{23} ICit (abcdef) ↔ AKG (abcde) + CO2 (f) + NADPH
v_{24} AKG (abcde) → SucCoA (bcde) + CO2 (a) + NADH
v_{25} SucCoA (abcd) ↔ Suc (½ abcd + ½ dcba) + ATP
v_{26} Suc (½ abcd + ½ dcba) ↔ Fum (½ abcd + ½ dcba) + FADH2
v_{27} Fum (½ abcd + ½ dcba) ↔ Mal (abcd)
v_{28} Mal (abcd) ↔ OAC (abcd) + NADH

Glyoxylate Shunt
v_{29} ICit (abcdef) ↔ Glyox (ab) + Suc (½ edcf + ½ fcde)
v_{30} Glyox (ab) + AcCoA (cd) → Mal (abdc)

Amphibolic Reactions
v_{31} Mal (abcd) → Pyr (abc) + CO2 (d) + NADPH
v_{32} Mal (abcd) → Pyr (abc) + CO2 (d) + NADH
v_{33} PEP (abc) + CO2 (d) → OAC (abcd)
v_{34} OAC (abcd) + ATP → PEP (abc) + CO2 (d)

Acetic Acid Formation
v_{35} AcCoA (ab) ↔ Ac (ab) + ATP

Amino Acid Biosynthesis
v_{36} AKG (abcde) + NADPH + NH3 → Glu (abcde)
v_{37} Glu (abcde) + ATP + NH3 → Gln (abcde)
v_{38} Glu (abcde) + ATP + 2 NADPH → Pro (abcde)
v_{39} Glu (abcde) + CO2 (f) + Gln (ghijk) + Asp (lmno) + AcCoA (pq) + 5 ATP + NADPH →
Arg (abcdef) + AKG (ghijk) + Fum (lmno) + Ac (pq)
v_{40} OAC (abcd) + Glu (efghi) → Asp (abcd) + AKG (efghi)
v_{41} Asp (abcd) + 2 ATP + NH3 → Asn (abcd)
v_{42} Pyr (abc) + Glu (defgh) → Ala (abc) + AKG (defgh)
v_{43} 3PG (abc) + Glu (defgh) → Ser (abc) + AKG (defgh) + NADH
v_{44} Ser (abc) ↔ Gly (ab) + MEETHF (c)
v_{45} Gly (ab) ↔ CO2 (a) + MEETHF (b) + NADH + NH3
v_{46} Thr (abcd) → Gly (ab) + AcCoA (cd) + NADH
v_{47} Ser (abc) + AcCoA (de) + 3 ATP + 4 NADPH + SO4 → Cys (abc) + Ac (de)
v_{48} Asp (abcd) + Pyr (efg) + Glu (hijkl) + SucCoA (mnop) + ATP + 2 NADPH →
LL-DAP (½ abcdgfe + ½ efgdcba) + AKG (hijkl) + Suc (½ mnop + ½ ponm)
v_{49} LL-DAP (½ abcdefg + ½ gfedcba) → Lys (abcdef) + CO2 (g)

v_{50} Asp (abcd) + 2 ATP + 2 NADPH → Thr (abcd)

v_{51} Asp (abcd) + METHF (e) + Cys (fgh) + SucCoA (ijkl) + ATP + 2 NADPH → Met (abcde) + Pyr (fgh) + Suc (½ ijkl + ½ lkji) + NH3

v_{52} Pyr (abc) + Pyr (def) + Glu (ghijk) + NADPH → Val (abcef) + CO2 (d) + AKG (ghijk)

v_{53} AcCoA (ab) + Pyr (cde) + Pyr (fgh) + Glu (ijklm) + NADPH → Leu (abdghe) + CO2 (c) + CO2 (f) + AKG (ijklm) + NADH

v_{54} Thr (abcd) + Pyr (efg) + Glu (hijkl) + NADPH → Ile (abfcdg) + CO2 (e) + AKG (hijkl) + NH3

v_{55} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Phe (abcefghij) + CO2 (d) + AKG (klmno)

v_{56} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Tyr (abcefghij) + CO2 (d) + AKG (klmno) + NADH

v_{57} Ser (abc) + R5P (defgh) + PEP (ijk) + E4P (lmno) + PEP (pqr) + Gln (stuvw) + 3 ATP + NADPH →
Trp (abcedklmnoj) + CO2 (i) + GAP (fgh) + Pyr (pqr) + Glu (stuvw)

v_{58} R5P (abcde) + FTHF (f) + Gln (ghijk) + Asp (lmno) + 5 ATP →
His (edcbaf) + AKG (ghijk) + Fum (lmno) + 2 NADH

One-Carbon Metabolism

v_{59} MEETHF (a) + NADH → METHF (a)

v_{60} MEETHF (a) → FTHF (a) + NADPH

Oxidative Phosphorylation

v_{61} NADH + ½ O2 → 2 ATP

v_{62} FADH2 + ½ O2 → 1 ATP

Transhydrogenation

v_{63} NADH ↔ NADPH

ATP Hydrolysis

v_{64} ATP → ATP:ext

Transport

v_{65} Ac (ab) → Ac.ext (ab)

v_{66} CO2 (a) → CO2.ext (a)

v_{67} O2.ext → O2

v_{68} NH3.ext → NH3

v_{69} SO4.ext → SO4

Biomass Formation

v_{70} 0.470 Ala + 0.281 Arg + 0.236 Asn + 0.236 Asp + 0.087 Cys + 0.280 Glu + 0.280 Gln + 0.432 Gly + 0.082 His + 0.215 Ile + 0.377 Leu + 0.279 Lys + 0.107 Met + 0.153 Phe + 0.178 Pro + 0.23 Ser + 0.244 Thr + 0.054 Trp + 0.131 Tyr + 0.315 Val + 0.270 G6P + 0.071 F6P + 0.851 R5P + 0.081 GAP + 0.628 3PG + 0.051 PEP + 0.083 Pyr + 1.690 AcCoA + 0.087 AKG + 0.380 OAC + 0.500 MEETHF + 33.601 ATP + 3.948 NADPH → 37.12 Biomass + 1.578 NADH

CO_2 Exchange

v_{71} CO_2.unlabeled (a) + CO_2 (b) → CO_2 (a) + CO_2.out (b)

Table B.2: Metabolic network model of ^{13}C-MFA of *E. coli* (upper metabolism)

Glycolysis

v_1 Gluc.ext (abcdef) + PEP (ghi) → G6P (abcdef) + Pyr (ghi)
v_2 G6P (abcdef) ↔ F6P (abcdef)
v_3 F6P (abcdef) → FBP (abcdef)
v_4 FBP (abcdef) ↔ DHAP (cba) + GAP (def)
v_5 DHAP (abc) ↔ GAP (abc)
v_6 GAP (abc) ↔ 3PG (abc)
v_7 3PG (abc) ↔ PEP (abc)
v_8 PEP (abc) → Pyr (abc)

Pentose Phosphate Pathway

v_9 G6P (abcdef) → 6PG (abcdef)
v_{10} 6PG (abcdef) → Ru5P (bcdef) + CO2 (a)
v_{11} Ru5P (abcde) ↔ X5P (abcde)
v_{12} Ru5P (abcde) ↔ R5P (abcde)
v_{13} X5P (abcde) ↔ TK-C2 (ab) + GAP (cde)
v_{14} F6P (abcdef) ↔ TK-C2 (ab) + E4P (cdef)
v_{15} S7P (abcdefg) ↔ TK-C2 (ab) + R5P (cdefg)
v_{16} F6P (abcdef) ↔ TA-C3 (abc) + GAP (def)
v_{17} S7P (abcdefg) ↔ TA-C3 (abc) + E4P (defg)

Entner-Doudoroff Pathway

v_{18} 6PG (abcdef) → KDPG (abcdef)
v_{19} KDPG (abcdef) → Pyr (abc) + GAP (def)

Xylose Metabolism

v_{20} Xyl.ext (abcde) → Xyl (abcde)

v_{21} Xyl (abcde) → X5P (abcde)

Out-fluxes

v_{22} PEP (abc) + PEP (def) + E4P (ghij) → Phe (abcefghij) + CO2 (d)

v_{23} Pyr → Lower Metabolism

v_{24} 2.040 G6P + 0.536 F6P + 7.463 R5P + 0.612 GAP + 1.400 E4P + 1.156 Phe → Biomass

Table B.3: Metabolic network model of ^{13}C-MFA of CHO cells (upper metabolism)

Glycolysis

v_1 Gluc.ext (abcdef) + PEP (ghi) → G6P (abcdef) + Pyr (ghi)

v_2 G6P (abcdef) ↔ F6P (abcdef)

v_3 F6P (abcdef) → FBP (abcdef)

v_4 FBP (abcdef) ↔ DHAP (cba) + GAP (def)

v_5 DHAP (abc) ↔ GAP (abc)

v_6 GAP (abc) ↔ 3PG (abc)

v_7 3PG (abc) ↔ PEP (abc)

Pentose Phosphate Pathway

v_8 G6P (abcdef) → 6PG (abcdef)

v_9 6PG (abcdef) → Ru5P (bcdef) + CO2 (a)

v_{10} Ru5P (abcde) ↔ X5P (abcde)

v_{11} Ru5P (abcde) ↔ R5P (abcde)

v_{12} X5P (abcde) ↔ TK-C2 (ab) + GAP (cde)

v_{13} F6P (abcdef) ↔ TK-C2 (ab) + E4P (cdef)

v_{14} S7P (abcdefg) ↔ TK-C2 (ab) + R5P (cdefg)

v_{15} F6P (abcdef) ↔ TA-C3 (abc) + GAP (def)

v_{16} S7P (abcdefg) ↔ TA-C3 (abc) + E4P (defg)

Out-fluxes

v_{17} PEP → Lower Metabolism

v_{18} 1.21 G6P + 1.01 R5P + 0.47 DHAP → Biomass

Appendix C

SUPPLEMENTARY MATERIAL FOR CHAPTER 7

Table C.1: Metabolic Network Model for Ch. 7

Glycolysis

v_1 Gluc.ext (abcdef) + PEP (ghi) → G6P (abcdef) + Pyr (ghi)
v_2 G6P (abcdef) ↔ F6P (abcdef)
v_3 F6P (abcdef) + ATP → FBP (abcdef)
v_4 FBP (abcdef) → F6P (abcdef) + Pi
v_5 FBP (abcdef) ↔ DHAP (cba) + GAP (def)
v_6 DHAP (abc) ↔ GAP (abc)
v_7 GAP (abc) ↔ 3PG (abc) + ATP + NADH
v_8 3PG (abc) ↔ PEP (abc)
v_9 PEP (abc) ↔ Pyr (abc) + ATP

Pentose Phosphate Pathway

v_{10} G6P (abcdef) → 6PG (abcdef) + NADPH
v_{11} 6PG (abcdef) → Ru5P (bcdef) + CO2 (a) + NADPH
v_{12} Ru5P (abcde) ↔ X5P (abcde)
v_{13} Ru5P (abcde) ↔ R5P (abcde)
v_{14} X5P (abcde) ↔ TK-C2 (ab) + GAP (cde)
v_{15} F6P (abcdef) ↔ TK-C2 (ab) + E4P (cdef)
v_{16} S7P (abcdefg) ↔ TK-C2 (ab) + R5P (cdefg)
v_{17} F6P (abcdef) ↔ TA-C3 (abc) + GAP (def)
v_{18} S7P (abcdefg) ↔ TA-C3 (abc) + E4P (defg)

Entner-Doudoroff Pathway

v_{19} 6PG (abcdef) → KDPG (abcdef)
v_{20} KDPG (abcdef) → Pyr (abc) + GAP (def)

TCA Cycle

v_{21} Pyr (abc) → AcCoA (bc) + CO2 (a) + NADH

v_{22} OAC (abcd) + AcCoA (ef) → Cit (dcbfea)
v_{23} Cit (abcdef) ↔ ICit (abcdef)
v_{24} ICit (abcdef) → AKG (abcde) + CO2 (f) + NADPH
v_{25} AKG (abcde) + CO2.int (f) + NADPH → ICit (abcdef)
v_{26} AKG (abcde) → SucCoA (bcde) + CO2 (a) + NADH
v_{27} SucCoA (bcde) + CO2.int (a) + NADH → AKG (abcde)
v_{28} SucCoA (abcd) ↔ Suc (½ abcd + ½ dcba) + ATP
v_{29} Suc (½ abcd + ½ dcba) ↔ Fum (½ abcd + ½ dcba) + FADH2
v_{30} Fum (½ abcd + ½ dcba) ↔ Mal (abcd)
v_{31} Mal (abcd) ↔ OAC (abcd) + NADH

Glyoxylate Shunt
v_{32} ICit (abcdef) ↔ Glyox (ab) + Suc (½ edcf + ½ fcde)
v_{33} Glyox (ab) + AcCoA (cd) → Mal (abdc)

Amphibolic Reactions
v_{34} Mal (abcd) → Pyr (abc) + CO2 (d) + NADPH
v_{35} Mal (abcd) → Pyr (abc) + CO2 (d) + NADH
v_{36} PEP (abc) + CO2.int (d) → OAC (abcd)
v_{37} OAC (abcd) + ATP → PEP (abc) + CO2 (d)

Acetic Acid Formation
v_{38} AcCoA (ab) ↔ Ac (ab) + ATP

Amino Acid Biosynthesis
v_{39} AKG (abcde) + NADPH + NH3 → Glu (abcde)
v_{40} Glu (abcde) + ATP + NH3 → Gln (abcde)
v_{41} Glu (abcde) + ATP + 2 NADPH → Pro (abcde)
v_{42} Glu (abcde) + CO2 (f) + Gln (ghijk) + Asp (lmno) + AcCoA (pq) + 5 ATP + NADPH →
Arg (abcdef) + AKG (ghijk) + Fum (lmno) + Ac (pq)
v_{43} OAC (abcd) + Glu (efghi) → Asp (abcd) + AKG (efghi)
v_{44} Asp (abcd) + 2 ATP + NH3 → Asn (abcd)
v_{45} Pyr (abc) + Glu (defgh) → Ala (abc) + AKG (defgh)
v_{46} 3PG (abc) + Glu (defgh) → Ser (abc) + AKG (defgh) + NADH
v_{47} Ser (abc) ↔ Gly (ab) + MEETHF (c)
v_{48} Gly (ab) → CO2 (a) + MEETHF (b) + NADH + NH3
v_{49} CO2.int (a) + MEETHF (b) + NADH + NH3 → Gly (ab)
v_{50} Thr (abcd) ↔ Gly (ab) + AcCoA (cd) + NADH

v_{51} Ser (abc) + AcCoA (de) + 3 ATP + 4 NADPH + SO4 → Cys (abc) + Ac (de)

v_{52} Asp (abcd) + Pyr (efg) + Glu (hijkl) + SucCoA (mnop) + ATP + 2 NADPH → LL-DAP (½ abcdgfe + ½ efgdcba) + AKG (hijkl) + Suc (½ mnop + ½ ponm)

v_{53} LL-DAP (½ abcdefg + ½ gfedcba) → Lys (abcdef) + CO2 (g)

v_{54} Asp (abcd) + 2 ATP + 2 NADPH → Thr (abcd)

v_{55} Asp (abcd) + METHF (e) + Cys (fgh) + SucCoA (ijkl) + ATP + 2 NADPH → Met (abcde) + Pyr (fgh) + Suc (½ ijkl + ½ lkji) + NH3

v_{56} Pyr (abc) + Pyr (def) + Glu (ghijk) + NADPH → Val (abcef) + CO2 (d) + AKG (ghijk)

v_{57} AcCoA (ab) + Pyr (cde) + Pyr (fgh) + Glu (ijklm) + NADPH → Leu (abdghe) + CO2 (c) + CO2 (f) + AKG (ijklm) + NADH

v_{58} Thr (abcd) + Pyr (efg) + Glu (hijkl) + NADPH → Ile (abfcdg) + CO2 (e) + AKG (hijkl) + NH3

v_{59} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Phe (abcefghij) + CO2 (d) + AKG (klmno)

v_{60} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Tyr (abcefghij) + CO2 (d) + AKG (klmno) + NADH

v_{61} Ser (abc) + R5P (defgh) + PEP (ijk) + E4P (lmno) + PEP (pqr) + Gln (stuvw) + 3 ATP + NADPH → Trp (abcedklmnoj) + CO2 (i) + GAP (fgh) + Pyr (pqr) + Glu (stuvw)

v_{62} R5P (abcde) + FTHF (f) + Gln (ghijk) + Asp (lmno) + 5 ATP → His (edcbaf) + AKG (ghijk) + Fum (lmno) + 2 NADH

v_{63} Ser (abc) → Pyr (abc) + NH3

One-Carbon Metabolism

v_{64} MEETHF (a) + NADH → METHF (a)

v_{65} MEETHF (a) → FTHF (a) + NADPH

Oxidative Phosphorylation

v_{66} NADH + ½ O2 → 2 ATP

v_{67} FADH2 + ½ O2 → 1 ATP

Transhydrogenation

v_{68} NADH ↔ NADPH

ATP Hydrolysis

v_{69} ATP → ATP:ext

Transport

v_{70} Ac (ab) → Ac.ext (ab)

v_{71} CO2 (a) → CO2.ext (a)
v_{72} O2.ext → O2
v_{73} NH3.ext → NH3
v_{74} SO4.ext → SO4

Biomass Formation

v_{75} Ala + Arg + Asn + Asp + Cys + Glu + Gln + Gly + His + Ile + Leu + Lys + Met + Phe + Pro + Ser + Thr + Trp + Tyr + Val + G6P + F6P + R5P + GAP + 3PG + PEP + Pyr + AcCoA + AKG + OAC + MEETHF + ATP + NADPH → Biomass + NADH
*strain-specific coefficients, based on the measured biomass composition, were used

CO_2 Exchange

v_{76} CO2.16G (a) → CO2.int (a) + X1
v_{77} CO2.12G (a) → CO2.int (a) + X1
v_{78} CO2.unlabeled (a) → CO2.int (a) + X1
v_{79} CO2.int (a) → CO2.out

Additional Reactions

ΔpfkA (glucose excretion)
v_{80} G6P (abcdef) → Glucose.out (abcdef)

ΔtpiA (methylglyoxal pathway)
v_{80} DHAP (abc) → Mglx (abc)
v_{81} Mglx (abc) → Lact (abc)
v_{82} Lact (abc) ↔ Pyr (abc)

SUPPLEMENTARY TEXT FOR CHAPTER 7

Physiological measurements

Physiological characteristics of these strains, including growth rate, yields, and biomass composition, were previously reported (Long et al., 2016b). Here, the growth rates from the tracer experiments were found to be consistent with the reported values, and used to update the values (Fig. S2). The acetate yield of *ΔrpiB* was updated. The biomass yield estimated from ^{13}C-MFA was used here, including in the calculation of glucose uptake rate. These values agreed well with the previously reported, directly measured values, assuming a constant dry weight to OD_{600} conversion of 0.32 g/L/OD_{600}.

Calculation of glucose secretion

Let c(t) [mM] be the concentration of medium glucose, X(t) be the labeling of medium glucose, X_{glg} be the labeling of intracellular G6P (i.e., glucose moiety of glycogen), v_1 be the rate of cellular glucose uptake (e.g., mM/h), and v_2 be the rate of glucose secretion.

$$\frac{dc}{dt} = v_2 - v_1 \qquad \text{(Eq. 1)}$$

$$\frac{d(c*X)}{dt} = v_2 * X_{glg} - v_1 * X \qquad \text{(Eq. 2)}$$

After substitution:

$$\frac{dX}{dt} = \frac{v_2}{c} * (X_{glg} - X) \qquad \text{(Eq. 3)}$$

Dividing (Eq. 1) by (Eq. 3):

$$\frac{dc}{dX} = \left(\frac{v_2 - v_1}{v_2}\right) * \left(\frac{c}{X_{glg} - X}\right) \quad \text{(Eq. 4)}$$

After integration:

$$\ln\left(\frac{X_{glg} - X}{X_{glg} - X_{t=0}}\right) = \alpha * \ln\left(\frac{c}{c_{t=0}}\right) \quad \text{(Eq. 5)}$$

where $\alpha = v_2/(v_1-v_2)$

Regression of measured values allows for the estimation of α, and relative secretion is calculated as:

$v_2/v_1 = \alpha/(1-\alpha)$

Random flux map generation (for Fig. 4B)

Random flux maps were generated as described previously (Crown et al., 2016b). The random flux maps captured a wide range of flux scenarios, with glycolysis flux ranging from 10 to 100 (normalized to glucose uptake of 100), oxPPP flux ranging 0 to 80, ED flux ranging 0 to 80, TCA cycle ranging 16 to 180, glyoxylate shunt ranging 0 to 30, acetate secretion ranging 0 to 70, and growth rate from 0.01 to 0.80 h^{-1}.

SUPPLEMENTARY FIGURES FOR CHAPTER 7

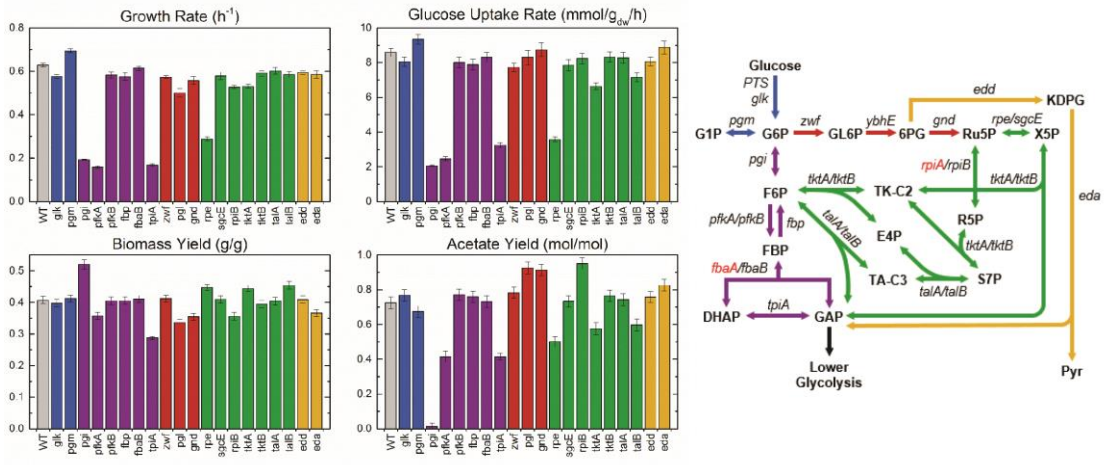


Figure C.1: Physiological characteristics of wild-type and 20 knockouts. Consensus growth rate (average, standard error of the mean) were calculated by combining previously reported measurement (Long et al., 2016b) and the tracer experiments performed in this study (see Fig. S2). Acetate yield was previously reported (Long et al., 2016b), with the value for *ΔrpiB* updated here. Biomass yields were estimated via ^{13}C-MFA here (standard errors shown), in good agreement with directly measured values (Long et al., 2016b) where a constant dry weight per OD_{600} is used (0.32 $g_{DW}/L/OD_{600}$). Glucose uptake rates were calculated as growth rate divided by biomass yield.

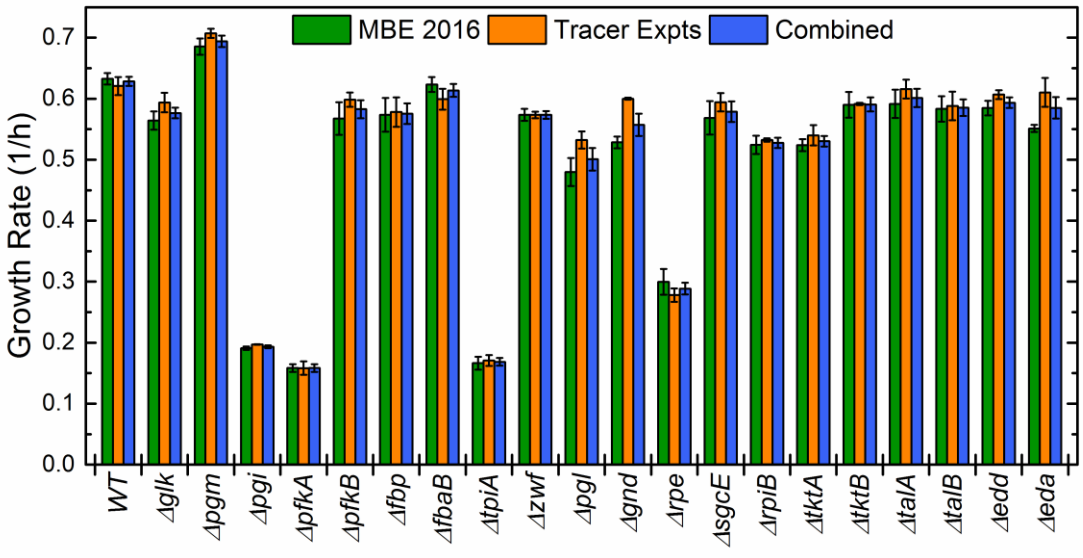


Figure C.2: Consensus measured growth rates for all strains. Growth rates were previously reported (Long et al., 2016b) (n=3), and are compared to those measured during the tracer experiments reported in this study (typically n=2; where additional tracer experiments were performed, e.g. for *ΔpfkA*, these were also included). Agreement was excellent, and all measurements were combined to calculate consensus growth rate values (blue bars, average and standard error of the mean shown).

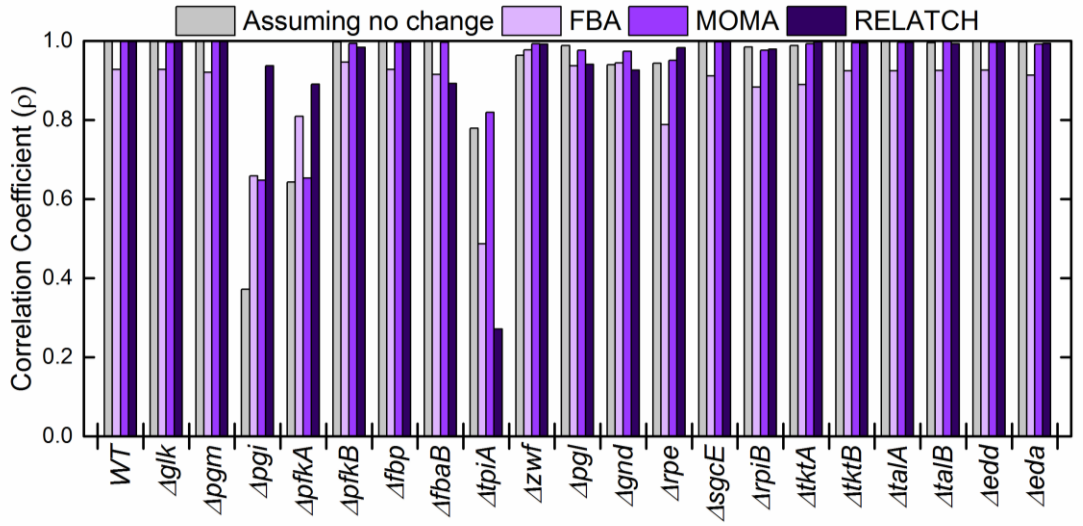


Figure C.3: Correlation coefficients of predictions and measurements for eleven key, normalized, intracellular fluxes. This is as shown in Fig. 4F, with values for all strains explicit. The gray shows the correlation of knockout values to the wild-type values (i.e., a trivial model assuming no change). This serves as a benchmark for judging COBRA methods, as fluxes change very little across many of the knockout strains. In the wild-type and these strains, FBA scored well but not perfectly (ρ=0.93), with the biggest disagreement being that FBA predicts higher oxPPP usage. MOMA and RELATCH use the measured wild-type fluxes as the reference, so agreement is excellent in the cases of minimal change. The cases of more substantial departure from the wild-type are highlighted in Fig. 7.7.

Appendix D

SUPPLEMENTARY MATERIAL FOR CHAPTER 9

Table D.1: Network model for ^{13}C-MFA.

Glycolysis

v_1 Gluc.ext (abcdef) + PEP (ghi) → G6P (abcdef) + Pyr (ghi)
v_2 G6P (abcdef) ↔ F6P (abcdef)
v_3 F6P (abcdef) + ATP ↔ FBP (abcdef)
v_4 FBP (abcdef) ↔ DHAP (cba) + GAP (def)
v_5 DHAP (abc) ↔ GAP (abc)
v_6 GAP (abc) ↔ 3PG (abc) + ATP + NADH
v_7 3PG (abc) ↔ PEP (abc)
v_8 PEP (abc) ↔ Pyr (abc) + ATP

Pentose Phosphate Pathway

v_9 G6P (abcdef) → 6PG (abcdef) + NADPH
v_{10} 6PG (abcdef) → Ru5P (bcdef) + CO2 (a) + NADPH
v_{11} Ru5P (abcde) ↔ X5P (abcde)
v_{12} Ru5P (abcde) ↔ R5P (abcde)
v_{13} X5P (abcde) ↔ TK-C2 (ab) + GAP (cde)
v_{14} F6P (abcdef) ↔ TK-C2 (ab) + E4P (cdef)
v_{15} S7P (abcdefg) ↔ TK-C2 (ab) + R5P (cdefg)
v_{16} F6P (abcdef) ↔ TA-C3 (abc) + GAP (def)
v_{17} S7P (abcdefg) ↔ TA-C3 (abc) + E4P (defg)

Entner-Doudoroff Pathway

v_{18} 6PG (abcdef) → KDPG (abcdef)
v_{19} KDPG (abcdef) → Pyr (abc) + GAP (def)

TCA Cycle

v_{20} Pyr (abc) → AcCoA (bc) + CO2 (a) + NADH
v_{21} OAC (abcd) + AcCoA (ef) → Cit (dcbfea)

v_{22} Cit (abcdef) ↔ ICit (abcdef)
v_{23} ICit (abcdef) → AKG (abcde) + CO2 (f) + NADPH
v_{24} AKG (abcde) + CO2.int (f) + NADPH → ICit (abcdef)
v_{25} AKG (abcde) → SucCoA (bcde) + CO2 (a) + NADH
v_{26} SucCoA (bcde) + CO2.int (a) + NADH → AKG (abcde)
v_{27} SucCoA (abcd) ↔ Suc (½ abcd + ½ dcba) + ATP
v_{28} Suc (½ abcd + ½ dcba) ↔ Fum (½ abcd + ½ dcba) + FADH2
v_{29} Fum (½ abcd + ½ dcba) ↔ Mal (abcd)
v_{30} Mal (abcd) ↔ OAC (abcd) + NADH

Glyoxylate Shunt
v_{31} ICit (abcdef) ↔ Glyox (ab) + Suc (½ edcf + ½ fcde)
v_{32} Glyox (ab) + AcCoA (cd) → Mal (abdc)

Amphibolic Reactions
v_{33} Mal (abcd) → Pyr (abc) + CO2 (d) + NADPH
v_{34} Mal (abcd) → Pyr (abc) + CO2 (d) + NADH
v_{35} PEP (abc) + CO2.int (d) → OAC (abcd)
v_{36} OAC (abcd) + ATP → PEP (abc) + CO2 (d)

Product Formation
v_{37} AcCoA (ab) ↔ Ac (ab) + ATP
v_{38} Pyr (abc) + NADH → Lact (abc)

Amino Acid Biosynthesis
v_{39} AKG (abcde) + NADPH + NH3 → Glu (abcde)
v_{40} Glu (abcde) + ATP + NH3 → Gln (abcde)
v_{41} Glu (abcde) + ATP + 2 NADPH → Pro (abcde)
v_{42} Glu (abcde) + CO2 (f) + Gln (ghijk) + Asp (lmno) + AcCoA (pq) + 5 ATP + NADPH →
Arg (abcdef) + AKG (ghijk) + Fum (lmno) + Ac (pq)
v_{43} OAC (abcd) + Glu (efghi) → Asp (abcd) + AKG (efghi)
v_{44} Asp (abcd) + 2 ATP + NH3 → Asn (abcd)
v_{45} Pyr (abc) + Glu (defgh) → Ala (abc) + AKG (defgh)
v_{46} 3PG (abc) + Glu (defgh) → Ser (abc) + AKG (defgh) + NADH
v_{47} Ser (abc) ↔ Gly (ab) + MEETHF (c)
v_{48} Gly (ab) → CO2 (a) + MEETHF (b) + NADH + NH3
v_{49} CO2.int (a) + MEETHF (b) + NADH + NH3 → Gly (ab)
v_{50} Thr (abcd) ↔ Gly (ab) + AcCoA (cd) + NADH

v_{51} Ser (abc) + AcCoA (de) + 3 ATP + 4 NADPH + SO4 → Cys (abc) + Ac (de)

v_{52} Asp (abcd) + Pyr (efg) + Glu (hijkl) + SucCoA (mnop) + ATP + 2 NADPH → LL-DAP (½ abcdgfe + ½ efgdcba) + AKG (hijkl) + Suc (½ mnop + ½ ponm)

v_{53} LL-DAP (½ abcdefg + ½ gfedcba) → Lys (abcdef) + CO2 (g)

v_{54} Asp (abcd) + 2 ATP + 2 NADPH → Thr (abcd)

v_{55} Asp (abcd) + METHF (e) + Cys (fgh) + SucCoA (ijkl) + ATP + 2 NADPH → Met (abcde) + Pyr (fgh) + Suc (½ ijkl + ½ lkji) + NH3

v_{56} Pyr (abc) + Pyr (def) + Glu (ghijk) + NADPH → Val (abcef) + CO2 (d) + AKG (ghijk)

v_{57} AcCoA (ab) + Pyr (cde) + Pyr (fgh) + Glu (ijklm) + NADPH → Leu (abdghe) + CO2 (c) + CO2 (f) + AKG (ijklm) + NADH

v_{58} Thr (abcd) + Pyr (efg) + Glu (hijkl) + NADPH → Ile (abfcdg) + CO2 (e) + AKG (hijkl) + NH3

v_{59} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Phe (abcefghij) + CO2 (d) + AKG (klmno)

v_{60} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Tyr (abcefghij) + CO2 (d) + AKG (klmno) + NADH

v_{61} Ser (abc) + R5P (defgh) + PEP (ijk) + E4P (lmno) + PEP (pqr) + Gln (stuvw) + 3 ATP + NADPH → Trp (abcedklmnoj) + CO2 (i) + GAP (fgh) + Pyr (pqr) + Glu (stuvw)

v_{62} R5P (abcde) + FTHF (f) + Gln (ghijk) + Asp (lmno) + 5 ATP → His (edcbaf) + AKG (ghijk) + Fum (lmno) + 2 NADH

v_{63} Ser (abc) → Pyr (abc) + NH3

One-Carbon Metabolism

v_{64} MEETHF (a) + NADH → METHF (a)

v_{65} MEETHF (a) → FTHF (a) + NADPH

Oxidative Phosphorylation

v_{66} NADH + ½ O2 → 2 ATP

v_{67} FADH2 + ½ O2 → 1 ATP

Transhydrogenation

v_{68} NADH ↔ NADPH

ATP Hydrolysis

v_{69} ATP → ATP:ext

Transport

v_{70} CO2 (a) → CO2.ext (a)

v_{71} O2.ext → O2
v_{72} NH3.ext → NH3
v_{73} SO4.ext → SO4
v_{74} Ac (ab) → Ac.ext (ab)
v_{75} Lact (abc) → Lact.ext (abc)
v_{76} Pyr (abc) → Pyr.ext (abc)
v_{77} Cit (abc) → Cit.ext (abc)
v_{78} Suc (abc) → Suc.ext (abc)

Biomass Formation

v_{79} Ala + Arg + Asn + Asp + Cys + Glu + Gln + Gly + His + Ile + Leu + Lys + Met + Phe + Pro + Ser + Thr + Trp + Tyr + Val + G6P + F6P + R5P + GAP + 3PG + PEP + Pyr + AcCoA + AKG + OAC + MEETHF + ATP + NADPH → Biomass + NADH
*strain-specific coefficients, based on the measured biomass composition, were used

CO_2 Exchange

v_{81} CO2.16G (a) → CO2.int (a) + X1
v_{82} CO2.12G (a) → CO2.int (a) + X1
v_{83} CO2.unlabeled (a) → CO2.int (a) + X1
v_{84} CO2.int (a) → CO2.out

Additional Reactions

ΔaceE (aspartate/fumarate exchange)

v_{80} Asp (abcd) → Fum (abcd)

Supplementary Figures for Chapter 9

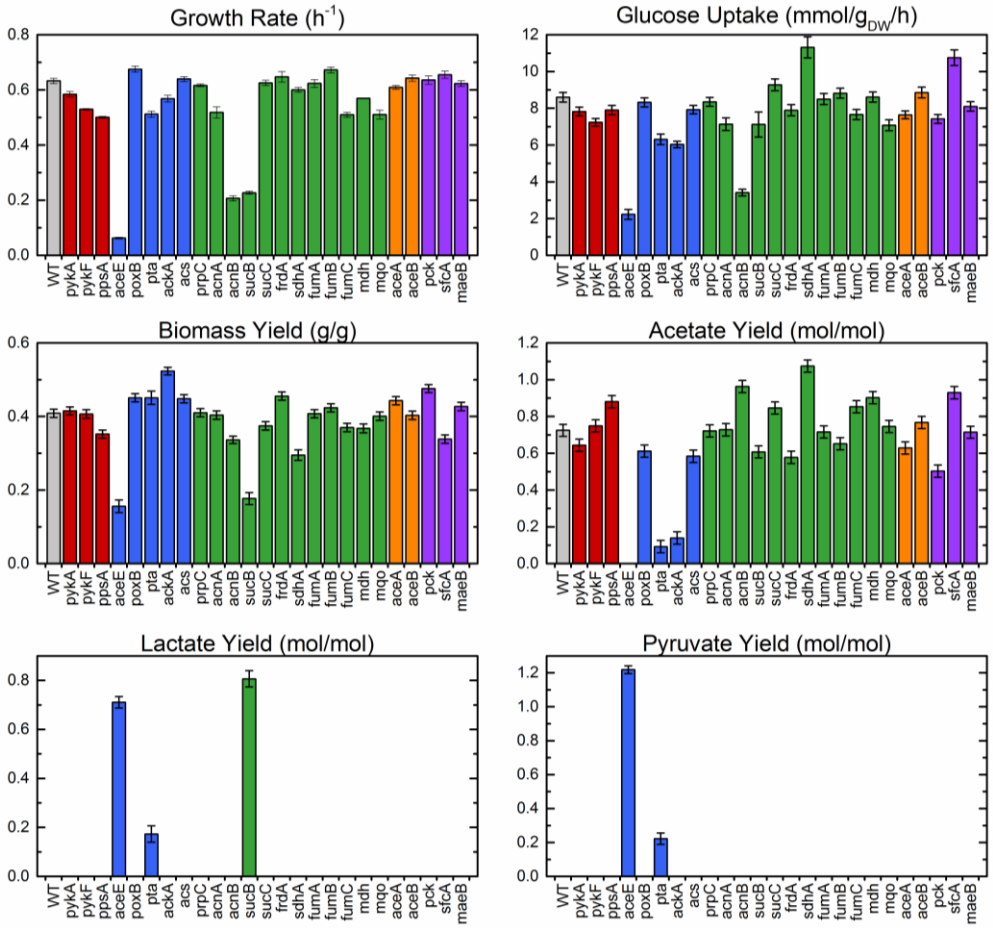


Figure D.1 Physiologies of wild-type and 25 knockout strains from lower central carbon metabolism, as estimated by ^{13}C-MFA.

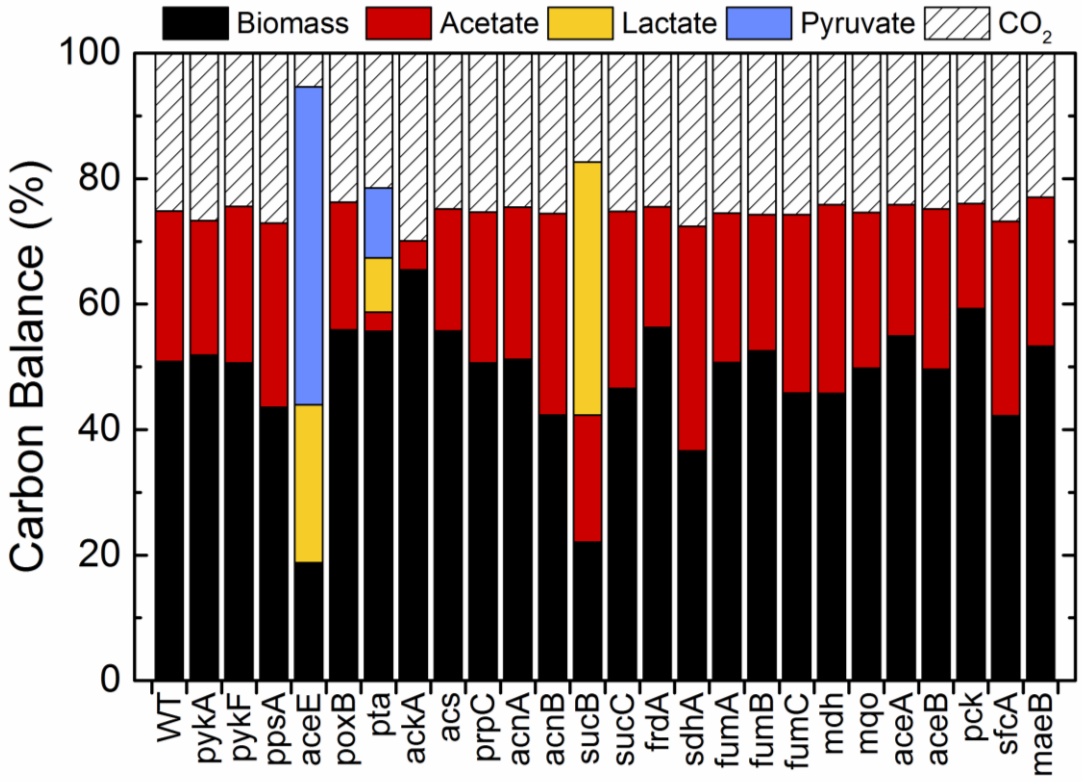


Figure D.2: Carbon balance as estimated by ^{13}C-MFA.

Appendix E

SUPPLEMENTARY MATERIAL FOR CHAPTER 10

Table E.1: Metabolic network model of *E. coli* used for ^{13}C-MFA

Glycolysis

v_1 Gluc.ext (abcdef) + PEP (ghi) → G6P (abcdef) + Pyr (ghi)
v_2 G6P (abcdef) ↔ F6P (abcdef)
v_3 F6P (abcdef) + ATP ↔ FBP (abcdef)
v_4 FBP (abcdef) ↔ DHAP (cba) + GAP (def)
v_5 DHAP (abc) ↔ GAP (abc)
v_6 GAP (abc) ↔ 3PG (abc) + ATP + NADH
v_7 3PG (abc) ↔ PEP (abc)
v_8 PEP (abc) ↔ Pyr (abc) + ATP

Pentose Phosphate Pathway

v_9 G6P (abcdef) → 6PG (abcdef) + NADPH
v_{10} 6PG (abcdef) → Ru5P (bcdef) + CO2 (a) + NADPH
v_{11} Ru5P (abcde) ↔ X5P (abcde)
v_{12} Ru5P (abcde) ↔ R5P (abcde)
v_{13} X5P (abcde) ↔ TK-C2 (ab) + GAP (cde)
v_{14} F6P (abcdef) ↔ TK-C2 (ab) + E4P (cdef)
v_{15} S7P (abcdefg) ↔ TK-C2 (ab) + R5P (cdefg)
v_{16} F6P (abcdef) ↔ TA-C3 (abc) + GAP (def)
v_{17} S7P (abcdefg) ↔ TA-C3 (abc) + E4P (defg)

Entner-Doudoroff Pathway

v_{18} 6PG (abcdef) → KDPG (abcdef)
v_{19} KDPG (abcdef) → Pyr (abc) + GAP (def)

TCA Cycle

v_{20} Pyr (abc) → AcCoA (bc) + CO2 (a) + NADH
v_{21} OAC (abcd) + AcCoA (ef) → Cit (dcbfea)

v_{22} Cit (abcdef) ↔ ICit (abcdef)
v_{23} ICit (abcdef) ↔ AKG (abcde) + CO2 (f) + NADPH
v_{24} AKG (abcde) → SucCoA (bcde) + CO2 (a) + NADH
v_{25} SucCoA (bcde) + CO2.int (a) + NADH → AKG (abcde)
v_{26} SucCoA (abcd) ↔ Suc (½ abcd + ½ dcba) + ATP
v_{27} Suc (½ abcd + ½ dcba) ↔ Fum (½ abcd + ½ dcba) + FADH2
v_{28} Fum (½ abcd + ½ dcba) ↔ Mal (abcd)
v_{29} Mal (abcd) ↔ OAC (abcd) + NADH

Glyoxylate Shunt
v_{30} ICit (abcdef) ↔ Glyox (ab) + Suc (½ edcf + ½ fcde)
v_{31} Glyox (ab) + AcCoA (cd) → Mal (abdc)

Amphibolic Reactions
v_{32} Mal (abcd) → Pyr (abc) + CO2 (d) + NADPH
v_{33} Mal (abcd) → Pyr (abc) + CO2 (d) + NADH
v_{34} PEP (abc) + CO2.int (d) → OAC (abcd)
v_{35} OAC (abcd) + ATP → PEP (abc) + CO2 (d)

Acetic Acid Formation
v_{36} AcCoA (ab) ↔ Ac (ab) + ATP

Amino Acid Biosynthesis
v_{37} AKG (abcde) + NADPH + NH3 → Glu (abcde)
v_{38} Glu (abcde) + ATP + NH3 → Gln (abcde)
v_{39} Glu (abcde) + ATP + 2 NADPH → Pro (abcde)
v_{40} Glu (abcde) + CO2 (f) + Gln (ghijk) + Asp (lmno) + AcCoA (pq) + 5 ATP + NADPH →
Arg (abcdef) + AKG (ghijk) + Fum (lmno) + Ac (pq)
v_{41} OAC (abcd) + Glu (efghi) → Asp (abcd) + AKG (efghi)
v_{42} Asp (abcd) + 2 ATP + NH3 → Asn (abcd)
v_{43} Pyr (abc) + Glu (defgh) → Ala (abc) + AKG (defgh)
v_{44} 3PG (abc) + Glu (defgh) → Ser (abc) + AKG (defgh) + NADH
v_{45} Ser (abc) ↔ Gly (ab) + MEETHF (c)
v_{46} Gly (ab) → CO2 (a) + MEETHF (b) + NADH + NH3
v_{47} CO2.int (a) + MEETHF (b) + NADH + NH3 → Gly (ab)
v_{48} Thr (abcd) ↔ Gly (ab) + AcCoA (cd) + NADH
v_{49} Ser (abc) + AcCoA (de) + 3 ATP + 4 NADPH + SO4 → Cys (abc) + Ac (de)
v_{50} Asp (abcd) + Pyr (efg) + Glu (hijkl) + SucCoA (mnop) + ATP + 2 NADPH →

LL-DAP (½ abcdgfe + ½ efgdcba) + AKG (hijkl) + Suc (½ mnop + ½ ponm)

v_{51} LL-DAP (½ abcdefg + ½ gfedcba) → Lys (abcdef) + CO2 (g)

v_{52} Asp (abcd) + 2 ATP + 2 NADPH → Thr (abcd)

v_{53} Asp (abcd) + METHF (e) + Cys (fgh) + SucCoA (ijkl) + ATP + 2 NADPH → Met (abcde) + Pyr (fgh) + Suc (½ ijkl + ½ lkji) + NH3

v_{54} Pyr (abc) + Pyr (def) + Glu (ghijk) + NADPH → Val (abcef) + CO2 (d) + AKG (ghijk)

v_{55} AcCoA (ab) + Pyr (cde) + Pyr (fgh) + Glu (ijklm) + NADPH → Leu (abdghe) + CO2 (c) + CO2 (f) + AKG (ijklm) + NADH

v_{56} Thr (abcd) + Pyr (efg) + Glu (hijkl) + NADPH → Ile (abfcdg) + CO2 (e) + AKG (hijkl) + NH3

v_{57} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Phe (abcefghij) + CO2 (d) + AKG (klmno)

v_{58} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Tyr (abcefghij) + CO2 (d) + AKG (klmno) + NADH

v_{59} Ser (abc) + R5P (defgh) + PEP (ijk) + E4P (lmno) + PEP (pqr) + Gln (stuvw) + 3 ATP + NADPH →
Trp (abcedklmnoj) + CO2 (i) + GAP (fgh) + Pyr (pqr) + Glu (stuvw)

v_{60} R5P (abcde) + FTHF (f) + Gln (ghijk) + Asp (lmno) + 5 ATP → His (edcbaf) + AKG (ghijk) + Fum (lmno) + 2 NADH

v_{61} Ser (abc) → Pyr (abc) + NH3

One-Carbon Metabolism

v_{62} MEETHF (a) + NADH → METHF (a)

v_{63} MEETHF (a) → FTHF (a) + NADPH

Oxidative Phosphorylation

v_{64} NADH + ½ O2 → 2 ATP

v_{65} FADH2 + ½ O2 → 1 ATP

Transhydrogenation

v_{66} NADH ↔ NADPH

ATP Hydrolysis

v_{67} ATP → ATP.ext

Transport

v_{68} Ac (ab) → Ac.ext (ab)

v_{69} CO2 (a) → CO2.ext (a)

v_{70} O2.ext → O2

v_{71} NH3.ext → NH3

v_{72} SO4.ext → SO4

Biomass Formation

v_{73} 0.488 Ala + 0.281 Arg + 0.229 Asn + 0.229 Asp + 0.087 Cys + 0.250 Glu + 0.250 Gln + 0.582 Gly + 0.090 His + 0.276 Ile + 0.428 Leu + 0.326 Lys + 0.146 Met + 0.176 Phe + 0.210 Pro + 0.205 Ser + 0.241 Thr + 0.054 Trp + 0.131 Tyr + 0.402 Val + 0.205 G6P + 0.071 F6P + 0.754 R5P + 0.129 GAP + 0.619 3PG + 0.051 PEP + 0.083 Pyr + 2.510 AcCoA + 0.087 AKG + 0.340 OAC + 0.443 MEETHF + 33.247 ATP + 5.363 NADPH → 39.68 Biomass + 1.455 NADH

CO_2 Exchange

v_{76} CO2.16G (a) → CO2.int (a) + X1

v_{77} CO2.12G (a) → CO2.int (a) + X1

v_{78} CO2.unlabeled (a) → CO2.int (a) + X1

v_{79} CO2.int (a) → CO2.out

Table E.2: Physiology of strains

	Growth Rate		Biomass Yield		Acetate Yield		Glucose Uptake Rate	
	1/h	S.E.	g/g	S.E.	mol/mol	S.E.	mmol/gdw/hr	S.E.
WT	0.72	0.02	0.38	0.02	0.70	0.05	10.47	0.52
pgi (unevolved)	0.14	0.00	0.48	0.01	0.00	0.02	1.59	0.05
ALE-1	0.37	0.01	0.43	0.01	0.34	0.02	4.80	0.15
ALE-2	0.46	0.01	0.52	0.01	0.10	0.01	4.94	0.16
ALE-3	0.51	0.02	0.49	0.01	0.24	0.01	5.73	0.24
ALE-4	0.35	0.00	0.52	0.01	0.05	0.00	3.72	0.08
ALE-5	0.33	0.01	0.50	0.01	0.15	0.00	3.66	0.10
ALE-6	0.49	0.03	0.52	0.01	0.12	0.01	5.32	0.28
ALE-7	0.42	0.02	0.53	0.01	0.11	0.02	4.44	0.20
ALE-8	0.41	0.00	0.52	0.01	0.05	0.05	4.39	0.07
ALE-9	0.49	0.01	0.46	0.01	0.15	0.03	5.83	0.24
ALE-10	0.43	0.01	0.50	0.01	0.10	0.01	4.77	0.18

Supplementary Figures for Chapter 10

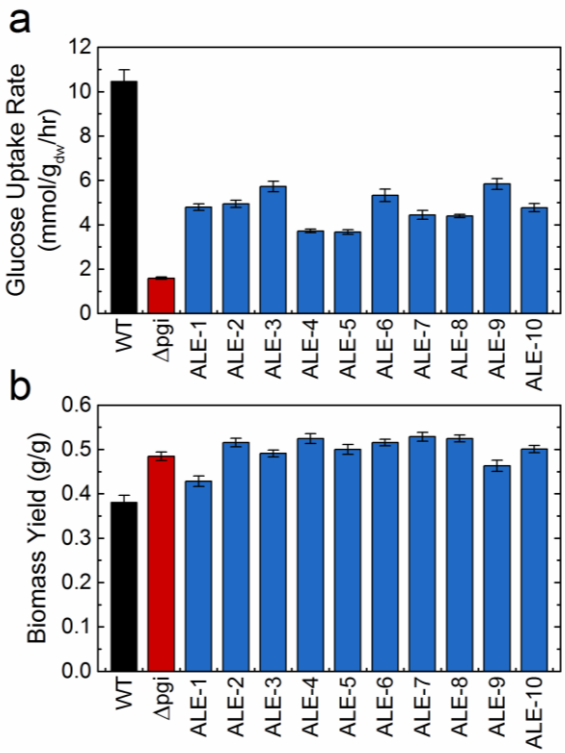


Figure E.1: Additional physiology of wild-type and *Δpgi* knockout strains. The estimated glucose uptake rates (a), and biomass yields (b) for all strains. The wild-type is shown in black, unevolved *Δpgi* in red, and *Δpgi* ALE strains in blue. The biomass yield was estimated as part of the full ^{13}C-MFA fitting, with the error bars reflecting standard deviations of the estimate. The estimated values were seen to be consistent with those estimated from culture measurements. Glucose uptake rates were calculated from the growth rates (Fig. 1a) and biomass yields (b), with propagated standard errors shown.

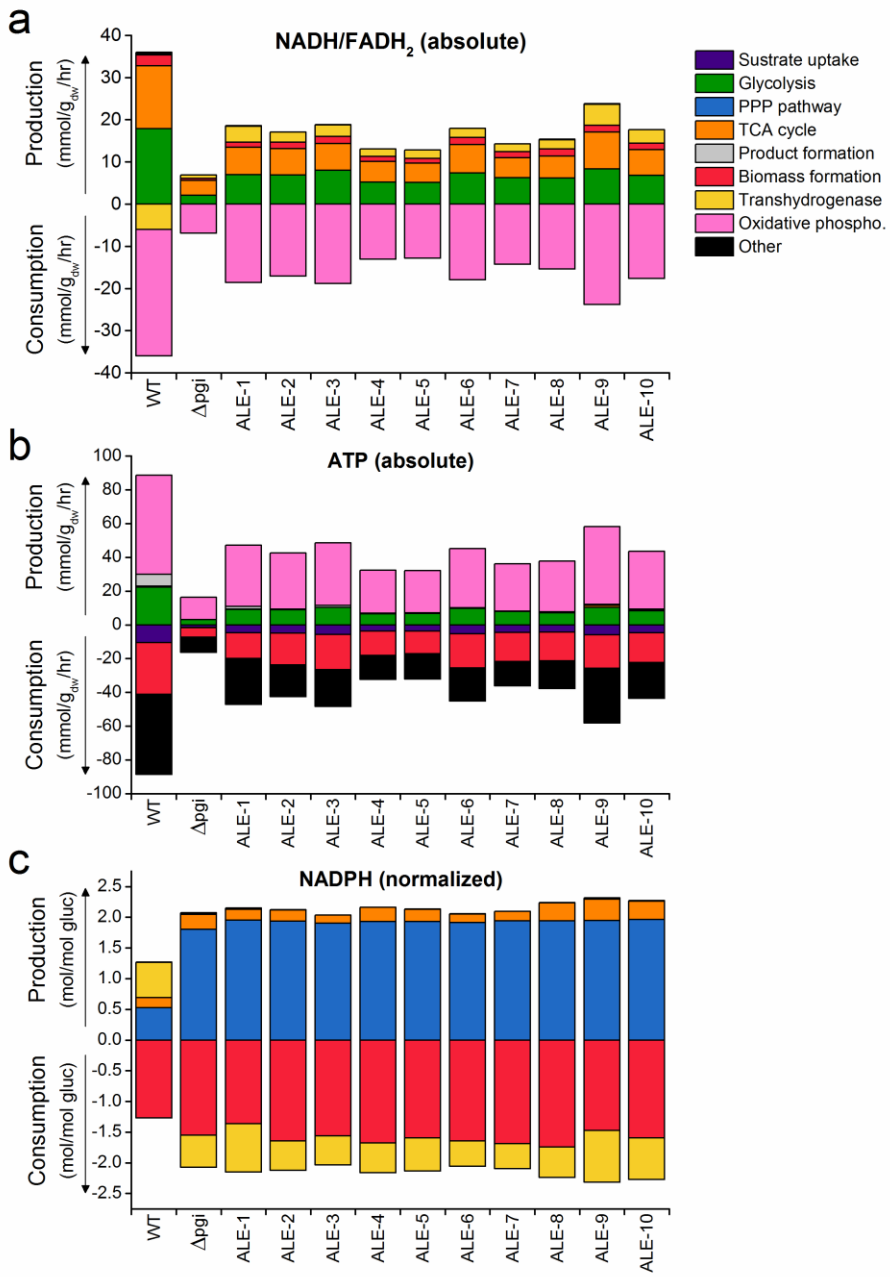


Figure E.2: Additional cofactor balances. Pathway specific contributions to cofactor production and consumption. Balances for NADH/FADH$_2$ (a) and ATP (b) are shown in absolute units (normalized contributions are shown in Fig. 3). The NADPH balance (c) is shown in normalized units (absolute contributions are shown in Fig. 4).

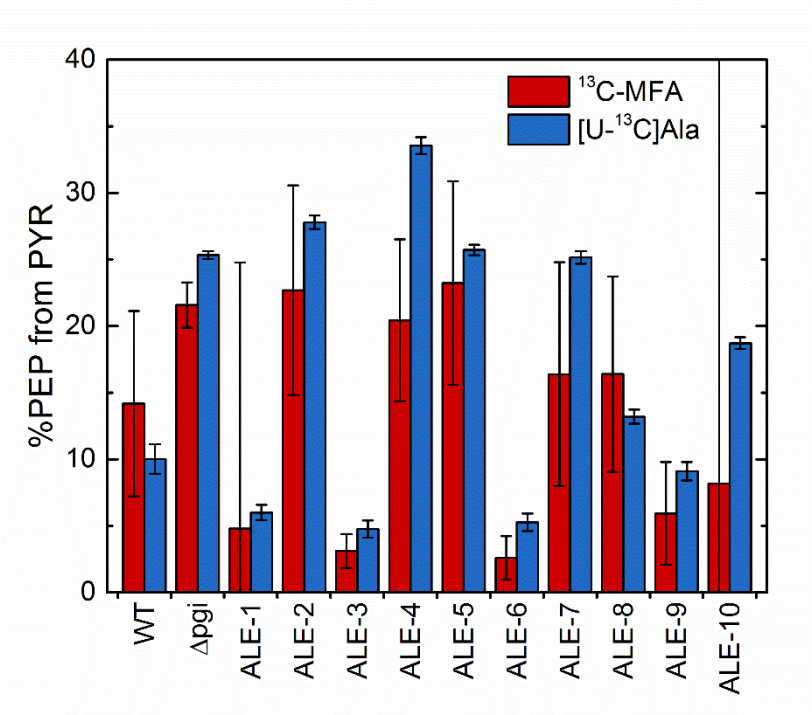


Figure E.3: PYR to PEP flux estimated independently via two ^{13}C tracer methods. The percentage of PEP derived from pyruvate, as calculated via ^{13}C-MFA and via the [U-^{13}C]alanine tracer method were in good overall agreement.

Appendix F

SUPPLEMENTARY MATERIAL FOR CHAPTER 11

Table F.1: Metabolic network model of *E. coli* used for ^{13}C-MFA

Glycolysis

v_1 Gluc.ext (abcdef) + PEP (ghi) → G6P (abcdef) + Pyr (ghi)
v_2 G6P (abcdef) ↔ F6P (abcdef)
v_3 F6P (abcdef) + ATP ↔ FBP (abcdef)
v_4 FBP (abcdef) ↔ DHAP (cba) + GAP (def)
v_5 DHAP (abc) ↔ GAP (abc)
v_6 GAP (abc) ↔ 3PG (abc) + ATP + NADH
v_7 3PG (abc) ↔ PEP (abc)
v_8 PEP (abc) ↔ Pyr (abc) + ATP

Pentose Phosphate Pathway

v_9 G6P (abcdef) → 6PG (abcdef) + NADPH
v_{10} 6PG (abcdef) → Ru5P (bcdef) + CO2 (a) + NADPH
v_{11} Ru5P (abcde) ↔ X5P (abcde)
v_{12} Ru5P (abcde) ↔ R5P (abcde)
v_{13} X5P (abcde) ↔ TK-C2 (ab) + GAP (cde)
v_{14} F6P (abcdef) ↔ TK-C2 (ab) + E4P (cdef)
v_{15} S7P (abcdefg) ↔ TK-C2 (ab) + R5P (cdefg)
v_{16} F6P (abcdef) ↔ TA-C3 (abc) + GAP (def)
v_{17} S7P (abcdefg) ↔ TA-C3 (abc) + E4P (defg)

Entner-Doudoroff Pathway

v_{18} 6PG (abcdef) → KDPG (abcdef)
v_{19} KDPG (abcdef) → Pyr (abc) + GAP (def)

TCA Cycle

v_{20} Pyr (abc) → AcCoA (bc) + CO2 (a) + NADH
v_{21} OAC (abcd) + AcCoA (ef) → Cit (dcbfea)

v_{22} Cit (abcdef) ↔ ICit (abcdef)
v_{23} ICit (abcdef) ↔ AKG (abcde) + CO2 (f) + NADPH
v_{24} AKG (abcde) → SucCoA (bcde) + CO2 (a) + NADH
v_{25} SucCoA (bcde) + CO2.int (a) + NADH → AKG (abcde)
v_{26} SucCoA (abcd) ↔ Suc (½ abcd + ½ dcba) + ATP
v_{27} Suc (½ abcd + ½ dcba) ↔ Fum (½ abcd + ½ dcba) + FADH2
v_{28} Fum (½ abcd + ½ dcba) ↔ Mal (abcd)
v_{29} Mal (abcd) ↔ OAC (abcd) + NADH

Glyoxylate Shunt
v_{30} ICit (abcdef) ↔ Glyox (ab) + Suc (½ edcf + ½ fcde)
v_{31} Glyox (ab) + AcCoA (cd) → Mal (abdc)

Amphibolic Reactions
v_{32} Mal (abcd) → Pyr (abc) + CO2 (d) + NADPH
v_{33} Mal (abcd) → Pyr (abc) + CO2 (d) + NADH
v_{34} PEP (abc) + CO2.int (d) → OAC (abcd)
v_{35} OAC (abcd) + ATP → PEP (abc) + CO2 (d)

Acetic Acid Formation
v_{36} AcCoA (ab) ↔ Ac (ab) + ATP

Amino Acid Biosynthesis
v_{37} AKG (abcde) + NADPH + NH3 → Glu (abcde)
v_{38} Glu (abcde) + ATP + NH3 → Gln (abcde)
v_{39} Glu (abcde) + ATP + 2 NADPH → Pro (abcde)
v_{40} Glu (abcde) + CO2 (f) + Gln (ghijk) + Asp (lmno) + AcCoA (pq) + 5 ATP + NADPH →
Arg (abcdef) + AKG (ghijk) + Fum (lmno) + Ac (pq)
v_{41} OAC (abcd) + Glu (efghi) → Asp (abcd) + AKG (efghi)
v_{42} Asp (abcd) + 2 ATP + NH3 → Asn (abcd)
v_{43} Pyr (abc) + Glu (defgh) → Ala (abc) + AKG (defgh)
v_{44} 3PG (abc) + Glu (defgh) → Ser (abc) + AKG (defgh) + NADH
v_{45} Ser (abc) ↔ Gly (ab) + MEETHF (c)
v_{46} Gly (ab) → CO2 (a) + MEETHF (b) + NADH + NH3
v_{47} CO2.int (a) + MEETHF (b) + NADH + NH3 → Gly (ab)
v_{48} Thr (abcd) ↔ Gly (ab) + AcCoA (cd) + NADH
v_{49} Ser (abc) + AcCoA (de) + 3 ATP + 4 NADPH + SO4 → Cys (abc) + Ac (de)
v_{50} Asp (abcd) + Pyr (efg) + Glu (hijkl) + SucCoA (mnop) + ATP + 2 NADPH →

LL-DAP (½ abcdgfe + ½ efgdcba) + AKG (hijkl) + Suc (½ mnop + ½ ponm)

v_{51} LL-DAP (½ abcdefg + ½ gfedcba) → Lys (abcdef) + CO2 (g)

v_{52} Asp (abcd) + 2 ATP + 2 NADPH → Thr (abcd)

v_{53} Asp (abcd) + METHF (e) + Cys (fgh) + SucCoA (ijkl) + ATP + 2 NADPH → Met (abcde) + Pyr (fgh) + Suc (½ ijkl + ½ lkji) + NH3

v_{54} Pyr (abc) + Pyr (def) + Glu (ghijk) + NADPH → Val (abcef) + CO2 (d) + AKG (ghijk)

v_{55} AcCoA (ab) + Pyr (cde) + Pyr (fgh) + Glu (ijklm) + NADPH → Leu (abdghe) + CO2 (c) + CO2 (f) + AKG (ijklm) + NADH

v_{56} Thr (abcd) + Pyr (efg) + Glu (hijkl) + NADPH → Ile (abfcdg) + CO2 (e) + AKG (hijkl) + NH3

v_{57} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Phe (abcefghij) + CO2 (d) + AKG (klmno)

v_{58} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH → Tyr (abcefghij) + CO2 (d) + AKG (klmno) + NADH

v_{59} Ser (abc) + R5P (defgh) + PEP (ijk) + E4P (lmno) + PEP (pqr) + Gln (stuvw) + 3 ATP + NADPH →
Trp (abcedklmnoj) + CO2 (i) + GAP (fgh) + Pyr (pqr) + Glu (stuvw)

v_{60} R5P (abcde) + FTHF (f) + Gln (ghijk) + Asp (lmno) + 5 ATP → His (edcbaf) + AKG (ghijk) + Fum (lmno) + 2 NADH

v_{61} Ser (abc) → Pyr (abc) + NH3

One-Carbon Metabolism

v_{62} MEETHF (a) + NADH → METHF (a)

v_{63} MEETHF (a) → FTHF (a) + NADPH

Oxidative Phosphorylation

v_{64} NADH + ½ O2 → 2 ATP

v_{65} FADH2 + ½ O2 → 1 ATP

Transhydrogenation

v_{66} NADH ↔ NADPH

ATP Hydrolysis

v_{67} ATP → ATP.ext

Transport

v_{68} Ac (ab) → Ac.ext (ab)

v_{69} CO2 (a) → CO2.ext (a)

v_{70} O2.ext → O2

v_{71} NH3.ext → NH3

v_{72} SO4.ext → SO4

Biomass Formation

v_{73} 0.488 Ala + 0.281 Arg + 0.229 Asn + 0.229 Asp + 0.087 Cys + 0.250 Glu + 0.250 Gln + 0.582 Gly + 0.090 His + 0.276 Ile + 0.428 Leu + 0.326 Lys + 0.146 Met + 0.176 Phe + 0.210 Pro + 0.205 Ser + 0.241 Thr + 0.054 Trp + 0.131 Tyr + 0.402 Val + 0.205 G6P + 0.071 F6P + 0.754 R5P + 0.129 GAP + 0.619 3PG + 0.051 PEP + 0.083 Pyr + 2.510 AcCoA + 0.087 AKG + 0.340 OAC + 0.443 MEETHF + 33.247 ATP + 5.363 NADPH → 39.68 Biomass + 1.455 NADH

CO_2 Exchange

v_{76} CO2.16G (a) → CO2.int (a) + X1

v_{77} CO2.12G (a) → CO2.int (a) + X1

v_{78} CO2.unlabeled (a) → CO2.int (a) + X1

v_{79} CO2.int (a) → CO2.out

Supplementary Figures for Chapter 11

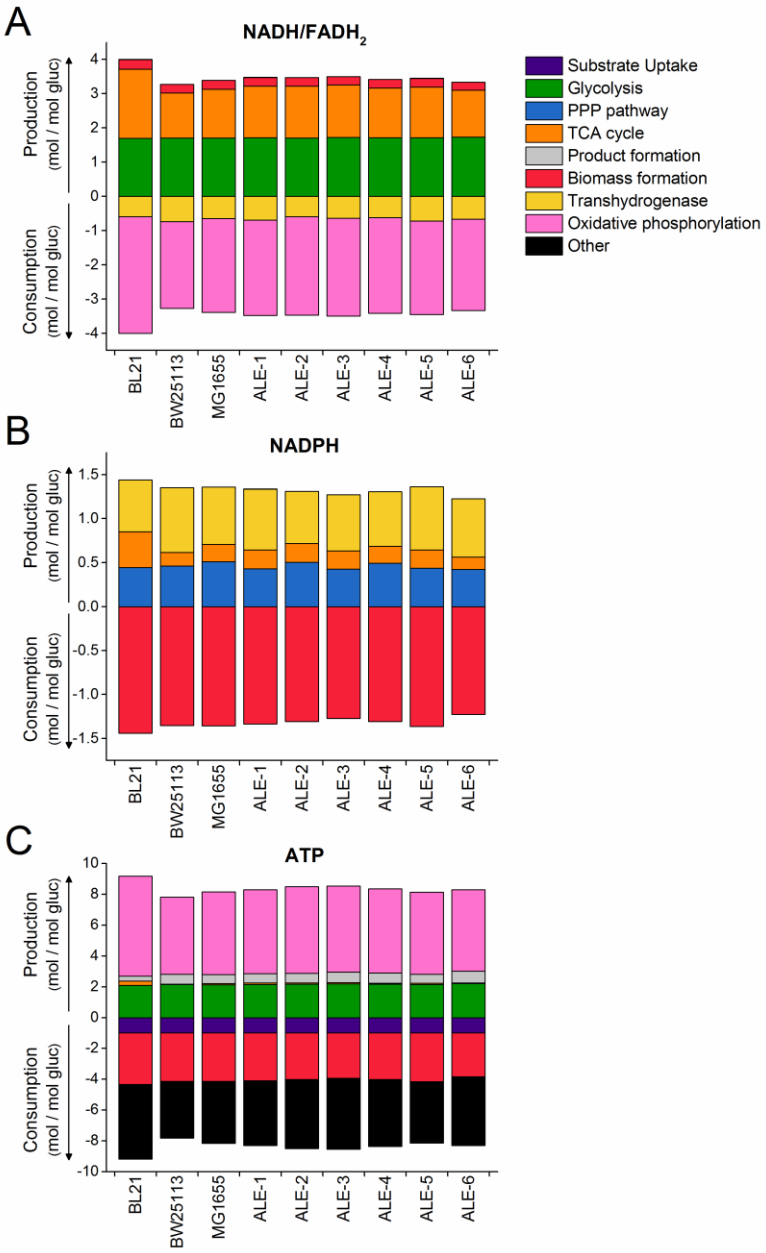


Figure F.1: Quantitative cofactor balances, normalized to glucose uptake. For each strain, the contributions of metabolic pathways to the production and consumption of cofactors are calculated. Positive values reflect production of cofactor, and negative values reflect consumption. Shown are balances for NADH/$FADH_2$ (lumped) (A), NADPH (B), and ATP (C). "Other" in the ATP panel represents the estimated ATP maintenance cost (here, assuming P/O ratio=2.0).

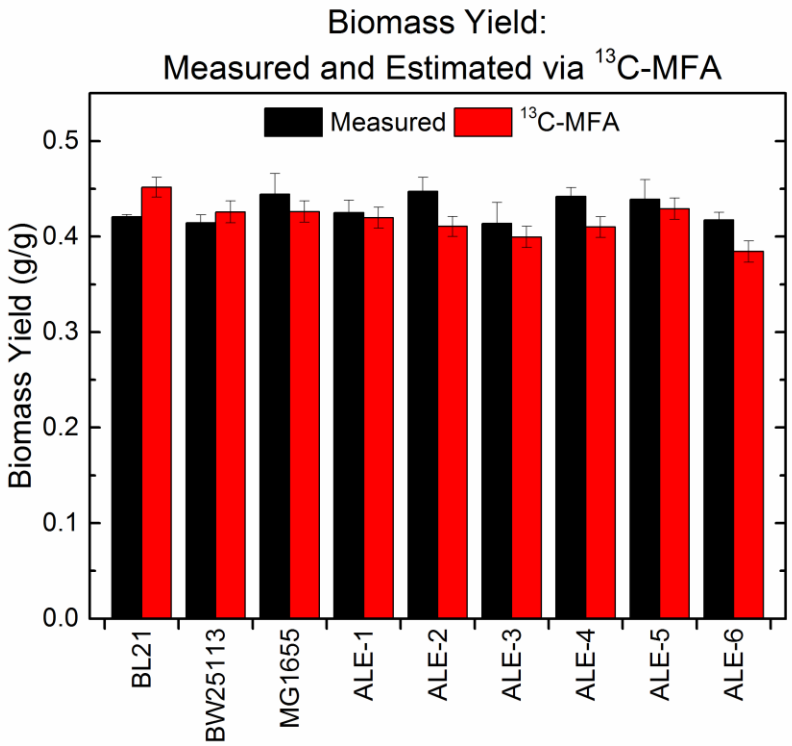


Figure F.2: Biomass yield (g_{DW}/g_{gluc}) as directly measured (as in Fig. 11.1) and as estimated via ^{13}C-MFA.

Appendix G

SUPPPLEMENTAL MATERIAL FOR CHAPTER 12

Table G.1: Metabolic network model of *Vibrio natriegens* used for ^{13}C-MFA

Glycolysis

v_1 Gluc.ext (abcdef) + PEP (ghi) → G6P (abcdef) + Pyr (ghi)
v_2 G6P (abcdef) ↔ F6P (abcdef)
v_3 F6P (abcdef) + ATP → FBP (abcdef)
v_4 FBP (abcdef) ↔ DHAP (cba) + GAP (def)
v_5 DHAP (abc) ↔ GAP (abc)
v_6 GAP (abc) ↔ 3PG (abc) + ATP + NADH
v_7 3PG (abc) ↔ PEP (abc)
v_8 PEP (abc) → Pyr (abc) + ATP

Pentose Phosphate Pathway

v_9 G6P (abcdef) → 6PG (abcdef) + NADPH
v_{10} 6PG (abcdef) → Ru5P (bcdef) + CO2 (a) + NADPH
v_{11} Ru5P (abcde) ↔ X5P (abcde)
v_{12} Ru5P (abcde) ↔ R5P (abcde)
v_{13} X5P (abcde) ↔ TK-C2 (ab) + GAP (cde)
v_{14} F6P (abcdef) ↔ TK-C2 (ab) + E4P (cdef)
v_{15} S7P (abcdefg) ↔ TK-C2 (ab) + R5P (cdefg)
v_{16} F6P (abcdef) ↔ TA-C3 (abc) + GAP (def)
v_{17} S7P (abcdefg) ↔ TA-C3 (abc) + E4P (defg)

Entner-Doudoroff Pathway

v_{18} 6PG (abcdef) → KDPG (abcdef)
v_{19} KDPG (abcdef) → Pyr (abc) + GAP (def)

TCA Cycle

v_{20} Pyr (abc) → AcCoA (bc) + CO2 (a) + NADH
v_{21} OAC (abcd) + AcCoA (ef) → Cit (dcbfea)

v_{22} Cit (abcdef) ↔ ICit (abcdef)
v_{23} ICit (abcdef) → AKG (abcde) + CO2 (f) + NADPH
v_{24} AKG (abcde) ↔ SucCoA (bcde) + CO2 (a) + NADH
v_{25} SucCoA (abcd) ↔ Suc (½ abcd + ½ dcba) + ATP
v_{26} Suc (½ abcd + ½ dcba) ↔ Fum (½ abcd + ½ dcba) + FADH2
v_{27} Fum (½ abcd + ½ dcba) ↔ Mal (abcd)
v_{28} Mal (abcd) ↔ OAC (abcd) + NADH

Glyoxylate Shunt
v_{29} ICit (abcdef) ↔ Glyox (ab) + Suc (½ edcf + ½ fcde)
v_{30} Glyox (ab) + AcCoA (cd) → Mal (abdc)

Amphibolic Reactions
v_{31} Mal (abcd) → Pyr (abc) + CO2 (d) + NADPH
v_{32} PEP (abc) + CO2 (d) → OAC (abcd)
v_{33} OAC (abcd) + ATP → PEP (abc) + CO2 (d)

Acetic Acid Formation
v_{34} AcCoA (ab) ↔ Ac (ab) + ATP

Amino Acid Biosynthesis
v_{35} AKG (abcde) + NADPH + NH3 → Glu (abcde)
v_{36} Glu (abcde) + ATP + NH3 → Gln (abcde)
v_{37} Glu (abcde) + ATP + 2 NADPH → Pro (abcde)
v_{38} Glu (abcde) + CO2 (f) + Gln (ghijk) + Asp (lmno) + AcCoA (pq) + 5 ATP + NADPH → Arg (abcdef) + AKG (ghijk) + Fum (lmno) + Ac (pq)
v_{39} OAC (abcd) + Glu (efghi) → Asp (abcd) + AKG (efghi)
v_{40} Asp (abcd) + 2 ATP + NH3 → Asn (abcd)
v_{41} Pyr (abc) + Glu (defgh) → Ala (abc) + AKG (defgh)
v_{42} 3PG (abc) + Glu (defgh) → Ser (abc) + AKG (defgh) + NADH
v_{43} Ser (abc) ↔ Gly (ab) + MEETHF (c)
v_{44} Gly (ab) ↔ CO2 (a) + MEETHF (b) + NADH + NH3
v_{45} Thr (abcd) → Gly (ab) + AcCoA (cd) + NADH
v_{46} Ser (abc) + AcCoA (de) + 3 ATP + 4 NADPH + SO4 → Cys (abc) + Ac (de)
v_{47} Asp (abcd) + Pyr (efg) + Glu (hijkl) + SucCoA (mnop) + ATP + 2 NADPH → LL-DAP (½ abcdgfe + ½ efgdcba) + AKG (hijkl) + Suc (½ mnop + ½ ponm)
v_{48} LL-DAP (½ abcdefg + ½ gfedcba) → Lys (abcdef) + CO2 (g)
v_{49} Asp (abcd) + 2 ATP + 2 NADPH → Thr (abcd)
v_{50} Asp (abcd) + METHF (e) + Cys (fgh) + SucCoA (ijkl) + ATP + 2 NADPH →

Met (abcde) + Pyr (fgh) + Suc (½ ijkl + ½ lkji) + NH3

v_{51} Pyr (abc) + Pyr (def) + Glu (ghijk) + NADPH → Val (abcef) + CO2 (d) + AKG (ghijk)

v_{52} AcCoA (ab) + Pyr (cde) + Pyr (fgh) + Glu (ijklm) + NADPH →
Leu (abdghe) + CO2 (c) + CO2 (f) + AKG (ijklm) + NADH

v_{53} Thr (abcd) + Pyr (efg) + Glu (hijkl) + NADPH → Ile (abfcdg) + CO2 (e) + AKG (hijkl) + NH3

v_{54} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH →
Phe (abcefghij) + CO2 (d) + AKG (klmno)

v_{55} PEP (abc) + PEP (def) + E4P (ghij) + Glu (klmno) + ATP + NADPH →
Tyr (abcefghij) + CO2 (d) + AKG (klmno) + NADH

v_{56} Ser (abc) + R5P (defgh) + PEP (ijk) + E4P (lmno) + PEP (pqr) + Gln (stuvw) + 3 ATP + NADPH →
Trp (abcedklmnoj) + CO2 (i) + GAP (fgh) + Pyr (pqr) + Glu (stuvw)

v_{57} R5P (abcde) + FTHF (f) + Gln (ghijk) + Asp (lmno) + 5 ATP →
His (edcbaf) + AKG (ghijk) + Fum (lmno) + 2 NADH

One-Carbon Metabolism

v_{58} MEETHF (a) + NADH → METHF (a)

v_{59} MEETHF (a) → FTHF (a) + NADPH

Oxidative Phosphorylation

v_{60} NADH + ½ O2 → 2 ATP

v_{61} FADH2 + ½ O2 → 1 ATP

Transhydrogenation

v_{62} NADH ↔ NADPH

ATP Hydrolysis

v_{63} ATP → ATP:ext

Transport

v_{64} Ac (ab) → Ac.ext (ab)

v_{65} CO2 (a) → CO2.ext (a)

v_{66} O2.ext → O2

v_{67} NH3.ext → NH3

v_{68} SO4.ext → SO4

Biomass Formation

v_{69} 0.418 Ala + 0.240 Arg + 0.188 Asn + 0.188 Asp + 0.074 Cys + 0.340 Glu + 0.340

Gln + 0.384 Gly + 0.071 His + 0.186 Ile + 0.314 Leu + 0.261 Lys + 0.104 Met + 0.134 Phe + 0.144 Pro + 0.203 Ser + 0.210 Thr + 0.046 Trp + 0.094 Tyr + 0.280 Val + 0.260 G6P + 0.071 F6P + 1.002 R5P + 0.129 GAP + 0.764 3PG + 0.051 PEP + 0.083 Pyr + 2.490 AcCoA + 0.087 AKG + 0.443 OAC + 0.588 MEETHF + 33.912 ATP + 5.482 NADPH $\rightarrow$ 38.093 Biomass + 1.891 NADH

CO_2 Exchange

v_{70} CO_2.unlabeled (a) + CO_2 (b) $\rightarrow$ CO_2 (a) + CO_2.out (b)

Table G.2: Biomass composition analysis of *V. natriegens* grown on glucose under aerobic conditions.

Overall biomass composition (%DW)	
Protein	46.5
RNA	28.6
Lipid	7.5
Glycogen	3.4

Amino acids (umol/gProtein)	
Ala	898
Arg	not measured
Asx	807
Cys	not measured
Glx	1464
Gly	825
His	153
Ile	399
Leu	675
Lys	560
Met	224
Phe	288
Pro	310
Ser	437
Thr	452
Trp	not measured
Tyr	203
Val	603

Fatty acids (umol/gFA)

C14:0	454
C16:1	1628
C16:0	1263
C18:1	469
C18:0	91

Table G.3 Metabolic fluxes for *Vibrio natriegens* as estimated by ^{13}C-MFA

Shown are the estimated net and exchange fluxes (normalized to substrate uptake rate of 100).
Accurate 95% confidence intervals of fluxes (LB95 = lower bound, UB95 = upper bound) were determined by evaluating the sensitivity of the minimized SSR to flux variations (Antoniewicz et al., 2006)
Number of fitted data sets: 2
Number of fitted measurements: 224
SSR: 188
Statistically Accepted: Yes

ABSOLUTE GLUCOSE UPTAKE RATE (mmol/g_{DW}/h): 21.4
FLUXES NORMALIZED TO 100 UNITS GLUCOSE UPTAKE

Rxn	Flux	best fit	LB95	UB95
1	Gluc.Ext + PEP -> G6P + Pyr	100.0	99.9	100.1
2	G6P <=> F6P (net)	80.2	79.2	81.1
3	F6P + ATP <=> FBP (net)	83.9	82.7	85.1
4	FBP <=> DHAP + GAP (net)	83.9	82.7	85.1
5	DHAP <=> GAP (net)	83.9	82.7	85.1
6	GAP <=> 3PG + ATP + NADH (net)	168.7	166.0	171.3
7	3PG <=> PEP (net)	153.8	149.7	157.8
8	PEP <=> Pyr + ATP (net)	22.3	15.6	28.2
9	G6P -> 6PG + NADPH	17.6	16.9	18.4
10	6PG -> Ru5P + CO2 + NADPH	17.0	16.3	17.9
11	Ru5P <=> X5P (net)	4.3	4.0	4.7
12	Ru5P <=> R5P (net)	12.7	11.8	13.7
13	X5P <=> GAP + E-C2 (net)	4.3	4.0	4.7
14	F6P <=> E4P + E-C2 (net)	-1.0	-1.3	-0.8
15	S7P <=> R5P + E-C2 (net)	-3.3	-3.5	-3.2
16	F6P <=> GAP + E-C3 (net)	-3.3	-3.5	-3.2

17	S7P <=> E4P + E-C3 (net)	3.3	3.2	3.5
18	6PG -> KDPG	0.6	0.3	0.9
19	KDPG -> GAP + Pyr	0.6	0.3	0.9
20	Pyr -> AcCoA + CO2 + NADH	109.7	101.3	117.8
21	AcCoA + OAC -> Cit	17.0	15.5	18.4
22	Cit <=> ICit (net)	17.0	15.5	18.4
23	ICit <=> AKG + CO2 + NADPH (net)	17.0	15.5	18.4
24	AKG <=> SucCoA + CO2 + NADH (net)	7.3	6.5	8.0
25	SucCoA <=> Suc + ATP (net)	4.3	3.5	4.9
26	Suc <=> Fum + FADH2 (net)	7.3	6.5	8.0
27	Fum <=> Mal (net)	10.0	9.0	10.8
28	Mal <=> OAC + NADH (net)	6.5	5.4	7.2
29	ICit <=> Glyox + Suc (net)	0.0	0.0	0.3
30	AcCoA + Glyox -> Mal	0.0	0.0	0.3
31	Mal -> Pyr + CO2 + NADPH	3.5	2.8	4.6
32	PEP + CO2 -> OAC	26.5	24.1	28.1
33	OAC + ATP -> PEP + CO2	0.0	0.0	1.2
34	AcCoA <=> Ac + ATP (net)	65.6	53.1	77.6
35	AKG + NADPH + NH3 -> Glu	46.4	42.0	50.8
36	Glu + ATP + NH3 -> Gln	5.9	5.3	6.4
37	Glu + 2 NADPH + ATP -> Pro	1.2	1.1	1.3
38	Glu + CO2 + Gln + NADPH + Asp + AcCoA + 5 ATP -> Arg + AKG + Fum + Ac	2.0	1.8	2.2
39	OAC + Glu -> Asp + AKG	12.2	11.0	13.3
40	Asp + 2 ATP + NH3 -> Asn	1.6	1.4	1.7
41	Pyr + Glu -> Ala + AKG	3.5	3.2	3.9
42	3PG + Glu -> Ser + NADH + AKG	8.4	7.7	9.2
43	Ser <=> Gly + MEETHF (net)	4.8	4.4	5.3
44	Gly <=> CO2 + MEETHF + NADH + NH3 (net)	1.6	1.4	1.7
45	Thr <=> Gly + AcCoA + NADH (net)	0.0	0.0	0.2
46	Ser + AcCoA + SO4 + 3 ATP + 4 NADPH + SO4 -> Cys + Ac	1.5	1.4	1.6
47	Asp + Pyr + Glu + 2 NADPH + ATP + SucCoA -> LL-DAP + AKG + Suc	2.2	2.0	2.4
48	LL-DAP -> Lys + CO2	2.2	2.0	2.4
49	Asp + 2 NADPH + 2 ATP -> Thr	3.3	3.0	3.6
50	Asp + METHF + Cys + 2 NADPH +	0.9	0.8	1.0

	ATP + SucCoA -> Met + Pyr + Suc + NH3			
51	2 Pyr + NADPH + Glu -> Val + CO2 + AKG	2.4	2.1	2.6
52	2 Pyr + AcCoA + Glu + NADPH -> Leu + 2 CO2 + AKG + NADH	2.6	2.4	2.9
53	Thr + Pyr + Glu + NADPH -> Ile + CO2 + AKG + NH3	1.6	1.4	1.7
54	E4P + 2 PEP + Glu + NADPH + ATP -> Phe + CO2 + AKG	1.1	1.0	1.2
55	E4P + 2 PEP + Glu + NADPH + ATP -> Tyr + CO2 + AKG + NADH	0.8	0.7	0.9
56	E4P + 2 PEP + R5P + Ser + Gln + NADPH + 3 ATP -> Trp + CO2 + Pyr + GAP + Glu	0.4	0.4	0.4
57	R5P + FTHF + Gln + Asp + 5 ATP -> His + 2 NADH + AKG + Fum	0.6	0.5	0.7
58	MEETHF + NADH -> METHF	0.9	0.8	1.0
59	MEETHF -> FTHF + NADPH	0.6	0.5	0.7
60	NADH + 0.5 O2 -> 2 ATP	253.0	236.2	268.9
61	FADH2 + 0.5 O2 -> ATP	7.3	6.5	8.0
62	NADH <=> NADPH (net)	68.9	59.1	78.2
63	ATP -> ATP.Ext	364.3	279.1	448.3
64	Ac -> Ac.Ext	69.1	57.0	80.7
65	CO2 -> CO2.Ext	141.4	134.2	148.2
66	O2.Ext -> O2	130.2	121.9	138.0
67	NH3.Ext -> NH3	49.8	45.1	54.6
68	SO4.Ext -> SO4	1.5	1.4	1.6
69	Biomass production (Cmol of biomass)	320.6	290.5	351.8
70	Intracellular CO2 dilution	19.7	6.2	33.6
EXCHANGE FLUXES				
2	G6P <=> F6P (exch)	84.5	57.2	124.9
4	FBP <=> DHAP + GAP (exch)	109.2	0.0	Inf
5	DHAP <=> GAP (exch)	202.6	0.0	Inf
6	GAP <=> 3PG + ATP + NADH (exch)	201.1	0.0	Inf
7	3PG <=> PEP (exch)	202.7	0.0	Inf
11	Ru5P <=> X5P (exch)	73.7	23.7	Inf
12	Ru5P <=> R5P (exch)	>1000	0.9	Inf
13	X5P <=> GAP + E-C2 (exch)	38.7	23.7	Inf

14	F6P <=> E4P + E-C2 (exch)	3.4	3.1	3.7
15	S7P <=> R5P + E-C2 (exch)	0.8	0.4	Inf
16	F6P <=> GAP + E-C3 (exch)	7.3	4.4	11.4
17	S7P <=> E4P + E-C3 (exch)	103.1	0.4	Inf
22	Cit <=> ICit (exch)	25.6	0.0	Inf
24	AKG <=> SucCoA + CO2 + NADH (exch)	1.8	1.3	2.8
25	SucCoA <=> Suc + ATP (exch)	>1000	17.7	Inf
26	Suc <=> Fum + FADH2 (exch)	>1000	14.4	Inf
27	Fum <=> Mal (exch)	>1000	>1000	Inf
28	Mal <=> OAC + NADH (exch)	3.6	1.0	7.1
29	ICit <=> Glyox + Suc (exch)	0.3	0.1	0.5
34	AcCoA <=> Ac + ATP (exch)	67.7	0.0	Inf
43	Ser <=> Gly + MEETHF (exch)	2.0	1.8	2.2
44	Gly <=> CO2 + MEETHF + NADH + NH3 (exch	0.0	0.0	0.0
62	NADH <=> NADPH (exch)	130.1	0.0	Inf

FRACTIONAL LABELING OF AMINO ACIDS (G-VALUES)				
	Fractional labeling of Gly (data set #1)	94%	93%	95%
	Fractional labeling of Val (data set #1)	89%	89%	90%
	Fractional labeling of Leu (data set #1)	87%	86%	87%
	Fractional labeling of Ile (data set #1)	88%	87%	89%
	Fractional labeling of Ser (data set #1)	97%	96%	98%
	Fractional labeling of Thr (data set #1)	92%	91%	93%
	Fractional labeling of Phe (data set #1)	86%	85%	86%
	Fractional labeling of Asp (data set #1)	94%	93%	94%
	Fractional labeling of Glu (data set #1)	94%	93%	94%
	Fractional labeling of Tyr (data set #1)	94%	93%	95%
	Fractional labeling of R5P (data set #1)	99%	98%	100%
	Fractional labeling of G6P (data set #1)	98%	98%	99%
	Fractional labeling of Ala (data set #2)	89%	88%	89%
	Fractional labeling of Gly (data set #2)	89%	0%	100%
	Fractional labeling of Val (data set #2)	84%	83%	84%
	Fractional labeling of Leu (data set #2)	80%	79%	80%
	Fractional labeling of Ile (data set #2)	81%	80%	81%
	Fractional labeling of Ser (data set #2)	97%	97%	98%

	Fractional labeling of Thr (data set #2)	88%	88%	88%
	Fractional labeling of Phe (data set #2)	79%	79%	79%
	Fractional labeling of Asp (data set #2)	91%	91%	92%
	Fractional labeling of Glu (data set #2)	92%	91%	92%
	Fractional labeling of Tyr (data set #2)	0%	0%	100%
	Fractional labeling of R5P (data set #2)	97%	97%	98%
	Fractional labeling of G6P (data set #2)	96%	96%	97%
	** data set #1 = [1,2]Glucose			
	** data set #2 = [1,6]Glucose			

CARBON BALANCE (mol/100 mol substrate)

	Substrate uptake	100.0
	Biomass production	320.6
	Acetate production	69.1
	CO2 production	141.4

CARBON BALANCE (Cmol/Cmol)

	Biomass production	53%
	Acetate production	23%
	CO2 production	24%

CO-FACTOR BALANCES (mol/100 mol substrate)

		NADH/FADH2	NADPH	ATP
	Sustrate uptake	0.0	0.0	100.0
	Glycolysis	168.7	0.0	207.1
	PPP pathway	0.0	34.7	0.0
	TCA cycle	130.8	17.0	4.3
	Product formation	0.0	0.0	65.6
	Biomass formation	29.7	-124.0	-326.0
	Transhydrogenase	-68.9	68.9	0.0
	Oxidative phosphorylation	-260.3	0.0	513.4
	Other	0.0	3.5	364.3

CO-FACTOR BALANCES (mmol/gDW/h)		NADH/FADH2	NADPH	ATP
	Sustrate uptake	0.0	0.0	-21.4
	Glycolysis	36.1	0.0	44.3
	PPP pathway	0.0	7.4	0.0
	TCA cycle	28.0	3.6	0.9
	Product formation	0.0	0.0	14.0
	Biomass formation	6.4	-26.5	-69.8
	Transhydrogenase	-14.7	14.7	0.0
	Oxidative phosphorylation	-55.7	0.0	109.9
	Other	0.0	0.7	-78.0

GLUCOSE UPTAKE AND BIOMASS DATA		
	Substrate uptake rate (mmol/gDW/h)	21.40
	Biomass reaction (ATP/Cmol)	-0.890
	Biomass reaction (NADPH/Cmol)	-0.144
	Biomass reaction (NADH/Cmol)	0.050

Supplementary Figures for Chapter 12

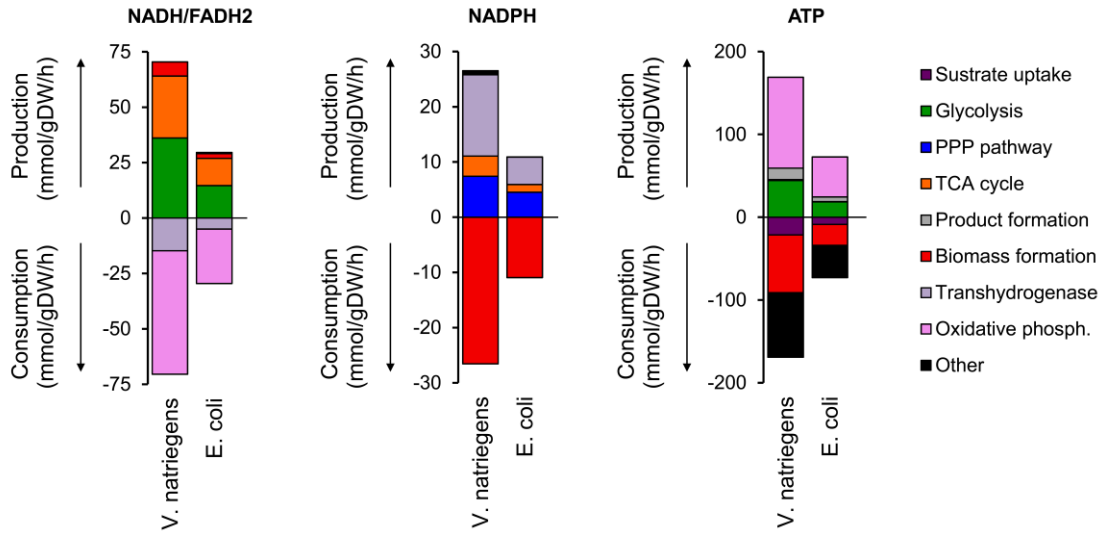


Figure G.1: Carbon and cofactor balances of *V. natriegens* and *E. coli*. (A) The overall carbon balances reflect the fates of glucose on an absolute (mmol/g_{DW}/h) basis. (B) The metabolic pathways responsible for the production and consumption of the cofactors NADH/$FADH_2$, NADPH, and ATP are summarized, on a normalized (per unit glucose) basis. For the ATP balance, the production rate from oxidative phosphorylation and the maintenance cost ('Other') are based on an assumed effective P/O ratio of 2. As this has not been reliably measured in *V. natriegens*, this result should be interpreted with caution.