1-1-2012

Development and Application of Fluxomics Tools for Analyzing Metabolisms in Non-Model Microorganisms

Xueyang Feng
Washington University in St. Louis

WASHINGTON UNIVERSITY

Department of Energy, Environmental and Chemical Engineering

Dissertation Examination Committee:
Yinjie J Tang, Chair
Pratim Biswas
Robert Blankenship
Daniel Giammar
Cynthia Lo
Costas Maranas
Himadri Pakrasi

Development and Application of Fluxomics Tools for Analyzing Metabolisms in Non-Model Microorganisms

by

Xueyang Feng

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

May 2012

Saint Louis, Missouri

ABSTRACT OF THE DISSERTATION

Development and Application of Fluxomics Tools for Analyzing Metabolisms in Non-Model Microorganisms

by

Xueyang Feng

Doctor of Philosophy in Energy, Environmental and Chemical Engineering

Washington University in St. Louis, 2012

Professor Yinjie Tang, Chair

Decoding microbial metabolism is of great importance in revealing the mechanisms governing the physiology of microbes and rewiring the cellular functions in metabolic engineering. Complementing the genomics, transcriptomics, proteinomics and metabolomics analysis of microbial metabolism, fluxomics tools can measure and simulate the *in vivo* enzymatic reactions as direct readouts of microbial metabolism. This dissertation develops and applies broad-scope tools in metabolic flux analysis to investigate metabolic insights of non-model environmental microorganisms.

^{13}C-based pathway analysis has been applied to analyze specific carbon metabolic routes by tracing and analyzing isotopomer labeling patterns of different metabolites after growing cells with ^{13}C-labeled substrates. Novel pathways, including Re-type citrate synthase in tricarboxylic acid cycle and

citramalate pathways as an alternate route for isoleucine biosynthesis, have been identified in *Thermoanaerobacter* X514 and other environmental microorganisms. Via the same approach, the utilizations of diverse carbon/nitrogen substrates and productions of hydrogen during mixotrophic metabolism in *Cyanothece* 51142 have been characterized, and the medium for a slow-growing bacterium, *Dehalococcoides ethenogenes* 195, has been optimized. In addition, ^{13}C-based metabolic flux analysis has been developed to quantitatively profile flux distributions in central metabolisms in a green sulfur bacterium, *Chlorobaculum tepidum*, and thermophilic ethanol-producing *Thermoanaerobacter* X514. The impact of isotope discrimination on ^{13}C-based metabolic flux analysis has also been estimated.

A constraint-based flux analysis approach was newly developed to integrate the bioprocess model into genome-scale flux balance analysis to decipher the dynamic metabolisms of *Shewanella oneidensis* MR-1. The sub-optimal metabolism and the time-dependent metabolic fluxes were profiled in a genome-scale metabolic network. A web-based platform was constructed for high-throughput metabolic model drafting to bridge the gap between fast-paced genome-sequencing and slow-paced metabolic model reconstruction. The platform provides over 1,000 sequenced genomes for model drafting and diverse customized tools for model reconstruction. The *in silico* simulation of flux distributions in both metabolic steady state and dynamic state can be achieved via

flux balance analysis and dynamic flux balance analysis embedded in this platform.

Cutting-edge fluxomics tools for functional characterization and metabolic prediction continue to be developed in the future. Broad-scope systems biology tools with integration of transcriptomics, proteinomics and fluxomics can reveal cell-wide regulations and speed up the metabolic engineering of non-model microorganisms for diverse bioenergy and environmental applications.

Acknowledgements

I would like to express my sincere gratitude to my advisor and mentor, Dr. Yinjie Tang, for personally training me in metabolic flux analysis. He tirelessly seeks opportunities to advance my academic career by helping me perfect research skills, collaborate with world-renowned scientists, and apply to academic fellowships. From him, I learned how to appreciate and cultivate my passion for research, maintain a positive attitude about a career in academia, and develop a meaningful philosophy about life. I am also thankful to Dr. Himadri Pakrasi for mentoring me during my Ph.D. research; Dr. Robert Blankenship and Dr. Joseph Tang for providing me opportunities to study novel microorganisms; Dr. Yixin Chen for teaching me the nonlinear optimization algorithms; Dr. Kirk Dolan for improving my computational skills to solve inverse problems; and Drs. Jizhong Zhou and Lisa Alveraz-Cohen for productive collaborations.

Very special thanks to my friend, Eric Xu, for helping me with my MicrobesFlux software development, and my lab mates, Bing Wu, Yi Xiao etc., for collegial discussions and help. My work is generously supported by the EECE department and DOE/NSF grants.

This dissertation is dedicated to my parents for their unconditional love and unwavering support. I can never adequately thank them for bringing out the best in me. I am forever indebted to them.

Table of Contents

List of Tables

Chapter 7

List of Figures

Chapter 1

Introduction

1.1 Systems biology of microbial metabolism

Microbial metabolism is a complex biological system that involves the interactions of thousands of genes, proteins and metabolites (1). Toward a system-level investigation of microbial metabolism, high-throughput "omics" tools are being developed and applied to capture the behaviors of different metabolic components. Transcriptomics, proteomics and metabolomics tools have been used to characterize the responses of gene expressions, protein productions and metabolite concentrations, respectively, to metabolic perturbations. In complementary to the aforementioned omics tools, fluxomics analysis has become instrumental in the systems biology of microbial metabolism. In fluxomics study, the metabolite turnover rates (i.e. metabolic fluxes) are measured or predicted as a direct readout of the functional output of metabolism (2). By developing and applying diverse fluxomics tools, the novel metabolic features in environmental microorganisms can be revealed and the knowledge-based rewiring of cellular functions can be achieved in metabolic engineering.

Fluxomics tools are often categorized into two groups based on the principles of design: constraint-based metabolic flux analysis and ^{13}C-based pathway and flux analysis (Figure 1.1). Constraint-based metabolic flux analysis

profiles the "optimal" microbial metabolisms for *in silico* simulations and predictions. Three components are utilized to find intracellular fluxes in this fluxomics approach: the stoichiometry of the metabolic reactions in a reconstructed metabolic model, an objective function based on prior knowledge of microbial metabolism, and a series of boundary conditions imposed on metabolic reactions (3). In comparison, ^{13}C-based pathway and flux analysis measures *in vivo* operation of a metabolic network. It is carried out by culturing microbes with defined ^{13}C-labeled carbon source, tracing the transitional paths of the labeled atoms between metabolites in the biochemical network, and analyzing the labeling profiles in various metabolites both qualitatively and quantitatively. ^{13}C-based pathway and flux analysis provides the most accurate readouts of microbial metabolisms (4).

1.2 Constraint-based metabolic flux analysis

Flux balance analysis (FBA) is commonly used for constraint-based metabolic flux analysis. To formulate a FBA problem, the metabolic model of a target microorganism, often in genome scale, needs to be reconstructed (3). A draft metabolic model is first constructed based on genome annotation, which is then iteratively refined by adjusting the metabolic functionality to fill the gaps between model simulations and experimental observations (5). From the reconstructed metabolic model, the stoichiometry matrix *S* is generated in FBA, with each row in *S* representing one metabolite and each column in *S* representing

one metabolic flux. The vector of metabolic fluxes, *v*, follows $S \cdot v = 0$ under metabolic steady state and is constrained within feasible ranges identified from the physical, chemical and biological characteristics (e.g., thermodynamic directionality, and enzyme capacity). Since the FBA problem is normally underdetermined with a larger number of variables (i.e., reactions) than that of equations, an objective function (e.g., maximizing growth rate) is assumed for the microbial metabolism to pinpoint the unique flux distributions in the metabolic network. The FBA is constructed as a linear programming (3,6):

$$\begin{aligned} &\max c^{T} \cdot v \\ &\text{s.t.} \quad S \cdot v = 0 \\ &\qquad lb \le v \le ub \end{aligned}$$

where *c* is a vector that represents the objective function in FBA.

As a mathematical approach for *in silico* simulation of microbial metabolism, FBA is equipped with a strong predictive capability to prove or disprove various assumptions in systems biology of microbial metabolism. However, since FBA relies on a biologically appropriate objective function to quantify the flow of metabolites through metabolic network, it cannot always give realistic metabolic readouts due to the lack of *a priori* knowledge of microbial metabolism (7).

1.3 ^{13}C-based metabolic pathway and flux analysis

The ^{13}C-based metabolic pathway analysis aims at qualitative elucidation of metabolic pathway activity by applying isotopic labeling approaches (8). In general, the microbes are cultured using a carbon source with a known distribution of ^{13}C isotopomers. As the ^{13}C-labeled carbon source is being used for different metabolic pathways, the metabolites synthesized in microbial metabolism (e.g. amino acids) acquire unique labeling patterns, which are measured by gas chromatography–mass spectrometry (GC-MS). To elucidate the route for metabolites synthesis, the "fingerprinting" labeling patterns of metabolites are generated by tracing the ^{13}C carbon transitions in the proposed pathway; and compared with the corresponding experimental measurements to reveal whether the metabolic pathway is active.

The ^{13}C-based metabolic flux analysis (^{13}C-MFA) is a quantitative approach to identify the metabolic flux distributions under metabolic and isotopic steady state, in which the pool size and the ^{13}C-labeling patterns of intracellular metabolites are invariable (4,9). By integrating isotopic labeling approach and *in silico* computation, ^{13}C-MFA is formulated as an inverse problem to find a set of fluxes that leads to the best match of the experimentally measured isotopomer abundances. The nonlinear programming of ^{13}C-MFA is formulated as follows (10,11):

$$\min \varepsilon(v_n) = \sum_{i=0}^{n} \left(\frac{M_i - N_i(v_n)}{\delta_i} \right)^2$$

$$\text{s.t.} \quad S \cdot v = 0$$

$$lb \le v \le ub$$

where v_n are the unknown fluxes to be optimized in the program, M_i are the measured isotopomer labeling patterns of metabolites from GC-MS, N_i are the corresponding model-simulated labeling patterns of metabolites, and δ_i are the corresponding standard deviation in GC-MS measurement. The unknown metabolic fluxes are searched to minimize ε.

The method of ^{13}C-based pathway and flux analysis allows for precise determinations of metabolic states in a particular growth condition. Compared to FBA, ^{13}C-based pathway and flux analysis cannot directly predict the metabolic behaviors. However, it provides a physiologically reliable description of microbial metabolism.

1.4 Recent advances in fluxomics

Fluxomics study is an emerging area in systems biology. In the past decade, various approaches have been developed and improved to investigate the cellular metabolic behaviors in response to environmental and genetic perturbations; which yield significant achievements in both mechanistic studies of cell physiology and rational designs in metabolic engineering.

1.4.1 Advances in constraint-based metabolic flux analysis

By integrating genome annotations, gene-protein-flux (GPR) correlations, and experimental validations, genome-scale metabolic models are powerful platforms in constraint-based flux analysis and have been constructed and applied for various microorganisms (5). Simulations from genome-scale FBA have been widely used in a number of studies (12), including interpreting various "omics" data (13), directing hypothesis-driven discoveries (14), investigating metabolic network properties (15), and predicting dynamic metabolisms of microorganisms (16,17). Advanced computational approaches, such as MOMA (18), OptKnock (19), and OptStrain (20), have been developed to guide the *in silico* design of genetic manipulations in metabolic engineering. Additionally, the multi-species relationships in mixed cultures are starting to be analyzed by constraint-based metabolic flux analysis (21).

1.4.2 Advances in ^{13}C-based pathway and flux analysis

State-of-the-art mass spectrometry technologies, such as liquid chromatography–mass spectrometry (LC-MS), have been developed to directly measure the ^{13}C-labeling patterns of low-abundant and fast-turnover metabolites. Mass spectrometry technologies could take the snapshot of transient metabolism and can be used for novel pathways identification (22) and kinetic flux profiling (23). As an improvement to conventional ^{13}C-MFA under metabolic and isotopic steady state, an advanced experimental and computational framework has been

proposed and employed to quantitatively investigate intracellular metabolism under metabolic steady state but isotopic non-steady state (24,25). To unravel the time profiles of metabolic fluxes under metabolic non-steady state, several approaches have been developed by tracing and analyzing the ^{13}C-labeling patterns of either fast-turnover intracellular metabolites (26), or slow-turnover proteinogenic amino acids (27). Additionally, pioneering fluxomics studies have recently been achieved to provide subpopulation-specific metabolic pathway usage in mixed cultures via ^{13}C-MFA of reporter proteins (28).

1.4.3 Advances in fluxomics software development

Multiple software toolboxes have been developed to facilitate fluxomics studies. The hallmark of software development in constraint-based metabolic flux analysis is the COBRA toolbox (29), which provides a computational platform for model constructions and FBA studies. Other toolboxes, such as OptFlux (30), have facilitated advanced FBA approaches for microbial strain design. A few software tools for ^{13}C-MFA are also available, including OpenFLUX (31) and FiatFlux (32). Moreover, web-based platforms have been constructed for metabolic model generations and FBA. For example, BioMet Toolbox is a web-based resource for FBA and transcriptome analysis (33). Webcoli supplies diverse approaches for users to reconstruct a genome-scale *E.coli* metabolic model (34). Model SEED (35) has been developed to automatically generate genome-scale metabolic models for different microbes.

1.5 Scope of dissertation

Decoding microbial metabolisms is of great importance in revealing the mechanisms governing the physiology of microbes and rewiring the microbial metabolism for bioenergy and biomaterial production. As one of the fundamental approaches to characterize microbial metabolism, the fluxomics tools have not yet been developed or applied as extensively as the other omics approaches in systems biology. Noteworthy hurdles are also present in fluxomics study of industrial biotechnology (Chapter 2). During my Ph.D. training, the ^{13}C-based pathway analysis has been applied to a wide range of environmental microorganisms, including *Cyanothece* 51142 (36,37), *Thermoanaerobacter* X514 (38), *Mycobacteria smegmatis* (39), *Shewanella oneidensis* MR-1 (40), *Roseobacter denitrificans* (41), *Heliobacterium modesticaldum* (42), and *Dehalococcoides ethenogenes* 195 (43), with a cluster of novel metabolic features uncovered to enrich the knowledge of microbial metabolisms (Chapter 3). The metabolic flux distributions in a green sulfur bacterium, *Chlorobaculum tepidumgreen* (10), have been investigated by ^{13}C-MFA to reveal the functionality of reversed TCA cycle; and the impact of isotope discrimination on ^{13}C-MFA has been rigorously evaluated using flux analysis of heterotrophic metabolism in *E.coli* (Chapter 4). A novel computational approach has been developed to decipher the dynamic metabolism of *Shewanella oneidensis* MR-1 by integrating the bioprocess model with dynamic flux balance analysis (Chapter 5). In addition,

MicrobesFlux, a web-based platform for metabolic model drafting has been constructed to bridge the gap between fast-paced genome-sequencing and slow-paced metabolic model reconstruction (Chapter 6). Several broad-scope studies have also been accomplished to supplement the systems biology of microbial metabolism; and advanced fluxomics tools are expected to be developed in future investigations (Chapter 7).

1.6 References

1. Heinemann M, Sauer U (2010) Systems biology of microbial metabolism. Curr Opin Microbiol 13: 1-7.

2. Dauner M (2010) From fluxes and isotope labeling patterns towards in silico cells. Curr Opin Biotechnol 21: 55-62.

3. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotechnol 28: 245-248.

4. Feng X, Page L, Rubens J, Chircus L, Varman A, et al. (2010) Bridging the Gap between Fluxomics and Industrial Biotechnology. Journal of Biomedicine and Biotechnology 2010.

5. Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 5: 93-121.

6. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol 60: 3724-3731.

7. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli.* Mol Syst Biol 3: 119.

8. You L, Page L, Berla B, Feng X, Pakrasi H, et al. (2011) Metabolic pathway discovery and quantification via ^{13}C-isotopomer analysis of proteinogenic amino acids. Journal of Visualized Experiments Accept.

9. Zamboni N, Fendt SM, Rühl M, Sauer U (2009) ^{13}C-based metabolic flux analysis. Nat Protoc 4: 878-892.

10. Feng X, Tang K-H, Blankenship RE, Tang YJ (2010) Metabolic flux analysis of the mixotrophic metabolisms in the green sulfur bacterium *Chlorobaculum tepidum*. J Biol Chem 285: 35104-35112.

11. Tang YJ, Chakraborty R, Martin HG, Chu J, Hazen TC, et al. (2007) Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-NTA. Appl Environ Microbiol 73: 3859-3864.

12. Oberhardt MA, Palsson BØ, Papin JA (2009) Applications of genome-scale metabolic reconstructions. Mol Syst Biol 5.

13. Moxley JF, Jewett MC, Antoniewicz MR, Villas-Boas SG, Alper H, et al. (2009) Linking high resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. Proc Natl Acad Sci 106: 6477~6482.

14. Segura D, Mahadevan R, Juarez K, Lovley DR (2008) Computational and experimental analysis of redundancy in the central metabolism of Geobacter sulfurreducens. . PLoS Comput Biol 4: e36.

15. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD (2004) Flux Coupling Analysis of Genome-Scale Metabolic Network. Genome Res 14: 301-312.

16. Mahadevan R, Edwards JS, Doyle FJ (2002) Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J 83: 1331-1340.

17. Hjersted JL, Henson MA, Mahadevan R (2007) Genome-scale analysis of Saccharomyces cerevisiae metabolism and ethanol production in fed-batch culture. Biotechnol Bioeng 97: 1190-1204.

18. Segrè D, Vitkup D, Church G (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 99: 15112-15117.

19. Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84: 647-657.

20. Pharkya P, Burgard AP, Maranas CD (2004) OptStrain: a computational framework for redesign of microbial production systems. Genome Res 14: 2367-2376.

21. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, et al. (2007) Metabolic modeling of a mutualistic microbial community. Mol Syst Biol 3: 92.

22. Amador-Noguez D, Feng X-J, Fan J, Roquet N, Rabitz H, et al. (2010) Systems-Level Metabolic Flux Profiling Elucidates a Complete, Bifurcated Tricarboxylic Acid Cycle in Clostridium acetobutylicum. Journal of Bacteriology 192: 4452–4461.

23. Yuan J, Bennett BD, Rabinowitz JD (2008) Kinetic flux profiling for quantitation of cellular metabolic fluxes. Nat Protoc 3: 1328-1340.

24. Nöh K, Grönke K, Luo B, Takors R, Oldiges M, et al. (2007) Metabolic flux analysis at ultra short time scale: isotopically non-stationary ^{13}C labeling experiments. J Biotechnol 129: 249-267.

25. Young JD, Shastri AA, Stephanopoulos G, Morgan JA (2011) Mapping photoautotrophic metabolism with isotopically nonstationary (13)C flux analysis. Metab Eng 13: 656-665.

26. Rühl M, Zamboni N, Sauer U (2010) Dynamic Flux Responses in Riboflavin Overproducing Bacillus subtilis to Increasing Glucose Limitation in Fed-Batch Culture. Biotechnol Bioeng 105: 795-804.

27. Antoniewicz MR, Kraynie DF, Laffend LA, González-Lergier J, Kelleher JK, et al. (2007) 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: 277-292.

28. Rühl M, Hardt WD, Sauer U (2011) Subpopulation-Specific Metabolic Pathway Usage in Mixed Cultures as Revealed by Reporter Protein-Based ^{13}C Analysis. Applied and Environmental Microbiology 77: 1816-1821.

29. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc 2: 727-738.

30. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, et al. (2010) OptFlux: an open-source software platform for in silico metabolic engineering. BMC Syst Biol 4: 45.

31. Quek L-E, Wittmann C, Nielsen LK, Krömer JO (2009) OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. Microb Cell Fact 8.

32. Zamboni N, Fischer E, Sauer U (2005) FiatFlux--a software for metabolic flux analysis from 13C-glucose experiments. BMC Bioinformatics 6.

33. Cvijovic M, Olivares-Hernández R, Agren R, Dahr N, Vongsangnak W, et al. (2010) BioMet Toolbox: genome-wide analysis of metabolism. Nucleic Acids Res 38: W144.

34. Jung T-S, Yeo HC, Reddy SG, Cho W-S, Lee D-Y (2009) WEbcoli: an interactive and asynchronous web application for in silico design and analysis of genome-scale *E.coli* model Bioinformatics 25: 2850-2852

35. Henry C, DeJongh M, Best A, Frybarger P, Linsay B, et al. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28: 977-982.

36. Feng X, Banerjee A, Berla B, Page L, Wu B, et al. (2010) Mixotrophic and photoheterotrophic metabolisms in Cyanothece sp. ATCC 51142 under continuous light. Microbiology 156: 2566 - 2574.

37. Wu B, Zhang B, Feng X, Rubens JR, Huang R, et al. (2010) Alternative isoleucine synthesis pathway in cyanobacterial species. Microbiology 156: 596-602.

38. Feng X, Mouttaki H, Lin L, Huang R, Wu B, et al. (2009) Characterization of the Central Metabolic Pathways in Thermoanaerobacter sp. X514 via Isotopomer-Assisted Metabolite Analysis. Appl Environ Microbiol 75: 5001-5008.

39. Tang YJ, Shui WQ, Myers S, Feng X, Bertozzi C, et al. (2009) Central metabolism in Mycobacterium smegmatis during the transition from O2-rich to O2-poor conditions as studied by isotopomer-assisted metabolite analysis. Biotechnol Lett 31: 1233-1240.

40. Tang YJ, Martin HG, Deutschbauer A, Feng X, Huang R, et al. (2009) Invariability of central metabolic flux distribution in *Shewanella oneidensis* MR-1 under environmental or genetic perturbations Biotechnol Prog 25: 1254-1259

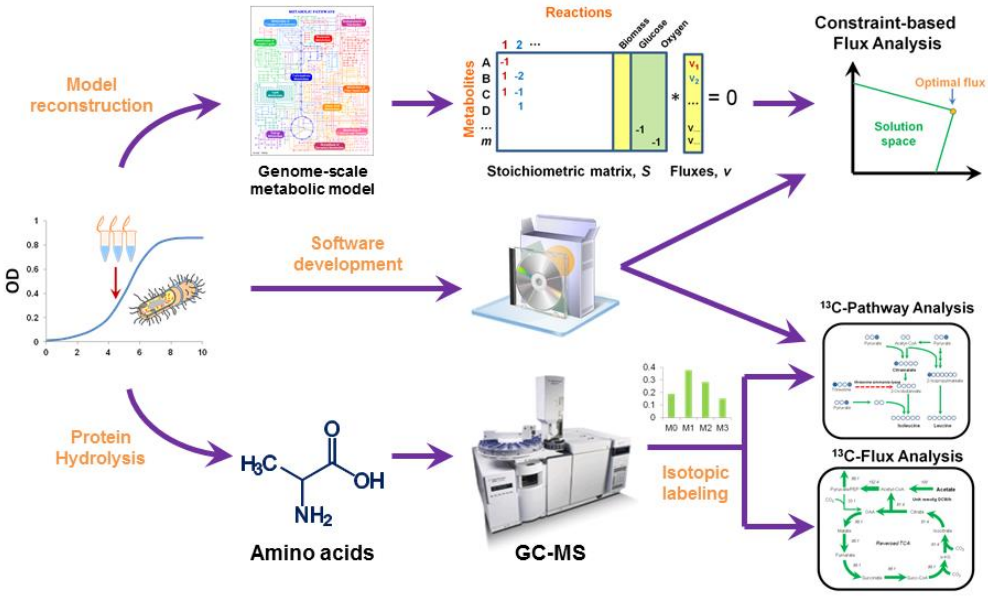
41. Tang K-H, Feng X, Tang YJ, Blankenship RE (2009) Carbohydrate metabolism and carbon fixation in Roseobacter denitrificans OCh114. PLoS One 4: e7233.

42. Tang K-H, Feng X, Zhuang W-Q, Alvarez-Cohen L, Blankenship RE, et al. (2010) Carbon flow of Heliobacterium modesticaldum is more related to Firmicutes than to the green sulfur bacteria. J Biol Chem 285: 35104-35112.

43. Zhuang WQ, Yi S, Feng X, Zinder SH, Tang YJ, et al. (2011) Selective utilization of exogenous amino acids by *Dehalococcoides ethenogenes*

strain 195 and the enhancement resulted to dechloronation activity. Appl Environ Microbiol 77: 7797-7803.

Figure 1. Fluxomic analysis of microbial metabolisms.



Chapter 2

Bridging the Gap between Fluxomics and Industrial Biotechnology

Xueyang Feng[1], Lawrence Page[2], Jacob Rubens[1], Lauren Chircus[1], Peter Colletti[1], Himadri B. Pakrasi[1, 2], Yinjie J. Tang[1,*]

[1]Department of Energy, Environmental and Chemical Engineering;

[2]Department of Biology, Washington University, St. Louis, MO 63130, USA

*:Corresponding author: yinjie.tang@seas.wustl.edu; phone: 314-935-3441

This chapter was previously published in

Feng X, Page L, Rubens J, Chircus L, Colletti P, Pakrasi HB, Tang YJ, "Bridging the gap between fluxomics and industrial biotechnology." J Biomed Biotechnol. 2010, 2010:460717.

2.1 Abstract

Metabolic flux analysis is a vital tool used to determine the ultimate output of cellular metabolism and thus detect biotechnologically relevant bottlenecks in productivity. ^{13}C-based metabolic flux analysis (^{13}C-MFA) and flux balance analysis (FBA) have many potential applications in biotechnology (1). However, noteworthy hurdles in fluxomics study are still present. First, several technical difficulties in both ^{13}C-MFA and FBA severely limit the scope of fluxomics findings and the applicability of obtained metabolic information. Second, the complexity of metabolic regulation poses a great challenge for precise prediction and analysis of metabolic networks, as there are gaps between fluxomics results and other omics studies. Third, despite identified metabolic bottlenecks or sources of host stress from product synthesis, it remains difficult to overcome inherent metabolic robustness or to efficiently import and express non-native pathways. Fourth, product yields often decrease as the number of enzymatic steps increases. Such decrease in yield may not be caused by rate-limiting enzymes, but rather is accumulated through each enzymatic reaction. Fifth, high throughput fluxomics tool hasn't been developed for characterizing non-model microorganisms and maximizing their application in industrial biotechnology. Refining fluxomics tools and understanding these obstacles will improve our ability to engineer highly-efficient metabolic pathways in microbial hosts.

Key words: pathway, ^{13}C-based, flux balance, post-transcriptional, robustness, rate-limiting

2.2 Introduction

Numerous chemical compounds, ranging from the anti-malaria drug artemisinin (2) to the "biofuel" butanol (3,4), have been produced with the aid of synthetic biology tools. The ability to efficiently synthesize natural or unnatural products requires a systems-level understanding of metabolism. Functional genomics tools such as genome sequencing, profiling of mRNA transcripts, and proteomics, are widely used to attain a comprehensive knowledge of how metabolic components (genes, proteins and metabolites) are regulated. In contrast to traditional omics tools, flux analysis (measurement of metabolite turnover rates) has become instrumental for physiological prediction and enzymatic rate quantification in metabolic networks (5). This technology also allows for the identification of metabolic interactions and the knowledge-based design of cellular functions. As such, one can utilize this tool to rationally modify biological hosts and analyze global physiological changes resulting from genetic modifications.

Fluxomics, the cell-wide quantification of intracellular metabolite turnover rates, was first performed via Flux Balance Analysis (FBA). This method uses the stoichiometry of the metabolic reactions in addition to a series of physical, chemical and biological characteristics (thermodynamics, energy balance, gene

regulation, etc.) to constrain the feasible fluxes under a given objective function (e.g., maximal biomass production). FBA is an underdetermined model (the number of constraints is smaller than the number of reactions in the metabolic network), which may give unrealistic metabolic readout. In spite of this limitation, FBA provides a useful framework for predicting a wide variety of cellular metabolisms. A complementary approach, ^{13}C-based metabolic flux analysis (^{13}C-MFA) allows for precise determinations of metabolic status under a particular growth condition. The key to ^{13}C-MFA is isotopic labeling, whereby microbes are cultured using a carbon source with a known distribution of ^{13}C. By tracing the transition path of the labeled atoms between metabolites in the biochemical network, one can quantitatively determine intracellular fluxes.

Flux analysis can not only provide genetic engineers with strategies for "rationally optimizing" a biological system, but also reveal novel enzymes useful for biotechnology applications (5). However, flux analysis platforms are still not routinely established in biotechnology companies. This review paper addresses current developments and challenges in the field of fluxomics, which may guide future study to bridge the gap between systems analysis of cellular metabolism and application in biotechnology.

2.3 Advances and limitations in metabolic flux analysis

2.3.1 Steady-state flux model

FBA and ^{13}C-MFA concentrate on the stoichiometric (rather than kinetic) properties of metabolic networks. FBA has been widely applied to predict cell growth rate, product yield using different feedstocks, lethality of gene knockouts, and advantageous pathway modifications (6). Such a model provides general guidelines for metabolic engineering and thus is a viable first step towards improving biosynthetic yield (7). The hallmark of large scale FBA is the COnstraint-Based Reconstruction and Analysis Toolbox (COBRA) (8), which provides a general platform for fluxomics studies.

A number of optimization algorithms and computational strategies for resolving *in silico* and *in vivo* inconsistencies have been proposed to improve the applicability of FBA (7,9). For example, incorporation of thermodynamic principles into FBA can constrain solution space (i.e., energy balance analysis) and obtain both stoichiometrically and thermodynamically feasible fluxes (1). To describe the "non-optimal" metabolic behaviors, FBA can use a bi-level optimization approach to estimate the potential trade-off between biomass accumulation and the yield of a desired product (10). FBA can also relax the objective function for maximization of the biomass and apply a Minimization of Metabolic Adjustment Algorithm to solve fluxes in mutant strains (1). Such an algorithm calculates fluxes by minimizing the difference between the wild-type flux distributions and the knockout-strain fluxes. Furthermore, FBA can be integrated with metabolic pathway analysis (MPA). MPA is a qualitative

technique that examines functional routes existing in a metabolic network without requiring knowledge of reaction rates (11). Combining MPA with FBA can quantitatively trace the plausible paths for optimal product synthesis, calculate cellular metabolism, and predict phenotypes under genetic manipulations or culture conditions (12). One main advantage of FBA is its capability for genome-scale modeling (including thousands of reactions), which bridge genomic annotation and functional metabolic output. Accordingly, the number of FBA models has increased exponentially since 1999 (13).

^{13}C-MFA aims to rigorously quantify pathway activities in intracellular metabolism by using both the isotopic labeling approach and *in silico* computation. ^{13}C-MFA is accomplished by feeding microbes a ^{13}C-labeled carbon source, measuring the enrichment pattern of the isotopomer in metabolites (e.g. amino acids), and deciphering the fluxes via computational routines (14). Since carbon fluxes through a metabolic network generate unique labeling patterns in metabolites, the overall flux distributions can be determined using isotopomer information. Advances in ^{13}C-MFA, including mass spectrometry-based metabolomics and isotopomer modeling approaches (such as novel algorithm using elementary metabolite units), have been discussed in recent papers (5,15,16).

Furthermore, open-source software has recently been published that facilitates *in silico* modelling. For example, WEbcoli is web-based software for

flux balance analysis of *E.coli* (17). In addition, OpenFLUX is computationally efficient software for ^{13}C-MFA (15), which incorporates the Elementary Metabolite Unit (EMU) framework for calculation of isotopomer balances (18). User-friendly software such as this allows biologists to perform fluxomics studies with little programming knowledge.

Methodologies for FBA and ^{13}C-MFA share two key characteristics: the use of a metabolic network and the assumption of a steady metabolic state (for internal metabolites). However, the two techniques have different purposes. FBA profiles the "optimal" metabolism for the desired performance; ^{13}C-MFA measures *in vivo* operation of a metabolic network. The two approaches to flux analysis are complementary when developing a rational metabolic engineering strategy. By comparing existent metabolic fluxes which were empirically determined via ^{13}C-MFA to the optimal metabolisms predicted by both FBA and other "omics" tools (such as transcription analysis), one can deduce gene targets for solving biotechnologically relevant productivity bottlenecks (19). Figure 2.1 shows that iterative flux analysis and genetic engineering of microbial hosts can remove competitive pathways or toxic byproducts, amplify genes encoding key metabolites, and balance energy metabolism (7).

2.3.2 Metabolic control and dynamic flux analysis

FBA and ^{13}C-MFA disregard dynamic intracellular behavior. This avoids the difficulties in developing kinetic models and performing intracellular

experimental measurements. However, many biological systems may not maintain a meaningful metabolic (or isotopic) steady state during the fermentation process (20,21,22). The description of metabolic perturbation and regulatory mechanisms requires kinetic modeling and control theories. For example, metabolic control analysis (MCA) couples local enzyme kinetics with systematic behavior to predict the control exerted on the targeted pathways by different components (e.g. transcription, enzymes) (23). Although MCA is not a quantitative measurement of flux, MCA can pinpoint bottle-neck enzymes (enzymes having the largest effect on the desired flux) in a pathway and allow the analysis of steady-state metabolism in response to changes in the cellular environment (24). In addition to MCA, the cybernetic approach (a model based on process dynamics and control) has been introduced for study of multi-enzyme systems and metabolic regulation (25). By incorporating both the enzyme kinetics in pathways and the enzyme synthesis kinetics, the cybernetic approach emphasizes microbial process dynamics and control during complicated fermentations (26).

Both MCA and the cybernetics approach focus on a simplified pathway network. To perform cell-wide quantitative analysis of a dynamic system, it is necessary to integrate the kinetic modeling with FBA and ^{13}C-MFA. Dynamic FBA (dFBA) has been developed to illuminate changing global enzyme activities (27,28). To avoid ordinary differential equations and dynamic optimization for describing intracellular metabolism, dFBA can use the Static Optimization

Approach (SOA) (29) which divides the time-course into numerous small intervals. At each time interval, a steady-state flux is calculated under the assumption of fast intracellular dynamics. By combining stoichiometric FBA for intracellular metabolism with dynamic mass balances on extracellular substrates and products, it is possible to reconstruct dFBA model for genome-scale analysis of microbial metabolisms in industrial fermentations, where product synthesis is often under dynamic control (30,31).

Recently, ^{13}C-dMFA (dynamic metabolic flux analysis) has been developed for isotopically nonstationary cultures. To profile the flux distributions for fed-batch cultures (slow dynamic metabolism), isotopic pseudo-steady state was assumed and two dilution parameters were introduced to account for isotopic transients. Another approach (Kinetic Flux Profiling) for solving intracellular fluxes is to create a sudden increase of the portion of ^{13}C in the substrate feed, then measure time-course samples as ^{13}C moves from the substrate into the metabolites (32). The fluxes can be calculated based on the rates of isotopic enrichment multiplied by the intracellular metabolite concentrations. A similar principle has been proposed for the flux analysis of photoautotrophic microorganisms (33) and *E. coli* in an isotopic transient phase (34). If the culture is under both metabolic and isotopic non-stationary state, exploratory and sophisticated ^{13}C-dMFA (dynamic ^{13}C-MFA) models have to be used to calculate both metabolic and isotopic kinetics (20,35,36). To solve the ^{13}C-dMFA problem

efficiently, a set of computational algorithms have been developed for tracing non-stationary isotopomer labeling in response to *in vivo* flux distributions (20,35,36). The EMU (elementary metabolite unit) framework has also been applied in ^{13}C-dMFA (18,37), because such algorithm can significantly improve computational times for tracing the labeling information (38). To avoid extensive simulation of dynamic isotopomer patterns, the SOA has to be applied by dividing the growth period into small time intervals (30~60 min), then the "mini" quasi-steady state ^{13}C-MFA can be applied at each time interval based on constraints from simultaneous isotopomer analysis of the fast turnover metabolites (39). By examining flux profiles over all time intervals, one can resolve the metabolic transients during the entire cultivation period.

2.3.3 Technical limitations of fluxomics

Cell-wide fluxomics tools (i.e., FBA and ^{13}C-MFA) have technical limitations. In genome-scale FBA models, the number of constraints (i.e., the availability of quantitative metabolite data) is much smaller than the number of reactions in the metabolic network. The calculation of such underdetermined systems depends on objective functions where one assumes that the metabolism optimizes its native ''goals'' (such as biomass or co-factor production) (40). This optimization principle has been questioned for several reasons. First, biological systems (e.g., *Bacillus subtilis*) seem to display sub-optimal growth performance (41). Second, a previous study examined 11 objective functions in *E. coli* and

found no single objective function that can perfectly describe flux states under various growth conditions (42). For example, unlimited aerobic growth on glucose is best described by a nonlinear maximization of the ATP yield per flux unit, but nutrient-limited continuous cultures favor biomass yield as the objective function. Third, some native cellular processes cannot be simply described by FBA. For example, cyanobacterial species (i.e., *Cyanothece* 51142) maintain their circadian rhythms (e.g., nitrogen fixation and light dependent reaction activities) under nutrient-sufficient and continuous light conditions (43,44).

The application of ^{13}C-MFA in industrial biotechnology also has several bottlenecks. The most prevalent constriction occurs because current techniques are insufficient for measuring large-scale metabolic networks. Obtaining labeling information of free metabolites rather than amino acids and solving large-scale nonlinear flux models pose two key challenges. As a result, most obtained flux information is limited to central metabolism. To date, only two large-scale ^{13}C-MFA (>300 reactions) have been reported, but many fluxes in their reports cannot be precisely determined due to insufficient constraints (45,46). The genome-scale ^{13}C-MFA is still in its infancy and requires further development of the relevant experimental techniques and computational tools (47). A second issue is that ^{13}C-dMFA is still poorly developed for determining dynamic metabolic behavior. It is difficult for rapid sampling and precise measurements of metabolites at short time intervals throughout the entire culture period. For example, to measure absolute

intracellular metabolite concentrations, one has to grow cell in fully ^{13}C-labeled medium, then the labeled cells are extracted with quenching solvent containing known concentrations of unlabeled internal standards (the concentrations of metabolites are calculated using the isotope ratio-based approach) (48). Such measurement requires extremely high cost of analytical efforts including quick sampling, rapid metabolite extraction, and a high resolution LC-MS instrument. Furthermore, the time-dependent model includes ordinary differential equations and significantly increases the computational complexity (20,34). Third, flux determination assumes that enzymatic reactions are homogenous inside the cell and that there are no transport limitations between metabolite pools. However, eukaryotes have organelles (compartments) that may have diffusion limitations or metabolite channeling (14,49). Compartmentalization of amino acid biosynthesis further clouds the obtained amino acid-based labeling information (50). Therefore, confident ^{13}C-MFA for eukaryotes not only requires the combination of different analytical tools (GC-MS, LC-MS and NMR) to obtain extensive labeling information (51), but also adequate sample processing and extraction methods (e.g. separation of compartments by ultracentrifugation). A fourth problem is that some industrial hosts and the great majority of environmental microbes resist cultivation in minimal media, and introducing other nutrient sources often significantly complicates metabolite labeling measurements and flux analyses (52). Finally, a microbial community demonstrates complex metabolic interactions between species. To date, only a few FBA models have

been developed for community studies (53,54). The exchange of metabolites among species is nearly impossible to unravel by ^{13}C-MFA because complete separation and measurement of metabolites from a single species in a microbial community is impossible (5). These technical limitations in both FBA and MFA models are responsible for the gap between fluxomics and its applicability in biotechnology.

2.4 Integration of fluxomics with other "omics"

It is desirable to integrate the concepts of systems biology (which combines the readouts from transcription as well as protein/metabolite profiling) with fluxomics (Figure 2.2) (47). For example, ^{13}C-MFA, enzyme activity assays, and RT-PCR analysis can be used together to study *E.coli* mutants' metabolism (55). Additionally, the responses of *E. coli* to genetic modification have been systematically examined by utilizing multiple high-throughput "omics" methods (56). The results illuminate relatively small changes in mRNA and proteins in response to genetic disruptions, which allow the cell to maintain a stable metabolic state under changing growth conditions. A similar approach to the study of *Synechocystis* 6803 has shown that the regulation of some enzymes is sensitive to light conditions (57). Many other regulatory mechanisms, however, still remain unknown. Furthermore, global regulators in industrial microorganisms have been successfully identified by correlating transcript/transduction levels and metabolic

fluxes (58,59,60,61). The discovery of functioning regulators provides insight to the entire regulation in metabolic network.

On the other hand, challenges in integrated "omics" studies are also present. The lack of understanding of metabolic regulation at different metabolic levels complicates the rational design of biological systems, which is a major barrier in industrial biotechnology. For example, post-transcriptional regulation poses a significant challenge in integrating fluxomics with other "omics" studies. It is well known that transcript and protein data correlate relatively well for specific pathways, yet this correlation can be poor in cell-wide analyses (62). Furthermore, most mRNA expression studies insufficiently predict enzyme activities or flux changes in many *E. coli* pathways (63). In studies on the adaptation of *E. coli* to environmental perturbations, the tricarboxylic acid cycle is found to correlate well with molecular changes at the transcriptional level, but flux alterations in other central metabolic pathways seem to be uncorrelated to changes in the transcriptional network (64). Because of the complexity of regulatory mechanisms spanning multiple cellular processes, fluxomics and other "omics" studies may have inconsistent observations which complicate systems-level analyses.

2.5 Fluxomics of microbes for industrial biotechnology

FBA allows *in silico* simulations of metabolism in "industrial workhorses," from which desired strains or targeted mutations can be identified.

^{13}C-MFA can assess *in vivo* metabolism of engineered strains under specific growth conditions and validate FBA results. Here, we summarize recent applications of FBA and ^{13}C-MFA for commonly-used industrial chassis (i.e., *E. coli*, *B. subtilis* and *S. cerevisiae*) and for non-model microorganisms (i.e., less-characterized or newly-discovered microorganisms).

2.5.1. *Escherichia coli* model

E. coli is the most commonly utilized species in fermentation industry. *E. coli* flux models were reported as early as the 1990s (65,66). For biotechnology applications, the Liao group first applied metabolic pathway analysis (MPA) to guide the genetic manipulation of *E.coli* strains and channel the metabolic fluxes from carbohydrate to the aromatic amino acid pathway (67). The Maranas group has integrated cell growth and product synthesis in the OptKnock toolbox (10) and applied it to construct high performance mutants. The computer-aided designs have shown improved lactic acid, succinate, and 1,3-propanediol production (68). FBA can predict lethality in a metabolic network where deletions of more than one non-essential gene mutants may trigger the death of the organism. For example, the Maranas group (69) analyzed the gene/reaction essentiality in a genome-scale model of *E. coli* and systemically identified possible pairs of synthetic lethals: non-essential genes whose simultaneous knockouts would have a potentially lethal effect. Incorporating information about synthetic lethality into the new model will curb the construction of ill-designed biological systems for

biotechnology. Furthermore, FBA can be used to find rate-limiting steps for product synthesis. For example, FBA revealed gene targets, and modification of those genes (i.e., knocking out the genes for pyruvate forming enzymes, over-expression of the glyoxylate shunt and glucose transport system) resulted in more than a ten-fold increase in succinate production (70,71,72). FBA has also been used to improve genetic strategies for the overproduction of secondary metabolites, such as amino acids (73) and lycopene (74).

Besides genetic strategies, FBA can provide useful information for the design of optimal fermentation conditions. For example, an FBA model was used to identify nutrient limitations during recombinant interleukin-2 (IL-2) production in *E. coli.* By supplementing specific amino acids, IL-2 production increased two-fold in fed-batch fermentation (75). Recently, a reactor-scale dFBA model was developed via a Static Optimization Approach to analyze *E. coli* metabolism for the production of a biopharmaceutical drug (27). dFBA contains a steady state FBA model embedded within a dynamic kinetic model that describes the time evolution of fermentation process variables (e.g., biomass growth, glucose consumption and products synthesis). Such a model provided guidelines for the optimization of fermentations at the scale of a 1000L process.

The ^{13}C-MFA model was first used to investigate metabolic regulation in *E.coli* under different genetic and environmental conditions (76). ^{13}C-MFA has also been used to examine various biotechnological processes involved in the

production of pharmaceuticals, amino acids and polymers. A large scale ^{13}C-MFA with over 300 reactions was successfully developed for amorphadiene (a precursor of the anti-malaria drug) producing *E. coli* strains (45). Another study revealed a growth phase-dependent metabolic shift in a lysine-producing *E. coli* strain (77). This work was performed in a fed-batch culture with rich medium (containing yeast extract), and metabolic fluxes in both exponential growth and stationary phases were estimated by measuring free metabolites. Metabolic analysis of the stationary phase is important since many products are synthesized during a non-growth phase. In a third example, ^{13}C-MFA of a 1,3-propanediol producing *E. coli* strain was conducted in fed-batch fermentation (78). The ^{13}C-MFA results showed a decrease in the split ratio between glycolysis and the pentose phosphate pathway over the time-course of the culture in response to increasing 1,3-propanediol fluxes.

2.5.2 *Bacillus subtilis* model

B. subtilis is the industrial organism of choice for the production of vitamins, antibiotics, enzymes, and nucleosides. The FBA model for *B. subtilis* was constructed based on a combination of genomic, biochemical, and physiological information (79). The FBA model was iteratively corrected and improved using information from high-throughput phenotypic screens of mutants, substrate utilization, gene essentiality, and sequence analyses. The *B. subtilis* flux model is mostly studied for riboflavin production, focusing on four aspects:

investigating phenotypes of wild type and knock-out strains, assessing production capacity, identifying the impact of different carbon sources on biosynthesis, and characterizing the cellular response to different culture conditions. The Sauer group has extensively investigated riboflavin-producing strains. They first used an FBA model to quantify growth maintenance coefficients, the maximum growth yield, and the specific riboflavin production rate in continuous cultivation (80). Later on, they applied ^{13}C-MFA to the same strain and found that genetic manipulations should target the NADPH balance and riboflavin biosynthetic pathways (81). In other studies on *B. subtilis*, they revealed several guidelines for high yield riboflavin production: 1. they compared the metabolic flux distributions and maintenance energy of eight *Bacillus* strains and discovered that *B. licheniformis* was the most suitable for industrial biotechnology (82); 2. they found that using malate as a substrate resulted in a suppressed respiratory TCA cycle and an enhanced overflow metabolism (83); 3. they found the pentose precursors of riboflavin were mainly synthesized via the non-oxidative pentose-phosphate pathway, so any suggested genetic modification should decrease the activity of the oxidative pentose phosphate pathway (84). Recently, they developed a ^{13}C-dMFA model for *B. subtilis* to identify the metabolic response of riboflavin overproduction under a glucose-limited fed-batch culture (39). This dynamic flux analysis was obtained by recording changes in labeling patterns of intracellular amino acids under a metabolic pseudo-steady state assumption.

2.5.3 *Saccharomyces cerevisiae* model

S. cerevisiae is a robust eukaryotic chassis used for the expression of a wide range of products. For example, flux analysis revealed target genes in two native pathways for the over-expression of succinate: the TCA and glyoxylate cycles (85). Another study showed the enhancement of sesquiterpene production via *in silico* driven metabolic engineering (86). Additionally, flux analysis has been extensively applied for improving ethanol production. First, a number of strategies were developed for the metabolic engineering of redox processes in *S. cerevisiae*, resulting in a decrease in the yield of glycerol by 40% and an increase in ethanol production under both glucose and xylose/glucose growth conditions (87). Second, Dikicioglu *et al.* (88) applied a genome-scale FBA model to analyze respiration-deficient mutants of *S. cerevisiae* for ethanol production. They found that many genetic manipulation strategies (e.g., the overexpression of the glutamate synthase gene) were unnecessary in a respiration-deficient metabolic background. This indicates that the rate limiting steps for ethanol production can change after the initial genetic manipulations of targeted genes. Third, a ^{13}C-MFA model was used to screen ethanol production in 14 hemiascomycetous yeast strains (50). This study suggests that *S. cerevisiae* is the ideal ethanol production candidate due to a strong NADPH-driven pentose phosphate pathway. Other ^{13}C-MFA studies characterized the metabolic shift between oxidative growth and fermentative growth with ethanol production (89), investigated alternative carbon

substrate (xylose) metabolisms (90), revealed key factors influencing biomass growth on xylose (91), and examined the consumption of ethanol and other storage carbohydrates in a glucose-limited chemostat culture (92).

Furthermore, a genome-scale FBA indicates an apparent enzyme dispensability, i.e., 80% of yeast genes seem to be non-essential for viability under laboratory conditions (93). The FBA illustrated the influence of non-essential genes on metabolic robustness and environmental fitness due to genetic buffering through alternative genes, while a ^{13}C-MFA (consisting of over 700 reactions) revealed a similar effect of metabolic network robustness on null mutations (46). Understanding the role of these redundant genes is important for a valid and efficient genetic modification.

2.5.4 Non-model microorganisms

Fluxomics is an important tool for the rigorous study of metabolism in less-characterized microbes that provides novel insights for application of these species to biotechnology. However, fluxomics have not been sufficiently applied to non-model microorganisms as compared to model microbial hosts. Table 2.1 summarizes some milestone papers in fluxomics studies on non-model species that are potentially useful for synthetic biology. Compared to the work done in the field of fluxomics for industrial workhorses, far fewer studies have been performed on non-model microorganisms. This is due to the complicated growth conditions, poorly-understood metabolic networks, and significant lack of genetic

and molecular biology tools. However, non-model environmental microorganisms are also important for industrial biotechnology because they often possess native biochemical pathways for chemical synthesis or the ability to utilize cheap substrates (94). Furthermore, flux analysis can be used to discover novel enzymes that can be cloned into industrial microbes to improve their capacity for product synthesis. For example, ^{13}C-MFA revealed a citramalate pathway for isoleucine biosynthesis (independent of the common threonine ammonia-lyase pathway) (95,96). Citramalate synthase, which has also been detected in some environmental bacteria (97,98,99), can be engineered into *E. coli* for 1-propanol and 1-butanol production. The new pathway bypasses threonine biosynthesis and represents the shortest keto-acid-mediated pathway; as such, it improved biofuel yield 9 to 22-fold (100). Currently, high-throughput genome sequencing methods are mapping genomes in novel microbes at a pace that far exceeds the pace of functional characterization of these species. Therefore, a high throughput ^{13}C-MFA technique is required for screening non-model microorganisms for new enzymes and maximizing their application in industrial biotechnology (5).

2.6 Finding bottlenecks for Industrial Biotechnology

One of the main goals of fluxomics is to indentify bottlenecks for industrial biotechnology and thereby assist in the creation of rational engineering strategies. Simple measurements of metabolism, however, are not enough to overcome unpredictable challenges in industrial biotechnology. Metabolic

regulation is very complex, and systems biology tools are incapable of revealing a general strategy for synthetic biology (101).

Bottlenecks in industrial biotechnology can be explained from the view of fluxomics. First, metabolic robustness (the ability to maintain metabolic performance under genetic or environmental perturbations) is a long-recognized key property of microbial systems (102). This basic mechanism is often responsible for the gap between computationally aided design and final experimental outcomes. For example, a ^{13}C-MFA study indicates that *E. coli* shows remarkable robustness in the central carbon metabolism in the presence of genetic variation, and is even more flexible in response to altered environmental conditions (e.g., different nutrients or oxygen levels) (76). Analyses of *E. coli* components at multiple "omics" levels also reveal unexpectedly small changes in messenger RNA, proteins and metabolite levels for most genetic disruptions. This is because *E. coli* actively regulates enzyme levels to maintain a stable metabolic state in the presence of perturbations (56,64). Similarly, *B. subtilis* shows rigidity and suboptimal performance for its flux regulation in over 137 genetic backgrounds (41). Furthermore, gene essentiality and pairwise genetic interactions have been investigated in *S. cerevisiae* (93,103). It has been found that a gene's function is buffered by duplication in *S. cerevisiae* genomic DNA or by an alternative biochemical pathway. Although only 13% of genes were suggested to be essential by single knockout experiments, simultaneous deletion

of pairs of non-essential genes (>70% of the total metabolic genes) were found to inhibit growth. Invariability of metabolic flux under mutagenic genotypes seems to be an important feature in many biological systems, and thus successful metabolic strategies highly depend on an understanding of robust cellular nature (104,105,106).

Metabolic engineering of industrial chassis is based on the premise that the yield of a desired product can be increased by identifying and over-expressing the enzymes that catalyze the rate-limiting steps in a given metabolic pathway. However, a method based on over-expressing rate-limiting enzymes will only work if these rate-limiting enzymes exist and remain rate-limiting when their activities are increased. Previous studies have shown that the commonly-believed "rate-limiting" enzymes may not exist in some industrial microbes and an increase in productivity has to be achieved by coordinated expression of entire pathways (107). Furthermore, rate-limiting steps in a metabolic network often shift after initial targets have been engineered. For example, phenotypic data in *S. cerevisiae* mutants revealed that some FBA-predicted gene targets for ethanol production are invalid if the cell's respiratory genes have been knocked-out (88). Another example of this phenomenon is highlighted by the metabolic consequences of the deletion of the methionine and cysteine biosynthesis repressor protein (McbR) in *Corynebacterium glutamicum*, which yielded no overproduction of methionine but drastic accumulation of homolanthionine (108).

The above evidence indicates that rate-limiting steps often shift after initial targets have been engineered. Additionally, simultaneous importation and expression of a few heterologous genes to improve the rate-limiting pathway may fail if the non-native pathway is incompatible with the host. These efforts often lead to metabolic imbalance and accumulation of toxic metabolites (3,4).

Based on the recent publications, we have constructed a linear regression model which shows that the yield of biosynthetic products decreases exponentially as a function of the steps away from central metabolism in *S. cerevisiae* (Figure 2.3). It is easier to achieve high carbon fluxes to the central metabolites, possibly because enzyme efficiency in central metabolism is usually high (109). However, the yields of secondary metabolites are smaller because each additional enzymatic step may not be perfectly efficient (model regression shows an average of ~67% efficiency in each enzymatic step in secondary metabolisms). This loss of yield is unavoidable due to the metabolism channeling the intermediates away from the desired product. Potential solutions to this problem include: 1. designing host-compatible enzymes with high product specificity (110); 2. feeding intermediates to the cell to reduce the number of enzymatic steps to final product (111); 3. creating synthetic protein scaffolds, which significantly improve intermediate conversion efficiency and overall biosynthetic yield (112).

In conclusion, fluxomics studies enable the quantification of intracellular metabolism. However, this tool is not fully developed, and it remains difficult to deduce cell-wide pathway bottlenecks and to provide effective strategies for biotechnology applications. Numerous technical difficulties in developing flux analysis methods and complicated metabolic regulatory mechanisms have severely limited the scope of fluxomics in industrial biotechnology. It is necessary for the future development of flux analysis to combine other advanced "omics" analysis and molecular biology techniques to resolve challenges in the fluxomics fields.

2.7 References

1. Boghigian BA, Seth G, Kiss R, Pfeifer BA (2010) Metabolic Flux Analysis for Pharmaceutical Production. Metab Eng 12: 81-95.

2. Ro DK, Paradise EM, Ouellet M, Fisher KJ, Newman KL, et al. (2006) Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature 440: 940-943.

3. Atsumi S, Higashide W, Liao JC (2009) Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nat Biotechnol 27: 1177-U1142.

4. Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451: 86-89.

5. Zamboni N, Sauer U (2009) Novel biological insights through metabolomics and 13C-flux analysis. Curr Opin Microbiol 12: 553-558.

6. Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ, et al. (2010) Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. Metab Eng 12: 173-186.

7. Blazeck J, Alper H (2010) Systems metabolic engineering: Genome-scale models and beyond. Biotechnol J

8. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. Nat Protoc 2: 727-738.

9. Kumar VS, Maranas CD (2009) GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. PLoS Comput Biol 5: e1000308. doi:1000310.1001371/journal.pcbi.1000308.

10. Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84: 647-657.

11. Trinh CT, Wlaschin A, Srienc F (2009) Elementary mode analysis: a useful metabolic pathway analysis tool for characterizing cellular metabolism. Appl Microbiol Biotechnol 81: 813-826.

12. Schilling CH, Edwards JS, Letscher D, Palsson BØ (2000-2001) Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnol Bioeng 71: 286-306.

13. Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, et al. (2009) A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189. PLoS Comput Biol 5: e1000285. .

14. Zamboni N, Fendt SM, Rühl M, Sauer U (2009) (13)C-based metabolic flux analysis. Nat Protoc 4: 878-892.

15. Quek L-E, Wittmann C, Nielsen LK, Krömer JO (2009) OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. Microb Cell Fact 8.

16. Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EK, et al. (2009) Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. Mass Spectrom Rev 28: 362-375.

17. Jung T-S, Yeo HC, Reddy SG, Cho W-S, Lee D-Y (2009) WEbcoli: an interactive and asynchronous web application for in silico design and analysis of genome-scale *E.coli* model Bioinformatics 25: 2850-2852

18. Antoniewicz MR, Kelleher JK, Stephanopoulos G (2007) Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. Metabolic Engineering 9: 68-86.

19. Tang YJ, Sapra R, Joyner D, Hazen TC, Myers S, et al. (2009) Analysis of Metabolic Pathways and Fluxes in a Newly Discovered Thermophilic and Ethanol-Tolerant Geobacillus Strain. Biotechnol Bioeng 102: 1377-1386.

20. Nöh K, Grönke K, Luo B, Takors R, Oldiges M, et al. (2007) Metabolic flux analysis at ultra short time scale: isotopically non-stationary ^{13}C labeling experiments. J Biotechnol 129: 249-267.

21. Wahl SA, Nöh K, Wiechert W (2008) ^{13}C labeling experiments at metabolic nonstationary conditions: an exploratory study. BMC Bioinformatics 9: 152.

22. Tang YJ, Meadows AL, Keasling JD (2007) A kinetic model describing *Shewanella oneidensis* MR-1 growth, substrate consumption, and product secretion. Biotechnol Bioeng 189: 894-901.

23. Wildermuth MC (2000) Metabolic control analysis: biological applications and insights. Genome Biol 1.

24. Hoefnagel MHN, Starrenburg MJC, Martens DE, Hugenholtz J, Kleerebezem M, et al. (2002) Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis. Microbiology 148: 1003-1013.

25. Namjoshi AA, Hu W-S, Ramkrishna D (2003) Unveiling steady-state multiplicity in hybridoma cultures: the cybernetic approach. Biotechnol Bioeng 81: 80-91.

26. Song H-S, Morgan JA, Ramkrishna D (2009) Systematic development of hybrid cybernetic models: application to recombinant yeast co-consuming glucose and xylose. Biotechnol Bioeng 103: 984-1002.

27. Meadows AL, Karnika R, Lama H, Forestella S, Snedecor B (2010) Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation. Metab Eng 12: 150-160.

28. Gianchandani EP, Chavali AK, Papin JA (2010) The application of flux balance analysis in systems biology. Systems Biology and Medicine 2: 372-382.

29. Mahadevan R, Edwards JS, Doyle FJ (2002) Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J 83: 1331-1340.

30. Hjersted JL, Henson MA, Mahadevan R (2007) Genome-scale analysis of Saccharomyces cerevisiae metabolism and ethanol production in fed-batch culture. Biotechnol Bioeng 97: 1190-1204.

31. Oddonea GM, Millsb DA, Block DE (2009) A dynamic, genome-scale flux model of Lactococcus lactis to increase specific recombinant protein expression. Metab Eng 11: 367-381.

32. Yuan J, Bennett BD, Rabinowitz JD (2008) Kinetic flux profiling for quantitation of cellular metabolic fluxes. Nat Protoc 3: 1328-1340.

33. Shastri AA, Morgan JA (2007) A transient isotopic labeling methodology for 13C metabolic flux analysis of photoautotrophic microorganisms. Phytochemistry 68: 2302-2312.

34. Schaub J, Mauch K, Reuss M (2008) Metabolic flux analysis in Escherichia coli by integrating isotopic dynamic and isotopic stationary 13C labeling data. Biotechnol Bioeng 99: 1170-1185.

35. Nöh K, Wahl A, Wiechert W (2006) Computational tools for isotopically instationary ^{13}C labeling experiments under metabolic steady state conditions. Metab Eng 8: 554-577.

36. Wahl SA, Nöh K, Wiechert W (2008) 13C labeling experiments at metabolic nonstationary conditions: an exploratory study. BMC Bioinformatics 9.

37. Young JD, Walther JL, Antoniewicz MR, Yoo H, Stephanopoulos G (2008) An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. Biotechnol Bioeng 99: 686-699.

38. Suthers PF, Chang YJ, Maranas CD (2010) Improved computational performance of MFA using elementary metabolite units and flux coupling. Metab Eng 12: 123-128.

39. Rühl M, Zamboni N, Sauer U (2009) Dynamic Flux Responses in Riboflavin Overproducing Bacillus subtilis to Increasing Glucose Limitation in Fed-Batch Culture. Biotechnol Bioeng 105: 795-804.

40. Stephanopoulos GN, Aristidou AA, Nielsen J (1998) Metabolic Engineering Principles and Methodologies. San Diego: Academic Press. 75, 120-130 p.

41. 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.

42. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli.* Mol Syst Biol 3: 119.

43. ColonLopez M, Sherman DM, Sherman LA (1997) Transcriptional and translational regulation of nitrogenase in light-dark- and continuous-light grown cultures of the unicellular cyanobacterium Cyanothece sp. strain ATCC 51142. Journal of Bacteriology 179: 4319-4327.

44. Toepel J, Welsh E, Summerfield TC, Pakrasi HB, Sherman LA (2008) Differential transcriptional analysis of the cyanobacterium Cyanothece sp strain ATCC 51142 during light-dark and continuous-light growth. J Bacteriol 190: 3904-3913.

45. Suthers PF, Burgard AP, Dasika MS, Nowroozi F, Van Dien S, et al. (2007) Metabolic flux elucidation for large-scale models using ^{13}C labeled isotopes. Metab Eng 9: 387-405.

46. Blank LM, Kuepfer L, Sauer U (2005) Large-scale ^{13}C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. Genome Biol 6: R49.

47. Dauner M (2010) From fluxes and isotope labeling patterns towards in silico cells. Curr Opin Biotechnol 21: 55-62.

48. Bennett BD, Yuan J, Kimball EH, Rabinowitz JD (2008) Absolute quantitation of intracellular metabolite concentrations by an isotope ratiobased approach. Nature Protocol 3: 1299-1311.

49. Malaisse WJ, Zhang Y, Jijakli H, Courtois P, Sener A (2004) Enzyme-to-enzyme channelling in the early steps of glycolysis in rat pancreatic islets. The International Journal of Biochemistry & Cell Biology 36: 1510-1520.

50. Blank LM, Lehmbeck F, Sauer U (2005) Metabolic-flux and network analysis in fourteen hemiascomycetous yeasts. FEMS Yeast Res 5: 545-558

51. Kleijn RJ, Geertman J-MA, Nfor BK, Ras C, Schipper D, et al. (2007) Metabolic flux analysis of a glycerol-overproducing Saccharomyces cerevisiae strain based on GC-MS, LC-MS and NMR-derived C-labelling data. FEMS Yeast Res 7: 216-231.

52. Kaeberlein T, Lewis K, Epstein SS (2002) Isolating "Uncultivable" Microorganisms in Pure Culture in a Simulated Natural Environment. Science 296: 1127-1129.

53. Taffs R, Aston JE, Brileya K, Jay Z, Klatt CG, et al. (2009) In silico approaches to study mass and energy flows in microbial consortia: a syntrophic case study. BMC Syst Biol 3: 114.

54. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, et al. (2007) Metabolic modeling of a mutualistic microbial community. Mol Syst Biol 3: 92.

55. Shimizu K (2004) Metabolic flux analysis based on 13C-labeling experiments and integration of the information with gene and protein expression patterns. Adv Biochem Eng Biotechnol 91: 1-449.

56. Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, et al. (2007) Multiple high-throughput analyses monitor the response of *E. coli* to perturbations. Science 316: 593-597.

57. Yang C, Hua Q, Shimizu K (2002) Metabolic flux analysis in *Synechocystis* using isotope distribution from ^{13}C-labeled glucose. Metab Eng 4: 202-216.

58. Moxley JF, Jewett MC, Antoniewicz MR, Villas-Boas SG, Alper H, et al. (2009) Linking high resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator Gcn4p. Proc Natl Acad Sci 106: 6477~6482.

59. Lemuth K, Hardiman T, Winter S, Pfeiffer D, Keller MA, et al. (2008) Global transcription and metabolic flux analysis of Escherichia coli in glucose-limited fed-batch cultivations. Appl Environ Microbiol 74: 7002-7015.

60. Tännler S, Fischer E, Coq DL, Doan T, Jamet E, et al. (2008) CcpN controls central carbon fluxes in Bacillus subtilis. J Bacteriol 190: 6178-6187.

61. Nanchen A, Schicker A, Revelles O, Sauer U (2008) Cyclic AMP-dependent catabolite repression is the dominant control mechanism of metabolic fluxes under glucose limitation in Escherichia coli. J Bacteriol 190: 2323-2330.

62. Mukhopadhyay A, Redding AM, Rutherford BJ, Keasling JD (2008) Importance of systems biology in engineering microbes for biofuel production. Curr Opin Biotechnol 19: 228-234.

63. Hua Q, Joyce AR, Palsson BO, Fong SS (2007) Metabolic characterization of *Escherichia coli* adapted to growth on lactate. Applied and Environmental Microbiology 73: 4639-4647.

64. Fong SS, Nanchen A, Palsson BO, 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.

65. Varma A, Palsson BO (1994) Metabolic Flux Balancing: Basic concepts, Scientific and Practical Use Bio/Technology 12: 994-998.

66. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol 60: 3724-3731.

67. Liao JC, Hou S-Y, Chao Y-P (1996) Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. Biotechnol Bioeng 52: 129-140.

68. Fong SS, Burgard AP, Herring CD, Knight EM, Blattner FR, et al. (2005) In silico design and adaptive evolution of Escherichia coli for production of lactic acid. Biotechnol Bioeng 91: 643-648.

69. Suthers PF, Zomorrodi A, Maranas CD (2009) Genome-scale gene/reaction essentiality and synthetic lethality analysis. Mol Syst Biol 5.

70. Lee SY, Hong SH, Moon SY (2002) In silico metabolic pathway analysis and design: succinic acid production by metabolically engineered Escherichia coli as an example. Genome Inform 13.

71. Lee SJ, Lee D-Y, Kim TY, Kim BH, Lee J, et al. (2005) Metabolic engineering of Escherichia coli for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. Appl Environ Microbiol 71: 7880-7887.

72. Wang Q, Chen X, Yang Y, Zhao X (2006) Genome-scale in silico aided metabolic analysis and flux comparisons of Escherichia coli to improve succinate production. Appl Microbiol Biotechnol 73: 887-894.

73. Park JH, Lee SY (2008) Towards systems metabolic engineering of microorganisms for amino acid production. Curr Opin Biotechnol 19: 454-460.

74. Alper H, Jin Y-S, Moxleya JF, Stephanopoulos G (2005) Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli. Metab Eng 7: 155-164.

75. Yegane‐Sarkandy S, Farnoud AM, Shojaosadati SA, Khalilzadeh R, Sadeghyzade M, et al. (2009) Overproduction of human interleukin-2 in recombinant Escherichia coli BL21 high-cell-density culture by the determination and optimization of essential amino acids using a simple stoichiometric model. Biotechnol Appl Biochem 54: 31-39.

76. Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, et al. (1999) Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism. J Bacteriol 181: 6679-6688.

77. Iwatani S, Van Dien S, Shimbo K, Kubota K, Kageyama N, et al. (2007) Determination of metabolic flux changes during fed-batch cultivation from measurements of intracellular amino acids by LC-MS/MS. J Biotechnol 128: 93-111.

78. Antoniewicz MR, Kraynie DF, Laffend LA, González-Lergier J, Kelleher JK, et al. (2007) 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: 277-292.

79. Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. J Biol Chem 282: 28791-28799.

80. Sauer U, Hatzimanikatis V, Hohmann HP, Manneberg M, Loon APv, et al. (1996) Physiology and metabolic fluxes of wild-type and riboflavin-producing Bacillus subtilis. Appl Environ Microbiol 62: 3687-3696.

81. Sauer U, Hatzimanikatis V, Bailey JE, Hochuli M, Szyperski T, et al. (1997) Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. Nat Biotechnol 15: 448-452.

82. Tannler S, Decasper S, Sauer U (2008) Maintenance metabolism and carbon fluxes in Bacillus species. Microbial Cell Factories 7.

83. Kleijn RJ, Buescher JM, Chat LL, Jules M, Aymerich S, et al. (2010) Metabolic fluxes during strong carbon catabolite repression by malate in Bacillus subtilis. J Biol Chem 285: 1587-1596.

84. Zamboni N, Fischer E, Muffler A, Wyss M, Hohmann H-P, et al. (2005) Transient expression and flux changes during a shift from high to low riboflavin production in continuous cultures of Bacillus subtilis. Biotechnol Bioeng 89: 219-232.

85. Otero JM, Olssona L, Nielsen J (2007) Metabolic engineering of Saccharomyces cerevisiae microbial cell factories for succinic acid production. J Biotechnol 131: 205.

86. Asadollahi MA, Maury J, Patil KR, Schalk M, Clark A, et al. (2009) Enhancing sesquiterpene production in Saccharomyces cerevisiae through in silico driven metabolic engineering. Metab Eng 11: 328-334.

87. Bro C, Regenberg B, Förster J, Nielsen J (2006) In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production. Metab Eng 8: 102-111.

88. Dikicioglu D, Pir P, Onsan ZI, Ulgen KO, Kirdar B, et al. (2008) Integration of metabolic modeling and phenotypic data in evaluation and improvement of ethanol production using respiration-deficient mutants of Saccharomyces cerevisiae. Appl Environ Microbiol 74: 5809-5816.

89. Frick O, Wittmann C (2005) Characterization of the metabolic shift between oxidative and fermentative growth in Saccharomyces cerevisiae by comparative 13C flux analysis. Microb Cell Fact 4: 30.

90. Grotkjæra T, Christakopoulos P, Nielsena J, Olsson L (2005) Comparative metabolic network analysis of two xylose fermenting recombinant Saccharomyces cerevisiae strains. Metab Eng 7: 437-444.

91. Sonderegger M, Jeppsson M, Hahn-Hägerdal B, Sauer U (2004) Molecular basis for anaerobic growth of Saccharomyces cerevisiae on xylose, investigated by global gene expression and metabolic flux analysis. Appl Environ Microbiol 70: 2307-2317.

92. Windena WAv, Dama JCv, Rasa C, Kleijna RJ, Vinkea JL, et al. (2005) Metabolic-flux analysis of Saccharomyces cerevisiae CEN.PK113-7D based on mass isotopomer measurements of (13)C-labeled primary metabolites. FEMS Yeast Res 5: 559-568.

93. Papp B, Pál C, Hurst LD (2004) Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. Nature 429: 661-664.

94. Alper H, Stephanopoulos G (2009) Engineering for biofuels: exploiting innate microbial capacity or importing biosynthetic potential? Nat Rev Microbiol 7: 715-723.

95. Risso C, Van Dien SJ, Orloff A, Lovley DR, Coppi MV (2008) Elucidation of an alternate isoleucine biosynthesis pathway in Geobacter sulfurreducens. J Bacteriol 190: 2266-2274.

96. Tang YJ, Chakraborty R, Martin HG, Chu J, Hazen TC, et al. (2007) Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-NTA. Appl Environ Microbiol 73: 3859-3864.

97. Wu B, Zhang B, Feng X, Rubens JR, Huang R, et al. (2010) Alternative isoleucine synthesis pathway in cyanobacterial species. Microbiology 156: 596-602.

98. Feng X, Mouttaki H, Lin L, Huang R, Wu B, et al. (2009) Characterization of the Central Metabolic Pathways in Thermoanaerobacter sp. X514 via Isotopomer-Assisted Metabolite Analysis. Appl Environ Microbiol 75: 5001-5008.

99. Howell DM, Xu H, White RH (1999) (R)-Citramalate Synthase in Methanogenic Archaea J Bacteriol 181: 331-333.

100. Atsumi S, Liao JC (2008) Directed Evolution of Methanococcus jannaschii Citramalate Synthase for Biosynthesis of 1-Propanol and 1-Butanol by Escherichia coli. Appl Environ Microbiol 74: 7802-7808.

101. Kwok R (2010) Five hard truths for synthetic biology. Nature 463: 288-290.

102. Stelling J, Sauer U, Szallasi Z, Doyle FJ, Doyle J (2004) Robustness of cellular functions. Cell 118: 675-685.

103. Deutscher D, Meilijson I, Kupiec M, Ruppin E (2006) Multiple knockout analysis of genetic robustness in the yeast metabolic network. Nat Genet 38: 993-998.

104. Kol S, Merlo ME, Scheltema RA, Vries Md, Vonk RJ, et al. (2010) Metabolomic characterization of the salt stress response in Streptomyces coelicolor. Appl Environ Microbiol 76: 2574-2581.

105. Tang YJ, Martin HG, Deutschbauer A, Feng X, Huang R, et al. (2009) Invariability of central metabolic flux distribution in *Shewanella oneidensis* MR-1 under environmental or genetic perturbations Biotechnol Prog 25: 1254-1259

106. Heinemann M, Sauer U (2010) Systems biology of microbial metabolism. Curr Opin Microbiol 13: 1-7.

107. Niederberger P, Prasad R, Miozzari G, Kacser H (1992) A strategy for increasing an in vivo flux by genetic manipulations. The tryptophan system of yeast. Biochem J 287: 473-479.

108. Krömer JO, Heinzle E, Schröder H, Wittmann C (2006) Accumulation of homolanthionine and activation of a novel pathway for isoleucine biosynthesis in Corynebacterium glutamicum McbR deletion strains. J Bacteriol 188: 609-618.

109. Colletti P, Goyal Y, Varman A, Feng X, Wu B, et al. (2011) Evaluating factors that influence microbial synthesis yields by regression with numerical and categorical variables. Biotechnology and Bioengineering 108: 893-901.

110. Yoshikuni Y, Ferrin TE, Keasling JD (2006) Designed divergent evolution of enzyme function. Nature 440: 1078-1082.

111. Eshkol N, Sendovski M, Bahalul M, Katz-Ezov T, Kashi Y, et al. (2009) Production of 2-phenylethanol from L-phenylalanine by a stress tolerant Saccharomyces cerevisiae strain. J Appl Microbiol 106: 534-542.

112. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, et al. (2009) Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27: 753-759.

113. Carlson R, Srienc F (2006) Effects of recombinant precursor pathway variations on poly[(R)-3-hydroxybutyrate] synthesis in Saccharomyces cerevisiae. J Biotechnol 124: 561-573.

114. Overkamp KM, Bakker BM, Kötter P, Luttik MAH, Dijken JPv, et al. (2002) Metabolic Engineering of Glycerol Production in Saccharomyces cerevisiae. Appl Environ Microbiol 68: 2814-2821.

115. Shiba Y, Paradise EM, Kirby J, Ro D-K, Keasling JD (2007) Engineering of the pyruvate dehydrogenase bypass in Saccharomyces cerevisiae for high-level production of isoprenoids. Metab Eng 9: 160-168.

116. Maris AJAv, Geertman J-MA, Vermeulen A, Groothuizen MK, Winkler AA, et al. (2004) Directed evolution of pyruvate decarboxylase-negative Saccharomyces cerevisiae, yielding a C2-independent, glucose-tolerant, and pyruvate-hyperproducing yeast. Appl Environ Microbiol 70: 159-166.

117. Tokuhiro K, Muramatsu M, Ohto C, Kawaguchi T, Obata S, et al. (2009) Overproduction of geranylgeraniol by metabolically engineered Saccharomyces cerevisiae. Appl Environ Microbiol 75: 5536-5543.

118. Szczebara FM, Chandelier C, Villeret C, Masurel A, Bourot S, et al. (2003) Total biosynthesis of hydrocortisone from a simple carbon source in yeast. Nat Biotechnol 21: 143-149.

119. Asadollahi MA, Maury J, Schalk M, Clark A, Nielsen J (2010) Enhancement of farnesyl diphosphate pool as direct precursor of sesquiterpenes through metabolic engineering of the mevalonate pathway in Saccharomyces cerevisiae. Biotechnol Bioeng 106: 86-96.

120. Verwaal R, Wang J, Meijnen J-P, Visser H, Sandmann G, et al. (2007) High-level production of beta-carotene in Saccharomyces cerevisiae by successive transformation with carotenogenic genes from Xanthophyllomyces dendrorhous. Appl Environ Microbiol 73: 4342-4350.

121. Vannellia T, Qib WW, Sweigardc J, Gatenbyd AA, Sariaslani FS (2007) Production of p-hydroxycinnamic acid from glucose in Saccharomyces cerevisiae and Escherichia coli by expression of heterologous genes from plants and fungi. Metab Eng 9: 142-151.

122. Jiang H, Wood KV, Morgan JA (2005) Metabolic engineering of the phenylpropanoid pathway in Saccharomyces cerevisiae. Appl Environ Microbiol 71: 2962-2969.

123. Toivari MH, Ruohonen L, Miasnikov AN, Richard P, Penttilä M (2007) Metabolic Engineering of Saccharomyces cerevisiae for Conversion of d-Glucose to Xylitol and Other Five-Carbon Sugars and Sugar Alcohols. Appl Environ Microbiol 73: 5471-5476.

124. Alper H, Moxley J, Nevoigt E, Fink GR, Stephanopoulos G (2006) Engineering yeast transcription machinery for improved ethanol tolerance and production. Science 314: 1565-1568.

125. Sauer M, Branduardi P, Valli M, Porro D (2004) Production of L-Ascorbic Acid by Metabolically Engineered Saccharomyces cerevisiae and Zygosaccharomyces bailii. Appl Environ Microbiol 70: 6086-6091.

126. Iwatani S, Yamada Y, Usuda Y (2008) Metabolic flux analysis in biotechnology processes. Biotechnol Lett 30: 791-799.

127. Shirai T, Fujimura K, Furusawa C, Nagahisa K, Shioya S, et al. (2007) Study on roles of anaplerotic pathways in glutamate overproduction of Corynebacterium glutamicum by metabolic flux analysis. Microb Cell Fact 6: 19.

128. McKinlay JB, Shachar-Hill Y, Zeikus JG, Vieille C (2007) Determining *Actinobacillus succinogenes* metabolic pathways and fluxes by NMR and GC-MS analyses of ^{13}C-labeled metabolic product isotopomers. Metab Eng 9: 177-192.

129. McKinlay JB, Vieille C (2008) ^{13}C-metabolic flux analysis of *Actinobacillus succinogenes* fermentative metabolism at different $NaHCO_3$ and H_2 concentrations. Metab Eng 10: 55-68.

130. Lee JY, Jang Y-S, Lee J, Papoutsakis ET, Lee SY (2009) Metabolic engineering of Clostridium acetobutylicum M5 for highly selective butanol production. Biotechnol J 4: 1432-1440.

131. Kleijn RJ, Winden WAv, Ras C, Gulik WMv, Schipper D, et al. (2006) 13C-labeled gluconate tracing as a direct and accurate method for determining the pentose phosphate pathway split ratio in Penicillium chrysogenum. Appl Environ Microbiol 72: 4743-4754.

132. Navarro E, Montagud A, Córdoba PFd, Urchueguía JF (2009) Metabolic flux analysis of the hydrogen production potential in Synechocystis sp. PCC6803. Int J Hydrogen Energy 34: 8828-8838.

133. Shastri AA, Morgan JA (2005) Flux balance analysis of photoautotrophic metabolism. Biotechnol Prog 21: 1617-1626.

134. Boyle NR, Morgan JA (2009) Flux balance analysis of primary metabolism in Chlamydomonas reinhardtii. BMC Syst Biol 3: 4.

135. Tsantili IC, Karim MN, Klapa MI (2007) Quantifying the metabolic capabilities of engineered Zymomonas mobilis using linear programming analysis. Microb Cell Fact 6: 8.

136. DeGraaf A, Striegel K, Wittig RM, Laufer B, Schmitz G, et al. (1999) Metabolic state of Zymomonas mobilis in glucose-, fructose-, and xylose-fed continuous cultures as analysed by 13C- and 31P-NMR spectroscopy. Arch Microbiol 171: 371-385.

137. Dong QL, Zhao XM, Ma HW, Xing XY, Sun NX (2006) Metabolic flux analysis of the two astaxanthin-producing microorganisms *Haematococcus pluvialis* and *Phaffia rhodozyma* in the pure and mixed cultures. Biotechnol J 1: 1283-1292.

Figure 2.1. An iterative approach of fluxomic analysis and rational metabolic engineering

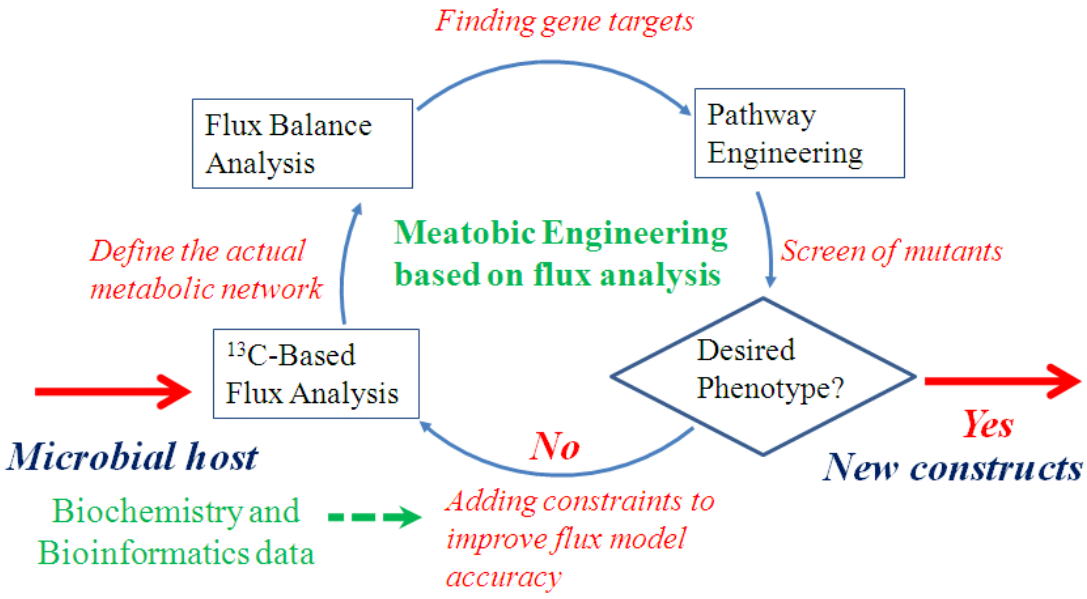


Figure 2.2. ^{13}C-assisted cellular metabolism analysis

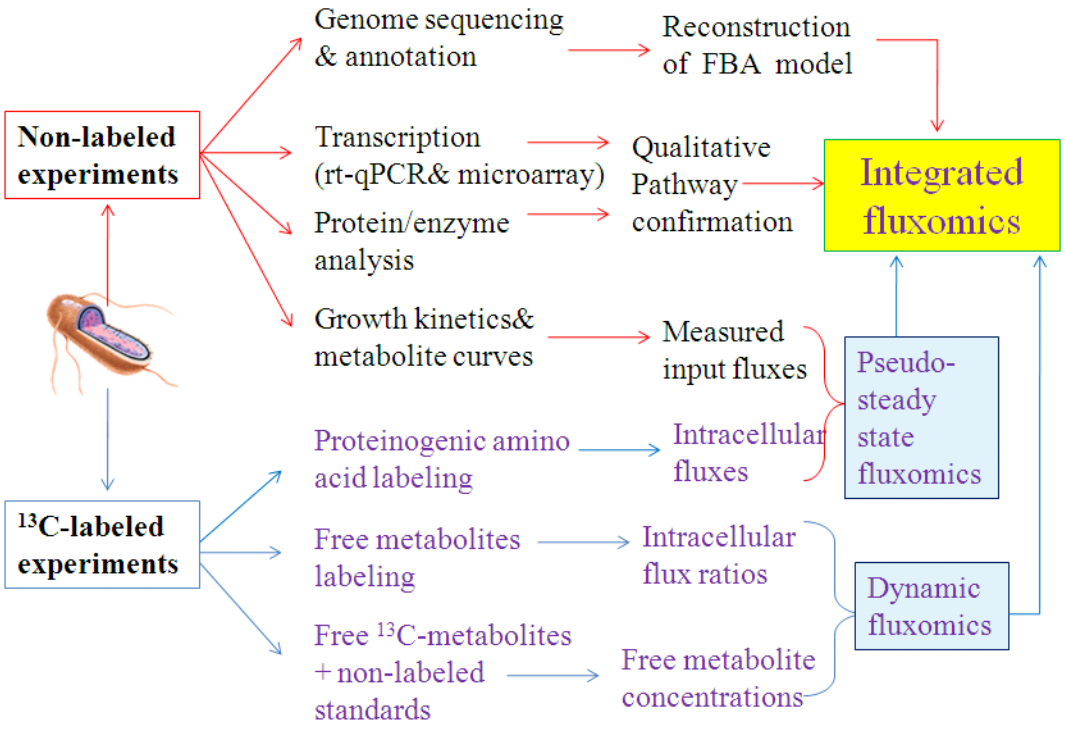


Figure 2.3. Product yields as a function of enzymatic steps from central metabolism. The solid line is the regression of published product yields by *S. cerevisiae* as a function of reaction steps from intermediate metabolites in central metabolism (including glycolysis, TCA cycle and pentose phosphate pathways). The yield declines exponentially as the number of reaction steps increases. The dotted lines are boundary curves with yield efficiencies of 30% and 70% respectively. All yield data from initial carbon sources are estimated from recent papers using our best judgment. The synthesized products and reaction steps are: Poly(R-3-hydroxybutyrate) (113) (steps=3); Glycerol (114) (steps=2); Artemisinic acid (2) (steps=10); Amorphadiene (115) (steps=9); Pyruvate (116) (steps=0); Geranylgeraniol (117) (steps=10); Hydrocortisone (118) (steps=19); Squalene (119) (steps=9); β-carotene (120) (steps=12); Lycopene (120) (steps=11); Phytoene (120) (steps=10); *p*-hydroxycinnamic acid (121) (steps=12); Naringenin (122) (steps=14); Pinocembrin (122) (steps=14); Xylitol and Ribitol (123) (steps=3); Ethanol (124) (steps=2); L-ascorbic acid (125) (steps=8).

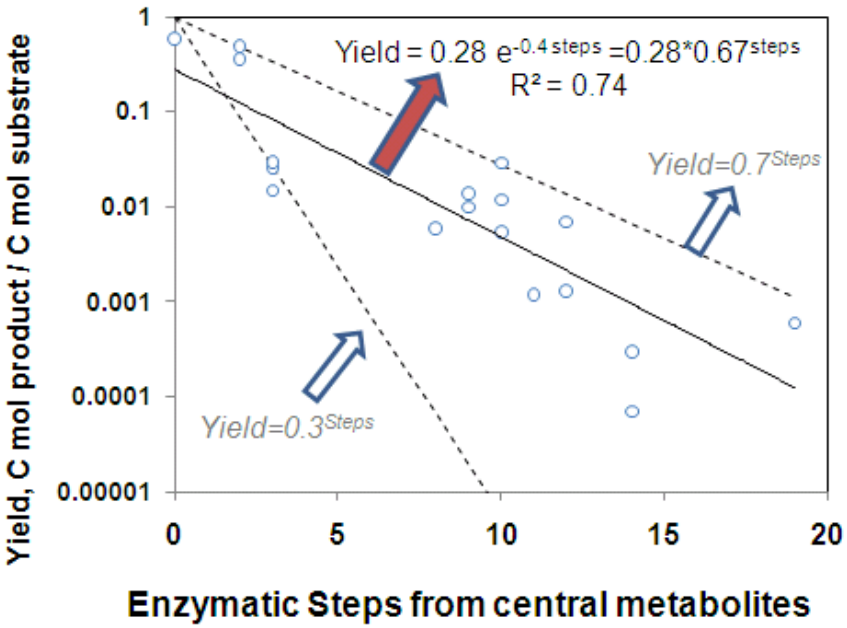

Yield = 0.28 e-0.4 steps =0.28*0.67steps
R² = 0.74
Yield=0.7Steps
Yield=0.3Steps
Yield, C mol product / C mol substrate
1
0.1
0.01
0.001
0.0001
0.00001
0
5
10
15
20
Enzymatic Steps from central metabolites

Table 2.1. Recent application of fluxomics of non-model microbes to bio-product synthesis

Species	Product	Substrate	Model description	Results from study	Reference
Corynebacterium glutamicum	Lysine	Glucose (sucrose, fructose)	^{13}C-MFA	MFA models (combining transcriptome, metabolome analysis) have been developed to study fluxes under different cultivation modes (mini-bioreactor, batch, fed-batch) using various carbon sources.	(126)
Corynebacterium glutamicum	Methionine	Glucose	^{13}C-MFA only focuses on flux distribution in the methionine pathway.	The *C. glutamicum* mutant (*mcbR)* showed no overproduction of methionine, but accumulation of homolanthionine.	(108)
Corynebacterium glutamicum	Glutamate	Glucose	^{13}C-MFA (focus on anaplerotic pathways)	The flux from phospho*enol*pyruvate to oxaloacetate catalyzed by phospho*enol*pyruvate carboxylase (PEPc) was active in the growth phase, whereas pyruvate carboxylase was inactive.	(127)
Actinobacillus succinogenes	Succinate formate and acetate	Glucose $NaHCO_3$	^{13}C-MFA (via NMR and GC-MS) and enzyme assay	The model indicated: 1. NADPH was produced primarily by transhydrogenase and/or by NADP-dependent malic enzyme; 2. oxaloacetate and malate were converted to pyruvate; 3. the effects of $NaHCO_3$ and H_2 on metabolic fluxes were quantified.	(128,129)

Geobacillus thermoglucosidasius	Ethanol	Glucose	FBA and ^{13}C-MFA	The model characterized the ethanol production under three oxygen conditions. The FBA analysis pointed out several gene targets for improving ethanol production.	(19)
Clostridium acetobutylicum	Butanol	Glucose	Genome scale-FBA	The engineered strain was able to produce 154 mM butanol with 9.9 mM acetone at pH 5.5, resulting in a butanol selectivity (a molar ratio of butanol to total solvents) of 0.84.	(130)
Penicillium chrysogenum	Penicillin	Gluconate /glucose	^{13}C -MFA (focus on pentose phose phase pathway and glycolysis)	The model determined the pentose phosphate pathway split ratio and estimated NADPH metabolism.	(131)
Synechocystis sp. PCC6803	Hydrogen	CO_2	FBA	The results included H_2 photoproduction, strategies to avoid oxygen inhibition, and analysis of hetero-, auto-, and mixotrophic metabolisms.	(132,133)
Synechocystis sp. PCC6803	Light energy & Biomass	Glucose/CO_2	^{13}C-MFA and dynamic ^{13}C -MFA	The model analyzed heterotrophic, autotrophic and mixotrophic metabolisms.	(33,57)
Chlamydomonas reinhardtii	Light energy & Biomass	CO_2	FBA model including three metabolically active compartments	The model indicated that heterotrophic growth had a low biomass yield on carbon, while mixotrophical and autotrophical growth had higher carbon utilization efficiency.	(134)
Zymomonas mobilis	Ethanol	Glucose/	FBA with	Model analyzed the	(135)

		xylose	various biological objectives	metabolic boundaries of *Z. mobilis*. The study indicated that ethanol and biomass production depend on anaerobic respiration stoichiometry and activity.	
Zymomonas mobilis	Ethanol	Glucose /frucose/ xylose	^{13}C–MFA via 1H-NMR 31P-NMR spectroscopy	The model characterized the intracellular metabolic state during growth on glucose, fructose and xylose in defined continuous cultures.	(136)
Coculture (*Desulfovibrio vulgaris* and *Methanococcus maripaludis)*	CH_4	Lactate	FBA analysis of microbial consortia	The model predicted the ratio of *D. vulgaris* to *M. maripaludis* cells during growth. It was possible to eliminate formate as an interspecies electron shuttle, but H_2 transfer was essential for syntrophic growth.	(54)
Community (oxygenic phototrophs, filamentous anoxygenic phototrophs, and sulfate-reducing bacteria).	Biomass and nitrogen fixation	CO_2	FBA and elementary mode analysis	The model predicted and described relative abundances of species, by-products, and the metabolic interactions.	(53)
Phaffia rhodozyma and *Haematococcus pluvialis*	Astaxanthin	Glucose with (peptone & yeast extract)	FBA analysis of mix culture	The two major astaxanthin-producing microorganisms exhibited elevated yields (2.8-fold) under mixed culture conditions compared to pure culture.	(137)

Chapter 3

^{13}C-Pathway Analysis of Environmental Microorganisms

3.1 ^{13}C-pathway analysis protocol

Novel metabolic features of environmental microorganisms can be revealed by ^{13}C-pathway analysis. As a complementary method to transcriptomics and proteomics studies, ^{13}C-pathway analysis can be used to uncover novel enzymes, investigate mixotrophic metabolism, and optimize the medium for slow-growing microbes. A protocol for ^{13}C-pathway analysis of environmental microorganisms is attached in Appendix 1. In general, ^{13}C-pathway analysis has three major steps.

The first step is to grow microbes on ^{13}C-labeled carbon substrates. Three key factors affect the ^{13}C-pathway analysis: medium composition, ^{13}C-labled carbon substrate, and culture modes. To avoid measurement noises from non-labeled carbon in nutrient supplements, a minimal medium with only one ^{13}C-labeled carbon source is required in tracer experiments. To collect the "fingerprinting" labeling patterns of metabolites synthesized from different pathways, singly labeled carbon substrates (>98% pure) are normally recommended in ^{13}C-pathway analysis. To guarantee both metabolite and isotopomer steady state are achieved in ^{13}C-pathway analysis, microbes can either

be cultured under chemostat mode, using typical bioreactors with continuous feeding, or batch mode, using shake flasks or small-scale mini-batch reactors.

The second step is to analyze amino acid labeling patterns using GC-MS. Amino acids are abundant in protein and thus can be obtained from biomass hydrolysis. To make the amino acids volatile enough for GC separation, amino acids are derivatized by N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide (TBDMS) prior to analysis. The TBDMS-derivatized amino acids can be fragmented by MS and result in different arrays of fragments, including four characteristic arrays of fragments that can reflect the ^{13}C-labeling patterns of amino acids: fragment $(M-15)^+$, which contains the entire amino acid but has lost a 15 MW methyl group; fragment $(M-57)^+$, which contains the entire amino acid but has lost a 57 MW tert-butyl group; fragment $(M-159)^+$, which has lost a 159 MW group that contains the 1^{st} carbon (α carboxyl group) of the amino acid; and fragment $(f302)^+$, which consists of a 302 MW group containing the amino acid without its R group. The final isotopic labeling patterns of proteinogenic amino acids are reported as mass fractions, such as M0, M1, M2, M3 and M4, to represent the percentage of fragments containing zero to four ^{13}C labeled carbons.

The third step is to trace ^{13}C carbon transition in the proposed pathways and to verify whether the labeling patterns in key metabolites are consistent with the annotated enzymatic reactions. Since errors or gaps within annotated pathways are common for poorly characterized microorganisms, the pathway

activity uncovered from ^{13}C-pathway analysis can be useful in functional characterization of metabolic pathway activities. Derived from the isotopic analysis of amino acids, labeling information about eight metabolites in the central carbon metabolism is provided and is used to reflect the functions of associated metabolic pathways.

^{13}C-pathway analysis is a powerful tool to probe active carbon metabolic pathways *in vivo*. Table 3.1 illustrates several carbon metabolic pathways that can be revealed by ^{13}C-labeled proteinogenic amino acids. In cooperation with *in vivo* physiological studies, transcription assays, and *in vitro* biochemical characterizations, new insights to carbon metabolisms of organisms can be illustrated.

Table 3.1 Fingerprinting ^{13}C-labeling patterns of amino acids for identifying novel metabolic pathways and enzymes.

Pathways and enzymes	**^{13}C-carbon substrate**	**Key amino acids**	**"Fingerprinting" labeling patterns**
The Entner-Duodroff (ED) pathway	[1-^{13}C] glucose, or [6-^{13}C] glucose	Ser & Ala	Notably lower (with [1-^{13}C] glucose) or higher (with [6-^{13}C] glucose) ^{13}C abundance in Ser
The reductive TCA cycle	Non-labeled CO_2 with [3-^{13}C] pyruvate	Ala, Asp, & Glu	Notably lower ^{13}C abundance in Asp and Glu than in Ala
The branched TCA cycle	[3-^{13}C] pyruvate, or [2-^{13}C] glycerol	Asp & Glu	Different labeling patterns in Asp versus in Glu
(*Re*)-citrate synthase in TCA cycle	[1-^{13}C] pyruvate	Glu	Non-labeled α-carboxyl group of Glu
The citramalate-pathway in isoleucine biosynthesis	[2-^{13}C] pyruvate or [1-^{13}C] acetate	Leu & Ile	Identical labeling patterns in Leu and Ile
The Calvin-Benson cycle	Non-labeled CO_2 with labeled organic carbon	His & Ser	Significantly low ^{13}C abundance in Ser and His
The oxidative pentose phosphate pathway	[1-^{13}C] glucose	Ala	non-labeled Ala >50%
The CO_2-anaplerotic pathways	^{13}C-bicarbonate with non-labeled organic carbon	Ala & Asp	Enriched ^{13}C abundance in Asp

3.2 Mixotrophic and photoheterotrophic metabolism in *Cyanothece* sp. ATCC 51142 under continuous light

Xueyang Feng[1], Anindita Bandyopadhyay[2], Bert Berla[1], Lawrence Page[2], Bing Wu[1], Himadri B. Pakrasi[1,2], and Yinjie J. Tang[1,*]

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

[2]Department of Biology, Washington University, St. Louis, MO 63130, USA

Corresponding author:

*: yinjie.tang@seas.wustl.edu; phone: 314-935-3441

This section was previously published in

Feng X, Banerjee A, Berla B, Page L, Wu B, Pakrasi H.B, and Tang YJ. Mixotrophic and photoheterotrophic metabolisms in Cyanothece sp. ATCC 51142 under continuous light." Microbiology. 2010, 156(Pt 8):2566-74

3.2.1 Abstract

The unicellular diazotrophic cyanobacterium, *Cyanothece* sp. ATCC 51142 (*Cyanothece* 51142), is able to grow aerobically in nitrogen-fixing conditions under alternating light-dark cycles or continuous illumination. This study investigated the impacts of carbon and nitrogen sources on *Cyanothece* 51142 metabolism via ^{13}C-assisted metabolite analysis and biochemical measurements. Under continuous light (50 μmol photons/m^2/s) and nitrogen-fixing conditions, we find glycerol addition promoted aerobic biomass growth (by twofold) and nitrogenase-dependent hydrogen production (up to 25 μmol H_2/mg chlorophyll/hr), but strongly reduced phototrophic CO_2 utilization. Under nitrogen-sufficient conditions, *Cyanothece* 51142 was able to metabolize glycerol photoheterotrophically, and the activity of light dependent reactions (e.g., oxygen evolution) was not significantly reduced. In contrast, *Synechocystis* sp. PCC 6803 showed apparent mixotrophic metabolism under similar growth conditions. Isotopomer analysis also detected that *Cyanothece* 51142 was able to fix CO_2 via anaplerotic pathways, and to uptake glucose and pyruvate for mixotrophic biomass synthesis.

Key words: ^{13}C-assisted, anaplerotic pathway, CO_2 utilization, glycerol, hydrogen, nitrogenase

3.2.2 Introduction

Rising concerns about global warming due to the greenhouse effect have renewed research focused on biological capture of CO_2. Cyanobacteria have versatile metabolic capabilities, which allow them to grow under autotrophic, heterotrophic, and mixotrophic conditions (1,2,3). More importantly, some cyanobacteria can capture solar energy to fix nitrogen and generate H_2, thereby serving as a source of biofertilizer and biofuel, while simultaneously consuming atmospheric CO_2 (4,5,6,7,8,9). *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium, is able to grow aerobically under nitrogen-fixing conditions and has been recognized for contributing to the marine nitrogen cycle. The recent sequencing of the *Cyanothece* 51142 genome and its transcriptional analysis uncovered the bacterium's diurnally oscillatory metabolism in alternating light-dark cycles (photosynthesis during the day and nitrogen fixation at night) (10,11,12). In general, cyanobacteria use spatial or temporal separation of the oxygen-sensitive nitrogen-fixation and the oxygen-evolving photosynthesis as a strategy for diazotrophic growth (9,13). Interestingly, *Cyanothece* 51142 demonstrates simultaneous N_2 fixation and O_2 evolution under continuous light conditions, though it appears to be unicellular (14,15). For example, a recent study on transcriptional and translational regulation of continuously-illuminated *Cyanothece* has revealed strong synthesis capability for nitrogenase and circadian expression of 10% of its genes (10,11). Furthermore, *Cyanothece* strains usually

utilize exogenous carbon substrates for mixotrophic growth under light and for heterotrophic growth under dark conditions (16). Carbon substrates are key factors controlling the efficiency of cyanobacterial aerobic growth and hydrogen production (7,16,17,18,19). Genome analysis studies have revealed that *Cyanothece* 51142 has a unique gene cluster on its linear chromosome containing key genes involved in glucose and pyruvate metabolism (12). However, the ability of this strain to metabolize glucose or pyruvate remains unknown.

To quantitatively examine the effect of carbon and nitrogen sources on *Cyanothece* central metabolism, this study investigated the impact of three carbon sources (glucose, glycerol, and pyruvate as representatives of sugar, lipid derivatives, and organic acids from central metabolic pathways, respectively) on *Cyanothece* 51142 growth and metabolism. Two nitrogen sources other than N_2, ammonia and nitrate, were also examined. Precise readouts on metabolic state and activity were based on ^{13}C-assisted metabolite analysis integrated with biochemical assays and the gene expression patterns obtained by reverse transcription PCR (RT-PCR) (20,21,22,23,24,25). Superior to the traditional ^{14}C method (1), the non-radioactive ^{13}C method can provide rich information about which carbons within a metabolite are labeled, and thus enable an in-depth understanding of carbon utilization and metabolic regulation in *Cyanothece* 51142.

3.2.3 Materials and methods

Bacterial strains and growth conditions. *Cyanothece* 51142 were first grown in 150 mL Erlenmeyer flasks fed with ASP2 medium (16) without nitrate. Ambient carbon dioxide provided the sole carbon source. For experiments examining the effect of nitrogen sources, 18 mM $NaNO_3$ or 17 mM NH_4Cl was added into the medium. Cultures were grown aerobically under continuous light (50 µmol photons·m^{-2}·s^{-1}) on a shaker at 150 rpm and 30°C. Cells at late mid-log phase were sub-cultured into different cultural media with various nitrogen and carbon sources. Isotopically-labeled carbon substrates (Cambridge Isotope Laboratories, Andover, MA) were used for mixotrophic growth, including 54 mM glycerol (2-^{13}C, >98%), 26 mM glucose (U-^{13}C, >98%) and 11 mM sodium pyruvate (3-^{13}C, >98%). For tracer experiments, a 3% inoculum from unlabeled stock culture was used to inoculate a 50 mL medium containing labeled carbon sources. At the mid-log phase of growth, a 3% inoculum from the first isotopic labeled culture was used to inoculate 50 mL sub-cultures with the same medium to remove the effect of unlabeled carbon introduced from the initial inoculum. Cell growth was monitored by a UV-Vis spectrometer (Genesys, Thermo Scientific, USA) at 730 nm. To perform a comparative study, a glucose tolerant *Synechocystis* strain PCC 6803 (a model cyanobacterium for studying fundamental processes of photosynthetic metabolism) was also cultured in BG11 medium (pH=7.6) under the same growth conditions (continuous light and 30 ^{o}C, (26). The BG11 medium was supplemented with 6 mM glucose (U-^{13}C, >98%) to support mixotrophic growth. *Synechocystis* PCC 6803 were also sub-cultured in

the same labeled medium twice before sampling for ^{13}C-labeled metabolite analysis.

Metabolite and photosynthetic activity analysis. To analyze metabolites in *Cyanothece* 51142, biomass was harvested at the middle-log phase of growth (~90 hours) by centrifugation at 7,000 rpm for 15 min at 10˚C. The concentrations of pyruvate, glucose and glycerol were analyzed with enzymatic assay kits (R-Biopharm, Germany). To measure hydrogen produced by *Cyanothece* 51142, 20 ml of culture solution was taken from the culturing flask after three days and transferred into 35.2 ml glass vials sealed with rubber septa and kept under continuous light (50 μmol photons·$m^{-2}s^{-1}$). A modified protocol was used to quantify hydrogen (27). Briefly, hydrogen that accumulated in the headspace of the sealed culture vials (for 5 hours) was withdrawn with a Hamilton gas-tight syringe and quantified on an Agilent 6890N Gas Chromatograph with a molseive 5A 60/80 column (inner dimensions 6'×1/8") and Thermal Conductivity Detector. Injection, oven, and detector temperatures were 100°C, 50°C, and 100°C, respectively. Argon was the carrier gas (flow rate of 65 ml·min^{-1}). All measurements included three biological replicates.

Photosynthesis activities were determined based on measurements of chlorophyll fluorescence and oxygen evolution. Chlorophyll fluorescence profiles of photosystem II (PSII) of *Cyanothece* 51142 under different nutrient conditions were detected by a FL100 fluorometer (Photon Systems Instruments, Brno, Czech

Republic) as detailed before (28). All samples taken for measurement were diluted to OD_{730} ~0.2 using cell-free ASP2 medium. The samples were first adapted for 3 min in total darkness. During the measurement (performed at room temperature), the fluorometer emitted saturating light pulses to determine samples' fluorescence yield. The photosynthesis activity was derived by the maximum quantum yield (F_v/F_m) according to the formula $F_v/F_m = (F_m - F_0)/F_m$, where F_0 is initial fluorescence and F_m is maximum fluorescence at the beginning of measurement (29).

Oxygen evolution rates of *Cyanothece* 51142 grown under different nutrient conditions were measured with a Hansatech oxygen electrode. Assays were performed at 30 °C on whole cells in ASP2 media with a saturating light intensity of 8,250 μmol photons·$m^{-2}s^{-1}$ for 2 mins in a 2.5-mL reaction volume. For each reaction, the chlorophyll concentration of each sample was diluted to ~6 μg·ml^{-1}. The oxygen evolution rates (μmol O_2•mg $chlorophyll^{-1}$•hr^{-1}) were then measured and normalized based on chlorophyll concentration.

RNA extraction and reverse transcription PCR (RT-PCR). The bacteria grown under different cultural conditions were harvested at mid-log phase according to the corresponding growth curves. The total RNA was extracted by using a PureLink™ RNA Mini Kit (Invitrogen, California), following the manufacturer's instruction. cDNA was synthesized from ~2 μg RNA by using a High-Capacity cDNA Reverse Transcription Kit (Invitrogen,

California). The primers for RT-PCR reactions were designed using Primer Premier 5 software (PREMIER Biosoft) and analyzed by OligoAnalyzer 3.0 software (Integrated DNA Technologies, Coralville, USA). The forward primer (AGCGGTGGAGTATGTGGT) and reverse primer (GGCTGGGTTTGATGAGATT) were employed to amplify a 16S rRNA gene as a control. The forward primer (CCGACTACACTCCGAAAG) and reverse primer (ACGTAACGCCCGTAATGC) were used to amplify the Rubisco (*rbcL*) gene and the forward primer (TAATCACGAAACGGGAG) and reverse primer (CACCACATCAGCGTATTG) to amplify the *prk* gene. The PCRs were conducted with the following cycle conditions: 2 min of activation of the polymerase at 94 °C followed by 30 cycles consisting of 1 min at 94 °C, 30 s at 53 °C and 2 min at 72 °C; finally, a 10 min extension process was performed at 72 °C. The final PCR product was observed directly on 2% agarose gels after electrophoresis.

Isotopic analysis. The preparation and isotopic analysis of proteogenic amino acids were performed as previously described (30,31). In brief, exponentially growing biomass from ~20 ml culture was collected by centrifugation (8,000×*g*, 10 min, 4˚C) and hydrolyzed in 6 M HCl at 100°C for 24 h. The amino acid mix was dried and derivatized in tetrahydrofuran (THF) and N-(tert-butyl dimethylsilyl)-N-methyl-trifluoroacetamide (Sigma-Aldrich, St. Louis, MO) at 70°C for 1 h. A gas chromatograph (GC) (Hewlett-Packard, model

7890A, Agilent Technologies Inc., Ballwin, MO) equipped with a DB5-MS column (J&W Scientific, Folsom, CA) and a mass spectrometer (MS) (5975C, Agilent Technologies Inc., Ballwin, MO) were used for analyzing amino acid labeling profiles. The ion $[M-57]^+$ from unfragmented amino acid was detected and mass fractions of key amino acids were calculated (32). The substrate utilization ratios R (reflecting the degree of mixotrophic metabolism) were calculated from the labeling patterns of proteogenic amino acids:

$$\text{Amino acid X: } \frac{0.98\times n\times V_{sub}+0.01\times V_{CO2}}{m\times V_{sub}+V_{CO2}}=\frac{(\sum_{i=1}^{C} i\times M_i)}{C}\Longrightarrow R=\frac{V_{sub}}{V_{CO2}} \quad (1)$$

where the ratio R indicated the utilization of labeled carbon substrate over unlabeled CO_2 for producing an amino acid X (and its precursors). M_i was the GC-MS isotopomer fraction for the given amino acid X (i.e., M_0 was the unlabeled fraction, M_1 was the singly labeled fraction, M_2 was the doubly labeled fraction, M_3 was the triply labeled fraction, etc). C was the total number of carbon atoms in the amino acid molecule. V_{sub} was the carbon flux from ^{13}C labeled substrate, V_{CO2} was the carbon flux from CO_2; *0.98* was the purity of the labeled carbon substrate; *0.01* was the natural abundance of ^{13}C, m was the total number of carbons in the substrate molecule, and n was the total number of labeled carbons in the substrate molecule. The ratio R indicated the amount of labeled carbon that percolated through the central metabolic networks (Figure 3.2.1).

3.2.4 Results

Cell growth with different carbon and nitrogen sources. Figure 3.2.2 and Supplementary Figure 3.2.S1 shows the effect of carbon and nitrogen substrates on the growth of *Cyanothece* 51142 under continuous light. Biomass growth was significantly enhanced by addition of glycerol to ASP2 medium. For example, glycerol addition doubled the specific growth rate from 0.28 day^{-1} to 0.63 day^{-1} under N_2 fixing conditions. These results are consistent with an earlier report on two *Cyanothece* strains (16). On the other hand, *Cyanothece* growth was not apparently enhanced by either glucose or pyruvate (Supplementary Figure 3.2.S1), and high concentration pyruvate (64 mM) inhibited *Cyanothece* growth. Compared to nitrogen fixing cultures, the presence of nitrate salts in the growth media increased *Cyanothece* autotrophic growth rates from 0.28 day^{-1} (N_2 fixation condition) to 0.37 day^{-1} (nitrate-sufficient condition). Similarly, the presence of glycerol enhanced growth rate by two-fold (from 0.60 day^{-1} to 1.02 day^{-1}). As expected, high concentrations of ammonium salts (17 mM) fully inhibited growth (data not shown) because of its well-known deleterious effect on photosystems of cyanobacteria (33,34).

Isotopic analysis of amino acids. ^{13}C enrichment patterns in key metabolites were used to estimate the relative utilization of labeled carbon substrates (i.e., glucose, pyruvate, or glycerol) and CO_2 for metabolite synthesis under mixotrophic growth. Figure 3.2.1 shows the central metabolic pathways in

Cyanothece 51142 (http://www.genome.jp/kegg/). Five amino acids were analyzed for their labeling: histidine (precursor: ribose-5-phosphate and 5,10 methyl-THF), synthesized from the Calvin Cycle and pentose phosphate pathway; serine (precursor: 3-phosphoglycerate, a product from the Calvin Cycle); alanine (precursor: pyruvate, originated from carbon substrate or CO_2 fixation); and aspartate and glutamate (precursors: oxaloacetate and 2-oxoglutarate, respectively, synthesized from the citric acid cycle). Under nitrate-sufficient conditions, glycerol could be used as the sole carbon source for synthesis of alanine, serine, and histidine (as indicated by approximately infinite R values). This indicates that the cell was undergoing completely heterotrophic metabolism. R values of some key amino acids in glucose and pyruvate cultures were positive and thus the two carbon sources were actually utilized for biomass synthesis (Table 3.2.1). However, their measured R values were between 0 and 0.3, which indicated that CO_2 was the main carbon source for metabolite synthesis. This result was consistent with the fact that glucose and pyruvate did not apparently improve the biomass growth. Compared to nitrogen-sufficient conditions, nitrogen fixing conditions further limited glucose and glycerol utilization, as shown by the decreased labeling fractions of three key amino acids (i.e., alanine, serine, and histidine) (Table 3.2.1).

Nitrogenase-dependent hydrogen production, photosynthesis and Calvin Cycle activity. Hydrogen production under continuous light with different

carbon substrates (N_2 as the sole nitrogen source) was measured in the exponential (day 4) and stationary (day 9) growth phases, respectively (Supplementary Figure 3.2.S2). In the exponential growth phase under nitrogen fixing conditions, hydrogen production rates were as follows: for glycerol (25±6 μmol H_2/mg chlorophyll/hr); for glucose (13±9 μmol H_2/mg chlorophyll/hr); for pyruvate (4±2 μmol H_2/mg chlorophyll/hr); and in the photoautotrophic condition (5±1 μmol H_2/mg chlorophyll/hr). Under all nitrate or ammonium chloride conditions, hydrogen production was not detected regardless of the carbon substrate.

The measurement of photosynthetic parameters (Figure 3.2.3) suggested that, compared to photoautotrophic conditions, addition of an exogenous carbon source (glycerol, glucose, or pyruvate) did not strongly suppress the maximal quantum yield of PSII (F_v/F_m) or the oxygen evolution rate. Nitrate-sufficient conditions enhanced the oxygen evolution rates by 2~3 fold compared to nitrogen-fixing conditions, while the change of quantum yields of PSII were much less significant (10~30%). Gene expression in the carbon fixation pathway was also determined (Figure 3.2.4). Reverse transcription PCR (RT-PCR) results indicated that two key enzymes in Calvin Cycle (ribulose-1,5-bisphosphate carboxylase oxygenase (Rubisco, *rbcL*) and phosphoribulokinase (*prk*)) were functional under growth conditions with glycerol or glucose. The above measurements confirm that the light dependent reactions were active under all

culture conditions, even though carbon substrates reduced the relative amount of CO_2 fixation for biomass synthesis.

3.2.5 Discussion

Carbon substrate utilization and regulation. In continuous light, *Cyanothece* 51142 can efficiently utilize glycerol for aerobic growth. Based on the measurement of carbon substrates in the culture medium during the exponential growth phase, the uptake rates of glycerol were 0.22±0.05 g·(g dry biomass•day)$^{-1}$ under nitrogen fixing and 0.35±0.06 g·(g dry biomass•day)$^{-1}$ under nitrate-sufficient conditions. Glycerol promoted *Cyanothece* 51142 growth because it provided carbon and energy sources. Under nitrate-sufficient conditions, the unlimited large value of R from serine, alanine and histidine labeling data indicated that 3-phosphoglycerate node, pyruvate node and ribose-5-phosphate node in the central metabolic pathways (Figure 3.2.1) were completely originating from glycerol, while the contribution of CO_2 photofixation to those metabolite nodes was negligible. As a comparison, a glucose-tolerant strain of *Synechocystis* sp. strain 6803 was cultured with fully labeled glucose under continuous light and nitrogen-sufficient conditions (Supplementary Figure 3.2.3). The measured R values (Table 3.2.1) for serine (0.87), alanine (0.92) and histidine (1.73) indicated that *Synechocystis* 6803 had a typical mixotrophic growth. In general, cyanobacterial heterotrophic growth has been reported only under three conditions: complete darkness, dim light, and pulses of light (35,36). When the

light is sufficient for photoautotrophy, *Cyanothece* photoheterotrophic growth was only achieved by addition of photosystem II inhibitors (16). This study shows that rapidly-growing *Cyanothece* 51142 cells can shift their metabolic strategies from mixotrophic or autotrophic growth to photoheterotrophic growth, possibly because maximal utilization of energy-rich carbon substrate (glycerol) can reduce energy costs related to CO_2 fixation (fixation of one CO_2 consumes two ATP and one NADPH) and building block synthesis so that maximal biomass growth can be achieved.

On the other hand, glucose was not apparently consumed by *Cyanothece* 51142 (the consumed concentrations were below 1mM in all experiments). In the [U-^{13}C] glucose experiments (Table 3.2.1), all five amino acids contained labeled carbons, which indicated that the labeled glucose had percolated through the entire central metabolic pathways, thereby confirming the ability of *Cyanothece* 51142 to metabolize glucose. The R values of all key amino acids were below 0.05 for both nitrogen fixation and nitrate-sufficient conditions, suggesting that a large fraction of the carbon in the biomass had originated from CO_2 fixation. In contrast, glucose was the most favorable carbon source for *Synechocystis* species (3), and the R values (Table 3.2.1) from key amino acids were around (0.4~1.7). While both *Synechocystis* 6803 and *Cyanothece* 51142 have completely annotated central pathways for glucose metabolism, *Synechocystis* 6803 contains a glucose transporter (gene code Sll0771) that shares a sequence relationship with

mammalian glucose transporters (1,37,38). So far, the presence of a glucose transporter in *Cyanothece* 51142 has not been rigorously verified. From the genome database (DOE Joint Genome Institute, www.jgi.doe.gov/), a gene (cce_3842) was identified as a glucose transport protein that shared weak (25%) amino acid identity with the Sll0771 protein of *Synechocystis* PCC6803. Based on the glucose-dependent growth data, we conclude that the enzymes involved in glucose transport or utilization in *Cyanothece* 51142 may not be as efficient as those of *Synechocystis* PCC6803.

Analysis of labeled pyruvate-grown *Cyanothece* cells showed that serine (whose precursor is 3-phosphoglycerate) and histidine (whose precursor is ribose-5-phosphate) were completely unlabeled (R=0). Such labeling profiles suggest that CO_2 was used as the sole carbon source for synthesis of metabolites in glycolysis and the pentose phosphate pathway (i.e., there was no gluconeogenesis activity). Pyruvate was used only to synthesize alanine (R=0.3~0.6) and metabolites in the TCA Cycle: (pyruvate→oxaloacetate→Asp) (pyruvate→acetyl-CoA→citrate→2-oxoglutarate→Glu), as reflected by the labeled carbon present in glutamate and aspartic acid. Interestingly, the R values of alanine (=0.60) and glutamate (=1.25) were higher under nitrogen-fixing conditions compared to the nitrate-sufficient conditions, indicating that relatively more labeled pyruvate was used for glutamate synthesis under these conditions. The nitrogen fixation was via nitrogenase: $N_2 + 6\ H^+ + 6\ e^- \rightarrow 2\ NH_3$, and the nitrogenase-generated

ammonium was assimilated into amino acids through the glutamine synthetase/glutamate synthase pathway (39). Utilization of supplemented pyruvate for glutamate synthesis could facilitate the nitrogen fixation process.

The enzyme ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) is known to be the rate-limiting factor in the Calvin Cycle for capturing CO_2 to synthesize three-carbon sugar (glycerate 3-phosphate) (40,41,42). We examined Rubisco (*rbcL*) and phosphoribulokinase (*prk*) gene expression to reveal the metabolic regulation in the Calvin Cycle at the transcriptional level. Under photoautotrophic, mixotrophic, and heterotrophic growth conditions, the expression of the two genes encoding ribulose-1,5-bisphosphate carboxylase oxygenase (*rbcL*) and phosphoribulokinase (*prk*) were clearly observed. Although Calvin Cycle genes were expressed, *Cyanothece* 51142 still grew heterotrophically in the presence of glycerol and nitrate based on the isotopomer data (no apparent incorporation of CO_2 from Calvin Cycle). These inconsistencies indicate that ^{13}C-assisted metabolite analysis provides direct readout on actual metabolic status, while gene expression results cannot be solely relied upon as there are many possible post-transcriptional regulations involved.

Furthermore, *Cyanothece* 51142 can fix CO_2 via anaplerotic pathways (i.e., C4 carbon fixation) (43). In the presence of glycerol and nitrate-sufficient conditions (Table 3.2.1), R ratios for aspartate synthesis were 1.53, much smaller than the R ratios ($R=\infty$) of Ala, Ser, and His. This indicates the utilization of CO_2

for the synthesis of C4 metabolites in the TCA Cycle via anaploric pathways, even though phototrophic CO_2 fixation was significantly inhibited: (1) PEP+CO_2→oxaloacetate (catalyzed by phosphoenolpyruvate carboxylase or phosphoenolpyruvate carboxykinase) or (2) pyruvate+CO_2→malate (catalyzed by malic oxidoreductase). Such anaplerotic pathways synthesized key TCA Cycle metabolites like oxaloacetate and succinate (precursors for chlorophyll).

Meanwhile, CO_2 was generated by two reactions (i.e., pyruvate→acetylCoA+CO_2; isocitrate→2-oxoglutarate+CO_2), which are essential steps for glutamate synthesis. These catabolic processes cause the loss of unlabeled carbon when 2nd position labeled glycerol is used as the main carbon source. Therefore, the coefficients V_{CO2} (CO_2 utilization flux) and R (carbon utilization ratio) were both negative for glutamate synthesis (Equation 1) in glycerol supplemented cultures (both nitrogen fixation and nitrate-sufficient conditions) (Table 3.2.1).

Photosynthesis activity. Photosynthesis activity was estimated by the Fv/Fm parameter (maximum quantum efficiency of photosystem II) (44). When glycerol or glucose were utilized, the maximum quantum yield F_v/F_m (i.e, efficiency of photosystem II) in *Cyanothece* 51142 was not significantly affected (changes are within ~30%, Figure 3.2.3a). Although chlorophyll fluorescence estimation is not an accurate method for determination of absolute PSII activity (45,46), we have used it in our study as a tool only to confirm active photon

capture in the light-harvesting antenna complexes of photosystem II under both heterotrophic and mixotrophic conditions.

Oxygen evolution was measured as one molecule of the pigment chlorophyll absorbs one photon and uses its energy to generate NADPH, ATP, and O_2 in the light-dependent reactions (47). The oxygen evolution rates in *Cyanothece* 51142 rose by 2~3 fold under all nitrate-sufficient conditions compared to corresponding nitrogen fixation conditions (Figure 3.2.3b). The significantly higher rates of oxygen evolution indicated that the photosynthetic process of water splitting was more active and provided more energy (ATP and NADPH) to support biomass growth under nitrate-sufficient conditions.

Finally, precise determination of the photosynthetic activity is difficult for *Cyanothece* 51442, since the metabolic behavior of *Cyanothece* 51442 fluctuates under continuous light due to its circadian rhythm (10,15). The photoreaction activities in Figure 3.2.3 are only qualitative (not quantitative) evidences to support the presence of active light-dependent reactions under all culture conditions.

Nitrogen utilization and nitrogenase-dependent hydrogen production. Under anaerobic condition (using argon gas to flush the culture), hydrogen production rates of *Cyanothece* 51142 were as high as 100 μmol/mg chlorophyll/hr (data not shown). Under aerobic conditions, hydrogen production enzyme (hydrogenase) was completely inactivated by oxygen (7). *Cyanothece*

51142 used nitrogenase for both nitrogen fixation and hydrogen production. Nitrate, ammonium or some amino acids inhibit nitrogenase activity and thus fully prohibit aerobic hydrogen production by cyanobacteria (48). Furthermore, NH_4^+ is a direct nitrogen source (nitrate is reduced to NH_4^+) that can be incorporated into biomass via glutamine / glutamate synthase (49). *Cyanothece* 51142, however, only grows with low concentration NH_4^+ (below 1 mM) because of an observed inhibition effect (48,50). Nitrogen fixation is an energy demanding process ($N_2+8H^++\mathbf{8e^-}+\mathbf{16ATP}\rightarrow 2NH_3+H_2+16ADP+16Pi$). Addition of glycerol reduces CO_2 fixation via the Calvin Cycle, so more energy (ATP and NADH) can be directed to nitrogen fixation and thus improve hydrogen production by 4~5 fold (4,5). Glucose and pyruvate cannot significantly promote hydrogen production because their utilization is very low and their effect on energy economy is limited. Hydrogen production rates dropped for all mixotrophic cultures of *Cyanothece* 51142 after 9 days, suggesting that inhibitory metabolites accumulated during the cultivation which reduced nitrogenase activities (40,41,42). Finally, the coexistence of oxygen-evolving photosynthesis and oxygen-sensitive nitrogen fixation (indicated by hydrogen evolution) is an attractive characteristic in some cyanobacteria (13,14). Unlike filamentous cyanobacterial species where nitrogen fixation and oxygenic photosynthesis are spatially segregated (51), *Cyanothece* 51142 is able to maintain activities for N_2 fixation, respiration, and photosynthesis within the same cell under continuous light. The strain not only has a strong ability to scavenge intracellular oxygen and

synthesize nitrogenase (9,15), but also develops a highly circadian mechanism for nitrogen fixation (52).

This study improves our understanding of *Cyanothece* 51142 physiology under different carbon and nitrogen sources as well as its potential application for hydrogen production applications. In general, exogenous carbon substrates may improve cellular growth, but have strong negative effects on CO_2 fixation. Continuously illuminated *Cyanothece* 51142 shows simultaneous oxygen evolution and nitrogenase-dependent hydrogen production, while hydrogen production can be significantly enhanced by the addition of glycerol. A comparison of metabolic status under autotrophic, mixotrophic and heterotrophic growth conditions indicated that *Cyanothece* 51142 has an inherent metabolic strategy for maximal biomass production at low energy cost. Finally, this study has further confirmed that ^{13}C-assisted metabolite analysis is a high throughput method which can provide new and precise information to understand a biological system.

Figure 3.2.1. Central metabolic pathways of *Cyanothece* 51142 with glucose, glycerol, and pyruvate as carbon substrates. The dashed line shows the metabolic pathway with glycerol as carbon substrate; the bold line indicates glucose; the solid line shows the common pathway for all carbon conditions. Abbreviations: ACCOA, acetyl-coenzyme A; Ala, alanine; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; GAP, glyceraldehyde 3-phosphate; 3PG, 3-phosphoglycerate; GLY, glycerol; GLU, glucose; His, histidine; ICIT, citrate/isocitrate; MAL, malate; OAA, oxaloacetate; OXO, 2-oxoglutarate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose-5-phosphate (or ribulose-5-phosphate); R15P, ribulose-1,5-bisphosphate; S7P, sedoheptulose-7-phosphate; Ser, serine; Xu5P: xylulose-5-phosphate.

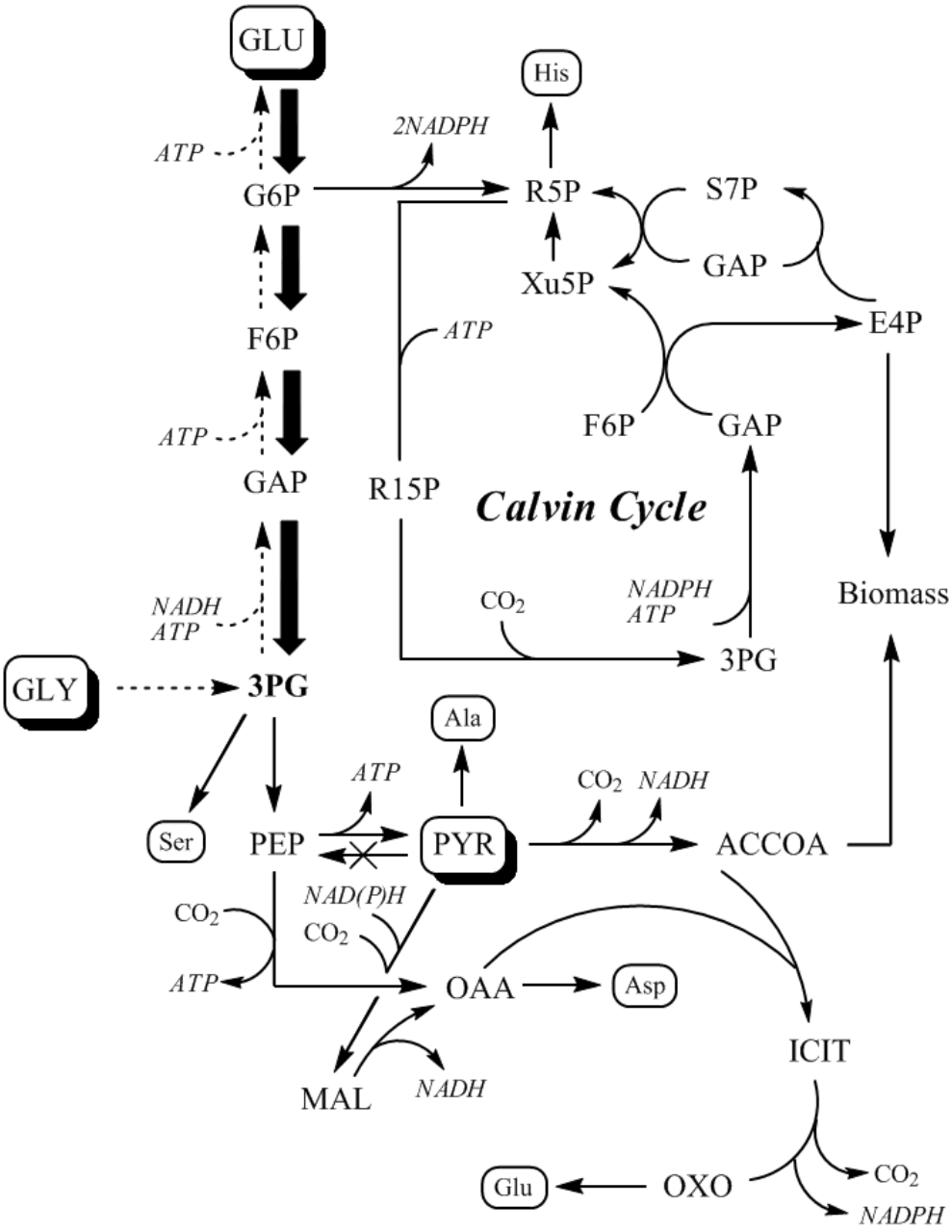

GLU
His
ATP
2NADPH
G6P
R5P
S7P
GAP
Xu5P
F6P
ATP
E4P
ATP
F6P
GAP
GAP
R15P
Calvin Cycle
NADH
ATP
CO2
NADPH
ATP
Biomass
GLY
3PG
3PG
Ala
ATP
CO2
NADH
Ser
PEP
PYR
ACCOA
NAD(P)H
CO2
CO2
ATP
OAA
Asp
ICIT
MAL
NADH
Glu
OXO
CO2
NADPH

Figure 3.2.2. *Cyanothece* 51142 growth curves under different nitrogen and carbon sources (biological replicates, n=3). ◊: Glycerol+Nitrate; □: Glycerol+N_2; Δ: CO_2+ Nitrate; ○: CO_2+N_2. The error bars are smaller than the symbols.

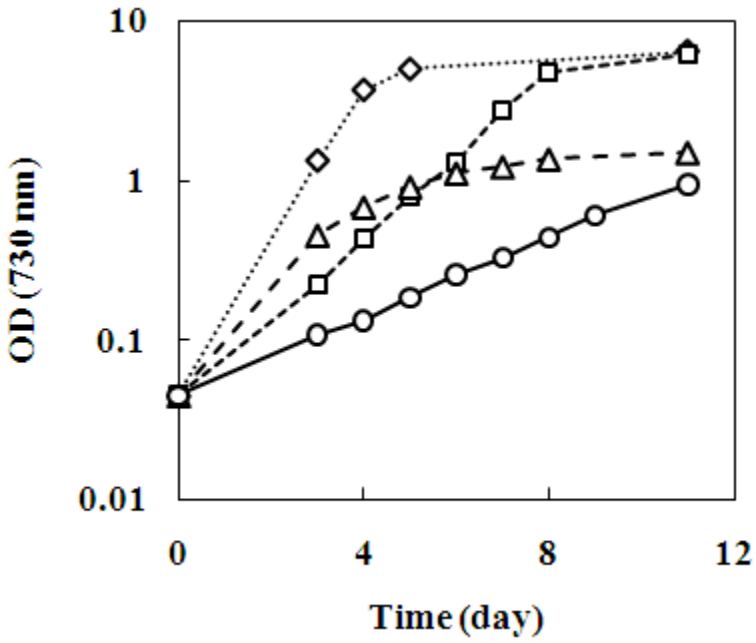
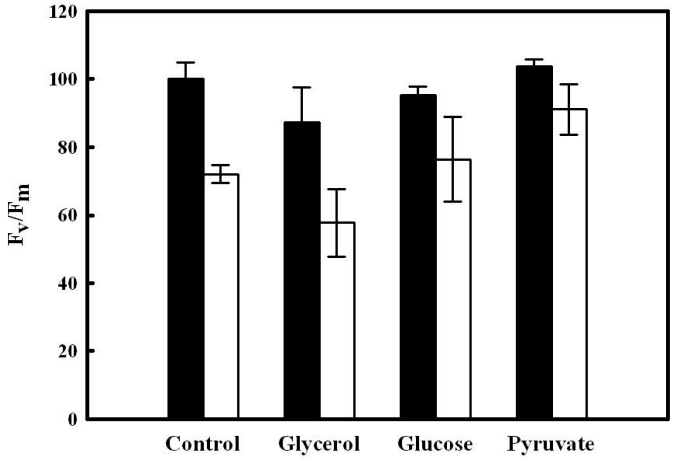
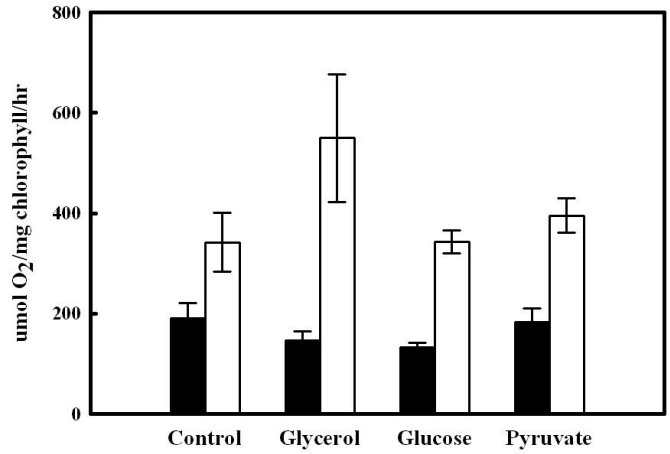


Figure 3.2.3. Maximum quantum yields (Figure 3.2.3a) of PSII and oxygen evolution rates (Figure 3.2.3b) in *Cyanothece* 51142 under different growth conditions (biological replicates, n=3). All samples were taken at the exponential growth phase based on the growth curve. Black column, N_2 as nitrogen source; white column, $NaNO_3$ as nitrogen source; Error bars, SD.

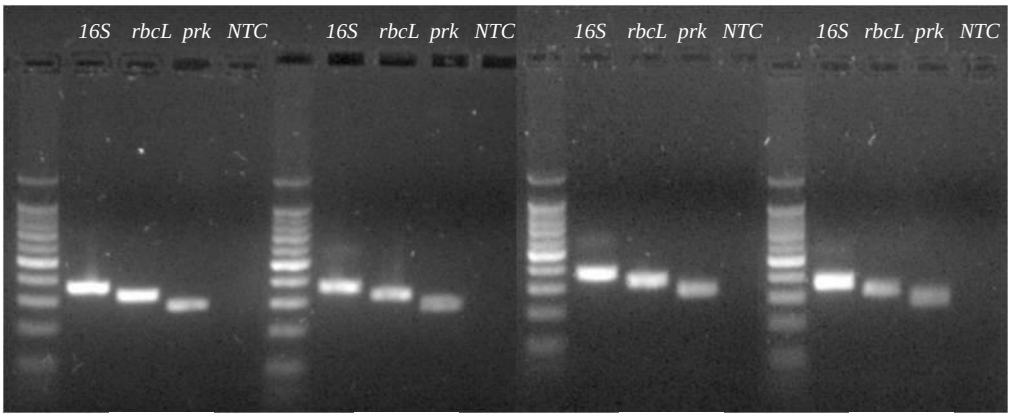


(a)



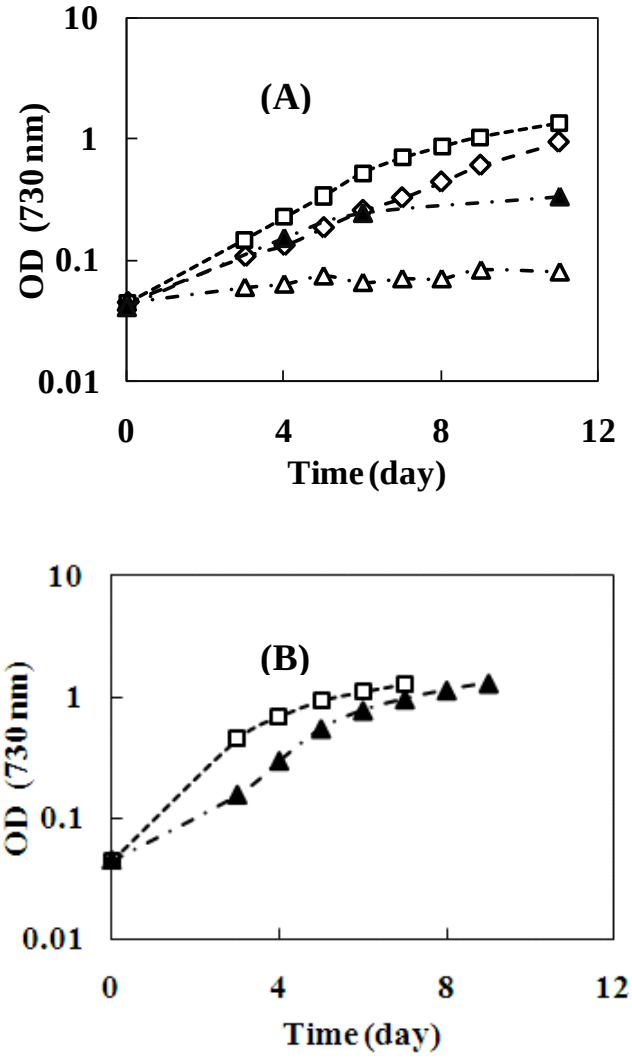
(b)

Figure 3.2.4. Reverse transcription PCR (RT-PCR) study for ribulose-1,5-bisphosphate carboxylase oxygenase (*rbcL*) and phosphoribulokinase (*prk*) under different mixotrophic growth conditions. (a) CO_2+N_2; (b) CO_2+$NaNO_3$; (c) glycerol+$NaNO_3$; (d) glucose+$NaNO_3$. The 16S rRNA gene was used as the internal reference; the no template control (NTC) was added under each mixotrophic growth conditions.

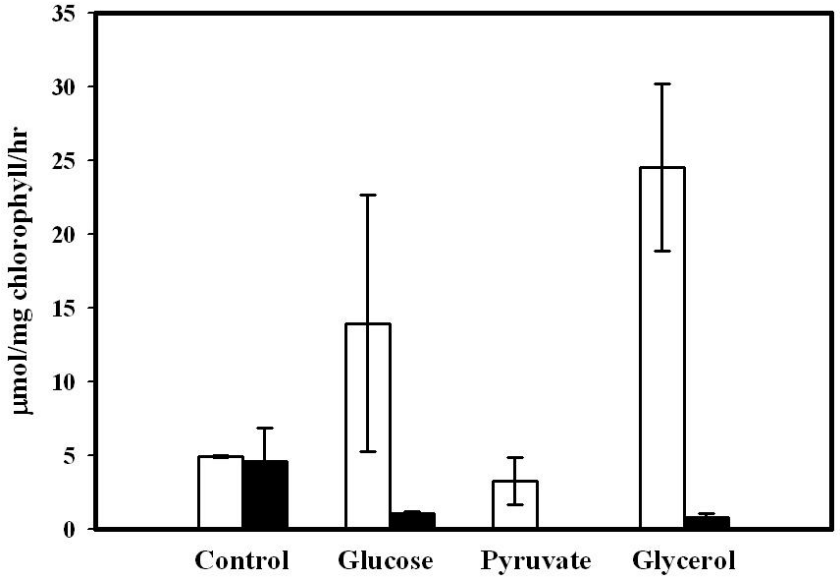


(a) **(b)** **(c)** **(d)**

Supplementary Figure 3.2.S1. The growth of Cyanothece 51142 in the presence of different carbon and nitrogen substrates under continuous light (n=3). (a) N_2 as nitrogen source; (b) $NaNO_3$ as nitrogen source. □, No carbon source; □, glucose (26 mM); □, pyruvate (64 mM); ▲, pyruvate (9 mM). The error bars are smaller than the symbols.



Supplementary Figure 3.2.S2. Hydrogen production under mixotrophic conditions. Open bars, hydrogen production at day 4; filled bars, hydrogen production at day 9. The asterisk indicates that hydrogen production was below the minimum detectable level.



Supplementary Figure 3.2.S3. Growth of *Synechocystis* 6803 in the presence of glucose (6 mM) under continuous light (n=3). ^{13}C-labelled samples were taken in the midexponential phase (~80 h), when there was still sufficient glucose in the medium. The error bars are smaller than the symbols.

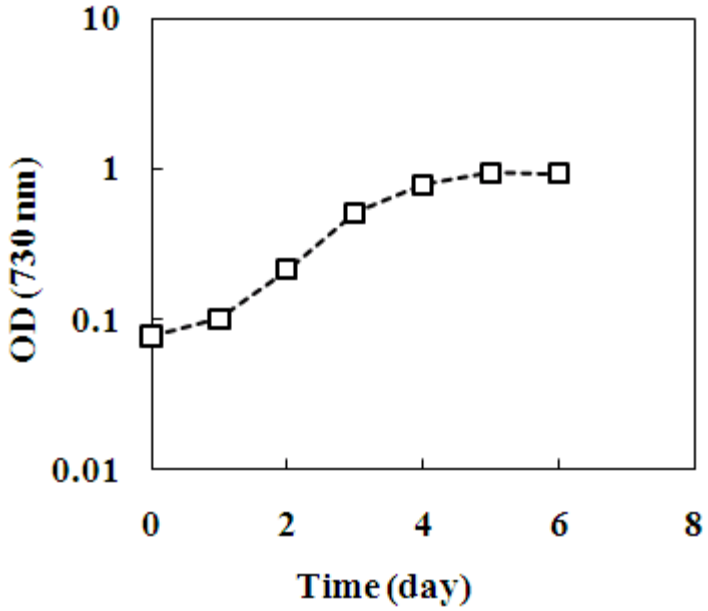


Table 3.2.1. Isotopic analysis of the labeling profiles of amino acids in *Cyanothece* 51142 and *Synechocystis* 6803 under different growth conditions (the standard error for GC-MS measurement was below 0.02, technical replicates, n=2)

Amino	$[M-57]^+$	N_2						$NaNO_3$						*Synechocystis* 6803 (nitrate-medium)	
Acids		Glucose	R^1	Pyruvate	R	Glycerol	R	Glucose	R	Pyruvate	R	Glycerol	R	Glucose	R
Ala	M_0	0.67		0.41		0.19		0.61		0.51		0.07		0.04	
	M_1	0.19	***0.032***	0.55	***0.597***	0.71	***4.2***	0.19	***0.042***	0.48	***0.327***	0.85	+∞	0.05	***0.92***
	M_2	0.11		0.03		0.10		0.17		0.01		0.07		0.28	
Ser	M_0	0.65		0.98		0.20		0.58		0.97		0.08		0.04	
	M_1	0.22	***0.033***	0.02	***0***	0.72	***3.7***	0.22	***0.046***	0.03	***0***	0.81	+∞	0.06	
	M_2	0.10		0		0.09		0.16		0		0.10		0.28	***0.87***
Asp	M_0	0.58		0.54		0.10		0.59		0.94		0.07		0.04	
	M_1	0.24	***0.030***	0.43	***0.195***	0.64	***2.2***	0.20	***0.032***	0.06	***0.005***	0.78	***1.53***	0.05	***0.44***
	M_2	0.11		0.04		0.25		0.17		0		0.15		0.19	
	M_3	0.06		0		0.01		0.03		0		0		0.47	
Glu[2]	M_0	0.43		0.15		0.02		0.38		0.47		0.01		0.02	
	M_1	0.26	***0.041***	0.44	***1.25***	0.14	***-1.78***	0.22	***0.051***	0.49	***0.170***	0.15	***-2.11***	0.02	***0.76***
	M_2	0.21		0.37		0.62		0.27		0.04		0.74		0.04	
	M_3	0.07		0.04		0.21		0.09		0		0.10		0.07	
	M_4	0.02		0		0.01		0.03		0		0		0.53	
His	M_0	0.44		0.91		0.05		0.33		0.92		0.01		0.01	
	M_1	0.28		0.08		0.28		0.24		0.08		0.21		0.01	
	M_2	0.17	***0.032***	0	***0***	0.50	***2.83***	0.22	***0.049***	0	***0***	0.55	+∞	0.02	***1.73***
	M_3	0.07		0		0.16		0.12		0		0.20		0.03	
	M_4	0.03		0		0		0.07		0		0.03		0.06	
	M_5	0		0		0		0.01		0		0		0.22	

Note: 1. Bold values were the carbon substrate (glycerol, pyruvate, or glucose) utilization ratios (substrate/CO_2 fixation) for amino acid synthesis calculated according to Equation (1). 2. The glutamate synthesis pathway involved the loss of two carbons from pyruvate to α-ketoglutarate. Such a microbial process changed the labeling enrichment, and the negative value indicated the net loss of unlabeled CO_2.

3.3 Characterization of the central metabolic pathways in *Thermoanaerobacter* sp. X514 via isotopomer-assisted metabolite analysis

Xueyang Feng[1], Housna Mouttaki[2], Lu Lin[2,3], Rick Huang[1], Bing Wu[1], Christopher L. Hemme[2], Zhili He[2], Baichen Zhang[4], Leslie M. Hicks[4], Jian Xu[3], Jizhong Zhou[2], and Yinjie J. Tang[1,*]

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

[2]Institute for Environmental Genomics and Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019, USA

[3] Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China

[4] Donald Danforth Plant Science Center, 975 N. Warson Rd., St. Louis, MO 63132, USA

*Corresponding author. Email: yinjie.tang@seas.wustl.edu; phone: 314-935-3441

This section was previously published in

Feng X, Mouttaki H, Lin L, Huang R, Wu B, Hemme CL, He Z, Zhang B, Hicks LM, Xu J, Zhou J, Tang YJ. Characterization of the central metabolic pathways in Thermoanaerobacter sp. strain X514 via isotopomer-assisted metabolite analysis. Appl Environ Microbiol. 2009, 75(15):5001-8.

3.3.1 Abstract

Thermoanaerobacter sp. X514 has great potential in biotechnology due to its capacity to ferment a range of C5 and C6 sugars to ethanol and other metabolites under thermophilic conditions. This study investigated the central metabolism of strain X514 via ^{13}C-labeled tracer experiments using either glucose or pyruvate as both the carbon and energy source. X514 grew on a minimal medium and thus contains complete biosynthesis pathways for all macromolecule building blocks. Based on genome annotation and isotopic analysis of amino acids, three observations can be obtained about the central metabolic pathways in X514. First, the oxidative pentose phosphate pathway in X514 is not functional, and the TCA cycle is incomplete under fermentative growth conditions. Second, X514 contains (*Re*)-type citrate synthase activity, although no gene homologous to the recently characterized (*Re*)-type citrate synthase of *Clostridium kluyveri* was found. Third, the isoleucine in X514 is derived from acetyl-CoA and pyruvate via the citramalate pathway rather than being synthesized from threonine via threonine ammonia-lyase. The functionality of the citramalate synthase gene (*CimA*, Teth514_1204) has been confirmed by enzymatic activity assay, while the

presence of intracellular citramalate has been detected by mass spectrometry. This study demonstrates the merits of combining ^{13}C-assisted metabolite analysis, enzyme assays, and metabolite detection not only to examine genome sequence annotations but also to discover novel enzyme activities.

Key words: ethanol, ^{13}C, (Re)-type citrate synthase, isoleucine, citramalate

3.3.2 Introduction

Rising global energy demand and the depletion of fossil energy resources have resulted in significant environmental, economic, and social impacts. Production of renewable, biomass-derived energy sources has been suggested as a partial solution to this problem. Among renewable energy sources, ethanol is an attractive short-term solution owing to its strong research foundation and its ready integration with the current petroleum-based infrastructure (53,54). Plant-based cellulose is the most attractive raw material for bioethanol production (55). However, the use of anaerobic cellulosic bacteria in consolidated bioreactors has been proposed as an efficient means of rapid conversion of cellulosic biomass to ethanol (56). Thermophilic bacteria of the genus *Thermoanaerobacter* have the ability to naturally ferment a wide variety of monomeric and polymeric carbohydrates, including D-xylose, into ethanol (57,58,59). While not cellulose-utilizing themselves, *Thermoanaerobacter* species in co-culture with thermophilic cellulose-utilizing *Clostridium* species have significantly higher yields of ethanol from both cellulose and hemicellulose than from monoculture alone (60,61).

Therefore, the investigation of carbon metabolism in *Thermoanaerobacter* sp. X514 has implications for understanding X514's potential in bioenergy production.

Despite the potential importance of X514 in biofuel production, a rigorous investigation of the central metabolic pathways in X514 has yet to be conducted. Although an array of functional genomics tools has been applied in predicting this species' metabolism (62,63,64), a precise description of cellular metabolism is complicated by misannotation and by post-transcriptional regulation of protein synthesis (20,65). The complete genome sequence of X514 from the KEGG database (http://www.genome.jp/kegg/) suggests a few gaps in several essential pathways involved in the biosynthesis of amino acids (e.g. isoleucine) and in the TCA cycle (e.g., citrate synthase). Therefore, X514 would not survive without supplements of isoleucine or other essential nutrients. However, X514 can actually grow in a completely minimal medium. Hence, the metabolism of X514 cannot be precisely revealed by genome sequence annotation alone. At this time, one of the most physiologically reliable methods for determining cell metabolism remains ^{13}C-based isotopic analysis (20,21,22,66). Based on ^{13}C-labeling patterns in key amino acids, the active pathways can be traced back, and new enzymes can be revealed. In this study, ^{13}C-based isotopic analysis was applied to accurately examine the annotated pathways in X514 and to investigate gaps in key biosynthetic pathways (22,31,67). Specifically, glucose (1^{st} or 6^{th} carbon labeled)

and pyruvate (1^{st} carbon labeled) were used respectively as the sole source of carbon to grow X514. By analyzing the mass spectra of different fragmentations in proteogenic amino acids derived from various pathways, we have determined the active fluxes of intracellular pathways (e.g., the pentose phosphate pathway and citric acid cycle). Meanwhile, misannotations in the genome sequence were checked and unknown enzymes involved in the pathway were identified. The isotopomer analysis linked the genome annotation to the final enzyme-functional output and thus significantly improved our understanding of the regulation of the central metabolism of X514.

3.3.3 Materials and methods

Medium and cultivation conditions. *Thermoananerobacter* sp. X514 was grown anaerobically at 60°C without shaking (68). The minimal medium contained (per liter) 1 g of NaCl, 0.5 g of $MgCl_2$, 0.2 g of KH_2PO_4, 0.3 g of NH_4Cl, 0.3 g of KCl, 0.015 g of $CaCl_2$, 0.25 mg of resazurin, 0.031 g of L-cysteine-HCl, 0.048 g of Na_2S, 2.52 g of $NaHCO_3$ and 1 ml trace element solution. One liter of trace element solution included: 10 ml 25% w/w HCl solution, 1.5 g of $FeCl_2$, 0.19 g of $CoCl_2$, 0.1 g of $MnCl_2$, 70 mg of $ZnCl_2$, 6 mg of H_3BO_3, 36 mg of Na_2MoO_4, 24 mg of $NiCl_2$, 2 mg of $CuCl_2$, 6 mg of Na_2SeO_3, 8 mg of Na_2WO_4, and 0.5 g of NaOH. The pH of the medium was adjusted with NaOH to 7.2-7.3. The vitamin solution was prepared according to the method developed by Wolin et al. (69). The rich medium was prepared by adding 0.1%

yeast extract to the minimal medium. Three types of ^{13}C-labeled carbon substrates were obtained from Cambridge Isotope Laboratories, Inc. (Andover, MA) and used for cell culture: pyruvate ([1-^{13}C], 98%), glucose ([1-^{13}C], 98%), or glucose ([6-^{13}C] , 98%). The medium was flushed with N_2 and was filter-sterilized. All gases, including nitrogen and a nitrogen-CO_2 mixture, were obtained from Airgas, Inc. (Radnor, PA). The strain was initially grown in a 50-ml culture medium with an unlabeled carbon source (glucose or pyruvate). At the mid-log phase of growth, a 3% inoculum was added to a 50-ml culture containing one of the following carbon sources: 2.1 g/L of [1-^{13}C] glucose, 2.1 g/L of [6-^{13}C] glucose, or 2.2 g/L 1st-position labeled pyruvate. At the mid-log phase of growth in this culture, 3% inoculum from the first ^{13}C-labeled culture medium was used to inoculate a 50-ml sub-culture (with the same labeled carbon source), which reduced the effect of unlabeled carbon from the initial stock.

Analytical methods. Biomass was harvested at the late-log phase of growth by centrifugation at 8,000 g for 15 min at 10°C. The concentrations of glucose, acetate, ethanol, and lactate were analyzed with a high performance liquid chromatography apparatus (Agilent Technologies, CA) equipped with a variable wavelength (190-600 nm) detector (VWD) (the UV absorption at 245 nm), and an ion exclusion column (Aminex HPX-87H, 300 mm × 7.8 mm, Bio-Rad Laboratories, CA) operating at 55°C. The mobile phase consisted of 0.025 % sulfuric acid at a flow rate of 0.6 ml/min.

Isotopic analysis. The preparation and isotopic analysis of proteogenic amino acids were performed as previously described (30,67). In brief, biomass was hydrolyzed in 6 M HCl at 100°C for 24 h. The amino acid solution was dried under air flush overnight, and amino acid samples were derivatized in tetrahydrofuran (THF) and N-(*tert*-butyl dimethylsilyl)-N-methyl-trifluoroacetamide (Sigma-Aldrich, St. Louis, MO) at 70°C for 1 h. A gas chromatograph (GC) (Hewlett-Packard, model 7890A, Agilent Technologies, CA) equipped with a DB5-MS column (J&W Scientific, Folsom, CA) and a mass spectrometer (MS) (model 5975C, Agilent Technologies, CA) was used for analyzing amino acid labeling profiles. Three types of charged fragments were detected by GC-MS for most amino acids: the $[M-57]^+$ group, which contained unfragmented amino acids; the $[M-159]^+$ group, which contained amino acids losing α carboxyl group; and the $[f302]^+$ group, which contained only 1^{st} (α carboxyl group) and 2^{nd} carbons in an amino acid backbone. (However, $[f302]^+$ cannot be detected in some amino acids.) Published algorithms were used to correct the effects of natural isotopes on the mass distributions of amino acids (32), and the final isotopomer distribution was shown in Tables 3.3.S1 and 3.3.S2. Ion mass fractions in Tables 3.3.S1 and 3.3.S2 were given for the amino acid fragments. M0, M1, M2… were fractions of unlabeled, singly labeled, and doubly labeled amino acids, respectively.

Confirmation of citramalate synthase. The citramalate synthase activity

was assayed by monitoring the pyruvate-dependent production of CoA over time (70,71). In brief, X514 cells from a 10-ml culture in the mid-log phase were centrifuged (19,000 g for 10 min, 4°C). The total protein content in the biomass was estimated using Bradford protein assay (Bio-Rad Laboratories, Inc, CA). X514 cell extracts were then prepared by sonication of X514 pellets for 3 minutes (30 sec on / 20 sec off) in 2 ml of a 0.1 M TES [N-tris (hydroxymethyl) methyl-2-aminoethanesulfonic acid, pH 7.5] buffer. Samples to be measured were brought to a final volume of 1000 µl by mixing with cell extracts (100 µl), the TES buffer (0.2 M, pH 7.5, 500 µl), pyruvate (10 mM, 100 µl), acetyl-CoA (50 mM, 20 µl), and distilled water. The resulting solutions were then incubated in the oven at 60°C for 2 h. At intervals of 20 min, 100 µl of either test samples or blank samples were taken from the oven and mixed with a 900-µl stop solution. The stop solution was prepared in distilled water with 50 µl of 10 mM DTNB [5, 5'-dithio-bis (2-nitrobenzoic acid)] in 0.1 M Tris-HCl and with 70 µl of 1 M Tris-HCl. The absorbance at 412 nm was recorded immediately and blanked against an identical incubation sample without pyruvate. The micromoles of HS-CoA produced were calculated from a standard curve generated with known concentrations (0 to 1 mM) of 2-mercaptoethanol and based on the linear function between the product formation and the amount of enzyme added over the 2-hour time period of the assay. All the chemicals employed in this measurement were from Sigma-Aldrich (MO, USA).

LC-MS/MS was further applied to confirm the presence of citramalate in X514. In this study, the free intracellular metabolites were extracted with a cold methanol-water solution (60%, v/v) and chloroform mixture (1:1), held at -20°C overnight. Subsequently, cold water was added to separate free metabolites from those components that might interfere with the results. The extracts were then lyophilized. After being dissolved with 1% formic acid in water, 50 μL samples were injected into the LC-MS/MS for separation and detection. The LC-MS/MS system was composed of a Shimadzu LC system (Shimadzu Corporation, Japan, http://www.shimadzu.com), a LEAP CTC PAL autosampler (LEAP Technologies, USA, http://www.leaptec.com), and an Applied Biosystems 4000 QTRAP mass spectrometer (Applied Biosystems, USA, http://www.appliedbiosystems.com) equipped with a TurbolonSpray electrospray ion source. 5 μM of citrate, malate, and citramalate standards (Sigma, MO, USA) in water were separately infused into the mass spectrometer to optimize compound-dependent parameters for multiple reaction monitoring (MRM) detection and to obtain corresponding MS/MS spectra. LC separation was achieved by coupling three 4.6×300 mm Onyx Monolith C18 columns (Phenomenex, CA) in tandem. The LC gradient was delivered at 1 ml/min with Solvent A: 0.1% formic acid in water, and Solvent B: 0.1% formic acid in methanol. The gradient started from 5% B and was kept isocratic for 4 min, then ramped to 20% within 7 min, and increased to 95% B within 1 min. Finally, after being held at 95% B for 8 min, the gradient was ramped down to 5% B, where it

remained for 4 min to re-equilibrate the column.

3.3.4 Results

Growth and metabolite curves. When X514 was cultivated in the minimal medium, the lag phase for X514 in the ^{13}C labeled glucose was 12 hours and was followed by an exponential growth phase with a doubling time of ~15-19 hours (Fig. 3.3.1a). This rate was much slower than X514's growth rate in the rich medium (with 0.1% yeast extract), which was measured a doubling time of 6 hours. Fig. 3.3.1b shows both the glucose consumption and the production of ethanol, acetate, and lactate when X514 was grown in the minimal glucose medium. No formate was detected during the culturing. The ability to grow in the minimal medium using glucose (or pyruvate, data not shown) indicates that X514 contains the necessary biosynthetic pathways for all macromolecule building blocks (i.e., for synthesizing amino acids).

Confirmation of amino acid biosynthetic pathways. According to the genome annotation from the KEGG resource (http://www.genome.jp/kegg/), two amino acid biosynthetic pathways (isoleucine and proline) were incomplete. To examine biosynthetic pathways in X514, the labeling profiles of 14 proteogenic amino acids were analyzed (Tables 3.3.S1 and 3.3.S2). Since pyruvate is the key metabolite in the central pathway (i.e., glycolysis, pentose phosphate pathway, and TCA cycle), the labeling profiles of amino acids from tracer experiments that use 1^{st}-position labeled pyruvate as the carbon source can easily identify the

precursors of several key amino acids (Table 3.3.S2). For example, alanine, valine, and serine demonstrated the same carbon molecule-labeling pattern as that of pyruvate (Table 3.3.S2), which confirms that pyruvate is the common precursor of these three amino acids (Fig. 3.3.2). The aromatic amino acids phenylalanine and tyrosine were derived from phosphoenolpyruvate (also synthesized from pyruvate). The labeling patterns of the $[f302]^+$ group (containing the 1st and 2nd carbons of fragmented amino acids) for alanine, phenylalanine, and tyrosine were all identical. This fact indicates that the 1st and 2nd carbons of all three amino acids are originally derived from the 1st and 2nd carbons of pyruvate.

In tracer experiments using the 1st-position labeled pyruvate, the similar labeling patterns of aspartate, methionine, and threonine in Table 3.3.S2 suggest the same precursor (oxaloacetate) for the three amino acids. Based on the KEGG-generated pathway map, oxaloacetate is expected to be synthesized from pyruvate (labeled at its first position), and CO_2 (labeled because it is cleaved from the 1st carbon of pyruvate), leading to the labeling of two carbons in oxaloacetate and aspartate (both α- and β- carboxyl group) (Fig. 3.3.2). In tracer experiments using [1-^{13}C] or [6-^{13}C] glucose as the carbon source, the labeling percentage of methionine was higher than that of aspartate from the same experiment, as a result of the addition of ^{13}C-enriched C1 pool (5, 10 methyl-THF) into the carbon backbone of methionine (synthesis route: aspartate + C1 pool→methionine). The labeled carbon entered the C1 pool via the glucose→serine→glycine+C1 pool

(Fig. 3.3.2), which caused more methionine to be labeled.

Alternate isoleucine pathway. Anaerobic bacteria such as *Methanococcus jannaschii* and *Leptospira* can biosynthesize isoleucine from citramalate by direct condensation of acetyl-CoA and pyruvate (70,72). Recently, Risso et al. (71) first used ^{13}C-assisted metabolic flux analysis and biochemistry assays to discover an alternate isoleucine pathway in *Geobacter sulfurreducens*. Interestingly, X514 may also contain a similar alternate isoleucine pathway. The labeling patterns of leucine and isoleucine in both glucose and pyruvate tracer experiments were similar. For example, the pyruvate experiment demonstrated that the M0 ($[M-159]^+$, without the carboxyl group) of both leucine and isoleucine was >91%, indicating that the C2~C6 carbons in both leucine and isoleucine were mostly unlabeled. Such labeling patterns in isoleucine are unexpected unless isoleucine shares the same precursors (pyruvate and acetyl-CoA) as leucine. According to the genome annotation, X514 lacks threonine ammonia-lyase (EC 4.3.1.19), which is necessary for the biosynthesis of isoleucine from threonine; this result is supported by our labeling data: when [1-^{13}C] pyruvate was used as the carbon source, threonine was labeled with two carbons (M2 $[M-57]^+$ >93%), while isoleucine (C2~C6) was not labeled (note: the GC-MS could not detect the labeling information for the first carbon of isoleucine due to the peak overlap (73)). We propose an alternate isoleucine synthesis pathway via the citramalate pathway that uses pyruvate and acetyl-CoA as precursors (Fig. 3.3.3). The key

enzyme regulating the citramalate pathway (citramalate synthase, *CimA*) is annotated in X514 (gene: Teth514_1204) (22,71).

In order to determine whether the citramalate pathway was active in X514, crude soluble extracts from mid-log phase cells were tested for the presence of citramalate synthase activity (about 27±9 nmol/mg protein/min). Furthermore, we applied high sensitivity mass spectrometry to detect whether there was any intracellular citramalate in X514. Fig. 3.3.S1 indicates that malate, citrate, and citramalate were clearly detected in the X514 soluble extracts by LC/MS/MS in the MRM mode. The retention time and the ratio of multiple MRM transitions selected for each of the three targeted compounds are in agreement with those of the respective authentic standards. Although CID spectra signals for the three targeted compounds in X514 soluble extracts were lower than the spectra signals of authentic standards, the major fragments and their relative abundance agreed with their corresponding standards (Fig. 3.3.S2). This result provides additional evidence to prove the presence of citramalate synthase in X514.

(*Re*)-type citrate synthase. In spite of the production of citrate by X514 (Fig. 3.3.S1), the genome annotation indicates that the enzyme in the first step of TCA cycle, i.e. the (*Si*)-type citrate synthase, is missing. Interestingly, the isotopomer data from the 1^{st}-position labeled pyruvate experiment suggests that the α-carboxyl group of glutamate is not labeled (Table 3.3.S2). Such an observation of the labeling pattern of glutamate is consistent with the main

characteristics of (*Re*)-type citrate synthase, which has been discussed in detail in our previous study on (*Re*)-type citrate synthase in *Desulfovibrio vulgaris Hildenborough* (22). Due to the lack of a regular (*Si*)-type citrate synthase, X514 may employ the (*Re*)-type citrate synthase. Fig. 3.3.2 shows the proposed carbon transition routes from labeled pyruvate (α-carboxyl group) to doubly labeled oxaloacetate (both α- and β- carboxyl group). The β-carboxyl group of 2-oxoglutarate and the β-carboxyl group of glutamate were ultimately derived from the β-carboxyl group of oxaloacetate.

Pentose phosphate pathway. About 50% of alanine was not labeled ($[M-57]^+$, M0 =0.51) when the [1-^{13}C] glucose was used as the carbon source (Table 3.3.S1). The glucose carbon backbone loses the first carbon as CO_2 when it is metabolized via the oxidative pentose phosphate pathway (PP pathway) to synthesize 5-carbon sugars (C5P) (Fig. 3.3.2). If the oxidative pentose phosphate pathway is not active, one glucose molecule (with either 1st or 6th carbon labeled) converts to two pyruvate molecules, and thus ~50% of pyruvate (inferred from alanine) is expected to be unlabeled. The fact that the fraction of unlabeled alanine was 51±2% indicates the very low activity of the oxidative pentose phosphate pathway, i.e., the carbon flux split ratio between G6P→C5P (oxidative pentose phosphate pathway) and G6P→G3P (glycolysis) was <3% (the algorithm for calculating the split ratio was provided in the supplementary material). Since the oxidative pentose phosphate pathway was not active, the isotopomer labeling

patterns of most amino acids in the [1-^{13}C] glucose experiment were identical to those in the [6-^{13}C] experiments (Table 3.3.S1).

3.3.5 Discussion

This study has examined the pentose phosphate pathway, amino acid biosynthesis and TCA cycle in *Thermoanaerobacter* sp. X514 by ^{13}C labeling experiments. X514 shows very low activity in the oxidative phase of the pentose phosphate pathway under glucose fermentation conditions. Such an observation is consistent with the missing of 6-phosphogluconolactonase (EC 3.1.1.31) gene, which catalyzes 6-phospho-D-glucono-1,5-lactone to 6-phospho-D-gluconate. Considering the important role of the oxidative pentose phosphate pathway in providing NADPH for biosynthesis, alternate NADPH pathways should be present in X514. Some bacteria utilize transhydrogenase PntAB or UdhA for NADPH generation (74)**,** but a BLASTP search (75) indicates that neither of the two transhydrogenases is encoded in the X514 genome (Fig. 3.3.S3). Although NADPH-dependent isocitrate dehydrogenase is annotated in X514 (Teth514_0327), it may provide only limited NADPH since the TCA cycle is branched and is mainly used for biosynthesis. On the other hand, ferredoxin-NADP$^+$ reductase (e.g., Teth514_0652) and pyruvate:ferredoxin oxidoreductase (e.g., Teth514_0781, the enzyme that catalyzes the production of ferredoxin from pyruvate) are annotated in the X514 genome and may be key sources of NADPH

in X514. Ferredoxin-$NADP^+$ reductase activity has been well documented in certain thermophilic anaerobes for NADPH production (76).

The discovery of an alternative pathway for isoleucine synthesis and (*Re*)-type citrate synthase activity demonstrates the unique metabolism of X514. Isoleucine is synthesized from the citramalate pathway so that both leucine and isoleucine share the same precursors (pyruvate and acetyl-CoA). Multiple lines of evidence support an alternate isoleucine biosynthesis pathway via citramalate in X514: 1) labeling patterns in key amino acids, 2) genomic evidence (i.e. presence of citramalate synthase gene but absence of threonine deaminase gene), 3) detection of citramalate synthase activity, 4) detection of citramalate via LC-MS/MS. In some organisms, citramalate is used for reactions other than isoleucine biosynthesis. For example, during *Rhodospirillum rubrum* phototrophic growth on acetate and CO_2, citramalate (formed via the condensation of acetate and pyruvate) can be degraded to glyoxylate and propionate (77). This pathway allows *R. rubrum* to assimilate acetate and synthesize intermediates in TCA cycle (i.e., succinate). On the other hand, Atsumi and Liao (78) introduced citramalate synthase (*CimA*) to *E.coli* and successfully evolved a new 2-oxobutanoate synthetic pathway for both 1-propanol and 1-butanol production (9- and 22-fold higher production, respectively). Therefore, the citramalate pathway in X514 could be potentially utilized in bio-butanol production.

A (*Re*)-type citrate synthase was recently reported in both *Clostridium kluyveri* (CKL 0973) and *Desulfovibrio* spp. (79). (*Re*)-type citrate synthase and (*Si*)-type citrate synthase are phylogenetically unrelated. (*Re*)-type citrate synthase is O_2 sensitive and thus restricted to anaerobic microorganisms. Using a JGI genome database search by May 2009 (http://img.jgi.doe.gov, with the BLAST search score >400 and with identity of amino acid sequences >40%), we found that ~200 microbial species may be annotated with citramalate synthase (a key step for an alternate isoleucine synthesis pathway), while ~40 stains may have (*Re*)-type citrate synthase (with the BLAST search score >290 and with identity >40%). A few species may contain both citramalate synthase and the (*Re*)-type citrate synthase (Table 3.3.1), including *Desulfovibrio desulfuricans* and *Desulfovibrio vulgaris Hildenborough*. *Thermoanaerobacter* sp. X514, *Dehalococcoides ethenogenes* and *Pelotomaculum thermopropionicum* species contain neither (*Si*)-type citrate synthase nor the documented (*Re*)-type citrate synthase, but all encode citramalate synthase. Table 3.3.1 indicates that the citramalate pathway and (*Re*)-type citrate synthase may be more widespread in microorganisms than originally thought.

(*Re*)-type citrate synthase-like activity was observed in X514. A candidate (*Re*)-type citrate synthase gene in X514 should be identifiable through a BLAST search of the polypeptide amino acid sequences (Fig. 3.3.S4). However, no gene candidate is identical to the reported (*Re*)-type citrate synthase in

Clostridium kluyveri (CKL 0973), and thus, X514 may contain an undocumented (*Re*)-type citrate synthase. However, X514 contains homocitrate synthase (Teth514_0415: 2-oxoglutarate+acetylCoA→homocitrate) and isopropylmalate synthase (Teth514_0014: 3-methyl-2-oxobutanoate+acetylCoA→2-isopropylmalate), which are phylogenetically related to the reported (*Re*)-type citrate synthase (oxalacetate+acetylCoA→citrate). More importantly, Teth514_0416 is annotated as aconitate hydratase (citrate→isocitrate), and this gene is in the same operon with homocitrate synthase (Teth514_0415). Therefore, homocitrate synthase (Teth514_0415) can be a potential (*Re*)-type citrate synthase candidate, and further experimentation is required to test this hypothesis. Interestingly, citramalate synthase (Teth514_1204: pyruvate+acetyl-CoA→citramalate) condenses acetyl-CoA and organic acids to form metabolites structurally similar to citrate. This enzyme also belongs to the isopropylmalate synthase/homocitrate synthase family.

In summary, ^{13}C isotopic analysis is a powerful tool to examine the metabolic networks of sequenced species and to predict novel enzymes. Our results suggest an inactive pentose phosphate pathway and an alternate isoleucine biosynthesis pathway via citramalate in X514. Furthermore, X514 also demonstrates the (*Re*)-type citrate synthase activity. A comprehensive understanding of metabolism in *Thermoanaerobacter* sp. X514 could have dual significance for both rational genetic engineering of microorganisms in biofuel

production (78) and the investigating of the evolution for phylogenetically related pathways.

3.3.6 Supporting information

Flux calculations in oxidative pentose phosphate pathway. The flux directed into the oxidative pentose phosphate (PP) pathway can be measured using the labeling profile of alanine. We assume that the flux through glycolysis is v_1 and that the flux through oxidative pentose phosphate pathway is v_2. The replenished route from the oxidative pentose phosphate pathway to glycolysis is v_3. Hence, the percentage of fluxes directed into the pentose phosphate pathway would satisfy:

$$\eta = \frac{v_2}{v_1 + v_2} \times 100\% .$$

In pure [1-^{13}C]-glucose experiments, the flux to unlabeled alanine through glycolysis is 0.5 v_1, while the flux to unlabeled alanine through pentose phosphate pathway is v_3. Hence, the percentage of unlabeled alanine satisfies:

$$\eta_{ala} = \frac{0.5v_1 + v_3}{v_1 + v_3} \times 100\%$$

When G6P enters the pentose phosphate pathway, the first carbon is lost as CO_2. Hence, only five carbons from the six-carbon glucose flow into C5P. C5P is then re-formed to synthesize G3P. Based on the mass balance, the equilibrium

should satisfy:

$$v_3 = \frac{5}{6} v_2 .$$

From the $[M-57]^+$ data in GC-MS, the percentage of unlabeled alanine is 51%. Hence,

$$\eta_{ala} = \frac{0.5v_1 + v_3}{v_1 + v_3} \times 100\% = 51\% .$$

With all the equations above, the percentage of flux directed into the pentose phosphate pathway is 2.4%.

Note: when the 6th carbon of glucose was labeled, the labeled carbon incorporated into E4P (the precursor of phenylalanine) via the reductive PP pathway. However, when the 1st carbon of glucose was labeled, the resulting E4P was not labeled via PP pathway, so the labeling percentage of phenylalanine was lower when the 1st carbon of glucose was used as carbon source (Table 3.3.S1).

Figure 3.3.1. *Thermoanaerobacter* sp. X514 growth and metabolite secretion. (a) Growth curves of *Thermoanaerobacter* sp. X514 in minimal glucose medium (symbol: ○) and rich medium (with 1g/L yeast extract) (symbol: □). (b) Glucose consumption and metabolite production by *Thermoanaerobacter* sp. X514 in minimal glucose medium: glucose (symbol: ■); acetate (symbol: ◊); lactate (symbol: □); ethanol (symbol: ○).

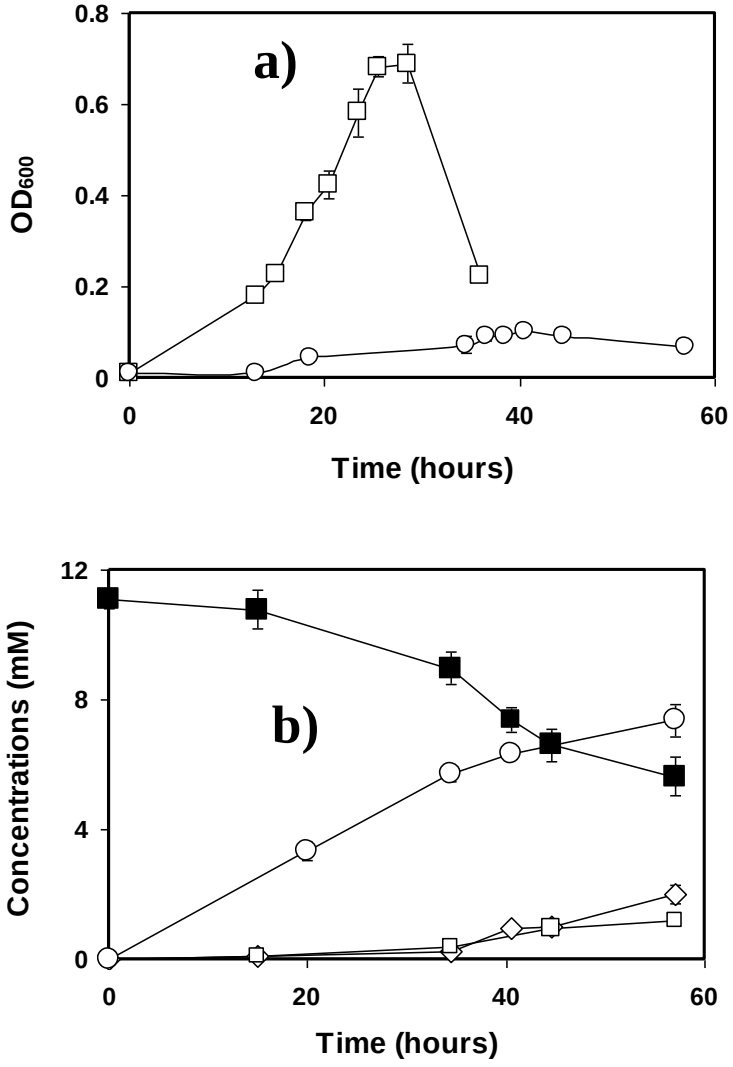


Figure 3.3.2. Proposed scheme of central metabolic pathways in *Thermoanaerobacter* sp. X514. ^{13}C labeled positions in metabolites are marked with an asterisk (*) for the 1st-position labeled pyruvate experiment. The inactive pathways are marked with dashed lines. Based on the genome annotation in KEGG database, some genes that are missed in the pathways are marked by dashed boxes. C5P: ribulose 5-phosphate, ribose 5-phosphate and xylulose 5-phosphate.

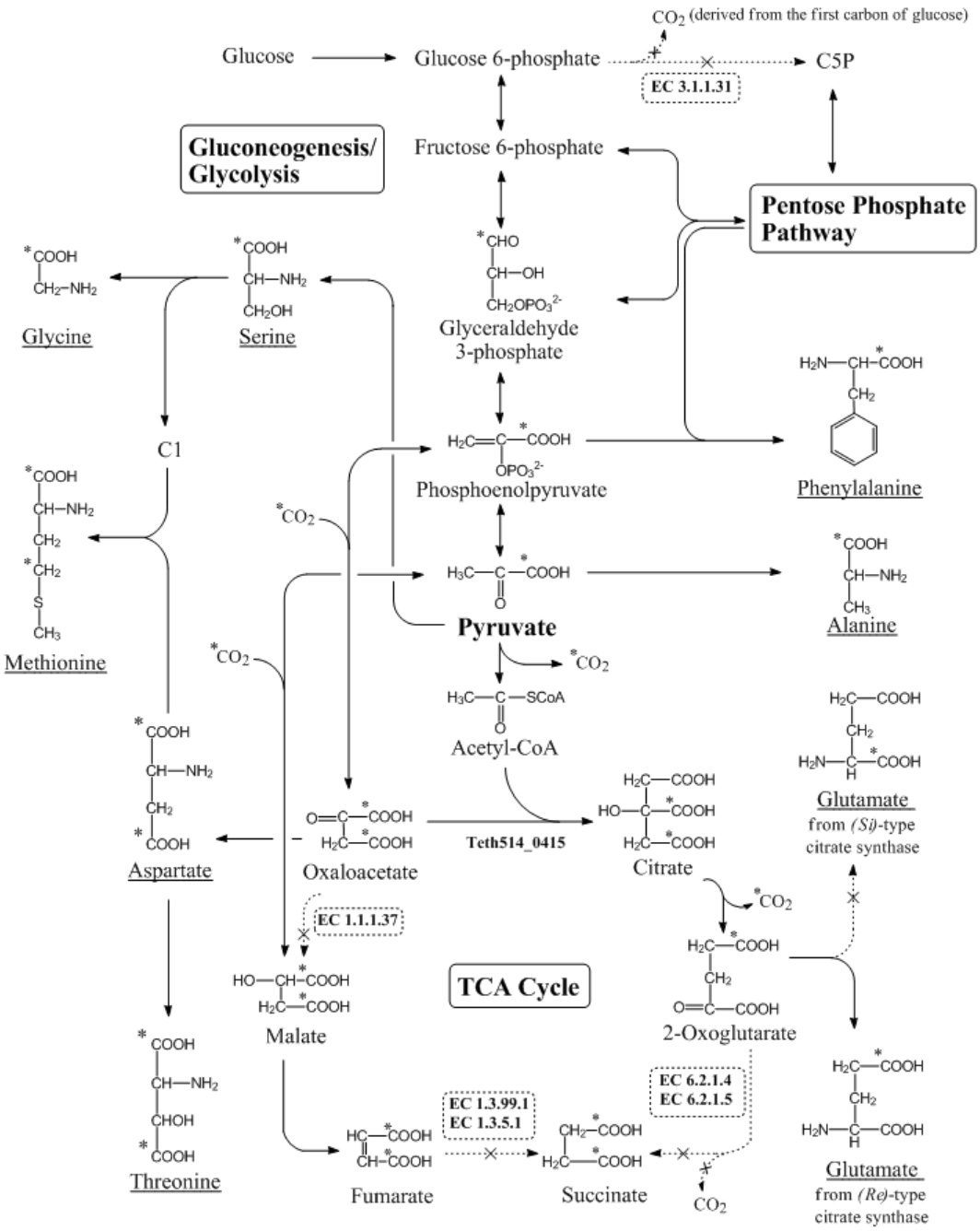

CO_2 (derived from the first carbon of glucose)
Glucose
Glucose 6-phosphate
C5P
EC 3.1.1.31
Gluconeogenesis/ Glycolysis
Fructose 6-phosphate
Pentose Phosphate Pathway
Glycine
Serine
Glyceraldehyde 3-phosphate
C1
Phosphoenolpyruvate
Phenylalanine
Methionine
Pyruvate
Alanine
Acetyl-CoA
Aspartate
Oxaloacetate
Teth514_0415
Citrate
Glutamate from (Si)-type citrate synthase
EC 1.1.1.37
Malate
TCA Cycle
2-Oxoglutarate
Threonine
EC 1.3.99.1
EC 1.3.5.1
EC 6.2.1.4
EC 6.2.1.5
Fumarate
Succinate
Glutamate from (Re)-type citrate synthase

Figure 3.3.3. Proposed scheme of isoleucine biosynthesis in *Thermoanaerobacter* sp. X514 (using 1st-position labeled pyruvate as the carbon source). ^{13}C labeled positions are marked with asterisks. The inactive pathway is marked by dashed line.

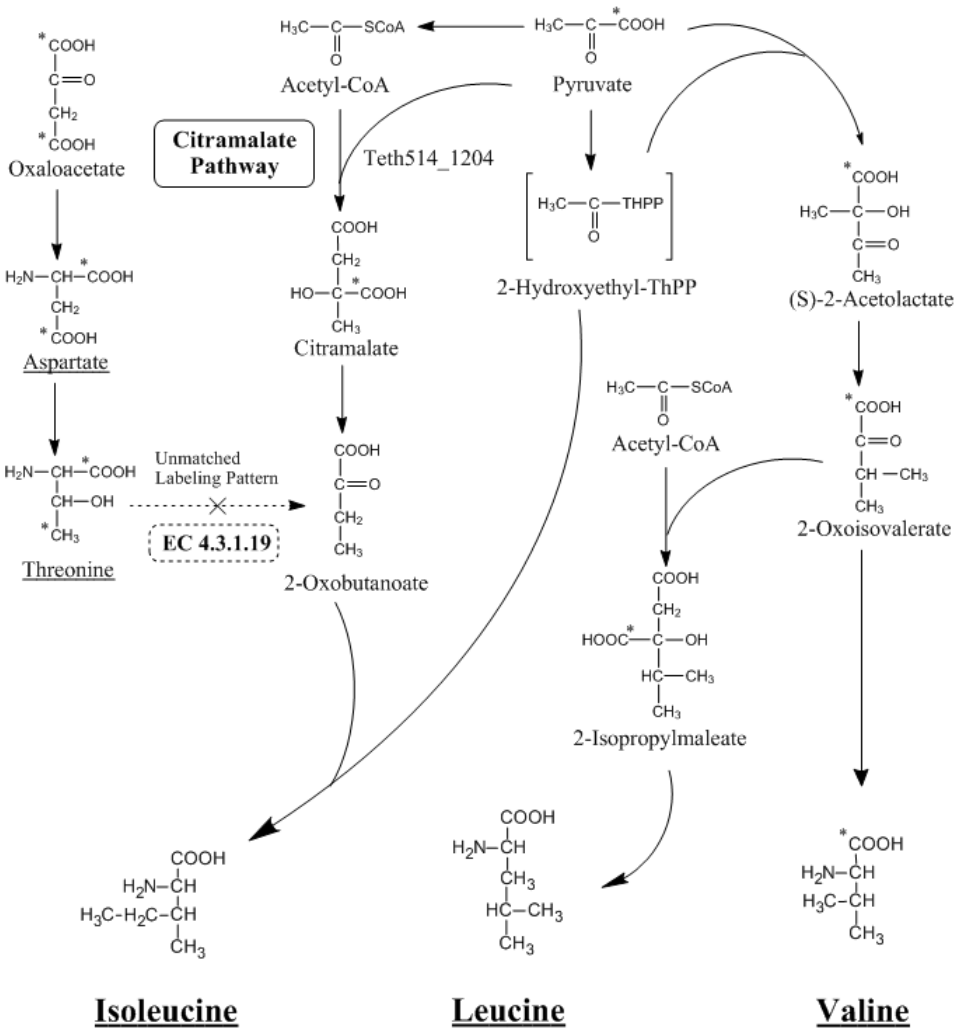


Figure 3.3.S1. Separation and detection of malate, citrate and citramalate in *Thermoanaerobacter* sp. X514 sample by LC-MS/MS in MRM mode. The multiple MRM transitions used for confident detection of the three target compounds in the sample are: 1) malate: 133/89, 133/71, 133/73, 133/43; 2) citrate: 191/129, 191/111, 191/87, 191/85; 3) citramalate: 147/87, 147/85, 147/57, 147/43, 147/41; 4) UN: unidentified compounds. The colors represent different transitions used in this method. Malate eluted at 5.6 min, citrate at 7.6 min and citramalate at 8.5 min. The peak at 9.0 min was an uncharacterized compound that shared the same MRM transitions as malate, but with different ratios.

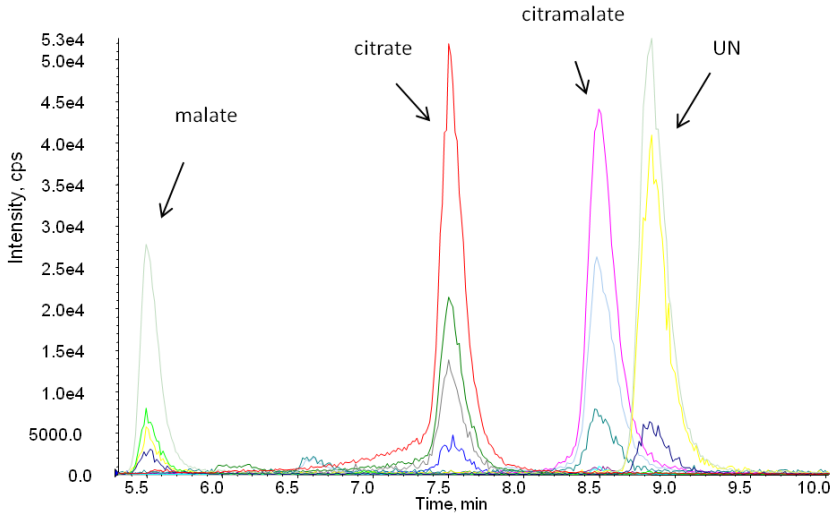


Figure 3.3.S2. Comparison of CID mass spectra of authentic standards with target compounds in sample X514. Left panel: CID mass spectra of authentic standards. Right panel: corresponding CID mass spectra in sample X514 after background subtraction to remove background noise.

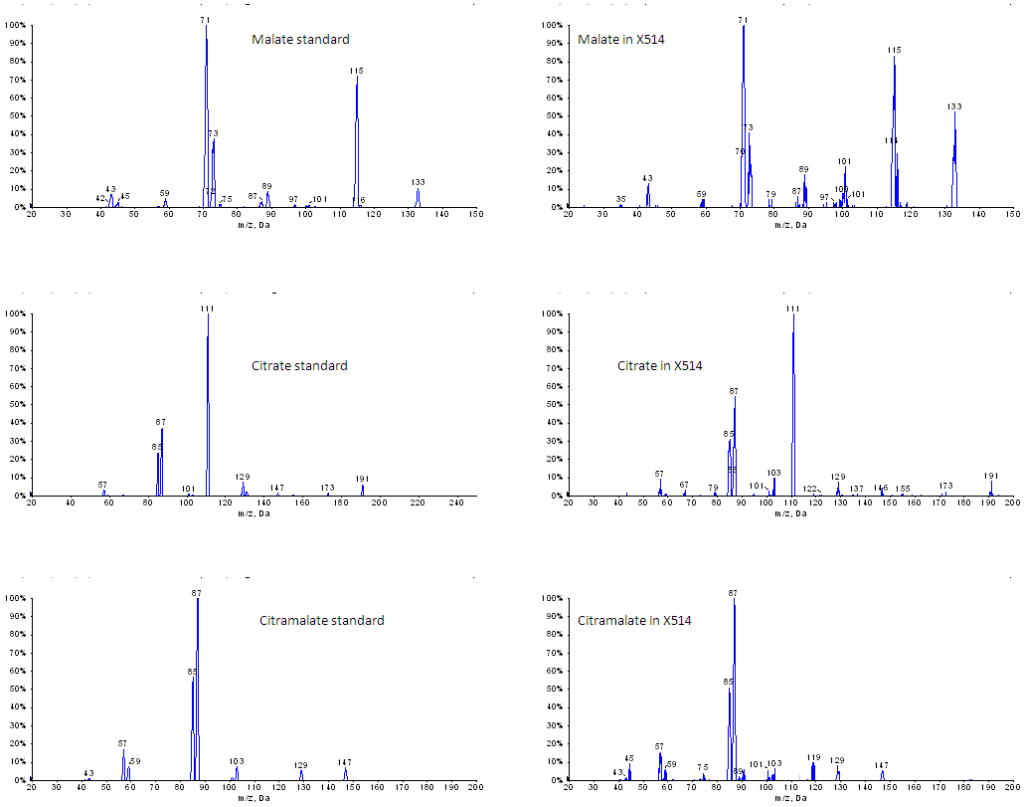


Figure 3.3.S3. The BLASTP (80) result of transhydrogenase PntA and UdhA in the genome of *Thermoanaerobacter* sp. X514. Neither gene was annotated.

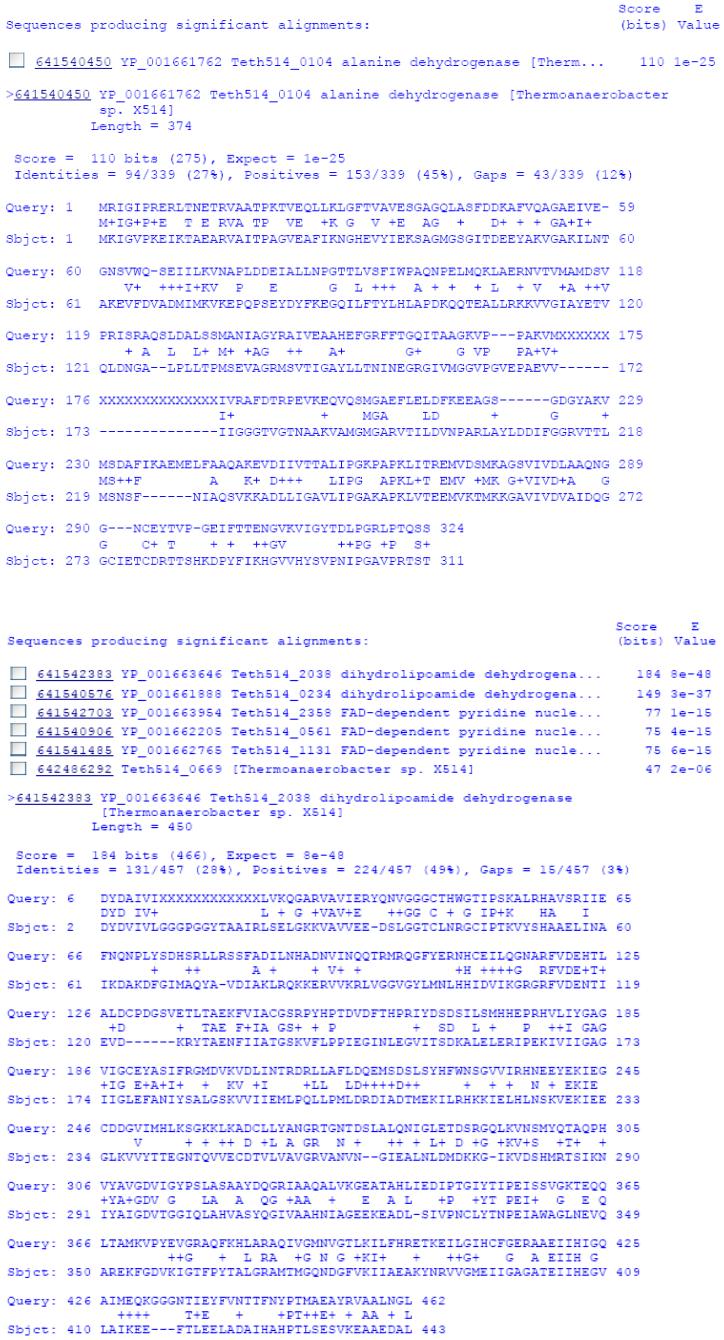
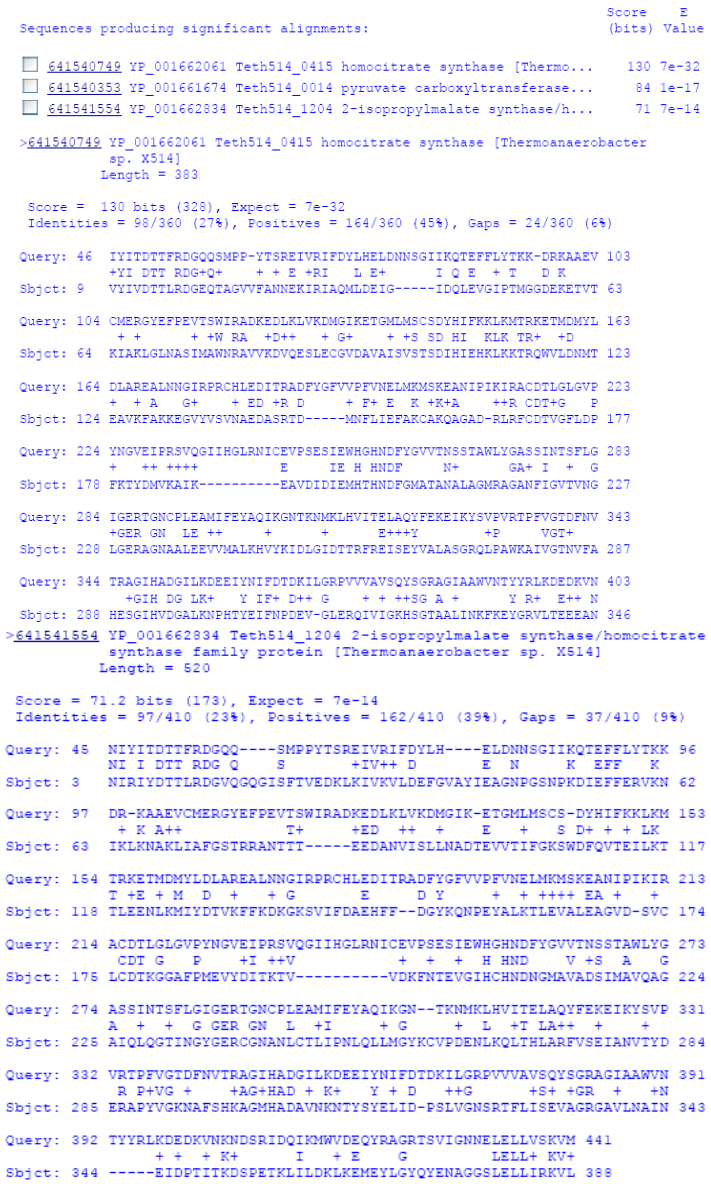


Figure 3.3.S4. Protein sequences of (*Re*)-type citrate synthase (CKL0973) were compared against the strain X514 genome (http://www.jgi.doe.gov) using BLAST search. The resulting polypeptide amino acid sequence identity was low: 27% for homocitrate synthase and 23% for 2-isopropylmalate synthase.



```
                                                                      Score    E
Sequences producing significant alignments:                           (bits) Value

 641540749 YP_001662061 Teth514_0415 homocitrate synthase [Thermo...    130 7e-32
 641540353 YP_001661674 Teth514_0014 pyruvate carboxyltransferase...     84 1e-17
 641541554 YP_001662834 Teth514_1204 2-isopropylmalate synthase/h...     71 7e-14

>641540749 YP_001662061 Teth514_0415 homocitrate synthase [Thermoanaerobacter
           sp. X514]
          Length = 383

 Score =  130 bits (328), Expect = 7e-32
 Identities = 98/360 (27%), Positives = 164/360 (45%), Gaps = 24/360 (6%)

Query: 46  IYITDTTFRDGQQSMPP-YTSREIVRIFDYLHELDNNSGIIKQTEFFLYTKK-DRKAAEV 103
           +YI DTT RDG+Q+    + + E +RI   L E+      I Q E  + T   D K
Sbjct: 9   VYIVDTTLRDGEQTAGVVFANNEKIRIAQMLDEIG-----IDQLEVGIPTMGGDEKETVT 63

Query: 104 CMERGYEFPEVTSWIRADKEDLKLVKDMGIKETGMLMSCSDYHIFKKLKMTRKETMDMYL 163
            + +      + +W RA  +D++   + G+    + +S SD HI  KLK TR+  +D
Sbjct: 64  KIAKLGLNASIMAWNRAVVKDVQESLECGVDAVAISVSTSDIHIEHKLKKTRQWVLDNMT 123

Query: 164 DLAREALNNGIRPRCHLEDITRADFYGFVVPFVNELMKMSKEANIPIKIRACDTLGLGVP 223
           +  + A   G+     + ED +R D     + F+ E  K +K+A    ++R CDT+G   P
Sbjct: 124 EAVKFAKKEGVYVSVNAEDASRTD-----MNFLIEFAKCAKQAGAD-RLRFCDTVGFLDP 177

Query: 224 YNGVEIPRSVQGIIHGLRNICEVPSESIEWHGHNDFYGVVTNSSTAWLYGASSINTSFLG 283
           +   ++ ++++          E     IE H HNDF     N+      GA+ I  +  G
Sbjct: 178 FKTYDMVKAIK----------EAVDIDIEMHTHNDFGMATANALAGMRAGANFIGVTVNG 227

Query: 284 IGERTGNCPLEAMIFEYAQIKGNTKNMKLHVITELAQYFEKEIKYSVPVRTPFVGTDFNV 343
           +GER GN  LE ++     +      E+++Y        +P     VGT+
Sbjct: 228 LGERAGNAALEEVVMALKHVYKIDLGIDTTRFREISEYVALASGRQLPAWKAIVGTNVFA 287

Query: 344 TRAGIHADGILKDEEIYNIFDTDKILGRPVVVAVSQYSGRAGIAAWVNTYYRLKDEDKVN 403
             +GIH DG LK+   Y IF+ D++ G    + + ++SG A +      Y R+  E++ N
Sbjct: 288 HESGIHVDGALKNPHTYEIFNPDEV-GLERQIVIGKHSGTAALINKFKEYGRVLTEEEAN 346
>641541554 YP_001662834 Teth514_1204 2-isopropylmalate synthase/homocitrate
           synthase family protein [Thermoanaerobacter sp. X514]
          Length = 520

 Score = 71.2 bits (173), Expect = 7e-14
 Identities = 97/410 (23%), Positives = 162/410 (39%), Gaps = 37/410 (9%)

Query: 45  NIYITDTTFRDGQQ----SMPPYTSREIVRIFDYLH----ELDNNSGIIKQTEFFLYTKK 96
           NI I DTT RDG Q     S       +IV++ D       E  N      K  EFF   K
Sbjct: 3   NIRIYDTTLRDGVQGQGISFTVEDKLKIVKVLDEFGVAYIEAGNPGSNPKDIEFFERVKN 62

Query: 97  DR-KAAEVCMERGYEFPEVTSWIRADKEDLKLVKDMGIK-ETGMLMSCS-DYHIFKKLKM 153
            + K A++           T+     +ED  ++  +    E   +    S D+ + + LK
Sbjct: 63  IKLKNAKLIAFGSTRRANTTT-----EEDANVISLLNADTEVVTIFGKSWDFQVTEILKT 117

Query: 154 TRKETMDMYLDLAREALNNGIRPRCHLEDITRADFYGFVVPFVNELMKMSKEANIPIKIR 213
           T +E + M   D  +   + G       E     D Y     +  + ++++ EA +    +
Sbjct: 118 TLEENLKMIYDTVKFFKDKGKSVIFDAEHFF--DGYKQNPEYALKTLEVALEAGVD-SVC 174

Query: 214 ACDTLGLGVPYNGVEIPRSVQGIIHGLRNICEVPSESIEWHGHNDFYGVVTNSSTAWLYG 273
            CDT G    P     +I ++V           +  +  +  H HND    V +S  A   G
Sbjct: 175 LCDTKGGAFPMEVYDITKTV----------VDKFNTEVGIHCHNDNGMAVADSIMAVQAG 224

Query: 274 ASSINTSFLGIGERTGNCPLEAMIFEYAQIKGN--TKNMKLHVITELAQYFEKEIKYSVP 331
           A  +  +  G GER GN  L  +I     + G     +  L  +T LA++  +    +  +
Sbjct: 225 AIQLQGTINGYGERCGNANLCTLIPNLQLLMGYKCVPDENLKQLTHLARFVSEIANVTYD 284

Query: 332 VRTPFVGTDFNVTRAGIHADGILKDEEIYNIFDTDKILGRPVVVAVSQYSGRAGIAAWVN 391
            R P+VG +    +AG+HAD + K+   Y + D   ++G      +S+ +GR  +   +N
Sbjct: 285 ERAPYVGKNAFSHKAGMHADAVNKNTYSYELID-PSLVGNSRTFLISEVAGRGAVLNAIN 343

Query: 392 TYYRLKDEDKVNKNDSRIDQIKMWVDEQYRAGRTSVIGNNELELLVSKVM 441
           T  + +  + K+       I   + E     G          LELL+ KV+
Sbjct: 344 TLGEIDPTITKDSPETKLILDKLKEMEYLGYQYENAGGSLELLIRKVL 388
```

Table 3.3.1. BLAST searches (May, 2009) for the key genes in an alternate isoleucine synthesis pathway (citramalate synthase, GSU 1798), for (*Si*)-type citrate synthase (EC 2.3.3.1 from *E.coli* K12) and for (*Re*)-type citrate synthase (CKL 0973 from *Clostridium kluyveri*) in the Joint Genome Institute (http://img.jgi.doe.gov).

Species	EC 2.3.3.1 *(Si)*-type citrate synthase	EC 2.3.3.3 (CKL 0973) *(Re)*-type citrate synthase	EC 2.3.1.182 GSU1798
**Thermoanaerobacter* sp. X514 (B)[F]	0%	27%	49%
Anaerocellum thermophilum DSM 6725 (B)[D]	0%	68%	51%
Caldicellulosiruptor saccharolyticus DSM 8903 (B)[F]	0%	68%	51%
Caldivirga maquilingensis IC-167 (A)[F]	39%	41%	47%
Candidatus Kuenenia stuttgartiensis (B)[F]	0%	55%	54%
Clostridium acetobutylicum ATCC 824 (B)[F]	0%	65%	50%
Clostridium botulinum A ATCC 3502 (B)[F]	0%	73%	22%
Clostridium kluyveri DSM 555 (B)[F]	28%	100%	29%
Clostridium cellulolyticum H10 (B)[F]	30%	60%	50%
Clostridium thermocellum ATCC 27405 (B)[F]	30%	60%	51%
Clostridium phytofermentans ISDg (B)[F]	30%	62%	25%
Desulfovibrio desulfuricans G20 (B)[F]	0%	48%	55%
Desulfovibrio vulgaris Hildenborough (B)[F]	0%	48%	55%
Desulfovibrio vulgaris DP4 (B)[F]	0%	48%	55%
Desulfovibrio vulgaris Miyazaki F (B)[F]	0%	49%	55%
**Dehalococcoides ethenogenes* 195 (B)[F]	0%	28%	53%
Syntrophus aciditrophicus SB (B)[F]	0%	49%	55%
Syntrophobacter fumaroxidans MPOB (B)[F]	32%	50%	59%
**Pelotomaculum thermopropionicum* SI (B)[F]	0%	29%	56%
Ignicoccus hospitalis KIN4/I (A)[F]	0%	50%	45%

Note:

1. Domains: (B)acteria, (A)rchaea.
2. Genome Completion: [F]inished, [D]raft.
3. Both (*Si*)- and (*Re*)-type citrate synthases in the species (with "*" in front) may be missing due to the low identity (<30%) of polypeptide amino acid sequences to the documented (*Si*)- and (*Re*)-type citrate synthases.

Table 3.3.S1. Isotopic analysis of amino acids in *Thermoanaerobacter* X514

Amino Acids	Precursors	Ions	$[M-57]^+$ $[1-^{13}C]/[6-^{13}C]$	$[M-159]^+$ $[1-^{13}C]/[6-^{13}C]$	$[f302]^+$ $[1-^{13}C]/[6-^{13}C]$
Ala	Pyruvate	M0	0.51/0.49	0.51/0.49	ND
		M1	0.47/0.50	0.45/0.47	
		M2	0.01/0.01	0.04/0.03	
Gly	Serine	M0	0.95/0.96	0.97/0.97	
		M1	0.05/0.03	0.03/0.03	ND
		M2	0/0		
Val	Pyruvate	M0	0.26/0.24	0.26/0.24	0.90/0.91
		M1	0.49/0.49	0.49/0.48	0.08/0.06
		M2	0.24/0.26	0.24/0.26	0.02/0.03
Leu	Pyruvate Acetyl-CoA	M0	Overlap Peak with f302	0.14/0.12	
		M1		0.38/0.37	ND
		M2		0.36/0.37	
		M3		0.12/0.13	
Ile	Pyruvate Acetyl-CoA	M0	Overlap Peak with f302	0.18/0.17	
		M1		0.41/0.41	ND
		M2		0.32/0.33	
		M3		0.08/0.09	
Met	Aspartate Methy-THF	M0	0.35/0.32	0.35/0.36	
		M1	0.48/0.51	0.49/0.48	ND
		M2	0.16/0.16	0.15/0.14	
		M3	0/0.01	0.01/0.02	
Ser	G3P	M0	0.54/0.52	0.55/0.53	0.96/0.99
		M1	0.44/0.48	0.45/0.47	0.04/0
		M2	0.02/0	0.01/0	0/0.01
Thr	Aspartate	M0	0.51/0.47	0.52/0.46	
		M1	0.47/0.52	0.48/0.51	
		M2	0.02/0	0/0.02	
Phe	PEP E4P	M0	0.19/0.05	0.20/0.06	0.97/0.98
		M1	0.42/0.29	0.42/0.29	0.03/0.02
		M2	0.30/0.44	0.30/0.43	0/0
		M3	0.08/0.21	0.08/0.21	
		M4	0.01/0.01	0/0.01	
Asp/Asn	OAA	M0	0.50/0.47	0.50/0.48	0.97/0.98

		M1	0.48/0.52	0.48/0.49	0.03/0.02
		M2	0.02/0.01	0.01/0.02	0/0
Glu/Gln	OXO	M0	0.26/0.24	0.27/0.24	ND
		M1	0.49/0.49	0.49/0.49	
		M2	0.25/0.26	0.23/0.26	
		M3	0.01/0.01	0/0	
Tyr	PEP E4P	M0	0.19/0.05	0.20/0.08	0.97/0.97
		M1	0.42/0.32	0.41/0.31	0.03/0.03
		M2	0.30/0.41	0.310.38	0/0
		M3	0.08/0.21	0.08/0.21	
		M4	0.01/0.01	0.01/0.02	

Notes for Table 3.3.S1 ~ 2:

1. Ion mass values represent amino acid molecules with specific fragmentation patterns: ($[M-57]^+$: No loss); ($[M-159]^+$: Loss of α carboxyl group); ($[f302]^+$: 1st and 2nd carbons in amino acids). ($[M-159]^+$: Loss of α carboxyl group) of threonine was replaced by ($[M-85]^+$: Loss of α carboxyl group) because the later ion had clearer signals.
2. $[f302]^+$ peaks of some amino acids overlapped with other peaks, so $[f302]^+$ only qualitatively reflects the labeling status in amino acids.
3. Asparagine and glutamate were converted into aspartate and glutamine during the protein hydrolysis.
4. Abbreviations: E4P, erythrose-4-phosphate; OAA, oxaloacetate; OXO, 2-ketoglutarate; PEP, phosphoenolpyruvate; G3P, 3-phosphoglycerate.
5. The standard deviations for measurement (n=2) of mass fractions are below 2%.

Table 3.3.S2. Isotopic analysis of amino acids in *Thermoanaerobacter* sp. X514 and their labeling positions ([1-^{13}C] pyruvate)

Amino Acids	Precursors	Ions	$[M-57]^+$	$[M-159]^+$	$[f302]^+$	Proposed ^{13}C enriched positions
Ala	Pyruvate	M0	0.02	0.93	0.03	C-C-*COOH
		M1	0.96	0.02	0.95	
		M2	0.02	0.04	0.02	
Gly	Serine	M0	0.03	0.99	ND	C-*COOH
		M1	0.96	0.01		
		M2	0.01			
Val	Pyruvate	M0	0.01	0.95	0.12	C-C-C-C-*COOH
		M1	0.94	0.04	0.87	
		M2	0.04	0	0.01	
Leu	Pyruvate Acetyl-CoA	M0	Overlap Peak with f302	0.93	ND	C-C-C-C-C-COOH C2~C6 were not labeled.
		M1		0.06		
		M2		0.01		
		M3		0.00		
Ile	Pyruvate Acetyl-CoA	M0	Overlap Peak with f302	0.91	ND	C-C-C-C-C-COOH C2 ~C6 were not labeled
		M1		0.08		
		M2		0.01		
		M3		0		
Met	Aspartate Methy-THF	M0	0	0.03	ND	C-S-*C-C-C-*COOH
		M1	0.03	0.93		
		M2	0.92	0.03		
		M3	0.04	0		
Ser	G3P	M0	0.02	0.97	0.03	C-C-*COOH
		M1	0.96	0.02	0.95	
		M2	0.01	0	0.02	
Thr	Aspartate	M0	0.01	0.03	ND	*C-C-C-*COOH
		M1	0.02	0.95		
		M2	0.93	0		
Phe	PEP E4P	M0	0.01	0.01	0.02	C-C-C-C-C-C-C-C-*COOH 1st carbon was labeled. The other two carbons' labeling positions could not be determined.
		M1	0.01	0.37	0.97	
		M2	0.38	0.56	0.01	
		M3	0.56	0.05		
		M4	0.04	0.01		

Asp/Asn	OAA	M0	0	0.02	0.02	*C-C-C-*COOH
		M1	0.03	0.95	0.98	
		M2	0.96	0.02	0	
Glu/Gln	OXO	M0	0.02	0.02	0.82	*C-C-C-C-COOH
		M1	0.93	0.94	0.18	
		M2	0.05	0.03	0	
		M3	0	0		
Tyr	PEP E4P	M0	0	0.02	0.02	C-C-C-C-C-C-C-C-*COOH 1st carbon was labeled. The other two carbons' labeling positions could not be determined
		M1	0.02	0.37	0.98	
		M2	0.37	0.54	0	
		M3	0.56	0.05		
		M4	0.04	0.01		

3.4 ^{13}C-pathway analysis of environmental microorganisms

^{13}C-pathway analysis has been applied to a number of environmental microorganisms, including 1) *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium (81); 2) *Heliobacterium modesticaldum*, a Gram-positive photosynthetic bacterium (82); 3) *Roseobacter denitrificans* OCh114, an aerobic anoxygenic phototrophic bacterium (83); 4) *Dehalococcoides ethenogenes* Strain 195, an important member of bioremediation communities to detoxify chloroethenes into the benign end product ethane (84); and 5) *Mycobacterium smegmatis*, a non-virulent bacterium that is often used to study the metabolism of a pathogenic bacterium, *Mycobacterium tuberculosis* (85). Novel metabolic features have been uncovered for the aforementioned environmental microorganisms via ^{13}C-pathway analysis. Diverse analytical methods were utilized to confirm the discoveries from ^{13}C-pathway analysis, including biochemical assays to detect *in vitr*o enzyme activities, quantitative RT-PCR to profile the gene expressions, and LC-MS/MS to measure fast-turnover metabolites.

3.4.1 *Cyanothece* 51142

Based on the identical labeling patterns of isoleucine and leucine in [2-^{13}C] glycerol cultures, an alternative isoleucine biosynthesis pathway via citramalate synthase was discovered in *Cyanothec*e 51142. Instead of using threonine as the precursor, *Cyanothec*e 51142 applied the citramalate pathway to

induce isoleucine biosynthesis by using the same precursors (i.e., pyruvate and acetyl-CoA) as those used in leucine biosynthesis pathway. The presence of citramalate pathway was confirmed by measuring the *in vitro* enzymatic activity of citramalate synthase via biochemical assays and by detecting the key intermediate in citramalate pathway, citramalte, via LC-MS/MS. A report about applications of ^{13}C-pathway analysis to identify citramalate pathway in *Cyanothece* 51142 is attached in Appendix 2.

3.4.2 *Heliobacterium modesticaldum*

Heliobacteria are the only cultured Gram-positive photosynthetic bacteria. The isotopomer data from proteinogenic amino acids was collected to probe the central carbon metabolism of *Heliobacterium modesticaldum*. The CO_2-anaplerotic pathway was found to be active during phototrophic growth and the citramalate pathway was employed by *Heliobacterium modesticaldum* for isoleucine biosynthesis. Furthermore, the oxidative TCA cycle was operative by using a putative (Re)-citrate synthase to produce citrate. The presence of (Re)-citrate synthase was confirmed by measuring activity of citrate synthase via an activity assay. The report of ^{13}C-pathway analysis for *Heliobacterium modesticaldum* is attached in Appendix 3.

3.4.3 *Roseobacter denitrificans* OCh114

Roseobacter denitrificans OCh114 is a model organism to study aerobic anoxygenic photosynthesis in bacteria. Through measurements of ^{13}C-isotopomer labeling patterns of proteinogenic amino acids in a series of tracer experiments, the metabolic routes for carbohydrate utilization, CO_2 assimilation, and amino acid biosynthesis were studied in *Roseobacter denitrifican*s OCh114. The anaplerotic pathways, mainly via the malic enzyme, were used to fix CO_2 by *R. denitrificans*. The Entner-Doudoroff (ED) pathway and the non-oxidative pentose phosphate pathway were employed by *R. denitrificans* in carbohydrate metabolism. The Embden-Meyerhof-Parnas (EMP, glycolysis) pathway was found to be inactive, which was confirmed by the absence of *in vitro* enzyme activity of 6-phosphofructokinase (PFK). In addition, isoleucine synthesis in *R. denitrificans* used both threonine-dependent (20% total flux) and citramalate pathway (80% total flux). The report of ^{13}C-pathway analysis for *Roseobacter denitrificans* OCh114 is attached in Appendix 4.

3.4.4 *Dehalococcoides ethenogenes* Strain 195

Dehalococcoides ethenogenes 195 is the only known bacteria that can fully degrade PCE to ethane (86,87). Isotopomer-based dilution analysis was applied to differentiate the amino acids in *Dehalococcoides ethenogenes* from two pathways: *de novo* synthesis from acetate, and amino acids import from environment. It was found that glutamate/glutamine and aspartate/asparagine were almost exclusively synthesized by *Dehalococcoides ethenogenes*, even when

provided in excess in the medium. In contrast, phenylalanine, isoleucine, leucine, and methionine were identified as the four amino acids that were most highly imported from environment. Adding either phenylalanine or the four highly imported amino acids to the defined mineral medium enhanced the growth rates, dechlorination activities, and yields of strain 195 in a similar level to supplementation with 20 amino acids. The transcriptional regulations of ABC-type amino acids transporters were also analyzed by qRT-PCR. The report of ^{13}C-pathway analysis for *Dehalococcoides ethenogenes* Strain 195 is attached in Appendix 5.

3.4.5 *Mycobacterium smegmatis*

Mycobacterium smegmatis is phylogenetically related to *Mycobacterium tuberculosis* (MTB) and is often used as a model for studying MTB metabolism since it is non-virulent. Isotopomer-assisted metabolite analysis was used to investigate the metabolic transition from normal growth to a non-replicating state under a hypoxic environment. The glyoxylate pathway and glycine dehydrogenase were induced as the bacillus encountered hypoxic stress. Meanwhile, the relative amount of acetyl-CoA entering the TCA cycle was doubled, whereas little entered the glycolytic and pentose phosphate pathways. The report of ^{13}C-pathway analysis for *Mycobacterium smegmatis* is attached in Appendix 6.

3.5 References

1. Bottomley PJ, Baalen CV (1978) Characteristics of Heterotrophic Growth in the Blue-Green Alga Nostoc sp. Strain Mac. Journal of General Microbiology 107: 309-318.

2. Eiler A (2006) Evidence for the Ubiquity of Mixotrophic Bacteria in the Upper Ocean: Implications and Consequences. Applied and Environmental Microbiology 72: 7431-7437.

3. Yang C, Hua Q, Shimizu K (2002) Metabolic flux analysis in *Synechocystis* using isotope distribution from ^{13}C-labeled glucose. Metab Eng 4: 202-216.

4. Dutta D, De D, Chaudhuri S, Bhattacharya SK (2005) Hydrogen production by Cyanobacteria. Microb Cell Fact 4.

5. Madamwar D, Garg N, Shah V (2000) Cyanobacterial hydrogen production. World J Microbiol Biotechnol 16: 757-767.

6. Bernat G, Waschewski N, Rogner M (2009) Towards efficient hydrogen production: the impact of antenna size and external factors on electron transport dynamics in Synechocystis PCC 6803. Photosynth Res 99: 205-216.

7. Tamagnini P, Leitao E, Oliveira P, Ferreira D, Pinto F, et al. (2007) Cyanobacterial hydrogenases: diversity, regulation and applications. Fems Microbiology Reviews 31: 692-720.

8. Tuli R, Naithani S, Misra HS (1996) Cyanobacterial photosynthesis and the problem of oxygen in nitrogen-fixation: A molecular genetic view. Journal of Scientific & Industrial Research 55: 638-657.

9. Fay P (1992) Oxygen relations of nitrogen fixation in cyanobacteria. Microbiological Reviews 56: 340-373.

10. Toepel J, Welsh E, Summerfield TC, Pakrasi HB, Sherman LA (2008) Differential transcriptional analysis of the cyanobacterium Cyanothece sp strain ATCC 51142 during light-dark and continuous-light growth. J Bacteriol 190: 3904-3913.

11. Stöckel J, Welsh EA, Liberton M, Kunnvakkam R, Aurora R, et al. (2008) Global transcriptomic analysis of Cyanothece 51142 reveals robust diurnal oscillation of central metabolic processes. Proc Natl Acad Sci 105: 6456-6461.

12. Welsh EA, Liberton M, Stöckel J, Loh T, Elvitigala T, et al. (2008) The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium

important in the marine nitrogen cycle. Proc Natl Acad Sci USA 105: 15094-15099.

13. Benemann JR, Weare NM (1974) Hydrogen Evolution by Nitrogen-Fixing Anabaena cylindrica Cultures. Science 184: 174-175.

14. Huang TC, Chow. TJ (1986) New type of N2-fixing unicellular cyanobacterium (blue-green alga). FEMS Microbiology Letter 36: 109-110.

15. ColonLopez M, Sherman DM, Sherman LA (1997) Transcriptional and translational regulation of nitrogenase in light-dark- and continuous-light grown cultures of the unicellular cyanobacterium Cyanothece sp. strain ATCC 51142. Journal of Bacteriology 179: 4319-4327.

16. Reddy KJ, Haskell JB, Sherman DM, Sherman LA (1993) Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. J Bacteriol 175: 1284-1292.

17. Neuer G, Bothe H (1985) Electron donation to nitrogenase in heterocysts of cyanobacteria. Archives of Microbiology 143: 185-191.

18. Dutta D, De D, Chaudhuri S, Bhattacharya SK (2007) Hydrogen production by Cyanobacteria. OMICS 11: 166-185.

19. Berman-Frank I, Lundgren P, Falkowski P (2003) Nitrogen fixation and photosynthetic oxygen evolution in cyanobacteria. Research in Microbiology 154: 157-164.

20. Fong SS, Nanchen A, Palsson BO, Sauer U (2006) Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. Journal of Biological Chemistry 281: 8024-8033.

21. Sauer U (2004) High-throughput phenomics: experimental methods for mapping fluxomes. Current Opinion in Biotechnology 15: 58-63.

22. Tang YJ, Pingitore F, Mukhopadhyay A, Phan R, Hazen TC, et al. (2007) Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using GC-MS and FT-ICR mass spectrometry. Journal of Bacteriology 189: 940-949.

23. Wu B, Zhang B, Feng X, Rubens JR, Huang R, et al. (2010) An Alternate Isoleucine Biosynthesis Pathway involves Citramalate Synthase in Cyanothece sp. ATCC 51142. Microbiology 156: 596-602.

24. Pingitore F, Tang YJ, Kruppa GH, Keasling JD (2007) Analysis of amino acid isotopomers using FT-ICR MS. Analytical Chemistry 79: 2483-2490.

25. Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EEK, et al. (2009) Advances in metabolic network and flux analysis of microorganisms via 13C isotopic labeling. Mass Spectrometry Reviews 28: 362-375.

26. Stanier RY, Kunisawa R, Mandel M, Cohen-Bazire G (1971) Purification and properties of unicellular blue-green algae (order Chroococcales). Bacteriol Rev 35: 171-205.

27. Rey FE, Heiniger EK, Harwood CS (2007) Redirection of Metabolism for Biological Hydrogen Production. Applied and Environmental Microbiology 73: 1665-1671.

28. Roose JL, Pakrasi HB (2004) Evidence that D1 processing is required for manganese binding and extrinsic protein assembly into photosystem II. J Biol Chem 279: 45417-45422.

29. Krause GH, Weis E (1991) Chlorophyll fluorescence and photosynthesis: the basics. Ann Rev Plant Physiol Plant Mol Biol 42: 313-349.

30. Tang YJ, Hwang JS, Wemmer D, Keasling JD (2007) The *Shewanella oneidensis* MR-1 fluxome under various oxygen conditions. Applied and Environmental Microbiology 73: 718-729.

31. Tang YJ, Meadows AL, Kirby J, Keasling JD (2007) Anaerobic central metabolic pathways in *Shewanella oneidensis* MR-1 reinterpreted in the light of isotopic metabolite labeling. Journal of Bacteriology 189: 894-901.

32. Wahl SA, Dauner M, Wiechert W (2004) New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnology and Bioengineering 85: 259-268.

33. Drath M, Kloft N, Batschauer A, Marin K, Novak J, et al. (2008) Ammonia triggers photodamage of photosystem II in the cyanobacterium Synechocystis sp strain PCC 6803. Plant Physiology 147: 206-215.

34. Dai G, Deblois CP, Liu S, Juneau P, Qiu B (2008) Differential sensitivity of five cyanobacterial strains to ammonium toxicity and its inhibitory mechanism on the photosynthesis of rice-field cyanobacterium Ge-Xian-Mi (Nostoc). Aquatic toxicology 89: 113-121.

35. Baalen CV, Hoare DS, Brandt E (1971) Heterotrophic growth of blue-green

algae in dim light. Journal of Bacteriology 105: 685-689.

36. Anderson SL, McIntosh L (1991) Light-activated heterotrophic growth of the cyanobacterium Synechocystis sp. strain PCC 6803: a blue-light-requiring process. Journal of Bacteriology 173: 2761-2767.

37. Flores E, Schmetterer G (1986) Interaction of fructose with the glucose permease of the cyanobacterium Synechocystis sp. strain PCC 6803. Journal of Bacteriology 66: 693-696.

38. Schmetterer GR (1990) Sequence conservation among the glucose transporter from the cyanobacterium Synechocystis sp. PCC 6803 and mammalian glucose transporters. Plant Molecular Biology 14: 697-706.

39. Postgate J, editor (1998) Nitrogen Fixation, 3rd Edition: Cambridge University Press, Cambridge UK.

40. Melnickia MR, Bianchib L, Philippisb RD, Melis A (2008) Hydrogen production during stationary phase in purple photosynthetic bacteria. International Journal of Hydrogen Energy </science/journal/03603199> 33: 6525-6534.

41. Nyström T (2004) Stationary phase physiology. Annual Review of Microbiology 58: 161-181.

42. Atsumi S, Higashide W, Liao JC (2009) Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nature Biotechnology 27: 1177-U1142.

43. Slack CR, Hatch MD (1967) Comparative studies on the activity of carboxylases and other enzymes in relation to the new pathway of photosynthetic carbon dioxide fixation in tropical grasses. Biochem J 103: 660-665.

44. Pirintsos SA, Munzi S, Loppi S, Kotzabasis K (2009) Do polyamines alter the sensitivity of lichens to nitrogen stress? Ecotoxicology and Environmental Safety 72: 1331-1336.

45. Ting CS, Owens TG (1992,) Limitations of the pulse-modulated technique for measuring the fluorescence characteristics of algae. Plant Physiol 100: 367-373.

46. Schreiber U, Endo T, Mi H, Asada K (1995) Quenching Analysis of Chlorophyll Fluorescence by the Saturation Pulse Method: Particular Aspects Relating to the Study of Eukaryotic Algae and Cyanobacteria. Plant and Cell Physiology 36: 873-882.

47. Kaftan D, Meszaros T, Whitmarsh J, Nedbal L (1999) Characterization of Photosystem II Activity and Heterogeneity during the Cell Cycle of the Green Alga Scenedesmus quadricauda. Plant Physiology 120: 433-441.

48. Rawson DM (1985) The effects of exogenous amino acids on growth and nitrogenase activity in the cyanobacterium Anabaena cylindrica PCC 7122. Journal of General Microbiology 134: 2549-2544.

49. Muro-Pastor MI, Reyes JC, Florencio FJ (2005) Ammonium assimilation in cyanobacteria. Photosynthesis Research 83: 135-150.

50. Galmozzi CV, Fernandez-Avila MJ, Reyes JC, Florencio FJ, Muro-Pastor MI (2007) The ammonium-inactivated cyanobacterial glutamine synthetase I is reactivated in vivo by a mechanism involving proteolytic removal of its inactivating factors. Mol Microbiol 65: 166-179.

51. Berman-Frank I, Lundgren P, Chen Y-B, Küpper H, Kolber Z, et al. (2001) Segregation of Nitrogen Fixation and Oxygenic Photosynthesis in the Marine Cyanobacterium Trichodesmium. Science 294: 1534-1537.

52. Elvitigala T, Stöckel J, Ghosh BK, Pakrasi HB (2009) Effect of continuous light on diurnal rhythms in Cyanothece sp. ATCC 51142. BMC Genomics 10: 226.

53. Herrera. S (2006) Bonkers about biofuels. Nature Biotechnology 24: 755-760.

54. Schubert C (2006) Can biofuels finally take center stage? Nature Biotechnology 24: 777-784.

55. van Zyl WH, Lynd LR, den Haan R, McBride JE (2007) Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. Adv Biochem Eng Biotechnol 108: 205-235.

56. Lynd LR, van Zyl WH, McBride JE, Laser M (2005) Consolidated bioprocessing of cellulosic biomass: an update. . Current Opinion in Biotechnology 16: 577-583.

57. Lynd LR (1989) Production of Ethanol from Lignocellulosic Materials Using Thermophilic Bacteria: Critical Evaluation of Potential and Review; Fiechter A, editor. Heidelberg: Springer-Verlag. 1-52 p.

58. Onyenwoke RU, Kevbrin VV, Lysenko AM, Wiegel J (2007) *Thermoanaerobacter pseudethanolicus* sp. nov., a thermophilic heterotrophic anaerobe from Yellowstone National Park. Int J Syst Evol Microbiol 57: 2191-2193.

59. Shaw AJ, Podkaminer KK, Desai SG, Bardsley JS, Rogers SR, et al. (2008) Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. Proc Natl Acad Sci USA 105: 13769-13774

60. Demain AL, Newcomb M, Wu JH (2005) Cellulase, clostridia, and ethanol Microbiol Mol Biol Rev 69: 124-154.

61. Ng TK, Ben-Bassat A, Zeikus JG (1981) Ethanol Production by Thermophilic Bacteria: Fermentation of Cellulosic Substrates by Cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. . Appl Environ Microbiol 41: 1337-1343.

62. Alm EJ, Huang KH, Price MN, Koche RP, Keller K, et al. (2005) The MicrobesOnline Web site for comparative genomics. Genome Res **15**: 1015-1022.

63. Tringe SG, Rubin EM (2005) Metagenomics: DNA sequencing of environmental samples. Nat Rev Genet 6: 805-814.

64. Warnecke F, Luginbühl P, Ivanova N, Ghassemian M, Richardson TH, et al. (2007) Metagenomic and functional analysis of hindgut microbiota of a wood-feeding higher termite. Nature 450: 560-565.

65. Hua Q, Joyce AR, Palsson BO, Fong SS (2007) Metabolic characterization of *Escherichia coli* adapted to growth on lactate. Applied and Environmental Microbiology 73: 4639-4647.

66. Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EK, et al. (2009) Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. Mass Spectrometry Reviews 28: 362-375.

67. Tang YJ, Sapra R, Joyner D, Hazen TC, Myers S, et al. (2009) Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. Biotechnology and Bioengineering 102: 1377-1386.

68. Roh Y, Liu SV, Li G, Huang H, Phelps TJ, et al. (2002) Isolation and Characterization of Metal-Reducing *Thermoanaerobacter* Strains from Deep Subsurface Environments of the Piceance Basin, Colorado. Applied and Environmental Microbiology 68: 6013-6020.

69. Wolin EA, Wolin MJ, Wolfe RS (1963) Formation of Methane by Bacterial Extracts. Journal of Biological Chemistry 238: 2882-2886.

70. Howell D, Xu H, White RH (1998) (R)-Citramalate Synthase in Methanogenic Archaea. J Bact 181: 331-333.

71. Risso C, Van Dien SJ, Orloff A, Lovley DR, Coppi MV (2008) Elucidation of an alternate isoleucine biosynthesis pathway in *Geobacter sulfurreducens*. Journal of Bacteriology 190: 2266-2274.

72. Xu H, Zhang YZ, Guo XK, Ren SX, Staempfli AA, et al. (2004) Isoleucine biosynthesis in *Leptospira interrogans* serotype lai strain 56601 proceeds via a threonine-independent pathway Journal of Bacteriology 186: 5400-5409.

73. Antoniewicz MR, Kelleher JK, Stephanopoulos G (2007) Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. Analytical Chemistry 79: 7554-7559.

74. 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*. Journal of Biological Chemistry 279: 6613-6619.

75. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25: 3389-3402.

76. Lacis LS, Lawford HG (1991) Thermoanaerobacter ethanolicus Growth and Product Yield from Elevated Levels of Xylose or Glucose in Continuous Cultures. Applied and Environmental Microbiology 57: 579-585.

77. Ivanovsky RN, Krasilnikova EN, Berg IA (2006) A proposed citramalate cycle for acetate assimilation in the purple non-sulfur bacterium *Rhodospirillum rubrum*. FEMS Microbiology Letters 153: 399-404.

78. Atsumi S, Liao JC (2008) Directed evolution of *Methanococcus jannaschii* citramalate synthase for 1-propanol and 1-butanol biosynthesis from *Escherichia coli*. Applied and Environmental Microbiology 74: 7802-7808.

79. Li F, Hagemeier CH, Seedorf H, Gottschalk G, Thauer RK (2007) Re-citrate synthase from *Clostridium kluyveri* is phylogenetically related to homocitrate synthase and isopropylmalate synthase rather than to Si-citrate synthase. Journal of Bacteriology 189: 4299-4304.

80. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389-3402.

81. Wu B, Zhang B, Feng X, Rubens JR, Huang R, et al. (2010) Alternative isoleucine synthesis pathway in cyanobacterial species. Microbiology 156:

596-602.

82. Tang K-H, Feng X, Zhuang W-Q, Alvarez-Cohen L, Blankenship RE, et al. (2010) Carbon flow of Heliobacterium modesticaldum is more related to Firmicutes than to the green sulfur bacteria. J Biol Chem 285: 35104-35112.

83. Tang K-H, Feng X, Tang YJ, Blankenship RE (2009) Carbohydrate metabolism and carbon fixation in Roseobacter denitrificans OCh114. PLoS One 4: e7233.

84. Zhuang WQ, Yi S, Feng X, Zinder SH, Tang YJ, et al. (2011) Selective utilization of exogenous amino acids by *Dehalococcoides ethenogenes* strain 195 and the enhancement resulted to dechloronation activity. Appl Environ Microbiol 77: 7797-7803.

85. Tang YJ, Shui WQ, Myers S, Feng X, Bertozzi C, et al. (2009) Central metabolism in Mycobacterium smegmatis during the transition from O2-rich to O2-poor conditions as studied by isotopomer-assisted metabolite analysis. Biotechnol Lett 31: 1233-1240.

86. Fennell DE, Nijenhuis I, Wilson SF, Zinder SH, Häggblom MM (2004) *Dehalococcoides ethenogenes* strain 195 reductively dechlorinates diverse chlorinated aromatic pollutants. Environmental Science and Technology 38: 2075-2081.

87. Maymó-Gatell X, Chien Y-t, Gossett J, Zinder S (1997) Isolation of a Bacterium That Reductively Dechlorinates Tetrachloroethene to Ethene Science 276: 1568-1571.

Chapter 4

^{13}C-Metabolic Flux Analysis of Central Carbon Metabolisms

4.1 ^{13}C-metabolic flux analysis (^{13}C-MFA) protocol

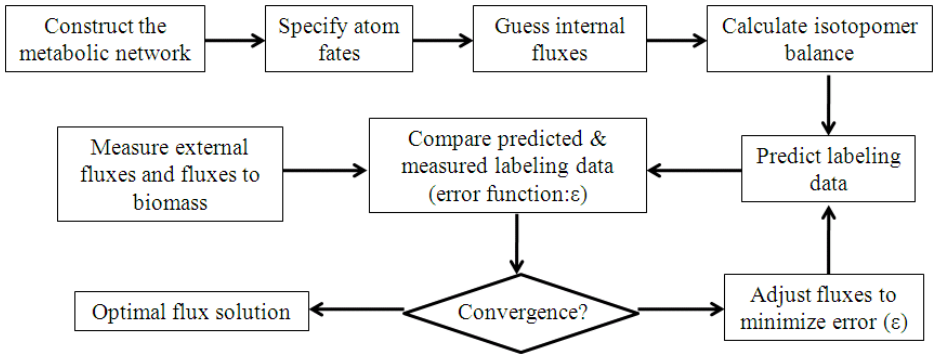
^{13}C-MFA integrates the isotopomer measurements and the computational optimization to quantify fluxes through a metabolic network. In general, ^{13}C-MFA is formulated as a nonlinear inverse problem, in which mass balance equations and isotopomer balance equations are used to find an optimal set of metabolic fluxes that has the smallest lack-of-fit between the simulated and measured isotopomer labeling patterns in proteinogenic amino acids (1). ^{13}C-MFA is usually applied to quantify the flux distributions in central metabolic network, including glycolysis, Entner–Doudoroff pathway, pentose phosphate pathway, the TCA cycle, futile pathways, and the glyoxylate shunt. At metabolic and isotopic steady state, labeling information of proteinogenic amino acids that are synthesized from these central metabolic pathways is collected for ^{13}C-MFA. The substrate uptake rates and products secretion rates are measured as the inflow and outflow fluxes in ^{13}C-MFA.

To quantify the intracellular metabolic fluxes, both mass and isotopomer balance equations are applied. At metabolic steady state, there is no net accumulation of the intracellular metabolites in the metabolic network. The mass balance in the given metabolic network is hence represented as $S \cdot v = 0$, where S

is the stoichiometric coefficients matrix and v is the flux distribution vector. For an intracellular metabolite synthesized from two reactant molecules at isotopic steady state, i.e. A+B→C, the isotopomer distribution of the product molecule can be represented as $IDV_C = (IMM_{A\rightarrow C} \cdot IDV_A) \otimes (IMM_{B\rightarrow C} \cdot IDV_B)$; where IDV is the isotopomer distribution vector for a metabolite, ⊗ represents element-wise multiplication, and $IMM_{A\rightarrow C}$ represents the isotopic mapping matrix that describes the isotopomer transitions from the precursor to the product. The isotopomer balance equations can be derived as $IDV \cdot \sum_{j=1}^{n} v_j = \sum_{j=1}^{n} \left(v_j \cdot IDV_j \right)$; where IDV are the isotopomer distributions for the target metabolite, v_j is the jth reaction that produces the target metabolite, and IDV_j are the isotopomer distributions of the target metabolites synthesized from v_j.

To simulate the isotopomer labeling patterns of amino acids that are measured by GC/MS, the isotopomer distributions are converted to mass distribution vectors (MDV) by the following MDV = M · IDV equation, where M is the conversion matrix. The difference between the simulated and measured MDV in proteinogenic amino acids is minimized by using a series of computational optimization algorithm, such as simulated annealing and evolutionary algorithms. Multiple sets of metabolic fluxes are tested and the flux distribution that best match the simulated data with the experimental data is provided by ^{13}C-MFA. The flowchart of ^{13}C-MFA is illustrated in Figure 4.1.

Figure 4.1. Framework for ^{13}C-metabolic flux analysis.



4.2 Metabolic flux analysis of the mixotrophic metabolisms in the green sulfur bacterium *Chlorobaculum tepidum*

Xueyang Feng[1,†], Kuo-Hsiang Tang[2,†], Robert E. Blankenship[2,3], Yinjie J Tang[1,*]

[1]Department of Energy, Environmental and Chemical Engineering,

[2]Department of Biology,

[3]Department of Chemistry, Washington University in St. Louis, St. Louis, MO 63130

*Address correspondence to: Tel: 314-935-3441, Fax: 314-935-7211; E-mail: yinjie.tang@seas.wustl.edu

† These authors contributed equally to the study.

This section was previously published in

Feng X, Tang K-H, Blankenship RE, Tang YJ. "Metabolic flux analysis of the mixotrophic metabolisms in the green sulfur bacterium Chlorobaculum tepidum." J Biol Chem. 2010, 285(50):39544-50.

4.2.1 Abstract

The photosynthetic green sulfur bacterium, *Chlorobaculum (Cba.) tepidum*, assimilates CO_2 and organic carbon sources (acetate or pyruvate) during mixotrophic growth conditions through a unique carbon and energy metabolism. Using a ^{13}C-labeling approach, this study examined biosynthetic pathways and flux distributions in the central metabolism of *Cba. tepidum*. The isotopomer patterns of proteinogenic amino acids revealed an alternate pathway for isoleucine synthesis (via citramalate synthase, CimA, CT0612). ^{13}C-based flux analysis indicated that carbons in biomass were mostly derived from CO_2 fixation via three key routes: the reductive tricarboxylic acid (RTCA) cycle, the pyruvate synthesis pathway via pyruvate:ferredoxin oxidoreductase (PFOR), and the CO_2-anaplerotic pathway via phosphoenolpyruvate carboxylase. During mixotrophic growth with acetate or pyruvate as carbon sources, acetyl-CoA was mainly produced from acetate (via acetyl-CoA synthetase) or citrate (via ATP citrate lyase). PFOR converted acetyl-CoA and CO_2 to pyruvate, and this growth-rate-control reaction is driven by reduced ferredoxin generated during phototrophic growth. Most reactions in the RTCA cycle were reversible. The relative fluxes through the RTCA cycle were 80~100 units for mixotrophic cultures grown on acetate and 200~230 units for cultures grown on pyruvate. Under the same light conditions, the flux results suggested a trade-off between energy-demanding CO_2 fixation and biomass growth rate: *Cba. tepidum* fixed more CO_2 and had higher biomass yield

($Y_{X/S}$: mole carbon in biomass/mole substrate) in pyruvate culture ($Y_{X/S}$=9.2) than in acetate culture ($Y_{X/S}$=6.4), but the biomass growth rate was slower in pyruvate culture than in acetate culture.

Key words: ^{13}C, citramalate, CO_2 fixation, ferredoxin, RTCA, light-harvesting

4.2.2 Introduction

Chlorobaculum tepidum (*Cba. tepidum*) is a representative green sulfur bacterium that is ecologically significant in global cycling of carbon, nitrogen and sulfur (2,3). The *Cba. tepidum* genome has been sequenced, and the genetic tools for creating *Cba. tepidum* mutant strains have been developed to make transposon-based mutations or targeted gene disruptions, which offer great potential to engineer *Cba. tepidum* for future applications (4). The annotated genome reveals unique aspects in carbon and energy metabolism in *Cba. tepidum*. Instead of using the Calvin-Benson Cycle for CO_2-assimilation as in most photosynthetic organisms, *Cba. tepidum* captures energy from light and uses it along with electrons, primarily derived from oxidation of sulfur compounds, to drive the reductive tricarboxylic acid cycle (RTCA) for synthesis of building block molecules (4). *Cba. tepidum* can grow mixotrophically with acetate or pyruvate as the organic carbon source (3). Although recent research has been performed on the carbon and energy metabolism of *Cba. tepidum* (5,6), rigorous quantification of the metabolic pathway activities has not yet been achieved. To provide quantitative readout of the metabolic functions and regulatory

mechanisms, this study has performed ^{13}C-based metabolic flux analysis of *Cba. tepidum* in the following steps: 1) growing cultures of *Cba. tepidum* with ^{13}C-labeled acetate or pyruvate, 2) using gas chromatography-mass spectrometry (GC-MS) to measure the resulting labeling pattern in key metabolites, and 3) deciphering *in vivo* metabolisms via a flux model (7). Isotopic labeling and metabolic flux analysis have been developed to identify the active biosynthesis pathways (8,9,10,11) and measure the global enzymatic reaction rates (12,13). Such a fluxomics approach can bridge the gap between genome annotations and final metabolic outputs, and has been applied for characterizing numerous environmental microorganisms, including *E. coli* (14,15), *Saccharomyces cerevisiae* (16,17), *Bacillus subtilis* (18), *Geobacter metallireducens* (19), *Shewanella oneidensis* (20), *Synechocystis* sp. (21), etc. This paper reports on the first studies of the fluxomics of mixotrophic metabolism in the green sulfur bacteria, and provides complementary information to previous genomic and proteomic studies.

4.2.3 Experimental procedure

^{13}C-labeled experiments. *Cba. tepidum* cultures were grown anaerobically at temperatures ranging from 46–50 $^{\circ}C$ in low-intensity light (10 ± 1 W/m^2). The medium composition (1L) was $Na_2EDTA\text{-}2H_2O$ (0.015 g), $MgSO_4\text{-}7H_2O$ (0.22 g), $CaCl_2\text{-}2H_2O$ (0.08 g), NaCl (0.45 g), NH_4Cl (0.45 g), $Na_2S_2O_3\text{-}5H_2O$ (2.6 g), KH_2PO_4 (0.57 g), MOPS buffer (2.4 g), trace element stock (1.2

mL), vitamin B_{12} stock (0.08 mg), $NaHCO_3$ (0.23 g), and Na_2S-$9H_2O$ (0.07 g). All chemicals were purchased from Sigma-Aldrich. The carbon source for the trace experiments was [1-^{13}C] acetate (purity > 98%, 0.83g/L) or [2-^{13}C] acetate (purity > 98%, 0.83g/L) for cultures growing on acetate, or [3-^{13}C] pyruvate (purity > 98%, 1.1 g/L) for cultures growing on pyruvate. The ^{13}C-labeled acetate and pyruvate were purchased from Cambridge Isotope Laboratories (www.isotope.com). Cell growth was monitored at OD_{625}. 1% cultures (100-fold dilution) in the exponential growth phase were used to inoculate fresh media with ^{13}C-labeled substrates. To reduce the effect of non-labeled carbon from the initial stock, cells were sub-cultured twice in the same labeled medium. Biomass was sampled at two time points (4 hr interval) in the middle-exponential growth phase, and the labeling patterns of proteinogenic amino acids in the biomass were measured. The invariability of amino acid labeling during the two time points confirmed the pseudo-steady-state metabolism in tracer experiments.

Metabolites and isotopomer analysis. The amount of pyruvate and acetate during the growth period was determined by enzymatic assays (22,23,24). For GC-MS measurement of amino acid labeling, the biomass was harvested by centrifugation and hydrolyzed using 6M HCl (24 hrs at 100 $^{\circ}$C) (20,25). The amino acids were derivatized in 0.2 mL tetrahydrofuran and 0.2 mL *N*-(tert-butyl dimethylsilyl)-*N*-methyl-trifluoroacetamide (Sigma-Aldrich, St. Louis, MO). A gas chromatograph (Hewlett-Packard model 7890A; Agilent Technologies, CA)

equipped with a DB5-MS column (J&W Scientific, Folsom, CA) and a mass spectrometer (model 5975C; Agilent Technologies, CA) were used for analyzing metabolite labeling profiles. Four types of charged fragments were detected by gas chromatography-mass spectrometry (GC-MS) for all the amino acids (Table 4.2.1 and Figure 4.2.1): the $[M-57]^+$ or $[M-15]^+$ group (containing unfragmented amino acids); the $[M-159]^+$ or $[M-85]^+$ group (containing amino acids that had lost an α-carboxyl group). For each type of fragments, the labeling patterns were represented by M_0, M_1, M_2, etc, which were fractions of unlabeled, singly labeled, and doubly labeled amino acids. The effects of natural isotopes on isotopomer labeling patterns were corrected by previously reported algorithms (26).

To compare the relative contributions of carbon substrates and CO_2 to mixotrophic biomass synthesis, the substrate utilization ratio R was calculated based on the labeling patterns of unfragmented amino acid X (e.g., alanine) (8):

$$\frac{0.98 \times n \times V_{sub} + 0.01 \times V_{CO_2}}{m \times V_{sub} + V_{CO_2}} = \frac{(\sum_{i=1}^{C} i \times M_i)}{C} \text{ and } R = \frac{mV_{sub}}{V_{CO_2}} \qquad \text{(Equation 1)}$$

where R ratio reflects the carbon flux ratio of labeled carbon substrate to unlabeled CO_2 for producing the corresponding amino acid X (mol carbon from substrate / mol carbon from CO_2); M_i is the GC-MS isotopomer fraction for a given amino acid. C is the total number of carbon atoms in the amino acid molecule. V_{sub} is the uptake of ^{13}C-labeled organic substrates, V_{CO2} is the uptake

of CO_2; 0.98 is the purity of the labeled carbon substrate; 0.01 is the natural abundance of ^{13}C, *m* is the total number of carbons in the substrate molecule, and *n* is the total number of labeled carbons in the substrate molecule.

Metabolic flux analysis. The pathway map of *Cba. tepidum* was generated based on genome annotation from the KEGG database (http://www.genome.jp/kegg/) and transcription analysis of several key pathways (27). The simplified pathway map includes the reductive tricarboxylic acid cycle, CO_2-anaplerotic pathway, gluconeogenesis pathway, and pentose phosphate pathway (Supplementary Figure 4.2.S1 and Table 4.2.S1). The development of a pseudo-steady-state flux model has been discussed before (20,28,29). In brief, the substrate (acetate or pyruvate) uptake rate was measured and normalized to 100 units. The biomass production was determined based on our previous paper (27). The biomass composition for macromolecules such as protein and fatty acids was assigned based on that of *E. coli* (14). The fluxes to biomass pools were loosely constrained by the estimated dry cell weight (DCW) and biomass compositions. These fluxes were used as initial inputs to the isotopomer model and optimized by isotopomer labeling information (Table 4.2.1 and Supplementary Table 4.2.S2). The remaining unknown intracellular fluxes were determined by reaction stoichiometry and atom/isotopomer mapping matrices in an iterative scheme (Supplementary Table 4.2.S3). The reaction reversibility was calculated using the exchange coefficient (19):

$$v_i^{exch} = \frac{exch_i}{1 - exch_i} \qquad \text{(Equation 2)}$$

where v_i^{exch} is the exchange flux defined as the smaller of the forward and backward fluxes, and $exch_i$ is the exchange coefficient with the range of [0,1] (If the reaction is irreversible, $exch_i$ is 0; if the reaction is "freely" reversible, $exch_i$ is close to 1). The flux combinations were searched to minimize of the objective function (20):

$$\varepsilon(v_n) = \sum_{i=1}^{a} \left(\frac{M_i - N_i(v_n)}{\delta_i} \right)^2 \qquad \text{(Equation 3)}$$

where v_n are the unknown fluxes to be optimized in the program, M_i is the measured MS data, N_i is the corresponding model-simulated MS data, and δ_i is the corresponding standard deviation in the GC-MS data (1~2%). The unknown metabolic fluxes were searched to minimize ε. The model was solved by IPOPT (Interior Point Optimizer, https://projects.coin-or.org/Ipopt), which is a software package for large-scale nonlinear optimization. To avoid getting trapped in a local optimal minimum, multiple initial guesses (>100) were used for model calculation to obtain the global solution.

To estimate the confidence interval for the calculated fluxes, a Monte Carlo approach was employed (15). In brief, the isotopomer concentration data sets were generated by adding 2% of normally distributed measurement noise to

actual measured isotopomer data. The same optimization routine was used to estimate flux distribution from these data sets. Confidence limits for each flux value were obtained from the probability distribution of calculated fluxes resulting from the simulated data sets (n=100). To determine the cellular metabolism's requirements for energy and reducing power, the relative consumption rates of ATP, NADH, NADPH and ferredoxin were quantified based on the flux distributions in the central metabolic pathways and biomass synthesis rates.

Calculation of biomass yield. The biomass yield (moles of carbon in biomass/moles of substrate) was calculated based on the substrate uptake rates (moles of substrate/L/hr) and biomass growth rates (g DCW/L/hr). The molecular formula for *Cba. tepidum* was assumed to be $CH_{1.8}O_{0.5}N_{0.2}P_{0.02}$ (molecular weight MW ≈ 25.5) (30). The biomass yield was calculated by:

$$\text{Yield} = \frac{\dfrac{\text{biomass growth rate (g/L/hr)}}{\text{MW of biomass (g/mol)}}}{\text{substrate uptake rate (mol/L/hr)}}$$

4.2.4 Results and discussion

Carbon utilization during mixotrophic growth. During exponential-growth with pyruvate and acetate, *Cba. tepidum* had the growth rates of 0.12 h^{-1} and 0.17 h^{-1}, respectively. In the trace experiments, the labeled carbons were detected in all amino acids (Table 4.2.1), suggesting utilization of acetate or

pyruvate to synthesize all building blocks. Table 4.2.2 shows the ratio of carbon utilization (organic carbon vs. CO_2) for mixotrophic synthesis of proteinogenic amino acids. For acetate culture, the substrate utilization ratio for Ala (precursor: pyruvate) was 0.60, which indicated that carbons in pyruvate were mainly derived from labeled acetate. Pyruvate is the precursor for both the gluconeogenesis pathway and the pentose phosphate pathway, so the R ratios (Equation 1) of Ser and Phe are similar to that of alanine. The substrate utilization ratio of His was high (0.96) for culture with [2-^{13}C]acetate because the His precursor C1 pool (N^5, N^{10}-methylene-tetrahydrofolate) was highly labeled. This carbon was derived from acetate by following route:

[2-^{13}C]acetate→[3-^{13}C]pyruvate→[3-^{13}C]Ser→[^{13}C] C1 pool.

In the mixotrophic culture with acetate, the R ratios for aspartate and glutamate (~0.2) were significantly lower than the other amino acids. Aspartate and glutamate were synthesized from the RTCA cycle, so the small R values indicated that the unlabeled CO_2 was the dominant carbon source for synthesizing metabolites in the RTCA cycle. In pyruvate culture, the substrate utilization ratios of amino acids (Ala, Phe, Ser, and Asp) were all lower than in acetate culture, indicating a higher contribution of CO_2 to synthesize building blocks.

Analysis of amino acid biosynthesis pathways. Growing in a completely defined medium, *Cba. tepidum* was able to synthesize all amino acids from CO_2 and organic carbon sources using the annotated pathways. Interestingly, the

labeling patterns of Leu and Ile from tracer experiments were found to be identical, which indicated that the two amino acids shared the same precursors. Ile is commonly synthesized via threonine ammonia-lyase, with Thr and pyruvate as the precursors, while Leu is synthesized from pyruvate and acetyl-CoA. In the [2-^{13}C] acetate trace experiment, both Thr and pyruvate are singly labeled, which could lead only to doubly labeled Ile, instead of the triply labeled Ile detected by GC-MS. Such an observation is consistent with the fact that a gene encoding threonine ammonia-lyase has not been annotated in the genome of *Cba. tepidum*. On the other hand, an alternative pathway for Ile biosynthesis (i.e., threonine-independent pathway) has been recently identified in several bacteria (9,10,25,31), where Ile is synthesized from acetyl-CoA and pyruvate (i.e., using the same precursors as Leu) through the formation of citramalate as the intermediate (Figure 4.2.1). A search in the genome of *Cba. tepidum* for the gene for citramalate synthase (*CimA*) as found in *Geobacter* (10) returned a high identity for CT0612 (~ 52%), supporting the presence of the citramalate pathway for Ile synthesis.

Central carbon metabolic pathways. The genome annotation from the KEGG database indicates: 1) pyruvate:ferredoxin oxidoreductase (PFOR) (*porA,* CT1628) converts pyruvate to acetyl-CoA (an end-product from RTCA cycle), 2) phosphoenolpyruvate (PEP) is derived from pyruvate via pyruvate phosphate dikinase (*ppd,* CT1682), and 3) PEP is fed into the TCA cycle via the CO_2-

anaplerotic pathway (Figure 4.2.2). The fraction of non-labeled Asp (M_0=0.64, Table 4.2.1) was much higher than that of Ala (M_0=0.45, Table 4.2.1) in trace experiments with [1-^{13}C] acetate, which indicated that non-labeled oxaloacetate was also generated through CO_2 fixation via the RTCA cycle. Furthermore, similar labeling patterns of Ala (precursor: pyruvate) and Ser (precursor: glycerate-3-P) confirmed that the carbon flux through the gluconeogenesis pathway was mainly from pyruvate.

The flux analysis results are shown in Figure 4.2.2. In general, predicted labeling patterns of all amino acids were good matches with the measured isotopomer data (Figure 4.2.3). Under acetate-growth conditions, acetyl-CoA could be generated from acetate uptake or as the end-product through the RTCA cycle. Fluxes through pyruvate:ferredoxin oxidoreductase (PFOR), the RTCA cycle, and the CO_2-anaplerotic pathway were ~152, 80-100, and ~30 units, respectively (Figure 4.2.2). Meanwhile, the flux ratio of the gluconeogenesis pathway to the RTCA cycle was close to 1:1. The oxidative pentose phosphate pathway (G6P → 6PG → Ru5P) was inactive under our experimental conditions. Compared to mixotrophic growth with acetate, the flux distribution during pyruvate growth demonstrated different patterns. In general, the fluxes into the RTCA cycle and CO_2-anaplerotic pathway were much higher (200-230 units and ~50 units, respectively) and thus more CO_2 was fixed. The flux ratio of the gluconeogenesis pathway to the RTCA cycle was reduced to 1:2, while acetyl-

CoA was generated only from the RTCA cycle. The oxidative pentose phosphate pathway was not active (flux<1 unit) under pyruvate mixotrophic growth (or acetate mixotrophic growth).

In the mixotrophic cultures grown on acetate or pyruvate, most reactions in the RTCA cycle were freely reversible (exchange coefficient close to 1). The reversibility is consistent with the physiological free energy changes in the RTCA cycle, where the standard Gibbs free energy for most reactions of the TCA cycle is positive so that the magnitude of the reaction equilibrium constants is small (32). Furthermore, PFOR catalyzes a reversible reaction between pyruvate and acetyl-CoA (pyruvate + CoA + 2 Fd_{ox} ↔ acetyl-CoA + CO_2 + 2 Fd_{red} + 2 H^+). However, metabolic flux analysis of the mixotrophic metabolism indicates that the net flux of this enzymatic reaction was from acetyl-CoA to pyruvate (33). The direction of the PFOR pathway indicates that the reduced ferredoxin from phototrophic processes made the reaction for pyruvate synthesis thermodynamically favorable.

Energy metabolism of *Chlorobaculum tepidum*. The mixotrophic metabolism of *Cba. tepidum* consumes energy harvested from light. Since the oxidative pentose phosphate pathway and normal TCA cycle are not fully functional, the energy (NADPH and NADH) generation mainly depends on the light reactions. Based on the metabolic flux distribution and previous measurement of the absolute uptake rates of carbon substrates and biomass

growth (27), the light-energy harvested for central metabolism and biomass growth was approximately quantified (Figure 4.2.4). In general, the photons absorbed by pigments (e.g. bacteriochlorophylls) initiate the electron flow and oxidize sulfide. The electrons are used to reduce ferredoxin and generate NAD(P)H (3,4). Concomitant with the oxidation of the reduced ferredoxin, the essential cofactors for energy metabolism (i.e., NADPH and NADH) are mainly generated by ferredoxin-$NAD(P)^+$ reductase (3,34). Meanwhile, the oxidation of sulfide also creates a proton motive force for ATP production. The generated ATP, cofactors, and reduced ferredoxin then become the "driver" for the mixotrophic metabolism and the RTCA cycle for energy-demanding CO_2 fixation. Figure 4.2.4 shows that the ATP, NADH and NADPH fluxes into biomass and the RTCA cycle are stronger in acetate-grown cultures than in pyruvate-grown cultures, supporting a higher growth rate during mixotrophic growth with acetate. On the other hand, the calculated biomass yield (moles of carbon in biomass/moles of substrate) was higher for pyruvate culture ($Y_{X/S}$=9.2) than for acetate ($Y_{X/S}$=6.4), based on our previous experiments (27), which is consistent with the fact that the normalized fluxes through RTCA are higher in pyruvate culture than acetate culture. Accordingly, pyruvate metabolism has more CO_2 fixation (but slower growth rate).

Figure 4.2.1. Citramalate pathway for isoleucine biosynthesis in *Cba. tepidum* (using [2-^{13}C] acetate and $NaHCO_3$ as the carbon sources). The asterisks indicate the positions of labeled carbon. The dashed lines indicate inactive pathways.

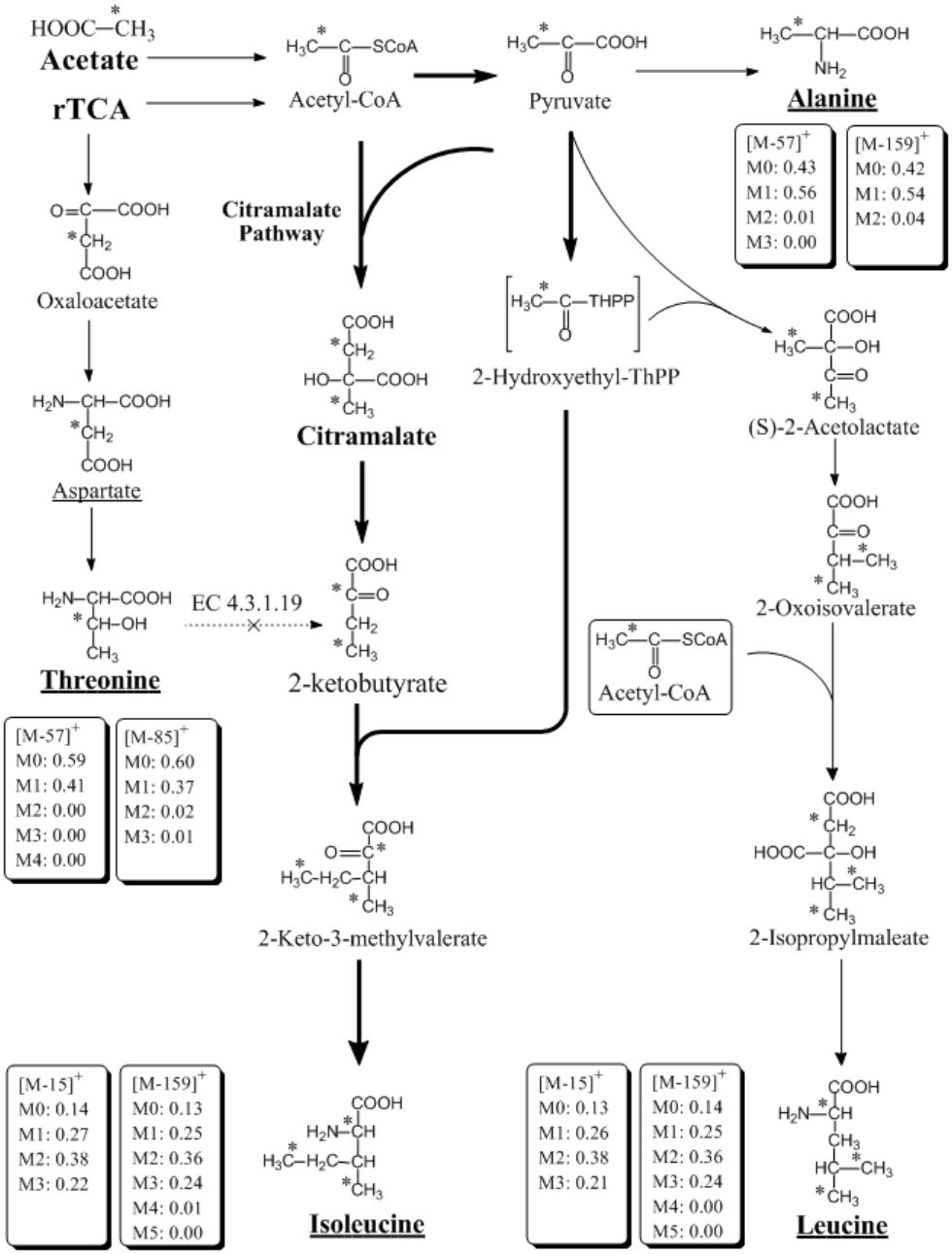
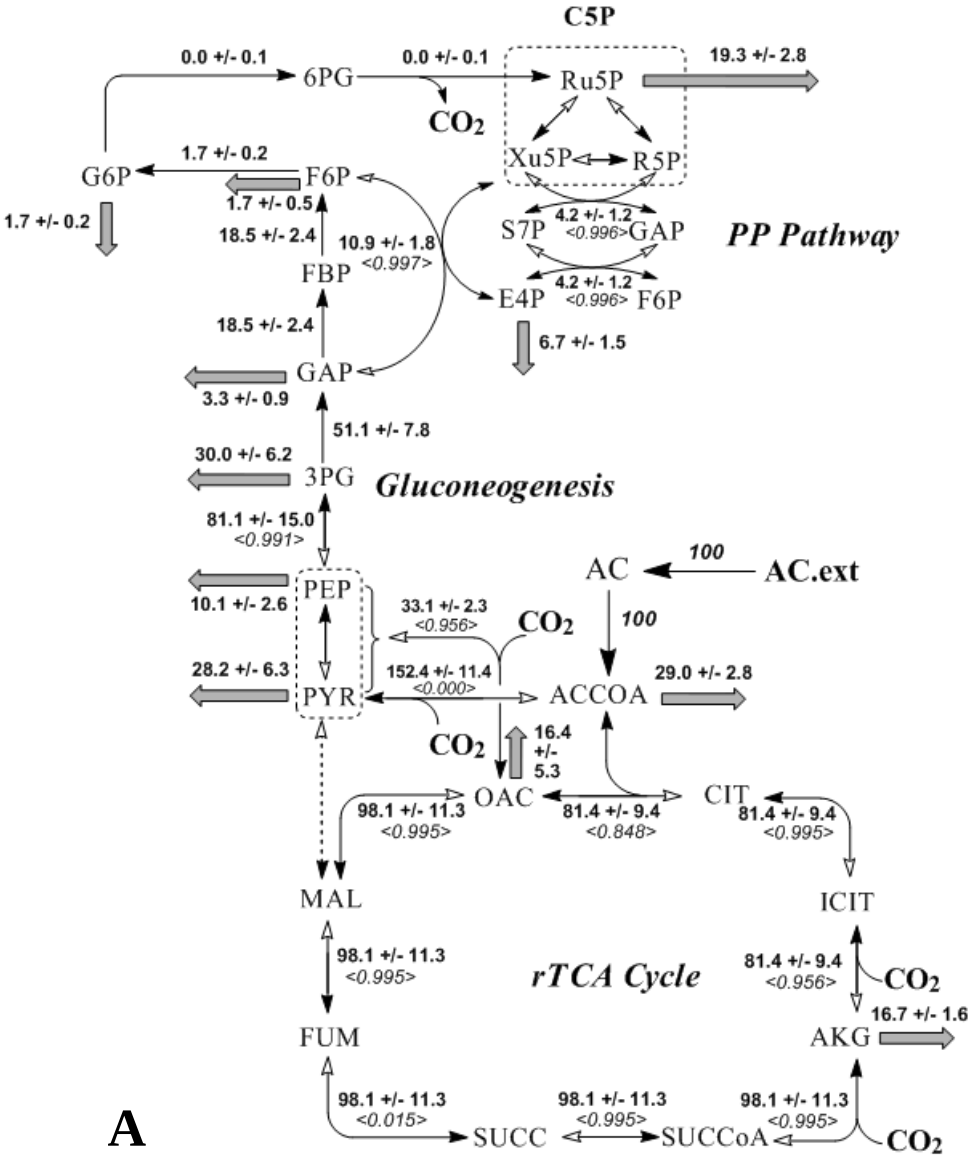
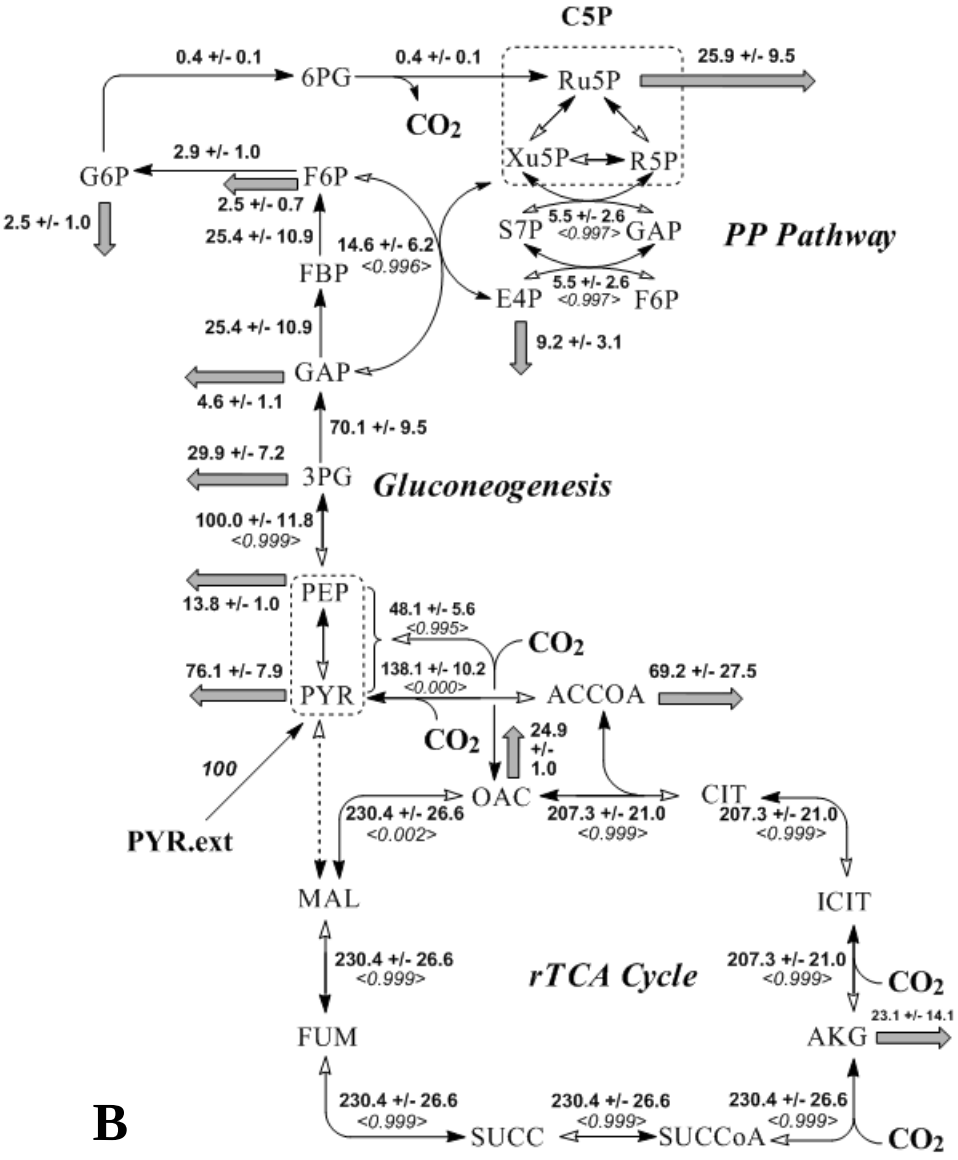


Figure 4.2.2. Metabolic flux distribution in *Cba. tepidum*. (A) Net flux distribution in acetate growth conditions (based on [2-^{13}C] acetate culture); (B) net flux distribution in pyruvate growth conditions. The standard derivation (std) and exchange coefficients are marked by flux +/- std and < exchange coefficient > respectively. The inactive pathways are marked with dashed lines. The calculated biomass yield (moles of carbon in biomass/moles of substrate): pyruvate culture ($Y_{X/S}$=9.2); acetate culture ($Y_{X/S}$=6.4). **Abbreviations:** 3PG, 3-phosphoglycerate; 6PG, 6-phosphogluconate; AC.ext, extracellular acetate; AC, intracellular acetate; ACCOA, acetyl-coenzyme A; AKG, α-ketoglutarate; CIT, citrate; DCW, dry cell weight; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; FBP, Fructose 1,6-bisphosphate; Fd_{red}, reduced ferredoxin; Fd_{ox}, oxidized ferredoxin; FNR, ferredoxin-NAD(P)$^+$ reductase; FUM, fumarate; G6P, glucose-6-phosphate; GAP, glyceraldehyde 3-phosphate; ICIT, isocitrate; KGOR, 2-ketoglutarate ferredoxin oxidoreductase; MAL, malate; OAC, oxaloacetate; PEP, phosphoenolpyruvate; PFOR, pyruvate:ferredoxin oxidoreductase; PYR, intracellular pyruvate; PYR.ext, extracellular pyruvate; R5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate; RTCA, the reductive tricarboxylic acid; S7P, sedoheptulose-7-phosphate; SUCC, succinate; SUCCoA, Succinyl-CoA; Xu5P, xylulose-5-phosphate.


C5P
0.0 +/- 0.1
6PG
0.0 +/- 0.1
Ru5P
19.3 +/- 2.8
CO2
1.7 +/- 0.2
G6P
F6P
Xu5P
R5P
1.7 +/- 0.5
1.7 +/- 0.2
18.5 +/- 2.4
10.9 +/- 1.8
<0.997>
S7P
4.2 +/- 1.2
<0.996>
GAP
PP Pathway
FBP
4.2 +/- 1.2
<0.996>
E4P
F6P
18.5 +/- 2.4
6.7 +/- 1.5
GAP
3.3 +/- 0.9
51.1 +/- 7.8
30.0 +/- 6.2
3PG
Gluconeogenesis
81.1 +/- 15.0
<0.991>
100
AC
AC.ext
PEP
10.1 +/- 2.6
33.1 +/- 2.3
<0.956>
CO2
100
28.2 +/- 6.3
152.4 +/- 11.4
<0.000>
29.0 +/- 2.8
PYR
ACCOA
CO2
16.4
+/-
5.3
98.1 +/- 11.3
<0.995>
OAC
81.4 +/- 9.4
<0.848>
CIT
81.4 +/- 9.4
<0.995>
MAL
ICIT
98.1 +/- 11.3
<0.995>
rTCA Cycle
81.4 +/- 9.4
<0.956>
CO2
16.7 +/- 1.6
FUM
AKG
98.1 +/- 11.3
<0.015>
98.1 +/- 11.3
<0.995>
98.1 +/- 11.3
<0.995>
SUCC
SUCCoA
CO2
A


C5P
0.4 +/- 0.1
6PG
0.4 +/- 0.1
Ru5P
25.9 +/- 9.5
CO2
2.9 +/- 1.0
G6P
F6P
Xu5P
R5P
2.5 +/- 1.0
2.5 +/- 0.7
25.4 +/- 10.9
14.6 +/- 6.2
<0.996>
S7P
5.5 +/- 2.6
<0.997>
GAP
PP Pathway
FBP
5.5 +/- 2.6
<0.997>
E4P
F6P
25.4 +/- 10.9
9.2 +/- 3.1
GAP
4.6 +/- 1.1
70.1 +/- 9.5
29.9 +/- 7.2
3PG
Gluconeogenesis
100.0 +/- 11.8
<0.999>
13.8 +/- 1.0
PEP
48.1 +/- 5.6
<0.995>
CO2
76.1 +/- 7.9
138.1 +/- 10.2
<0.000>
69.2 +/- 27.5
PYR
ACCOA
CO2
24.9
+/-
1.0
100
OAC
CIT
230.4 +/- 26.6
<0.002>
207.3 +/- 21.0
<0.999>
207.3 +/- 21.0
<0.999>
PYR.ext
MAL
ICIT
230.4 +/- 26.6
<0.999>
rTCA Cycle
207.3 +/- 21.0
<0.999>
CO2
23.1 +/- 14.1
FUM
AKG
230.4 +/- 26.6
<0.999>
230.4 +/- 26.6
<0.999>
230.4 +/- 26.6
<0.999>
CO2
SUCC
SUCCoA

B

Figure 4.2.3. Model quality test for (A) acetate metabolism; (B) pyruvate metabolism. ● alanine data, ○ serine data, ▼ aspartate data, Δ glutamate data, ■ leucine data, □ histidine data, ◊ phenylalanine data, and ♦ glycine data.

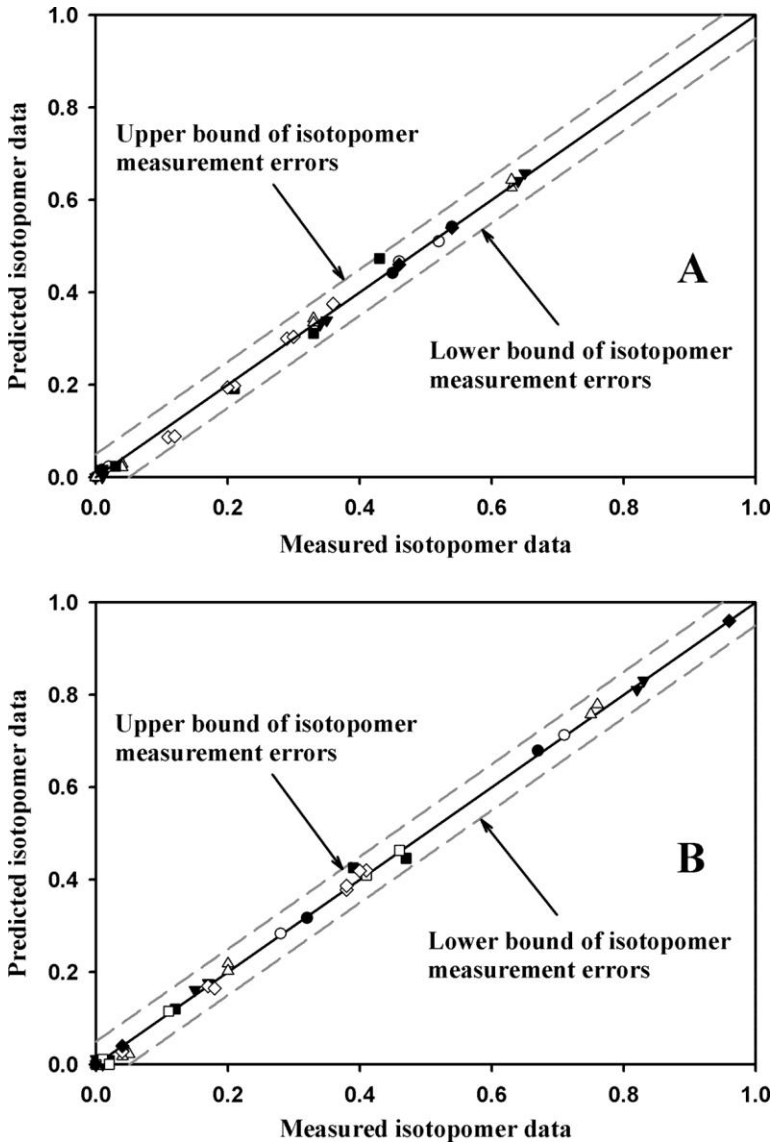
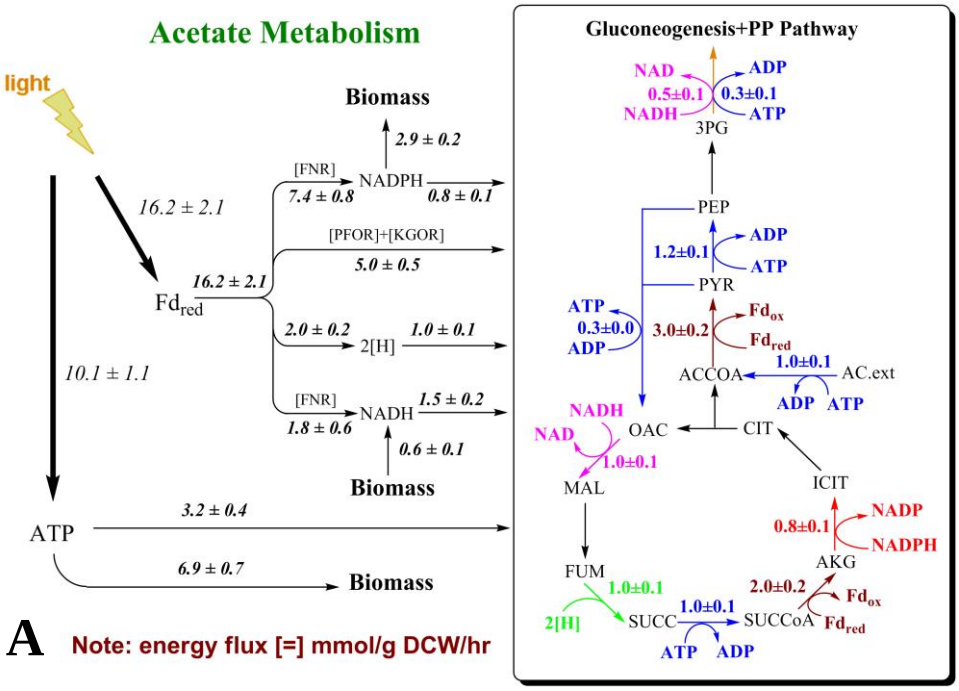


Figure 4.2.4. Proposed energy metabolism in *Cba. tepidum*. (A) energy requirement (mmol/gDCW/hr) in acetate growth conditions; (B) energy requirement (mmol/gDCW/hr) in pyruvate growth conditions. The intracellular energy metabolism was quantified in the framed figures based on the relative flux distributions (The detailed calculations are in Supplementary Tables 4.2.S4A and S4B). Arrows pointing to the framed figure indicated the energy demand of intracellular metabolism. Arrows pointing to biomass indicate the energy demand of biomass accumulation. Arrows pointing from light indicated the entire energy harvested by *Cba. tepidum*. The light reaction produces reduced ferredoxin and ATP. NADPH and NADH are mainly generated by ferredoxin-$NAD(P)^+$ reductase. Biomass (protein) synthesis can also generate a small amount of NADH, as indicated in the figure.


Acetate Metabolism
light
Biomass
2.9 ± 0.2
[FNR]
7.4 ± 0.8
NADPH
0.8 ± 0.1
16.2 ± 2.1
[PFOR]+[KGOR]
5.0 ± 0.5
16.2 ± 2.1
Fdred
2.0 ± 0.2
2[H]
1.0 ± 0.1
10.1 ± 1.1
[FNR]
1.8 ± 0.6
NADH
1.5 ± 0.2
0.6 ± 0.1
Biomass
3.2 ± 0.4
ATP
6.9 ± 0.7
Biomass
A
Note: energy flux [=] mmol/g DCW/hr
Gluconeogenesis+PP Pathway
NAD
0.5±0.1
NADH
ADP
0.3±0.1
ATP
3PG
PEP
1.2±0.1
ADP
ATP
PYR
ATP
0.3±0.0
ADP
3.0±0.2
Fdox
Fdred
ACCOA
1.0±0.1
AC.ext
ADP
ATP
NADH
NAD
OAC
CIT
1.0±0.1
MAL
ICIT
0.8±0.1
NADP
NADPH
AKG
FUM
1.0±0.1
2[H]
2.0±0.2
Fdox
SUCC
1.0±0.1
SUCCoA
Fdred
ATP
ADP

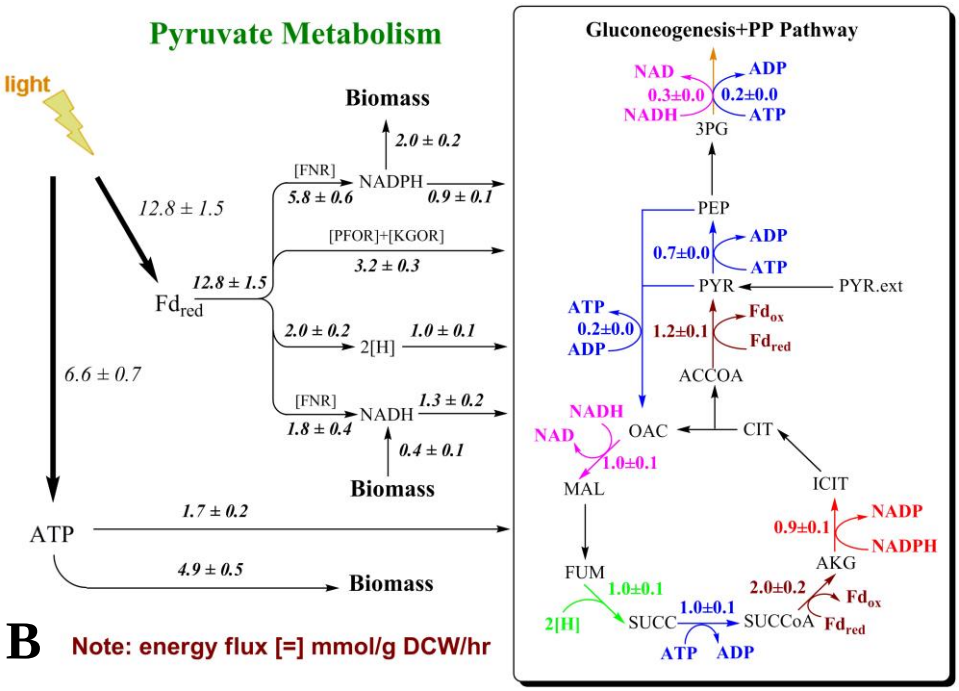

Pyruvate Metabolism
light
Biomass
2.0 ± 0.2
[FNR]
5.8 ± 0.6
NADPH
0.9 ± 0.1
12.8 ± 1.5
[PFOR]+[KGOR]
3.2 ± 0.3
12.8 ± 1.5
Fdred
2.0 ± 0.2
2[H]
1.0 ± 0.1
6.6 ± 0.7
[FNR]
1.8 ± 0.4
NADH
1.3 ± 0.2
0.4 ± 0.1
Biomass
1.7 ± 0.2
ATP
4.9 ± 0.5
Biomass
B
Note: energy flux [=] mmol/g DCW/hr
Gluconeogenesis+PP Pathway
NAD
0.3±0.0
NADH
ADP
0.2±0.0
ATP
3PG
PEP
0.7±0.0
ADP
ATP
PYR
PYR.ext
ATP
0.2±0.0
ADP
1.2±0.1
Fdox
Fdred
ACCOA
NADH
NAD
OAC
CIT
1.0±0.1
MAL
ICIT
0.9±0.1
NADP
NADPH
AKG
FUM
1.0±0.1
2[H]
2.0±0.2
Fdox
SUCC
1.0±0.1
SUCCoA
Fdred
ATP
ADP

Figure 4.2.S1. Pathway map for mixotrophic metabolism of *Chlorobaculum tepidum*

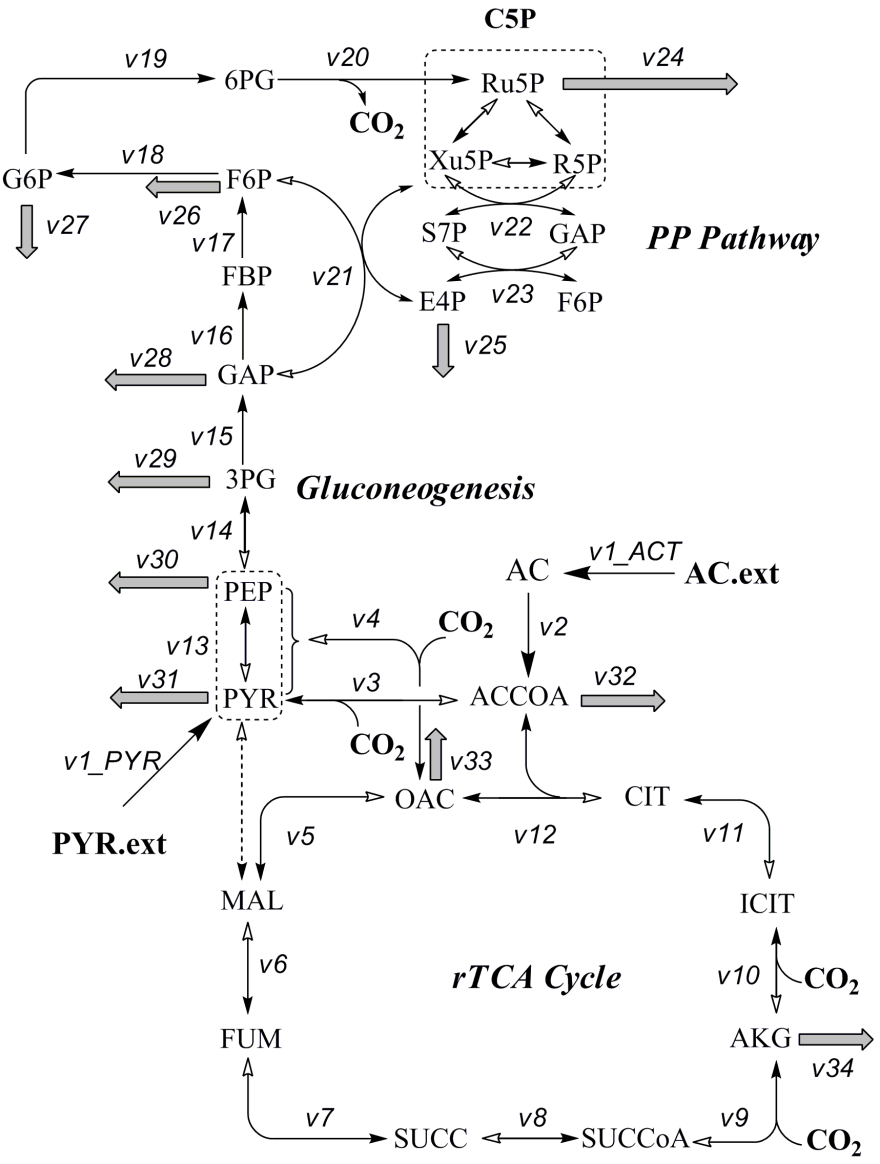


Table 4.2.1. Isotopic labeling patterns in amino acids from *Chlorobaculum tepidum*[a]

Carbon source	Amino acids	Fragments	M_0	M_1	M_2
[1-^{13}C] acetate	Ala	$[M-57]^+$	0.45	0.54	0.01
		$[M-159]^+$	0.44	0.53	0.03
	Gly	$[M-57]^+$	0.54	0.46	0.00
		$[M-85]^+$	0.54	0.46	
	Ser	$[M-57]^+$	0.46	0.52	0.02
		$[M-159]^+$	0.47	0.53	0.00
	Leu	$[M-159]^+$	0.21	0.43	0.33
	Ile	$[M-159]^+$	0.21	0.43	0.33
	Asp	$[M-57]^+$	0.64	0.35	0.01
		$[M-159]^+$	0.65	0.34	0.01
	Glu	$[M-57]^+$	0.63	0.33	0.04
		$[M-159]^+$	0.63	0.33	0.04
	Phe	$[M-57]^+$	0.11	0.29	0.36
		$[M-159]^+$	0.12	0.30	0.36
[2-^{13}C] acetate	Ala	$[M-57]^+$	0.43	0.56	0.01
		$[M-159]^+$	0.42	0.54	0.04
	Gly	$[M-57]^+$	0.94	0.06	0.00
		$[M-85]^+$	0.95	0.05	
	Ser	$[M-57]^+$	0.44	0.55	0.01
		$[M-159]^+$	0.45	0.55	0.00
	Leu	$[M-159]^+$	0.13	0.25	0.36
	Ile	$[M-159]^+$	0.14	0.25	0.36
	Asp	$[M-57]^+$	0.61	0.38	0.00
		$[M-159]^+$	0.62	0.37	0.00
	Glu	$[M-57]^+$	0.62	0.35	0.03
		$[M-159]^+$	0.63	0.34	0.03
	Phe	$[M-57]^+$	0.13	0.26	0.36
		$[M-159]^+$	0.14	0.26	0.35
	His	$[M-57]^+$	0.16	0.35	0.36
		$[M-159]^+$	0.28	0.48	0.22
[3-^{13}C] pyruvate	Ala	$[M-57]^+$	0.67	0.32	0.01
		$[M-159]^+$	0.65	0.30	0.05
	Gly	$[M-57]^+$	0.96	0.04	0.00
		$[M-85]^+$	0.97	0.03	
	Ser	$[M-57]^+$	0.71	0.28	0.01
		$[M-159]^+$	0.71	0.28	0.01
	Leu	$[M-159]^+$	0.47	0.38	0.12
	Ile	$[M-159]^+$	0.47	0.39	0.12
	Asp	$[M-57]^+$	0.82	0.17	0.00
		$[M-159]^+$	0.83	0.15	0.01
	Glu	$[M-57]^+$	0.75	0.20	0.05
		$[M-159]^+$	0.76	0.20	0.04
	Phe	$[M-57]^+$	0.38	0.41	0.17

	His	$[M-159]^+$	0.38	0.40	0.18
		$[M-57]^+$	0.46	0.41	0.11
		$[M-159]^+$	0.56	0.34	0.08

a. the standard deviations for GC-MS measurement were based on the duplicate experiments (n=2), with the standard deviation below 2%.

Table 4.2.2. Carbon substrate utilization ratios[a] in amino acids from *Chlorobaculum tepidum*

Amino acids	Carbon sources		
	[1-^{13}C] acetate + $NaHCO_3$	[2-^{13}C] acetate + $NaHCO_3$	[3-^{13}C] pyruvate + $NaHCO_3$
Ala	0.60	0.60	0.48
Ser	0.58	0.58	0.42
Asp	0.20	0.22	0.12
Glu	0.18	0.18	0.21
His	NA[b]	0.96	0.38
Phe	0.62	0.64	0.26

a. the carbon source utilization ratio (organic substrate/CO_2 fixation) for amino acid synthesis was calculated according to Equation (1); b. isotopomer labeling pattern for histidine was not detected under [1-^{13}C] acetate + $NaHCO_3$ due to the weak signal to noise ratio in GC-MS.

Table 4.2.S1. Metabolic pathways in mixotrophic metabolism of *Chlorobaculum tepidum*

Pathway ID	Reactions	Notes
v1_ACT	Extracellular acetate→Intracellular acetate	Confirmed by qRT-PCR (*acsA*, *ackA*)
v1_PYR	Extracellular pyruvate→Intracellular pyruvate	
v2	Acetate→Acetyl-CoA (-ATP)	
v3	Acetyl-CoA+CO2↔Pyruvate (-2 Fd_{red})	Confirmed by qRT-PCR (*porA*)
v4	PEP/Pyruvate+CO2↔OAC (+ATP)	Confirmed by qRT-PCR (*ppc*, *pckA*)
v5	OAC↔MAL (-NADH)	
v6	MAL↔FUM	
v7	FUM↔SUCC (-2[H])	
v8	SUCC↔SUCCoA (-ATP)	
v9	SUCCoA+CO2↔AKG (-2 Fdr_{ed})	Confirmed by qRT-PCR (*korA*, *korB*)
v10	AKG+CO2↔ICIT (-NADPH)	Confirmed by qRT-PCR (*icd*)
v11	ICIT↔CIT	Confirmed by qRT-PCR (*acn*)
v12	CIT↔ACCOA+OAC	Confirmed by qRT-PCR (*gltA*, *aclA*, *aclB*)
v13	PYR↔PEP (-ATP)	Confirmed by qRT-PCR (*ppd*)
v14	PEP↔3PG	
v15	3PG→GAP (-ATP) (-NADH)	
v16	2 GAP→FBP	
v17	FBP→F6P (+ATP)	
v18	F6P→GAP	
v19	GAP→6PG (+NADPH)	
v20	6PG→C5P+CO2 (+NADPH)	
v21	F6P+GAP↔XU5P+E4P	
v22	E4P+F6P↔S7P+GAP	
v23	S7P+GAP↔XU5P+R5P	
v24	R5P→Biomass	Energy cost is listed in Table.S2
v25	E4P→Biomass	Energy cost is listed in Table.S2
v26	F6P→Biomass	Energy cost is listed in Table.S2
v27	G6P→Biomass	Energy cost is listed in Table.S2
v28	GAP→Biomass	Energy cost is listed in Table.S2
v29	3PG→Biomass	Energy cost is listed in Table.S2
v30	PEP→Biomass	Energy cost is listed in Table.S2
v31	PYR→Biomass	Energy cost is listed in Table.S2
v32	ACCOA→Biomass	Energy cost is listed in Table.S2
v33	OAC→Biomass	Energy cost is listed in Table.S2
v34	AKG→Biomass	Energy cost is listed in Table.S2

Table 4.2.S2. Precursors for building blocks synthesis in *Chlorobaculum tepidum*

Building block	**Cost of making 1 mol of each of these building blocks (mol/mol)**
	Precursors
Protein amino acids	
Alanine	1 PYR
Arginine	1 AKG, (-5 ATP) (-1 NADPH)
Asparagine	1 Asp, (-2 ATP)
Aspartate	1 OAC
Cysteine	1 Ser (-3 ATP) (-4 NADPH)
Glutamate	1 AKG (-1 NADPH)
Glutamine	1 Glu (-1 ATP)
Glycine	1 Ser
Histidine	1 C5P, 1 C1 unit (-5 ATP) (+2 NADH)
Isoleucine	2 PYR, 1 ACCOA (-1 NADPH) (+1 NADH)
Leucine	2 PYR, 1 ACCOA (-1 NADPH) (+1 NADH)
Lysine	1 OAC, 1 PYR (-1 NADPH)
Methionine	1 Asp,1 C1 unit (-1 ATP) (-2 NADPH)
Phenylalanine	1 E4P, 2 PEP (-1 ATP) (-1 NADPH)
Proline	1 Glu (-1 ATP) (-2 NADPH)
Serine	1 3PG (+1 NADH)
Threonine	1 OAC (-2 ATP) (-2 NADPH)
Tryptophan	1 C5P, 1 E4P,1 PEP (-3 ATP) (-1 NADPH)
Tyrosine	1 E4P, 2 PEP (-1 ATP) (-1 NADPH) (+1 NADH)
Valine	2 PYR (-1 NADPH)
RNA nucleotides	
ATP	1 C5P, 1 3PG
GTP	1 C5P, 1 3PG
CTP	1 C5P, 1 OAC
UTP	1 C5P, 1 OAC
DNA nucleotides	
dATP	1 C5P, 1 3PG
dGTP	1 C5P, 1 3PG
dCTP	1 C5P, 1 OAC
dTTP	1 C5P, 1 OAC
Average fatty acid	8.2 ACCOA
Average carbohydrate	1 G6P

Table 4.2.S3. Atom mapping matrices for metabolic reactions

Pyr→PEP 1 0 0 0 1 0 0 0 1	PEP→3PG 1 0 0 0 1 0 0 0 1	3PG→GAP 1 0 0 0 1 0 0 0 1	Acetate→ ACCOA 1 0 0 1	C1→GLY 0 1
ICIT→CO2 0 0 0 0 0 1	ACCOA→Pyruvate 0 0 1 0 0 1	AKG→CO2 1 0 0 0 0	AKG →GLU 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1	AKG→SUCC 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1
ASP→MET 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0	ACCOA → LEU 1 0 0 1 0 0 0 0 0 0 0 0	PYR→ LEU 0 0 0 0 0 0 0 1 0 0 1 0 0 0 1 0 0 1	C1→MET 0 0 0 0 1	C1→HIS 0 0 0 0 0 1
E4P→TYR 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1	E4P→F6P 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1	E4P→PHE 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1	E4P→S7P 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1	E4P→GAP 0 1 0 0 0 0 1 0 0 0 0 1
CO2→OAA 0 0 0 1	F6P→E4P 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1	F6P→G6P 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1	F6P→S7P 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	GAP→F6P 0 0 1 0 1 0 1 0 0 1 0 0 0 1 0 0 0 1
F6P→C5P 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	ICIT→AKG 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0	F6P→G6P 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1	G6P→C5P 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1	OAC→MAL 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1

			0 0 0 0 0 1 0 0 0 0 0 0 1	
MAL→FUM 0.5 0 0 0.5 0 0.5 0.5 0 0 0.5 0.5 0 0.5 0 0 0.5	FUM→SUCC 0.5 0 0 0.5 0 0.5 0.5 0 0 0.5 0.5 0 0.5 0 0 0.5	SUCC→SUCCoA 0.5 0 0 0.5 0 0.5 0.5 0 0 0.5 0.5 0 0.5 0 0 0.5	PYR→CO2 1 0 0	PYR→ACCOA 0 1 0 0 0 1
SUCCoA→SUCC 0.5 0 0 0.5 0 0.5 0.5 0 0 0.5 0.5 0 0.5 0 0 0.5	CO2→ MAL 0 0 0 1	PYR→OAA 1 0 0 0 1 0 0 0 1 0 0 0	SUCC→FUM 0.5 0 0 0.5 0 0.5 0.5 0 0 0.5 0.5 0 0.5 0 0 0.5	FUM→MAL 0.5 0 0 0.5 0 0.5 0.5 0 0 0.5 0.5 0 0.5 0 0 0.5
AKG→SUCCoA 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1	SUCCoA →AKG 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1	MAL(OAA)→CO2 0 0 0 1	MAL→OAC 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1	PYR→MAL 1 0 0 0 1 0 0 0 1 0 0 0
CIT → ACCOA 0.5 0 0 0 0.5 0 0 0.5 0 0.5 0 0	ACCOA → CIT 0.5 0 0 0.5 0 0 0 0.5 0.5 0 0 0	CIT→OAC 0 0 0 0 0 1 0 0 1 0 0 0 0 0.5 0 0.5 0 0 0.5 0 0 0 0.5 0	OAC→CIT 0 0 0 0.5 0 0 0.5 0 0 1 0 0 0 0 0.5 0 0 0 0 0.5 1 0 0 0	PYR→ALA 1 0 0 0 1 0 0 0 1
OAC→ASP 1 0 0 0 0 1 0 0 0 0 1 0 0 0 0 1	OAC→CO2 0 0 0 1	PEP→OAA 1 0 0 0 1 0 0 0 1 0 0 0	OAC→PEP 1 0 0 0 0 1 0 0 0 0 1 0	PEP→PYR 1 0 0 0 1 0 0 0 1
C5P→F6P 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	C5P→S7P 1 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1	G6P→CO2 1 0 0 0 0 0	PEP→TYR 1 0 0 0 1 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0	PEP→PHE 1 0 0 0 1 0 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
GAP→C5P 0 0 0 0 0 0 1 0 0 0 1 0 0 0 1	PYR→LEU 0 0 0 0 0 0 0 1 0 0 2 0 0 0 2 0 0 1	PYR→Lys 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 1	PYR→Val 1 0 0 0 1 0 0 2 0 0 0 2 0 0 1	C5P→His 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0
C5P→S7P 1 0 0 0 0	C5P→GAP 0 0 1 0 0	S7P→E4P 0 0 0 1 0 0 0	S7P→F6P 1 0 0 0 0 0 0	S7P→ C5P 1 0 1 0 0 0 0

0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1	0 0 0 1 0 0 0 0 0 1	0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1	0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 1 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 1
SER→C1 0 0 1	SER→GLY 1 0 0 0 1 0	ICIT→AKG 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 1 0	AKG→ICIT 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0	GAP→E4P 0 0 0 1 0 0 0 1 0 0 0 1
CIT→ICIT 0.5 0 0 0 0.5 0 0 0.5 0 0.5 0 0 0 0 1 0 0 0 0 0.5 0 0.5 0 0 0.5 0 0 0 0.5 0 0 0 0 0 0 1	ICIT→CIT 0.5 0 0 0 0.5 0 0 0.5 0 0.5 0 0 0 0 1 0 0 0 0 0.5 0 0.5 0 0 0.5 0 0 0 0.5 0 0 0 0 0 0 1			

Table 4.2.S4A. Energy demand of *Chlorobaculum tepidum* in acetate growth conditions

Acetate Metabolism							
Flux ID	Flux (units)	Flux (mmol/g/hr)	ATP	NADH	NADPH	2[H]	Fd_{red}
v1_ACT	100.0±0.0	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v2	100.0±0.0	1.0±0.1	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v3	152.4±11.4	1.5±1.2	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	3.0±0.2
v4	33.1±2.3	0.3±0.0	-0.3±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v5	98.1±11.3	1.0±0.1	0.0±0.0	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0
v6	98.1±11.3	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v7	98.1±11.3	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	1.0±0.1	0.0±0.0
v8	98.1±11.3	1.0±0.1	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v9	98.1±11.3	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	2.0±0.2
v10	81.4±9.4	0.8±0.1	0.0±0.0	0.0±0.0	0.8±0.1	0.0±0.0	0.0±0.0
v11	81.4±9.4	0.8±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v12	81.4±9.4	0.8±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v13	124.2±10.4	1.2±0.1	1.2±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v14	81.1±15.0	0.8±0.2	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v15	51.1±7.8	0.5±0.1	0.5±0.1	0.5±0.1	0.0±0.0	0.0±0.0	0.0±0.0
v16	18.5±2.4	0.2±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v17	18.5±2.4	0.2±0.0	-0.2±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v18	1.7±0.2	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v19	0.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v20	0.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v21	10.9±1.8	0.1±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v22	4.2±0.2	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v23	4.2±0.2	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Sum	-	-	3.2±0.4	1.5±0.2	0.8±0.1	1.0±0.1	5.0±0.5
Biomass	-	μ=0.17 hr^{-1} ±0.02 hr^{-1}	40.86·0.17 =6.9±0.7	-3.55·0.17 =0.6±0.1	16.87·0.17 =2.9±0.3	0.00·0.17 =0.0±0.0	0.00·0.17 =0.0±0.0

Table 4.2.S4B. Energy demand of *Chlorobaculum tepidum* in pyruvate growth conditions

Pyruvate Metabolism							
Flux ID	Flux (units)	Flux (mmol/g/hr)	ATP	NADH	NADPH	2[H]	Fd_{red}
v1_PYR	100.0±0.0	0.4±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v2	100.0±0.0	0.4±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v3	138.1±10.2	0.6±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	1.2±0.1
v4	48.1±5.6	0.2±0.0	-0.2±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v5	230.4±26.6	1.0±0.1	0.0±0.0	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0
v6	230.4±26.6	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v7	230.4±26.6	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	1.0±0.1	0.0±0.0
v8	230.4±26.6	1.0±0.1	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v9	230.4±26.6	1.0±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	2.0±0.2
v10	207.3±21.0	0.9±0.1	0.0±0.0	0.0±0.0	0.9±0.1	0.0±0.0	0.0±0.0
v11	207.3±21.0	0.9±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v12	207.3±21.0	0.9±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v13	161.1±4.9	0.7±0.0	0.7±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v14	100.0±11.8	0.4±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v15	70.1±9.5	0.3±0.0	0.3±0.0	0.3±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v16	25.4±10.9	0.1±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v17	25.4±10.9	0.1±0.0	-0.1±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v18	2.9±1.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v19	0.4±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v20	0.4±0.1	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v21	14.6±6.2	0.1±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v22	5.5±2.6	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
v23	5.5±2.6	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0	0.0±0.0
Sum	-	-	1.7±0.2	1.3±0.1	0.9±0.1	1.0±0.1	3.2±0.3
Biomass	-	μ=0.12 hr^{-1} ±0.01 hr^{-1}	40.86·0.12 =4.9±0.5	-3.55·0.12 =-0.4±0.1	16.87·0.12 =2.0±0.2	0.00·0.12 =0.0±0.0	0.00·0.12 =0.0±0.0

Biomass formation: 2.833 PYR+ 1.078 AKG+1.787 OAC+1.493 3PG+0.898 R5P+2.938 ACCOA+0.719 PEP+0.361 E4P+0.205 G6P+0.071 F6P+0.129 GAP +16.866 NADPH+ 40.860 ATP→39.68 Biomass+3.547 NADH

4.3 ^{13}C-metabolic flux analyses of *Thermoanaerobacter* species

^{13}C-MFA has been applied to quantify and compare metabolic flux distributions in two *Thermoanaerobacter* species: *Thermoanaerobacter* sp. X514 and *Thermoanaerobacter pseudethanolicus* 39E (35). Both species were cultured anaerobically with 2 g/L of [1-^{13}C] glucose or 2 g/L of [1-^{13}C] xylose. The metabolic network of *Thermoanaerobacter* species included pentose phosphate pathway, glycolysis, the branched TCA cycle, and the futile pathways.

When using glucose as the carbon substrate, the flux through the oxidative pentose phosphate pathway, which was often used for NADPH production, was moderately higher in 39E than that in X514. In spite of very different growth rates between X514 and 39E in glucose and xylose cultures, the intracellular carbon flux distributions, after normalization by carbon substrate consumption rates, were not significantly different between these two strains. This suggested a similar regulation of central metabolic pathways in these two phylogenetically closely related strains. The report of ^{13}C-MFA of *Thermoanaerobacter* species is attached in Appendix 7. Beside, ^{13}C-MFA has also been used in investigating the metabolic robustness of *Shewanella oneidensis* MR-1 (36) (Appendix 8).

4.4 Evaluation of isotope discrimination in ^{13}C-based metabolic flux analysis

Xueyang Feng and Yinjie J. Tang*

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

Corresponding author:

*: yinjie.tang@seas.wustl.edu; phone: 314-935-3441

This section was previously published in

Feng X and Tang YJ. "Evaluation of isotope discrimination in 13C-based metabolic flux analysis." Anal Biochem. 2011, 417(2): 295-297.

4.4.1 Abstract

In a ^{13}C-experiment for metabolic flux analysis (^{13}C-MFA), we examined isotope discrimination by measuring the labeling of glucose, amino acids and hexose monophosphates via mass spectrometry. When *Escherichia coli* grew in a mix of 20% fully labeled and 80% naturally labeled glucose medium, the cell metabolism favored light isotopes, and the measured isotopic ratios (δ^{13}C) were in the range of -35 to -92. Glucose transporters might play an important role in such isotopic fractionation. Flux analysis showed that both isotopic discrimination and isotopic impurities in labeled substrates could impact the accuracy and precision of ^{13}C-MFA.

Key words: amino acid, glucose transporter, δ^{13}C, mass spectrometer, isotopic impurity

4.4.2 Methods and results

^{13}C-metabolic flux analysis (^{13}C-MFA) experiments consist of feeding microbes with enriched ^{13}C-labeled substrate, then measuring the isotopic distribution in the resulting metabolites (often amino acids) to quantify absolute fluxes through the metabolic network. In a typical ^{13}C-MFA experiment, uniformly ^{13}C-labeled substrates are often used in a mixture with the non-labeled form to create efficient scrambling of labeled carbons in the backbone of all metabolites (37,38). Such an experiment assumes that the microbial activity is equal for all isotopomers, and that ^{13}C tracing will not change the metabolic

kinetics. However, a kinetic isotopic effect may be present in all bio-reactions, wherein light isotopes react faster than heavy isotopes due to the lower activation energies for enzymatic breaking of the ^{12}C-^{12}C bond than the ^{13}C-^{13}C bond (39,40,41). In the microbial process, the isotopic effect depends on isotopomer diffusion kinetics, intrinsic enzyme characteristics, and microbial growth conditions (such as pH, micronutrients, and growth factors) (41,42). However, most isotopic effect studies are in the fields of biogeochemistry and ecology, where the ^{13}C is naturally labeled (around 1.1%).

This study investigated the isotopic effect in ^{13}C-MFA experiments when highly enriched ^{13}C-susbtrate was used. *E.coli* BL21 (DE3) was first grown in M9 medium with 22.2 mM of glucose (shaking flasks, 150 rpm, 37°C), which contained non-labeled glucose (Sigma-Aldrich, USA) and fully labeled glucose (99% purity, Cambridge Isotope, USA) in approximately a 4:1 molar ratio. To reduce the non-labeled carbon from the initial stock, cells were then sub-cultured (0.1% inoculation ratio) in the same ^{13}C-labeled medium. All sub-cultured samples from the ^{13}C-MFA experiments were harvested at the mid-log phase (8 hrs) for isotopomer analysis.

We first employed an isotope ratio mass spectrometer and a gas chromatograph-mass spectrometer (GC-MS) to analyze isotopic fractionation in cell metabolism. The exact ratio of $^{13}C/^{12}C$ (i.e. $^{13}C/^{12}C=0.2497$) in the initial carbon substrate (glucose) of ^{13}C-MFA experiments was determined using an

isotope ratio mass spectrometer after combustion of glucose to CO_2 (performed by the Stable Isotope Biogeochemistry Laboratory, Washington University). Since the mass detector could be saturated for ^{13}C enriched compounds (^{13}C >5%) (43), we made a tenfold dilution of the samples with naturally labeled glucose before measurement (39,40). To determine the ratio of $^{13}C/^{12}C$ in the biomass (reflected by proteinogenic amino acids), a gas chromatograph (Hewlett-Packard model 7890A; Agilent Technologies, CA) and a mass spectrometer (model 5975C; Agilent Technologies, CA) were employed. The sample preparation and isotopic analysis of proteinogenic amino acids were performed as previously described (20,25). The GC-MS was carefully tuned and the measurement errors were determined using naturally labeled amino acids (Supplementary Table 4.4.S1)(44). In a separate ^{13}C-MFA experiment, we used a liquid chromatograph-mass spectrometer (LC-MS) to investigate the isotopic effect. The first metabolite in the glycolysis, hexose monophosphate (mainly glucose 6-phosphate, G6P), was extracted by cold methanol. The labeling pattern was determined using ion pair reverse phase LC-TOF(a Quadrupole Time-of-Flight, Agilent) with electrospray ionization (ESI) in negative mode. The measurement was performed by the Proteomics & Mass Spectrometry Facility at the Donald Danforth Plant Science Center) (45).

The isotopic ratio ($\delta^{13}C$) is a standard parameter for isotopic fractionation and was calculated as:

$$\delta^{13}C = \frac{(^{13}C/^{12}C \text{ metabolite}) - (^{13}C/^{12}C \text{ glucose})}{^{13}C/^{12}C \text{ glucose}} \times 1000 \qquad \text{(Equation 1)}$$

The $^{13}C/^{12}C$ ratio of glucose used in ^{13}C-MFA experiments can be determined by an isotope ratio mass spectrometer or LC-TOF. In this work, the $^{13}C/^{12}C$ ratios in metabolites were calculated from the labeling patterns of metabolites. Due to their associated pathways, each metabolite has a different value for $^{13}C/^{12}C$, given below:

$$^{13}C/^{12}C \text{ metabolite} = \frac{\sum_{i=1}^{C} i \times M_i}{C - (\sum_{i=1}^{C} i \times M_i)} \qquad \text{(Equation 2)}$$

where M_i is the isotopomer fraction for a given metabolite (i.e., M_0 is the non-labeled fraction, M_1 is the singly labeled fraction, M_2 is the doubly labeled fraction, , etc). C is the total number of carbon atoms in the molecule.

Previous research has reported that when a naturally labeled carbon source was digested, the isotopic ratio for key metabolites (including amino acids) resulting from enzymatic conversion of glucose was around -20 (46). When *E. coli* was grown with 20% [U-^{13}C] glucose and 80% non-labeled glucose, the growth curve was similar to that for non-labeled glucose (supplementary Figure 4.4.S1). Based on GC-MS measurement, $\delta^{13}C$ values from key amino acids were all negative (i.e., light isotopes were favored) and significantly above the instrument errors (Figure 4.4.1). The $\delta^{13}C$ values also suggest that pathways may

have different isotope discrimination. Phenylalanine and histidine are derived from the pentose phosphate pathway. Their $\delta^{13}C$ values were between -47 and -45, similar to the $\delta^{13}C$ values of serine and alanine (both of which are derived from glycolysis). On the other hand, leucine has acetyl-CoA as one of its precursors. The oxidation of pyruvate to acetyl-CoA has been reported to have a strong isotopic effect (47), so the leucine synthesis route showed a higher microbial isotopic effect ($\delta^{13}C$ = -79) than that of alanine (converted from pyruvate). Besides, aspartate had the highest isotopic ratio ($\delta^{13}C$ = -92) among all amino acids, because the anaplerotic reaction (PEP + CO_2 → Oxaloacetate) used non-labeled CO_2 for oxaloacetate (the precursor of aspartate) synthesis, leading to additional dilution of the ^{13}C pool of aspartate.

To confirm the GC-MS results, we performed independent ^{13}C experiments and examined the labeling patterns of G6P and glucose in the ^{13}C culture medium. Table 4.4.1 gives the LC-MS-based labeling fraction data for G6P (M_0~M_6, ion m/z=259~265) and glucose (M_0~M_6, ion m/z=179~185). These results indicate that cells selectively metabolize more light glucose than heavy glucose ($\delta^{13}C$=-77±27) during the middle-log phase. The LC-MS-based $\delta^{13}C$ values were close to the $\delta^{13}C$ values obtained by GC-MS (via analysis of amino acid labeling). The G6P data suggests that the selectivity of glucose uptake may also be a reason for isotopic fractionation. In general, heavy glucose (fully labeled, molecular weight=186 Da) has a slower diffusion rate than ^{12}C glucose

(non-labeled, molecular weight=180 Da). Thus the glucose transporter may favor light glucose and cause isotopic fractionation during the glucose uptake process.

Finally, a ^{13}C-MFA model estimated the impact of the isotopic effect on the accuracy of flux calculations. The pathway map of *E. coli* BL21 (DE3) includes the glycolysis, citric acid cycle, pentose phosphate, and anaplerotic pathways (supplementary Fig 4.4.S2). The development of the ^{13}C-MFA has been discussed previously (20). In brief, glucose, acetate, and biomass pools were measured, and glucose uptake rates were normalized to 100 units. The unknown fluxes were determined in an iterative scheme based on the reaction stoichiometry and atom transition routes through the defined metabolic network. The flux optimization was performed using the "fmincon" function in MATLAB (MathWorks). Because local solutions may be found, the model was run 50 times from different initial guesses to generate the solution space for each flux. By using the amino acid labeling data (supplementary Table 4.4.S2), the flux distributions were estimated with three sets of initial glucose-labeling patterns. The first glucose set (case 1) was comprised of 20% fully labeled glucose mixed with 80% non-labeled glucose, which represented the traditional approach for ^{13}C-MFA. The second glucose set (case 2) was adjusted to 19.2% universally labeled glucose mixed with 80.8% non-labeled glucose, which corrected for bias in glucose utilization (i.e., assuming that *E. coli* utilizes relatively more ^{12}C-glucose than ^{13}C-glucose). Such reduced labeling was based on a microbial isotopic fractionation ratio ($\delta^{13}C = -47$). The third glucose set (case 3) was based

on Table 4.4.1, and the labeling distribution was M_0=74%, M_1=5%, M_2=1%, M_5=1% and M_6=18%. Since commercial labeled glucose is not perfectly pure ($M_0 \neq 80\%$, $M_6 \neq 20\%$), the differences in fluxes from case 1 and case 3 represent a common error in conventional GC-MS-based ^{13}C-MFA analysis. The estimated global net flux distributions in the three sets of labeling data showed that the isotopic effect slightly changed the relative fluxes in the central pathways (~1 units) (Supplementary Fig. 4.4.S3). Beside, consideration of the isotopic impurity of glucose improved the precision of the flux calculation (i.e. smaller solution space on case 3 than that on case 1). The flux difference in key pathways between case 1 and 3 could be up to 5 units. Therefore, the isotopic impurity, rather than isotopic fractionation, should be the concern for flux estimation in conventional labeling experiments.

Figure 4.4.1. Isotopic effect in *E. coli* BL21 (DE3) with 20% [U-^{13}C] glucose and 80% non-labeled glucose. The biased isotopic ratio is defined as - δ^{13}C, based on GC-MS measurement. The higher the biased isotopic ratio, the more ^{12}C atoms are preferred in the amino acids synthesis.

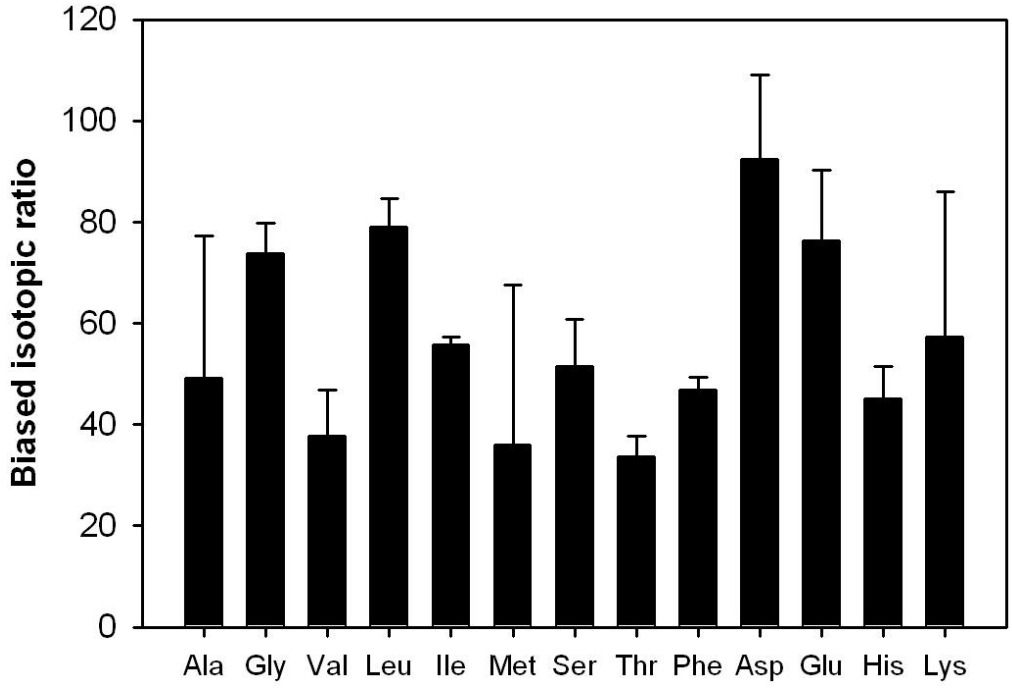


Figure 4.4.S1. Growth curve of *E. coli* BL21 (DE3) with non-labeled glucose (●, solid line) or 20% [U-^{13}C] glucose and 80% non-labeled glucose (○, dot-dashed line).

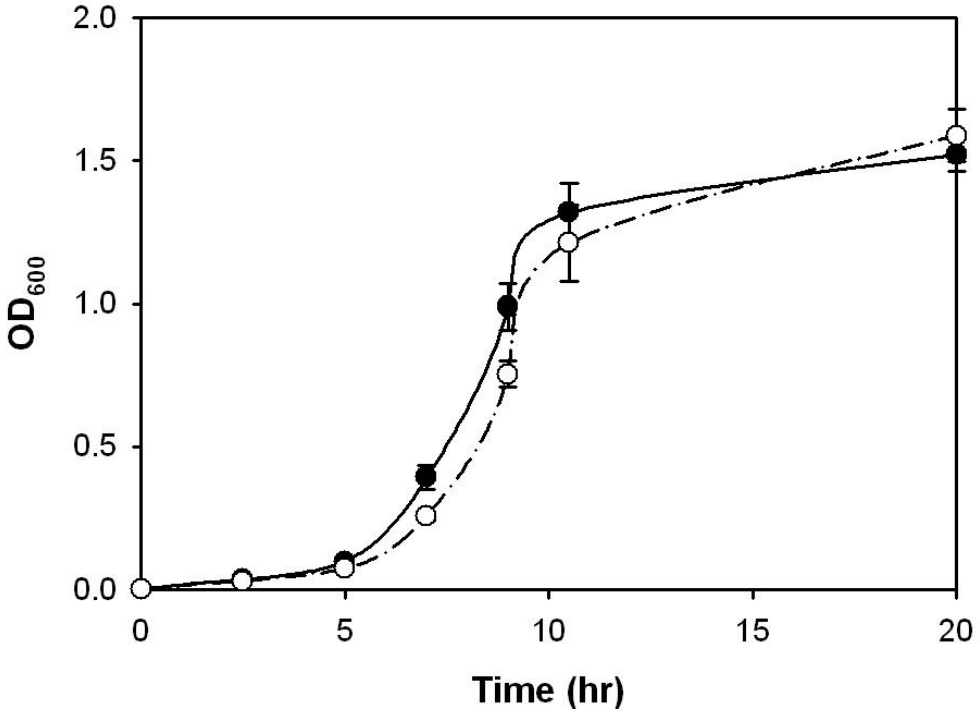


Figure 4.4.S2. Flux analysis of central metabolism in *E. coli* BL21 (DE3). The grey arrows indicate fluxes to biomass. Abbreviations: 6PG, 6-phosphogluconate; ACCOA, acetyl-coenzyme A; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; GAP, glyceraldehyde 3-phosphate; 3PG, 3-phosphoglycerate; CIT, citrate/isocitrate; MAL, malate; OAA, oxaloacetate; AKG, 2-oxoglutarate; PEP, phosphoenolpyruvate; PYR, pyruvate; P5P, ribose-5-phosphate (or xylulose-5-phosphate, or ribulose-5-phosphate); S7P, sedoheptulose-7-phosphate.

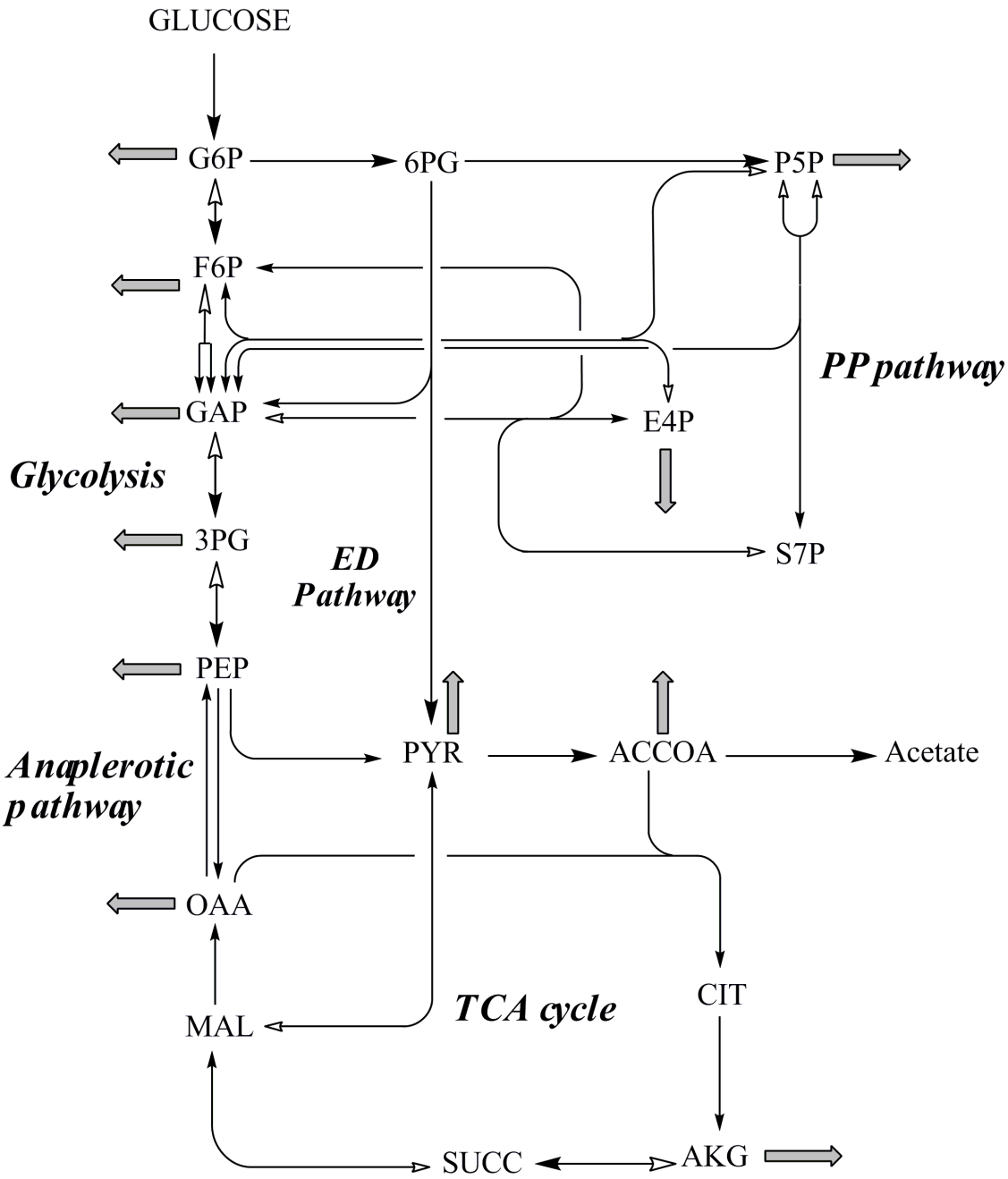

GLUCOSE
G6P
6PG
P5P
F6P
PP pathway
GAP
E4P
Glycolysis
3PG
ED
Pathway
S7P
PEP
PYR
ACCOA
Acetate
Anaplerotic
pathway
OAA
CIT
TCA cycle
MAL
SUCC
AKG

Figure 4.4.S3. Boxplot of metabolic flux results from MATLAB fmincon optimization using 50 random guesses of initial fluxes. Three sets of glucose-labeling patterns were used for flux calculations. 1) not biased: 20% [U-^{13}C] glucose mixed with 80% non-labeled glucose; 2) biased: 19.2% [U-^{13}C] labeled glucose mixed with 80.8% non-labeled glucose; 3) actual: glucose labeling patterns obtained from LC-MS measurement as listed in Table 4.4.1. All the fluxes were normalized by taking the glucose uptake flux as 100 units, while the optimal flux values were marked in the Figure.

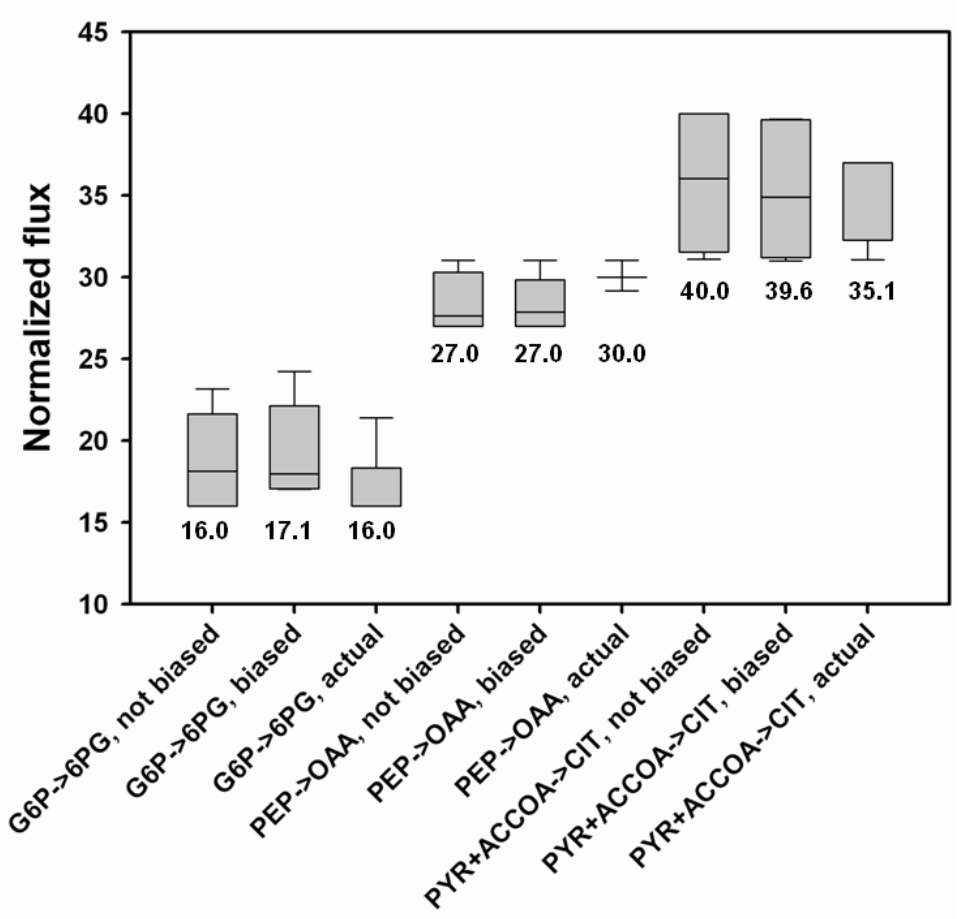


Table 4.4.1. Calculation of $\delta^{13}C$ for G6P derived from LC-MS

Fragmentations	G6P in *E. coli*		Glucose in medium	
	Replicate 1	**Replicate 2**	**Replicate 1**	**Replicate 2**
$\mathbf{M_0}$	0.663	0.655	0.743	0.743
$\mathbf{M_1}$	0.091	0.095	0.051	0.049
$\mathbf{M_2}$	0.054	0.056	0.008	0.012
$\mathbf{M_3}$	0.059	0.066	0.000	0.000
$\mathbf{M_4}$	0.014	0.000	0.000	0.000
$\mathbf{M_5}$	0.000	0.000	0.015	0.013
$\mathbf{M_6}$	0.119	0.128	0.183	0.183
$(^{13}C/^{12}C)$	0.236	0.243	0.260	0.260
	0.240±0.005		0.260±0.000	
$(\delta^{13}C)$	-77			

Note: Two independent culture experiments were performed with five technical repeated measurements. The instrument error was 0.4%. $^{13}C/^{12}C$ was calculated by equation 2; $\delta^{13}C$ was calculated by equation 1.

Table 4.4.S1. GC/MS calibration for eight amino acids fragments.

Ala 260 ($C_{11}H_{26}O_2NSi_2$); $\delta^{13}C$=-20.4

m/z	theory	data	difference
260	100	100	0.0
261	23.2	22.6	-0.6
262	9.5	9.4	-0.1
263	1.5	1.4	-0.1
264	0.3	0.2	-0.1
265	0.0	0.0	0.0
266	0.0	0.0	0.0
267	0.0	0.0	0.0
268	0.0	0.0	0.0

Gly 246 ($C_{10}H_{24}O_2NSi_2$); $\delta^{13}C$=-19.2

m/z	theory	data	difference
246	100	100	0.0
247	22.0	21.4	-0.6
248	9.2	9.1	-0.1
249	1.3	1.3	0.0
250	0.2	0.2	0.0
251	0.0	0.0	0.0
252	0.0	0.0	0.0
253	0.0	0.0	0.0
254	0.0	0.0	0.0

Leu 344 ($C_{17}H_{38}O_2NSi_2$); $\delta^{13}C$=-27.6

m/z	theory	data	difference
344	100	100	0.0
345	30.0	29.3	-0.7
346	11.3	11.2	-0.1
347	2.1	2.0	-0.1
348	0.4	0.3	-0.1
349	0.0	0.1	0.0
350	0.0	0.0	0.0

Ser 390 ($C_{17}H_{40}O_3NSi_3$); $\delta^{13}C$=-20.4

m/z	theory	data	difference
390	100	100	0.0
391	35.1	34.3	-0.8
392	16.3	15.9	-0.4
393	3.8	3.7	-0.1
394	0.9	0.9	0.0
395	0.1	0.1	0.0
396	0.0	0.0	0.0
397	0.0	0.0	0.0
398	0.0	0.0	0.0
399	0.0	0.0	0.0
400	0.0	0.0	0.0

Phe 336 ($C_{17}H_{30}O_2NSi_2$); $\delta^{13}C$=-14.6

m/z	theory	data	difference
336	100	100	0.0
337	29.9	29.3	-0.6
338	11.2	11.0	-0.2
339	2.1	2.1	0.0
340	0.4	0.4	0.0
341	0.0	0.0	0.0
342	0.0	0.0	0.0
343	0.0	0.0	0.0
344	0.0	0.0	0.0

Asp 418 ($C_{18}H_{40}O_4NSi_3$); $\delta^{13}C$=-18.8

m/z	theory	data	difference
418	100	100	0.0
419	36.1	35.1	-1.0
420	16.9	16.6	-0.3
421	4.0	3.9	-0.1
422	1.0	0.9	-0.1
423	0.2	0.1	-0.1
424	0.0	0.0	0.0

425	0.0	0.0	0.0
426	0.0	0.0	0.0
427	0.0	0.0	0.0
428	0.0	0.0	0.0

Glu 432 ($C_{19}H_{42}O_4NSi_3$); $\delta^{13}C$=-27.1

m/z	**theory**	**data**	**difference**
432	100	100	0.0
433	37.4	36.4	-1.0
434	17.4	16.9	-0.5
435	4.3	4.1	-0.2
436	1.0	1.0	0.0
437	0.2	0.1	-0.1
438	0.0	0.0	0.0
439	0.0	0.0	0.0
440	0.0	0.0	0.0
441	0.0	0.0	0.0
442	0.0	0.0	0.0

His 440 ($C_{20}H_{42}O_2N_3Si_3$); $\delta^{13}C$=-24.3 (‰)

m/z	**theory**	**data**	**difference**
440	100	100	0.0
441	39.2	38.3	-0.9
442	17.7	17.2	-0.5
443	4.4	4.2	-0.2
444	1.0	0.9	-0.1
445	0.2	0.1	-0.1
446	0.0	0.0	0.0
447	0.0	0.0	0.0
448	0.0	0.0	0.0
449	0.0	0.0	0.0
450	0.0	0.0	0.0

Note: Table 4.4.S1 determined the instrumental errors of GC-MS using derivatized amino acids (naturally labeled with ^{13}C, ^{18}O, ^{15}N and $^{29,30}Si$). $\delta^{13}C$ values in the table represent the measurement bias for labeled amino acids during ^{13}C-MFA.

Reference: M.R. Antoniewicz, J.K. Kelleher, and G. Stephanopoulos, Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. Anal. Chem. 79 (2007) 7554-9.

Table 4.4.S2. Measured and model fitted isotopomer data from TBDMS-derivatized amino acids

Amino Acids	Fragment	Origin	Isotopomer Enrichment				
			M0	M1	M2	M3	M4
Ala	$[M-57]^+$	Exp	0.78	0.04	0.02	0.17	
		Cal_*Unbiased*	0.77	0.04	0.03	0.17	
		Cal_*Biased*	0.77	0.04	0.03	0.16	
		Cal_*Actual*	0.75	0.05	0.02	0.17	
Gly	$[M-57]^+$	Exp	0.79	0.05	0.16		
		Cal_*Unbiased*	0.79	0.05	0.16		
		Cal_*Biased*	0.79	0.05	0.16		
		Cal_ *Actual*	0.79	0.05	0.16		
Leu	$[M-159]^+$	Exp	0.51	0.15	0.24	0.06	0.03
		Cal_*Unbiased*	0.49	0.16	0.24	0.07	0.03
		Cal_*Biased*	0.51	0.15	0.24	0.06	0.03
		Cal_ *Actual*	0.48	0.18	0.24	0.07	0.03
Ser	$[M-57]^+$	Exp	0.74	0.08	0.05	0.13	
		Cal_*Unbiased*	0.74	0.07	0.06	0.14	
		Cal_*Biased*	0.74	0.07	0.06	0.13	
		Cal_ *Actual*	0.73	0.08	0.05	0.14	
Asp	$[M-57]^+$	Exp	0.61	0.17	0.11	0.09	0.02
		Cal_*Unbiased*	0.60	0.15	0.12	0.10	0.03
		Cal_*Biased*	0.61	0.16	0.11	0.09	0.03
		Cal_ *Actual*	0.63	0.15	0.10	0.10	0.03
Glu	$[M-57]^+$	Exp	0.52	0.16	0.21	0.07	0.02
		Cal_*Unbiased*	0.49	0.16	0.24	0.07	0.03
		Cal_*Biased*	0.50	0.17	0.23	0.06	0.03
		Cal_ *Actual*	0.51	0.19	0.21	0.06	0.02
Phe	$[M-57]^+$	Exp	0.45	0.09	0.12	0.16	0.10
		Cal_*Unbiased*	0.44	0.07	0.13	0.17	0.09
		Cal_*Biased*	0.45	0.08	0.13	0.17	0.09
		Cal_ *Actual*	0.42	0.09	0.12	0.15	0.11
His	$[M-57]^+$	Exp	0.53	0.16	0.10	0.11	0.03
		Cal_*Unbiased*	0.53	0.16	0.10	0.11	0.03
		Cal_*Biased*	0.53	0.16	0.10	0.11	0.03
		Cal_ *Actual*	0.53	0.16	0.11	0.11	0.03

Note:
Cal_*Unbiased* isotopic data are calculated by assuming $\delta^{13}C = 0$.
Cal_*Biased* isotopic data are calculated by assuming $\delta^{13}C = -47$.
Cal_ *Actual* isotopic data are calculated based on the labeling pattern of glucose detected by LC-MS.

4.5 References

1. Zamboni N, Fendt SM, Rühl M, Sauer U (2009) ^{13}C-based metabolic flux analysis. Nat Protoc 4: 878-892.

2. Wahlund TM, Woese CR, Castenholz RW, Madigan MT (1991) A thermophilic green sulfur bacterium from New Zealand hot springs, Chlorobium tepidum sp. nov. Arch Microbiol 156: 81-90.

3. Frigaard N-U, Chew AGM, Li H, Maresca JA, A.Bryant D (2003) Chlorobium tepidum: insights into the structure, physiology, and metabolism of a green sulfur bacterium derived from the complete genome sequence. Photosynth Res 78: 93-117.

4. Eisen JA, Nelson KE, Paulsen IT, Heidelberg JF, Wu M, et al. (2002) The complete genome sequence of Chlorobium tepidum TLS, a photosynthetic, anaerobic, green-sulfur bacterium. Proc Natl Acad Sci U S A 99: 9509-9514.

5. Wahlund T, Tabita F (1997) The reductive tricarboxylic acid cycle of carbon dioxide assimilation: initial studies and purification of ATP-citrate lyase from the green sulfur bacterium Chlorobium tepidum. J Bacteriol 179: 4859-4867.

6. Charnock C, Refseth UH, Sirevåg R (1992) Malate dehydrogenase from Chlorobium vibrioforme, Chlorobium tepidum, and Heliobacterium gestii: purification, characterization, and investigation of dinucleotide binding by dehydrogenases by use of empirical methods of protein sequence analysis. J Bacteriol 174: 1307-1313.

7. Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EK, et al. (2009) Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. Mass Spectrom Rev 28: 362-375.

8. Feng X, Banerjee A, Berla B, Page L, Wu B, et al. (2010) Mixotrophic and photoheterotrophic metabolisms in Cyanothece sp. ATCC 51142 under continuous light. Microbiology In press.

9. Wu B, Zhang B, Feng X, Rubens JR, Huang R, et al. (2010) Alternative isoleucine synthesis pathway in cyanobacterial species. Microbiology 156: 596-602.

10. Risso C, Van Dien SJ, Orloff A, Lovley DR, Coppi MV (2008) Elucidation of an alternate isoleucine biosynthesis pathway in Geobacter sulfurreducens. J Bacteriol 190: 2266-2274.

11. Pingitore F, Tang YJ, Kruppa GH, Keasling JD (2007) Analysis of amino acid isotopomers using FT-ICR MS. Analytical Chemistry 79: 2483-2490.

12. Tang YJ, Meadows AL, Kirby J, Keasling JD (2007) Anaerobic central metabolic pathways in *Shewanella oneidensis* MR-1 reinterpreted in the light of isotopic metabolite labeling. Journal of Bacteriology 189: 894-901.

13. Tang YJ, Martin HG, Dehal PS, Deutschbauer A, Llora X, et al. (2009) Metabolic flux analysis of Shewanella spp. reveals evolutionary robustness in central carbon metabolism. Biotechnol Bioeng 102: 1161-1169.

14. Fischer E, Zamboni N, Sauer U (2004) High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints. Anal Biochem 325: 308-316.

15. Zhao J, Shimizu K (2003) Metabolic flux analysis of *Escherichia coli* K12 grown on 13C-labeled acetate and glucose using GC-MS and powerful flux calculation method. Journal of Biotechnology 101: 101-117.

16. Kleijn RJ, Geertman JM, Nfor BK, Ras C, Schipper D, et al. (2007) Metabolic flux analysis of a glycerol-overproducing *Saccharomyces cerevisiae* strain based on GC-MS, LC-MS and NMR-derived ^{13}C-labelling data. FEMS Yeast Res 7: 216-231.

17. Sonderegger M, Jeppsson M, Hahn-Hägerdal B, Sauer U (2004) Molecular basis for anaerobic growth of Saccharomyces cerevisiae on xylose, investigated by global gene expression and metabolic flux analysis. Appl Environ Microbiol 70: 2307-2317.

18. 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.

19. Tang YJ, Chakraborty R, Martin HG, Chu J, Hazen TC, et al. (2007) Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-NTA. Appl Environ Microbiol 73: 3859-3864.

20. Tang YJ, Hwang JS, Wemmer D, Keasling JD (2007) The *Shewanella oneidensis* MR-1 fluxome under various oxygen conditions. Applied and Environmental Microbiology 73: 718-729.

21. Yang C, Hua Q, Shimizu K (2002) Metabolic flux analysis in *Synechocystis* using isotope distribution from ^{13}C-labeled glucose. Metab Eng 4: 202-216.

22. Tang KH, Feng X, Tang YJ, Blankenship RE (2009) Carbohydrate metabolism and carbon fixation in *Roseobacter denitrificans* OCh114. PLoS One 4: e7233.

23. Howell BF, McCune S, Schaffer R (1979) Lactate-to-pyruvate or pyruvate-to-lactate assay for lactate dehydrogenase: a re-examination. Clin Chem 25: 269-272.

24. Tang KH, Yue H, Blankenship RE (2010) Energy metabolism of Heliobacterium modesticaldum during phototrophic and chemotrophic growth. BMC Microbiol 10: 150.

25. Feng X, Mouttaki H, Lin L, Huang R, Wu B, et al. (2009) Characterization of the Central Metabolic Pathways in Thermoanaerobacter sp. X514 via Isotopomer-Assisted Metabolite Analysis. Appl Environ Microbiol 75: 5001-5008.

26. Wahl SA, Dauner M, Wiechert W (2004) New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnology and Bioengineering 85: 259-268.

27. Tang K-H, Blankenship RE (2010) Diverse metabolic pathways employed by the green sulfur bacterium Chlorobaculum tepidum. Journal of Biological Chemistry Submitted.

28. Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, et al. (1999) Metabolic flux ratio analysis of genetic and environmental modulations of Escherichia coli central carbon metabolism. Journal of Bacteriology 181: 6679-6688.

29. Tang YJ, Sapra R, Joyner D, Hazen TC, Myers S, et al. (2009) Analysis of Metabolic Pathways and Fluxes in a Newly Discovered Thermophilic and Ethanol-Tolerant Geobacillus Strain. Biotechnol Bioeng 102: 1377-1386.

30. Stephanopoulos GN, Aristidou AA, Nielsen J (1998) Metabolic Engineering Principles and Methodologies. San Diego: Academic Press. 75, 120-130 p.

31. Tang YJ, Pingitore F, Mukhopadhyay A, Phan R, Hazen TC, et al. (2007) Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using GC-MS and FT-ICR mass spectrometry. Journal of Bacteriology 189: 940-949.

32. Voet D, Voet JG, Pratt CW (2008) Fundamentals of Biochemistry, 3rd edition, p.586. New Jersey: John Wiley&Sons, Inc.

33. Furdui C, Ragsdale SW (2000) The role of pyruvate ferredoxin oxidoreductase in pyruvate synthesis during autotrophic growth by the Wood-Ljungdahl pathway. J Biol Chem 275: 28494-28499.

34. Seo D, Sakurai H (2002) Purification and characterization of ferredoxin-NAD(P)(+) reductase from the green sulfur bacterium Chlorobium tepidum. Biochim Biophys Acta 1597: 123-132.

35. Hemme C, Fields M, He Q, Deng Y, Lin L, et al. (2011) Correlation of genomic and physiological traits of thermoanaerobacter species with biofuel yields. Appl Environ Microbiol 77: 7998-8008.

36. Tang YJ, Martin HG, Deutschbauer A, Feng X, Huang R, et al. (2009) Invariability of central metabolic flux distribution in *Shewanella oneidensis* MR-1 under environmental or genetic perturbations Biotechnol Prog 25: 1254-1259

37. Zamboni N, Fendt SM, Rühl M, Sauer U (2009) (13)C-based metabolic flux analysis. Nat Protoc 4: 878-892.

38. Suthers PF, Burgard AP, Dasika MS, Nowroozi F, Van Dien S, et al. (2007) Metabolic flux elucidation for large-scale models using ^{13}C labeled isotopes. Metab Eng 9: 387-405.

39. Heinzle E, Yuan Y, Kumar S, Wittmann C, Gehre M, et al. (2008) Analysis of 13C labeling enrichment in microbial culture applying metabolic tracer experiments using gas chromatography–combustion–isotope ratio mass spectrometry. Anal Biochem 380: 202-210.

40. Yuan Y, Yang TH, Heinzle E (2010) 13C metabolic flux analysis for larger scale cultivation using gas chromatography-combustion-isotope ratio mass spectrometry. Metab Eng 12: 392-400.

41. Cleland WW (2003) The use of isotope effects to determine enzyme mechanisms. J Biol Chem 278: 51975-51984.

42. Henn MR, Gleixner G, Chapela IH (2002) Growth-Dependent Stable Carbon Isotope Fractionation by Basidiomycete Fungi: δ13C Pattern and Physiological Process. Appl Environ Microbiol 68: 4956-4964.

43. J. Thomas Brenna, Thomas N. Corso, Herbert J. Tobias, Caimi RJ (1997) High-precision continuous-flow isotope ratio mass spectrometry. Mass Spectrom Rev 16: 227-258.

44. Antoniewicz MR, Kelleher JK, Stephanopoulos G (2007) Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. Anal Chem 79: 7554-7559.

45. Coulier L, Bas R, Jespersen S, Verheij E, Werf MJvd, et al. (2006) Simultaneous Quantitative Analysis of Metabolites Using Ion-Pair Liquid Chromatography−Electrospray Ionization Mass Spectrometry. Anal Chem 78: 6573-6582.

46. Blair N, Leu A, Muñoz E, Olsen J, Kwong E, et al. (1985) Carbon Isotopic Fractionation in Heterotrophic. Microbial Metabolism. Appl Environ Microbiol 1985 50: 996-1001.

47. DeNiro M, Epstein S (1977) Mechanism of carbon isotope fractionation associated with lipid synthesis. Science 197: 261-263.

Chapter 5

Integrating Flux Balance Analysis into Kinetic Models to Decipher the Dynamic Metabolism of *Shewanella oneidensis*

Xueyang Feng[1], You Xu[2], Yixin Chen[2], Yinjie J. Tang[1,*]

1. Department of Energy, Environmental and Chemical Engineering,

2. Department of Computer Science and Engineering, Washington University, St. Louis, MO 63130

*Correspondence to: Tel: 314-935-3441, Fax: 314-935-7211; E-mail: yinjie.tang@seas.wustl.edu

This chapter was previously published in

Feng X, Xu Y, Chen Y, Tang YJ. "Integrating Flux Balance Analysis into Kinetic Models to Decipher the Dynamic Metabolism of Shewanella oneidensis MR-1" PLoS Comput Biol. In press. DOI:10.1371/journal.pcbi.1002376

5.1 Abstract

Shewanella oneidensis MR-1 sequentially utilizes lactate and its waste products (pyruvate and acetate) during batch culture. To decipher MR-1 metabolism, we integrated genome-scale flux balance analysis (FBA) into a multiple-substrate Monod model to perform the dynamic flux balance analysis (dFBA). The dFBA employed a static optimization approach (SOA) by dividing the batch time into small intervals (i.e., ~400 mini-FBAs), then the Monod model provided time-dependent inflow/outflow fluxes to constrain the mini-FBAs to profile the pseudo-steady-state fluxes in each time interval. The mini-FBAs used a dual-objective function (a weighted combination of "maximizing growth rate" and "minimizing overall flux") to capture trade-offs between optimal growth and minimal enzyme usage. By fitting the experimental data, a bi-level optimization of dFBA revealed that the optimal weight in the dual-objective function was time-dependent: the objective function was constant in the early growth stage, while the functional weight of minimal enzyme usage increased significantly when lactate became scarce. The dFBA profiled biologically meaningful dynamic MR-1 metabolisms: 1. the oxidative TCA cycle fluxes increased initially and then decreased in the late growth stage; 2. fluxes in the pentose phosphate pathway and gluconeogenesis were stable in the exponential growth period; and 3. the glyoxylate shunt was up-regulated when acetate became the main carbon source for MR-1 growth.

5.2 Authors' summary

This study integrates two modeling approaches, a Monod kinetic model and genome-scale flux balance analysis, to analyze the dynamic metabolism of an environmentally important bacterium (*S. oneidensis* MR-1). The modeling results reveal that MR-1 metabolism is suboptimal for biomass growth, while MR-1 continuously reprograms the intracellular flux distributions in adaption to nutrient conditions. This innovative dFBA framework can be widely used to investigate transient cell metabolisms in response to environmental variations. Furthermore, the dFBA is able to simulate metabolite-labeling dynamics in ^{13}C-tracer experiments, and thus can serve as a springboard to advanced ^{13}C-assisted dynamic metabolic flux analysis by using labeled proteinogenic amino acids to improve flux results.

Key words: ^{13}C, enzyme usage, genome-scale, Monod, objective functions, suboptimal, TCA cycle

5.3 Introduction

Cell metabolisms are highly dependent on environmental conditions, so the metabolic state often shifts during the cultivation period (1,2,3). Characterizing the transience of metabolic fluxes is important for understanding how cells responded to environmental changes. Bioprocess models (e.g., a Monod-based kinetic model) (4) have been widely applied to predict microbial

dynamics, but they cannot directly obtain the intracellular flux distributions. On the other hand, flux balance analysis (FBA) profiles the rates of enzymatic reactions based on stoichiometric mass balance, knowledge of reaction constraints, and measurements of inflow/outflow fluxes (5,6). As an underdetermined model, FBA requires an objective function (e.g., "maximizing growth rate") for flux calculation. However, since cells may show suboptimal metabolism and reprogram their metabolic fluxes under different environmental conditions, the commonly used objective function is insufficient to describe cell physiologies (7,8,9). Furthermore, FBA assumes steady-state metabolic conditions, and thus is unable to directly analyze the transience of cell metabolism (10,11,12).

This study developed an FBA framework that integrates Monod kinetics and FBA to decipher the dynamic metabolism of MR-1 (Figure 5.1). MR-1 is a facultative anaerobic bacterium, which not only plays an important ecological role in carbon cycling and metal reduction, but also has been widely used for *in situ* bioremediation and microbial fuel cell applications (13,14,15). MR-1 has a diverse carbon utilization capability and shifts its metabolism during batch cultivation (16). MR-1 uses lactate for initial growth and produces acetate and pyruvate. In the late growth stage, MR-1 metabolizes less energy-favorable pyruvate and acetate. To describe such kinetic behavior, we used unsegregated Monod equations to simulate cell growth, lactate utilization, and metabolite

secretion and reuse. The standard Monod model was incorporated into a genome-scale FBA model, iSO783 (17), to formulate the dynamic FBA (dFBA) framework (11), which enabled quantitative predictions of the MR-1 metabolism.

5.4 Results

5.4.1 Monod model

MR-1 growth displayed an apparent lag phase (~7.1 h) in 30 mM lactate medium (0.1% inoculation). By incorporating a time delay function for the lag growth phase, a standard Monod model consisting of four ordinary differential equations was built to describe the extracellular metabolite curves and growth kinetics (Figure 5.2). The parameters of the Monod model were estimated by fitting the experimental data. Table 5.1 indicates that the lactate-based biomass yield was higher than that for either pyruvate or acetate, confirming the preferential utilization of lactate as an energy-favorable carbon substrate for MR-1. Similarly, the lactate-based growth rate (μ_{max}) was much higher than that for either pyruvate or acetate, indicating that lactate was the major carbon substrate for biomass growth at the early growth stage. Table 5.1 lists the rate coefficients (k_{pl}, k_{ap} and k_{al}) for waste products (pyruvate and acetate) synthesis and reuse, which indicates that MR-1 quickly consumed lactate, producing significant metabolic overflows to the waste products. Such a strategy illustrates an advantageous ecological niche for MR-1 in competing for favorable carbon sources. Finally, although our standard Monod model reasonably well described

MR-1 growth data, its results showed some lack-of-fit with statistical analysis (Table 5.S1). Such a discrepancy was possibly due to the model's simplification and to measurement noises. In this study, the kinetic model represents a compromise between complexity and practical simplicity.

5.4.2 Link kinetic model to FBA

To resolve the flux dynamics, the static optimization approach (SOA) divided the cultivation phase into numerous pseudo-steady states so that a conventional genome-scale MR-1 framework (iSO783, containing 774 reactions and 634 metabolites) was able to calculate the flux distributions (17) in each five-minute time interval. Such dFBA model consisted of ~400 mini-FBAs. To avoid repeated and tedious measurements of biomass and metabolite concentrations for each mini-FBA, we used the Monod model to determine the inflow/outflow fluxes of lactate, acetate, and pyruvate in each time interval. The mini-FBAs could be resolved by an objective function of "maximizing growth rate", but this function severely overestimated the actual biomass growth (Figure 5.3). To account for the suboptimal metabolic features (7,8,9), we used a dual-objective function in dFBA: a combination of "maximize growth rate" and "minimize overall flux". By appropriately weighing both objectives, we explored the trade-offs between optimal cell growth and minimal enzyme usage. Specifically, the Monod model determined the transient growth rate for each time interval, which tuned the weights in the dual-objective functions for the mini-FBAs so that the biomass

growth curve simulated by dFBA was in agreement with experimental observations. Figure 5.3 showed that the optimal dual-objective function in mini-FBAs was time-dependent. In general, these dual-objective functions were invariable before the carbon substrate switched from lactate to acetate/pyruvate. When lactate became scarce, the weight of "minimizing overall flux" in the objective function increased significantly, indicating an intracellular reduction of enzyme synthesis and minimization of intracellular fluxes.

5.4.3 Dynamic flux distributions in MR-1

The dynamic flux distributions in MR-1 were calculated using the bi-level optimization (Figure 5.4). The carbon flows to extracellular acetate and pyruvate were high when lactate was sufficient (~33% and ~25% of the lactate uptake flux before the carbon substrate switch, respectively). Fluxes into the gluconeogenesis pathway, reductive PP pathway, and ED pathway were mainly for biomass synthesis, and remained approximately constant during the exponential growth phase. In the middle log phase (22~25 hours), when the growth rate reached the maximum, fluxes in the oxidative TCA cycle reached a peak (e.g., ~6 mmol/g DCW/h for succinyl-CoA synthetase) to generate energy and building blocks. When lactate became scarce (25~30 hrs), MR-1 had to utilize its waste metabolites (acetate and pyruvate). During this metabolic shift, most intracellular fluxes started to decrease. In the late log growth phase (30~34 hrs), it was also observed that the glyoxylate shunt was up-regulated compared to TCA cycle

fluxes after acetate became the main carbon source for MR-1 growth. The glyoxylate shunt reduced the oxidation of carbon substrate for CO_2 production by diverting the carbon flow into a shorter metabolic route. The glyoxylate shunt activity was further confirmed by *in vitro* enzyme assays at both the mid-log phase (malate synthase activity was 0.18±0.11 μmol/g DCW/min) and the late-log phase (malate synthase activity was 0.37±0.17 μmol/g DCW/min).

5.4.4 Simulation of dynamic ^{13}C-labeling in proteinogenic amino acids

In ^{13}C-labeled tracer experiments, dFBA can be used to predict the isotopomer dynamics in slow turnover metabolites, such as proteinogenic amino acids. During MR-1 growth with [3-^{13}C] lactate, the dynamic metabolism led to variations of labeling patterns in intracellular metabolites (biomass precursors) so that the isotopic labeling in proteinogenic amino acids was continuously changing during cell growth (18). Here, we simulated the time-integrative labeling patterns in proteinogenic amino acids based on fluxes from dFBA. The predicted isotopomer labeling patterns of five proteinogenic amino acids (Ala, Ser, Glu, Asp, and Gly, at t = 24 and 30 h) are illustrated in Figure 5.5A and Figure 5.S1. Compared with the experimental measurements, the labeling patterns predicted by dFBA are consistent with the measured labeling patterns, but some lack-of-fit still exists. One of the limitations of the FBA model is that the intracellular pathway is treated as unidirectional, so the effect of exchange fluxes on isotopomer data is neglected. Considering that some *in vivo* reactions could be bi-directional, we

implemented exchange coefficients for four pathways (e.g., the anaplerotic pathway: pyruvate → malate) in the model to improve the simulation of ^{13}C-labeling (Table 5.2). After introduction of the exchange coefficients, the measured and the simulated isotopomer data for proteinogenic amino acids matched (R^2 = 0.9619, Figure 5.5B).

5.5 Discussion

dFBA models have been developed to describe the dynamic metabolism of *E.coli* (10), *Saccharomyces cerevisiae* (19), *Lactococcus lactis* (12), and even for a more complicated coculture system of *E.coli* and *Saccharomyces cerevisiae* (20). In this study, we developed dFBA for analyzing metabolic states of *S. oneidensis* MR-1. The time-dependent inflow/outflow fluxes for dFBA can be constrained by either a Monod model or other empirical models (such as polynomial-fitting to the measurement data (21)). The Monod model is suitable to uncover kinetic properties of a scale-up bioprocess and empowers the dFBA to correlate the bioprocess parameters (such as nutrient concentrations and inhibition coefficients) with intracellular metabolism analysis. The integration of the Monod model and dFBA can decipher and predict cell metabolisms in response to batch fermentation conditions.

To describe biological realities, a physiologically reasonable objective function is important for FBA. For *E.coli* metabolism, 11 objective functions have been systemically investigated under different cultivation conditions (7). It turns

out that no single objective function can describe metabolic states accurately for all conditions. A recent study of MR-1 indicated that futile cycles may be operational, in which less energetically efficient enzymes are expressed at higher levels than their counterparts and decrease biomass yield (17). Such suboptimal metabolic features in MR-1 make the conventional objective functions difficult to use in predicting actual cell physiology. To bridge the gap between the *in silico* simulations and experimental observations, we assigned dual-objective functions to resolve mini-FBAs. Using dual-objective functions, dFBA accurately predicted the elevated flux ratio of the glyoxylate shunt (represented by malate synthase activity) to the oxidative TCA cycle (represented by fumarase activity) when acetate started to be used as the main carbon substrate (Figure 5.S2). The up-regulation of the glyoxylate shunt and down-regulation of the oxidative TCA pathways were consistent with a previous ^{13}C-metabolic flux analysis of MR-1 (18). In comparison, this transient metabolic shift in the glyoxylate shunt could not be captured by a single objective function, such as maximal biomass growth. Moreover, our dFBA results showed the weight of the two objective functions remained relatively constant when lactate was sufficient. At the early stage of MR-1 growth, such a pseudo-steady state has been experimentally verified by previous isotopomer-based analysis (18). Under nutrient scarcity conditions, MR-1 metabolism may reduce synthesis and usage of enzymes to achieve a compromise between minimization of general physiological activities and maintenance of essential cellular functions (22).

The dFBA model can also simulate time-dependent isotopomer enrichment in proteinogenic amino acids. In turn, the isotopomer results (Figure 5.5) can be used to validate and improve the dFBA model predictions. For example, our dFBA model predicted low fluxes through malic enzyme during the exponential growth because these pathways may reduce biomass production, while the genetic analysis indicates a high functionality of malic enzyme (17). In the dynamic isotopomer simulations, we found that the fitting of isotopomer labeling patterns was significantly improved by introducing the bi-directional fluxes through the pathway Malate $\leftrightarrow$ CO_2 + Pyruvate, while keeping the net flux minimal. Such reversible reactions suggest metabolic flexibility. The activity of malic enzyme was also confirmed by *in vitro* enzyme assays at both the mid-log phase (malic enzyme activity was 0.90±0.18 μmol/g DCW/min) and late-log phase (malic enzyme activity was 1.73±0.81 μmol/g DCW/min).

Proteinogenic amino acids are abundant in biomass and can easily be measured by GC-MS. Complementing this instrumental data, ^{13}C-metabolic flux analysis (MFA) offer analytic insight into the cell metabolisms in fermentation processes (3,23,24). However, the turnover rate of protein is much slower than that of intracellular metabolites, so ^{13}C-MFA is useful only for analyzing the steady-state central metabolism. To perform ^{13}C-MFA of dynamic flux distributions, the fast turnover metabolites have to be extracted and analyzed at multiple time points (1,3), which requires significant sampling efforts and high-

sensitivity analytical measurement of low-abundance/unstable metabolites. Moreover, the calculation of dynamic fluxes with isotopomer data formulates an inverse nonlinear optimization problem, which is computationally challenging. Due to insufficient methods for analyzing low abundance metabolites (25), as well as limitations in computational algorithms, dynamic ^{13}C-MFA cannot resolve the flux distributions in a large-scale metabolic network. To overcome these difficulties, this study illustrates a proof-of-concept approach that exploits the synergy between proteinogenic-amino-acid-based ^{13}C-MFA and genome-scale dynamic flux balance analysis.

In our dFBA, the Monod model is solved first independently of the FBA. As an alternative approach, we also tested to integrate the kinetic models with FBA (integrative Flux Balance Analysis, iFBA). iFBA simultaneously optimizes the kinetic model parameters and solves the dynamic cell metabolism in MR-1 (Supplementary Text S1). We found that iFBA also requires a dual objective function, a weighted combination of "maximizing growth rate" and "minimizing overall flux", to improve the model accuracy, similar to the dFBA approach (Figure 5.S3 and Table 5.S2). Such results indicate that it is difficult to use a single objective to describe the flux states under all growth conditions, while the time-dependent trade-off objective functions are effective for analyzing the dynamic suboptimal metabolism.

In conclusion, as in other FBA studies, the dFBA framework proposed in this study links macroscopic bioprocess variables (such as nutrient concentrations) to microscopic intracellular metabolism analysis. It predicts metabolic responses under dynamic culture conditions, and reveals the impact of the kinetic parameters (such as μ_{max}) on intracellular flux distributions. Furthermore, dFBA can identify the objective functions that are possibly used by microorganisms in adaption to environmental variations. Finally, by simulating and comparing the isotopomer labeling patterns of different metabolites, the proposed dFBA framework can potentially improve dynamic flux resolutions by incorporating the isotopomer data from labeled proteinogenic amino acids.

5.6 Materials and methods

Culture conditions, analytical methods, and isotopomer analysis. *S. oneidensis* MR-1 (ATCC 70050) was first grown in LB medium in shake flasks overnight. A 0.1% inoculum volume was then cultured into modified MR-1 defined medium (26) in shake flasks (100 mL culture for each of triplicate experiments, shaken at 150 rpm) at 30°C. The initial carbon source was ~30 mM lactate. The growth curve was monitored by dried biomass weight. The concentrations of lactate and acetate in the medium were measured using enzyme kits (r-Biopharm, Darmstadt, Germany). The concentration of pyruvate in the medium was measured with the enzyme assay developed by Marbach and Weli (27).

To analyze the activity of malate synthase and malic enzyme, samples were taken at early middle log phase (biomass of MR-1 ~0.08 g/L) and late log phase (biomass of MR-1 ~0.23 g/L). The harvested cells were centrifuged and re-suspended in 100 mM Tris buffer. The samples were then ultra-sonicated for 5 min to release the enzymes. Malate synthase activity was gauged based on the reaction of CoASH with DTNB (Acetyl-CoA+glyoxylate→Malate+CoASH; CoASH+DTNB→CoA-TNB+TNB), as described by Dixon and Kornberg (28). In general, 20 μL acetyl-CoA (5 mM), 10 μL DTNB (10 mM), 50 μL cell extract, and 500 μL of a solution containing 0.1 M potassium phosphate and 10 mM $MgCl_2$ were mixed with water. The mixture was then added with 20 μL 100 mM glyoxylate. The difference in absorbance at OD_{412} before and after glyoxylate addition reflected the activity of malate synthase, in which one unit ΔOD_{412} corresponded to 70.6 nmol of TNB produced (in a 1 mL reaction solution). Furthermore, the activity of malic enzyme was detected based on increased absorbance at 340 nm due to the reduction of NAD^+ to NADH (29). In brief, 400 μL 250 mM Tris HCl, 20 μL 50 mM $MnCl_2$, 25 μL 40 mM NH_4Cl, 100 μL 1M KCl, 50 μL 20 mM NAD^+, 100 μL 100 mM malate, and 50 μL cell extract were mixed with water (1 mL reaction solution). The change in absorbance at OD_{340} reflected the activity of malic enzyme.

In the labeling experiment, MR-1 was first grown overnight in the LB medium in shake flasks. A 0.1% inoculum volume was then cultured into 100 mL

of modified MR-1 defined medium at 30°C, with the initial carbon source as 30 mM [3-^{13}C] lactate (purity>98%) purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). The biomass was harvested at ~24 h (before lactate was depleted) and ~30 h (after the substrate had switched from lactate to waste products). To analyze the labeling pattern of proteinogenic amino acids, we hydrolyzed the biomass with 6M HCl at 100 ^{o}C. The isotopic analysis of proteinogenic amino acids was performed using a GC-MS based TBDMS method, as previously described (30,31,32). Ions of $[M-57]^{+}$ (unfragmented amino acid) were used for the ^{13}C-simulations (33).

Monod model development. A multiple-substrate Monod model was developed to describe the cell growth, lactate consumption and secretion, and reuse of waste products (acetate and pyruvate).

$$\frac{dLACT}{dt} = (-\frac{X \cdot \mu_L}{Y_{X/L}} - r_{P,L} - r_{A,L}) \cdot S(t - t_L) \quad (1)$$

$$\frac{dACT}{dt} = (r_{A,L} + r_{A,P} - \frac{X \cdot \mu_A}{Y_{X/A}}) \cdot S(t - t_L) \quad (2)$$

$$\frac{dPYR}{dt} = (r_{P,L} - \frac{X \cdot \mu_P}{Y_{X/P}} - r_{A,P}) \cdot S(t - t_L) \quad (3)$$

$$\frac{dX}{dt} = X \cdot (\mu_A + \mu_P + \mu_L - k_e) \cdot S(t - t_L) \quad (4)$$

where *X* is biomass (g DCW/L); *LACT*, *ACT*, and *PYR* are lactate, acetate, and pyruvate concentrations (mmol/L), respectively; μ_L, μ_A, and μ_P are the specific growth rates (h^{-1}) on lactate, acetate, and pyruvate, respectively; k_e is the

endogenous metabolism rate constant (h^{-1}); $Y_{X/L}$, $Y_{X/A}$, and $Y_{X/P}$ are the biomass yield coefficients (g DCW/mol substrate) of lactate, acetate, and pyruvate respectively; $r_{P,L}$ and $r_{A,L}$ are the production rates (mmol/L/h) of acetate and pyruvate from lactate, respectively. $r_{A,P}$ is the production rates (mmol/L/h) of acetate from pyruvate. $S(t-t_L)$ is the dimensionless unit-step time delay function (S=0 when $t < t_L$; S=1 when $t = t_L$) which described the lag phase after inoculation.

The specific cell growth rate was described by Monod equations:

$$\mu_L = \frac{\mu_{max,L} \cdot LACT}{K_{s,l} + LACT} \quad (5)$$

$$\mu_A = \frac{\mu_{max,A} \cdot ACT}{K_{s,a} + ACT} \quad (6)$$

$$\mu_P = \frac{\mu_{max,P} \cdot PYR}{K_{s,p} + PYR} \quad (7)$$

where $\mu_{max,L}$, $\mu_{max,A}$, and $\mu_{max,P}$ are the maximal growth rates (h^{-1}) for fully aerobic growth on lactate, acetate, and pyruvate, respectively; and $K_{s,l}$, $K_{s,a}$, and $K_{s,p}$ are Monod constants (mmol/L) for lactate, acetate, and pyruvate, respectively. The acetate and pyruvate production rates are assumed to be proportional to the biomass (16), as expressed below:

$$r_{A,L} = k_{al} \cdot LACT \cdot X \quad (8)$$

$$r_{P,L} = k_{pl} \cdot LACT \cdot X \quad (9)$$

$$r_{A,P} = k_{ap} \cdot PYR \cdot X \quad (10)$$

where k_{al} and k_{pl} are rate constants of acetate and pyruvate production, respectively (L· (h·g DCW)$^{-1}$). k_{ap} is the rate constant of acetate production from pyruvate (L· (h·g DCW)$^{-1}$).

The kinetic model (Equations 1~10) contained 14 kinetic parameters. To estimate the model parameters, an ordinary least squares (OLS) method (34) was applied to solve the inverse problem. OLS aimed to minimize the residual sum of the squares (R) between model simulations and experimental measurements, expressed as

$$R = [Y - \eta(t;\beta)]^T [Y - \eta(t;\beta)] \quad (11)$$

where η represents four dependent variables simulated by the kinetic model; β represents the vector of the parameters to be estimated; and Y is the vector of the measured value of the dependent variables. Since the scales of the dependent variables were different (e.g., the scale of the biomass measurement was < 1 g/L, while the lactate measurement >10 mmol/L), the direct application of OLS would overemphasize the fitting of dependent variables with large scales. Therefore, we normalized dependent variables by the corresponding maximum concentrations measured in the experiments.

The "ode23" command in MATLAB (R2009a, Mathworks) was used to solve ODEs, and the "fmincon" command was used to find suitable settings of the parameters. Figure 5.S4 is the histogram of normalized model residuals. The

standard deviations of the parameters were derived from a bootstrap analysis, in which the experimental measurements were randomly re-sampled 1000 times and the corresponding parameters were simulated with the same parameter estimation approach. The 1000 re-sampling was found to be adequate since the variation of parameters converged to within a desired tolerance of 0.1%.

Bi-level dFBA study. The growth phase was divided into 408 pseudo-steady-state intervals with instantaneous transitions between the two adjacent intervals (11). At each pseudo-steady state (~five minutes) (35), a mini-FBA was formulated by a dual-objective function comprised of "maximizing the growth rate" and "minimizing overall flux"; and inflow/outflow fluxes (for lactate, acetate, and pyruvate) derived from the Monod model. The inflow/outflow fluxes were calculated from:

$$\frac{dLACT}{dt} = -v_{lac\_inflow} \cdot X \quad (12)$$

$$\frac{dACT}{dt} = v_{act\_outflow} \cdot X \quad (13)$$

$$\frac{dPYR}{dt} = v_{pyr\_outflow} \cdot X \quad (14)$$

where v_{lac_inflow}, $v_{act\_outflow}$, and $v_{pyr\_outflow}$ are transient lactate inflow flux, acetate outflow flux, and pyruvate outflow flux, respectively.

At each pseudo steady state, the mini-FBA followed a bi-level optimization framework similar to ObjFind (36). The internal optimization was an

FBA with a combined objective function in which the weighting factor of "minimizing overall flux" ranged from zero to one. The difference between the transient growth rate simulated from the FBA and that derived from the Monod model was minimized in the external optimization, by tuning the weighting factor in the combined objective function. The bi-level optimization determined a trade-off between maximizing growth rate and maximizing enzyme efficiency under the specified growth conditions. The bi-level optimization was formulated as:

$$
\begin{aligned}
&\min\ (\mu_{\text{Monod}} - \mu_{\text{FBA}})^2 \\
&s.t. \\
&\left[\begin{array}{l}
\min\ [w \cdot \sum_i v_i^2 - (1-w) \cdot \mu_{FBA}] \\
s.t. \\
\quad S \cdot v = 0 \\
\quad lb \le v \le ub \\
\quad v_{lac\_\inf low},\ v_{act\_outflow},\ v_{pyr\_outflow} \text{ from the Monod model}
\end{array}\right] \\
&0 \le w \le 1
\end{aligned}
\qquad (15)
$$

where μ_{monod} and μ_{FBA} are transient growth rates derived from the Monod model and the dFBA study, respectively; w is the weight of "minimizing overall flux" in the combined objective function; v is the vector of the intracellular fluxes; S is the stoichiometry matrix; lb and ub are the lower and upper boundaries for intracellular flux.

The internal optimization was a typical quadratic programming (QP) problem and was solved using the CPLEX solver in the TOMLAB optimization toolbox (TOMLAB Optimization Inc, Seattle, WA) within MATLAB (R2009a).

The external optimization problem (i.e., search for weight) was solved by a grid search. Since the QP problem in this study was convex, the locally searched results were also the global solutions (37).

Simulation of dynamic ^{13}C-labeling in proteinogenic amino acids. Our previous ^{13}C-MFA study of MR-1 showed that the labeling patterns in its proteinogenic amino acids changed during the late-stage of batch growth (18). Using the dynamic flux distributions from dFBA, we could now simulate dynamics of isotopomer labeling patterns in proteinogenic amino acids using the algorithm below.

$$
\begin{aligned}
&\min \sum_{i=1}^{5} \frac{(MDV_{i,\exp}\big|_{t=24h} - MDV_{i,sim}\big|_{t=24h})^2}{\sigma_i^{\,2}} + \sum_{i=1}^{5} \frac{(MDV_{i,\exp}\big|_{t=30h} - MDV_{i,sim}\big|_{t=30h})^2}{\sigma_i^{\,2}} \\
&s.t.\ \ v_{exch} = \frac{exch}{1-exch} \\
&\quad 0 \le exch \le 1 \\
&\quad f \in [v,\ v_{exch}] \\
&\quad IDV(t) = \frac{\sum_{i=1}^{n} f_i(t) \cdot (IMM_{k1} \cdot IDV_{k1} \otimes IMM_{k2} \cdot IDV_{k2} \cdots)}{\sum_{i=1}^{n} f_i(t)} \\
&\quad IDV_{\text{int}}\big|_{t=t'} = \frac{\int_0^{t'} p(t) \cdot IDV(t) \cdot \Delta t \cdot \Delta X(t)}{\int_0^{t'} p(t) \cdot \Delta t \cdot \Delta X(t)} \\
&\quad MDV_{i,sim}\big|_{t=24h} = M \cdot IDV_{\text{int}}\big|_{t=24h} \\
&\quad MDV_{i,sim}\big|_{t=30h} = M \cdot IDV_{\text{int}}\big|_{t=30h}
\end{aligned}
\qquad (16)
$$

Since FBA neglects flux reversibility, we implemented exchange coefficients *exch* to account for the reversibility of four key metabolic pathways (ME2, GHMT, GLYCL, and SUCOAS in iSO783, Table 5.2). In Equation 16, v_{exch} is the vector of exchanged fluxes in the reversible reactions; *v* is the vector of the transient fluxes at the *t'* interval, simulated from the mini-FBA; *p(t)* are fluxes to proteinogenic amino acids at each time interval; Δt is the scale of a pseudo steady state (5 min); $\Delta X(t)$ is the biomass produced at each time interval; *IMM* is the isotopomer mapping matrices describing the carbon atoms transitions from reactants to products in a reaction; *IDV(t)* are the isotopomer distribution vectors of transient intracellular metabolites at each time interval, which is calculated based on the different labeling patterns of precursors from *n* pathways; $IDV_{int}|_{t=t'}$ are the isotopomer distribution vectors in proteinogenic amino acids at the end of the *t'* time interval; $MDV_{i,sim}$ and $MDV_{i,exp}$ are the mass distribution vectors for each of the five proteinogenic amino acids, as simulated and as measured by GC-MS, respectively; *M* is the matrix for converting *IDV* to *MDV*; σ_i is the standard deviation of the GC-MS measurement (error ~2%), which is assumed to be constant in this study. The carbon transitions in the reactions involved were given in Supplementary Text S2. The "fmincon" command in MATLAB was used to find the exchange coefficients, and the "nlparci" command in MATLAB was used to find the asymptotic confidence intervals of the exchange coefficients. These exchange coefficients significantly improved dFBA simulation of the labeling patterns in proteinogenic amino acids.

5.7 Supplementary Text S1: framework of integrative Flux Balance Analysis (iFBA)

We have tested an alternative model framework: iFBA. The dynamic cultivation process was decomposed into numerous pseudo-steady-state time intervals. At each time interval, the inflow/outflow fluxes in the FBA were derived from the Monod equations; while the biomass increase during the time interval was predicted by mini-FBAs using proper objective functions. At the end of each time interval, the predicted biomass increase was incorporated into the Monod equations to estimate the metabolite and substrate concentrations at the next time interval. Then, we obtained the inflow/outflow fluxes of mini-FBA in the next time interval ($t+\Delta t$). To improve the model accuracy, iFBA employed a dual-objective function *w(i)*, a weighted combination of "maximizing growth rate" and "minimizing overall flux". The time-dependent weight in the dual-objective function and the kinetic parameters in Monod equations were determined by minimizing the differences between iFBA predicted MR-1 growth kinetics and the experimentally measured data. The iFBA was formulated as below:

$$\min\ R = [Y' - \eta'(t;\beta)]^T [Y' - \eta'(t;\beta)]$$

s.t.

$$\begin{bmatrix}
X(0),\ LACT(0),\ ACT(0),\ PYR(0) \\
for\ i = 1 : num_timepoint \\
\quad t = i \cdot dt \\
\quad v_{lac_inflow}(i) = f_1(LACT, ACT, PYR) \\
\quad\quad = -[\ \frac{\mu_{\max,L} \cdot LACT(i)}{K_{s,l} + LACT(i)} / Y_{X/L} - k_{pl} \cdot LACT(i) - k_{al} \cdot LACT(i)] \cdot S(t - t_L) \\
\quad v_{act_outflow}(i) = f_2(LACT, ACT, PYR) \\
\quad\quad = [\ - \frac{\mu_{\max,A} \cdot ACT(i)}{K_{s,a} + ACT(i)} / Y_{X/A} + k_{ap} \cdot PYR(i) + k_{al} \cdot LACT(i)] \cdot S(t - t_L) \\
\quad v_{pyr_outflow}(i) = f_3(LACT, ACT, PYR) \\
\quad\quad = [\ - \frac{\mu_{\max,A} \cdot ACT(i)}{K_{s,a} + ACT(i)} / Y_{X/A} + k_{ap} \cdot PYR(i) + k_{al} \cdot LACT(i)] \cdot S(t - t_L) \\
\quad \begin{bmatrix}
\min\ [w(i) \cdot \sum_i v(i)_i^2 - (1 - w(i)) \cdot \mu(i)] \\
s.t. \\
\quad S \cdot v(i) = 0 \\
\quad lb \le v(i) \le ub \\
\quad v_{lac_inflow}(i),\ v_{act_outflow}(i),\ v_{pyr_outflow}(i) \\
\quad w(i) = p_1 \cdot \exp(p_2 \cdot S(t - p_3))
\end{bmatrix} \\
\quad X(i+1) = X(i) + \mu(i) \cdot X(i) \cdot dt \\
\quad LACT(i+1) = LACT(i) + f_1 \cdot X(i) \cdot dt \\
\quad ACT(i+1) = ACT(i) + f_2 \cdot X(i) \cdot dt \\
\quad PYR(i+1) = PYR(i) + f_3 \cdot X(i) \cdot dt
\end{bmatrix}$$

$$lb \le \beta,\ p_1,\ p_2 \le ub$$

The new symbols introduced in iFBA are as following: *num_timepoint* is the number of time intervals decomposed during the entire cultivation process, which is 408; *dt* is the time of each time interval, which is 1/12 h. p_1, p_2 and p_3 are

three parameters used to simulate the dynamic weighting factors in the dual objective function. The internal dFBA problem was solved using the CPLEX solver in TOMLAB optimization toolbox (TOMLAB optimization Inc, Seattle, WA) within MATLAB (R2009a). The external optimization problem (i.e. search for weight) was solved by SNOPT solver in TOMLAB optimization toolbox within MATLAB (R2009a). The histogram of normalized residuals in growth kinetics simulated by iFBA was shown in Figure 5.S5. The Lack-of-fit test for iFBA was shown in Table 5.S3.

5.8 Supplementary Text S2: reactions involved in ^{13}C-labeling simulations

TCA cycle and metabolites transport

J1 (L-LACD2) LAC (abc)→pyr(abc)

J2 (-ACt6) accoa(ab)↔ACT (ab)

J3 (-PYRt2) pyr(abc)↔PYR(abc)

J4 (PDH) pyr(abc)→accoa(bc)+co2(a)

J5 (CS) oaa(abcd)+accoa(ef)→cit(dcbfea)

J6 (ACONT) cit(abcdef)→icit(abcdef)

J7 (ICDHy) icit (abcdef)→akg(abcde)+co2(f)

J8 (AKGD) akg(abcde)→succoa(bcde)+co2(a)

J9 (SUCD7) succ(1/2abcd+1/2dcba)→fum(1/2abcd+1/2dcba)

J10 (FUM) fum(1/2abcd+1/2dcba)→mal-L(abcd)

J11 (MDH) mal-L(abcd)→oaa(abcd)

J12 (PPC) pep(abc)+co2(d)→oaa(abcd)

J13 (PPCK) oaa(abcd)→pep(abc)+co2(d)

Glyoxylate shunt

J14 (ICL) icit(abcdef)→glx(de)+succ(1/2abcf+1/2fcba)

J15 (MALS) accoa(ab)+glx(cd)→mal-L(dcab)

Reversible net fluxes and C1 metabolism

J16 (GHMT) ser-L(abc)↔glx(ab)+mlthf(c)

J17 (GLYCL) glx(ab)↔co2(a)+mlthf(b)

J18 (ME2) mal-L(abcd)↔pyr(abc)+co2(d)

J19 (-SUCOAS) succoa(abcd)↔succ(1/2abcd+1/2dcba)

Gluconeogenesis

J20 (PPS) pyr(abc)→pep(abc)

J21 (-ENO) pep(abc)→2pg(abc)

J22 (-PGM) 2pg(abc)→3pg(abc)

Amino acid biosynthesis

J23 3pg(abc)→ser-L(abc)

J24 akg(abcde)→glu-L(abcde)

J25 pyr(abc)→ala-L(abc)

J26 oaa(abcd)→asp-L(abcd)

Notes:

1) All abbreviations are referred to iSO783, except LAC, PYR and ACT; which represent extracellular lactate, pyruvate and acetate, respectively.

2) The reaction IDs in iSO783 are listed with brackets. The negative sign indicates that the net flux of the pathway is in the opposite direction as set by iSO783. The amino acids biosynthesis pathways are lumped; hence no ID in iSO783 is available.

5.9 References

1. Wahl SA, Nöh K, Wiechert W (2008) ^{13}C labeling experiments at metabolic nonstationary conditions: an exploratory study. BMC Bioinformatics 9: 152.

2. Wiechert W, Nöh K (2005) From stationary to instationary metabolic flux analysis. Adv Biochem Eng Biotechnol 92: 145-172.

3. Rühl M, Zamboni N, Sauer U (2010) Dynamic Flux Responses in Riboflavin Overproducing Bacillus subtilis to Increasing Glucose Limitation in Fed-Batch Culture. Biotechnol Bioeng 105: 795-804.

4. Kovárová-Kovar K, Egli T (1998) Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. Microbiol Mol Biol Rev 62: 646-666.

5. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol 60: 3724-3731.

6. Orth JD, Thiele I, Palsson BØ (2010) What is flux balance analysis? Nat Biotechnol 28: 245-248.

7. Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. Mol Syst Biol 3: 119.

8. Burgard AP, Pharkya P, Maranas CD (2003) Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. Biotechnol Bioeng 84: 647-657.

9. Segrè D, Vitkup D, Church G (2002) Analysis of optimality in natural and perturbed metabolic networks. Proc Natl Acad Sci U S A 99: 15112-15117.

10. Meadows AL, Karnika R, Lama H, Forestella S, Snedecor B (2010) Application of dynamic flux balance analysis to an industrial Escherichia coli fermentation. Metab Eng 12: 150-160.

11. Mahadevan R, Edwards JS, Doyle FJ (2002) Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J 83: 1331-1340.

12. Oddonea GM, Millsb DA, Block DE (2009) A dynamic, genome-scale flux model of Lactococcus lactis to increase specific recombinant protein expression. Metab Eng 11: 367-381.

13. Myers CR, Nealson KH (1988) Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. Science 240: 1319-1321.

14. Tiedje JM (2002) *Shewanella*- the environmentally versatile genome. Nature Biotechnology 20: 1093-1094.

15. Logan BE, Murano C, Scott K, Gray ND, Head IM (2005) Electricity generation from cysteine in a microbial fuel cell. Water Research 39: 942-952.

16. Tang YJ, Meadows AL, Keasling JD (2007) A kinetic model describing *Shewanella oneidensis* MR-1 growth, substrate consumption, and product secretion. Biotechnol Bioeng 189: 894-901.

17. Pinchuk GE, Hill EA, Geydebrekht OV, Ingeniis JD, Zhang X, et al. (2010) Constraint-based model of Shewanella oneidensis MR-1 metabolism: a tool for data analysis and hypothesis generation. PLoS Comput Biol 6: e1000822.

18. Tang YJ, Martin HG, Dehal PS, Deutschbauer A, Llora X, et al. (2009) Metabolic flux analysis of Shewanella spp. reveals evolutionary robustness in central carbon metabolism. Biotechnol Bioeng 102: 1161-1169.

19. Hjersted JL, Henson MA, Mahadevan R (2007) Genome-scale analysis of Saccharomyces cerevisiae metabolism and ethanol production in fed-batch culture. Biotechnol Bioeng 97: 1190-1204.

20. Hanly T, Henson M (2011) Dynamic flux balance modeling of microbial co-cultures for efficient batch fermentation of glucose and xylose mixtures. Biotechnol Bioeng 108: 376-385.

21. Lequeux G, Beauprez J, Maertens J, Horen EV, Soetaert W, et al. (2010) Dynamic metabolic flux analysis demonstrated on cultures where the limiting substrate is changed from carbon to nitrogen and vice versa. J Biomed Biotechnol 621645.

22. Holzhütter HG (2004) The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. Eur J Biochem 271: 2905-2922.

23. Zamboni N, Fendt SM, Rühl M, Sauer U (2009) ^{13}C-based metabolic flux analysis. Nat Protoc 4: 878-892.

24. Antoniewicz MR, Kraynie DF, Laffend LA, González-Lergier J, Kelleher JK, et al. (2007) 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: 277-292.

25. Reaves M, Rabinowitz J (2011) Metabolomics in systems microbiology. Curr Opin Biotechnol 22: 17-25.

26. Tang YJ, Laidlaw D, Gani K, Keasling JD (2006) Evaluation of the effects of various culture conditions on Cr(VI) reduction by *Shewanella oneidensis* MR-1 in a novel high-throughput mini-bioreactor. Biotechnology and Bioengineering 95: 176-184.

27. Marbach EP, Weli MH (1967) Rapid enzymatic measurement of blood lactate and pyruvate. Use and significance of metaphosphoric acid as a common precipitant. Clin Chem 13: 314-325.

28. Dixon GH, Kornberg HL (1959) Assay methods for key enzymes of the glyoxylate cycle. Biochem J 72.

29. Spina J, Bright HJ, Rosenbloom J (1966) Purification and Properties of L-malic enzyme from Escherichia coli. Biochemistry 9: 3794-3729.

30. Feng X, Banerjee A, Berla B, Page L, Wu B, et al. (2010) Mixotrophic and photoheterotrophic metabolisms in Cyanothece sp. ATCC 51142 under continuous light. Microbiology 156: 2566 - 2574.

31. Feng X, Mouttaki H, Lin L, Huang R, Wu B, et al. (2009) Characterization of the Central Metabolic Pathways in Thermoanaerobacter sp. X514 via

Isotopomer-Assisted Metabolite Analysis. Appl Environ Microbiol 75: 5001-5008.

32. Feng X, Tang K-H, Blankenship RE, Tang YJ (2010) Metabolic flux analysis of the mixotrophic metabolisms in the green sulfur bacterium *Chlorobaculum tepidum*. J Biol Chem 285: 35104-35112.

33. Wahl SA, Dauner M, Wiechert W (2004) New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnology and Bioengineering 85: 259-268.

34. Beck JV, Arnold KJ (1977) Parameter estimation: John Wiley & Sons.

35. Stephanopoulos GN, Aristidou AA, Nielsen J (1998) Metabolic Engineering Principles and Methodologies. San Diego: Academic Press. 75, 120-130 p.

36. Burgard AP, Maranas CD (2003) Optimization-based framework for inferring and testing hypothesized metabolic objective functions. Biotechnol Bioeng 82: 670-677.

37. Boyd S, Vandenberghe L, editors (2004) Convex Optimization: Cambridge University Press.

Figure 5.1. Flowchart of dFBA to decipher the dynamic metabolism of *S. oneidensis* MR-1.

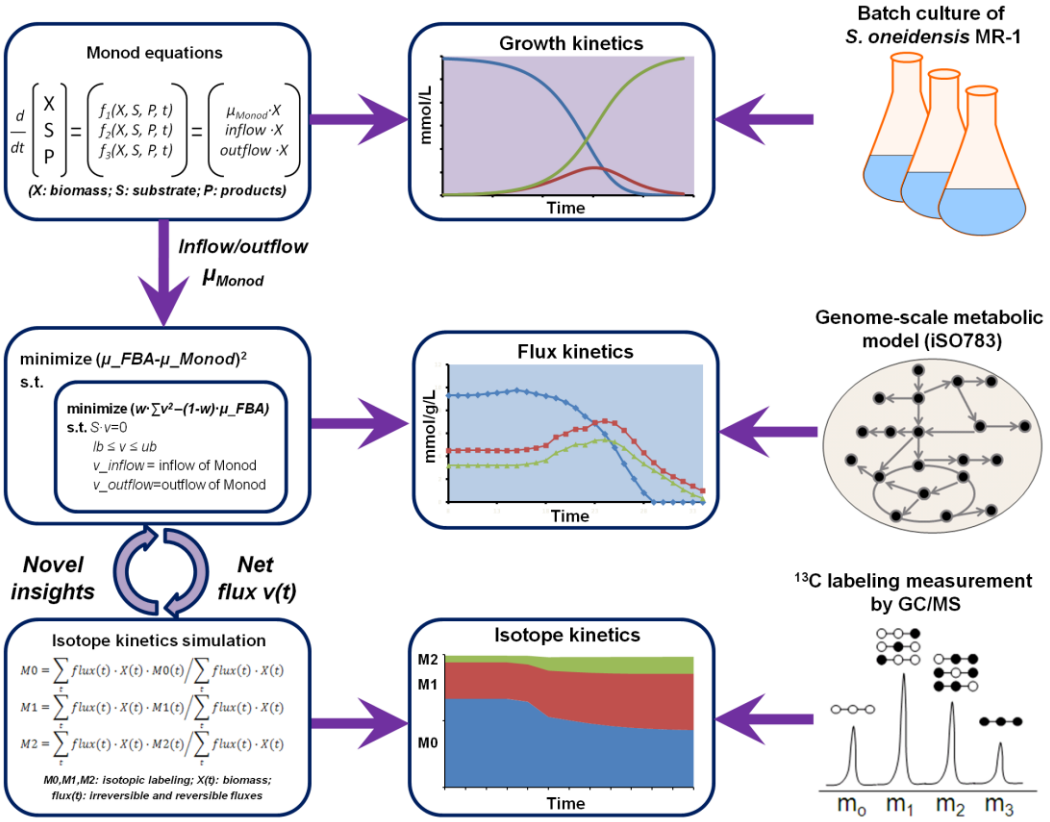


Figure 5.2. Monod model for growth kinetics. The green dots are the measurements, and the blue lines are the simulated growth by the empirical Monod model.

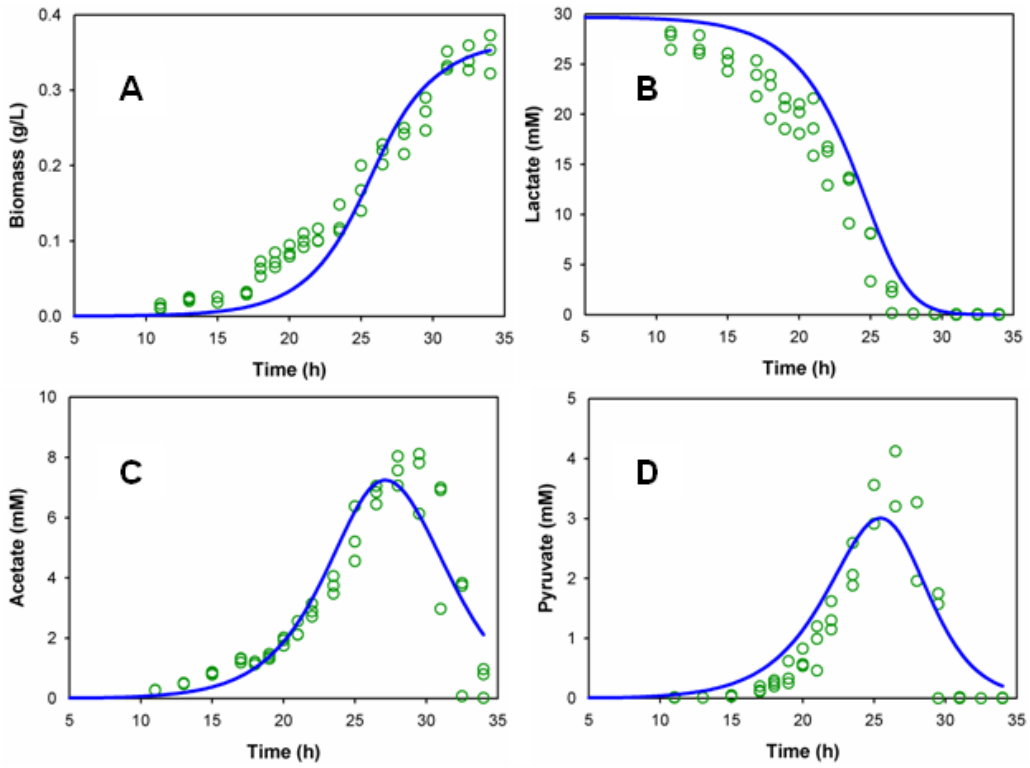


Figure 5.3. Prediction of growth rates (h^{-1}). Blue ○: growth rate determined by the Monod model. Red ○: dFBA prediction using the objective function (maximization of growth rate). Green □: dFBA prediction using dual-objective functions (maximization of growth rate and minimization of overall flux). Yellow ◊: the weight of the dual-objective functions that predicted the measured growth rates. Note: the summation of the square of fluxes ($\sum v^2$) was a very large number (total 774 fluxes), so the magnitude of weight *w* was small.

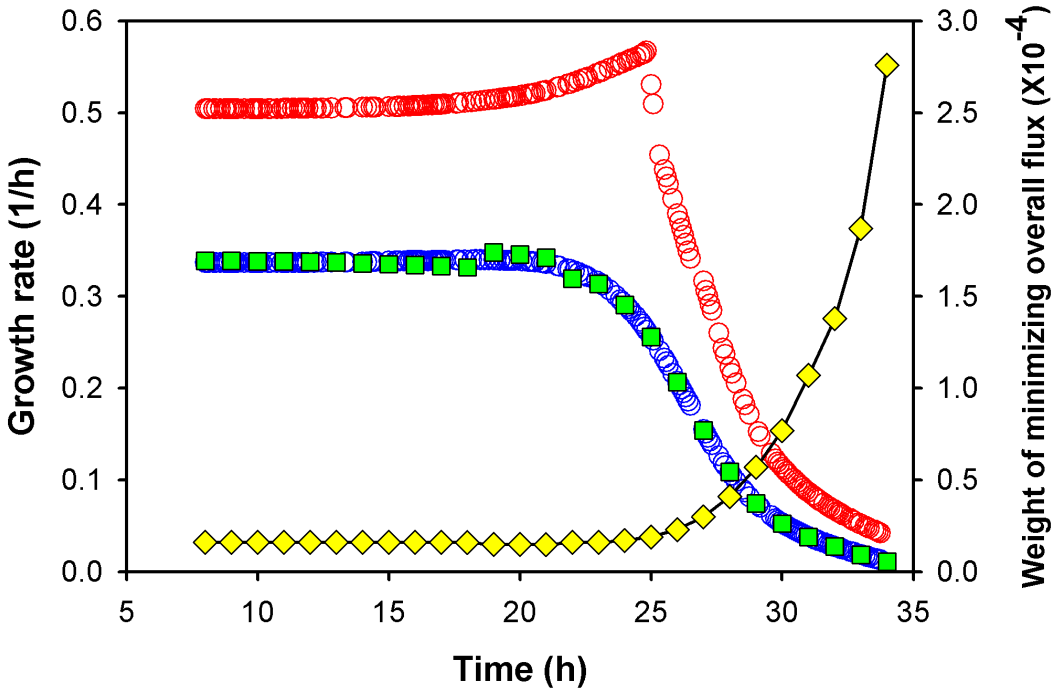


Figure 5.4. Dynamic flux distributions (unit: mmol/g DCW/h) in central metabolic pathways. The yellow filled cycles are intracellular metabolites; the blue filled cycles are substrates and extracellular metabolites (LAC: extracellular lactate, PYR: extracellular pyruvate, ACT: extracellular acetate); the dashed lines indicate inactive pathways; the green filled boxes are reactions listed in iSO783. All the abbreviations refer to iSO783 (7,8,9).

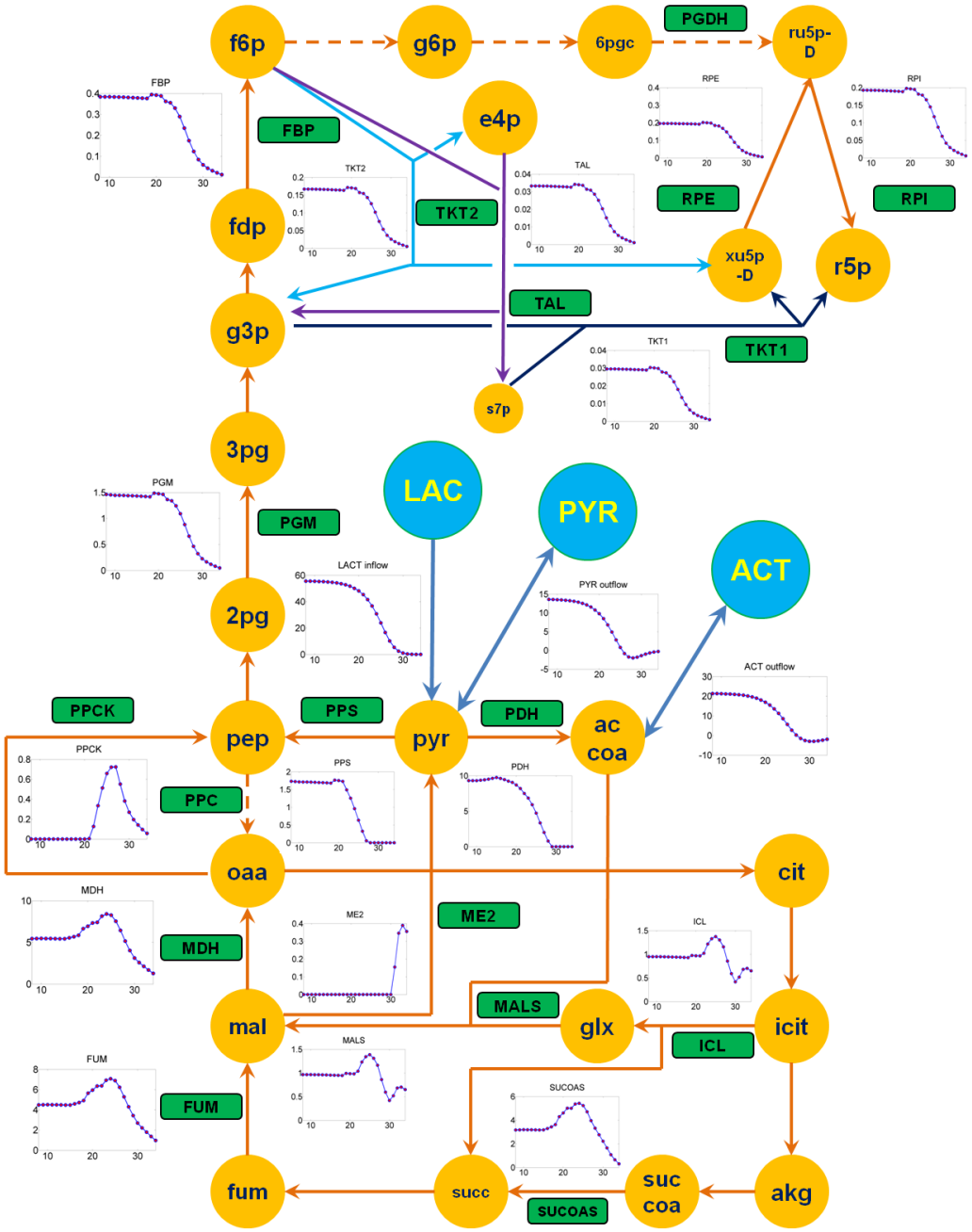

f6p
g6p
6pgc
PGDH
ru5p-D
e4p
FBP
TKT2
TAL
RPE
RPI
fdp
xu5p-D
r5p
g3p
TKT1
s7p
3pg
LAC
PYR
ACT
PGM
2pg
LACT inflow
PYR outflow
ACT outflow
PPCK
PPS
PDH
pep
pyr
ac coa
PPC
oaa
cit
MDH
ME2
ICL
MALS
mal
glx
icit
FUM
fum
succ
SUCOAS
suc coa
akg

Figure 5.5. Experimentally observed and simulated isotopomer labeling patterns $[M-57]^+$ in proteinogenic amino acids. The standard error for GC-MS measurement was below 0.02. **A1**: dynamic isotopomer simulation for glutamate from dFBA without considering reaction reversibility (dFBA w/o reversibility). **A2**: dynamic isotopomer simulation for glutamate from dFBA considering reaction reversibility (dFBA w/ reversibility). Bar plot: comparison of experimentally observed isotopomer labeling to simulated isotopomer labeling patterns of glutamate (**A1**: without considering reaction reversibility; **A2**: considering reaction reversibility). **B**: The model fitting of the isotopomer labeling data of five key amino acids (Ala, Gly, Ser, Asp, and Glu) at t = 24 and 30 h.

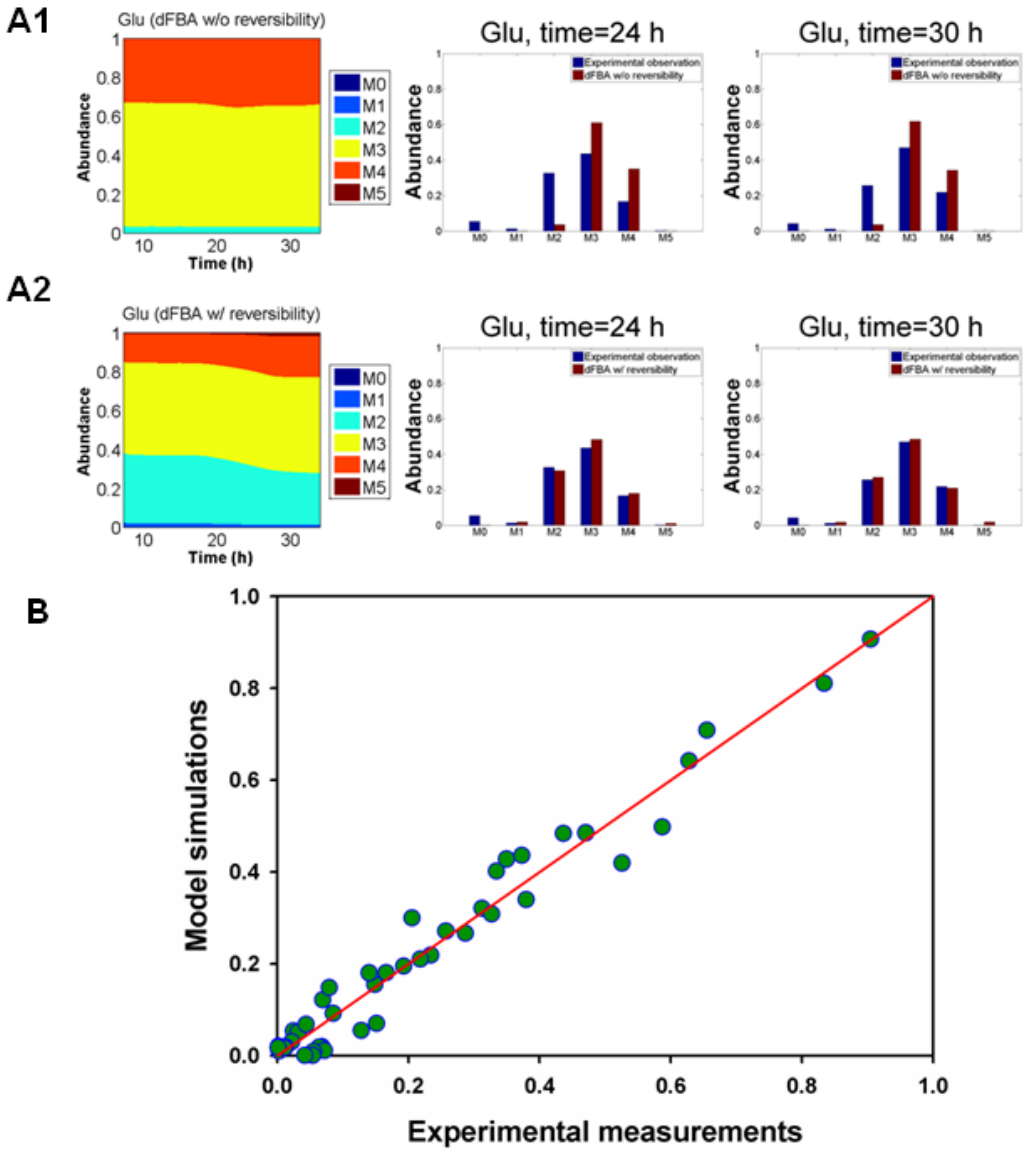
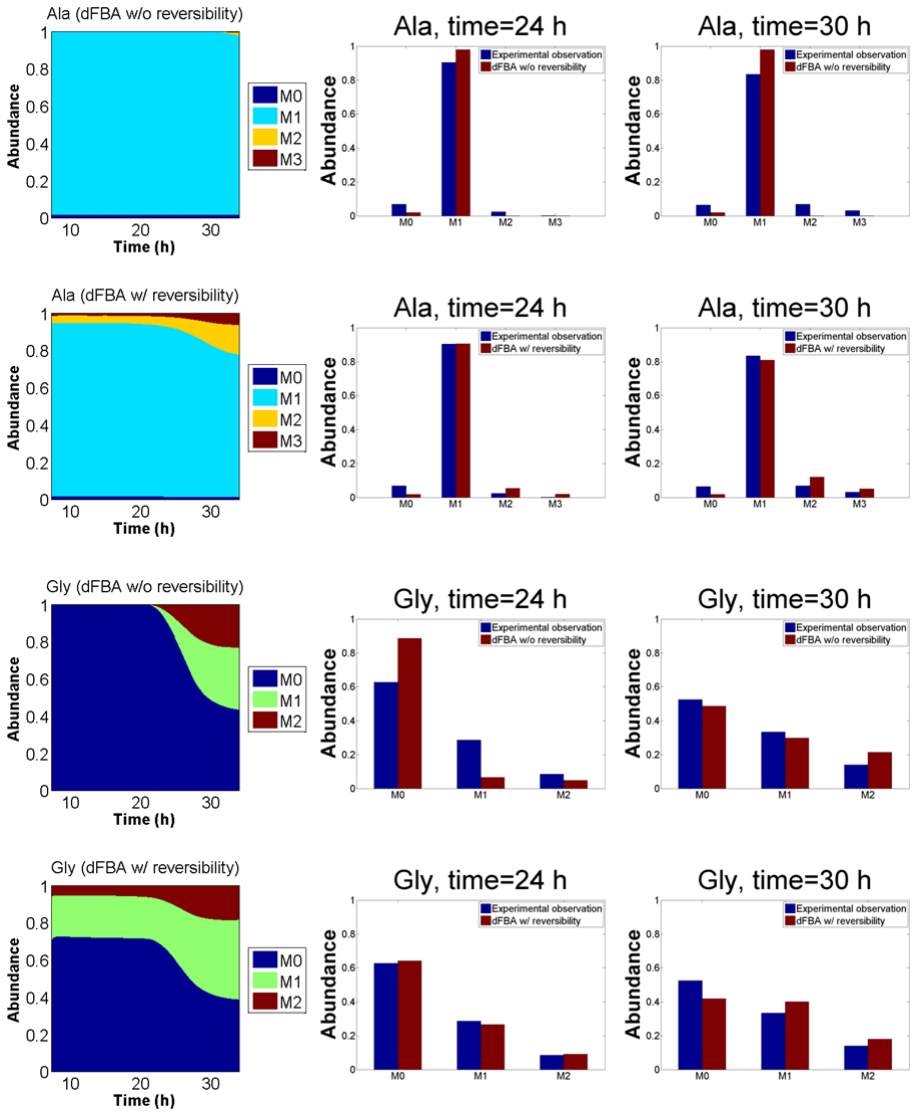

A1
Glu (dFBA w/o reversibility)
Abundance
Time (h)
M0
M1
M2
M3
M4
M5
Glu, time=24 h
Glu, time=30 h
Experimental observation
dFBA w/o reversibility
A2
Glu (dFBA w/ reversibility)
dFBA w/ reversibility
B
Model simulations
Experimental measurements

Figure 5.S1. Experimental observed and simulated isotopomer labeling patterns $[M-57]^+$ in key proteinogenic amino acids. The standard error for GC-MS measurement was ~0.02. Area plot: dynamic isotopomer simulation (case 1: simulation without considering reaction reversibility; case 2: simulation considering reaction reversibility). Bar plot: comparison of experimental data to simulated isotopomer labeling patterns (case 1: without considering reaction reversibility; case 2: considering reaction reversibility).


Ala (dFBA w/o reversibility)
Abundance
Time (h)
M0
M1
M2
M3
Ala, time=24 h
Ala, time=30 h
Experimental observation
dFBA w/o reversibility
Ala (dFBA w/ reversibility)
dFBA w/ reversibility
Gly (dFBA w/o reversibility)
Gly, time=24 h
Gly, time=30 h
Gly (dFBA w/ reversibility)

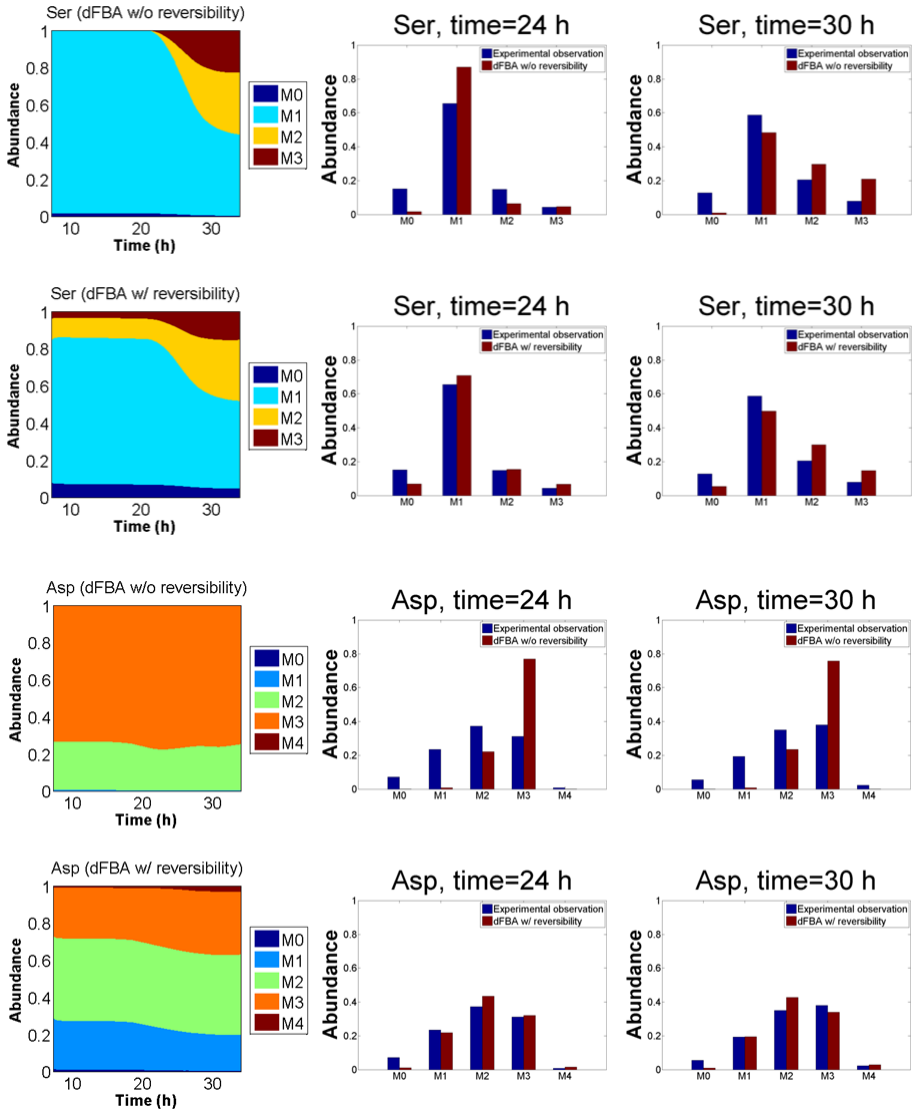

Ser (dFBA w/o reversibility)
Abundance
Time (h)
M0
M1
M2
M3
Ser, time=24 h
Ser, time=30 h
Experimental observation
dFBA w/o reversibility
Ser (dFBA w/ reversibility)
dFBA w/ reversibility
Asp (dFBA w/o reversibility)
M4
Asp, time=24 h
Asp, time=30 h
Asp (dFBA w/ reversibility)

Figure 5.S2. Flux ratio of malate synthase (MALS) and fumarase (FUM) in dynamic metabolism of *Shewanella oneidensis* MR-1. Blue ■: time profiles of flux ratio using "maximizing growth rate" as the objective function in dFBA; red ▲: time profiles of flux ratio using dual-objective function in dFBA: a combination of "maximize growth rate" and "minimize overall flux". The entire growth of MR-1 was divided into three phases. In phase I, lactate was mainly used as the carbon substrate. In phase II, lactate, acetate and pyruvate were used as the carbon substrates. In phase III, acetate was used as the carbon substrate.

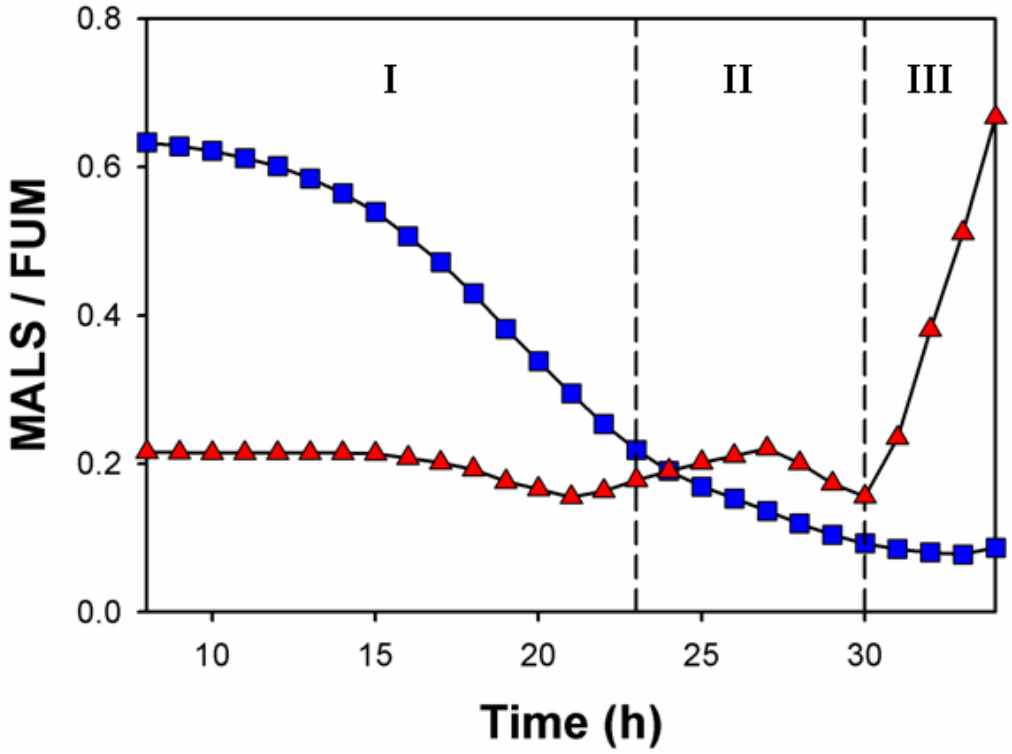


Figure 5.S3. Growth kinetics simulated by iFBA using "maximizing growth rate" as the objective function (red line) or using the dual-objective function (green line). The blue dots are the measurements.

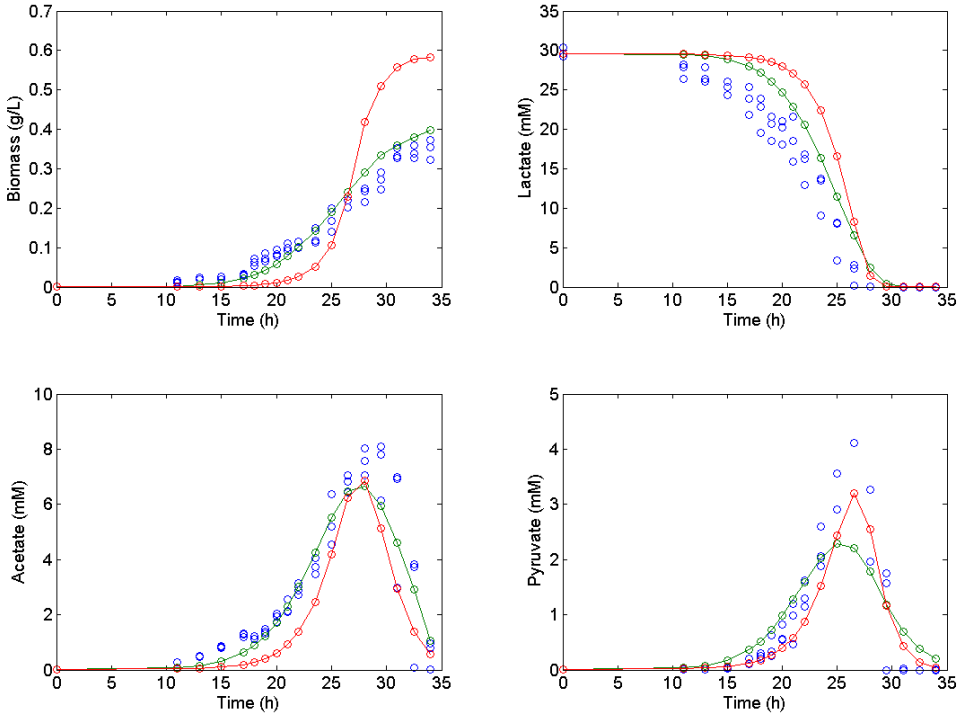


Figure 5.S4. Histogram of normalized Monod model residuals.

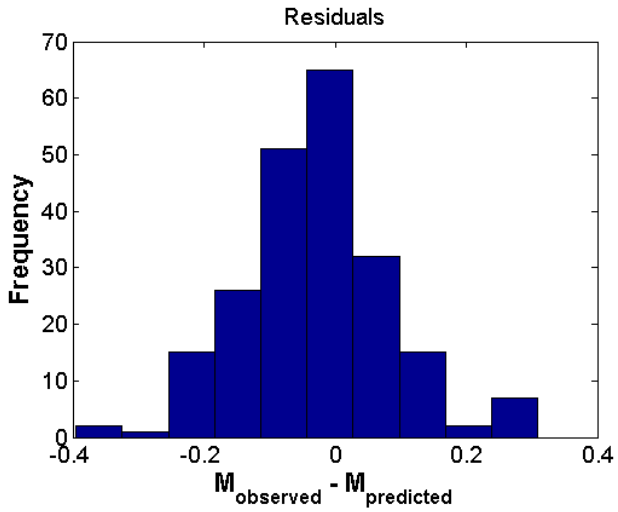


Figure 5.S5. Histogram of normalized residuals in growth kinetics simulated by iFBA.

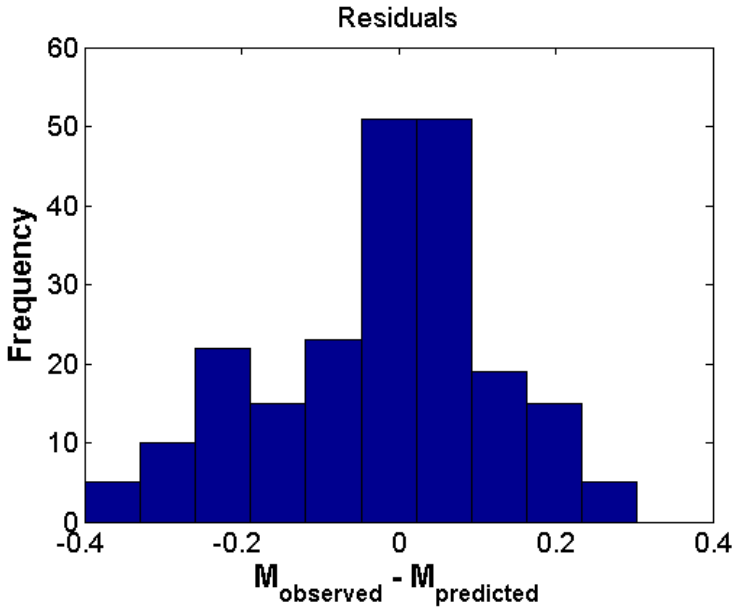


Table 5.1. Parameters estimated in the empirical Monod model

Symbols	Notation	Unit	Value
$\mu_{max,L}$	Maximum specific growth rate using lactate	h^{-1}	0.57±0.11
$\mu_{max,P}$	Maximum specific growth rate using pyruvate	h^{-1}	0.14±0.02
$\mu_{max,A}$	Maximum specific growth rate using acetate	h^{-1}	0.13±0.02
$Y_{X/L}$	Apparent biomass yield coefficient from lactate	g DCW/mol lactate	17.0±1.3
$Y_{X/P}$	Apparent biomass yield coefficient from pyruvate	g DCW/mol pyruvate	16.7±1.3
$Y_{X/A}$	Apparent biomass yield coefficient from acetate	g DCW/mol actate	11.1±4.7
$K_{s,l}$	Monod lactate saturation constant	mM	19.4±7.9
$K_{s,p}$	Monod pyruvate saturation constant	mM	19.4±8.1
$K_{s,a}$	Monod acetate saturation constant	mM	10.1±2.2
k_{al}	Acetate production coefficient from lactate	$L\cdot (h\cdot g\ DCW)^{-1}$	0.71±0.06
k_{pl}	Pyruvate production coefficient from lactate	$L\cdot (h\cdot g\ DCW)^{-1}$	0.45±0.04
k_{ap}	Acetate production coefficient from pyruvate	$L\cdot (h\cdot g\ DCW)^{-1}$	0.94±0.08
k_e	Endogenous metabolism rate constant	h^{-1}	0.013±0.016
t_L	Lag time in growth	h	7.10±0.01

Table 5.2. Exchange coefficients for key metabolic pathways of MR-1

Pathways	Abbreviation	Exchange coefficients	Confidence intervals
Malate ↔ CO_2 + Pyruvate	ME2	0.862	[0.803 0.921]
Serine ↔ Glycine + C1 unit	GHMT	0.270	[0.062 0.477]
Glycine ↔ C1 unit + CO_2	GLYCL	0.109	[0.061 0.157]
Succinate ↔ Succinyl-CoA	SUCOAS	0.944	[0.906 0.983]

Table 5.S1. Lack-of-fit test for the Monod model.

Model Name	**Lack-of-fit sum of squares, SSLOF**	**Degree of freedom, df_1**	**Pure-error sum of squares, SSPE**	**Degree of freedom, df_1**	$\frac{SSLOF / df_1}{SSPE / df\ 2}$	$F(df_1,df_2)$
Monod model	2.064	58	0.813	144	6.303	1.390

Notes:

In order to test whether or not a model could fit the data well, we applied the lack-of-fit test. It partitioned the total sum of squares of residuals in the model (SSE) into two sources of variability: 1) the sum of squares from pure error variability (SSPE) and 2) the sum of squares from lack of fit (SSLOF), i.e. SSE=SSPE+SSLOF.

For testing the null hypothesis H0: the nonlinear model was adequate versus hypothesis Ha: the nonlinear model was inadequate, the test statistic is: $F = \frac{SSLOF / df_1}{SSPE / df\ 2}$. For a specified α-level of significance, we rejected the null hypothesis if $F > F_\alpha$, where F_α is the α-level critical value (α=0.95 in this study) corresponding to an $F(df_1; df_2)$ distribution. The *F*-test indicates our nonlinear model can be further improved by including new parameters. The improved kinetic model is presented in another manuscript under review.

Table 5.S2. Parameters estimated in iFBA

Symbols	Notation	Unit	iFBA
$\mu_{max,L}$	Maximum specific growth rate using lactate	h^{-1}	0.53
$\mu_{max,P}$	Maximum specific growth rate using pyruvate	h^{-1}	0.14
$\mu_{max,A}$	Maximum specific growth rate using acetate	h^{-1}	0.14
$Y_{X/L}$	Apparent biomass yield coefficient from lactate	g DCW/mol lactate	17.5
$Y_{X/P}$	Apparent biomass yield coefficient from pyruvate	g DCW/mol pyruvate	15.5
$Y_{X/A}$	Apparent biomass yield coefficient from acetate	g DCW/mol actate	10.9
$K_{s,l}$	Monod lactate saturation constant	mM	19.4
$K_{s,p}$	Monod pyruvate saturation constant	mM	19.4
$K_{s,a}$	Monod acetate saturation constant	mM	10.1
k_{al}	Acetate production coefficient from lactate	$L\cdot (h\cdot g\ DCW)^{-1}$	0.70
k_{pl}	Pyruvate production coefficient from lactate	$L\cdot (h\cdot g\ DCW)^{-1}$	0.42
k_{ap}	Acetate production coefficient from pyruvate	$L\cdot (h\cdot g\ DCW)^{-1}$	0.94
k_e	Endogenous metabolism rate constant	h^{-1}	0.013
t_L	Lag time in growth	h	7.10
p_1	Parameters used in tradeoff objective function	dimensionless	5.3×10^{-6}
p_2	Parameters used in tradeoff objective function	h^{-1}	0.33
p_3	Parameters used in tradeoff objective function	h	26.7

Table 5.S3. Lack-of-fit test for iFBA

Model Name	Lack-of-fit sum of squares, SSLOF	Degree of freedom, df_1	Pure-error sum of squares, SSPE	Degree of freedom, df_1	$\frac{SSLOF / df_1}{SSPE / df\,2}$	$F(df_1,df_2)$
iFBA	3.726	56	0.813	144	11.78	1.396

Note: In order to test whether or not a model could fit the data well, we applied the lack-of-fit test. The *F*-test indicates that the iFBA model should be further improved to describe the experimental data.

Chapter 6

MicrobesFlux: a Web Platform for Drafting Metabolic Models from KEGG Database

Xueyang Feng[1,†,$], You Xu[2,$], Yixin Chen[2] and Yinjie J. Tang[1,*]

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

$: XF and YX are equally contributed

Contact: †: fengx@seas.wustl.edu

***:**yinjie.tang@seas.wustl.edu

This chapter was manuscript in preparation for *BMC Systems Biology*

6.1 Abstract

Background: Parallel to the efforts currently underway in mapping microbial genomes using high-throughput sequencing methods, systems biologists are building metabolic models to characterize and predict cell metabolisms. The key step in building a metabolic model is querying multiple databases to collect and assemble essential information about genome-annotations and the architecture of metabolic network for a specific organism. To speed up metabolic model development for a large number of microorganisms, we need a user-friendly platform to construct metabolic network and perform constraint-based flux balance analysis based on genome databases and experimental results.

Results: We have developed a web-based platform (MicrobesFlux) for generating and reconstructing metabolic models for annotated microorganisms. The MicrobesFlux is able to load the metabolic network (including enzymatic reactions and metabolites) of over 1,100 species from KEGG database (Kyoto Encyclopedia of Genes and Genomes) and then automatically converting it to a metabolic model. The platform also provides diverse customized tools, such as gene knockouts and introduction of heterologous pathways, for users to redefine the model network. The reconstructed metabolic network can be formulated to a constraint-based flux model to predict and analyze the carbon fluxes in the metabolisms. The simulation results can be output in SBML format (The Systems Biology Markup Language). Furthermore, we also validate and demonstrate the

platform functions by developing an FBA model (including 229 reactions) for a recent annotated bioethanol producer, *Thermoanaerobacter sp*. strain X514, to predict its biomass growth and ethanol productions.

Conclusion: The MicrobesFlux is an installation-free and open-source platform that enables biologists with little programming knowledge to develop metabolic models for newly sequenced microorganisms. Our system allows users to construct metabolic networks of organisms directly from the KEGG database. It also provides users with predictions of microbial metabolism via flux balance analysis. This prototype platform can be a springboard to advanced and broad-scope metabolic modeling of complex biological systems by integrating other "omics" data or ^{13}C-assisted metabolic flux analysis results. MicrobesFlux is available at http://tanglab.engineering.wustl.edu/static/MicrobesFlux.html.

6.2 Background

Arising interests have focused on systems analysis of cell metabolisms (1,2,3,4,5,6). Metabolic flux analysis is a key systems biology approach that determines the final *in vivo* enzyme activities in a metabolic network and links genetics to biological functions. In the past decade, over 100 genome-scale metabolic models have been constructed for *E.coli* (7,8,9), *Bacillus subtilis* (10,11), and *Saccharomyces cerevisiae* (12,13,14), to expand our understanding of their physiologies. While important, the pace of metabolic model reconstructions is still much slower than the pace of high throughput genome

sequencing of diverse microorganisms (15) due to three reasons. First, reconstructing metabolic models is normally a tedious, slow and labor-intensive process, including over 90 steps from assembling genome annotations of target organisms to validating the metabolic model by various "omics" studies (16). Second, a systemic reconstruction of metabolic model often relies on commercial software (e.g. MATLAB) and demands proficient programming skills of the researchers. The majority of microbiologists, who know the physiology of environmental microorganisms well, may not gain access to commercial software or programming skills. Third, to study less-characterized environmental organisms, it needs to efficiently convert a vast amount of experimental data into model constraints to reduce solution space and improve the model predictability. Therefore, programming-free and user-friendly software need to be developed to overcome rate-limiting steps in metabolic model reconstructions.

Currently, only a few software tools are available to assist biologists for metabolic modeling. SimPheny is commercial software for genome-scale Flux Balance Analysis (www.genomatica.com). Webcoli supplies diverse approaches for users to reconstruct a genome-scale *E.coli* metabolic model (17). OpenFLUX is computationally efficient software for ^{13}C-assisted metabolic flux (18). OptFlux is open-source and modular software for FBA and microbial strain design using an evolutionary optimization algorithm (19). BioMet Toolbox is a web-based resource for FBA and transcriptome analysis (20). Model SEED (21) has been

developed to automatically generate genome-scale metabolic models for different microbes based on RAST sequencing. Complementing these tools, we are developing MicrobesFlux, a web platform to draft and reconstruct metabolic models. This system has several features: 1) automatically generate metabolic models of over 1,100 microbes sequenced in KEGG dataase (www.genome.jp/kegg/), 2) fine tune the metabolic models according to user-defined requests, and 3) perform constraint-based flux analysis under both metabolic steady states (i.e. flux balance analysis, FBA) and dynamic states (i.e. dynamic flux balance analysis, dFBA). The marriage of high-throughput model generation and the customized genome reconstruction is of great benefit since biologists can easily validate or disprove various hypotheses in microbial metabolism by drafting and comparing numerous metabolic models. Besides, this prototype platform can potentially link to other software (e.g. OptFlux (19), COBRA (22)) to perform broad-scope metabolic modeling of complex biological systems.

6.3 Implementation

MicrobesFlux is an open-source platform that does not require mandatory registration. It has three components: *logic level*, *application level*, and *achievement level*. The *logic level* includes two fundamental databases used in MicrobesFlux, where KGML is for organism-specific metabolic networks and KEGG LIGAND is for general enzymatic reactions and metabolites. The basic

principles for metabolic model reconstruction and constraint-based flux analysis are summarized in logic level (Figure 6.1). In the application level, organism-specific metabolic network is loaded from KEGG database to generate the metabolic model, which is pursued for customized reconstruction. The reconstructed metabolic model is then formulated as either an FBA or a dFBA problem to determine the flux distributions under metabolic steady or dynamic state, respectively. The constraint-based flux analysis will be accomplished in the achievement level, by using state-of-the-art optimization solvers, such as IPOPT (Interior Point OPTimizer) and cloud computations. The calculated flux distributions and the reconstructed metabolic network are recorded in SBML format (The Systems Biology Markup Language), and the metabolic networks are visualized using Scalable Vector Graphics (SVG). Both results are sent to users in the output module. In summary, three key features are embedded in MicrobesFlux: 1) high-throughput and automatic generation of metabolic model; 2) customized reconstruction of metabolic models; and 3) constraint-based flux analysis in steady and dynamic metabolic states.

6.3.1 High-throughput and automatic generation of metabolic model

MicrobesFlux can serve as a platform to build metabolic models for 1,194 organisms sequenced in KEGG database. To generate a metabolic model for a specific microorganism, the gene annotation information are loaded from the corresponding KGML files, and cross-referred to KEGG LIGAND database. The

generated pathway network serves as a seed model and subject to further reconstruction based on user-defined requests. The pipeline for metabolic model generation is automatic, time-saving and high-throughput.

6.3.2 Customized reconstruction of metabolic models

The model of designated microorganisms, automatically generated from KEGG database, should be reconstructed to fill the gaps between genome annotation and the functional metabolism. Based on experimental knowledge, users can manually adjust mis-annotated pathways in the original metabolic network. Furthermore, users can knock out the native pathways or introduce heterologous pathways in metabolic models for engineered microbial strains. The reaction can also be set as reversible in the MicrobesFlux platform. In addition, the uptake of carbon substrates and the extracellular secretion of metabolites can be experimentally measured, which are constructed as "inflow" pathways and "outflow" pathways in the metabolic model, respectively. Since numerous metabolites are used as building blocks for biomass production, an independent biomass formation pathway can be introduced by allowing users to input the organism-specific biomass composition information. The metabolic pathways can be grouped by different metabolic features to provide an intuitive manipulation of the metabolic network by users.

6.3.3 Constraint-based flux analysis in steady and dynamic metabolic states

For each reconstructed metabolic network, users can choose to perform either FBA or dFBA to determine flux distribution under metabolic steady or dynamic states. In general, the objective functions used in constraint-based flux analysis can either be the commonly used "maximizing biomass", or a user-defined objective function. The users can also define the upper and lower bounds of each flux. MicrobesFlux performs the constraint-based flux analysis of the reconstructed model using optimization solver (i.e., IPOPT). As a unique feature, dynamic simulation of metabolic fluxes can be achieved in MicrobesFlux, following the static optimization approach (SOA) (23). In specific, the users can assume that the entire dynamic microbial metabolism decomposed into numerous pseudo-steady states. At each pseudo-steady state, a conventional FBA problem is formulated with the user-defined inflow and outflow fluxes, so a dFBA problem will be converted to multiple mini-FBAs that are subject to constraints from the measurement of time-dependent inflow (substrate uptake) and outflow fluxes (metabolite production). To avoid extensive analytical efforts to measure inflow and outflow fluxes at each time interval, the users can use an empirical or kinetic model to estimate the time-dependent inflow fluxes and outflow fluxes for mini-FBAs through entire growth period based on limited measurement data. The dynamic flux simulation is of particular industrial interest since many biological systems cannot maintain a meaningful metabolic steady state during fermentation process.

6.4 Results

We have applied MicrobesFlux to a few studies in drafting metabolic models. We first drafted a TOY model (Figure 6.2), which has 10 metabolites and 16 fluxes, as a demonstration of MicrobesFlux workflow. We then constructed a medium-scale stoichiometric model with 196 metabolites and 229 reactions for *Thermoanaerobacter* sp. strain X514, a thermophilic bacterium that is of great interests in cellulosic ethanol production (24). The functionality of MicrobesFlux has been proved in both case studies.

6.4.1 Case study 1: a toy model

To demonstrate the use of MicrobesFlux platform, a simple toy model was constructed, only including central metabolic pathways, namely glycolysis, the pentose phosphate pathway, the TCA cycle, and the anaplerotic pathway. Glucose represented the carbon substrate and acetate represented the extracellular metabolite product. The TOY model was loaded from MicrobesFlux (Figure 6.3), which included 10 reactions that described the intracellular fluxes and lumped biomass production. The toy model was then reconstructed by introducing the inflow flux: "Glucose → G6P" and the outflow flux: "AcCoA → Acetate". The drafted TOY model was then pursued for constraint-based flux analysis, by setting the objective function as "maximizing biomass" and fixing the inflow and outflow fluxes as 11.0 and 6.4 mmol/g/h. The simulated results from

MicrobesFlux were confirmed by an independent linear optimization of the same TOY model via "linprog" in MATLAB.

6.4.2 Case study 2: a medium-scale metabolic model for *Thermoanaerobacter* sp. strain X514

Based on the similar flowchart, we have drafted a medium-scale metabolic model of *Thermoanaerobacter* sp. strain X514. *Thermoanaerobacter* species are thermophilic bacteria that can covert both pentose and hexose simultaneously to ethanol with high yield (25,26). It can also be co-cultured with cellulosic *Clostridium thermocellum* species to produce ethanol from cellulose. Strain X514 has been sequenced recently and a comprehensive study on the central carbon metabolism has been accomplished by using ^{13}C-assisted pathway analysis (25). Using the experimental data of strain X514 reported in (25) and (26), we apply MicrobesFlux to construct a metabolic model that can describe the carbohydrate metabolism and amino acids biosynthesis in Strain X514. The drafted model is consisted of 196 metabolites and 229 reactions (162 intracellular reactions, 19 inflow/outflow reactions, 39 gap-filling reactions, and 9 biomass-producing reactions). The intracellular reactions in the drafted model are derived from genome-annotation, only considering the carbon and cofactor balance. The inflow/outflow reactions were introduced into the metabolic model if the transporters of the specific substrates or extracellular products have been reported by published research. Moreover, two algorithms were employed to fill the gaps

in the metabolic pathways. First, we have implemented un-annotated novel pathways identified from ^{13}C-assisted pathway analysis (e.g. the Re-type citrate synthase in TCA cycle) into the metabolic model of strain X514 (25). Since the draft model is a simplified model by focusing on carbohydrate metabolism and amino acids biosynthesis, the gaps may also be generated due to the lack of consideration of the metabolites exchange between the pathways that were included in the model (e.g. carbohydrate metabolism) and other pathways that were not included in the model (e.g. purine metabolism). Accordingly, we employ the principle of introducing the metabolite-exchange-reactions for filling the gaps. For example, UTP is involved in both carbohydrate metabolism and RNA synthesis. Since we did not include the RNA synthesis pathway in the MicrobesFlux model, we instead introduced a UTP exchange pathway to fill the gap. The gap-filling reactions were evaluated carefully to make sure that each one of them was necessary for feasible predictions of biomass production.

The biomass composition of *Thermoanaerobacter* sp. strain X514 is not yet available. In the drafted model, we used the reported biomass composition of a close species, *Clostridium acetobutylicum* (27). Besides the biomass composition, the growth-associated maintenance (GAM) and non-growth associated maintenance (NGAM) energies have been found to play an important role in simulating the growth rate of organisms (27,28,29). In our model simulation, the NGAM was chosen as 7.6 mmol ATP/g DCW, as reported

previously (29). To identify the GAM, we plotted the relations between growth rate and GAM (Figure 6.4), and found the value of GAM that can match the experimental measurement (i.e. growth rate of strain X514 as 0.043 h^{-1} reported in (26)) was 220.0 mmol ATP/g DCW, and this fitted value is qualitatively consistent to previous reported GAM value (150.0 mmol ATP/g DCW) in a thermopilic ethanol producer *Clostridium thermocellum* (30). Based on this drafted model, we also predicted the correlations between growth rate, ethanol production, and waste product outflow (Figure 6.5). The prediction shows a trade-off relation between ethanol production and growth rate. The ethanol production can be increased by 25% (i.e. increased from 6.3 mmol/g/h to 7.8 mmol/g/h) while halving the growth rate (i.e. decreased from 0.045 h^{-1} to 0.027 h^{-1}). By inhibiting the acetate production from 2.0 mmol/g/h to 0.6 mmol/g/h, the ethanol production under the optimal growth conditions could be improved by 33%. Therefore, we have shown the platform is able to make reasonable predictions for the biomass growth in response to metabolites synthesis.

6.5 Discussion and conclusion

MicrobesFlux is designed for high-throughput drafting of metabolic models for environmental organisms based on genome annotations in KEGG. Unlike the model organisms that have been systemically studied via different "omics" approaches, the environmental organisms have more complex metabolic features but fewer measurements from laboratories. Therefore, MicrobesFlux can

be used as a platform to construct models for predicting cell metabolism. On the other hand, validation of a genome-scale metabolic model can be a challenge due to requirement for manually fine tuning metabolic pathways and due to lack of experimental data to confirm functional pathways. For example, *Thermoanaerobacter* sp. strain X514 used in the MicrobesFlux case study is capable of growing under high temperature and converting sugars to ethanol, which can be simulated by a medium-scale metabolic model. Such a convenient FBA model has decent predictive power for *in silico* studies of a non-model environmental microorganism for bioethanol production at different growth rates (Figure 6.5). However, many other metabolic features (e.g., secondary metabolisms) in the strain X514 are not fully understood yet. It requires extensive experimental studies, including transcriptomics and proteomics analysis, for a precise reconstruction of genome-scale model. Although the scale of drafted MicrobesFlux model in this study is still small (<300 reactions) so that users can easily fine tune the metabolic model and manually fill annotation gaps in carbon metabolic pathways, MicrobesFlux can serve as a springboard to future advanced genome-scale metabolic model by adding more reconstruction steps, such as the gene-protein-reaction (GPR) correlation, charge balance, and metabolites exchange among metabolic compartments (e.g. mitochondria in eukaryotic cells).

In summary, we have developed MicrobesFlux to draft the metabolic models of environmental organisms from KEGG database and serves as a high-

throughput tool for systems biology. MicrobesFlux extricates microbiologists from programming and commercial software that are normally required to build a metabolic model. It covers a large number of sequenced organisms in KEGG database and provides multiple approaches to assist the model generation and reconstruction. Both FBA and dFBA can be achieved in MicrobesFlux to simulate the metabolic behaviors of organisms in metabolic steady state and dynamic state. The drafted model can be further used by other software for genome-scale model reconstruction and *in silico* predictions. In the future, we will implement broad fluxomic approaches (e.g. ^{13}C-metabolic flux analysis) in MicrobesFlux to improve the accuracy and predictive power of drafted metabolic models.

6.6 Availability and requirements

- Project name: MicrobesFlux
- Project homepage:

  http://tanglab.engineering.wustl.edu/static/MicrobesFlux.html
- Operating systems: Platform independent.
- Programming language: Java and Python.
- License: MicrobesFlux is freely available for non commercial purposes.
- Any restrictions to use by non-academics: none

6.7 References

1. Chang RL, Ghamsari L, Manichaikul A, Home EFY, Balaji S, et al. (2011) Metabolic network reconstruction of Chlamydomonas offers insight into light-driven algal metabolism. Molecular Systems Biology 7.

2. Saha R, Suthers PF, Maranas CD (2011) Zea mays iRS1563: A Comprehensive Genome-Scale Metabolic Reconstruction of Maize Metabolism. PLoS ONE 6: e21784.

3. Lewis N, Schramm G, Bordbar A, Schellenberger J, Andersen M, et al. (2010) Large-scale in silico modeling of metabolic interactions between cell types in the human brain. Nature Biotechnology 28: 1279-1285.

4. Mahadevan R, Palsson BO, Lovley DR (2011) In situ to in silico and back: elucidating the physiology and ecology of Geobacter spp. Using genome-scale modeling. Nature Reviews Microbiology 9: 39-50.

5. Satish Kumar V, Ferry JG, Maranas CD (2011) Metabolic reconstruction of the archaeon methanogen Methanosarcina acetivorans. BMC Systems Biology 5.

6. Imam S, Yilmaz S, Sohmen U, Gorzalski AS, Reed JL, et al. (2011) iRsp1095: A genome-scale reconstruction of the Rhodobacter sphaeroides metabolic network. BMC Systems Biology 5.

7. Edwards JS, Palsson BO (2000) The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. PNAS 97: 5528–5533.

8. Reed J, Vo T, Schilling C, Palsson B (2003) An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). Genome Biol 4: R54.

9. Feist A, Henry C, Reed J, Krummenacker M, Joyce A, et al. (2007) A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. Mol Syst Biol 3.

10. Oh Y-K, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale reconstruction of metabolic network in Bacillus subtilis based on high-throughput phenotyping and gene essentiality data. J Biol Chem 282: 28791-28799.

11. Henry CS, Zinner JF, Cohoon MP, Stevens RL (2009) iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. Genome Biology 10.

12. Förster J, Famili I, Fu P, Palsson B, Nielsen J (2003) Genome-scale reconstruction of the Saccharomyces cerevisiae metabolic network. Genome Res 13: 244-253.

13. Duarte N, Herrgård M, Palsson B (2004) Reconstruction and validation of Saccharomyces cerevisiae iND750, a fully compartmentalized genome-scale metabolic model. Genome Res 14: 1298-1309.

14. Herrgård M, Swainston N, Dobson P, Dunn W, Arga K, et al. (2008) A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. Nat Biotechnol 26: 1155-1160.

15. Suthers PF, Dasika MS, Kumar VS, Denisov G, Glass JI, et al. (2009) A genome-scale metabolic reconstruction of Mycoplasma genitalium, iPS189. PLoS Comput Biol 5: e1000285. .

16. Thiele I, Palsson BO (2010) A protocol for generating a high-quality genome-scale metabolic reconstruction. Nature Protocols 5: 93-121.

17. Jung T-S, Yeo HC, Reddy SG, Cho W-S, Lee D-Y (2009) WEbcoli: an interactive and asynchronous web application for in silico design and analysis of genome-scale *E.coli* model Bioinformatics 25: 2850-2852

18. Quek L-E, Wittmann C, Nielsen LK, Krömer JO (2009) OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. Microb Cell Fact 8.

19. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, et al. (2010) OptFlux: an open-source software platform for in silico metabolic engineering. BMC Syst Biol 4: 45.

20. Cvijovic M, Olivares-Hernández R, Agren R, Dahr N, Vongsangnak W, et al. (2010) BioMet Toolbox: genome-wide analysis of metabolism. Nucleic Acids Res 38: W144.

21. Henry C, DeJongh M, Best A, Frybarger P, Linsay B, et al. (2010) High-throughput generation, optimization and analysis of genome-scale metabolic models. Nat Biotechnol 28: 977-982.

22. Becker SA, Feist AM, Mo ML, Hannum G, Palsson BØ, et al. (2007) Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox. Nat Protocols 2: 727-738.

23. Mahadevan R, Edwards JS, Doyle FJ (2002) Dynamic flux balance analysis of diauxic growth in Escherichia coli. Biophys J 83: 1331-1340.

24. He Q, Hemme CL, Jiang H, He Z, Zhou J (2011) Mechanisms of enhanced cellulosic bioethanol fermentation by co-cultivation of *Clostridium* and *Thermoanaerobacter* spp. Bioresource Technology 102: 9586-9592.

25. Feng X, Mouttaki H, Lin L, Huang R, Wu B, et al. (2009) Characterization of the Central Metabolic Pathways in Thermoanaerobacter sp. X514 via Isotopomer-Assisted Metabolite Analysis. Appl Environ Microbiol 75: 5001-5008.

26. Hemme C, Fields M, He Q, Deng Y, Lin L, et al. (2011) Correlation of genomic and physiological traits of thermoanaerobacter species with biofuel yields. Appl Environ Microbiol 77: 7998-8008.

27. Lee J, Yun H, Feist AM, Palsson BØ, Lee SY (2008) Genome-scale reconstruction and in silico analysis of the Clostridium acetobutylicum ATCC 824 metabolic network. Appl Microbiol Biotechnol 80: 849-862.

28. Feist A, Scholten J, Palsson B, Brockman F, Ideker T (2006) Modeling methanogenesis with a genome-scale metabolic reconstruction of Methanosarcina barkeri. . Mol Syst Biol 2: 2006.0004.

29. Varma A, Palsson BO (1994) Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. Appl Environ Microbiol 60: 3724-3731.

30. Roberts S, Gowen C, Brooks J, Fong S (2010) Genome-scale metabolic analysis of Clostridium thermocellum for bioethanol production. BMC Syst Biol 4.

Figure 6.1. Architecture of MicrobesFlux.

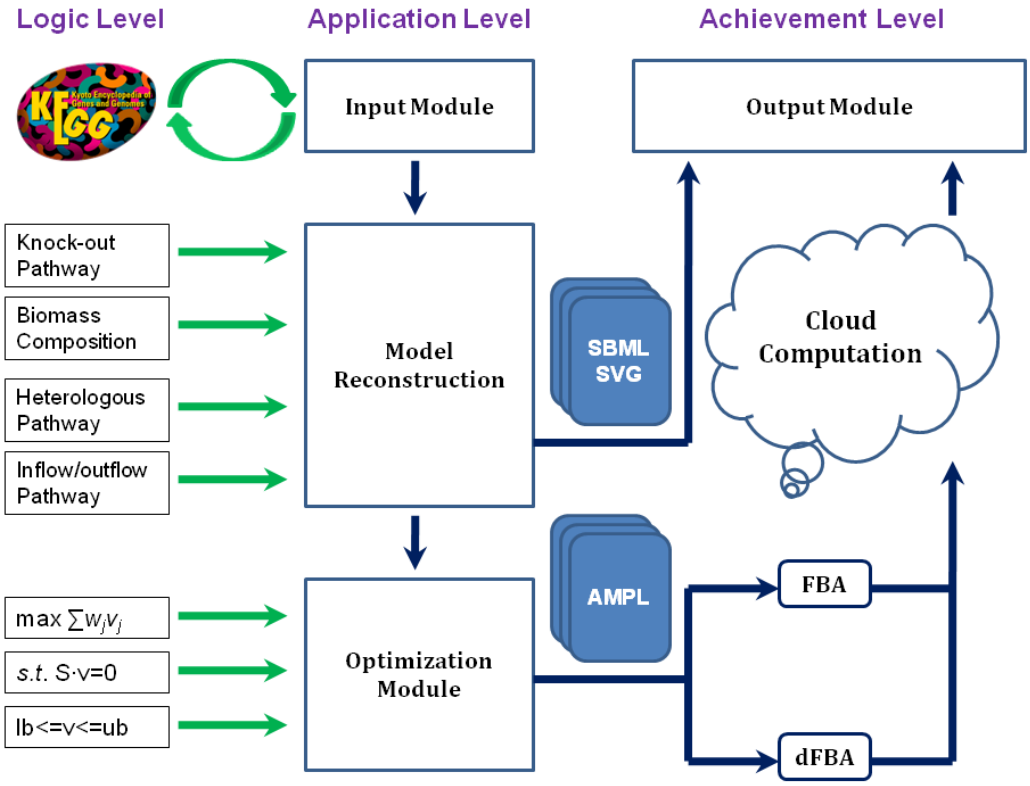
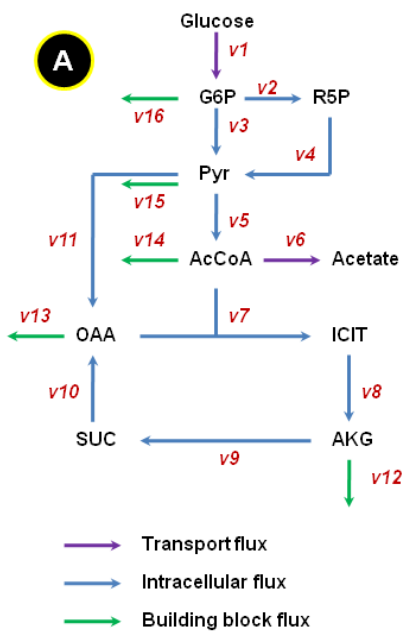
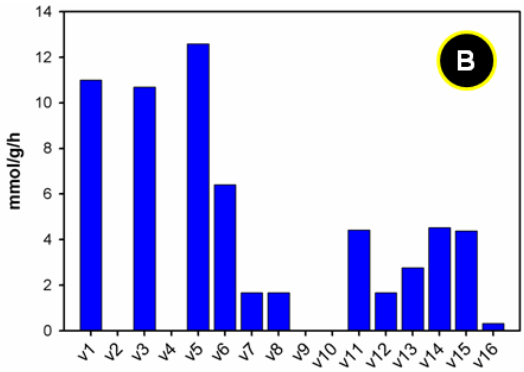


Figure 6.2. (**A**) Pathway network of TOY model used in MicrobesFlux, and (**B**) simulated flux distribution of TOY model used in MicrobesFlux. The same results were obtained by using "linprog" in MATLAB.





Biomass production:

1.078 AKG + 1.786 OAA + 2.928 AcCoA + 2.833 Pyr + 0.205 G6P → Biomass

Maximum growth rate predicted by MicrobesFlux:
Growth rate = 1.54 h^{-1}

Figure 6.3. Screenshot of reconstructing TOY model by using MicrobesFlux. (**A**) load TOY model from MicrobesFlux; (**B**) pathway information of TOY model; (**C**) customized reconstruction of TOY model; and (**D**) constraint-based flux balance analysis of TOY model.

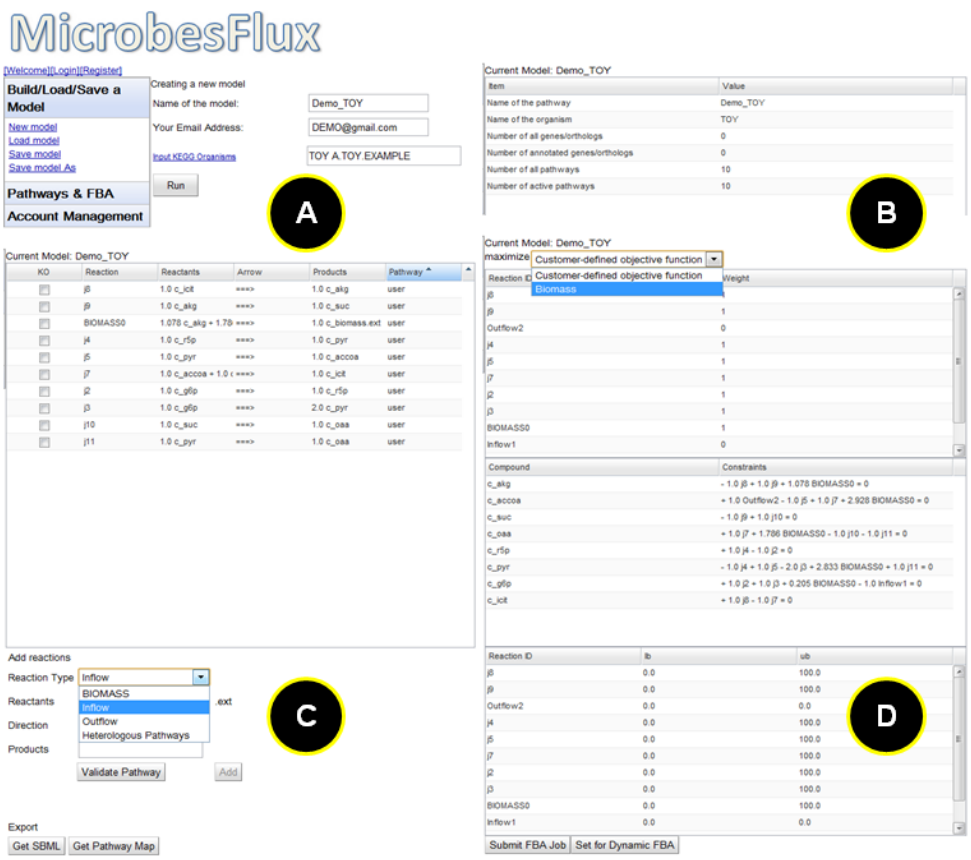


Figure 6.4. Estimation of the growth-associated maintenance (GAM) in *Thermoanaerobacter* sp. strain X514. From this comparison, the GAM value (red dotted line) consistent with experimental data (i.e. growth rate was 0.043 h^{-1}) could be estimated and was indicated as a 220.0 mmol ATP/ g DCW.

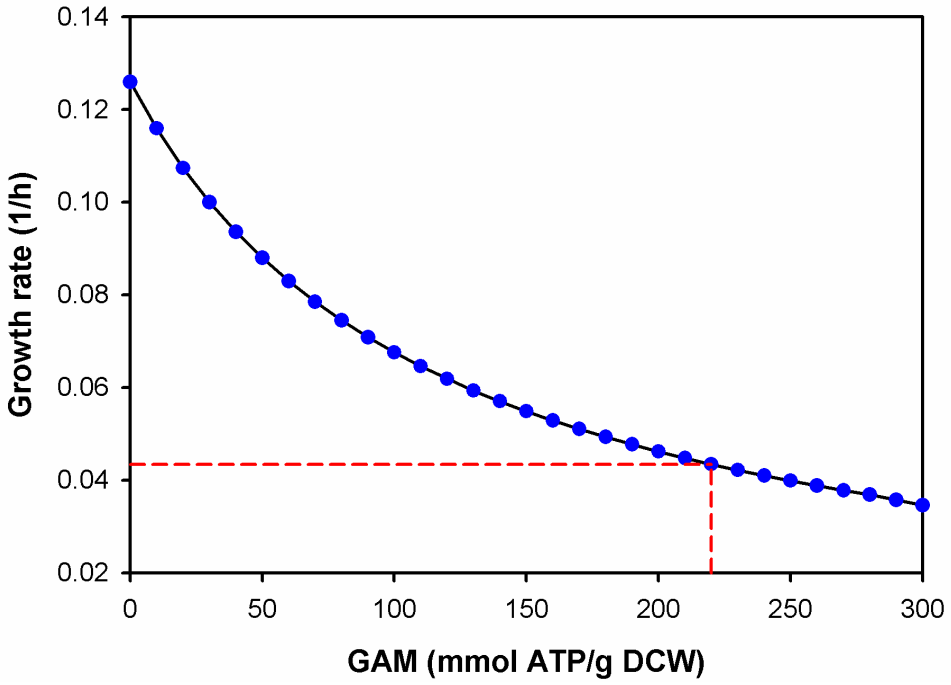
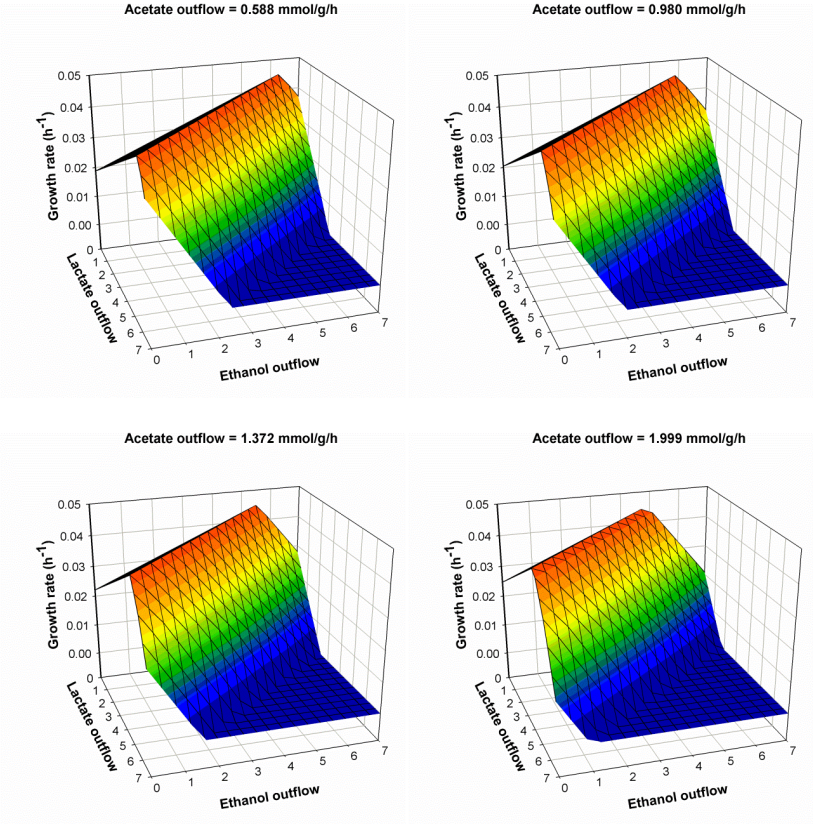


Figure 6.5. Predictions of the relationship between growth rate and outflow fluxes [Unit: mmol/g/h] in *Thermoanaerobacter* sp. strain X514. The glucose inflow flux is fixed as 3.92 mmol/g/h.



Chapter 7

Conclusions and Future Investigations

7.1 Summary

I have developed and applied multiple fluxomics tools during my Ph.D. training. ^{13}C-pathway analysis has been systemically used to discover novel pathways, investigate the mixotrophic metabolism, and optimize medium for slow-growing bacteria. The achievements from ^{13}C-pathway analysis are summarized in Table 7.1. Meanwhile, ^{13}C-MFA has extended the applications from quantifying metabolic fluxes in industrial workhorses to those in environmental microorganisms. The accurate metabolic readouts of the flux distribution have been determined for *Chlorobaculum tepidum*, as the first study in quantifying the mixotrophic metabolism of green sulfur bacteria. The isotope discriminations have been confirmed in heterotrophic metabolism of *E.coli*, which could affect the solution space in flux calculations by ^{13}C-MFA. Moreover, to investigate the dynamic metabolism of environmental microorganisms, a framework that integrates bioprocess models, dFBA, and isotopomer simulations has been applied in *Shewanella oneidensis* MR-1. The dynamic flux distributions have been profiled, which revealed a few underlying metabolic features during the dynamic metabolism. Such dynamic flux analysis approach can bypass the measurement of low-abundant and fast-turnover metabolites, and can address the

suboptimal microbial metabolisms that are often adopted by environmental microorganisms. As one of the steps in bridging the gaps between fast-paced genome sequencing and slow-paced metabolic model constructions, a web-based platform, MicrobesFlux, is being built based on KEGG database for high-throughput metabolic models drafting. A large database including 1,192 sequenced genomes is available in MicrobesFlux. Diverse customized tools are provided for model reconstruction. Both FBA and dFBA can be achieved to simulate the metabolic fluxes of environmental microorganisms in metabolic steady state and dynamic state.

Beside fluxomics investigation on the metabolisms of environmental microorganisms, several projects have also been accomplished by studying bioprocess engineering, metabolic engineering, and nano-toxicity. In general, a parsimonious kinetic model has been constructed to capture the growth kinetics of *Shewanella oneidensis* MR-1 (Appendix 9). A bioprocess model was built to describe the isobutanol fermentation of recombinant *E.coli* strains (Appendix 10). The broad scope research also involved constructing a statistical model that analyzed influential factors in production yield of metabolic engineering (Appendix 11) (1), and characterizing the bacterial responses to nanoparticles (Appendix 12) (2).

7.2. Current challenges in fluxomics analysis

One of the main applications of fluxomics analysis is to identify the bottlenecks in biochemical production by engineered microorganisms and to assist the rational design of mutants in metabolic engineering. Unfortunately, several barriers in current fluxomics analysis stalled the inception of fluxomics-based biotechnology.

On one hand, the fluxomics can measure the carbon flows in central metabolic pathways to the final product and predict the optimal operation of metabolic network. However, the flux measurements and predictions have not yet been coordinated systemically and left many key questions unanswered in metabolic engineering. For example, metabolic robustness, the ability of microorganisms to maintain the similar metabolic performances under genetic or environmental perturbations, is a long-recognized property of microbial systems. Comparative studies on wide-type and genetically manipulated mutants via ^{13}C-MFA often fail to tell the differences in flux distributions of central metabolism, and yield no information or predictions for the mechanisms of microbial robustness. Constraint-based flux analysis can *in silico* test different hypothesis about microbial robustness. However, most of the hypotheses cannot be proved due to lack of accurate flux measurements in FBA.

On the other hand, metabolic engineering of industrial chassis is based on the premise that the yield of a desired product can be increased by identifying and over-expressing the enzymes that catalyze the rate-limiting steps in a given

metabolic pathway. However, the rate-limiting enzymes may not always exist in reality due to the complicated cell-wide regulations of microbial metabolisms that involve both the well-known gene-protein-flux correlations, and the less well-known post-transcriptional and post-translational regulations. In this case, the increases in productivity are often achieved by coordinated expression of multiple genes in the metabolic network, in which fluxomics analysis cannot provide reasonable guidelines for rational design in metabolic engineering, until it is cross-talked with other "omics" data in systems biology.

7.3 Future development of advanced fluxomics tools

Fluxomics analysis is playing an increasingly important role in systems biology and bioengineering. One of current limits in fluxomics studies is the tradeoff between model predictive capacity and model accuracy. FBA is well-known for the strong predictive power by adopting genome-scale metabolic models. However, FBA lacks the accuracy in describing the cellular metabolism. Conversely, ^{13}C-MFA provides accurate metabolic readouts in a simplified metabolic network. Unfortunately, it cannot predict metabolic behaviors well. To further refine fluxomics analysis, my future research will focus on developing an integrated modeling approach that not only describes the cellular metabolism accurately in genome-scale, but also predicts the flux distributions based on diverse assumptions. The barriers between model prediction and model accuracy

will be removed, which can be of specific help in rational design of metabolically engineered mutants.

Another topic in my future investigation is to cross-talk fluxomics with transcriptomics and proteomics to unravel the cell-wide post-transcriptional and post-translational regulations. Little is known about the post-transcriptional and post-translational network in cellular metabolism except their complex characteristics. The fluxomics study can provide reliable and high-throughput of cellular metabolism. By quantitatively correlating the input (i.e., transcriptomics and proteomics) and output (i.e. fluxomics) of metabolic system, the topology and activities of metabolic regulatory network that controls the cellular metabolism can be empirically deciphered. The decoding of metabolic regulatory mechanisms will be of great value in understanding the physiology, especially the pathophysiology of diverse biological systems.

Finally, open-source software for integrative and broad scope fluxomics analysis is planned to be developed in future. It is expected that genome-scale metabolic models can be automatically generated from multiple genome sequencing databases. The drafted metabolic models can be efficiently and successively validated based on heterogeneous databases on systems biology experiments (e.g. transcriptional analysis, lethality tests). State-of-the-art flux analysis approaches can be achieved in the genome-scale metabolic models,

which provide precise readouts on metabolic flux distributions and predictive guidelines for *in silico* genetic manipulations.

7.4 References

1. Colletti P, Goyal Y, Varman A, Feng X, Wu B, et al. (2011) Evaluating factors that influence microbial synthesis yields by regression with numerical and categorical variables. Biotechnology and Bioengineering 108: 893-901.

2. Wu B, Huang R, Sahu M, Feng X, Biswas P, et al. (2010) Bacterial responses to Cu-doped TiO(2) nanoparticles. Science of the Total Environment 408: 1755-1758.

Table 7.1. Novel metabolic features identified by ^{13}C-pathway analysis

Microbes	Interests	Novel metabolic features	Collaborations
Shewanella oneidensis MR-1	Bioremediation and microbial fuel cells.	Invariability of central metabolism under environmental or genetic perturbations	Dr. Adam Arkin
Mycobacteria smegmatis	Relative of pathogenic *Mycobacterium tuberculosis*	Induced glyoxylate pathway and glycine dehydrogenase under hypoxic stress	Dr. Carolyn Bertozzi and Dr. Jay D. Keasling
Thermoanaerobacter X514	Ethanol producing at high temperature	(Re)-type citrate synthase and citramalate pathway	Dr. Jizhong Zhou
Roseobacter denitrificans OCh114	Aerobic anoxygenic phototrophs, CO_2 fixation	Inactive glycolysis and active Entner–Doudoroff pathway	Dr. Robert Blankenship
Cyanothece 51142	Biofuel producing, nitrogen fixation	Citramalate pathway for isoleucine synthesis; inhibition of CO_2 fixation by glycerol	Dr. Himadri Pakrasi
Heliobacterium modesticaldum	Gram-positive, thermophilic, photosynthetic bacteria	(Re)-type citrate synthase in TCA cycle	Dr. Robert Blankenship
Dehalococcoides ethenogenes 195	TCE degradation	Exogenous amino acids are selectively utilized	Dr. Lisa Alvarez-Cohen

Appendix 1

Metabolic Pathway Confirmation and Discovery Through

^{13}C-labeling of Proteinogenic Amino Acids

Video Article

Metabolic Pathway Confirmation and Discovery Through ^{13}C-labeling of Proteinogenic Amino Acids

Le You[1], Lawrence Page[2], Xueyang Feng[1], Bert Berla[1], Himadri B. Pakrasi[3], Yinjie J. Tang[1]

[1]Department of Energy, Environmental and Chemical Engineering, Washington University
[2]Department of Biology, Washington University
[3]Department of Energy, Environmental and Chemical Engineering and Department of Biology, Washington University

Correspondence to: Yinjie J. Tang at yinjie.tang@seas.wustl.edu

URL: http://www.jove.com/details.php?id=3583
DOI: 10.3791/3583

Keywords: GC-MS, novel pathway, metabolism, labeling, phototrophic microorganism,

Date Published: //

Citation: You, L., Page, L., Feng, X., Berla, B., Pakrasi, H.B., Tang, Y.J. Metabolic Pathway Confirmation and Discovery Through 13C-labeling of Proteinogenic Amino Acids. J. Vis. Exp. (), e3583, DOI : 10.3791/3583 ().

Abstract

Microbes have complex metabolic pathways that can be investigated using biochemistry and functional genomics methods. One important technique to examine cell central metabolism and discover new enzymes is ^{13}C-assisted metabolism analysis 1. This technique is based on isotopic labeling, whereby microbes are fed with a ^{13}C labeled substrates. By tracing the atom transition paths between metabolites in the biochemical network, we can determine functional pathways and discover new enzymes.

As a complementary method to transcriptomics and proteomics, approaches for isotopomer-assisted analysis of metabolic pathways contain three major steps [2]. ***First***, we grow cells with ^{13}C labeled substrates. In this step, the composition of the medium and the selection of labeled substrates are two key factors. To avoid measurement noises from non-labeled carbon in nutrient supplements, a minimal medium with a sole carbon source is required. Further, the choice of a labeled substrate is based on how effectively it will elucidate the pathway being analyzed. Because novel enzymes often involve different reaction stereochemistry or intermediate products, in general, singly labeled carbon substrates are more informative for detection of novel pathways than uniformly labeled ones for detection of novel pathways[3, 4]. ***Second***, we analyze amino acid labeling patterns using GC-MS. Amino acids are abundant in protein and thus can be obtained from biomass hydrolysis. Amino acids can be derivatized by N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide (TBDMS) before GC separation. TBDMS derivatized amino acids can be fragmented by MS and result in different arrays of fragments. Based on the mass to charge (m/z) ratio of fragmented and unfragmented amino acids, we can deduce the possible labeled patterns of the central metabolites that are precursors of the amino acids. ***Third***, we trace 13C carbon transitions in the proposed pathways and, based on the isotopomer data, confirm whether these pathways are active [2]. Measurement of amino acids provides isotopic labeling information about eight crucial precursor metabolites in the central metabolism. These metabolic key nodes can reflect the functions of associated central pathways.

^{13}C-assisted metabolism analysis via proteinogenic amino acids can be widely used for functional characterization of poorly-characterized microbial metabolism[1]. In this protocol, we will use *Cyanothece* 51142 as the model strain to demonstrate the use of labeled carbon substrates for discovering new enzymatic functions.

Video Link

The video component of this article can be found at http://www.jove.com/details.php?id=3583

Protocol

1. Cell culture (Figure 1)

1. Grow cells in minimal medium with trace elements, salts, vitamins, and specifically labeled carbon substrates that are best for pathway investigation. Use either shaking flasks or bioreactors for cell culture. Organic nutrients, such as yeast extract, may interfere with the measurement of amino acid labeling and thus cannot be present in the culture medium.
2. Monitor cell growth by the optical density of the culture at an optimal wavelength (e.g., OD_{730} for *Cyanothece* 51142) with a UV/Vis spectrophotometer.
3. Cells can first be grown in a non-labeled medium. The middle-log growth phase cells are preferred to be used for inoculation (3% (v/v) by volume inoculation ratio) of the labeled medium. The labeled culture should be sub-cultured (3% v/v inoculation ratio) in the same labeled medium to avoid the introduction of non-labeled carbon from the initial inoculum.

2. Amino acid extraction

1. Harvest sub-cultured cells (10mL) in the middle-log growth phase by centrifugation (10 min, 8000×g).
2. Resuspend the pellet in 1.5mL of 6M HCl and transfer it to a clear glass, screw-top GC vial. Cap the vials and place them in a 100°C oven for 24 hours to hydrolyze the biomass proteins into amino acids. Hydrolysis of biomass pellets can yield 16 of the 20 common amino acids (Figure 2) [5]. Cysteine and tryptophan are degraded, and glutamine and asparagine are converted to glutamate and aspartate, respectively.

3. Centrifuge the amino acid solution at 20,000×g for 5 min using 2 ml Eppendorf tubes, and transfer the supernatants to new GC vials. This step removes solid particles in the hydrolysis solution.
4. Remove the GC vial lids and dry the samples completely under a stream of air using a Thermo Scientific Reacti-Vap evaporator (note: a freeze dryer can also be used to dry samples). This step can be done overnight.

3. Amino acid derivatization and GC-MS conditions

Analysis of amino acids or charged/highly polar metabolites via GC requires that these metabolites be derivatized, so that the amino acids are volatile and can be separated by gas chromatography [2].

1. Dissolve the dried samples with 150 µL of tetrahydrofuran (THF) and 150 µL of N-(tert-butyldimethylsilyl)-N-methyltrifluoroacetamide derivatization reagent.
2. Incubate all samples in an oven or a water bath between 65 and 80°C for 1 hour. Vortex occasionally to make sure the metabolites in the vial are dissolved.
3. Centrifuge the samples at 20,000×g for 10 min, and then transfer the supernatant to new GC vials. The supernatant should be a clear and yellowish solution. Due to saturation of the detectors, GC-MS measurement accuracy can be affected by the high concentration of injected TBDMS derivatized amino acids (these samples often shows dark brown color), therefore, we should dilute these samples using THF before GC-MS measurement [6].
4. Analyze the samples by GC-MS (use a 1:5 or 1:10 split ratio, injection volume = 1 µL, carrier gas helium = 1.2 mL/min). Use the following GC temperature program: hold at 150°C for 2 minutes, increase at 3°C per min to 280°C, increase at 20°C per min to 300°C, and then hold for 5 minutes. Solvent delay can be set as ~5 min (for a 30 meter GC column). The range of the mass to charge ratio (m/z) in MS can be set between 60 and 500.

4. GC-MS data analysis

1. TBDMS derivatized amino acid measurement can also be affected by isotope discrimination in GC separation. Light isotopes move slightly faster than heave isotopes in GC column. To reduce the potential measurement bias, we may average the mass spectrum of the whole amino acid peak range [6]
2. The GC and MS spectra of TBDMS derivatized metabolites have been reported before[7]. The GC retention time and the unique m/z peaks for each amino acid are illustrated in Figure 3.
3. Derivatization of amino acids or central metabolites introduces significant amounts of naturally-labeled isotopes, including ^{13}C (1.13%), ^{18}O (0.20%), ^{29}Si (4.70%), and ^{30}Si (3.09%). The measurement noise from natural isotopes in the raw mass isotopomer spectrum can be corrected by using published software[5, 8]. The final isotopic labeling data are reported as mass fractions, e.g., M_0, M_1, M_2, M_3 and M_4 (representing fragments containing zero to four ^{13}C labeled carbons).
4. Measurement of amino acids can provide isotopic labeling information about eight crucial precursor metabolites: 2-oxo-glutarate, 3-P-glycerate, acetyl-CoA, erythrose-4-P, oxaloacetate, phosphoenolpyruvate, pyruvate, and ribose-5-P. The labeling patterns in these metabolites can be used to identify several central metabolic pathways (Figure 2) [9]. The outcome of the labeling experiments can be further confirmed using other biochemistry methods (e.g., RT-PCR).

5. Pathway analysis using labeled amino acid data

By investigating only a few key amino acids produced from well-designed ^{13}C tracer experiments, we may reveal several unique pathways or enzyme activities without performing sophisticated ^{13}C-metabolic flux analysis of entire central metabolism.

1. Entner–Doudoroff pathway: [1-^{13}C] glucose can be used as the carbon source. If the pathway is active, serine labeling will be significantly lower than labeling in alanine [10].
2. Branched TCA cycle: [1-^{13}C] pyruvate can be used as the carbon source. If the TCA cycle is broken, aspartate can be labeled by two carbons, while glutamate is labeled with only one carbon [11, 12].
3. CO_2 fixation by Calvin-Benson-Bassham cycle in a mixotrophic metabolism: Non-labeled CO_2 and labeled carbon substrates are both used as the carbon sources. If Calvin cycle is functional, serine and histidine labeling will be significantly diluted, comparing to other amino acids. Such method can determine the relative CO_2 fixation when organic carbon sources are present in the medium [13].
4. Oxidative pentose phosphate pathway: [1-^{13}C] glucose can be used as the carbon source. If the pathway is active, non-labeled alanine will be >50% [12].
5. Anaplerotic pathway (e.g., PEP + CO_2 à oxaloacetate): $^{13}CO_2$ and non-labeled carbon substrates (e.g., glycerol or pyruvate) can be used as the carbon source. If the pathway is active, aspartate labeling will be significantly enriched, comparing to alanine and serine [13].
6. *Re*-citrate synthase: [1-^{13}C] pyruvate can be used as the carbon source. If the enzyme is active, glutamate is labeled in β-carboxyl group [3, 4].
7. Citramalate pathway: [1-^{13}C] pyruvate, [2-^{13}C] glycerol, or [1-^{13}C] acetate can be used as the carbon source. If the pathway is active, leucine and isoleucine labeling amounts are identical [14].
8. Serine-isocitrate lyase cycle: [1-^{13}C] pyruvate or [1-^{13}C] lactate can be used as the carbon source. If the pathway is active, the third position carbon in serine will be labeled [15].
9. Utilization of nutrients (i.e., exogenous amino acids): a culture medium with fully labeled carbon substrates and non-labeled amino acids can be used. If the cells selectively utilize these supplemented non-labeled nutrients, we will see significant labeling dilution of these amino acids in the biomass. This method can be used to investigate which nutrient supplements are preferred by the cell [16].

6. Representative results:

Recent bioenergy studies have revived interests in using novel phototrophic microorganisms for bioenergy production and CO_2 capture. In the past years, quite a few 13C-metabolism analyses, including advanced 13C-Metabolic Flux Analyses (13C-MFA), have been applied to investigate central metabolisms in phototrophic bacteria, because biochemical knowledge of the central metabolic pathways is not well-founded in these non-model organisms[10, 11, 17-20]. Here, we present an example of the discovery of an alternate isoleucine pathway in *Cyanothece* 51142 [21]. *Cyanothece* 51142 does not contain the enzyme (EC 4.3.1.19, threonine ammonia-lyase), which catalyzes conversion of threonine to 2-ketobutyrate in the typical isoleucine synthesis pathway. To resolve the isoleucine pathway, we grow *Cyanothece* 51142 (20 mL) in ASP2 medium [22] with 54 mM glycerol

(2-13C, >98%). *Cyanothece* 51142 utilizes 2nd position labeled glycerol as the main carbon source. We observe that threonine and alanine have one labeled carbon, while isoleucine is labeled with three carbons. Therefore, synthesis in *Cyanothece* 51142 cannot be derived from the threonine route employed by most organisms (Figure 4). On the other hand, leucine and isoleucine have identical labeling patterns based on fragment (M-15)$^+$ and fragment (M-159)$^+$. For example, the isotopomer data from [M-15]$^+$ (containing unfragmented amino acids) show identical labeling for leucine (M0=0.01, M1=0.03, M2=0.21, M3=0.69) and isoleucine (M0=0.01, M1=0.03, M2=0.24, M3=0.67). Thus leucine and isoleucine must be synthesized from the same precursors (i.e., pyruvate and acetyl-CoA). This observation is consistent with the labeled carbon transition in the citramalate pathway for isoleucine synthesis. To confirm this pathway, we search the Joint Genome Institute database and find the presence of a citramalate synthase *CimA* (cce_0248) in *Cyanothece*.

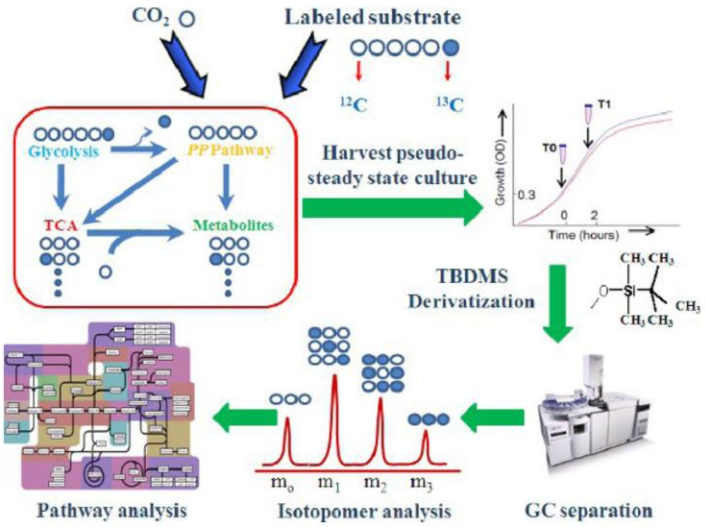


Figure 1. The ^{13}C-assisted pathway analysis steps.

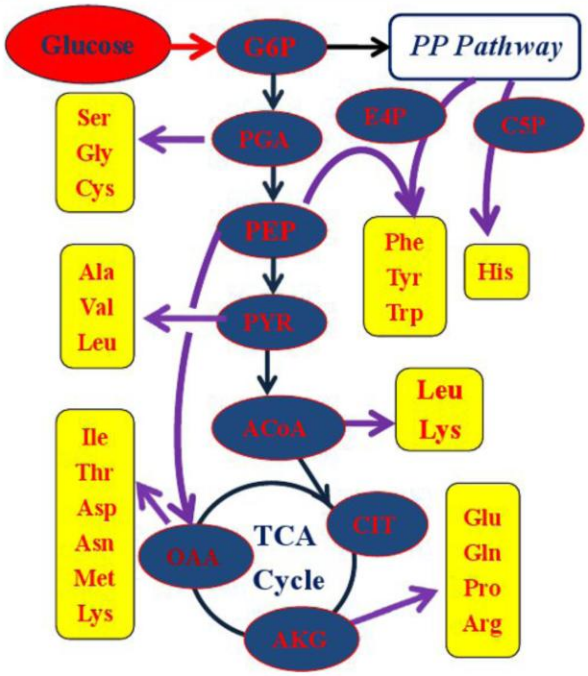


Figure 2. Amino acids used for acquiring the labeling pattern of their metabolic precursors. ACoA, acetyl-CoA; AKG, α-Ketoglutarate; C5P, ribose 5-phosphate; CIT, citrate; E4P, erythrose 4-phosphate; G6P, glucose 6-phosphate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PGA, 3-phosphoglycerate; PYR, pyruvate.

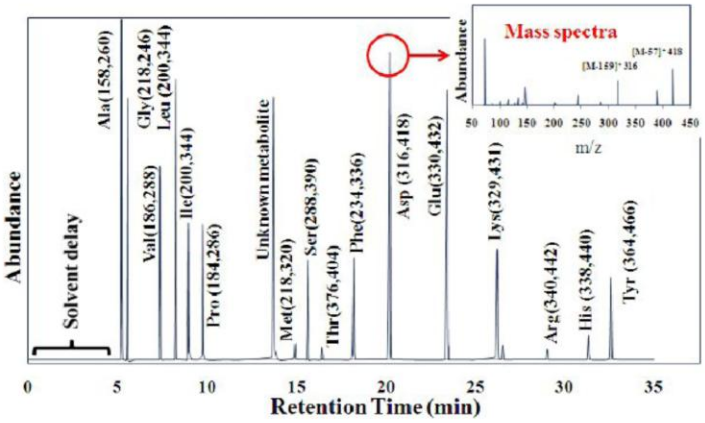


Figure 3. GC peaks for 16 amino acids. TBDMS derivatized amino acids are cracked by MS into two fragments: $(M-57)^+$, containing the entire amino acid, and $(M-159)^+$, which lacks the α carboxyl group of the amino acid. For leucine and isoleucine, the (M-57)+ was overlapped by other mass peaks. We suggest using fragment (M-15)+ to analyze the entire amino acid labeling. The (f302)+ group is detected in most amino acids, which contains only the first (α-carboxyl group) and second carbons in an amino acid backbone. Because this MS peak often has high noise-to-signal ratios, $(f302)^+$ is not recommended for quantitatively analyzing the metabolic fluxes[7].

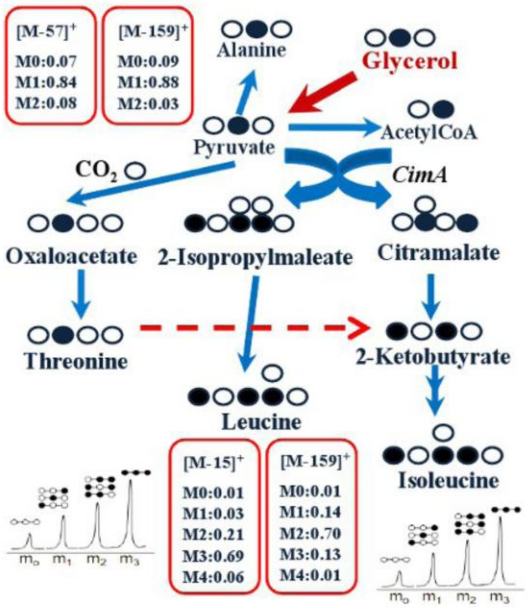


Figure 4. Labeling transitions in isoleucine pathways in Cyanothece 51142 (modified from our previous paper)[21].

Discussion

This protocol consists of feeding the cell with a labeled substrate and measuring the resulting isotopic labeling patterns in the amino acids via GC-MS. Since MS data (m/z ratios) give just the overall amount of labeling of MS ions, we have to assess the isotopomer distributions of amino acids by examining the m/z ratios of both unfragmented $(M-57)^+$ and fragmented amino acids (i.e., $(M-159)^+$ and $(f302)^+$). Furthermore, we can perform several cell cultures with a chemically identical medium but substrates that have different labeling patterns (1^{st} position labeled, 2^{nd} position labeled, etc.). The labeling information about metabolites from these experiments can be integrated to decode the actual carbon transition routes through the central metabolic pathways.

For pathway analysis, the choice of a labeled substrate is important. In general, singly labeled carbon substrates are easier to use in tracing the fate of the labeled carbon when ^{13}C percolates through central pathways, while multiple-carbon labeled substrates may confound carbon tracing. Also, singly labeled substrates are more informative to elucidate unique molecule structures in metabolites than uniformly labeled substrates.[4] For example, the (*Re*)-type citrate synthase shows different reaction stereochemistry from normal citrate synthase, and thus causes citrate to have different molecular chirality. On the other hand, substrates are different in their suitability to detect their associated pathways. Glucose is best for detecting the split ratio between the glycolysis and pentose phosphate pathways, while pyruvate or acetate are best for analyzing the TCA cycle and some amino acid pathways. Therefore, it is necessary to use different substrates to investigate the overall picture of cell metabolisms.

^{13}C-isotope labeling is a useful technique for determining functional pathways in microorganisms. However, this technique has several limitations. ***First***, it is suited only for analysis of carbon metabolism using organic substrates, as it cannot directly resolve metabolism in autotrophic metabolisms if CO_2 is used as the sole carbon source. Autotrophic cultures using CO_2 label all amino acids to the same extent as the input $^{12}CO_2$/$^{13}CO_2$ mixture [23]. This makes pathway analysis difficult, as metabolism analysis has to be inferred from a rearrangement of ^{13}C carbons in metabolites by different metabolic pathways. ***Second***, this paper presents solely qualitative results discriminating between "active" and "non-active" pathways. Precise quantification of metabolism requires a sophisticated modeling approach (i.e., ^{13}C-MFA) to decipher metabolic fluxes from isotopomer data. ***Third***, the scope of ^{13}C-metabolism analysis is limited by technical challenges in determining low abundance and unstable metabolites. Broader metabolic network can be probed by analyzing free metabolites besides amino acids. Measurement of free metabolites requires both highly-efficient metabolite extraction methods and highly-sensitive analytical platforms. LC-MS, FT-ICR MS, and CE-MS have been used for identifying the labeling patterns of free metabolites, and provide more insight into cell metabolisms[2]. ***Fourth***, ^{13}C-assisted pathway analysis is best done in minimal medium, because addition of non-labeled nutrient supplements leads to falsely lower labeling concentrations and make quantitative ^{13}C-MFA studies not so straightforward. Also, cells may utilize exogenous amino acids extensively for protein synthesis, and thus give very weak labeling signals for these proteinogenic amino acids [24]. If a rich medium is required to grow cells, measurement of intracellular metabolites, instead of amino acids, can effectively reduce the interference in labeling data that arises from exogenous non-labeled carbon nutrients.

Finally, an increasing number of genome sequences for non-model microbial species are being published each year. However, functional characterization of these species has lagged far behind the pace of genomic sequencing. ^{13}C-labeling approaches can play important roles in the

confirmation and discovery of metabolic pathways in many non-model organisms. Furthermore, the labeling information can be integrated with metabolic modeling (^{13}C-MFA) to decipher absolute carbon fluxes in microorganisms [25]. Therefore, this technique can be widely used in analyzing biological systems related to biofuel, ecological and medical applications.

Disclosures

No conflicts of interest declared.

Acknowledgements

This study was supported by an NSF Career Grant (MCB0954016) and a DOE Bioenergy Research Grant (DEFG0208ER64694).

References

1. Zamboni, N., & Sauer, U. Novel biological insights through metabolomics and ^{13}C-flux analysis. *Curr Opin Microbiol.* **12**, 553-558 (2009).
2. Tang, Y.J. *et al.* Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. *Mass Spectrom Rev* **28**, 362-375 (2009).
3. Tang, Y.J. *et al.* Investigation of carbon metabolism in "*Dehalococcoides ethenogenes*" strain 195 via isotopic and transcriptomic analysisa. *J. Bacteriol.* **191**, 5224-5231 (2009).
4. Tang, Y.J. *et al.* Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using GC-MS and FT-ICR mass spectrometry. *Journal of Bacteriology* **189**, 940-949 (2007).
5. Dauner, M.S., U. GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. *Biotechnology Progress* **16**, 642-649 (2000).
6. Wittmann, C. Fluxome analysis using GC-MS. *Microbial Cell Factories* **6**, 6 (2007).
7. Antoniewicz, M.R., Kelleher, J.K., & Stephanopoulos, G. Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. *Anal. Chem.* **79**, 7554-7559 (2007).
8. Wahl, S.A., Dauner, M., & Wiechert, W. New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. *Biotechnology and Bioengineering* **85**, 259-268 (2004).
9. Shaikh, A., Tang, Y.J., Mukhopadhyay, A., & Keasling, J.D. Isotopomer distributions in amino acids from a highly expressed protein as a proxy for those from total protein. *Analytical Chemistry* **80(3)**, 886-890 (2008).
10. Tang, K.-H., Feng, X., Tang, Y.J., & Blankenship, R.E. Carbohydrate metabolism and carbon fixation in *Roseobacter denitrificans* OCh114. *PLoS One* **4**, e7233 (2009).
11. Tang, K.-H. *et al.* Carbon flow of Heliobacterium modesticaldum is more related to Firmicutes than to the green sulfur bacteria. *J Biol Chem* **285**, 35104-35112 (2010).
12. Feng, X. *et al.* Characterization of the Central Metabolic Pathways in Thermoanaerobacter sp. X514 via Isotopomer-Assisted Metabolite Analysis. *Appl. Environ. Microbiol.* **75**, 5001-5008 (2009).
13. Feng, X. *et al.* Mixotrophic and photoheterotrophic metabolisms in Cyanothece sp. ATCC 51142 under continuous light. *Microbiology* **156**, 2566 - 2574 (2010).
14. Tang, Y.J. *et al.* Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-NTA. *Appl. Environ. Microbiol.* **73**, 3859-3864 (2007).
15. Tang, Y.J., Meadows, A.L., Kirby, J., & Keasling, J.D. Anaerobic central metabolic pathways in *Shewanella oneidensis* MR-1 reinterpreted in the light of isotopic metabolite labeling. *Journal of Bacteriology* **189**, 894-901 (2007).
16. Zhuang, W.Q. *et al.* Selective utilization of exogenous amino acids by *Dehalococcoides ethenogenes* strain 195 and the enhancement resulted to dechloronation activity. *Appl Environ Microbiol*, Accept (2011).
17. Feng, X., Tang, K.-H., Blankenship, R.E., & Tang, Y.J. Metabolic flux analysis of the mixotrophic metabolisms in the green sulfur bacterium *Chlorobaculum tepidum*. *J Biol Chem* **285**, 35104-35112 (2010).
18. McKinlay, J.B., & Harwood, C.S. Carbon dioxide fixation as a central redox cofactor recycling mechanism in bacteria. *Proc Natl Acad Sci U S A* **107** (2010).
19. McKinlay, J.B., & Harwood, C.S. Calvin cycle flux, pathway constraints, and substrate oxidation state together determine the H2 biofuel yield in photoheterotrophic bacteria. *MBio* **2** (2011).
20. Erb, T.J. *et al.* Synthesis of C5-dicarboxylic acids from C2-units involving crotonyl-CoA carboxylase/reductase: The ethylmalonyl-CoA pathway. *PNAS* **104**, 10631-10636 (2007).
21. Wu, B. *et al.* Alternative isoleucine synthesis pathway in cyanobacterial species. *Microbiology* **156**, 596-602 (2010).
22. Reddy, K.J., Haskell, J.B., Sherman, D.M., & Sherman, L.A. Unicellular, aerobic nitrogen-fixing cyanobacteria of the genus Cyanothece. *J Bacteriol.* **175**, 1284-1292 (1993).
23. Shastri, A.A., & Morgan, J.A. A transient isotopic labeling methodology for ^{13}C metabolic flux analysis of photoautotrophic microorganisms. *Phytochemistry* **68**, 2302-2312 (2007).
24. Tang, Y.J. *et al.* Invariability of central metabolic flux distribution in *Shewanella oneidensis* MR-1 under environmental or genetic perturbations *Biotechnol Prog.* **25**, 1254-1259 (2009).
25. Zamboni, N., Fendt, S.M., Ruhl, M., & Sauer, U. ^{13}C-based metabolic flux analysis. *Nature Protocols* **4**, 878-892 (2009).

Appendix 2

Alternative Isoleucine Synthesis Pathway in

Cyanobacterial Species

Microbiology (2010), **156**, 596–602 DOI 10.1099/mic.0.031799-0

Alternative isoleucine synthesis pathway in cyanobacterial species

Bing Wu,[1]† Baichen Zhang,[2]† Xueyang Feng,[1]† Jacob R. Rubens,[3] Rick Huang,[1] Leslie M. Hicks,[2] Himadri B. Pakrasi[1,3] and Yinjie J. Tang[1]

Correspondence
Yinjie J. Tang
yinjie.tang@seas.wustl.edu

[1]Department of Energy, Environmental and Chemical Engineering, Washington University, St Louis, MO 63130, USA

[2]Donald Danforth Plant Science Center, 975 N. Warson Rd, St Louis, MO 63132, USA

[3]Department of Biology, Washington University, St Louis, MO 63130, USA

Cyanothece sp. ATCC 51142 is an aerobic N_2-fixing and hydrogen-producing cyanobacterium. Isotopomer analysis of its amino acids revealed an identical labelling profile for leucine and isoleucine when *Cyanothece* 51142 was grown mixotrophically using 2-^{13}C-labelled glycerol as the main carbon source. This indicated that *Cyanothece* 51142 employs the atypical alternative citramalate pathway for isoleucine synthesis, with pyruvate and acetyl-CoA as precursors. Utilization of the citramalate pathway was confirmed by an enzyme assay and LC-MS/MS analysis. Furthermore, the genome sequence of *Cyanothece* 51142 shows that the gene encoding the key enzyme (threonine ammonia-lyase) in the normal isoleucine pathway is missing. Instead, the cce_0248 gene in *Cyanothece* 51142 exhibits 53 % identity to the gene encoding citramalate synthase (CimA, GSU1798) from *Geobacter sulfurreducens*. Reverse-transcription PCR indicated that the cce_0248 gene is expressed and its transcriptional level is lower in medium with isoleucine than in isoleucine-free medium. Additionally, a BLAST search for citramalate synthase and threonine ammonia-lyase implies that this alternative isoleucine synthesis pathway may be present in other cyanobacteria, such as *Cyanothece* and *Synechococcus*. This suggests that the pathway is more widespread than originally thought, as previous identifications of the citramalate pathway are limited to mostly anaerobic bacteria or archaea. Furthermore, this discovery opens the possibility that such autrotrophic micro-organisms may be engineered for robust butanol and propanol production from 2-ketobutyrate, which is an intermediate in the isoleucine biosynthesis pathway.

Received 16 June 2009
Revised 23 October 2009
Accepted 28 October 2009

INTRODUCTION

Cyanothece sp. ATCC 51142 is an aerobic unicellular marine cyanobacterium that has a robust diurnal cycle for photosynthesis and nitrogen fixation and is also capable of evolving hydrogen (Colón-López & Sherman, 1998; Stöckel *et al.*, 2008; Toepel *et al.*, 2008). Previous transcriptional network studies on *Cyanothece* sp. have shown unique periodic changes in its central metabolism throughout the day–night cycle (Stöckel *et al.*, 2008). The genome of this strain has recently been sequenced (Welsh *et al.*, 2008) and the enzyme (threonine ammonia-lyase, EC 4.3.1.19) catalysing the conversion of threonine to 2-ketobutyrate in the normal isoleucine synthesis pathway is not annotated based on KEGG pathway maps (http://www.genome.jp/kegg/). However, the key intermediate in isoleucine biosynthesis pathways, 2-ketobutyrate, can also be synthesized from citramalate. While most bacteria employ the threonine pathway to form isoleucine, some anaerobic bacteria and archaea, such as *Methanococcus jannaschii* and *Geobacter sulfurreducens*, can synthesize isoleucine from citramalate via condensation of acetyl-CoA and pyruvate catalysed by citramalate synthase (CimA) (Howell *et al.*, 1999; Risso *et al.*, 2008; Tang *et al.*, 2009b; Xu *et al.*, 2004).

In this study, we investigated the actual route for isoleucine synthesis in *Cyanothece* 51142 via ^{13}C-based isotopomer analysis. These experiments consisted of feeding the culture a defined ^{13}C-labelled substrate and measuring the isotopic enrichment in metabolites once the labelled carbon had percolated through the metabolic network. The isotopomer labelling information allowed us to decipher carbon transitions in the isoleucine synthesis pathways (Tang *et al.*, 2009a). The identification of such enzymes can

†These authors contributed equally to this work.

Abbreviations: CID, collision-induced dissociation; MRM, multiple reaction monitoring.

Two supplementary figures are available with the online version of this paper.

improve our understanding of carbon metabolism in *Cyanothece* 51142 as well as in other recently sequenced cyanobacterial strains. Furthermore, because an intermediate in the citramalate pathway is 2-ketobutyrate (the precursor of 1-propanol and 1-butanol: Atsumi & Liao, 2008), this discovery may have potential applications to CO_2-based biofuel research by making cyanobacteria a viable vehicle for the autotrophic production of butanol and propanol.

METHODS

Culture conditions. *Cyanothece* 51142 was grown in 150 ml Erlenmeyer flasks fed with ASP2 medium (Toepel *et al.*, 2008). The medium composition for 1 litre was as follows: Na_2EDTA, 0.03 g; K_2HPO_4, 0.05 g; KCl, 0.6 g; NaCl, 18 g; $NaNO_3$, 1.5 g; $MgSO_4$, 2.24 g; $CaCl_2.2H_2O$, 0.37 mg; $FeCl_3.6H_2O$, 3.9 mg; HCl, 0.1 µmol; TAPS, 1 g; TAPSO, 1 g; H_3BO_3, 34.3 mg; $MnCl_2.4H_2O$, 4.32 mg; $ZnSO_4.7H_2O$, 0.668 mg; $CuSO_4.5H_2O$, 3 µg; $CoCl_2.6H_2O$, 12 µg; $Na_2MoO_4.2H_2O$, 44 µg. The carbon source was [2-^{13}C]glycerol (99 %, Cambridge Isotope Laboratories), at an initial concentration of 7 g l^{-1}. The strain was initially grown in 20 ml labelled culture medium (with 3 % inoculum volume). At the mid-exponential phase of growth, a 3 % inoculum was added to a 50 ml subculture containing the same labelled medium, which reduced the effect of unlabelled carbon from the initial stock. All cultures were grown aerobically at 30 °C under continuous light (50 µmol photons m^{-2} s^{-1}) in 125 ml flasks shaking at 150 r.p.m.

Analytical measurements. During the growth of *Cyanothece* 51142, cell density was monitored on a UV-Vis spectrophotometer (Genesys, Thermo Scientific) at 730 nm. At each time point, samples were collected and glycerol concentrations were measured using a glycerol enzymic assay kit (R-Biopharm). When the growth of the strain approached the mid-exponential phase, the biomass was harvested by centrifugation (10 000 ***g***, 10 min), and the cell pellet was hydrolysed in 6 M HCl at 100 °C for 24 h. After air-drying overnight, the dried samples (containing free amino acids) were derivatized in tetrahydrofuran and *N*-(tert-butyldimethylsilyl)-*N*-methyltrifluoroacetamide (Sigma-Aldrich) at 70 °C for 1 h. Isotopomer measurements were made on a gas chromatograph (Hewlett Packard, model 6890, Agilent Technologies) equipped with a DB5-MS column (J&W Scientific) and a mass spectrometer (5975, Agilent Technologies). Four types of charged fragments were detected by GC-MS for the amino acids: the $[M-57]^+$ or $[M-15]^+$ group, which contains unfragmented amino acids; and the $[M-159]^+$ or $[M-85]^+$ group, which contains amino acids having lost α-carboxyl groups. Because the $[M-57]^+$ peaks in leucine and isoleucine overlap with other peaks, the $[M-15]^+$ group was used to obtain the isotopomer labelling information of the unfragmented leucine and isoleucine. When amino acids are derivatized for GC separation, derivatization groups introduce the natural isotopes, including ^{13}C (1.13 %), ^{18}O (0.20 %), ^{29}Si (4.70 %) and ^{30}Si (3.09 %). Published algorithms were used to correct for the effects of those natural isotopes on the mass distributions of amino acids (Wahl *et al.*, 2004)

LC-MS/MS (composed of a Shimadzu HPLC system with a LEAP PAL autosampler and a 4000 QTRAP with TurboIonSpray ion source, Applied Biosystems) was used to confirm the production of citramalate by *Cyanothece* 51142. A 5 ml sample of culture taken during the exponential growth phase was centrifuged for 10 min at 8000 ***g***. The intracellular metabolites were extracted from the cell pellet using 2 ml cold methanol and 2 ml chloroform. Then 1 ml distilled water (0 °C) was added to partition the solution. The hydrophobic layer at the bottom was removed, and extracted metabolites in the aqueous solution were lyophilized and reconstituted in 1 % formic acid. After vortexing and sonication, the samples were centrifuged at room temperature for 4 min at 13 200 ***g*** to remove water-insoluble components. The LC gradient was optimized by injecting a citramalate standard that was detected by MS in multiple reaction monitoring (MRM) mode (Heinig & Henion, 1999). Under the optimized LC-MS/MS conditions, 50 µl of test sample was injected into the LC-MS/MS. Polar acidic components in the sample were resolved by an Onyx Monolith column (300 mm × 4.6 mm, Phenomenex) in gradient elution mode. Negative electrospray ionization (ESI) mode was used for detecting citramalate in both MRM and MS/MS scan modes. Collision-induced dissociation (CID) spectra of the citramalate parent ion *m/z* 147 $(M-H)^-$ and MRM parameters were optimized for sensitive detection.

An identical LC-MS setup was used for measurement of citramalate's labelling pattern with the following modifications: (1) the HPLC system was changed to the Agilent 1200 HPLC system; (2) the injection volume was set to 40 µl. After obtaining citramalate mass patterns, ^{13}C-labelled citramalate was subjected to MS/MS fragment analysis using the QTRAP 4000 system. If citramalate was labelled, the pseudo-molecular parent ion of citramalate (*m/z* 147, no loss of carbons) and its fragment ions (*m/z* 57, 85 and 87) would have a higher mass-to-charge ratio. Such information could be used to determine the isotopic labelling pattern of citramalate.

Enzymic activity assay. The methods of Howell *et al.* (1999) and Risso *et al.* (2008) were used. A wet weight of 100–200 mg of cell pellet was suspended in 0.1 M TES buffer, pH 7.5. The cell extract was prepared by sonication (MISONIX) of the cell pellet for 3 min with a 30 s on/20 s off cycle. A 100 µl sample of cell extract was mixed with TES buffer (0.2 M, pH 7.5, 500 µl), pyruvate (10 mM, 100 µl) and acetyl-CoA (50 mM, 20 µl), then topped up with distilled water to a final volume of 1000 µl. Blank samples were prepared as for the test samples but without pyruvate. The resulting test sample and blank sample solutions were then incubated at 30, 45 or 60 °C in an oven for 2 h. At intervals of 20 min, a 100 µl test sample and a blank sample were taken from the incubator and added to a solution containing 50 µl 10 mM DTNB [5,5′-dithiobis(2-nitrobenzoic acid)] in 0.1 M Tris/HCl (pH 8.0), 70 µl 1 M Tris/HCl (pH 8.0), and distilled water to a final volume of 0.9 ml. The A_{412} values for the test sample and blank sample were recorded. The production of SH-CoA was determined based on a standard curve generated with 2-mercaptoethanol. The enzyme activity was assayed by detecting the production of SH-CoA over 2 h. All the chemicals employed in this assay were from Sigma-Aldrich.

RNA extraction and reverse transcription PCR (RT-PCR). In order to confirm the expression of cce_0248 (encoding CimA), *Cyanothece* 51142 was cultured in ASP2 medium with glycerol (7 g l^{-1}). In the exponential growth phase (OD_{730} ~0.6), isoleucine was added to the culture to a final concentration of 10 mM. The control experiment (without addition of isoleucine) was performed in parallel and the both culture samples were collected after 30 min. The RT-PCR protocol was based on a previous report (Johnson *et al.*, 2005). In brief, total RNA was extracted from the frozen cell pellets using RNAwiz reagent (Ambion) according to the product manual. Contaminant DNA was removed by DNase treatment using an RQ1 RNase-Free DNase kit (Promega) in accordance with the manufacturer's protocol. The extracted RNA was transcribed to cDNA by a SuperScript II Reverse Transcriptase kit (Invitrogen), also according to the manufacturer's protocol. The forward primer TTAGCTGCAGGAACACGATG and reverse primer TCTCGAACAATGCGACTGAC were employed to amplify the cce_0248 gene, the forward primer AGAGGATGAGCAGCCACACT and reverse primer TAATTCCGGATAACGCTTGC to amplify the *rrn* 16Sa gene (i.e. 16S

rRNA) as a positive control gene, and the forward primer GACCCCCATTAAAGCGAGAA and reverse primer TTAACCAAGGAGGCGGATTT to amplify the *nifX* gene (as a negative control gene) in the cDNA by PCR using the Platinum *Pfx* DNA Polymerase kit (Invitrogen) (Welsh *et al.*, 2008). PCRs were conducted with the following cycle conditions: 2 min of activation of the polymerase at 94 °C followed by 30 cycles consisting of 1 min at 94 °C, 30 s at 58 °C and 2 min at 72 °C; finally, a 10 min extension process was performed at 72 °C. The final PCR product was observed directly on agarose gels after electrophoresis.

RESULTS AND DISCUSSION

Cyanothece 51142 exhibited a maximum specific growth rate of ~0.9 day^{-1} in ASP2 medium in the presence of glycerol during the exponential phase (Fig. 1). The glycerol was quickly utilized for biomass production during the exponential growth phase and most key amino acids were highly labelled (Fig. 2). For example, >90 % of alanine was labelled with one or more carbons, indicating that a significant amount of labelled glycerol was directed into the central metabolism (glycolysis, TCA and pentose phosphate pathway), which produced the building blocks for amino acids and other metabolites. In comparison, CO_2 (unlabelled carbon) fixation via the Calvin cycle was minimal when glycerol was present in the medium (only 7 % of alanine was unlabelled).

The isotopomer data from $[M-159]^+$ (loss of the α-carboxyl group) showed that the labelling profiles for leucine (M0=0.01, M1=0.14, M2=0.70, M3=0.13; M0, M1, M2... are fractions of unlabelled, singly labelled and doubly labelled amino acids, respectively) and isoleucine (M0=0.01, M1=0.14, M2=0.71, M3=0.13) were similar (the standard error for isotopomer analysis is <0.05). Furthermore, the isotopomer data from $[M-15]^+$ (containing unfragmented amino acids) also showed identical labelling for leucine (M0=0.01, M1=0.03, M2=0.21, M3=0.69) and isoleucine (M0=0.01, M1=0.03, M2=0.24, M3=0.67). These data suggest that leucine and isoleucine are synthesized from the same precursors (i.e. pyruvate and acetyl-CoA), and thus that isoleucine synthesis in *Cyanothece* 51142 takes place via an alternative pathway to the threonine-pyruvate pathway employed by most organisms. Further supporting this conclusion, isoleucine could not be labelled with three carbons ($[M-15]^+$: M3=0.67) if threonine and pyruvate were its precursors: threonine was mostly labelled with one carbon ($[M-57]^+$, M1=0.64) and alanine (pyruvate as its precursor) was also mostly labelled with one carbon (alanine $[M-57]^+$, M1=0.84) (Fig. 2).

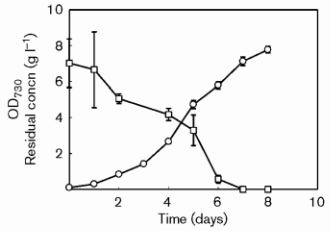


Fig. 1. Growth of *Cyanothece* 51142 and utilization of glycerol. ○, OD_{730}; □, glycerol concentration.

The citramalate pathway via citramalate synthase is an alternative to the common threonine pathway for the synthesis of isoleucine. Citramalate synthase (CimA), which catalyses the synthesis of the isoleucine precursor 2-ketobutyrate, has been discovered in other organisms via isotopomer analysis, enzyme chemistry or cloning methods (Table 1). Two types of genes encoding CimA have been identified and both genes are homologous to homocitrate synthase. CimA was previously reported in an archaeal species (*Methanococcus jannaschii*, gene MJ1392) (Howell *et al.*, 1999) and recently a new CimA gene homologue (*Geobacter sulfurreducens*, gene GSU1798) has been discovered (Risso *et al.*, 2008); further CimA identifications are listed in Table 1. According to the genome sequence of *Cyanothece* 51142 (from the Joint Genome Institute: http://www.jgi.doe.gov), the gene for threonine-ammonia lyase (EC 4.3.1.19) catalysing the conversion of threonine to 2-ketobutyrate is missing. We also confirmed the absence of this gene from the *Cyanothece* 51142 genome by BLAST analysis (Table 2). In contrast, BLAST search results indicate the presence of a complete citramalate pathway for isoleucine synthesis and the gene for CimA is identified as cce_0248, which has 53 % identity to the reported citramalate synthase gene (GSU1798) in *G. sulfurreducens*.

To confirm citramalate synthase activity in *Cyanothece* 51142, the specific activities of this enzyme were measured. The enzyme activities in crude cell extracts at elevated temperatures (30, 45 and 60 °C) were 8.54 ± 2.24, 5.69 ± 1.79 and 2.79 ± 0.07 nmol min^{-1} (mg dry biomass)$^{-1}$, respectively. In comparison, the citramalate synthase from the thermophilic *M. jannaschii* shows higher thermostability (Howell *et al.*, 1999). However, this enzyme assay does not conclusively prove the presence of citramalate synthase (Risso *et al.*, 2008) and therefore high-sensitivity and selectivity LC-MS/MS was used to detect intracellular citramalate concentrations in *Cyanothece* 51142. Initially, the LC-MS/MS conditions were optimized using the signal of an authentic citramalate standard. MRM-based methods then confidently identified citramalate from cell extracts in our unlabelled cultures (Fig. 3). We spiked the cell extracts with an authentic citramalate standard and the spiked standard co-eluted with endogenous citramalate as a single peak. The measured intracellular concentration of citramalate [0.12 ± 0.03 μmol (g dry biomass)$^{-1}$] was smaller than the reported pools of central metabolites [such as α-ketoglutarate and malate, ~2–4 μmol (g dry biomass)$^{-1}$] in cyanobacteria (such as *Synechocystis*) (Shastri & Morgan,

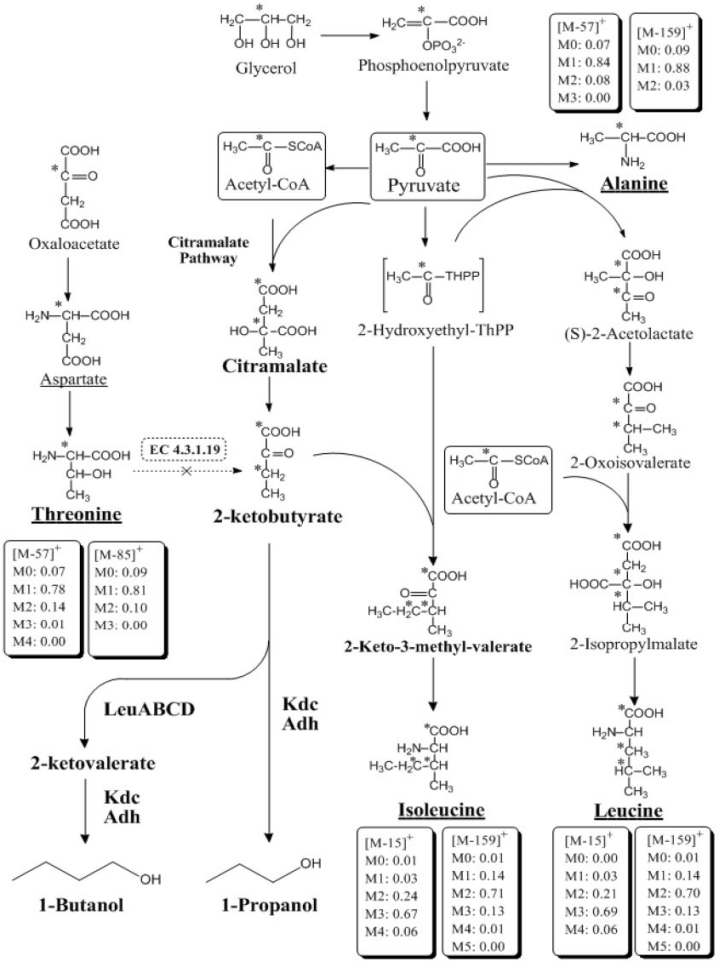


Fig. 2. An alternative isoleucine pathway in *Cyanothece* 51142. The standard errors of the GC-MS data from labelling experiments are below 0.05 (n=2). * represents the labelled carbon. Kdc, 2-keto acid decarboxylase; Adh, alcohol dehydrogenase; LeuABCD, isopropylmalate-related enzymes. The detected labelling patterns in pyruvate, threonine, isoleucine and leucine exclude a biosynthetic pathway via threonine as an intermediate but are in line with the citramalate pathway.

2007). This lower concentration is consistent with the proposed role for citramalate as a secondary metabolite in *Cyanothece* 51142. Furthermore, the citramalate labelling pattern resulting from the tracer experiment (using [2-^{13}C]glycerol as the carbon source) was also investigated. The spectra of the parent ion $(M-H)^-$ had a mass-to-

Table 1. Reported threonine-independent isoleucine synthesis pathways

Species	Reference(s)	Main methods	*cimA* similarity (GSU 1798)	*cimA* similarity (MJ1392)	Growth conditions
Methanobacterium thermoautotrophicum	Eikmanns *et al.* (1983)	^{14}C-labelling experiments	MTH1481 28 %	MTH723 58 %	Anaerobic
Methanococcus jannaschii	Howell *et al.* (1999)	*cimA* gene cloned and expressed in *E. coli*	ND	MJ1392 100 %	Aerobic (*E. coli*)
Leptospira interrogans	Charon *et al.* (1974); Westfall *et al.* (1983) Xu *et al.* (2004)	^{14}C-labelling experiments *cimA* gene cloned and expressed in *E. coli*	LIC11726 26 %	LIC11726 41 %	Aerobic (*E. coli*)
Thermoproteus neutrophilus	Schäfer *et al.* (1989)	^{13}C NMR	Tneu_0320 45 %	Tneu_0832 55 %	Anaerobic, thermophilic
Ignicoccus hospitalis	Jahn *et al.* (2007)	Enzyme assays; ^{13}C-labelling experiments	Igni_0645 45 %	Igni_0983 52 %	Anaerobic, thermophilic
Geobacter sulfurreducens	Risso *et al.* (2008)	^{13}C GC-MS and gene/enzyme assays	GSU1798 100 %	GSU1906 41 %	Anaerobic
Geobacter metallireducens	Tang *et al.* (2007)	^{13}C GC-MS	Gmet_1879 92 %	Gmet_1265 42 %	Anaerobic
Serratia marcescens	Kisumi *et al.* (1977)	Enzyme assays	Spro_0745 (26 %)*	Spro_0745 (37 %)*	Anaerobic
Thermoanaerobacter sp. X514	Feng *et al.* (2009)	^{13}C GC-MS; enzyme assays	Teth514_1204 49 %	Teth514_0415 45 %	Anaerobic, thermophilic
Dehalococcoides ethenogenes 195	Tang *et al.* (2009c)	^{13}C GC-MS; qPCR	DET0825 53 %	DET0830 41 %	Anaerobic

ND, Not determined.
*Sequence similarity is based on *Serratia proteamaculans* 568 because the *Serratia marcescens* sequence is not available.

charge ratio of 149 (from the ^{13}C-labelled sample) instead of 147 (from the unlabelled sample), which indicates that citramalate was mostly labelled with two carbons in this experiment. The two labelled carbons in citramalate were derived from pyruvate and acetyl-CoA (Fig. 2). The carbon labelling positions in citramalate were determined by

Table 2. BLAST searches of amino acid sequences for key genes similar to the target genes (GSU1798 in *G. sulfurreducens* and MJ1392 in *M. jannaschii*, encoding CimA) and threonine ammonia-lyase (b3772 in *E. coli* K-12) in several sequenced cyanobacterial species

Cyanobacterium	Strain	GSU1798		MJ1392		b3772	
		Gene	Identity (%)	Gene	Identity (%)	Gene	Identity (%)
Cyanothece sp.	ATCC 51142	cce_0248	53	cce_4008	41	cce_1455	28
	CCY 0110	CY 0110_21607	53	CY0110_30221	41	CY0110_21225	29
	PCC 7424	PCC7424_3413	53	PCC7424_1126	41	PCC7424_0305	34
	PCC 7425	Cyan7425_0344	25	Cyan7425_0344	41	Cyan7425_3269	50
	PCC 7822	Cyan7822DRAFT_4501	53	Cyan7822DRAFT_3533	41	Cyan7822DRAFT_1627	29
	PCC 8801	PCC8801_0449	55	PCC8801_0832	41	PCC8801_1025	28
	PCC 8802	Cyan8802DRAFT_1361	54	Cyan8802DRAFT_0632	41	Cyan8802DRAFT_3061	28
Synechococcus sp.	*elongatus* PCC 6301	syc 1089_d	53	syc0145_d	40	syc1498_c	33
	elongatus PCC 7942	Synpcc7942_0428	53	Synpcc7942_1410	40	Synpcc7942_2612	33
	JA-2-3 Ba (2-13)	CYB_1307	51	CYB_1442	43	CYB_2298	27
	JA-3-3 Ab	CYA_0566	51	CYA_2223	43	CYA_0143	51
	PCC 7002	SYNPCC7002_A0730	51	SYNPCC7002_A1356	41	SYNPCC7002_A1616	50
Synechocystis sp.	PCC 6803	sll1564	52	slr0186	41	slr2072	49

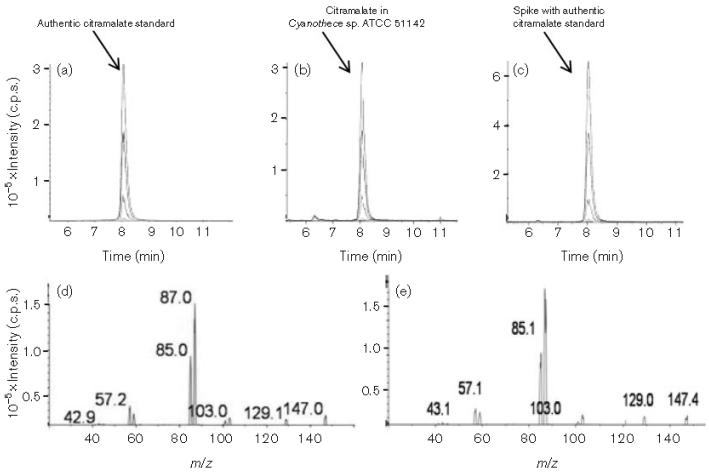


Fig. 3. LC-MS/MS spectra for citramalate: (a) MRM signals of authentic citramalate standard; (b) MRM signals of cell extracts; (c) MRM signals of cell extracts spiked with authentic citramalate standard; (d) MS/MS CID spectra from authentic citramalate standard; (e) MS/MS CID spectra from extracted citramalate in *Cyanothece* 51142.

analysing the fragments of the parent ion (see Supplementary Fig. S1, available with the online version of this paper). It was concluded that citramalate was labelled predominantly at the C-2 and C-4 positions, based on MRM experiments and MS/MS fragmentation of ^{13}C-labelled citramalate (parent ion *m/z* 149) (Fig. 2).

Finally, expression of cce_0248, the key gene in *Cyanothece* 51142 that encodes the putative citramalate synthase, was confirmed by RT-PCR. The RT-PCR products from the amplification of cce_0248 and the 16S rRNA gene (as a positive control) were distinct and as expected, although the expression of cce_0248 was lower than that of the16S rRNA gene (Supplementary Fig. S2). The *nifX* gene (a gene involved in the fixation of atmospheric nitrogen) was not expressed under either of the culture conditions (with or without isoleucine), because the ASP2 medium contained nitrate as the nitrogen source. So the *nifX* gene was treated as a negative control gene. RT-PCR products also showed that the presence of isoleucine (10 mM) in the medium reduced the expression of cce_0248 relative to the control experiments. This result could be explained by a previous report that citramalate synthase was subject to feedback inhibition by 5 mM isoleucine (Xu *et al.*, 2004).

Our findings led us to search for further homologues of CimA of *G. sulfurreducens* in other species and we found that homologues are present in a wide diversity of cyanobacteria. BLAST searches for citramalate synthase (GSU1798) and threonine ammonia-lyase (b3772) were conducted in *Synechocystis* sp. PCC 6803, seven strains of *Cyanothece* sp. and 16 strains of *Synechococcus* sp., based on amino acid sequence similarity. The BLAST results (Table 2) indicate that at least six strains of *Cyanothece* sp. and five strains of *Synechococcus* sp. have high probabilities of employing the citramalate pathway to synthesize isoleucine rather than the normal isoleucine pathway via threonine ammonia-lyase (CimA identity >50 %). *Synechococcus* JA-3-3 Ab and PCC 7002, and *Synechocystis* PCC 6803, may be capable of employing two different pathways for isoleucine synthesis. Twelve strains of *Prochlorococcus* sp. have also been examined (data not shown), and all of them appeared not to have the genes encoding the identical CimA based on the KEGG pathway map (http://www.genome.jp/kegg/); these strains may only have the normal isoleucine pathway via threonine ammonia-lyase. The predictions in Table 2 require further experimental verification. For comparison, the two other genes closest to GSU1798 in sequence identity within the cyanobacteria have relatively low sequence identities (20–30 %; data not shown) to GSU1798; these genes encode 2-isopropylmalate synthase (which participates in the biosynthesis of L-leucine) and homocitrate synthase.

Conclusions

Our isotopomer information, enzyme activity measurement, metabolite analysis and gene expression data unambiguously

indicate that the citramalate pathway is utilized for isoleucine synthesis in *Cyanothece* 51142. Furthermore, genome sequences suggest that quite a few other cyanobacterial strains may also employ the citramalate pathway for isoleucine synthesis. To our knowledge, citramalate pathways have previously been identified only in some anaerobic bacteria and archaea. The discovery of this alternative isoleucine pathway in cyanobacteria may help us understand their unique metabolic regulation related to photosynthesis and nitrogen fixation. Moreover, the citramalate pathway may be utilized to synthesize 2-ketobutyrate, a precursor of both butanol and propanol. As such, the discovery of citramalate synthase in cyanobacteria may provide a new biofuel synthesis route utilizing autotrophic micro-organisms.

ACKNOWLEDGEMENTS

The authors would like to express their gratitude to Shan Yi (University of California, Berkeley), Bert Berla, Jana Stöckel and LeeAnn Perry (Washington University in St. Louis) for their help in RT-PCR experiments. This study was supported in part by a DOE bioenergy research grant (DEFG0208ER64694) and in part by a grant from I-CARES (International Center for Advanced Renewable Energy and Sustainability) at Washington University in St Louis to the Tang lab. Funding for the 4000 QTRAP LC-MS/MS was provided by NSF-MRI grant no. DBI-0521250 to the Hicks lab, and that for the Agilent GC-MS was provided by the Missouri Life Sciences Trust Fund to the Pakrasi lab.

REFERENCES

Atsumi, S. & Liao, J. C. (2008). Directed evolution of *Methanococcus jannaschii* citramalate synthase for biosynthesis of 1-propanol and 1-butanol by *Escherichia coli*. *Appl Environ Microbiol* **74**, 7802–7808.

Charon, N. W., Johnson, R. C. & Peterson, D. (1974). Amino acid biosynthesis in the spirochete *Leptospira*: evidence for a novel pathway of isoleucine biosynthesis. *J Bacteriol* **117**, 203–211.

Colón-López, M. S. & Sherman, L. A. (1998). Transcriptional and translational regulation of photosystem I and II genes in light-dark- and continuous-light-grown cultures of the unicellular cyanobacterium *Cyanothece* sp. strain ATCC 51142. *J Bacteriol* **180**, 519–526.

Eikmanns, B., Linder, D. & Thauer, R. K. (1983). Unusual pathway of isoleucine biosynthesis in *Methanobacterium thermoautotrophicum*. *Arch Microbiol* **136**, 111–113.

Feng, X., Mouttaki, H., Lin, L., Huang, R., Wu, B., Hemme, C. L., He, Z., Zhang, B., Hicks, L. M. & other authors (2009). Characterization of the central metabolic pathways in *Thermoanaerobacter* sp. X514 via isotopomer-assisted metabolite analysis. *Appl Environ Microbiol* **75**, 5001–5008.

Heinig, K. & Henion, J. (1999). Fast liquid chromatographic-mass spectrometric determination of pharmaceutical compounds. *J Chromatogr B Biomed Sci Appl* **732**, 445–458.

Howell, D. M., Xu, H. M. & White, R. H. (1999). (*R*)-Citramalate synthase in methanogenic archaea. *J Bacteriol* **181**, 331–333.

Jahn, U., Huber, H., Eisenreich, W., Hugler, M. & Fuchs, G. (2007). Insights into the autotrophic CO_2 fixation pathway of the archaeon *Ignicoccus hospitalis*: comprehensive analysis of the central carbon metabolism. *J Bacteriol* **189**, 4108–4119.

Johnson, D. R., Lee, P. K. H., Holmes, V. F., Fortin, A. C. & Alvarez-Cohen, L. (2005). Transcriptional expression of the *tceA* gene in a *Dehalococcoides*-containing microbial enrichment. *Appl Environ Microbiol* **71**, 7145–7151.

Kisumi, M., Komatsubara, S. & Chibata, I. (1977). Pathway for isoleucine formation form pyruvate by leucine biosynthetic enzymes in leucine-accumulating isoleucine revertants of *Serratia marcescens*. *J Biochem* **82**, 95–103.

Risso, C., Van Dien, S. J., Orloff, A., Lovley, D. R. & Coppi, M. V. (2008). Elucidation of an alternate isoleucine biosynthesis pathway in *Geobacter sulfurreducens*. *J Bacteriol* **190**, 2266–2274.

Schäfer, S., Paalme, T., Vilu, R. & Fuchs, G. (1989). ^{13}C-NMR study of acetate assimilation in *Thermoproteus neutrophilus*. *Eur J Biochem* **186**, 695–700.

Shastri, A. A. & Morgan, J. A. (2007). A transient isotopic labeling methodology for ^{13}C metabolic flux analysis of photoautotrophic microorganisms. *Phytochemistry* **68**, 2302–2312.

Stöckel, J., Welsh, E. A., Liberton, M., Kunnvakkam, R., Aurora, R. & Pakrasi, H. B. (2008). Global transcriptomic analysis of *Cyanothece* 51142 reveals robust diurnal oscillation of central metabolic processes. *Proc Natl Acad Sci U S A* **105**, 6156–6161.

Tang, Y. J., Chakraborty, R., Martin, H. G., Chu, J., Hazen, T. C. & Keasling, J. D. (2007). Flux analysis of central metabolic pathways in *Geobacter metallireducens* during reduction of soluble Fe(III)-NTA. *Appl Environ Microbiol* **73**, 3859–3864.

Tang, Y. J., Martin, H. G., Myers, S., Rodriguez, S., Baidoo, E. E. & Keasling, J. D. (2009a). Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. *Mass Spectrom Rev* **28**, 362–375.

Tang, Y. J., Sapra, R., Joyner, D., Hazen, T. C., Myers, S., Reichmuth, D., Blanch, H. & Keasling, J. D. (2009b). Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. *Biotechnol Bioeng* **102**, 1377–1386.

Tang, Y. J., Yi, S., Zhuang, W.-Q., Zinder, S. H., Keasling, J. D. & Alvarez-Cohen, L. (2009c). Investigation of carbon metabolism in "*Dehalococcoides ethenogenes*" strain 195 via isotopic and transcriptomic analyses. *J Bacteriol* **191**, 5224–5231.

Toepel, J., Welsh, E., Summerfield, T. C., Pakrasi, H. B. & Sherman, L. A. (2008). Differential transcriptional analysis of the cyanobacterium *Cyanothece* sp strain ATCC 51142 during light-dark and continuous-light growth. *J Bacteriol* **190**, 3904–3913.

Wahl, S. A., Dauner, M. & Wiechert, W. (2004). New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. *Biotechnol Bioeng* **85**, 259–268.

Welsh, E. A., Liberton, M., Stöckel, J., Loh, T., Elvitigala, T., Wang, C., Wollam, A., Fulton, R. S., Clifton, S. W. & other authors (2008). The genome of *Cyanothece* 51142, a unicellular diazotrophic cyanobacterium important in the marine nitrogen cycle. *Proc Natl Acad Sci U S A* **105**, 15094–15099.

Westfall, H. N., Charon, N. W. & Peterson, D. E. (1983). Multiple pathways for isoleucine biosynthesis in the spirochete *Leptospira*. *J Bacteriol* **154**, 846–853.

Xu, H., Zhang, Y. Z., Guo, X. K., Ren, S., Staempfli, A. A., Chiao, J., Jiang, W. & Zhao, G. (2004). Isoleucine biosynthesis in *Leptospira interrogans* serotype lai strain 56601 proceeds via a threonine-independent pathway. *J Bacteriol* **186**, 5400–5409.

Edited by: C. C. Zhang

Alternative isoleucine synthesis pathway in cyanobacterial species

Bing Wu, Baichen Zhang, Xueyang Feng, Jacob R. Rubens, Rick Huang, Leslie M. Hicks, Himadri B. Pakrasi and Yinjie J. Tang

Microbiology (2010), **156**, 596–602

SUPPLEMENTARY DATA

For the unlabelled citramalate, the parent ion *m/z* was 147 (Fig. S1-a). The results from the ^{13}C-labelled sample show that the 149/88 transition was dominant (Fig. S1-b), followed by the 149/86 transition. The ratio between 149/86 and 147/85 indicates that most of the citramalate was labelled with two ^{13}C carbons. According to the MS2 spectrum of unlabelled citramalate, the fragment ions 87 and 85 contained different carbon skeletons: C-1, C-2 and C-5 in the 87 ion and C-1, C-2, C-3, C-5 in the 85 ion (Fig. S1-b). The dominance of the 149/88 and 149/86 MRM transitions confirmed that the two ^{13}C-labelled carbons were distributed separately on the pyruvate moiety (C-1, C-2, C-5) and the remaining acetate moiety (C-3, C-4) of the citramalate molecule (please refer to Fig. 2 in the main text for the citramalate pathway), and that the C-3 carbon was excluded from being labelled by ^{13}C. Therefore, the only labelled carbon in the acetate portion of citramalate was in the C-4 position. On the other hand, in the ion spectrum of the unlabelled citramalate, the *m/z* 57 fragment ion contains C-2, C-3 and C-5 and the *m/z* 43 fragment ion contains C-2 and C-3 only. The dominance of the *m/z* 44 fragment ion in the ^{13}C-labelled sample readily indicates that the labelled carbon in the pyruvate part of the molecule was in the C-2 position. Therefore, it was concluded that citramalate was labelled dominantly by two ^{13}C carbons and they were possibly located at the C-2 and C-4 positions based on MRM experiments and MS/MS fragmentation of ^{13}C-labelled citramalate (parent ion *m/z* 149).

Fig. S1. Determination of citramalate labelling pattern. (a) LC-MS/MS spectra for unlabelled citramalate. (b) LC-MS/MS spectra for labelled citramalate and their mass fragments.

Fig. S2. Gel profiles of RT-PCR products (two conditions: 1, exponential growth culture without isoleucine; 2, addition of 10 mM isoleucine into exponential growth culture for 30 min). (a) Expression of cce_0248 and 16S rRNA. Lane 1: DNA ladder; lane 2: culture without Ile (amplify cce_0248 gene, the expected size of the RT-PCR band is 302 bp); lane 3: culture with 10 mM Ile (amplify cce_0248 gene); lane 4: negative control (without cDNA, amplify cce_0248 gene) ; lane 5: culture without Ile (amplify 16S rRNA gene, the expected size of the RT-PCR band is 243 bp); lane 6: culture with 10 mM Ile (amplify 16S rRNA gene); lane 7: negative control (without cDNA, amplify 16S rRNA gene). (b) Expression of *nifX* gene and 16S rRNA gene. Lane 1: culture without Ile (amplify 16S rRNA gene); lane 2: culture with 10 mM Ile (amplify 16S rRNA gene); lane 3: negative control (without cDNA, amplify 16S rRNA gene); lane 4: culture without Ile (amplify n*ifX* gene, the expected size of the RT-PCR band is 95 bp); lane 5: culture with 10 mM Ile (amplify *nifX*

gene); lane 6: negative control (without cDNA, amplify NifX gene); lane 7: DNA ladder. Note: no PCR products were amplified from the *nifX* gene in both culture conditions.

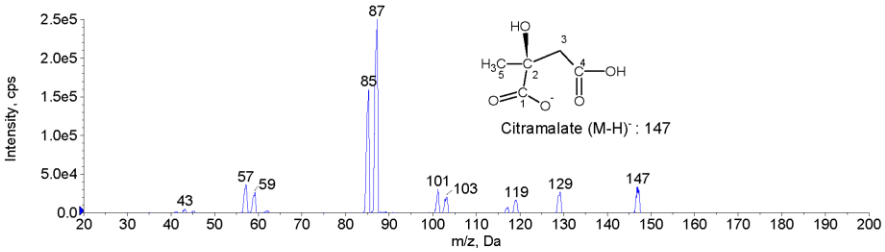


Figure S1-a. LC-MS/MS spectra for unlabeled citramalate.

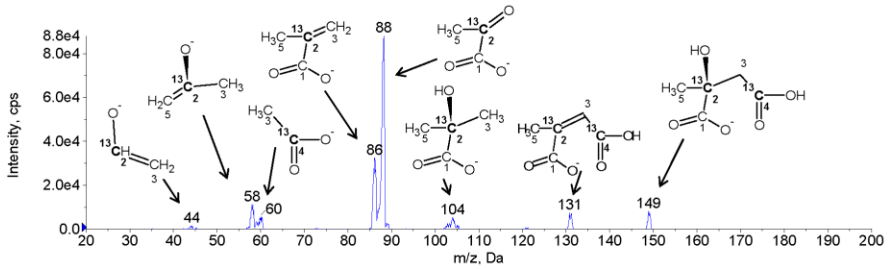


Figure S1-b. LC-MS/MS spectra for labeled citramalate and their mass fragments

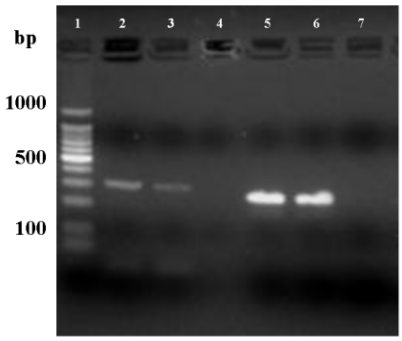


Figure S2-a. Expression of cce_0248 and 16S rRNA.

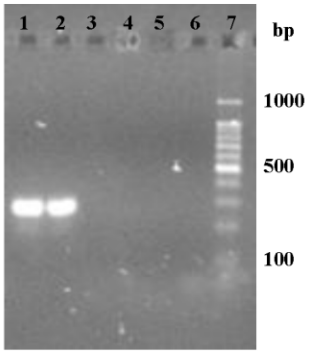


Figure S2-b: Expression of NifX gene and 16S rRNA.

Appendix 3

Carbon Flow of Heliobacteria Is Related More to Clostridia than to the Green Sulfur Bacteria

THE JOURNAL OF BIOLOGICAL CHEMISTRY VOL. 285, NO. 45, pp. 35104–35112, November 5, 2010

Carbon Flow of Heliobacteria Is Related More to *Clostridia* than to the Green Sulfur Bacteria*[S]

Received for publication, July 14, 2010, and in revised form, August 30, 2010 Published, JBC Papers in Press, August 31, 2010, DOI 10.1074/jbc.M110.163303

Kuo-Hsiang Tang[‡§1,2], **Xueyang Feng**[¶1], **Wei-Qin Zhuang**[‖], **Lisa Alvarez-Cohen**[‖], **Robert E. Blankenship**[‡§], **and Yinjie J. Tang**[¶3]

From the Departments of [‡]*Biology,* [§]*Chemistry, and* [¶]*Energy, Environment, and Chemical Engineering, Washington University, St. Louis, Missouri 63130 and the* [‖]*Department of Civil and Environmental Engineering, University of California, Berkeley, California 94720*

The recently discovered heliobacteria are the only Gram-positive photosynthetic bacteria that have been cultured. One of the unique features of heliobacteria is that they have properties of both the photosynthetic green sulfur bacteria (containing the type I reaction center) and *Clostridia* (forming heat-resistant endospores). Most of the previous studies of heliobacteria, which are strict anaerobes and have the simplest known photosynthetic apparatus, have focused on energy and electron transfer processes. It has been assumed that like green sulfur bacteria, the major carbon flow in heliobacteria is through the (incomplete) reductive (reverse) tricarboxylic acid cycle, whereas the lack of CO_2-enhanced growth has not been understood. Here, we report studies to fill the knowledge gap of heliobacterial carbon metabolism. We confirm that the CO_2-anaplerotic pathway is active during phototrophic growth and that isoleucine is mainly synthesized from the citramalate pathway. Furthermore, to our surprise, our results suggest that the oxidative (forward) TCA cycle is operative and more active than the previously reported reductive (reverse) tricarboxylic acid cycle. Both isotopomer analysis and activity assays suggest that citrate is produced by a putative (*Re*)-citrate synthase and then enters the oxidative (forward) TCA cycle. Moreover, in contrast to (*Si*)-citrate synthase, (*Re*)-citrate synthase produces a different isomer of 2-fluorocitrate that is not expected to inhibit the activity of aconitase.

Heliobacteria are a relatively newly discovered group of anaerobic photosynthetic bacteria. All of the cultured heliobacteria require organic carbon for anoxygenic growth, and several of the species can fix nitrogen (1, 2). Heliobacteria are the only cultured Gram-positive photosynthetic bacteria and are phylogenetically related to the bacterial phylum *Firmicutes*, such as the aerobic *Bacillus* and anaerobic *Clostridia* (1). Among 10 cultured heliobacteria (2), *Heliobacterium modesticaldum* is the only one that can grow at temperatures of >50 °C. Madigan and co-workers (1, 3) reported that pyruvate, lactate, acetate ($+HCO_3^-$), or yeast extract can support the phototrophic growth of *H. modesticaldum*, and our recent studies demonstrated that D-sugars can also support the growth of *H. modesticaldum* (4).

An unusual feature of heliobacteria is that they have properties of both green sulfur bacteria (containing the type I reaction center) and *Clostridia* (forming heat-resistant endospores) (1). Heliobacteria have the simplest photosynthetic apparatus among the photosynthetic organisms (5), and most of the research on them has been focused on understanding the photosynthetic machinery and energy transfer processes (1, 6–8). In contrast, our knowledge of phototrophic carbon metabolism by heliobacteria is still poorly developed.

Only two reported studies have experimentally probed the carbon metabolism of heliobacteria; one is our recent study with *H. modesticaldum* (4), and the other is the 1994 report by Kelly and co-workers (9) on *Heliobacterium* strain HY-3. Our studies highlighted the critical roles that pyruvate plays during phototrophic and chemotrophic growth of *H. modesticaldum*, reported on three new carbon sources for *H. modesticaldum*, and verified the genomic information. Kelly and co-workers (9) assayed activity for the enzymes in the TCA cycle and employed ^{13}C NMR to analyze ^{13}C-labeled protein-based amino acids using [2-^{13}C]pyruvate or [2-^{13}C]acetate. Furthermore, the genes *aclAB*, encoding ATP citrate lyase (ACL),[4] and *gltA*, encoding citrate synthase (CS), have not been annotated in the *H. modesticaldum* genome (5), and neither we (4) nor Kelly and co-workers (9) detected the enzymatic activities for ACL or CS. Together, previous studies suggested that the RTCA cycle is not complete in *H. modesticaldum* (Fig. 1*A*) (4, 5) and *Heliobacterium* strain HY-3 (9).

Note that ^{13}C NMR may be insufficient for generating complete labeling information of key metabolites, as ^{13}C NMR cannot directly detect labeled carbon on the carboxyl group of an amino acid if its α-carbon is not labeled (10). More importantly, essential questions remain unresolved for carbon metabolism of heliobacteria. (*a*) If the (incomplete) RTCA cycle is employed

* This work was supported by National Science Foundation Career Grant MCB0954016 (to Y. J. T.) and Exobiology Program of NASA Grant NNX08AP62G (to R. E. B.).

[S] The on-line version of this article (available at http://www.jbc.org) contains supplemental Tables S1–S5, Fig. S1, and additional references.

[1] Both authors contributed equally to this work.

[2] To whom correspondence may be addressed: Depts. of Biology and Chemistry, Campus Box 1137, Washington University, St. Louis, MO 63130. Tel.: 314-935-6343; Fax: 314-935-4432; E-mail: j.tang@wustl.edu.

[3] To whom correspondence may be addressed: Dept. of Energy, Environment, and Chemical Engineering, Campus Box 1180, Washington University, St. Louis, MO 63130. Tel.: 510-295-5651; Fax: 314-935-5464; E-mail: yinjie.tang@seas.wustl.edu.

[4] The abbreviations used are: ACL, ATP citrate lyase; CS, citrate synthase; FAc, fluoroacetate; 2-FC, 2-fluorocitrate; α-KG, α-ketoglutarate; OAA, oxaloacetate; OTCA cycle, oxidative (forward) TCA cycle; PEP, phosphoenolpyruvate; RTCA cycle, reductive (reverse) TCA cycle; BChl *g*, bacteriochlorophyll *g*; 8^1-OH-Chl a_F, 8^1-hydroxychlorophyll *a* with a farnesol tail.

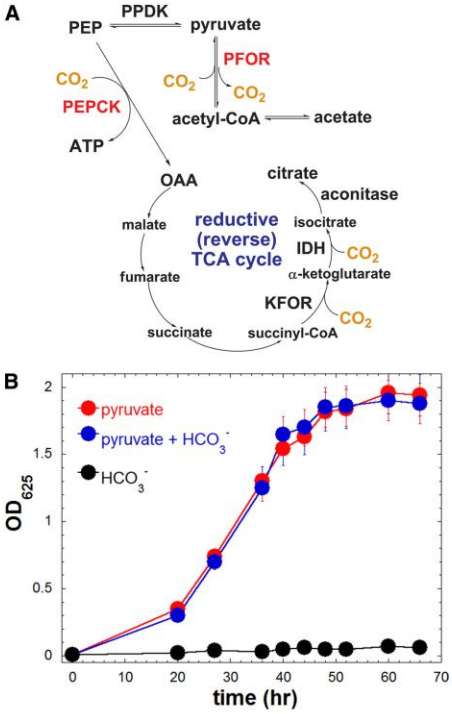


FIGURE 1. *A*, previously proposed carbon flow in *H. modesticaldum* from genomic information and enzymatic activity assays. *B*, phototrophic growth curve of *H. modesticaldum* grown on pyruvate, HCO_3^-, or pyruvate and HCO_3^- as the defined carbon source. Cell density was estimated by measuring the optical density (*OD*) at 625 nm (see the "Experimental Procedures"). *PFOR*, pyruvate:ferredoxin oxidoreductase; *PEPCK*, phosphoenolpyruvate carboxykinase; *KFOR*, α-KG:ferredoxin oxidoreductase; *IDH*, isocitrate dehydrogenase.

for biomass production in heliobacteria (Fig. 1*A*), then CO_2 is essential for phototrophic growth (a metabolic type similar to the green sulfur bacteria that employ an RTCA cycle) (11). However, no apparent CO_2-enhanced growth has been observed for pyruvate-grown heliobacterial cultures. (*b*) Some ^{13}C-labeled patterns in the study by Kelly and co-workers (9) suggested that the oxidative (forward) TCA (OTCA) cycle is operative. If this is the case, it is difficult to understand how the OTCA cycle could possibly be initiated without the encoding genes or enzymatic activities of ACL or CS.

This study is intended to apply both biochemical methods and ^{13}C-based metabolite analysis via mass spectrometry to address those essential questions by investigating how the TCA cycle and amino acid biosynthesis are performed by *H. modesticaldum*. Our results identify the missing link for the carbon metabolism of *H. modesticaldum* (and perhaps *Heliobacterium* strain HY-3) and suggest that the carbon flow of *H. modesticaldum* is more akin to *Clostridia* than to the green sulfur bacteria.

EXPERIMENTAL PROCEDURES

Materials—Chemicals and enzymes were obtained from Sigma. The ^{13}C-labeled sodium bicarbonate, [1-^{13}C]- and [3-^{13}C]pyruvate were obtained from Cambridge Isotope Laboratories, Inc. The DNA oligomers were obtained from Integrated DNA Technology without further purification.

Cell Cultures—The cell cultures employed in this study are listed below. Cell density was estimated by measuring the optical density (OD) of a suspension of cells at 625 nm (OD_{625}) for the cell growth of both *H. modesticaldum* and *Roseobacter denitrificans* because absorbance of photosynthetic pigments is minimal around 625 nm (4), and OD_{600} was measured to estimate the growth of *Desulfovibrio vulgaris* Hildenborough (DvH), in which no photosynthetic pigment was present. Autoclaved subcultures were used as negative controls, and all of the experiments were performed in triplicate.

H. modesticaldum—The *H. modesticaldum* strain Ice1^T culture was provided by Dr. Michael T. Madigan (Southern Illinois University, Carbondale) and was grown phototrophically in a minimal medium (1, 4) with pyruvate (20 mM) and (±)HCO_3^- (20 mM) as the carbon sources. The fresh medium was inoculated with 1–2% cell culture in the late exponential growth phase. Physiological studies with fluoroacetate (FAc) for the pyruvate-grown cultures were performed with 20 mM pyruvate and 20 mM FAc included in the growth medium. The cultures were grown inside the anaerobic chamber (Coy) in low intensity light (100 ± 10 μmol/m^2/s) at 46–48 °C.

R. denitrificans—The *R. denitrificans* OCh114 culture, obtained from Dr. J. T. Beatty (University of British Columbia, Vancouver, Canada), was grown aerobically in a defined medium with pyruvate (20 mM), FAc (20 mM), or pyruvate (20 mM) + FAc (20 mM) as the sole carbon sources. The defined growth medium was prepared as reported (12). The cultures were grown at 28 °C (20–30%-filled Erlenmeyer flask) with shaking at 150 rpm in the dark.

D. vulgaris Hildenborough (DvH)—DvH was kindly provided by Dr. Terry C. Hazen (Lawrence Berkeley National Laboratory, Berkeley, CA). DvH was grown in a defined growth medium containing (per liter) NaCl (1.0 g), $MgCl_2{\cdot}6H_2O$ (0.5 g), KH_2PO_4 (0.2 g), NH_4Cl (0.3 g), KCl (0.3 g), $CaCl_2{\cdot}2H_2O$ (0.015 g), $MgSO_4{\cdot}7H_2O$ (0.2 g), a trace element solution (1 ml) (13), a Na_2SeO_3/Na_2WO_4 solution (1 ml) (14), resazurin (10 mg), lactate (5 mM), Na_2SO_4 (5 mM), and with a N_2/CO_2 (80:20, v/v) headspace. After autoclaving and cooling to room temperature, the medium was supplied with a vitamin mixture solution (0.5 ml) (15). DvH was subcultured into fresh medium in triplicate with 1% (v/v) inoculate, and cell cultures were grown at 30 °C anaerobically. FAc amended cultures were prepared in the same content except FAc (1.5 mM) was added before inoculation.

RNA Extraction and Quantitative Real Time PCR—The transcriptomic profiling data were generated using the approaches described elsewhere (16). The primer sequences for QRT-PCR are listed in supplemental Table S1.

Isotopomer Analysis via GC-MS—The isotopomer analysis was performed using the methods reported previously (17) and is described in the supplemental material.

Activity Assays for Citrate Synthase—Enzymatic activities were performed in triplicate with cell-free extracts that were prepared using the procedure as described earlier (4). Protein concentration in cell extracts was determined by the Bradford assay (18) using bovine serum albumin as the standard. The reaction turnover for the reaction catalyzed by CS was followed by the formation of 5,5′-dithiobis-2-nitrobenzoic acid or Ellman's reagent-modified CoA using the method reported previously (9, 19) with minor modification. To minimize oxygen in the assay mixture for measuring CS activity, all of the reagents (5,5′-dithiobis-2-nitrobenzoic acid, metal ions), substrates (acetyl-CoA, oxaloacetate), and buffer were prepared with oxygen-free water in an anaerobic chamber (Coy), and the assay components were mixed in a tightly sealed cuvette inside the anaerobic chamber prior to removal for the spectral measurements. The reaction was initiated by the addition of oxaloacetate (OAA) (0.5 mM), and the increase at A_{412} was followed to monitor the activity. The amount of CoA production was estimated using a standard calibration curve with β-mercaptoethanol. The control experiment was performed with the same procedure with water rather than OAA added to initiate the reaction.

RESULTS

Carbon Metabolism in H. modesticaldum

In this study, we probed the central carbon metabolic pathway with the following approaches: (i) physiological studies with fluoroacetate; (ii) isotopomer data; (iii) mass spectrum of photosynthetic pigments; and (iv) transcriptomic profiles. Pyruvate is the best known organic carbon source for supporting the phototrophic growth of *H. modesticaldum* (3, 4) and several other heliobacteria (1), and it was used for probing the carbon metabolism of *H. modesticaldum*.

Physiological Studies with Fluoroacetate—FAc has been reported as a metabolic toxin. The toxicity of FAc is generally recognized to arise from the fact that the carbon flow in the OTCA cycle is blocked through the inhibition of aconitase by (−)-*erythro*-(2*R*,3*R*)-2-F-citrate (2-FC), which is synthesized from F-acetyl-CoA and OAA by CS (20). Consistent with this hypothesis, an aerobic anoxygenic photoheterotrophic bacterium *R. denitrificans*, which has been known to have an active OTCA cycle (12), was notably inhibited by FAc (Fig. 2*A*). In contrast, the growth of *H. modesticaldum* on pyruvate was similar with FAc *versus* without FAc (Fig. 2*B*). It is known that FAc, an acetate analog, can be taken up by *H. modesticaldum* and converted into F-acetyl-CoA, because the *acsA* gene, encoding acetyl-CoA synthetase (ACS), has been annotated and the enzymatic activity of ACS has been reported (4). Furthermore, acetate ($+HCO_3^-$) is also known to support the phototrophic growth of *H. modesticaldum* (1, 3, 4). Consistent with the physiological studies, the transcript levels of genes for carbon metabolism are similar (within ~2 $\Delta\Delta C_T$) for cultures grown on pyruvate (20 mM) and with or without the addition of FAc (20 mM) (Table 1). Together, our studies indicate little effect of

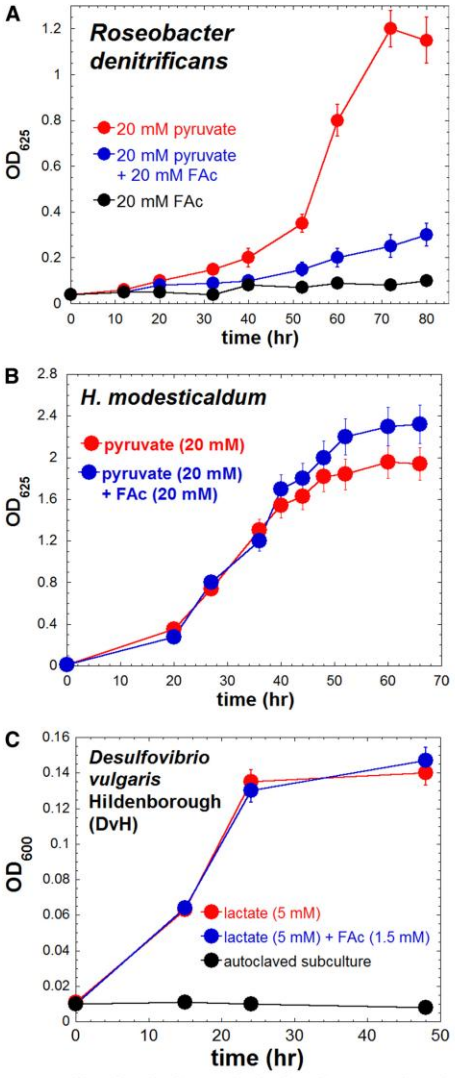


FIGURE 2. **Effect of FAc for the growth of *R. denitrificans*, *H. modesticaldum*, and *D. vulgaris* Hildenborough (*DvH*).** *A*, growth curve of *R. denitrificans* grown on pyruvate, FAc, or pyruvate and FAc as the carbon source; *B*, growth curve of *H. modesticaldum* during photoheterotrophic growth on pyruvate with or without FAc; *C*, growth curve of DvH grown on lactate with or without FAc.

TABLE 1
The transcript level of the genes for carbon metabolism of *H. modesticaldum* during photoheterotrophic growth on pyruvate with or without FAc
The genes encoding the enzymes in the (R)TCA cycle are highlighted in boldface italic.

Gene	ΔC_T[a] (pyruvate)	ΔC_T[a] (pyruvate + FAc)	$\Delta\Delta C_T$ (ΔC_T in pyruvate − ΔC_T in pyruvate + FAc)	relative expression level (pyruvate vs. pyruvate + FAc) = $2^{\Delta\Delta C_T}$
pykA (HM1_0076, pyruvate kinase)	11.0 ± 0.2	12.7 ± 0.2	1.7 ± 0.4	3.2 ± 1.3
acn (HM1_0105, aconitase)	10.7 ± 0.1	12.6 ± 0.1	1.8 ± 0.2	3.5 ± 1.1
fdxR (HM1_0289, Fd-NADP$^+$ reductase, FNR)	12.2 ± 0.0	14.2 ± 0.1	2.0 ± 0.1	4.0 ± 1.0
porA (HM1_0807, pyruvate, Fd oxidoreductase, PFOR)	9.8 ± 0.2	11.7 ± 0.1	1.8 ± 0.3	3.5 ± 1.2
acsA (HM1_0951, acetyl-CoA synthetase)	12.1 ± 0.0	14.3 ± 0.2	2.1 ± 0.2	4.3 ± 1.1
icd (HM1_1471, isocitrate dehydrogenase)	10.7 ± 0.2	12.4 ± 0.0	1.6 ± 0.2	3.0 ± 1.1
mdh (HM1_1472, malate dehydrogenase	10.9 ± 0.0	12.4 ± 0.0	1.5 ± 0.0	2.8 ± 0.0
ackA (HM1_2157, acetate kinase)	11.6 ± 0.1	13.6 ± 0.0	2.0 ± 0.1	4.0 ± 1.0
ppdK (HM1_2461, pyruvate phosphate dikinase)	11.3 ± 0.1	13.2 ± 0.1	1.9 ± 0.2	3.7 ± 1.1
korC (HM1_2762, KFOR, γ subunit)	9.9 ± 0.0	10.8 ± 0.2	0.9 ± 0.2	1.9 ± 1.1
oorB (HM1_2763, KFOR, β subunit)	9.2 ± 0.0	10.9 ± 0.1	1.7 ± 0.1	3.2 ± 1.0
korA (HM1_2766, KFOR, α subunit)	8.9 ± 0.1	10.7 ± 0.0	1.8 ± 0.1	3.5 ± 1.0
korD (HM1_2767, KFOR, δ subunit)	8.9 ± 0.0	11.0 ± 0.1	2.1 ± 0.1	4.3 ± 1.0
nifV (HM1_0858, homocitrate synthase)	15.0 ± 0.1	15.4 ± 0.2	0.4 ± 0.3	1.3 ± 1.2
aksA (HM1_2993, homocitrate synthase)	10.5 ± 0.2	12.0 ± 0.2	1.5 ± 0.4	2.8 ± 1.3
pckA (HM1_2773, PEP carboxykinase)	11.9 ± 0.0	13.5 ± 0.1	1.6 ± 0.1	3.0 ± 1.0

[a] $\Delta C_T = C_T$ (the threshold cycle) of the target gene $- C_T$ of the 16 S rRNA gene.

FAc on the growth of *H. modesticaldum*, implying that (−)-*erythro*-2-FC cannot be synthesized in *H. modesticaldum*, in agreement with the findings that no genes encoding CS and ACL have been annotated in the genome (5), and no enzymatic activities of CS or ACL have been detected for *H. modesticaldum* (4).

Moreover, when the anaerobic sulfate-reducing bacterium *D. vulgaris* Hildenborough (DvH) was grown with a 1:1 molar ratio of [2-^{13}C]acetate and nonlabeled lactate, significant amounts of labeled carbon were detected in the biomass (supplemental Table S2), indicating that DvH can utilize acetate for producing biomass. Furthermore, isotopomer analysis using [2-^{13}C]acetate (supplemental Table S2) and a previous DvH flux analysis (21) indicate that acetate ↔ acetyl-CoA is reversible in DvH. No FAc inhibition was detected during the growth of DvH (Fig. 2*C*).

Isotopomer Analysis by GC-MS—Previous studies have established that the cell growth of *H. modesticaldum* is best supported by pyruvate (1, 3, 4). We used ^{13}C-labeled pyruvate and characterized the protein-based amino acids for probing the central carbon metabolic pathways. *H. modesticaldum* is recognized as a photoheterotrophic bacterium (3, 4), and CO_2 neither supports nor enhances its phototrophic growth (Fig. 1*B*). When *H. modesticaldum* was grown on [1-^{13}C]pyruvate, aspartate was mainly double-labeled, and glutamate was primarily single-labeled (Table 2). In the (incomplete) RTCA cycle (Fig. 1), α-KG, the precursor of glutamate, is synthesized from OAA, the precursor of aspartate, which results in the incorporation of all the carbons from aspartate into the carbon backbones of glutamate. Also, assuming that some of the CO_2 molecules assimilated through the RTCA cycle are $^{13}CO_2$ generated from the reaction [1-^{13}C]pyruvate + CoA → acetyl-CoA + $^{13}CO_2$ (Fig. 1), ^{13}C-labeled content is expected to be higher for glutamate than for aspartate. Consequently, the multiply labeled instead of the singly labeled glutamate would eventually become dominant through the (incomplete) RTCA cycle. However, in this study, the observed lower ^{13}C-labeled content of glutamate compared with aspartate with [1-^{13}C]pyruvate-grown cells cannot have been generated by cells utilizing solely the (incomplete) RTCA cycle. Furthermore, the carbon flux of *H. modesticaldum* cannot primarily go through the (incomplete) RTCA cycle but must instead mainly transform via the OTCA cycle. Moreover, aspartate was dominantly single-labeled (M1 >0.8), and glutamate was mostly double-labeled (M2 >0.75) for cells using [3-^{13}C]pyruvate (supplemental Table S3), in agreement with significant carbon flow through the OTCA cycle.

Mass Spectrum of Photosynthetic Pigments—We further analyzed the mass spectrum of BChls we reported earlier using MALDI-TOF (matrix-assisted laser desorption ionization time-of-flight) mass spectrometry (4), in which ^{13}C-labeled photosynthetic pigments (bacteriochlorophyll *g* (BChl *g*) and 8^1-hydroxychlorophyll *a* with a farnesol tail (8^1-OH-Chl a_F)) were detected using [3-^{13}C]pyruvate. Both BChl *g* and 8^1-OH-Chl a_F were synthesized from eight molecules of glutamate and eight molecules of acetyl-CoA (for generating the farnesyl tail with a 15-carbon unit at C17^3 position), and thus ^{13}C-labeled BChl *g* and 8^1-OH-Chl a_F are expected using [3-^{13}C]pyruvate. The *m*/*e* for BChl *g* and 8^1-OH-Chl a_F is 796.7 and 812.7, respectively, with unlabeled pyruvate, and for BChl *g* and 8^1-OH-Chl a_F is 816.8 and 832.8, respectively, with [3-^{13}C]pyruvate. The number of labeled carbons for the labeled BChl *g* or 8^1-OH-Chl a_F was estimated to be 21–22 using the program IsoPro 3.1. Using [3-^{13}C]pyruvate as the carbon source, glutamate is expected to be double-labeled through the OTCA cycle because citrate is double-labeled via condensation of [3-^{13}C]OAA and [2-^{13}C]acetyl-CoA, where glutamate is single-labeled through the RTCA cycle (from [3-^{13}C]OAA). Thus, 16 carbons are labeled through the RTCA cycle (eight single-labeled glutamates and eight single-labeled acetyl-CoAs) and 24-labeled carbons through the OTCA cycle (eight double-labeled glutamates and eight single-labeled acetyl-CoAs) using [3-^{13}C]pyruvate. Thus, observations of higher ^{13}C contents on BChl *g* and 8^1-OH-Chl a_F in the mass spectrum than expected from the RTCA cycle would suggest that the OTCA cycle contributes to the formation of glutamate and α-KG. Consistent with this hypothesis, glutamate was found to be mostly double-

TABLE 2
Isotopomer labeling patterns of protein-based amino acids in the cultures grown on [1-^{13}C]pyruvate

Amino acid	Precursor	Ion[a]	M-57[b]	M-159[c]	f302[d]	Proposed ^{13}C-enriched positions (shown as asterisk)
Ala	Pyruvate	M0	0.20	0.85	0.20	C-C-*COOH
		M1	0.77	0.09	0.79	
		M2	0.03	0.06	0.01	
Gly	Serine	M0	0.24	0.98	0.12	C-*COOH
		M1	0.75	0.02	0.48	
		M2	0.01		0.40	
Val	Pyruvate	M0	0.20	0.89	0.28	C-C-C-C-*COOH
		M1	0.77	0.09	0.72	
		M2	0.03	0.01	0.00	
Leu	Pyruvate	M0	0.60	0.83	0.63	Nonlabeled
	Acetyl-CoA	M1	0.36	0.14	0.36	
		M2	0.03	0.02	0.01	
		M3	0.00	0.00		
Ile	Pyruvate	M0	0.66	0.83	0.69	Nonlabeled
	Threonine	M1	0.25	0.16	0.25	
		M2	0.08	0.01	0.06	
		M3	0.01	0.00		
Met	Aspartate	M0	0.09	0.22	0.17	C-S-*C-C-C-*COOH
	Methyl-THF[e]	M1	0.24	0.58	0.51	
		M2	0.51	0.20	0.32	
		M3	0.15	0.01		
Ser	3-Phosphoglycerate	M0	0.19	0.94	0.21	C-C-*COOH
		M1	0.77	0.06	0.78	
		M2	0.04	0.00	0.01	
Thr	Aspartate	M0	0.11	0.26	0.11	*C-C-C-*COOH
		M1	0.29	0.72	0.46	
		M2	0.59	0.02	0.43	
Phe	PEP	M0	0.08	0.13	0.22	C-C-*C-*C-C-C-C-C-*COOH
	Erythrose 4-phosphate	M1	0.11	0.29	0.77	
		M2	0.31	0.53	0.01	
		M3	0.46	0.04		
		M4	0.04	0.01		
Asp	OAA	M0	0.11	0.27	0.27	*COOH-C-C-*COOH
		M1	0.31	0.71	0.73	
		M2	0.56	0.01	0.01	
Glu	α-Ketoglutarate	M0	0.26	0.28	0.82	C-C-C-C-*COOH
		M1	0.64	0.68	0.18	
		M2	0.09	0.04	0.00	
		M3	0.01	0.00		
His	Ribose 5-phosphate	M0	0.13	0.13	0.92	N-C-N-C-*C-*C-C-COOH
		M1	0.41	0.41	0.06	
		M2	0.32	0.34	0.02	
		M3	0.10	0.07		
Lys	Aspartate	M0	0.14	0.28	0.92	C-C-*C-C-C-*COOH
	Pyruvate	M1	0.26	0.68	0.08	
		M2	0.58	0.03	0.00	
		M3	0.02	0.00		
		M4	0.00	0.00		

[a] M0, M1, M2, M3, and M4 indicate mass fraction of the unlabeled, single-labeled, double-labeled, triple-labeled, and quadruple-labeled amino acid.
[b] [M-57]$^+$ indicates un-fragmented amino acid detected by GC-MS.
[c] [M-159]$^+$ indicates an amino acid minus the α−carboxyl group detected by GC-MS.
[d] f302 is the fragment of the first two carbons in a derivatized amino acid (17).
[e] THF is tetrahydrofolate.

labeled (M2 >75%) using [3-^{13}C]pyruvate (supplemental Table S3).

Transcriptomic Profiles—All of the genes encoding the enzymes in the RTCA cycle, except for ACL, have been annotated in the *H. modesticaldum* genome (5). Table 1 shows that genes in the RTCA cycle (highlighted in **boldface italic**) are expressed and that the transcript level for the genes responsible for carbon metabolism is similar (at most a 4-fold difference) in the pyruvate-grown cultures with and without FAc.

Taking all of the experimental evidence together, our studies indicate that in addition to the incomplete RTCA cycle, the OTCA cycle is also employed by *H. modesticaldum*. Like the RTCA cycle, the OTCA cycle is not complete but mainly contributes to the formation of α-KG, as indicated by isotopomer analysis and mass spectrometry of photosynthetic pigments. The active OTCA cycle is consistent with the observed lack of CO_2-enhanced pyruvate growth of *H. modesticaldum* (Fig. 1*B*), because CO_2 is produced through the OTCA cycle.

Citrate Synthase Activity Detected Anaerobically

Table 2 shows that over 50% aspartate and 60% glutamate were labeled in the β-carboxyl group with [1-^{13}C]pyruvate as the carbon source, implying that citrate formation is possibly catalyzed by (*Re*)-citrate synthase ((*Re*)-CS) (Fig. 3). Both (*Re*)-CS and normal CS (*i.e.* (*Si*)-CS) catalyze the formation of citrate through aldol condensation of OAA and acetyl-CoA, whereas the acetyl-CoA moiety is added to the "pro-*R*" and "pro-*S*" arm of citrate through catalysis of (*Re*)-CS and (*Si*)-CS, respectively (Fig. 3).

We have previously showed that no CoA production was detected with acetyl-CoA and OAA added in the cell extracts of *H. modesticaldum* under aerobic conditions and without diva-

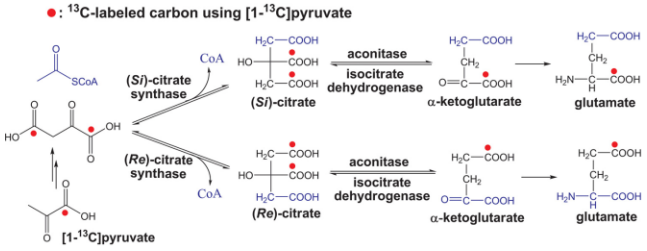


FIGURE 3. **Reactions catalyzed by the (*Re*)- *versus* (*Si*)-citrate synthase, aconitase, and isocitrate dehydrogenase.** ^{13}C labeling distributions in citrate, α-ketoglutarate, and glutamate using [1-^{13}C]pyruvate are shown.

lent metal ion supplied, confirming that (*Si*)-CS is not produced by *H. modesticaldum* (4). Anaerobic conditions have not been reported to be required for the activity of (*Si*)-CS. To test if citrate can be produced by *H. modesticaldum* via the catalysis of (*Re*)-CS, rigorous efforts were made to minimize the oxygen content in the assay mixtures (described under "Experimental Procedures"), because oxygen-sensitive and divalent metal ion (Mn^{2+}, Mg^{2+}, or Co^{2+}) dependences were reported for the activity measurements of (*Re*)-CS (19). The increase of A_{412} for 5,5′-dithiobis-2-nitrobenzoic acid-modified CoA can be detected with Mn^{2+}, OAA, and acetyl-CoA under anaerobic conditions, suggesting the presence of the CS activity in *H. modesticaldum*. The catalytic activity of the novel CS, likely (*Re*)-CS, was estimated to be 50 ± 20 nmol/min/mg protein.

Anaplerotic-CO_2 Fixation Pathways

The isotopomer labeling experiment with ^{13}C-labeled HCO_3^- and unlabeled pyruvate as the carbon source showed that over 50% alanine was labeled (supplemental Table S4) and that all the labeled alanine was labeled at the carboxyl group. The labeling pattern suggests that the reaction catalyzed by pyruvate:ferredoxin oxidoreductase (acetyl-CoA + CO_2 + $2Fd_{red}$ + $2H^+$ ↔ pyruvate + CoA + $2Fd_{ox}$) is freely reversible and very active so that the labeled bicarbonate is incorporated into pyruvate. The isotopomer analysis is in agreement with our recent physiological studies that pyruvate:ferredoxin oxidoreductase plays a central role in carbon metabolism of *H. modesticaldum* (4). Moreover, when using [1-^{13}C]pyruvate as the carbon source, aspartate can be synthesized either from the CO_2-anaplerotic pathways with pyruvate and/or phosphoenolpyruvate (PEP) and $^{13}CO_2$ (from decarboxylation of [1-^{13}C]pyruvate), leading to double-labeled aspartate, or through the OTCA cycle, in which aspartate is not expected to be double labeled. The predominance of double-labeled aspartate indicates the high carbon flow via the CO_2-anaplerotic pathway and low flux from the OTCA cycle, consistent with observed activity of PEP carboxykinase illustrated in our recent studies (4), and confirming that the OTCA cycle is not complete as genes encoding the enzymes specific for the OTCA cycle have not been annotated in the genome (5).

Amino Acid Biosynthesis

The isotopomer analysis for biosynthesis of several amino acids is illustrated as follows.

Alternative Isoleucine Biosynthesis Pathway—Isoleucine is typically synthesized through the threonine pathway (Fig. 4). Using [1-^{13}C]pyruvate as the carbon source, threonine was mainly double-labeled (M2 >58%), so isoleucine would be expected to be double-labeled via the threonine pathway. Table 2 shows that isoleucine was largely nonlabeled (M0 >60%) with [1-^{13}C]pyruvate, suggesting that isoleucine was mostly synthesized through the citramalate pathway, although all of the genes in the threonine pathway have been annotated. In the citramalate pathway, citramalate synthase (CimA) catalyzes the formation of D-*erythro*-3-methylmalate (*i.e.* citramalate) through condensation of pyruvate and acetyl-CoA (Fig. 4). Two gene loci, HM1_1519 and HM1_1515, encoding putative CimA in the citramalate pathway have been annotated in the genome. The amino acid sequence encoding by gene locus HM1_1519 shows >50% identity to the recently reported CimA (Teth514_1204) in *Thermoanaerobacter* sp. X514 (22), which is a close relative to the genus *Clostridium*. The supplemental Table S5 lists the bacteria with the citramalate pathway identified.

Normal Pathways for Alanine, Serine, Phenylalanine, and Lysine Biosynthesis—The isotopomer pattern suggested alanine and serine are synthesized from pyruvate and PEP, respectively. Also, using [3-^{13}C]pyruvate as the carbon source, the labeling pattern in supplemental Table S3 indicates that phenylalanine is synthesized from the common biosynthetic pathway with erythrose 4-phosphate and PEP as precursors in agreement with the genomic information (5). Furthermore, Pickett *et al.* (9) also reported the same results regarding pathways of alanine, serine, and phenylalanine biosynthesis for *Heliobacterium* strain HY-3.

The labeling patterns of lysine in the isotopomer analysis suggest that lysine is synthesized through the common diaminopimelate pathway with pyruvate and aspartate as the precursors, rather than through the α-amino adipate pathway (supplemental Fig. S1), in which 2-ketoglutarate and acetyl-CoA are condensed and converted to lysine. This conclusion is consistent with the fact that all genes in the α-aminoadipate pathway are missing, except for the genes encoding putative homocitrate synthase (*nifV* (HM1_0858) and *aksA* (HM1_2993)). This enzyme catalyzes the formation of homocitrate ((*R*)-2-hydroxybutane-1,2,4-tricarboxylate) by condensing α-KG and acetyl-CoA. Note that the *askA* gene is expressed with higher transcript level (Table 1). The possible function of the homocitrate synthase for carbon metabolism will be discussed later.

DISCUSSION

Growth of H. modesticaldum with FAc

Fig. 2 shows that there is no detectable difference in the growth of *H. modesticaldum* with or without FAc, suggesting that (−)-*erythro*-2-FC cannot be synthesized by CS or/and ACL

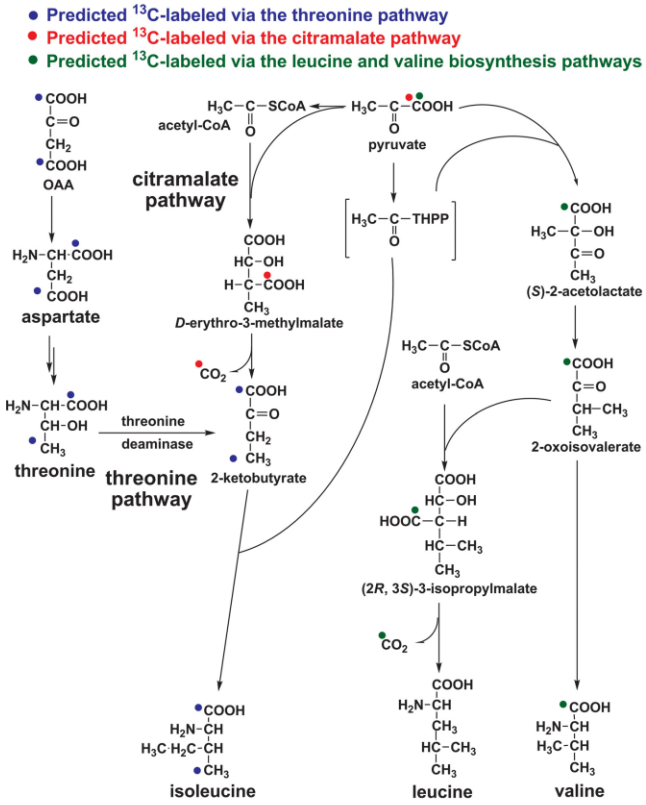


FIGURE 4. **Proposed biosynthesis pathways for isoleucine, including the citramalate and threonine pathways, leucine, and valine.** Predicted ^{13}C labeling distributions using [1-^{13}C]pyruvate are shown.

in *H. modesticaldum*. However, the lack of genes encoding ACL and CS along with the lack of the enzymatic activities is inconsistent with our presented isotopomer data, which suggest that at least a partial OTCA cycle is active and an oxygen-sensitive novel CS is produced by *H. modesticaldum* to initiate the OTCA cycle. If a novel CS is produced by *H. modesticaldum*, then the lack of FAc inhibition during the growth of *H. modesticaldum* needs to be explained. Two working hypotheses can be considered.

High Flux from Pyruvate to Acetyl-CoA to Compete with F-acetyl-CoA—Similar activity of F-acetyl-CoA *versus* acetyl-CoA for normal CS has been reported previously (23). Assuming that (−)-*erythro*-2-FC is synthesized by a putative novel CS, acetyl-CoA from pyruvate could compete with F-acetyl-CoA from FAc for the interactions of the putative CS. A high flux from pyruvate to acetyl-CoA is suggested from our recent studies showing that conversion of pyruvate to acetyl-CoA, catalyzed by pyruvate:ferredoxin oxidoreductase, is very active in pyruvate-grown *H. modesticaldum* (4) and the isotopomer data presented in this report. However, Fig. 2*A* shows that the pyruvate-grown *R. denitrificans*, which is also known to exhibit high flux from pyruvate to acetyl-CoA (12), was notably reduced by the presence of FAc, suggesting that other factors must contribute to the lack of FAc-inhibition on the phototrophic growth of *H. modesticaldum*.

Different Stereoisomer of 2-FC Is Synthesized—Alternatively, it is possible that the putative CS catalyzes the formation of a 2-FC isomer other than (−)-*erythro*-2-FC. Note that only one of four possible 2-FC isomers, (−)-*erythro*-2-FC, catalyzed by (*Si*)-CS, is a potent inhibitor of aconitase (20). Thus, if the putative CS does not catalyze the formation of (−)-*erythro*-2-FC but a different 2-FC isomer, then the toxicity of FAc is not expected to be observed for *H. modesticaldum*. This hypothesis is further elaborated below.

Different Isomers of 2-FC Produced by (Re)- Versus (Si)-CS

Although there is no difference in the stereochemistry of citrate from condensation of acetyl-CoA and OAA catalyzed by (*Si*)-CS *versus* (*Re*)-CS (Fig. 3), different isomers of 2-F-citrate (2-FC) are produced from condensation of F-acetyl-CoA and OAA catalyzed by (*Si*)-CS *versus* (*Re*)-CS (Fig. 5). Lauble *et al.* (20) proposed that interactions of (−)-*erythro*-2-FC with aconitase produce an intermediate that inhibits aconitase, whereas interactions of (+)-*erythro*-2-FC with aconitase leads to formation of α-KG, instead of inhibiting aconitase.

To test this hypothesis, we performed physiological studies with the sulfate-reducing bacterium *D. vulgaris* Hildenborough (DvH) using FAc. Acetate can serve as a carbon source for DvH in the presence of reducing power, which can be generated during lactate-supported growth (supplemental Table S2). Thus, FAc, like acetate, can be assimilated by DvH using the reducing power generated by lactate oxidation. Like *H. modesticaldum*, no genes encoding ACL and (*Si*)-CS have been annotated in the DvH genome, whereas the activity of (*Re*)-CS and contribution of CS to the carbon flow of the OTCA cycle have been identified for DvH (24). Fig. 2*C* shows no inhibition for the

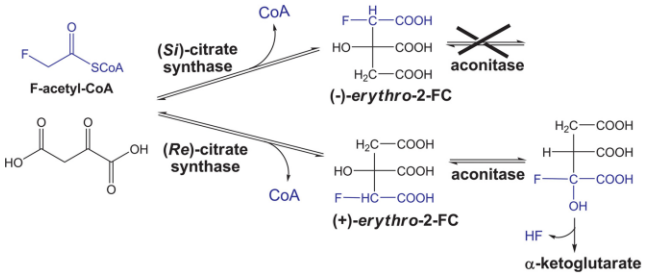


FIGURE 5. **Reaction of F-acetyl-CoA and OAA catalyzed by (*Re*)- *versus* (*Si*)-citrate synthase, and interactions of (−)-*erythro*-2-FC *versus* (+)-*erythro*-2-FC with aconitase.**

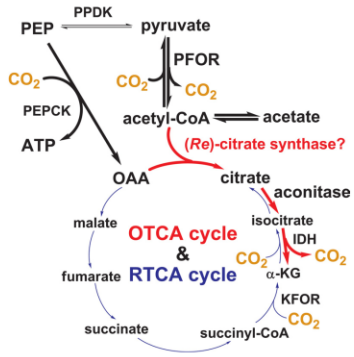


FIGURE 6. **New view for carbon flow in *H. modesticaldum*.** The proposed carbon flow is through the OTCA cycle (*red*) and the RTCA cycle (*blue*) with a stronger flux through the OTCA cycle (acetyl-CoA → α-KG, shown in *bold*). The proposed role of the putative (*Re*)-citrate synthase is shown.

growth of DvH with FAc, consistent with the hypothesis by Lauble *et al.* (20). Consequently, we believe that the lack of FAc inhibition on the growth of *H. modesticaldum* is in agreement with generation of (+)-*erythro*-2-FC, instead of (−)-*erythro*-2-FC, by (*Re*)-CS.

Homocitrate Synthase May Function as (Re)-CS in H. modesticaldum

(*Re*)-CS has been identified in *Clostridium kluyveri* (19), AND THE ACTIVITY OF (*Re*)-CS has been reported in *Thermoanaerobacter* sp. X514 (22), DvH (24), *Dehalococcoides ethenogenes* 195 (26), *Ignicoccus hospitalis* (27), and several anaerobic bacteria. Because different stereoisomers of citrates are generated through the catalysis of (*Re*)- *versus* (*Si*)-CS, it is not surprising that these two types of CS are rather phylogenetically distinct (19, 28).

As mentioned under "Results," the product turnover catalyzed by an oxygen-sensitive CS, possibly a putative (*Re*)-CS, has been detected with Mn^{2+} ions supplied while minimizing oxygen content in the assay solution. Although the gene encoding (*Re*)-CS has not been annotated in the *H. modesticaldum* genome, homocitrate synthase has been suggested to be phylogenetically related to (*Re*)-CS (19). It is possible that homocitrate synthase catalyzes not only the formation of homocitrate from acetyl-CoA and α-KG but also the formation of citrate from acetyl-CoA and OAA (*i.e.* (*Re*)-CS). Two genes in the *H. modesticaldum* genome possibly encode homocitrate synthase, *nifV* (HM1_0858) and *aksA* (HM1_2993). A BLAST search shows that proteins encoded by both genes, particularly by *askA*, share high identity with the reported putative homocitrate synthase/(*Re*)-CS, in agreement with the observed higher transcript level (*i.e.* lower ΔC_T) of *askA* compared with that of *nifV* (Table 1). Given the data presented in this study, we propose that the putative homocitrate synthase, an enzyme that has been reported to function as (*Re*)-CS in several bacteria, is likely responsible for synthesizing citrate in *H. modesticaldum*.

New View for Carbon Flow of H. modesticaldum

Fig. 6 represents a new model for the carbon flow of *H. modesticaldum* that reflects improved understanding resulting from isotopomer and activity assays; the major carbon flux of *H. modesticaldum* is through the OTCA cycle. (*Re*)-citrate synthase, identified in several *Clostridia* and other anaerobic bacteria, is likely employed by *H. modesticaldum* to produce citrate for entering the OTCA cycle. The finding of the OTCA cycle herein fills the knowledge gap for the carbon flow of *H. modesticaldum*. It is intriguing to learn that the major carbon flux of *H. modesticaldum* is switched from the RTCA cycle to the OTCA cycle when the only gene (ATP citrate lyase) missing in the RTCA cycle is replaced by a gene with novel function ((*Re*)-citrate synthase) in the OTCA cycle. Together, our studies suggest that the carbon flow of *H. modesticaldum* (and perhaps *Heliobacterium* strain HY-3) is more akin to *Clostridia* than to the green sulfur bacteria, which employ the RTCA cycle for CO_2 assimilation and biomass production (25).

REFERENCES

1. Madigan, M. T. (2006) *Prokaryotes* **4,** 951–964
2. Asao, M., and Madigan, M. T. (2010) *Photosynth. Res.* **104,** 103–111
3. Kimble, L. K., Mandelco, L., Woese, C. R., and Madigan, M. T. (1995) *Arch. Microbiol.* **163,** 259–267
4. Tang, K. H., Yue, H., and Blankenship, R. E. (2010) *BMC Microbiol.* **10,** 150
5. Sattley, W. M., Madigan, M. T., Swingley, W. D., Cheung, P. C., Clocksin, K. M., Conrad, A. L., Dejesa, L. C., Honchak, B. M., Jung, D. O., Karbach, L. E., Kurdoglu, A., Lahiri, S., Mastrian, S. D., Page, L. E., Taylor, H. L., Wang, Z. T., Raymond, J., Chen, M., Blankenship, R. E., and Touchman, J. W. (2008) *J. Bacteriol.* **190,** 4687–4696
6. Heinnickel, M., and Golbeck, J. H. (2007) *Photosynth. Res.* **92,** 35–53
7. Baymann, F., and Nitschke, W. (2010) *Photosynth. Res.* **104,** 177–187
8. Golbeck, J. H. (2010) *Photosynth. Res.* **104,** 101–102
9. Pickett, M. W., Williamson, M. P., and Kelly, D. J. (1994) *Photosynth. Res.* **41,** 75–88
10. Tang, Y. J., Hwang, J. S., Wemmer, D. E., and Keasling, J. D. (2007) *Appl. Environ. Microbiol.* **73,** 718–729
11. Overmann, J. (2006) *Prokaryotes* **7,** 359–378

12. Tang, K. H., Feng, X., Tang, Y. J., and Blankenship, R. E. (2009) *PLoS One.* **4,** e7233
13. Widdel, F., and Pfennig, N. (1984) in *Bergey's Manual of Systematic Bacteriology* (Krieg, N. R., and Holt, J. G., eds) Vol. 1, pp. 663–679, Williams & Wilkins, Baltimore
14. Brysch, K., Schneider, C., Fuchs, G., and Widdel, F. (1987) *Arch. Microbiol.* **148,** 264–274
15. Wolin, E. A., Wolin, M. J., and Wolfe, R. S. (1963) *J. Biol. Chem.* **238,** 2882–2886
16. Tang, K. H., Wen, J., Li, X., and Blankenship, R. E. (2009) *J. Bacteriol.* **191,** 3580–3587
17. Wahl, S. A., Dauner, M., and Wiechert, W. (2004) *Biotechnol. Bioeng.* **85,** 259–268
18. Bradford, M. M. (1976) *Anal. Biochem.* **72,** 248–254
19. Li, F., Hagemeier, C. H., Seedorf, H., Gottschalk, G., and Thauer, R. K. (2007) *J. Bacteriol.* **189,** 4299–4304
20. Lauble, H., Kennedy, M. C., Emptage, M. H., Beinert, H., and Stout, C. D. (1996) *Proc. Natl. Acad. Sci. U.S.A.* **93,** 13699–13703
21. Tang, Y., Pingitore, F., Mukhopadhyay, A., Phan, R., Hazen, T. C., and Keasling, J. D. (2007) *J. Bacteriol.* **189,** 940–949
22. Feng, X., Mouttaki, H., Lin, L., Huang, R., Wu, B., Hemme, C. L., He, Z., Zhang, B., Hicks, L. M., Xu, J., Zhou, J., and Tang, Y. J. (2009) *Appl. Environ. Microbiol.* **75,** 5001–5008
23. Marcus, A., and Elliott, W. B. (1956) *J. Biol. Chem.* **218,** 823–830
24. Pingitore, F., Tang, Y., Kruppa, G. H., and Keasling, J. D. (2007) *Anal. Chem.* **79,** 2483–2490
25. Tang, K. H., and Blankenship, R. E. (July 22, 2010) *J. Biol. Chem.* 10.1074/jbc.M110.157834
26. Tang, Y. J., Yi, S., Zhuang, W. Q., Zinder, S. H., Keasling, J. D., and Alvarez-Cohen, L. (2009) *J. Bacteriol.* **191,** 5224–5231
27. Jahn, U., Huber, H., Eisenreich, W., Hügler, M., and Fuchs, G. (2007) *J. Bacteriol.* **189,** 4108–4119
28. Lamzin, V. S., Dauter, Z., and Wilson, K. S. (1995) *Curr. Opin. Struct. Biol.* **5,** 830–836

Supporting Information

Isotopomer analysis. Cell pellets were hydrolyzed in 6 M HCl at 100°C for 24 hrs. After air-drying overnight, the dried samples containing free amino acids were derivatized with *N*-(tert-butyl-dimethylsilyl)-*N*-methyl-trifluoroacetamide in tetrahydrofuran at 70°C for 1 hr. Isotopomer measurements were made on a GC (Hewlett-Packard, model 6890, Agilent Technologies, Palo Alto, CA) equipped with a DB5-MS column (J&W Scientific, Falsom, CA) and a mass spectrometer (MS) (5975, Agilent Technologies, Palo Alto, CA). Several groups of charged fragments were detected by GC-MS for the amino acids: the $[M-57]^+$ or $[M-15]^+$ group, which contains un-fragmented amino acids; the $[M-159]^+$ or $[M-85]^+$ group, which contains amino acids losing α carboxyl group; and the f302 group, which contains the fragments of the first two carbons from derivatized amino acids. The $[M-57]^+$ peaks in leucine and isoleucine overlap with other peaks. Published algorithms were used to correct the effects of natural isotopes on the mass distributions of amino acids (1) mass fractions (i.e. M0, M1, M2… which are fractions of unlabeled, single-labeled, and double-labeled amino acids...).

Table S1. Sequences of primers described in this report.

Gene (loci number, predicted/ reported function)	Forward primer (5'→ 3')	Reverse primer (5'→ 3')
16S rRNA gene	GCAACGCGAAGAACCTTACC	GGGCACCCTCGCATCTC
pykA (HM1_0076, pyruvate kinase)	GCCCGAATCATCTCCATCAG	AACGCCCCGCACGAA
acn (HM1_0105, aconitase)	AATCAGCCTGTGGTCCCTGT	CAGAGACACCGCCCGAGTT
fdxR (HM1_0289, ferredoxin-$NADP^+$ reductase, FNR)	CCTGCTCCCGGTCAAAATC	TTCTTCGCGCCGATGAA
porA (HM1_0807, pyruvate: Fd oxidoreductase, PFOR)	GAAGCCTGCAACCCCTACTATAAG	GGTGAGTTTGCCGATCTCCTT
acsA (HM1_0951, acetyl-CoA synthetase)	TCCAAACCTGAAATCCTATGAAGAG	AGAACTCGCGCTCCACATCT
icd (HM1_1471, isocitrate dehydrogenase)	TCAACCCCGGATCGGTC	CTGCCAGCCGAGGTGTTC
mdh (HM1_1472, malate dehydrogenase	CGGCTATGAGGGCATCTACAC	CGGTCAGCTCGATCTCAAAGA
ackA (HM1_2157, acetate kinase)	CCCGCGTCGGTGACAT	CGTCAATCCCTCTTTTTCCATC
ppdK (HM1_2461, pyruvate phosphate dikinase)	AGATGTCGTTGCCGGTATCC	AAGCATTCGGGCAGTTCTTC
korC (HM1_2762, KFOR, γ subunit)	GGGAAGCCCTCGAAAAAGC	TGTTCATCTCCTCGGTCCCT
oorB (HM1_2763, KFOR, β subunit)	AAAGGGACCACCGCTCCT	GCCAGGTTGCAGATGTCGA
korA (HM1_2766, KFOR, α subunit)	CGGCGACCATCCTGTCAT	CCGTCAGGTTGAAGCATTCC
korD (HM1_2767, KFOR, δ subunit)	AAGTGTTGGGCGCTGACG	TGCACTTGGTACATTTTTTCGG
nifV (HM1_0858, homocitrate synthase)	GAAGCCTATCCGCCCGA	GCTGTATTTGCCAAAGGCGA
aksA (HM1_2993, homocitrate synthase)	CGCTTCCCGTTCTGATATTGA	CGCCCAGCTTTTTGGCTT
pckA (HM1_2773, PEP carboxykinase)	GATGCCATCTTCCACGAGGTA	CAGTCCCTGTTACGTGTCGAAA

Table S2. Isotopomer labeling patterns of protein-based amino acids in the *Desulfovibrio vulgaris* Hildenborough (DvH) cultures grown on non-labeled lactate and [2-^{13}C]acetate.[a]

Amino acid	Precursor	Ion	M-57	M-159
Ala	pyruvate	M0	0.89	0.85
		M1	**0.10**	**0.12**
		M2	**0.01**	**0.03**
Leu	pyruvate acetyl-CoA	M0		0.34
		M1	peaks	**0.49**
		M2	overlapped	**0.14**
Phe	phosphoenolpyruvate erythrose-4-phosphate	M0	0.65	0.63
		M1	**0.28**	**0.29**
		M2	**0.05**	**0.06**
Asp	oxaloacetate	M0	0.85	0.86
		M1	**0.14**	**0.10**
		M2	**0.01**	**0.02**
Glu	α-ketoglutarate	M0	0.43	0.42
		M1	**0.49**	**0.50**
		M2	**0.03**	**0.06**
His	ribose-5-phosphate	M0	0.78	0.81
		M1	**0.13**	**0.13**
		M2	**0.06**	**0.03**

[a] While DvH cannot grow using acetate as the sole carbon and energy source, DvH can grow with a mixed-substrate (lactate and acetate). When [2-^{13}C]acetate and non-labeled lactate with 1:1 molar ratio were presented in DvH growth medium, a doubling time ~ 9 hours was observed during the middle-log phase. The labeling carbons were significantly enriched in its protein-based amino acids for the cultures harvested during the late-log growth phase (see Table above), indicating that labeled acetate was partially utilized for protein synthesis in the mixed-substrate culture. The labeling data presented above also illustrates that acetate ↔ acetyl-CoA ↔ pyruvate is a reversible pathway. The DvH culture condition was described in the Experimental Procedures.

Table S3. Isotopomer labeling patterns of protein-based amino acids in the *H. modesticaldum* cultures grown on [3-^{13}C]pyruvate.

Amino acid	Precursor	Ion	M-57	M-159	f302	Proposed 13C enriched positions
Ala	pyruvate	M0	0.06	0.09	0.13	*C-C-COOH
		M1	0.89	0.80	0.85	
		M2	0.05	0.11	0.02	
Gly	serine	M0	0.70	0.72	0.28	*C-COOH
		M1	0.29	0.28	0.35	
		M2	0.01		0.37	
Val	pyruvate	M0	0.06	0.08	0.73	*C-*C-C-C-COOH
		M1	0.03	0.05	0.18	
		M2	0.87	0.78	0.10	
Leu	pyruvate	M0	0.06	0.08	0.28	*C-*C-C-C-*C-COOH
	acetyl-CoA	M1	0.02	0.02	0.30	
		M2	0.15	0.27	0.42	
		M3	0.75	0.57		
Ile	pyruvate	M0	0.05	0.05	0.22	*C-*C-C-C-*C-COOH
	threonine	M1	0.02	0.02	0.48	
		M2	0.27	0.34	0.29	
		M3	0.64	0.57		
Met	aspartate	M0	0.06	0.06	0.29	*C-S-C-*C-C-COOH
	methyl-THF	M1	0.37	0.39	0.54	
		M2	0.52	0.52	0.17	
		M3	0.05	0.04		
Ser	3-phospho-	M0	0.09	0.09	0.75	*C-C-COOH
	glycerate	M1	0.87	0.89	0.24	
		M2	0.04	0.02	0.01	
Thr	aspartate	M0	0.07	0.08	0.11	C-*C-C-COOH
		M1	0.83	0.86	0.64	
		M2	0.11	0.07	0.25	
Phe	PEP	M0	0.06	0.06	0.76	*C-*C-C-C-C-C-*C-C-COOH (clockwise)
	erythrose-4-	M1	0.01	0.01	0.23	
	phosphate	M2	0.04	0.05	0.01	
		M3	0.74	0.75		
		M4	0.14	0.13		
Asp	OAA	M0	0.06	0.06	0.58	C-*C-C-COOH
		M1	0.82	0.87	0.39	
		M2	0.12	0.07	0.03	

Glu	α-KG	M0	0.07	0.07	0.71	C-*C-C-*C-COOH
		M1	0.10	0.22	0.24	
		M2	0.76	0.67	0.05	
		M3	0.07	0.05		
His	ribose-5-	M0	0.06	0.13	0.14	N-*C-N-C-C-C-C-*COOH
	phosphate	M1	0.12	0.43	0.78	
		M2	0.53	0.35	0.08	
		M3	0.28	0.08		
Lys	aspartate	M0	0.07	0.07	0.91	C-*C-C-*C-C-COOH
	pyruvate	M1	0.05	0.04	0.09	
		M2	0.79	0.81	0.00	
		M3	0.09	0.06		
		M4	0.00	0.01		

Table S4. Isotopomer labeling patterns of protein-based amino acids in the *H. modesticaldum* cultures grown on ^{13}C-labeled sodium bicarbonate and non-labeled pyruvate.

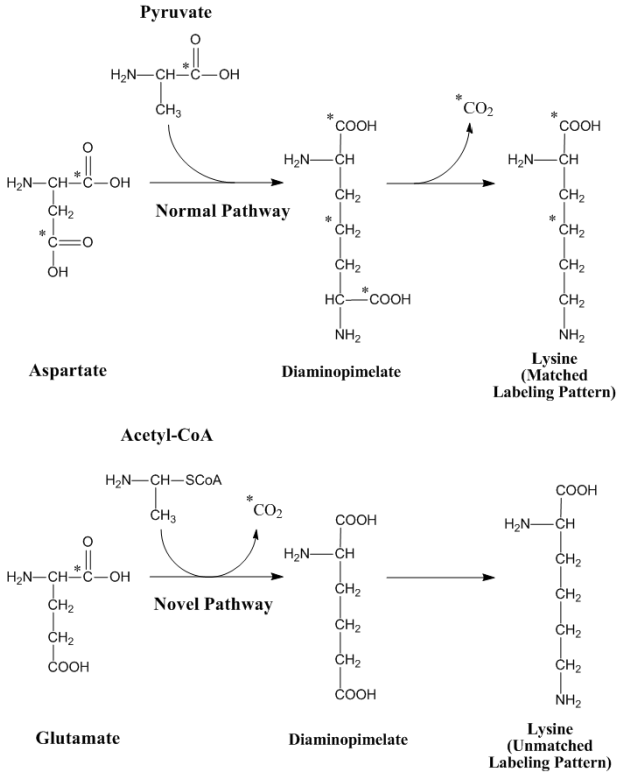
Amino acid	Precursors	Ions	M-57	M-159	f302	Proposed 13C enriched positions
Ala	pyruvate	M0	0.47	0.92	0.47	C-C-*COOH
		M1	0.52	0.03	0.53	
		M2	0.01	0.05	0.00	
Gly	serine	M0	0.49	0.99	0.24	C-*COOH
		M1	0.51	0.01	0.29	
		M2	0.01		0.47	
Val	pyruvate	M0	0.43	0.93	0.51	C-C-C-C-*COOH
		M1	0.54	0.05	0.48	
		M2	0.03	0.01	0.01	
Leu	pyruvate	M0	0.74	0.92	0.80	Non-labeled
	acetyl-CoA	M1	0.20	0.07	0.19	
		M2	0.02	0.01	0.01	
		M3	0.04	0.00		
Ile	pyruvate	M0	0.64	0.86	0.65	Non-labeled
	threonine	M1	0.29	0.13	0.31	
		M2	0.07	0.01	0.04	
		M3	0.00	0.00		
Met	aspartate	M0	0.23	0.42	0.28	C-S-*C-C-C-*COOH
	methyl-THF	M1	0.45	0.52	0.42	
		M2	0.29	0.04	0.30	
		M3	0.03	0.01		
Ser	3-phospho-	M0	0.45	0.96	0.48	C-C-*COOH
	glycerate	M1	0.54	0.03	0.51	
		M2	0.02	0.01	0.01	
Thr	aspartate	M0	0.23	0.43	0.20	*C-C-C-*COOH
		M1	0.47	0.57	0.49	
		M2	0.30	0.02	0.31	
Phe	PEP	M0	0.14	0.27	0.41	C-C-*C-*C-C-C-C-C-*COOH
	erythrose-4-	M1	0.32	0.45	0.59	
	phosphate	M2	0.35	0.25	0.00	
		M3	0.18	0.03		
		M4	0.01	0.00		
Asp	OAA	M0	0.24	0.47	0.47	*COOH-C-C-*COOH
		M1	0.46	0.52	0.53	
		M2	0.31	0.01	0.00	
Glu	α-KG	M0	0.38	0.43	0.78	C-C-C-C-*COOH

		M1	0.50	0.54	0.21	
		M2	0.11	0.02	0.01	
		M3	0.00	0.00		
Lys	aspartate	M0	0.26	0.42	0.87	C-C-*C-C-C-*COOH
	Pyruvate	M1	0.41	0.54	0.13	
		M2	0.34	0.03	0.00	
		M3	0.00	0.01		
		M4	0.01	0.00		

Table S5. List of bacteria with the citramalate pathway reported.

Species	References	CimA similarity (GSU1798)	CimA similarity (MJ1392)	Growth conditions
Roseobacter denitrifican	(2)	RD1_3182, 45%	RD1_1121, 40%	aerobic
Methanobacterium thermoautotrophicum	(3)	MTH1481, 28%	MTH723, 58%	anaerobic
Methanococcus jannaschii	(4)	ND	MJ1392, 100%	aerobic (*E. coli*)
Leptospira interrogans	(5,6), (7)	LIC11726, 26%	LIC11726, 41%	aerobic (*E. coli*)
Thermoproteus neutropkilccs	(8)	Tneu_0320, 45%	Tneu_0832, 55%	anaerobic, thermophilic
Ignicoccus hospitalis	(9)	Igni_0645, 45%	Igni_0983, 52%	anaerobic, thermophilic
Geobacter sulfurreducens	(10)	GSU1798, 100%	GSU1906, 41%	anaerobic
Geobacter metallireducens	(11)	Gmet_1879, 92%	Gmet_1265, 42%	anaerobic
Serratia marcescens	(12)	Spro_0745, 26%	Spro_0745, 37%	anaerobic
Thermoanaerobacter sp. X514	(13)	Teth514_1204 49%	Teth514_0415 45%	anaerobic, thermophilic
Dehalococcoides ethenogenes 195	(14)	DET0825, 53%	DET0830, 41%	anaerobic
Heliobacterium modesticaldum	(15) and this report	HM_1519, 55%	HM_1515, 45%	anaerobic

Fig. S1. The normal versus the novel pathway for lysine biosynthesis in the *H. modesticaldum* cultures grown on [1-^{13}C]pyruvate. Predicted labeled carbons are marked by asterisks.



References in Supporting Information

1. Wahl, S. A., Dauner, M., and Wiechert, W. (2004) *Biotechnol. Bioeng.* **85**(3), 259-268
2. Tang, K. H., Feng, X., Tang, Y. J., and Blankenship, R. E. (2009) *PLoS One* **4**(10), e7233
3. Eikmanns, B., Linder, D., and Thauer, R. K. (1983) *Arch. Microbiol.* **136**(2), 111-113
4. Howell, D. M., Xu, H. M., and White, R. H. (1999) *J. Bacteriol.* **181**(1), 331-333
5. Charon, N. W., Johnson, R. C., and Peterson, D. (1974) *J Bacteriol* **117**(1), 203-211
6. Westfall, H. N., Charon, N. W., and Peterson, D. E. (1983) *J. Bacteriol* **154**(2), 846-853
7. Xu, H., Zhang, Y. Z., Guo, X. K., Ren, S. X., Staempfli, A. A., Chiao, J. S., Jiang, W. H., and Zhao, G. P. (2004) *J. Bacteriol.* **186**(16), 5400-5409
8. Schäfer, S., Paalme, T., Vilu, R., and Fuchs, G. (1989) *Eur J Biochem* **186**, 695-700
9. Jahn, U., Huber, H., Eisenreich, W., Hugler, M., and Fuchs, G. (2007) *J. Bacteriol.* **189**(11), 4108-4119
10. Risso, C., Van Dien, S. J., Orloff, A., Lovley, D. R., and Coppi, M. V. (2008) *J. Bacteriol.* **190**(7), 2266-2274
11. Tang, Y. J., Chakraborty, R., Martin, H. G., Chu, J., Hazen, T. C., and Keasling, J. D. (2007) *Appl. Environ. Microbiol.* **73**(12), 3859-3864
12. Kisumi, M., Komatsubara, S., and Chibata, I. (1977) *J. Biochem.* **82**(1), 95-103
13. Feng, X., Mouttaki, H., Lin, L., Huang, R., Wu, B., Hemme, C. L., He, Z., Zhang, B., Hicks, L. M., Xu, J., Zhou, J., and Tang, Y. J. (2009) *Appl. Environ. Microbiol.* **75**(15), 5001-5008
14. Tang, Y. J., Yi, S., Zhuang, W. Q., Zinder, S. H., Keasling, J. D., and Alvarez-Cohen, L. (2009) *J. Bacteriol.* **191**(16), 5224-5231
15. Pickett, M. W., Williamson, M. P., and Kelly, D. J. (1994) *Photosynth. Res.* **41**, 75-88

Appendix 4

Carbohydrate Metabolism and Carbon Fixation in

Roseobacter denitrificans OCh114

OPEN ACCESS Freely available online

PLoS one

Carbohydrate Metabolism and Carbon Fixation in *Roseobacter denitrificans* OCh114

Kuo-Hsiang Tang[1], Xueyang Feng[2], Yinjie J. Tang[2], Robert E. Blankenship[1]*

1 Departments of Biology and Chemistry, Washington University in St. Louis, St. Louis, Missouri, United States of America, **2** Department of Energy, Environment and Chemical Engineering, Washington University in St. Louis, St. Louis, Missouri, United States of America

Abstract

The *Roseobacter* clade of aerobic marine proteobacteria, which compose 10–25% of the total marine bacterial community, has been reported to fix CO_2, although it has not been determined what pathway is involved. In this study, we report the first metabolic studies on carbohydrate utilization, CO_2 assimilation, and amino acid biosynthesis in the phototrophic *Roseobacter* clade bacterium *Roseobacter denitrificans* OCh114. We develop a new minimal medium containing defined carbon source(s), in which the requirements of yeast extract reported previously for the growth of *R. denitrificans* can be replaced by vitamin B_{12} (cyanocobalamin). Tracer experiments were carried out in *R. denitrificans* grown in a newly developed minimal medium containing isotopically labeled pyruvate, glucose or bicarbonate as a single carbon source or in combination. Through measurements of ^{13}C-isotopomer labeling patterns in protein-derived amino acids, gene expression profiles, and enzymatic activity assays, we report that: (1) *R. denitrificans* uses the anaplerotic pathways mainly via the malic enzyme to fix 10–15% of protein carbon from CO_2; (2) *R. denitrificans* employs the Entner-Doudoroff (ED) pathway for carbohydrate metabolism and the non-oxidative pentose phosphate pathway for the biosynthesis of histidine, ATP, and coenzymes; (3) the Embden-Meyerhof-Parnas (EMP, glycolysis) pathway is not active and the enzymatic activity of 6-phosphofructokinase (PFK) cannot be detected in *R. denitrificans*; and (4) isoleucine can be synthesized from both threonine-dependent (20% total flux) and citramalate-dependent (80% total flux) pathways using pyruvate as the sole carbon source.

Citation: Tang K-H, Feng X, Tang YJ, Blankenship RE (2009) Carbohydrate Metabolism and Carbon Fixation in *Roseobacter denitrificans* OCh114. PLoS ONE 4(10): e7233. doi:10.1371/journal.pone.0007233

Editor: Sebastian D. Fugmann, National Institute on Aging, United States of America

Received July 12, 2009; **Accepted** September 4, 2009; **Published** October 1, 2009

Funding: This work was supported by a grant from the Exobiology program of NASA to R. E. Blankenship, and in part supported by a DOE bioenergy research grant (DEFG0208ER64694) and by Missouri Life Sciences Research Board to YJ Tang. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: blankenship@wustl.edu

Introduction

Two of the most important sources of carbon sinks known in nature are absorption of CO_2 by the oceans and photosynthesis by photosynthetic organisms. The marine *Roseobacter* clade are potentially major contributors to global CO_2 fixation as they make up 10 25% of the total microbial community in some surface ocean ecosystems [1 4]. Some members of the *Roseobacter* clade belong to a group known as Aerobic Anoxygenic Phototrophs (AAPs), the only known organisms performing photosynthesis requiring oxygen but not producing oxygen, while other members are non-phototrophic. CO oxidation was confirmed experimentally for the non-phototrophic *Roseobacter* clade bacterium *Silicibacter pomeroyi* [5] and other *Roseobacter* clade [6], and CO_2 fixation was suggested in several marine AAPs [7], while bioinformatic analysis in *Roseobacter* clade with completed genome sequence indicated that the genes encoding ribulose bisphosphate carboxylase/oxygenase (RUBISCO) and phosphoribulokinase required in the Calvin cycle for carbon fixation, as well as genes for other autotrophic CO_2 fixation pathways, are missing in these bacteria [5,8 10]. Thus, it has been of great interest to determine how *Roseobacter* clade bacteria can fix CO_2, if they indeed fix CO_2.

Anaplerotic pathways have been proposed as an alternative mechanism for CO_2 fixation in *Roseobacter* clade [5,8,11], but have not been verified experimentally. Given that organisms in the *Roseobacter* clade are known to require organic carbon sources for growth [12,13], understanding how *Roseobacter* clade bacteria utilize organic carbon and assimilate CO_2 will help us understand the bio-energy metabolism, production of bioactive metabolites, and roles of global carbon cycle in these wide-spread marine bacteria.

Here, we report metabolic and biochemical studies of *Roseobacter denitrificans* OCh114 [14], which can denitrify, as its name indicates [13], and produce bacteriochlorophyll *a* (BChl *a*) aerobically. The genomic sequence of *R. denitrificans* [8] suggested that the tricarboxylic acid (TCA) cycle and anaplerotic pathways are complete, and that most of the genes for carbohydrate metabolism in the Embden-Meyerhof-Parnas (EMP, glycolysis), Entner-Doudoroff (ED), and pentose phosphate (PP) pathways are present and annotated. To understand the contributions of these enzymes and reaction pathways in the metabolism of *R. denitrificans*, we developed a minimal growth medium containing only defined carbon source(s), optimized the growth conditions of *R. denitrificans* in different defined carbon sources, and used isotopomer assisted metabolite analyses, biochemical approaches, and gene expression profiles to investigate the carbon assimilation, carbohydrate utilization, and amino acid biosynthesis pathways in *R. denitrificans*.

Results

Growth of R. denitrificans OCh114 in different growth conditions

Similar growth curves and spectral features were obtained in the *R. denitrificans* OCh114 cultures grown aerobically in the light (20 W/m^2), dark, and day-night cycles (Figure S1a and S1b), and slightly higher OD_{600} were reached in the dark and day-light cycles compared to in the continuous light growth conditions (Figure S1c). It is consistent with the reports that light has a negative effect on pigment formation in some AAPs [15]. No growth was observed anaerobically regardless of the light intensity (data not shown). Multiple organic carbon sources have been tested for the growth of *R. denitrificans* OCh114 [13]. Here, we report studies of three carbon sources: pyruvate, D-glucose, and CO_2 (or HCO_3^-). Spectral features of the photosynthesis system and light-harvesting complexes for the cultures grown in the minimal medium containing pyruvate or glucose are similar to the cultures grown in the rich medium (Figure 1a and 1b) as well as the results reported previously [13,16]. The image of *R. denitrificans* cells grown in the minimal medium with pyruvate was examined by OLYMPUS FV1000/BX61 high-resolution confocal microscope, and the morphology and average cell size is consistent with previous reports [13]. Uptake of pyruvate, $2.5\times10^{-2}\pm5\times10^{-4}$ mmole per hour (Figure 1d and S1d), is approximately 2 to 3-fold faster than uptake of glucose by *R. denitrificans* (Figure 1e), consistent with the faster growth in pyruvate than in D-glucose (Figure 1c). Higher OD_{600} and better cell growth can be obtained using 0.2% pyruvate in the minimal medium (data not shown). No differences in the ^{13}C-isotopomer abundances of the protein-derived amino acids for cultures grown in the minimal medium containing pyruvate were observed under either dark or illuminated conditions (Table S1a and S1b). Similar results were also observed in the cultures grown in the minimal medium supplied with glucose.

Vitamin B_{12} is required for the growth of R. denitrificans

The previous studies of the growth of *R. denitrificans* OCh114 or other AAPs have been performed in either a rich medium or a medium containing either 0.02 g/liter [13] or 0.1 g/liter [17] of yeast extract (undefined carbon sources). We confirmed the necessity of yeast extract, as poor growth of the *R. denitrificans* cultures without yeast extract was observed (data not shown). Although it has been recognized that yeast extract contains rich vitamin mixtures, it also includes amino acids and other undefined carbon sources. Ideally, the ^{13}C-isotopic labeling studies require a minimal medium containing only defined carbon source(s), thus it is desirable to optimize the growth conditions by eliminating the yeast extract. We found that vitamin B_{12} (cyanocobalamin) can serve as an alternative to yeast extract for growing *R. denitrificans* in defined carbon sources, as cultures with $OD_{600}\geq3$ can be reached in our minimal growth medium with glucose as the sole carbon source (Figure 1e). Vitamin B_{12} and different forms of cobalamin are required for methionine and protein synthesis, deoxyribonucleotide triphosphate synthesis, amino acid metabolisms, and CO_2 fixation (in methanogens, although there is no such pathway identified in *R. denitrificans*), and are included in the growth media of many photoautotrophic, photoheterotrophic and chemoheterotrophic bacteria. Requirement for vitamin B_{12} in the growth of *R. denitrificans* may partially explain why yeast extract was necessary to be included in the minimal medium containing either pyruvate or glucose.

The carbohydrate utilization pathways in R. denitrificans

In the cultures grown in the minimal medium containing either D-[1-^{13}C]glucose or D-[6-^{13}C]glucose, the isotopomer labeling data of serine (the precursor is 3-phosphoglycerate) and alanine (the precursor is pyruvate) were different (Table S1e, S1f, S1i, and S3). Three metabolic pathways can be employed for sugar utilization by *R. denitrificans* and need to be considered to account for the isotopomer abundance in these protein-derived amino acids: (1) the Embden-Meyerhof-Parnas (EMP) pathway (glycolysis), by which one [1-^{13}C]glucose molecule is cleaved into two glyceraldehyde-3-phosphate (GAP) molecules: one is [3-^{13}C]-labeled and the other is unlabeled using either D-[1-^{13}C]glucose or D-[6-^{13}C]glucose; (2) the Entner-Doudoroff (ED) pathway [18], by which one molecule of glucose generates one molecule of GAP, in which the third carbon is labeled using D-[6-^{13}C]glucose, and one molecule of pyruvate, where the first carbon is labeled using D-[1-^{13}C]glucose (Figure 2a); (3) the pentose phosphate (PP) pathway, the first carbon of glucose is removed as CO_2, if the oxidative PP pathway is active, to generate ribose-5-phosphate, which can be converted to GAP through the non-oxidative PP pathway (Figure S2). As shown in Figure 2a, regardless of the pathways, GAP is converted into 3-phosphoglycerate (3-PGA), the precursor of glycine, serine, and cysteine, and then to pyruvate, the precursor of alanine, valine, leucine, and isoleucine.

If *R. denitrificans* predominantly uses the EMP pathway for carbohydrate metabolism, one will expect similar ^{13}C-isotopomer abundance in serine and alanine (converted from pyruvate through alanine aminotransferase). However, this is not consistent with the higher M+0 value (more unlabeled carbon) in serine (0.74) than in alanine (0.56) using D-[1-^{13}C]glucose in our studies (Table S1e and S3). Alternatively, genes encoding two key enzymes in the ED pathway can be found in *R. denitrificans*: *eda* (RD1_2878), encoding 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase (EC 4.1.2.14), and *edd* (RD1_2879), encoding phosphogluconate dehydratase (EC 4.2.1.12). If *R. denitrificans* uses the ED pathway as one of the alternative carbohydrate utilization pathways, [1-^{13}C]glucose is converted into [1-^{13}C]-2-keto-3-deoxy-6-phosphogluconate, which is cleaved into [1-^{13}C]pyruvate and unlabeled GAP, leading to unlabeled 3-PGA (Figure 2b). In this case, serine is expected to be mostly unlabeled, while half of the alanine, derived from [1-^{13}C] pyruvate, is labeled, consistent with our experimental data.

The QRT-PCR results indicate that both *eda* (RD1_2878) and *edd* (RD1_2879) genes are expressed and the transcript level of these genes is higher in the minimal medium containing either pyruvate or glucose than in the rich medium, similar to the gene expression profiles of other genes responsible for carbon fixation and carbon metabolism examined in this report (Figure 3). Moreover, the activity of 2-keto-3-deoxy-phosphogluconate (KDPG) aldolase and phosphogluconate dehydrase can be detected in cell-free extracts. Alternatively, the gene encoding 6-phosphofructokinase (PFK, EC 2.7.1.11), an essential enzyme for the EMP pathway, is not annotated, and no PFK activity can be detected in cell free extracts.

The proposed ED pathway was further tested using D-[6-^{13}C]glucose, in which [6-^{13}C]KDPG is cleaved into unlabeled pyruvate and [3-^{13}C]GAP in the ED pathway (Figure 2b). Our data showed a smaller M+0 value in serine using D-[6-^{13}C]glucose (0.30) than using D-[1-^{13}C]glucose (0.74), compared to slightly higher M+0 value in alanine using D-[1-^{13}C]glucose (0.56) versus D-[6-^{13}C] glucose (0.48) (Table S1e and S1i). The M+0 value in serine using D-[1-^{13}C]glucose or D-[6-^{13}C]glucose is not close to 1 or 0, respectively, suggesting that in addition to the ED pathway, the PP pathway is also active and generating GAP with different labeling pattern using D-[6-^{13}C]glucose or D-[1-^{13}C] glucose (Figure 2a). Based on the isotopomer abundance of serine using D-[1-^{13}C]glucose and D-[6-^{13}C]glucose, we estimated 25~30% of GAP is produced from the (non-oxidative) PP pathway (Table S3).

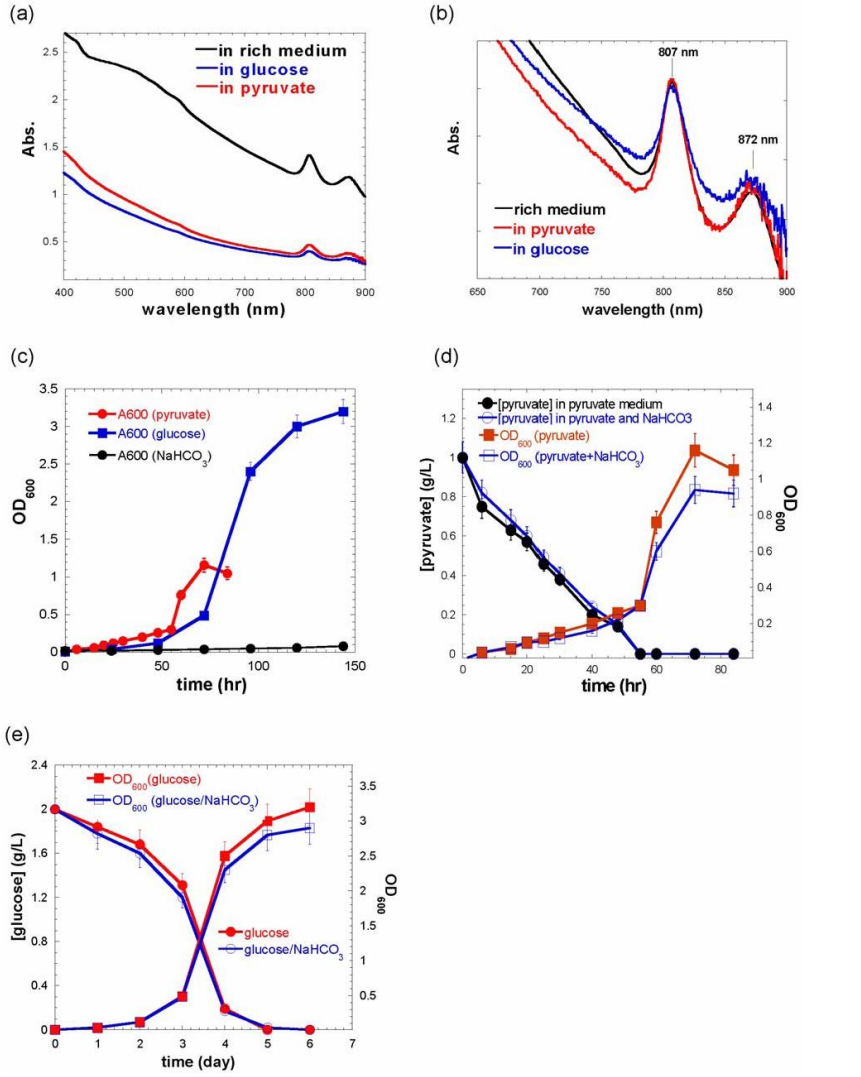


Figure 1. Spectra, image and organic carbon uptake of *R. denitrificans* OCh114. The spectra of cultures grown in the rich medium vs. in the minimal medium supplied with either 0.1% pyruvate or 0.1% glucose (a). No undefined carbon sources were included in the minimal medium reported in this paper. The normalized spectra in the 650 nm–900 nm range (b), the cell growth in the minimal medium with pyruvate, glucose, or HCO_3^- (c), and the cell growth curve and uptake of pyruvate (d) or glucose (e) with or without the addition of 0.2% $NaHCO_3$ in the minimal medium containing pyruvate or glucose. More than ten biological replicates were performed in every growth conditions.
doi:10.1371/journal.pone.0007233.g001

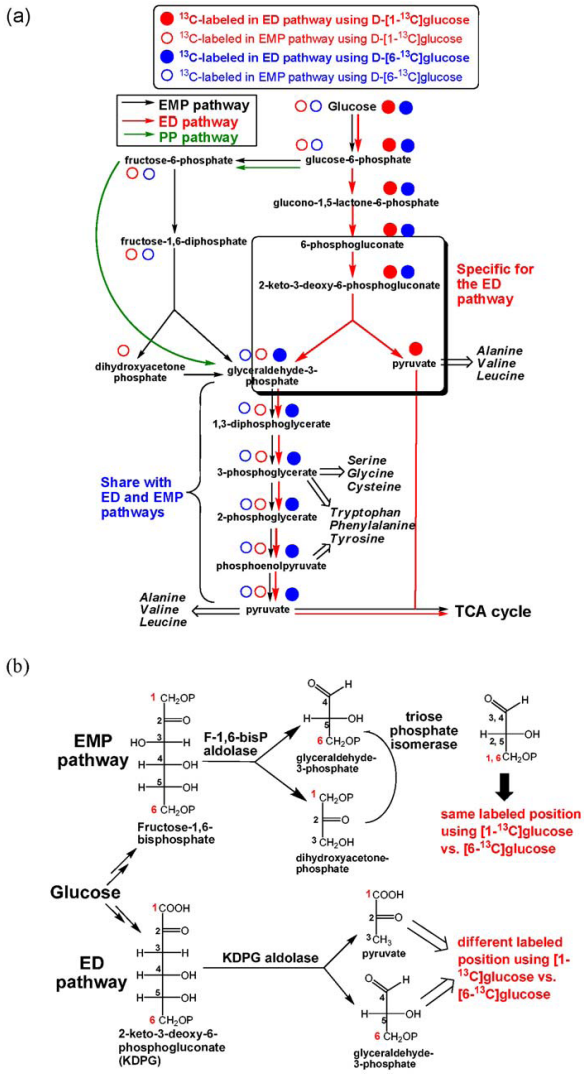


Figure 2. Proposed carbohydrate metabolism and amino acid biosynthesis pathways in *R. denitrificans*. Predicted ^{13}C-labeling distributions using D-[1-^{13}C]glucose and D-[6-^{13}C]glucose in the EMP and ED pathways are shown (a), and the cleavage of a hexose molecule in the EMP vs. ED pathway (b). Abbreviations: EMP, Embden-Meyerhof-Parnas (glycolysis); ED, Entner-Doudoroff; and PP, pentose phosphate. Experimentally identified ^{13}C-labeling patterns are reported in the context.
doi:10.1371/journal.pone.0007233.g002

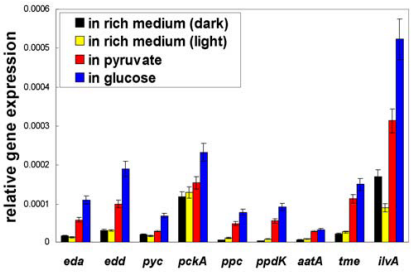


Figure 3. QRT-PCR studies. Gene name (encoding enzyme and gene location number): *eda* (2-keto-3-deoxy-6-phosphogluconate aldolase, RD1_2878), *edd* (6-phosphogluconate dehydratase, RD1_2879), *pyc* (pyruvate carboxylase, RD1_3376), *pckA* (phosphoenolpyruvate carboxykinase, RD1_1376), *ppc* (phosphoenolpyruvate carboxylase RD1_4248), *ppdK* (pyruvate phosphate dikinase, RD1_1948), *aatA* (aspirate aminotransferase, RD1_3892), *tme* (malic enzyme, RD1_0421) and *ilvA* (threonine deaminase, RD1_0416). Relative gene expression value of each gene is calculated with $2^{-\Delta Ct}$, where $\Delta Ct = Ct_{target\ gene} - Ct_{16S\ rRNA\ gene}$, and the 16S rRNA gene was used as the internal reference. Three biological replicates and eighteen technical replicates were preformed for every gene.
doi:10.1371/journal.pone.0007233.g003

The ED pathway also leads to different predicted labeling in the aromatic amino acids. Phosphoenolpyruvate (PEP), synthesized from 3-PGA, and erythrose-4-phosphate (E4P), the intermediate in the non-oxidative PP pathway, are the precursors of phenylalanine, tyrosine and tryptophan. Labeled E4P in the non-oxidative PP pathway (Figure S2) leads to lower M+0 value in the aromatic amino acids (0.50+0.03) compared to serine (0.74) using D-[1-^{13}C]glucose. The ED and PP pathways can explain much higher M+0 value in phenylalanine (0.48) and tyrosine (0.53) using D-[1-^{13}C]glucose than the value in these amino acids (0.04) using D-[6-^{13}C]glucose, and more labeled carbon can be incorporated into PEP and E4P using D-[6-^{13}C]glucose (Figure 2a and 4).

The Pentose Phosphate pathway in R. denitrificans

A smaller M+0 value in histidine (0.28) than other amino acids (0.42–0.89) was detected using D-[1-^{13}C]glucose (Table S1e). Histidine is synthesized from 5-phospho-ribosyl-α-pyrophosphate (PRPP), the purine ring of ATP (the carbon from the formyl group of N^{10}-formyl-tetrahydrofolate (THF)), and glutamine (the nitrogen source). The 6-phosphogluconate, an intermediate of the ED/oxidative PP pathway, is the precursor of PRPP. In the oxidative PP pathway (Figure 4), the first (labeled) carbon in D-[1-^{13}C]glucose is released as CO_2 during the conversion of 6-phosphogluconate into D-ribulose-5-phosphate by 6-phosphogluconate dehydrogenase (PGD, EC 1.1.1.44), and it is unknown if the oxidative PP pathway is complete because the *pgd* gene in *R. denitrificans* is not annotated. Thus, the ^{13}C-isotopomer abundance in histidine using D-[1-^{13}C]glucose is possible from the following sources: (1) from the N^{10}-formyl moiety of THF if the oxidative PP pathway is active, (2) from D-ribose-5-phosphate or D-ribulose-5-phosphate derived from the non-oxidative PP pathway, where the labeled carbon is maintained (Figure 4), or both. The N^{10}-formyl moiety of THF is generated from THF and formate by transformylase, and various pathways can lead to formate biosynthesis in *R. denitrificans*. It is less likely that the labeled histidine using D-[1-^{13}C]glucose was produced exclusively from the ^{13}C-labeled formate through the oxidative PP pathway.

Alternatively, the non-oxidative PP pathways can also generate D-ribose-5-phosphate through steps of isomerization, epimerization and transketolation (Figure S3). If the metabolic flux goes through the non-oxidative PP pathway, more than one ^{13}C-labeled sources are available using D-[6-^{13}C]glucose (i.e. labeled GAP from the ED pathway and labeled fructose-6-phosphate (F6P)), while only F6P is labeled using D-[1-^{13}C]glucose (GAP is not labeled via the ED pathway using D-[1-^{13}C]glucose) (Figure 4). Since the EMP pathway in *R. denitrificans* is likely to be inactive, based on the experimental data and genomic information presented above, the labeled F6P is only led into the non-oxidative PP pathway. It is the most straightforward explanation for the detected ^{13}C-labeling abundance in histidine (M+0 value is 0.28) with D-[1-^{13}C]glucose (Table S1e) and for smaller M+0 value in histidine using D-[6-^{13}C]glucose (0.10) (Table S1i) than using D-[1-^{13}C]glucose (0.28). Together, our studies imply that *R. denitrificans* uses the non-oxidative PP pathway for histidine, ATP, coenzymes, and nucleic acids biosynthesis, as well as supplying GAP for the ED pathway (Figure 2a and 4).

Carbon fixation in R. denitrificans

Previous reports suggested that *R. denitrificans* [19] and other AAPs [11,20] can fix CO_2. In our studies, no growth of *R. denitrificans* was observed in the minimal medium with CO_2 (or bicarbonate) as the sole carbon source (Figure 1c), consistent with the finding that the key genes required for the Calvin cycle are missing in the genome of *R. denitrificans* OCh114 [8]. Moreover, no CO_2-enhanced growth was obtained in the minimal medium containing pyruvate or glucose (Figure 1d and 1e), and ^{13}C-isotopomer distribution of protein-based amino acids suggests that the presence of bicarbonate has a rather small effect on pyruvate and glucose metabolism in *R. denitrificans* (Table S1c and S1f). These experimental data indicate a carbon-fixation pathway other than the Calvin cycle is utilized by *R. denitrificans* OCh114. Swingley et al. [8] proposed that *R. denitrificans* could fix CO_2 by pyruvate or/and PEP to form oxaloacetate (OAA). Using uniformly ^{13}C-labeled D-glucose (D-[U-$^{13}C_6$] glucose) (Table S1j and S1k), our studies show less fully labeled amino acids derived from OAA (aspartate (M+4, 0.83), methionine (M+5, 0.78) and threonine (M+4, 0.85)) than the amino acid derived from pyruvate (alanine (M+3, 0.93)) (Table S1k). Also, a smaller M+0 value can be seen in threonine (0.86) and aspartate (0.87) than alanine (0.94) using ^{13}C-$NaHCO_3$ and unlabeled pyruvate (Table S1d). Similar results are also obtained using ^{13}C-$NaHCO_3$ and unlabeled glucose (Table S1g). Together, our studies imply that *R. denitrificans* fixes approximately 10–15% of protein carbon from CO_2 via the anaplerotic pathways, despite lacking the Calvin cycle. These experimental evidences support the CO_2 fixation pathways by *R. denitrificans* proposed by Swingley et al. [8].

The TCA cycle and anaplerotic pathways

Our studies imply that *R. denitrificans* has an active TCA cycle and metabolic fluxes in the anaplerotic pathways. PEP can be converted into OAA, the precursor of aspartate, asparagine, methionine, threonine, lysine and isoleucine. OAA can be also generated from pyruvate and through the TCA cycle. Using D-[1-^{13}C] glucose, the M+0 value is similar in aspartate (0.58), threonine (0.47), methionine (0.49) and alanine (0.56) but smaller than serine (0.74) (Table S1e), suggesting that ^{13}C-isotopomer abundance in the amino acids derived from OAA is largely from pyruvate (via the anaplerotic pathway) and/or the TCA cycle flux.

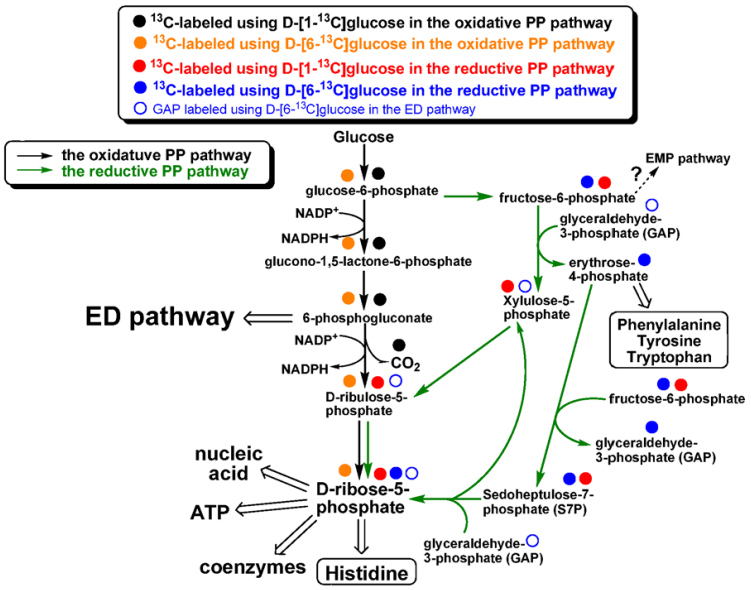


Figure 4. Proposed pentose phosphate pathway in *R. denitrificans*. All of the steps are reversible, except expulsion of CO_2 catalyzed by 6-phosphogluconate dehydrogenase. Predicted ^{13}C-labeling distributions using D-[1-^{13}C]glucose and D-[6-^{13}C]glucose in the oxidative PP and non-oxidative PP pathways are shown. Experimentally identified ^{13}C-labeling patterns are reported in the context.
doi:10.1371/journal.pone.0007233.g004

Four anaplerotic enzymes are known to assimilate CO_2 and replenish the intermediates of the TCA cycle, and genes encoding these anaplerotic enzymes can be all found in the genome of *R. denitrificans* OCh114; malic enzyme (*me*, RD1_0421, EC 1.1.1.40, pyruvate⇌malate), PEP carboxykinase (*pckA*, RD1_1376, EC 4.1.1.49, PEP⇌OAA), pyruvate carboxylase (*pyc*, RD1_3376, EC 6.4.1.1, pyruvate⇌OAA) and PEP carboxylase (*ppc*, RD1_4248, EC 4.1.1.31, PEP⇌OAA). To further investigate carbon fixation by *R. denitrificans*, we used gene expression studies to confirm if the pathways are active and to identify the transcriptional level of the genes encoding enzymes for the pathways under various growth conditions, as well as ^{13}C-isotopomer abundance to probe the anaplerotic flux. In the conditions we tested, the transcript level is higher for *pckA* (RD1_1376) and *me* (RD1_0421) than for *pyc* (RD1_3376) and *ppc* (RD1_4248) (Figure 3). The activity of these four anaplerotic enzymes can be detected in cell-free crude extracts, consistent with the gene expression profiles. The activity is higher in PEP carboxykinase (6 umole/min·mg protein) and pyruvate carboxylase (2.5 umole/min·mg protein) than in PEP carboxylase (0.7 umole/ min·mg protein) in the cell extracts from the rich dark cultures, while the activity of pyruvate carboxylase and PEP carboxylase are comparable (1–2 umole/min·mg protein) in the minimal medium supplied with pyruvate.

To investigate metabolic fluxes through the anaplerotic pathways, [3-^{13}C]pyruvate was used and the ^{13}C-isotopomer abundance in aspartate (from OAA), serine (from PEP, 3-PGA) and alanine (from pyruvate) was analyzed (Table S1h). Both QRT-PCR and ^{13}C-labeling pattern suggest that the flux of pyruvate to PEP is weak (different isotopomer distribution between alanine and serine), consistent with low gene expression of *ppdK* (RD1_1948, pyruvate phosphate dikinase, EC 2.7.9.1, pyruvate→PEP) and that the pathway from OAA to PEP is likely to be active and pyruvate to OAA is not as strong. This is in agreement with the gene expression profile of *pckA* (RD1_1376) and *pyc* (RD1_3376). Hence, most of the ^{13}C-labeling in OAA likely comes from the TCA cycle flux (from malate), in which the pyruvate to malate pathway (via malic enzyme) should be rather active, compatible with higher expression of *me* (RD1_0421) than *pyc* (RD1_3376). The proposed anaplerotic flux is summarized in Figure 5.

The isoleucine and leucine biosynthesis pathways

In many organisms, isoleucine is converted from threonine by threonine deaminase (EC 4.3.1.19) in the threonine-dependent pathway [21], and leucine is made from valine. The ^{13}C-isotopomer abundance in [1-^{13}C]pyruvate cultures showed higher M+0 value in leucine (0.91) and isoleucine (0.86) compared to valine (0.25) and threonine (0.70) (Table S1a). Most of the leucine should be unlabeled because the first (labeled) carbon in ^{13}C-(2*R*,3*S*)-3-isopropylmalate is released as CO_2 catalyzed by 3-isopropylmalate dehydrogenase (LeuB, EC 1.1.1.85) (Figure 6). However, highly unlabeled isoleucine is not expected if 2-oxobutanoate, a precursor

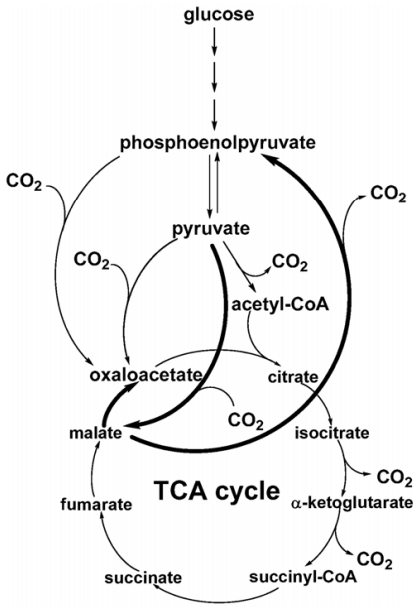


Figure 5. Proposed metabolic flux for carbon fixation by *R. denitrificans*. Bold curves represent proposed stronger metabolic fluxes.
doi:10.1371/journal.pone.0007233.g005

of isoleucine biosynthesis, is derived from threonine in the threonine-dependent pathway (Figure 6). The results suggest that isoleucine is not exclusively derived from threonine. If it is, the M+0 values in threonine and isoleucine should be similar.

Alternatively, a citramalate-dependent pathway, which is a threonine-independenet pathway, was recently reported in several bacteria [22–24]. These microorganisms can synthesize isoleucine from 2-oxobutanoate either exclusively from the citramalate-dependent pathway [24] or from both threonine- and citramalate-dependent pathways [23]. The higher M+0 value in isoleucine in our studies (Table S1a) can be explained if a significant portion of isoleucine is synthesized through the citramalate-dependent pathway, in which LeuB removes the first (labeled) carbon in ^{13}C-D-*erythro*-3-methylmalate as CO_2 (Figure 6). Similarly, the M+0 value is higher in isoleucine (0.90) than in threonine (0.77) using [1-^{13}C]pyruvate and unlabeled $NaHCO_3$ (Table S1c).

In the studies of the threonine-dependent pathway, QRT-PCR showed that the gene encoding theronine deaminase (*ilvA*, RD1_0416) was active in various growth conditions (Figure 3). The activity of citramalate synthase (CimA, EC 2.3.1.182) (3+0.5 nmole/min·mg protein) and theronine deaminase (60+5 nmole/min·mg protein) can be detected in the cell-free extracts of the minimal medium containing pyruvate. Further, smaller M+0 values in isoleucine (0.83–0.85) than in leucine (0.90–0.91) using [1-^{13}C]pyruvate (Table S1a and S1b) imply that some isoleucine molecules are derived from the threonine rather than pyruvate and acetyl-CoA. If *R. denitrificans* utilizes the citramalate-dependent pathway exclusively to generate isoleucine, the ^{13}C-labeling pattern in leucine and isoleucine will be identical (and the difference is expected to be within the error of GC-MS measurements: ~0.01), since the same precursors (pyruvate and acetyl-CoA) are employed in the biosynthesis of both amino acids. Therefore, both threonine- and citramalate-dependent pathways should be contributed to synthesize isoleucine, considering that the difference in ^{13}C-isotopomer abundance between leucine and isoleucine (0.06–0.07) is larger than the instrumental error (~0.01). Further, the calculated values of all ^{13}C-isotopomer abundance (from M+0 to M+5) of isoleucine indicate that 80% of the *in vivo* metabolic flux for isoleucine biosynthesis comes from the citramalte-dependent pathway, 20% from the threonine-dependent pathway, when [3-^{13}C]pyruvate is the carbon source (Table S1h and Figure S3). Together, our studies suggest that both the citramalate- and threonine-dependent pathways are active for isoleucine biosynthesis in *R. denitrificans*, similar to the report in *Geobacter sulfurreducens* [23]. As the putative *cimA* gene is not annotated in *R. denitrificans*, identification of the proposed citramalate synthase is in progress.

Discussion

CO_2 fixation in R. denitrificans and other AAPs

In this paper, we reported that *R. denitrificans* uses the anaplerotic pathways to fix 10–15% of protein carbon from CO_2. The amount of CO_2 fixation is lower in *R. denitrificans* than in autotrophic bacteria, supporting previous observations that *R. denitrificans* cannot grow autotrophically. Other than *R. denitrificans*, CO_2 fixation in several AAPs producing BChl *a* aerobically was reported previously by $^{14}CO_2$ incorporation studies: *Erythrobacter sibiricum* (a very low level of CO_2 (0.4%) assimilated) [7], *Erythrobacter longus* [7], *Acidiphilium rubrum* [20] and *Erythrobacter* sp. strains [15]. Some difference in CO_2 fixation can be detected between *R. denitrificans* and *A. rubrum*. In *A. rubrum*, CO_2 fixation is completely inhibited with the glucose-growth culture, and is enhanced 3- to 5-fold in the light (~3.0 nmoles/mg dry cell weight) [20]. In contrast, our studies suggest that CO_2 incorporation can be detected in *R. denitrificans* with either pyruvate- or glucose-growth cultures in either light or dark growth, and is not enhanced by light exposure.

Among the wide-spread marine *Roseobacter* clade, many of them cannot generate photosynthetic pigments. To the best of our knowledge, no CO_2 assimilation data have been reported for these unpigmented strains of the *Roseobacter* clade. It will be very useful to compare the difference in the CO_2 assimilation between *R. denitrificans* and un-pigmented *Roseobacter* clade bacteria.

Carbohydrate metabolism in R. denitrificans

The ED pathway has generally been viewed as an alternative to the EMP pathway. In examining the carbohydrate catabolism of both *Rhodobacter capsulatus* [25] and *Rhodobacter sphaeroides* [26] under anaerobically phototrophic and aerobically heterotrophic growth conditions, Conrad and Schlegel reported that fructose degradation was shifted to the ED pathway from the EMP pathway during the transition from anaerobically phototrophic to aerobically heterotrophic growth conditions, and that glucose catabolism was going through the ED pathway under both phototrophic and heterotrophic growth. In agreement with their experimental evidences, all of the genes in the EMP and ED pathways can be found in the genome of several *R. sphaeroides* strains, while the 6-phosphogluconate dehydrogenase gene (*pgd*) in the oxidative PP pathway cannot be annotated.

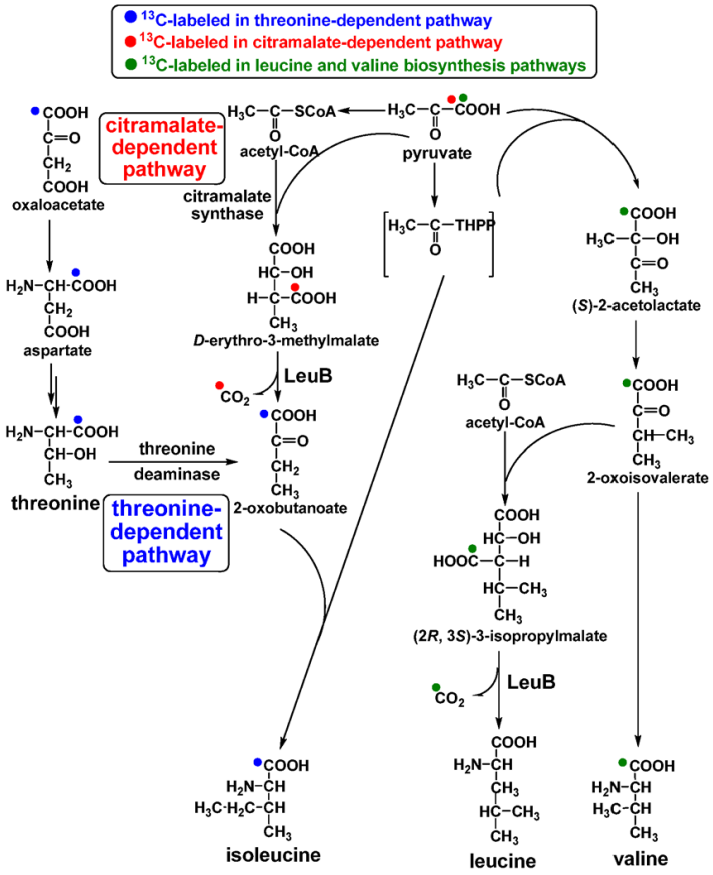


Figure 6. Proposed valine, leucine, and isoleucine biosynthesis pathways in *R. denitrificans*. Predicted ^{13}C-labeling distributions using [1-^{13}C]pyruvate in various pathways are shown. Experimentally identified ^{13}C-labeling patterns are reported in the context.
doi:10.1371/journal.pone.0007233.g006

Our studies in this report indicate that the ED pathway plays a significant role in glucose metabolism of *R. denitrificans*, and that the non-oxidative PP pathway contributes to histidine, ATP, coenzyme and nucleic acid biosynthesis, and that no metabolic flux going through the EMP pathway and the oxidative PP pathway. Our experimental evidences are consistent with the genomic information of *R. denitrificans*, in which the ED pathway and the non-oxidative PP pathway are complete, but the *pfk* gene, essential for the EMP pathway, and the *pgd* gene, required for the oxidative PP pathway, are not annotated.

Searching the genomes of the *Roseobacter* clade [5,8–10], all of them have genes in the ED pathway annotated. The ED pathway is identified mainly in prokaryotes, and almost all of them are aerobes and facultative bacteria, consistent with the physiological features of AAPs. Moreover, note that the *pgd* gene is also missing in *Silicibacter pomeroyi* DSS-3, *Silicibacter* sp. TM1040 and *Dinoroseobacter shibae* DFL12, and that the *pfk* gene is not found in *Jannaschia* sp. CCS1, *Silicibacter pomeroyi* DSS-3, and *Silicibacter* sp. TM1040. Alternatively, *Dinoroseobacter shibae* DFL12 has all of the genes in the EMP pathway, and *Jannaschia* sp. CCS1 has complete

oxidative PP pathway. It will be interesting to learn carbohydrate metabolisms in these two bacteria.

Possible physiological roles of the ED pathway in R. denitrificans

The physiological significance for *R. denitrificans* and possibly some other resourceful heterotrophic marine *Roseobacter* clade [12] to depend on the ED pathway, which produces 1 ATP, 1 NADH and 1 NADPH, compared to 2 ATP and 2 NADH from the more common EMP pathway, is an interesting question. The most straightforward answer could be that *R. denitrificans* demands more NADPH than ATP in the cells when the oxidative PP pathway is not active. The oxidative PP pathway, not the non-oxidative PP pathway, can generate two molecules of NADPH from one molecule of glucose, and can produce approximately 60% of the NADPH required inside the cells.

NADPH is not only the reducing power for synthesizing energy-rich molecules and biofuels, but also prevents oxidative stress by reducing glutathione via glutathione reductase, which converts reactive H_2O_2 into H_2O by glutathione peroxidase [27]. The reactive oxygen species inside the cell can be fatal for *R. denitrificans* since it is an aerobic anoxygenic bacterium that contains highly absorbing pigments that can sensitize reactive oxygen species formation. When the oxidative PP pathway is not active, the ED pathway may be the best option to generate NADPH required by the cells. It was not established whether an anoxygenic photosynthetic bacterium, like *R. denitrificans*, can produce glutathione, which can be synthesized by oxygenic phototrophs, such as cyanobacteria and higher plants. Note that genes for glutathione biosynthesis (glutamate-cysteine ligase, *gshA*, RD1_4077, glutathione synthetase, *gshB*, RD1_1192), glutathione reductase (*gor*, RD1_1919), glutathione peroxidase (*gpo*, RD1_0599), and some glutathione transferases can be found in *R. denitrificans* genome, suggesting possible roles of glutathione in reducing oxidative stress in *R. denitrificans*.

The other possibility for *R. denitrificans* adapting the ED pathway could be that many marine *Roseobacter*-lineage bacteria live in coastal seawater, an environment enriched with a wider variety of organic acids, and the ED pathway is a better option for digesting the hexuronic acids (e.g. glucuronic acid). This hypothesis may be supported by the broad substrate specificity of KDPG aldolase identified biochemically [28] and structurally [29]. As gene sequencing is currently in progress for 40 different *Roseobacter* strains [2], it will be interesting to learn if the patterns of carbohydrate metabolism revealed in this work may be a general theme in the *Roseobacter* clade, especially the significance of the ED/non-oxidative PP pathway in these physiologically diverse, widespread and high abundance marine microorganisms.

Gene expression under different growth conditions

Throughout the QRT-PCR data analysis, we observed that the transcript level for all of the genes tested is higher in cells grown in the minimal medium containing either pyruvate or glucose than in the rich medium (Figure 3). The discrepancies may be attributed to the following possible scenarios: (1) metabolic regulation: it has been thought in some bacteria, *E. coli* included, that the anaplerotic and metabolic pathways are not as much in demand in the rich medium as in the minimal medium containing defined carbon source, or some pathways may be shut down due to metabolic regulation [30,31], so genes in those metabolic pathways are not highly expressed in rich growth medium; and/or (2) the abundance of the 16S rRNA gene in different cell growth conditions: as reported in previous studies [32 34], the abundance of the 16S rRNA gene is dependent on the growth rate of the cells. If it is the case for *R. denitrificans*, lower copy numbers of the 16S rRNA gene will be expected in the minimal medium with glucose (grow slowest), and highest number in the rich medium grown under dark conditions (grow fastest). Together, both possible scenarios can explain lower transcript level for all of the target genes we tested in the rich medium. Further investigations are required for addressing this important issue.

Requirement of yeast extract or exogenous vitamin B_{12}

In this report, we demonstrate that vitamin B_{12} can successfully replace the undefined carbon source, yeast extract, for the growth of *R. denitrificans* in our developed minimal medium containing defined organic carbon source. While it is not known why the undefined carbon source yeast extract is required for *R. denitrificans* and several other photosynthetic bacteria, it is not clear why exogenous vitamin B_{12} is required for *R. denitrificans*, because all of the genes required for vitamin B_{12} biosynthesis are annotated in *R. denitrificans* (based on the information in the KEGG (Kyoto Encyclopedia of Genes and Genomes) database). It is also unclear how cobalamin can be utilized by *R. denitrificans*, since no known cobalamin transporter genes (*btuB*, *btuC*, *btuD* and *btuF*) can be recognized in the genome. Nonetheless, the effect of exogenous vitamin B_{12} is clear and its inclusion permitted most of the results reported in this paper. More experimental evidences are required for understanding the roles of yeast extract and exogenous vitamin B_{12} reported herein for enhancing cell growth.

A common feature in Roseobacter clade?

Our studies suggest approximately 10 15% CO_2 fixation in protein carbon through the anaplerotic enzymes by *R. denitrificans*, supporting the proposal by Swingley et al. [8]. We also confirmed that unlike carbon fixation by autotrophs, *R. denitrificans* cannot live with CO_2 as the sole carbon source, consistent with lack of genes for key enzymes of the Calvin cycle. Missing genes in the Calvin cycle are also common features in the other four *Roseobacter* clade bacteria with complete sequences [2,5,9], suggesting that carbon fixation mechanisms other than the Calvin cycle are needed for the AAPs.

Other than the ED pathway, our studies demonstrate that *R. denitrificans* uses the non-oxidative PP pathway as the other alternative pathway for sugar degradation to make the ribose-5-phosphate for histidine, cofactors, ATP and nucleic acid biosynthesis, as well as to produce GAP for carbon metabolism. In contrast to the oxidative PP pathway, no CO_2 is released in the non-oxidative PP pathway, which is an important feature for carbon fixation by autotrophs using the Calvin cycle. In these autotrophs, the non-oxidative PP pathway is part of the Calvin cycle. Although *R. denitrificans* cannot use the Calvin cycle for fixing CO_2, it does have a complete non-oxidative PP pathway and relies on the pathway for converting hexose phosphate into pentose phosphate.

Materials and Methods

Materials

The DNA oligomers in this report are from Integrated DNA Technology (IDT) and were used without further purification. The ^{13}C-labeled glucose, pyruvate and sodium bicarbonate were purchased from Cambridge Isotope Laboratories (CIL), Inc (MA, USA).

The growth of bacterial strains

R. denitrificans OCh114 (a gift from the laboratory of Dr. Beatty at University of British Columbia, Vancouver, Canada) was grown aerobically (20 30%-filled Erlenmeyer flasks) on either a nutrition-

rich medium (Difco™ Marine Broth 2216; Becton, Dickson and Company) (pH 7.5) or a reported minimal medium [13] with modifications (illustrated in next paragraph). No undefined carbon sources (yeast extract) was included in our minimal growth medium. The seawater source in our minimal medium was prepared by 3.2% Instant Ocean® artificial sea salts (United Pet Group, Cincinnati, OH). The minimal growth medium was adjusted to pH 7.5. All of the cultures were grown at 28°C and the growth rates were estimated by OD_{600}. We used 2% cultures (50-fold dilution) in the late exponential growth phase grown in the rich medium to inoculate the minimal medium containing defined carbon source(s). To minimize the carryover of the rich medium, inoculated cells were centrifuged and supernatant liquid was removed. Cell pellets were washed with the minimal medium (no carbon source included) twice, resuspended in the minimal medium, and then transferred to the minimal medium containing defined carbon source(s).

In the minimal medium supplied with defined ^{13}C-isotopically labeled carbon source, 0.1% pyruvate (labeled on either the first or third carbon position), 0.1% glucose (on the first, sixth or uniform labeled), or 0.2% $NaH^{13}CO_3$ were added. Due to the requirement of a growth medium without including undefined carbon sources for GC/MS studies, ferric citrate [13] was replaced by ferric chloride hexahydrate, and yeast extract was substituted by 20 μg/L vitamin B_{12}. The cultures were grown to an exponential phase, representing the pseudo steady-state, and then harvested for analyzing the isotopomer distribution in key metabolites.

RNA extraction and quantitative real-time PCR (QRT-PCR)

The methods used to extract RNA and perform QRT-PCR were described previously [35]. QRT-PCR was performed to profile the gene expression under different growth conditions of *R. denitrificans* OCh114. cDNA was synthesized from 1 μg RNA and 100 μM random 9mer DNA using Superscript III reverse transcriptase (Invitrogen). The QRT-PCR reactions were performed via ABI 7500 Real-Time PCR System (Applied Biosystems). The primers (shown in Table S2) for QRT-PCR reactions were designed using Primer Express 2.0 software (PE Applied Biosystems) and analyzed by OligoAnalyzer 3.0 software (IDT). The Power SYBR Green Master Mix (PE Applied Biosystem) was used for amplifying DNA [35]. The cycle threshold (Ct) was calculated as the cycle number at which a fluorescence threshold (ΔRn) crossed the baseline. Data were normalized by analyzing ΔCt = Ct of the target gene Ct of the internal control gene (16S rRNA), and each relative gene expression value is calculated with $2^{-(\Delta Ct)}$. Three biological replicates, with six technical replicates for each biological sample, and total eighteen technical replicates were performed for validation, and the mean value was reported (Figure 3). The amplified DNA fragments were verified by 1% agarose gel electrophoresis, and a single fragment was obtained for all amplicons (data not shown).

Measurements of pyruvate/glucose uptake

The amount of D-glucose in solution was determined by hexokinase/glucose-6-phosphate dehydrogenase coupling assay with a D-glucose assay kit (Roche, Mannheim, Germany), and the reduction of NAD^+ was followed spectrophotometrically by the increase at 340 nm in absorbance. The amount of pyruvate in the culture was estimated by lactate dehydrogenase (EC 1.1.2.3) (Sigma-Aldrich) with 25 μM NADH in 0.1 M Tris-HCl buffer at pH 8.0 [36]. The reaction was followed by the oxidation of NADH (the decrease at 340 nm in absorbance).

Enzyme assays

Enzymatic activities were performed with cell-free crude extracts prepared as follows. Cells were harvested by centrifugation at 5,000×*g* for 15 min at 4°C and washed with 20 mM Tris-HCl buffer at pH 8.0. Cell pellet was resuspended in the same buffer containing 1 mM phenylmethanesulfonyl fluoride (PMSF). Resuspended cells were disrupted by sonication, and cell debris was removed with centrifugation at 20,000×*g* for 30 min. Protein concentration in cell extracts was determined by the Bradford assay [37] using bovine serum albumin as a standard. The activity of pyruvate carboxylase, phosphoenolpyruvate carboxylase, phosphoenolpyruvate carboxykinase, malic enzyme, citramalate synthase, 6-phosphofructokinase, 2-keto-3-deoxy-6-phosphogluconate aldolase, 6-phosphogluconate dehydrase, isopropylmalate synthase and threonine deaminase in cell extracts were assayed as described previously [22,23,38 46].

Gas chromatography (GC)/mass spectrometry (MS)

The methods for GC/MS are reported previously [44,47 49]. In brief, biomass was collected from 50 mL culture by centrifuge and then was hydrolyzed in 6 M HCl at 100°C. The resulting amino acids were dried and derivatized in tetrahydrofuran and N-(*tert*-butyl dimethylsilyl)-N-methyl-trifluoroacetamide (Sigma-Aldrich) at 70°C for 1 h. A gas chromatograph (GC) (Hewlett-Packard, model 7890A, Agilent Technologies Inc., Ballwin, MO) equipped with a DB5-MS column (J&W scientific, Folsom, CA) and a mass spectrometer (5975C, Agilent Technologies Inc.) was used for analyzing metabolite labeling profiles. We report four types of MS fragment data: (1) the $[M\text{-}15]^+$ group, which loses a methyl group; (2) the $[M\text{-}57]^+$ group, containing un-fragmented amino acids; (3) the $[M\text{-}159]^+$ group, which contains amino acids losing α carboxyl group; and (4) the $[M\text{-}302]^+$ group, containing only the 1st (α carboxyl group) and 2nd carbons in the amino acid backbone (Note that $[M\text{-}302]^+$ and $[M\text{-}15]^+$ cannot be detected in some amino acids). Three labeling data were used to trace the carbon metabolic route and to identify active pathways. Throughout the article, M+0, M+1, M+2… stands for the value of unlabeled (M+0) and of additional ^{13}C-labeled protein-derived amino acids (M+1, M+2….).

Supporting Information

Figure S1 Spectra, culture, and growth curves of R. denitrificans under different growth conditions. The culture of R. denitrificans grown in the rich medium under the day-light cycles (a). The spectrum (b) and growth curve (c) of the cells grown in the rich medium under dark versus under light. The spectrum of the cells grown under dark and day-light cycle is shown in the inset. The formation of the light harvesting antenna complex II antenna complex (807-nm peak) is repressed in the light, while the level of the reaction center-light harvesting complex I (RC-LHI) core complex (872-nm peak) is similar. Data fit of pyruvate uptake with or without NaHCO3 in the defined medium, and the uptake rate is 2.5×10−2±5×10−4 mmole per hour (d).
Found at: doi:10.1371/journal.pone.0007233.s001 (1.12 MB TIF)

Figure S2 The non-oxidative pentose phosphate pathway. All of the reaction steps are reversible. Possible 13C-labeling using [1-13C]glucose or [6-13C]glucose is shown in green or red dots, respectively.
Found at: doi:10.1371/journal.pone.0007233.s002 (0.47 MB TIF)

Figure S3 Experimental values of the isotopomer abundance of isoleucine (red bar) using [3-13C]pyruvate as the defined carbon source, and predicted contributions of the threonine (back bar) and citramalate (blue bar) pathways for isoleucine biosynthesis in R. denitrificans.
Found at: doi:10.1371/journal.pone.0007233.s003 (0.62 MB TIF)

Table S1 13C-isotopomer abundances of tert-butyl dimethylsilyl (TBDMS) derivatives of protein-derived amino acids from R. denitrificans OCh114 grown in the defined medium containing different carbon sources. Except in (b), all of the cultures were grown in the dark. The unlabeled molecule is shown as M+0, and additional mass comes from 13C-labeled carbon source. All of the data, except leucine and isoleucine, are [M-57]+ data (tert-butyl group is removed). The [M-15]+ data (methyl group is cleaved) for leucine and isoleucine are shown, because the [M-57]+ data are overlapped in GC-MS.
Found at: doi:10.1371/journal.pone.0007233.s004 (0.27 MB DOC)

Table S2 Primers used for QRT-PCR studies reported in this paper
Found at: doi:10.1371/journal.pone.0007233.s005 (0.04 MB DOC)

Table S3 The theoretical and experimental M+0 values in alanine and serine using D-[1-13C] glucose and D-[6-13C]glucose[a,b]
Found at: doi:10.1371/journal.pone.0007233.s006 (0.03 MB DOC)

Acknowledgments

We appreciate the positive comments and helpful suggestions from the editor and the expert reviewers for improving the manuscript.

Author Contributions

Conceived and designed the experiments: KHT YT REB. Performed the experiments: KHT XF. Analyzed the data: KHT YT. Contributed reagents/materials/analysis tools: KHT XF. Wrote the paper: KHT REB.

References

1. Beja O, Suzuki MT, Heidelberg JF, Nelson WC, Preston CM, et al. (2002) Unsuspected diversity among marine aerobic anoxygenic phototrophs. Nature 415: 630 633.
2. Brinkhoff T, Giebel HA, Simon M (2008) Diversity, ecology, and genomics of the *Roseobacter* clade: a short overview. Arch Microbiol 189: 531 539.
3. Wagner-Dobler I, Biebl H (2006) Environmental biology of the marine Roseobacter lineage. Annu Rev Microbiol 60: 255 280.
4. Selje N, Simon M, Brinkhoff T (2004) A newly discovered Roseobacter cluster in temperate and polar oceans. Nature 427: 445 448.
5. Moran MA, Buchan A, Gonzalez JM, Heidelberg JF, Whitman WB, et al. (2004) Genome sequence of Silicibacter pomeroyi reveals adaptations to the marine environment. Nature 432: 910 913.
6. Buchan A, Gonzalez JM, Moran MA (2005) Overview of the marine *Roseobacter* lineage. Appl Environ Microbiol 71: 5665 5677.
7. Yurkov VV, Beatty JT (1998) Aerobic anoxygenic phototrophic bacteria. Microbiol Mol Biol Rev 62: 695 724.
8. Swingley WD, Sadekar S, Mastrian SD, Matthies HJ, Hao J, et al. (2007) The complete genome sequence of Roseobacter denitrificans reveals a mixotrophic rather than photosynthetic metabolism. J Bacteriol 189: 683 690.
9. Moran MA, Belas R, Schell MA, Gonzalez JM, Sun F, et al. (2007) Ecological genomics of marine Roseobacters. Appl Environ Microbiol 73: 4559 4569.
10. Biebl H, Allgaier M, Tindall BJ, Koblizek M, Lunsdorf H, et al. (2005) Dinoroseobacter shibae gen. nov., sp. nov., a new aerobic phototrophic bacterium isolated from dinoflagellates. Int J Syst Evol Microbiol 55: 1089 1096.
11. Kolber ZS, Plumley FG, Lang AS, Beatty JT, Blankenship RE, et al. (2001) Contribution of aerobic photoheterotrophic bacteria to the carbon cycle in the ocean. Science 292: 2492 2495.
12. Moran MA, Miller WL (2007) Resourceful heterotrophs make the most of light in the coastal ocean. Nat Rev Microbiol 5: 792 800.
13. Shiba T (1991) Roseobacter litoralis gen. nov., sp. nov. and Roseobacter denitrificans sp. nov., aerobic pink-pigmented bacteria which contain bacteriochlorophyll a. Syst Appl Microbial 14: 140 145.
14. Shiba T, Simidu U, Taga N (1979) Distribution of aerobic bacteria which contain bacteriochlorophyll a. Appl Environ Microbiol 38: 43 45.
15. Koblizek M, Beja O, Bidigare RR, Christensen S, Benitez-Nelson B, et al. (2003) Isolation and characterization of Erythrobacter sp. strains from the upper ocean. Arch Microbiol 180: 327 338.
16. Kortluke C, Breese K, Gad'on N, Labahn A, Drews G (1997) Structure of the puf operon of the obligately aerobic, bacteriochlorophyll alpha-containing bacterium Roseobacter denitrificans OCh114 and its expression in a Rhodobacter capsulatus puf puc deletion mutant. J Bacteriol 179: 5247 5258.
17. Biebl H, Allgaier M, Lunsdorf H, Pukall R, Tindall BJ, et al. (2005) *Roseovarius* mucosus sp. nov., a member of the *Roseobacter* clade with trace amounts of bacteriochlorophyll a. Int J Syst Evol Microbiol 55: 2377 2383.
18. Entner N, Doudoroff M (1952) Glucose and gluconic acid oxidation of Pseudomonas saccharophila. J Biol Chem 196: 853 862.
19. Shiba T, Harashima K (1986) Aerobic photosynthetic bacteria. Microbiol Sci 3: 376 378.
20. Kishimoto N, Fukaya F, Inagaki K, Sugio T, Tanaka H, et al. (1995) Distribution of bacteriochlorophyll a among aerobic and acidophilic bacteria and light-enhanced CO2-incorporation in Acidiphilium rubrum. FEMS Microbiol Ecol 16: 291 296.
21. Umbarger HE (1978) Amino acid biosynthesis and its regulation. Annu Rev Biochem 47: 532 606.
22. Howell DM, Xu H, White RH (1999) (R)-citramalate synthase in methanogenic archaea. J Bacteriol 181: 331 333.
23. Risso C, Van Dien SJ, Orloff A, Lovley DR, Coppi MV (2008) Elucidation of an alternate isoleucine biosynthesis pathway in Geobacter sulfurreducens. J Bacteriol 190: 2266 2274.
24. Xu H, Zhang Y, Guo X, Ren S, Staempfli AA, et al. (2004) Isoleucine biosynthesis in Leptospira interrogans serotype lai strain 56601 proceeds via a threonine-independent pathway. J Bacteriol 186: 5400 5409.
25. Conrad R, Schlegel HG (1977) Different degradation pathways for glucose and fructose in Rhodopseudomonas capsulata. Arch Microbiol 112: 39 48.
26. Conrad R, Schlegel HG (1977) Influence of aerobic and phototrophic growth conditions on the distribution of glucose and fructose carbon into the Entner-Doudoroff and Embden-Meyerhof pathways in Rhodopseudomonas sphaeroides. J Gen Microbiol 101: 277 290.
27. Kruger NJ, von Schaewen A (2003) The oxidative pentose phosphate pathway: structure and organisation. Curr Opin Plant Biol 6: 236 246.
28. Griffiths JS, Wymer NJ, Njolito E, Niranjanakumari S, Fierke CA, et al. (2002) Cloning, isolation and characterization of the Thermotoga maritima KDPG aldolase. Bioorg Med Chem 10: 545 550.
29. Allard J, Grochulski P, Sygusch J (2001) Covalent intermediate trapped in 2-keto-3-deoxy-6-phosphogluconate (KDPG) aldolase structure at 1.95-A resolution. Proc Natl Acad Sci U S A 98: 3679 3684.
30. Lin H, San KY, Bennett GN (2005) Effect of Sorghum vulgare phosphoenolpyruvate carboxylase and Lactococcus lactis pyruvate carboxylase coexpression on succinate production in mutant strains of Escherichia coli. Appl Microbiol Biotechnol 67: 515 523.
31. Motter AE, Gulbahce N, Almaas E, Barabasi AL (2008) Predicting synthetic rescues in metabolic networks. Mol Syst Biol 4: 168.
32. Holmes DE, Nevin KP, Lovley DR (2004) Comparison of 16S rRNA, nifD, recA, gyrB, rpoB and fusA genes within the family Geobacteraceae fam. nov. Int J Syst Evol Microbiol 54: 1591 1599.
33. Case RJ, Boucher Y, Dahllof I, Holmstrom C, Doolittle WF, et al. (2007) Use of 16S rRNA and rpoB genes as molecular markers for microbial ecology studies. Appl Environ Microbiol 73: 278 288.
34. Pavlekovic M, Schmid MC, Schmider-Poignee N, Spring S, Pilhofer M, et al. (2009) Optimization of three FISH procedures for in situ detection of anaerobic ammonium oxidizing bacteria in biological wastewater treatment. J Microbiol Methods 78: 119 126.
35. Tang KH, Wen J, Li X, Blankenship RE (2009) Role of the AcsF protein in Chloroflexus aurantiacus. J Bacteriol 191: 3580 3587.
36. Thompson JS, Richardson KE (1968) Determination of pyruvate in enzyme-catalyzed reactions in the presence of glyoxylate. Anal Biochem 24: 197 201.
37. Bradford MM (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.
38. Guillouet S, Rodal AA, An G, Lessard PA, Sinskey AJ (1999) Expression of the Escherichia coli catabolic threonine dehydratase in Corynebacterium glutamicum and its effect on isoleucine production. Appl Environ Microbiol 65: 3100 3107.
39. Mukhopadhyay B, Stoddard SF, Wolfe RS (1998) Purification, regulation, and molecular and biochemical characterization of pyruvate carboxylase from Methanobacterium thermoautotrophicum strain deltaH. J Biol Chem 273: 5155 5166.
40. Ray TB, Black CC (1976) Characterization of Phosphoenolpyruvate Carboxykinase from Panicum maximum. Plant Physiol 58: 603 607.
41. Meyer CR, Rustin P, Wedding RT (1988) A simple and accurate spectrophotometric assay for phosphoenolpyruvate carboxylase activity. Plant Physiol 86: 325 328.
42. Van Schaftingen E, Jett MF, Hue L, Hers HG (1981) Control of liver 6-phosphofructokinase by fructose 2,6-bisphosphate and other effectors. Proc Natl Acad Sci U S A 78: 3483 3486.

43. Hatfield GW, Umbarger HE (1970) Threonine deaminase from Bacillus subtilis. II. The steady state kinetic properties. J Biol Chem 245: 1742 1747.
44. Tang YJ, Sapra R, Joyner D, Hazen TC, Myers S, et al. (2009) Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant Geobacillus strain. Biotechnol Bioeng 102: 1377 1386.
45. Meloche HP, Wood WA (1964) The mechanism of 6-phosphogluconic dehydrase. J Biol Chem 239: 3505 3510.
46. Meloche HP, Wood WA (1964) The mechanism of 2-keto-3-deoxy-6-phosphogluconic aldolase. J Biol Chem 239: 3511 3514.
47. Tang Y, Pingitore F, Mukhopadhyay A, Phan R, Hazen TC, et al. (2007) Pathway confirmation and flux analysis of central metabolic pathways in Desulfovibrio vulgaris hildenborough using gas chromatography-mass spectrometry and Fourier transform-ion cyclotron resonance mass spectrometry. J Bacteriol 189: 940 949.
48. Pingitore F, Tang Y, Kruppa GH, Keasling JD (2007) Analysis of amino acid isotopomers using FT-ICR MS. Anal Chem 79: 2483 2490.
49. Wahl SA, Dauner M, Wiechert W (2004) New tools for mass isotopomer data evaluation in (13)C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnol Bioeng 85: 259 268.

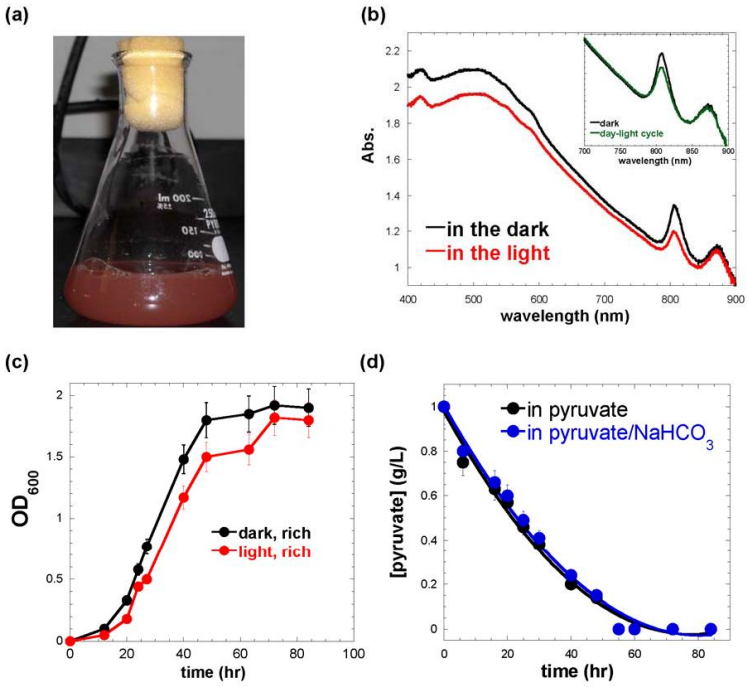

(a)
(b)
Abs.
wavelength (nm)
dark
day-light cycle
in the dark
in the light
(c)
OD600
time (hr)
dark, rich
light, rich
(d)
[pyruvate] (g/L)
time (hr)
in pyruvate
in pyruvate/NaHCO3

Figure S1

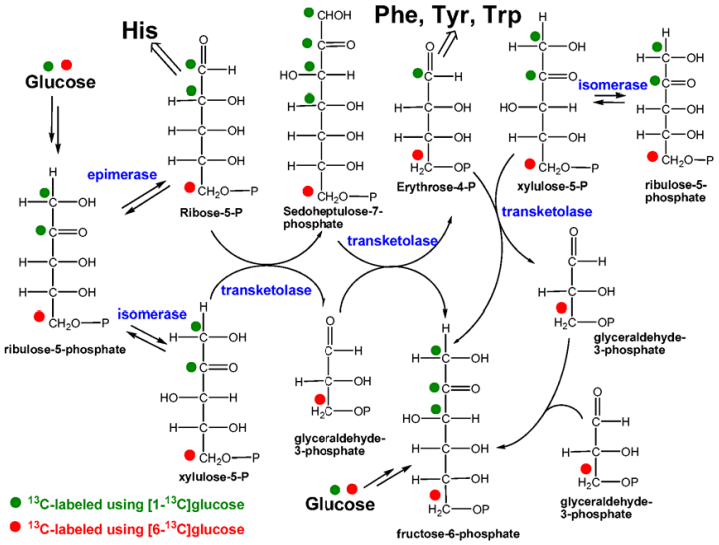
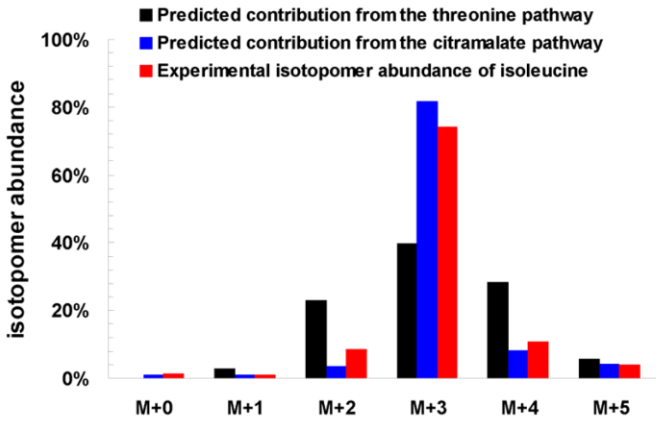

His
Phe, Tyr, Trp
Glucose
epimerase
Ribose-5-P
Sedoheptulose-7-phosphate
Erythrose-4-P
xylulose-5-P
isomerase
ribulose-5-phosphate
transketolase
transketolase
transketolase
isomerase
ribulose-5-phosphate
xylulose-5-P
glyceraldehyde-3-phosphate
glyceraldehyde-3-phosphate
glyceraldehyde-3-phosphate
Glucose
fructose-6-phosphate
13C-labeled using [1-13C]glucose
13C-labeled using [6-13C]glucose

Figure S2



Isoleucine Mass Isotopomer[a]	Experimental labeling pattern[b]	Prediction[c]		
		Threonine-dependent pathway	Citramalate-dependent pathway	Both Pathway
M+0	0.0133	0.0007	0.0115	0.0094
M+1	0.0100	0.0298	0.0091	0.0130
M+2	0.0868	0.2301	0.0376	0.0742
M+3	0.7411	0.3995	0.8160	0.7368
M+4	0.1077	0.2831	0.0822	0.1204
M+5	0.0412	0.0569	0.0436	0.0461
Total Error[d]	0.0197			

[a] M+0, M+1, M+2…: the value of unlabeled and ^{13}C-labeled isoleucine.
[b] Experimental labeling pattern shows the isotopomer distribution of isoleucine that lost α-carboxyl group (M-159).
[c] Predicted values were calculated using the optimization routine with the assumption that each pathway functioned separately, and an optimization routine was used to select the best-fitting ratio between them.
[d] Errors are the sum of squared differences between predicted and experimental values for isoleucine labeling pattern.

Flux calculation for citramalate pathway. Mass distribution vectors (MDVs) contain mole fractions for groups of isotopomers for isoleucine, leucine and threonine were constructed first. Simulated annealing algorithm was applied to search for the ratio of the citramalate dependent pathway: the threonine dependent pathway. The initial temperature was chosen to be 100°C and the optimizations were run until the routine converged on a minimum of the option tolerance. The objective value was the sum-of-squares difference between the measured and calculated MDV values for isoleucine. Several routines were repeated in order to avoid the local minimum. All the isotopomer optimization routines were coded in MATLAB (The Mathworks, Natick, MA).

Figure S3

Table S1

(a) 0.1% [1-^{13}C]pyruvate, dark

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.2787	0.7838	0.2475	0.9146	0.8518	0.7598	0.6679	0.7014	0.7241	0.8398	0.6804	0.5188	0.5166
M+1	0.7185	0.2178	0.7289	0.0700	0.1008	0.2322	0.2203	0.1846	0.1757	0.1577	0.2899	0.3119	0.3131
M+2	0.0053	0	0.0247	0.0009	0.0327	0.0092	0.1073	0.1152	0.1014	0.001	0.0218	0.1399	0.156
M+3	0		0	0.0143	0.0034	0	0.005	0.0047	0	0.0017	0.0071	0.0292	0.0044
M+4			0.0008	0.0050	0		0.0053	0	0	0.0001	0.003	0.0002	0.012
M+5			0.0003	0.0029	0.0005		0			0	0	0	0.0012
M+6				0	0.0004						0	0.0007	0
M+7												0	0.0004
M+8												0	0.0003
M+9												0	0

(b) 0.1% [1-^{13}C]pyruvate, light

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.2891	0.8038	0.2604	0.8997	0.8295	0.8032	0.711	0.7448	0.7616	0.8445	0.7204	0.547	0.5421
M+1	0.6954	0.1953	0.7073	0.0559	0.1342	0.2	0.1964	0.1719	0.169	0.1481	0.2527	0.3333	0.3314
M+2	0.0146	0.0008	0.033	0.0108	0.0376	0	0.1846	0.0816	0.0667	0.0076	0.0286	0.1051	0.1152
M+3	0.0009		0	0.0326	0.0068	0	0.0059	0	0.0023	0	0	0.0141	0.0101
M+4			0	0.0060	0		0.0009	0.0036	0.0003	0	0	0.0001	0
M+5			0	0	0.0004		0.0012			0.0002	0	0.0009	0.001
M+6				0.0001	0.0005						0	0	0.0008
M+7												0	0
M+8												0	0
M+9												0	0

(c) 0.1% [1-^{13}C]pyruvate and 0.2% $NaHCO_3$

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.2542	0.8021	0.073	0.9364	0.9042	0.7971	0.7471	0.7713	0.7903	0.8972	0.7348	0.5095	0.5378
M+1	0.7334	0.1972	0.8895	0.0484	0.0843	0.2038	0.1993	0.1807	0.1651	0.1022	0.2354	0.3576	0.3537
M+2	0.0128	0.0007	0.0426	0.0215	0.0135	0	0.0558	0.038	0.0412	0	0.0333	0.1155	0.0967
M+3	0		0	0	0.0026	0	0	0.01	0.002	0	0	0.0198	0.0121
M+4			0	0	0		0	0	0.0014	0	0	0	0
M+5			0	0.0006	0.0004		0			0	0	0	0
M+6				0.0002	0.0003						0	0	0
M+7												0	0
M+8												0	0
M+9												0	0

(d) 0.2% ^{13}C-$NaHCO_3$ and 0.1% unlabeled pyruvate

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.9427	0.9473	0.951	0.8878	1.000	0.9189	0.9149	0.8633	0.8683	0.8864	0.9517	0.854	0.8383
M+1	0.05	0.0554	0.0478	0.0638	0	0.0762	0.0742	0.1379	0.1328	0.1186	0.0023	0.1382	0.1194
M+2	0.005	0	0.0074	0.0363	0	0.0029	0.0225	0	0	0	0.109	0.0111	0.06
M+3	0.0024		0	0.0237	0	0.002	0	0	0	0	0	0.0006	0
M+4			0	0	0		0	0	0	0	0	0	0
M+5			0	0	0		0			0	0	0.0003	0
M+6				0.008	0						0	0.0003	0
M+7												0	0
M+8												0	0
M+9												0	0

(e) 0.1% D-[1-^{13}C]glucose

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.5629	0.8943	0.4146	0.4191	0.4677	0.7369	0.4889	0.4666	0.5764	0.3947	0.2788	0.4793	0.5286
M+1	0.3994	0.0983	0.4438	0.3457	0.3002	0.2437	0.2883	0.3756	0.3139	0.3754	0.5393	0.3197	0.2929
M+2	0.0365	0.0074	0.1305	0.1949	0.1813	0.02	0.1674	0.1448	0.0978	0.1858	0.1464	0.1524	0.1472
M+3	0.0012		0.0086	0.0329	0.0521	0	0.0276	0.0128	0.012	0.0395	0.0347	0.039	0.0382
M+4			0.0026	0.0074	0.0087		0.0145	0.0003	0	0.005	0	0.0083	0.0017
M+5			0	0.0022	0		0.0134			0	0	0.0015	0.0007
M+6				0	0.0003						0	0	0.0007
M+7												0	0
M+8												0	0
M+9												0	0

(f) 0.1% D-[1-^{13}C]glucose and 0.2% $NaHCO_3$

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.5833	0.9159	0.4607	0.5373	0.5329	0.7873	0.5698	0.5782	0.6266	0.4452	N/A[a]	0.5917	0.5887
M+1	0.3894	0.0784	0.4101	0.2586	0.2925	0.2048	0.2958	0.3172	0.2831	0.3432		0.2517	0.2734
M+2	0.029	0.0057	0.1192	0.1439	0.1042	0.0154	0.1261	0.0847	0.0831	0.1728		0.1227	0.0909
M+3	0		0.0068	0.0683	0.0739	0	0.01	0.02	0.0051	0.0335		0.0295	0.0431
M+4			0.0036	0.0007	0.0069		0	0	0.0022	0.0053		0.0034	0.0084
M+5			0	0	0		0			0		0	0
M+6				0.0007	0.0005							0.001	0
M+7												0	0
M+8												0	0
M+9												0	0

[a] Peaks overlapped in GC/MS.

(g) 0.2% ^{13}C-$NaHCO_3$ and 0.1% unlabeled D-glucose

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.9696	0.9723	0.9496	0.9183	0.9863	0.9722	0.8411	0.9373	0.9326	0.9449	0.921	0.9155	0.8951
M+1	0.0321	0.0264	0.0546	0.0793	0.0343	0.0323	0.154	0.0511	0.0671	0.0545	0.0743	0.0826	0.1024
M+2	0	0.0013	0	0	0	0	0.0551	0.0209	0.0016	0	0	0.0039	0
M+3	0		0	0.0266	0	0	0	0.0039	0	0	0.0033	0	0.0057
M+4			0	0	0.0013		0	0	0.0008	0	0	0	0.004
M+5			0	0	0.0009		0			0.0002	0	0	0
M+6				0.0005	0						0.0004	0.0001	0
M+7												0	0.0001
M+8												0	0.0001
M+9												0	0

(h) [3-^{13}C]pyruvate

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.0171	0.1988	0.0146	0	0	0.0208	0	0.0149	0.0126	0.0109	0.0098	0.0107	0.007
M+1	0.7182	0.5877	0.0207	0	0	0.3075	0	0.2059	0.1813	0.0154	0.0268	0.0026	0.0032
M+2	0.1695	0.2135	0.821	0.0981	0.0872	0.4777	0.1773	0.4048	0.3769	0.2857	0.082	0.0079	0.0112
M+3	0.0952		0.0773	0.7987	0.7212	0.194	0.4638	0.3738	0.4103	0.4502	0.2275	0.0635	0.0592
M+4			0.0496	0.0585	0.1201		0.3075	0.0006	0.0189	0.2261	0.3265	0.1728	0.1626
M+5			0.0169	0.0471	0.0674		0.0514			0.0118	0.2523	0.2604	0.2582
M+6				0	0.0041						0.0759	0.2437	0.2667
M+7												0.155	0.1449
M+8												0.0701	0.0831

M+9												0.0132	0.0039

(i) D-[6-^{13}C]glucose

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.4789	0.769	0.1835	0.1100	0.1340	0.2987	0.175	0.3021	0.2203	0.1371	0.0975	0.037	0.0368
M+1	0.4605	0.2022	0.4193	0.3021	0.3788	0.6092	0.3827	0.4597	0.4419	0.358	0.4136	0.1868	0.1971
M+2	0.0514	0.0288	0.3092	0.3988	0.3489	0.0783	0.3077	0.1964	0.2679	0.353	0.4056	0.3678	0.3712
M+3	0.0092		0.0682	0.1674	0.1098	0.0137	0.1245	0.0445	0.0655	0.1302	0.0543	0.3033	0.3004
M+4			0.0166	0.0124	0.0276		0.0159	0	0.0044	0.0206	0.0246	0.0757	0.0644
M+5			0.0031	0.0071	0.0078		0			0.0011	0.005	0.0224	0.0233
M+6				0.0023	0						0	0.006	0.0051
M+7												0.001	0.0009
M+8												0	0.0001
M+9												0	0.0006

(j) 25% D-[U-$^{13}C_6$]glucose/75% unlabeled D-glucose

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.7301	0.7797	0.5643	0.4838	0.4563	0.7094	0.4734	0.5455	0.5712	0.4659	0.4779	0.4444	0.4198
M+1	0.1056	0.0935	0.1326	0.1426	0.1960	0.1206	0.2742	0.2223	0.2324	0.261	0.237	0.1446	0.1394
M+2	0.049	0.1268	0.1545	0.2605	0.2140	0.0727	0.1545	0.1538	0.1328	0.194	0.0952	0.1205	0.1293
M+3	0.1153		0.1118	0.0512	0.0822	0.0974	0.0736	0.0626	0.055	0.0603	0.093	0.1412	0.1475
M+4			0.0166	0.0547	0.0423		0.022	0.0158	0.0086	0.0165	0.0541	0.0771	0.0849
M+5			0.0201	0.0044	0.0112		0.0022			0.0024	0.035	0.0347	0.0399
M+6				0.0028	0						0.0077	0.0234	0.0216
M+7												0.0101	0.0115
M+8												0.0022	0.0031
M+9												0.0017	0.0029

(k) 100% D-[U-$^{13}C_6$]glucose

	Ala	Gly	Val	Leu	Ile	Ser	Met	Thr	Asp	Glu	His	Phe	Tyr
M+0	0.0255	0.0673	0.0434	0.0440	0.0472	0.033	0.049	0.0347	0.0454	0.0173	0.0301	0.0739	0.0449
M+1	0.0048	0.0353	0.004	0.0041	0.0023	0.0099	0.016	0.0031	0.0037	0.0018	0.0042	0.0069	0.0062
M+2	0.0433	0.8974	0.0035	0.0114	0.0144	0.0458	0.0047	0.0078	0.0126	0.0036	0.0015	0.0004	0.0012
M+3	0.9265		0.0044	0.0040	0	0.9113	0.0159	0.1045	0.1071	0.0125	0.0023	0.0003	0.0008
M+4			0.0668	0.0130	0.0160		0.1322	0.8499	0.8311	0.1185	0.0069	0.0011	0.0024
M+5			0.878	0.1413	0.1429		0.7821			0.8463	0.0885	0.002	0.0014
M+6				0.7822	0.7814						0.8665	0.0014	0.0035
M+7												0.0085	0.0105
M+8												0.1024	0.1086
M+9												0.803	0.8206

Table S2

Gene name (gene location)	Proteins	Forward primers	Reverse primers
16S rRNA		TGTTCGGAATTACTGGGCG	TCGGGATTTCACCCCTAACTT
pyc (RD1_3376)	pyruvate carboxylase	CCTTGGGCTTGCGGATC	CATCTGGTTCACCTCGGCA
pckA (RD1_1376)	PEP carboxykinase	ATGCCATGCCAGTTCGGTA	CGGTGGTTCGGACCCC
ppc (RD1_4248)	PEP carboxylase	GGCTCGATATTGCGGAACTC	ATCCTCTGCCATCGCCC
tme (RD1_0421)	malic enzyme	(1)ACCCCCGGAAAGTTCGAG (2)CGCCTATGATTACACCAACAAGG	(1)AAGACTGAGGTCCCGCTGC (2)CCCAAGACCGCTGTCCC
ppk (RD1_1948)	pyruvate phosphate dikinase	CAGGGTGTCTTTCCCGAAGA	AATCTGCAAAAGTTCCCCCA
aatA (RD1_3892)	aspirate aminotrans-ferase	TCGAGCGGGCATAGGAAA	GGCTTGATTTTGACCCGGT
eda (RD1_2878)	KDPG aldolase	CCAGAAGTGGTAATTCCAGCG	TTCACCCGGCGCGAC
edd (RD1_2879)	phosphogluconate dehydrase	CGCACGGTGCTTTTTTCG	GTTCCTGCCAGCGGGTC
ilvA (RD1_0416)	threonine deaminase	CACATGAAAGCGACGCCC	GCGTGATGTTTTTGTCTGCG

Table S3

	Theoretical $(M+0)^{Ala}$ value	**Experimental $(M+0)^{Ala}$ value**	**Theoretical $(M+0)^{ser}$ value**	**Experimental $(M+0)^{Ser}$ value**
D-[1-^{13}C]glucose	0.5	0.56	1.0	0.74
D-[6-^{13}C]glucose	0.5	0.48	0	0.30

p.s.

[a] $(M+0)^{Ala}$, M+0 value for alanine; $(M+0)^{Ser}$, M+0 value for serine

[b] The theoretical $(M+0)^{Ala}$ and $(M+0)^{Ser}$ values were estimated by the ED pathway.

Estimate the contribution of the PP pathway for carbohydrate metabolism. The difference between the theoretical and experimental (M+0) values in serine can be considered the contribution of the non-oxidative PP pathway, and approximately 25-30% GAP is estimated to be produced by the non-oxidative PP pathway. The difference between the theoretical and experimental M+0 values in alanine, 2-6%, can be considered as the experimental error.

Appendix 5

Selective Utilization of Exogenous Amino Acids by *Dehalococcoides ethenogenes* Strain 195 and Its Effects on Growth and Dechlorination Activity

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 2011, p. 7797–7803 Vol. 77, No. 21
0099-2240/11/$12.00 doi:10.1128/AEM.05676-11

Selective Utilization of Exogenous Amino Acids by *Dehalococcoides ethenogenes* Strain 195 and Its Effects on Growth and Dechlorination Activity▽†

Wei-Qin Zhuang,[1]‡ Shan Yi,[1]‡ Xueyang Feng,[2] Stephen H. Zinder,[3] Yinjie J. Tang,[2] and Lisa Alvarez-Cohen[1,4]*

Department of Civil and Environmental Engineering, University of California, Berkeley, California 94720[1]; Energy, Environmental, and Chemical Engineering, Washington University, One Brookings Drive, Box 1180, St. Louis, Missouri 63130[2]; Section of Microbiology, Cornell University, Ithaca, New York 14853[3]; and Earth Sciences Division, Lawrence Berkeley National Laboratory, Cyclotron Rd., Berkeley, California 94720[4]

Received 30 May 2011/Accepted 23 August 2011

Bacteria of the genus *Dehalococcoides* are important members of bioremediation communities because of their ability to detoxify chloroethenes to the benign end product ethene. Genome-enabled studies conducted with *Dehalococcoides ethenogenes* 195 have revealed that two ATP-binding cassette (ABC)-type amino acid transporters are expressed during its exponential growth stages. In light of previous findings that Casamino Acids enhanced its dechlorination activity, we hypothesized that strain 195 is capable of importing amino acids from its environment to facilitate dechlorination and growth. To test this hypothesis, we applied isotopomer-based dilution analysis with ^{13}C-labeled acetate to differentiate the amino acids that were taken up by strain 195 from those synthesized *de novo* and to determine the physiological changes caused by the significantly incorporated amino acids. Our results showed that glutamate/glutamine and aspartate/asparagine were almost exclusively synthesized by strain 195, even when provided in excess in the medium. In contrast, phenylalanine, isoleucine, leucine, and methionine were identified as the four most highly incorporated amino acids, at levels >30% of respective proteinogenic amino acids. When either phenylalanine or all four highly incorporated amino acids were added to the defined mineral medium, the growth rates, dechlorination activities, and yields of strain 195 were enhanced to levels similar to those observed with supplementation with 20 amino acids. However, genes for the putative ABC-type amino acids transporters and phenylalanine biosynthesis exhibited insignificant regulation in response to the imported amino acids. This study also demonstrates that using isotopomer-based metabolite analysis can be an efficient strategy for optimizing nutritional conditions for slow-growing microorganisms.

Bacteria of the genus *Dehalococcoides* play a key role in bioremediation of chlorinated ethenes due to their ability to completely transform these toxic groundwater contaminants into benign ethene (4, 10, 24). However, bioremediation applications are limited by several physiological characteristics of *Dehalococcoides* that make these bacteria challenging to stimulate in the environment or to cultivate in the laboratory, including long doubling times (0.8 to 2.4 days), low cell yields (0.24 to 4.9 g of protein/mol of Cl^- released) (8), and specific requirements for a variety of exogenously supplied compounds (e.g., hydrogen as an electron donor, acetate as a carbon source, and cobalamin as a cofactor) (9, 24, 35). Further, it has been reported that generating large quantities of active *Dehalococcoides* in a timely manner can represent a significant challenge (36, 41). Recent genome sequencing and annotation of several *Dehalococcoides* isolates have enabled novel molecular and biochemical approaches to improve our understanding of their physiology (16, 17, 20, 39), which in turn can lead to the identification of biological mechanisms that limit bioremediation applications using *Dehalococcoides* species.

Genome annotations of four sequenced *Dehalococcoides* strains reveal that each strain possesses nine genes encoding two putative ATP binding cassette (ABC)-type amino acid uptake carriers (19, 25, 35) (www.img.jgi.doe.gov) (see Table S1 in the supplemental material). For strain 195, one of the transporters encoded by six genes (DET0938, DET0941 to DET0945), is annotated as a high-affinity branched-chain amino acid transporter (35), which might be expressed to uptake leucine, isoleucine, and valine across the cytoplasmic membrane of cells. The other transporter encoded by three genes (DET0417 to DET0419) is likely a polar amino acid transporter (35). However, the putative substrates of this transporter are still indeterminate, in that each subunit of the DET0417 to DET0419 transporter seemingly has various substrates (see Table S1 in the supplemental material), and functional annotation based on sequence is often inexact. Nevertheless, a previous transcriptomics study with strain 195 demonstrated the expression of these genes during exponential growth stages (16), suggesting a potential role in the uptake of exogenous amino acids. Maymo-Gatell et al. (24) reported that the addition of Casamino Acids (a complex acid hydrolysate of casein including most amino acids except tryptophan [27]) re-

* Corresponding author. Mailing address: 760 Davis Hall, Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720-1710. Phone: (510) 643-5969. Fax: (510) 642-7483. E-mail: alvarez@ce.berkeley.edu.

‡ W.-Q.Z. and S.Y. contributed equally to this study.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 2 September 2011.

sulted in the enhancement of strain 195 dechlorination. A recent paper reported that, in contrast to the initial sequence annotation of strain 195, this strain synthesizes all 20 common amino acids for biomass protein production with acetate and CO_2 as carbon sources (39). Taken together, previous studies raised two questions about the utilization of exogenous amino acids by *Dehalococcoides* spp.: (i) whether the presence and expression of the ABC transporters lead to the phenotypic function of importing amino acids and (ii) whether any of the imported amino acids plays an important role in facilitating robust and sustainable *Dehalococcoides* growth.

Interestingly, a very recent study using a constraint-based model demonstrated an *in silico* enhancement of biomass yield of *Dehalococcoides* as a result of unlimited uptake of exogenous amino acids (15). The extensive computational model was largely based on the genome sequences and associated annotations of *Dehalococcoides* spp., which may or may not accurately reflect the specific phenotypic functionalities related to the import and incorporation of amino acids. The study described here used isotopomer-based metabolite analysis to quantitatively differentiate the amino acids that were preferentially incorporated from those that were *de novo* synthesized by *D. ethenogenes* strain 195. Further, physiological changes of strain 195 caused by the imported amino acids were quantified. Reverse transcription-quantitative PCR (RT-qPCR) was used to assess whether the genes involved in biosynthesis and transport of the favorable amino acids were differentially expressed in response to exogenous amino acid amendment.

MATERIALS AND METHODS

Chemicals. Trichloroethene (TCE), *cis*-dichloroethene (cDCE), vinyl chloride (VC), and ethene were purchased from Fisher Scientific Co. (Pittsburgh, PA), Acros Chemical Co. (Pittsburgh, PA), Fluka Chemical Co. (Ronkonkoma, NY), and Alltech Associate, Inc. (Deerfield, IL), respectively. [U-^{13}C]glucose (fully labeled, >98% purity) and [1-^{13}C]sodium acetate (>99% purity) were purchased from Cambridge Isotope Laboratories, Inc. (Andover, MA). The 20 amino acids were purchased from Sigma-Aldrich, Inc. (St. Louis, MO). All gases, including medical air, nitrogen, hydrogen, and hydrogen-CO_2 mixture, were obtained from Praxair Inc (Oakland, CA).

Bacterial strain and culture conditions. *D. ethencgenes* strain 195 was grown in batch cultures at 34°C in 160-ml glass serum bottles containing 100 ml of defined mineral salt medium and a H_2/CO_2 headspace (80/20 [vol/vol]) as described previously (39). The medium was amended with a modified Wolin vitamin solution containing a 50-μg/liter final B_{12} concentration (9). Sodium acetate (2 mM) and liquid TCE (7 μl) were added to the culture medium as the carbon source and electron acceptor, respectively. To identify the amino acids that strain 195 took up from the medium, tracer experiments were conducted using [1-^{13}C]sodium acetate as the carbon source, with unlabeled 20 common amino acids (20 mg/liter each) as supplements. In order to observe amino acid uptake by fully acclimatized cultures and reduce the effect of unlabeled carbon from the inoculum (3%), strain 195 was subcultured five times in the same experimental conditions before being harvested for isotopomer-based metabolite analysis. Because of the presence of at least one uptake system that was found or predicted for each amino acid (23), *Escherichia coli* K-12 MG1655 was used as a positive control in order to verify the analytical methods. Strain MG1655 was cultured aerobically with [U-^{13}C]glucose (2 g/liter) in M63 minimal medium (34), with or without amendment of 20 amino acids (50 mg/liter each). Strain MG1655 was grown from a 0.5% inoculum and harvested in the late exponential growth phase (optical density at 600 nm = 1).

In order to reveal the long-term growth effects of amino acid amendments, strain 195 was grown with different amino acid amendments over five subcultures (with similar initial cell densities) prior to quantifying dechlorination and growth rates for comparison. Four cultivation conditions were compared: AA20, amended with all 20 common amino acids (20 mg of each/liter); AA4, amended with phenylalanine, leucine, isoleucine, and methionine (20 mg of each/liter); AA1, amended with 20 mg of phenylalanine/liter; and AA0, an unamended culture. All subcultures were inoculated with similar initial cell densities in order to avoid the effect of various initial cell concentrations.

Analytical methods. For isotopomer-based metabolite analysis, approximately 2 liters of liquid culture (~7.7×10^7 cells/ml) from 20 individual 125-ml bottles were aseptically harvested by centrifugation (22,000 × *g*, 4°C, 15 min) for each experimental condition. Cell pellets were washed with sterile MilliQ water three times and stored at −80°C before use. The preparation and isotopomeric analysis of proteinogenic amino acids were performed as previously described (31, 38). In brief, biomass was hydrolyzed in 6 M HCl at 100°C for 24 h. During the hydrolysis, tryptophan and cysteine are degraded, while asparagine and glutamine are converted into aspartate and glutamate, respectively. The amino acid solution was dried under air flush overnight. Amino acid samples were derivatized in tetrahydrofuran (THF) and *N*-(*tert*-butyl dimethylsilyl)-*N*-methyl-trifluoroacetamide (Sigma-Aldrich) at 70°C for 1 h. A gas chromatograph (Hewlett-Packard, model 6890; Agilent Technologies, Palo Alto, CA) equipped with a DB5-MS column (J&W Scientific, Folsom, CA) and a mass spectrometer (model 5975; Agilent Technologies) was used for isotopomer analysis. Two types of charged fragments were clearly detected by gas chromatography-mass spectrometry for the derivatized amino acids: $[M\text{-}57]^+$, which contains the entire amino acid, and $[M\text{-}159]^+$, which contains the amino acid without the first carbon (α carboxyl group) (39). The $[M\text{-}57]^+$ peaks in leucine, isoleucine, and proline overlap with other peaks, so the $[M\text{-}159]^+$ group was used to obtain the isotopomer labeling information of those amino acids. The final isotopomer labeling fractions were indicated as follows: M0 (unlabeled fraction), M1 (singly labeled fraction), M2 (fraction with two labeled carbons), M3 (fraction with labeled three carbons), and so forth.

For determining the compositions of the proteinogenic amino acids of strain 195, approximately 100 mg (wet weight) of cell pellets ($n = 2$) was harvested from different amino acid amendment conditions by centrifugation as specified above for isotopomer-based metabolite analysis. The analyses were performed by the Molecular Structure Facility, University of California, Davis (http://msf.ucdavis.edu/), using a Beckman 6300 amino acid analyzer (Beckman Coulter, Inc., Fullerton, CA).

Ethenes in culture headspace were measured using a gas chromatograph (Hewlett-Packard, model 5890; Agilent Technologies) with a 30-m J&W capillary column and a flame ionization detector as described previously (9, 38).

Quantification of cell numbers with different amino acid amendments. To compare the growth rates of strain 195 under different amino acid amended conditions, 2-ml cultures were periodically sampled from three replicate culture bottles for DNA extraction. Cells were pelleted by centrifugation (14,000 rpm and 15 min at 4°C) and stored at −20°C until extraction. Genomic DNA (gDNA) was isolated by using a DNeasy Blood & Tissue kit (Qiagen, Inc., Valencia, CA) according to the manufacturer's instructions. Since each strain 195 cell contains only one copy of the *tceA* gene (35), cell numbers were represented as *tceA* gene copies per ml of culture medium. *tceA* gene copies were quantified from triplicate qPCRs performed on each of the triplicate DNA samples by using a StepOnePlus real-time PCR system (Applied Biosystems, Foster City, CA). Briefly, each qPCR mixture contained 2 μl of sample gDNA or a 5-fold serially diluted standard, 1× Fast SYBR green Master Mix (Applied Biosystems), and 0.5 μM concentrations of each forward and reverse primer (see Table S2 in the supplemental material). Standard curves for qPCR analysis of *tceA* genes were generated from known concentrations of strain 195 gDNA with linear regression R^2 values of 99.9% and amplification efficiencies of 99.5 to 105.3%.

RNA extraction, reverse transcription, and qPCR. When ca. 67% of added TCE was dechlorinated, 100 ml of active strain 195 cells (mid-exponential-phase cultures) were collected for RNA extraction from each of triplicate culture bottles representing AA0, AA4, and AA20. Cell samples were collected by filtration through sterile 0.2-μm-pore-size membrane filters, flash-frozen in liquid nitrogen, and stored at −80°C for up to 2 weeks before RNA extraction (16). Total RNA was prepared from frozen cells using an RNeasy minikit (Qiagen) according to the manufacturer's directions. Contaminating DNA in RNA samples was removed with three consecutive rigorous DNase I treatments using a DNA-free kit (Ambion, Austin, TX) according to the manufacturer's directions. Purified RNA was stored at −80°C prior to further use. RNA integrity was checked on a 2100 Bioanalyzer RNA 6000 Nano LabChip (Agilent Technologies), and RNA quantification was determined using a Ribogreen assay on a NanoDrop 3300 Fluorospectrometer (Thermo Scientific, Wilmington, DE). The total mass of purified RNA extracted from the 100-ml cultures ranged from 1.7 to 3.4 μg. cDNA was synthesized in 40-μl reaction mixes containing 50-ng RNA templates, 0.5 μM concentrations of random hexamer, and 50 U of reverse transcriptase by using a TaqMan reverse transcription reagent kit (Applied Biosystems) as described previously (21).

cDNA samples from the reverse transcriptions were diluted 5-fold with nu-

clease-free water and quantified in three replicate qPCRs using Applied Biosystems SYBR green fast mix and primers specific to strain 195's genes of interest (see Table S2 in the supplemental material) as previously described for quantifying *tceA* gene copies. Relative standard curves for qPCR analysis were generated from 5- or 10-fold serially diluted pooled cDNA with linear regression R^2 values of >99.0% and amplification efficiencies within a range of 100% ± 10% (see Table S2 in the supplemental material). Three optimal reference genes—16S rRNA gene, DET0435 (*prs*), and DET1606 (*recA*)—representing the most stable reference genes from a set of tested candidate reference genes were selected by the GeNorm software for normalization (42). REST 2009 software (30) was used to normalize the genes of interest to the geometric mean of the three reference genes, to calculate the relative expression ratios between AA4, AA20, and control (AA0) samples, and to identify statistically significant differences in transcript levels.

Statistical significance for the gene expression data of the branched-chain amino acid transporter (DET0941, DET0942, and DET0944), *tceA* (DET0079) and *pheA* (DET1547) were tested in three conditions (AA0, AA4, and AA20) using pairwise comparison, in order to evaluate their response to the presence of the putative substrates. The *P* values obtained from these tests were then adjusted to correct for multiple comparisons by applying a sequentially rejective procedure using Holm tests (7). The polar amino acid transport genes (DET0417 to DET0419) were assessed for only two experimental conditions (AA0 and AA20), since AA4 probably does not contain the putative substrates for this transporter. *P* values of <0.05 were considered statistically different for DET0417 to DET0419. Only genes exhibiting differential expression of >2-fold, i.e., $\log_2$(relative expression ratio) > 2, were considered biologically significant (16).

RESULTS

Incorporation of amino acids by strain 195. In order to determine which amino acids were imported and incorporated by strain 195 during growth, experiments were performed using ^{13}C-labeled acetate as the carbon source with amendment of 20 unlabeled amino acids, and isotopomer analysis was applied to quantify dilution of labeled amino acids within cells. Since strain 195 possesses all pathways needed for complete *de novo* amino acid biosynthesis (39), the rationale behind this experiment was to evaluate whether strain 195 prefers to import some amino acids from the medium rather than synthesizing them *de novo*.

In order to validate this isotopomer dilution method, we first applied it to *E. coli* strain K-12 MG1655, a strain known to be capable of both complete *de novo* amino acid synthesis and amino acid uptake from medium (3, 23, 40), as a positive control. When this strain was grown with [U-^{13}C]glucose in the presence of 20 unlabeled amino acids, isotopomer results (mass fraction M0) showed 30 to 98% dilution of all tested amino acids in the exponential growth biomass (Fig. 1). This result confirmed the isotopomer dilution method for differentiating amino acid uptake from the medium from *de novo* synthesis. In addition to tryptophan and cysteine that are degraded during protein hydrolysis, Fig. 1 does not contain labeling information for arginine, since the mass spectra for arginine fragments (either $[M-57]^+$ or $[M-159]^+$) are not accurate due to peak overlap and rearrangement (2).

When strain 195 was grown in medium with labeled acetate and unlabeled amino acids, a total of 13 amino acids exhibited dilutions greater than 5% (Fig. 1). Eight of these are hydrophobic in nature, while the other five are hydrophilic with polar side chains (see Table S3 in the supplemental material). Only background level dilution rates (<5%) were detected for glutamate, glutamine, aspartate, and asparagine, indicating that they were not significantly imported by strain 195. The dilutions of serine and glycine were relatively low (~10%), indicating that they are also mostly synthesized *de novo* (from 3-P-glycerate or C1 metabolism) (29). Four amino acids exhibited >35% dilution (Fig. 1), with phenylalanine (Phe) exhibiting the highest dilution of over 80%. Isoleucine (Ile), leucine (Leu), and methionine (Met) were the next highest imported amino acids, with 46 to 37% dilution. Although tyrosine shares a similar biosynthesis pathway to phenylalanine, it exhibited a lower labeling dilution (~20%), indicating that tyrosine was mainly synthesized *de novo*.

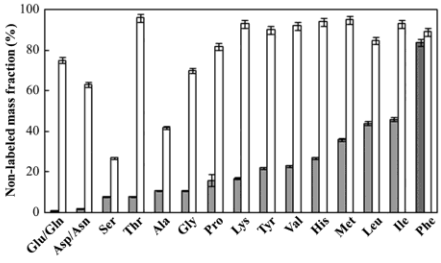


FIG. 1. Mass fraction of unlabeled proteinogenic amino acids (M0 fraction) of *D. ethenogenes* strain 195 (▤) and *E. coli* K-12 MG1655 (□).

Effects of amino acids on physiological traits of *D. ethenogenes* strain 195. The high dilution of Phe, Leu, Ile, and Met indicated that strain 195 was importing these amino acids under the experimental conditions. In order to assess the effects of the amendments on strain 195 physiology, a comparison of growth and TCE degradation was performed for strain 195 amended with different amino acid combinations, including no amino acids (AA0), Phe only (AA1), Phe, Iso, Leu, and Met (AA4), and all 20 common amino acids (AA20). TCE dechlorination patterns demonstrated that the AA0 culture dechlorinated ~65 μmol TCE after 15 days, while the AA1, AA4, and AA20 cultures depleted the TCE within 10 days (Fig. 2). The experiment also showed that AA1 generated less ethene than AA4 and AA20, and thus more vinyl chloride remained in the end. Furthermore, all three amino acid-amended cultures grew faster than the unamended culture AA0 (Fig. 3). During the exponential phase, the amino acid-amended cultures exhibited shorter doubling times than did the unamended culture (Table 1). Amino acid-amended cultures reached their highest cell number before day 8, while the unamended culture reached its highest cell number after day 10. In addition, the cell yields of the amino acid amended cultures were two to three times that of the unamended culture (Table 1). All amino acid amended cultures reproducibly exhibited very similar growth and degradation profiles, suggesting that the incorporated amino acids, especially phenylalanine, play an important role in enhancing biomass growth.

To test the hypothesis that variations in availability of amino acids for import and/or variations in growth rates may affect amino acid ratios within cells (32), the proteinogenic amino acids in strain 195 were analyzed in AA0, AA4, and AA20. In each condition, overall amino acid composition of stain 195 exhibited negligible differences (see Table S4 in the supple-

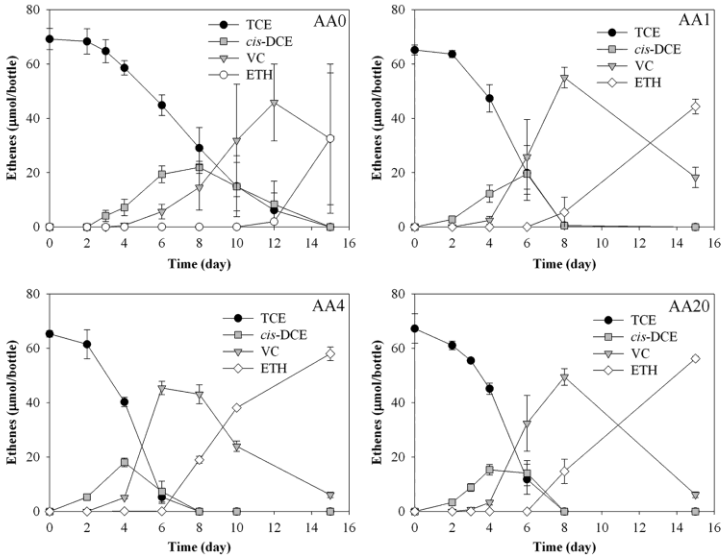


FIG. 2. Dechlorination profiles by *D. ethenogenes* strain 195 amended with 0 (AA0), 1 (AA1), 4 (AA4), or 20 (AA20) amino acids after five subcultures. The data shown are the mean values from triplicate biotic experiments, with error bars representing one standard deviation.

mental material). Interestingly, we also found that the most abundant proteinogenic amino acids (i.e., alanine, aspartate/asparagine, glycine, and glutamate/glutamine) did not correspond to the amino acids that were most preferentially imported from the medium (phenylalanine, isoleucine, leucine, and methionine).

Transcript levels of the genes involved in ABC transporters and phenylalanine synthesis. RT-qPCR was used to investigate the transcriptional response of the putative ABC-type transporter genes to the presence of the putative substrates. Here, both ABC transporter genes exhibited no significant response to the presence of putative amino acid substrates in either the AA4 or the AA20 culture (Table 2). Although DET0944, the gene encoding the substrate-binding protein for the branched-chain amino acid transporter, exhibited down-regulation that was significant according to the *P* value, the fold change in relative expression ratio was not significant.

Since a major fraction of phenylalanine (>80%) was imported by strain 195, the potential regulation of imported phe-

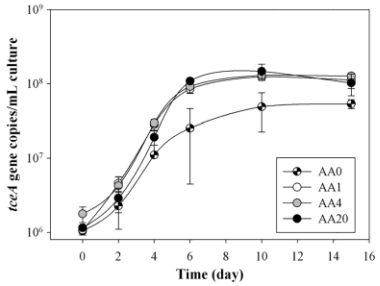


FIG. 3. Growth curves of *D. ethenogenes* strain 195 amended with 0 (AA0), 1 (AA1), 4 (AA4), or 20 (AA20) amino acids after five subcultures. The data shown are the mean values from triplicate biotic experiments, with error bars representing one standard deviation.

TABLE 1. Doubling times and cell yields of *D. ethenogenes* strain 195 using TCE and amended with 0 (AA0), 1 (AA1), 4 (AA4), or 20 (AA20) amino acids after five subcultures

Culture	Mean ± SD[a]	
	Doubling time (days)[b]	Cell yield (10^7/μmol of Cl^- released)[c]
AA0	1.3 ± 0.3	3.8 ± 0.6
AA1	0.95 ± 0.04	7.8 ± 0.3
AA4	0.91 ± 0.16	10.0 ± 0.6
AA20	0.76 ± 0.08	7.5 ± 2.5

[a] F tests were performed to determine statistically significant differences ($P < 0.05$) in variances of the four conditions.
[b] The doubling times were calculated from cell numbers during exponential growth on days 2 and 6 (*tceA* copies).
[c] The Cl^- released was computed using only metabolic dechlorination from TCE to VC as described by Duhamel and Edwards (6).

TABLE 2. Differential transcript levels of genes involved in ABC-type amino acid transporters and phenylalanine synthesis[a]

Gene	Relative expression ratio[b]		*P*[c]		Annotated functions
	AA4	AA20	AA4	AA20	
DET0417	NA[d]	0.65	NA	0.68	ATP binding protein of the polar amino acid transporter
DET0418	NA	0.73	NA	0.63	Permease subunit of the polar amino acid transporter
DET0419	NA	0.64	NA	0.28	Substrate binding protein of the polar amino acid transporter
DET0941	0.69	0.45	0.16	0.05	ATP binding protein of the branched-chain amino acid transporter
DET0942	0.73	0.52	0.22	0.10	Permease subunits of the branched-chain amino acid transporter
DET0944	0.82	0.84	0.05*	0.01*	Substrate binding protein of the branched-chain amino acid transporter
pheA (DET1547)	0.86	0.78	0.42	0.21	Prephenate dehydratase
tceA (DET0079)	0.78	0.49	0.41	0.20	Reductive dehalogenase

[a] All measurements are averages of triplicate qPCR results from triplicate biological samples.
[b] Relative expression ratios were generated from REST 2009 software using the mathematical model described by Pfaffl et al. (32).
[c] *P* values were generated by using REST 2009 software. *, Statistically significant *P* values in pairwise comparison that fulfilled the criterion in the Holm test.
[d] NA, not applicable.

nylalanine on its biosynthesis pathway was investigated. The gene *pheA* (DET1547) that encodes the enzyme transforming prephenate to phenylpyruvate (see Fig. S1 in the supplemental material) contains an ACT domain. The ACT domain, derived from aspartokinase, chorismate mutase and TyrA (prephenate dehydrogenase), has a putative role as a regulatory module involved in a feedback inhibition regulated by the presence of phenylalanine on the catalytic enzyme(s) (22). The transcript levels of this gene showed no significant differences between amino acid amended cultures and control AA0, indicating the absence of transcriptional feedback inhibition to phenylalanine synthesis by imported amino acids (Table 2). Similarly, the reductive dehalogenase gene *tceA* was not regulated under AA4 and AA20 conditions.

DISCUSSION

Although strains of *Dehalococcoides* have been found in many different subsurface environments (11), researchers have observed that the growth of *Dehalococcoides* is usually more robust in complex microbial communities than in isolation (9). Additions of cofactors such as vitamin B_{12} and complex nutrient supplements have been found to enhance both the dechlorination and growth of *Dehalococcoides* (24). Identification of other important growth factors for the robust and sustained cultivation of *Dehalococcoides* could lead to improved understanding of the physiology of this organism and to potentially more successful bioremediation strategies (36, 41). Conventional cultivation-dependent screening techniques, such as a monofactorial analysis in batch culture experiments by stepwise omission of single amino acids, can be used for such identification as demonstrated by Holliger et al. (13) in their study with *Dehalobacter restrictus*.

In the present study, we used results of isotopomer dilution analysis to differentiate imported and synthesized amino acids by *D. ethenogenes* 195 and to further assay the physiological effects of these amino acids on this strain. Four amino acids—phenylalanine, isoleucine, leucine, and methionine—were successfully identified as effective nutrient factors for enhancing the growth yield and dechlorination rates of strain 195. By importing amino acids rather than synthesizing them *de novo*, strain 195 may save energy and alleviate bottlenecks associated with anabolic pathways that limit the growth rate of *Dehalococcoides*, as has been recently demonstrated *in silico* using a constraint-based model (15). Based on estimations made for *E. coli* using acetate as a sole carbon and energy source, it costs about 13 to 87 high-energy phosphate bonds for a cell to synthesize one amino acid (1). Our isotopomer dilution results revealed that strain 195 could import some energetically costly amino acids (e.g., 35 to 58 phosphate bonds per amino acid in *E. coli*), such as phenylalanine, isoleucine, leucine, and methionine (37). As a comparison, the energy cost is estimated to be 1 to 2 ATPs per amino acid for importing them from the growth medium (12, 26, 28).

In addition, aromatic precursors (e.g., chorismate) of phenylalanine are likely the same precursors of ubiquinones that are integrated in large amounts into *Dehalococcoides* membranes, possibly involved in protecting against free radicals during reductive dechlorination (43). Therefore, the availability of large amounts of exogenous phenylalanine might reduce the anabolic energy cost compared to biosynthesis of these high-energy consuming precursors. Interestingly, tyrosine shares some of the same precursor, prephenate, as phenylalanine (see Fig. S1 in the supplemental material), but tyrosine was mainly synthesized *de novo* by strain 195. Tyrosine synthesis is via prephenate dehydrogenase (EC 1.3.1.12; DET0460): prephenate + $NAD^+ \rightarrow$ 4-hydroxyphenylpyruvate + CO_2 + NADH, whereas phenylalanine synthesis occurs via prephenate dehydratase: prephenate $\rightarrow$ phenylpyruvate + H_2O + CO_2. Consequently, the biosynthesis of tyrosine requires NAD^+ as a reactant while producing NADH, which is predicted to be an important limiting factor for *Dehalococcoides* growth (15). Therefore, *Dehalococcoides* cells may be able to grow more efficiently by allowing tyrosine synthesis to occur (generating NADH) while importing available phenylalanine from the medium in order to diminish competition for the prephenate precursor. Furthermore, although tryptophan is the most energy-intensive amino acid to synthesize (1, 37), isotopomer analysis using current protocols is unable to measure this amino acid. Nevertheless, comparison of TCE dechlorination and the growth of strain 195 in AA4 and AA20 demonstrates negligible differences, indicating that the amino acids beyond the four most preferentially imported ones (including tryptophan) provide little additional benefit to strain 195 under the conditions tested in the present study.

The amino acid dilution results indicated that strain 195 completely synthesized aspartate, asparagine, glutamate, and glutamine, even though these amino acids were available in the medium. This is distinctly different from fast-growing bacteria, e.g., *E. coli* strain K-12 MG1655, which showed significant import and utilization of all exogenous amino acids, as shown in Fig. 1. This difference might be associated with the environmental origins of these two bacteria, where *E. coli* likely evolved a metabolism for fast growth in nutrient-rich environments, such as the guts of animals (14), whereas *Dehalococcoides* is common in nutrient-scarce environments such as groundwater aquifers, where its metabolism may have evolved to rely heavily on *de novo* amino acid biosynthesis (5). In addition, a total of 13 amino acids exhibited dilutions from 5% to >80% demonstrated that strain 195 can only selectively import and incorporate certain amino acids with different efficiencies. This finding is different from an assumption in the previous modeling study that *Dehalococcoides* were capable of unselectively taking in all amino acids (15). The discrepancy between our results and the previous study indicates the understanding of *Dehalococcoides* physiology is still limited. Thus, the knowledge obtained here could be integrated into future *in silico* models of *Dehalococcoides* for a better simulation.

The genome annotations of strain 195 and other sequenced *Dehalococcoides* strains predict that the two ABC-type amino acid transporters would likely uptake amino acids with different characteristics of hydrophobicity and polarity (see Table S1 in the supplemental material). Although data from the isotopomer analysis confirmed 195's ability to import amino acids with different chemical characteristics (see Table S3 in the supplemental material) (e.g., hydrophilic Thr, Lys, and His, together with hydrophobic Pro, Ala, Tyr, Val, Met, Leu, Ile, and Phe), the actual substrate(s) for the transporters have not yet been positively identified. For example, the predicted branched-chain transporter (DET0938, DET0941 to DET0944) was annotated to have a broad substrate specificity, and bioinformatics findings suggest that this transporter is similar to the LIV-I/LS system in *E. coli* (33), which catalyzes the translocation of not only Leu, Ile, and Val but also Phe and Tyr from the periplasm to the cytoplasm (18). An analogous broad specificity was observed in *E. coli* strain K-12 MG1655 (Fig. 1). However, these hypothetical substrates were not equally imported by strain 195, and a differential expression analysis revealed that the transporter genes were not regulated by the presence of their hypothetical substrates. The same was observed for the putative polar amino acid transporters of strain 195. It has been reported that transporters with broad substrate specificities tend to be constitutively expressed regardless of the presence of their substrates (23) and that may be the case in the present study.

In general, we demonstrate here that the specific amino acids that have stimulatory effects on slow-growing bacteria can be effectively identified by an isotopomer-based metabolite analysis. Moreover, although *D. ethenogenes* 195 exhibited no metabolic regulation for import of amino acids under the experimental conditions, exogenous amendment with one or more amino acids resulted in enhanced growth and dechlorination.

ACKNOWLEDGMENTS

This research was funded by the Superfund Basic Research Program under grant NIEHS ES04705, the Strategic Environmental Research and Development Program at U.S. Department of Energy (project ER-1587). This study was also partially supported by an NSF Career Grant (MCB0954016).

REFERENCES

1. **Akashi, H., and T. Gojobori.** 2002. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. Proc. Natl. Acad. Sci. U. S. A. **99:**3695–3700.
2. **Antoniewicz, M. R., J. K. Kelleher, and G. Stephanopoulos.** 2007. Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. Anal. Chem. **79:**7554–7559.
3. **Blattner, F. R., et al.** 1997. The complete genome sequence of *Escherichia coli* K-12. Science **277:**1453–1462.
4. **Cupples, A. M., A. M. Spormann, and P. L. McCarty.** 2003. Growth of a *Dehalococcoides*-like microorganism on vinyl chloride and *cis*-dichloroethene as electron acceptors as determined by competitive PCR. Appl. Environ. Microbiol. **69:**953–959.
5. **Devlin, J. F., D. Katic, and J. F. Barker.** 2004. *In situ* sequenced bioremediation of mixed contaminants in groundwater. J. Contam. Hydrol. **69:**233–261.
6. **Duhamel, M., and E. A. Edwards.** 2007. Growth and yields of dechlorinators, acetogens, and methanogens during reductive dechlorination of chlorinated ethenes and dihaloelimination of 1,2-dichloroethane. Environ. Sci. Technol. **41:**2303–2310.
7. **Glantz, S. A.** 2001. Primer of biostatistics, 5th ed. McGraw-Hill, New York, NY.
8. **He, J., Y. Sung, R. Krajmalnik-Brown, K. M. Ritalahti, and F. E. Löffler.** 2005. Isolation and characterization of *Dehalococcoides* sp. strain FL2, a trichloroethene (TCE)- and 1,2-dichloroethene-respiring anaerobe. Environ. Microbiol. **7:**1442–1450.
9. **He, J. Z., V. F. Holmes, P. K. H. Lee, and L. Alvarez-Cohen.** 2007. Influence of vitamin B_{12} and cocultures on the growth of *Dehalococcoides* isolates in defined medium. Appl. Environ. Microbiol. **73:**2847–2853.
10. **He, J. Z., K. M. Ritalahti, K. L. Yang, S. S. Koenigsberg, and F. E. Löffler.** 2003. Detoxification of vinyl chloride to ethene coupled to growth of an anaerobic bacterium. Nature **424:**62–65.
11. **Hendrickson, E. R., et al.** 2002. Molecular analysis of *Dehalococcoides* 16S ribosomal DNA from chloroethene-contaminated sites throughout North America and Europe. Appl. Environ. Microbiol. **68:**485–495.
12. **Higgins, C. F., and K. J. Linton.** 2004. The ATP switch model for ABC transporters. Nat. Struct. Mol. Biol. **11:**918–926.
13. **Holliger, C., et al.** 1998. *Dehalobacter restrictus* gen. nov. and sp. nov., a strictly anaerobic bacterium that reductively dechlorinates tetra- and trichloroethene in an anaerobic respiration. Arch. Microbiol. **169:**313–321.
14. **Hudault, S., J. Guignot, and A. L. Servin.** 2001. *Escherichia coli* strains colonizing the gastrointestinal tract protect germfree mice against *Salmonella typhimurium* infection. Gut **49:**47–55.
15. **Islam, A. M., E. A. Edwards, and R. Mahadevan.** 2010. Characterizing the metabolism of *Dehalococcoides* with a constraint-based model. PLoS Comput. Biol. **6:**e1000887.
16. **Johnson, D. R., et al.** 2008. Temporal transcriptomic microarray analysis of "*Dehalococcoides ethenogenes*" strain 195 during the transition into stationary phase. Appl. Environ. Microbiol. **74:**2864–2872.
17. **Johnson, D. R., A. Nemir, G. L. Andersen, S. H. Zinder, and L. Alvarez-Cohen.** 2009. Transcriptomic microarray analysis of corrinoid responsive genes in *Dehalococcoides ethenogenes* strain 195. FEMS Microbiol. Lett. **294:**198–206.
18. **Koyanagi, T., T. Katayama, H. Suzuki, and H. Kumagai.** 2004. Identification of the LIV-I/LS system as the third phenylalanine transporter in *Escherichia coli* K-12. J. Bacteriol. **186:**343–350.
19. **Kube, M., et al.** 2005. Genome sequence of the chlorinated compound respiring bacterium *Dehalococcoides* species strain CBDB1. Nat. Biotechnol. **23:**1269–1273.
20. **Lee, P. K. H., J. He, S. H. Zinder, and L. Alvarez-Cohen.** 2009. Evidence for nitrogen fixation by "*Dehalococcoides ethenogenes*" strain 195. Appl. Environ. Microbiol. **75:**7551–7555.
21. **Lee, P. K. H., T. W. Macbeth, K. S. Sorenson, R. A. Deeb, and L. Alvarez-Cohen.** 2008. Quantifying genes and transcripts to assess the in situ physiology of "*Dehalococcoides*" spp. in a trichloroethene-contaminated groundwater site. Appl. Environ. Microbiol. **74:**2728–2739.
22. **Liberles, J. S., M. Thórólfsson, and A. Martínez.** 2005. Allosteric mechanisms in ACT domain containing enzymes involved in amino acid metabolism. Amino Acids **28:**1–12.
23. **Marin, K., and R. Krämer.** 2007. Amino acid transport systems in biotechnologically relevant bacteria, p. 289–325. *In* V. F. Wendisch (ed.), Amino acid biosynthesis: pathways, regulation, and metabolic engineering. Springer-Verlag, Heidelberg, Germany.

24. **Maymo-Gatell, X., Y. T. Chien, J. M. Gossett, and S. H. Zinder.** 1997. Isolation of a bacterium that reductively dechlorinates tetrachloroethene to ethene. Science **276:**1568–1571.
25. **McMurdie, P. J., et al.** 2009. Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. PLoS Genet. **5:**e1000714.
26. **Mimmack, M. L., et al.** 1989. Energy coupling to periplasmic binding protein-dependent transport-systems: stoichiometry of ATP hydrolysis during transport *in vivo*. Proc. Natl. Acad. Sci. U. S. A. **86:**8257–8261.
27. **Mueller, J. H., and E. R. Johnson.** 1941. Acid hydrolysates of casein to replace peptone in the preparation of bacteriological media. J. Immunol. **40:**33–38.
28. **Patzlaff, J. S., T. van der Heide, and B. Poolman.** 2003. The ATP/Substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. J. Biol. Chem. **278:**29546–29551.
29. **Perrenoud, A., and U. Sauer.** 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.
30. **Pfaffl, M. W., G. W. Horgan, and L. Dempfle.** 2002. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Res. **30:**e36.
31. **Pingitore, F., Y. Tang, G. H. Kruppa, and J. D. Keasling.** 2007. Analysis of amino acid isotopomers using FT-ICR MS. Anal. Chem. **79:**2483–2490.
32. **Pramanik, J., and J. D. Keasling.** 1998. Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. Biotechnol. Bioeng. **60:**230–238.
33. **Saier, M. H., M. R. Yen, K. Noto, D. G. Tamang, and C. Elkan.** 2009. The transporter classification database: recent advances. Nucleic Acids Res. **37:** D274–D278.
34. **Sambrook, J., E. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
35. **Seshadri, R., et al.** 2005. Genome sequence of the PCE-dechlorinating bacterium *Dehalococcoides ethenogenes*. Science **307:**105–108.
36. **Simpkin, T. J., and R. D. Norris.** 2010. Engineering and implementation challenges for chlorinated solvent remediation, p. 109–143. *In* H. F. Stroo and C. H. Ward (ed.), *In situ* remediation of chlorinated solvent plumes. Springer, New York, NY.
37. **Stephanopoulos, G. N., A. A. Aristidou, and J. P. Nielsen.** 1998. Metabolic engineering principles and methodologies, vol. 75. Academic Press, Inc., San Diego, CA.
38. **Tang, Y., et al.** 2007. Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using gas chromatography-mass spectrometry and Fourier transform-ion cyclotron resonance mass spectrometry. J. Bacteriol. **189:**940–949.
39. **Tang, Y. J., et al.** 2009. Investigation of carbon metabolism in "*Dehalococcoides ethenogenes*" strain 195 by use of isotopomer and transcriptomic analyses. J. Bacteriol. **191:**5224–5231.
40. **Tao, H., C. Bausch, C. Richmond, F. R. Blattner, and T. Conway.** 1999. Functional genomics: expression analysis of *Escherichia coli* growing on minimal and rich media. J. Bacteriol. **181:**6425–6440.
41. **Vainberg, S., C. Condee, and R. Steffan.** 2009. Large-scale production of bacterial consortia for remediation of chlorinated solvent-contaminated groundwater. J. Ind. Microbiol. Biotechnol. **36:**1189–1197.
42. **Vandesompele, J., et al.** 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. **3:**0034.1.
43. **White, D. C., et al.** 2005. Phospholipid furan fatty acids and ubiquinone-8: lipid biomarkers that may protect *Dehalococcoides* strains from free radicals. Appl. Environ. Microbiol. **71:**8426–8433.

Supporting Information

Title: Selective Utilization of Exogenous Amino Acids by *Dehalococcoides ethenogenes* strain 195 and the Effects on Growth and Dechlorination Activity

Authors: Wei-Qin Zhuang[1#], Shan Yi[1#], Xueyang Feng[2], Stephen H. Zinder[3], Yinjie J. Tang[2]and Lisa Alvarez-Cohen[1,4,*]

Affiliation:

[1]Department of Civil and Environmental Engineering, University of California, Berkeley, CA 94720; [2]Energy, Environmental and Chemical Engineering, Washington University One Brookings Drive, Box 1180, St. Louis, MO 63130; [3]Section of Microbiology, Cornell University, Ithaca, New York 14853;[4]Earth Sciences Division, Lawrence Berkeley National Laboratory, Cyclotron Rd., Berkeley, CA 94720;

* **Correspondent:** E-mail: alvarez@ce.berkeley.edu, Phone: (510) 643-5969, Fax: (510) 642-7483.

[#]**W.-Q.Z.** and **S.Y.** contributed equally to this study.

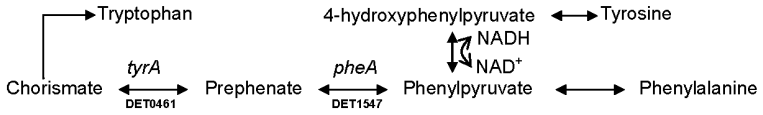


Fig. S1. Biosynthetic pathway of phenylalanine annotated in *D. ethenogenes* strain 195 (3).

Table S1. Genes annotated as putative ABC-type amino acid transporters in four sequenced *Dehalococcoides* strains and their amino acid sequence similarities to strain 195 homologs.

ABC-system	strains	Permease subunits (% similarity to 195 homologs)	Substrate binding protein (% similarity to 195 homologs)	ATP binding proteins (% similarity to 195 homologs)	Putative Substrate (references)
Polar amino acid transporter	195	DET0418	DET0419	DET0417	DET0418:His/Glu/Gln/Arg/opine family (6); polar amino acid (K02028)(3);
	BAV1	DehaBAV1_0397 (91.21%)	DEhaBAV1_0398 (88.08%)	DehaBAV1_0396 (95.26%)	(5)
	VS	DeVSDRAFT_1003 (94.87%)	DeVSDRAFT_1002 (95.38%)	DeVSDRAFT_1004 (95.65%)	polar amino acid (5)
	CBDB1	cbdb_A370 (91.58%)	cbdb_A371 (88.08%)	cbdb_A369 (95.26%)	Unspecified (4)
Branched chain amino acid transporter	195	DET0944 DET0945	DET0942 DET0943	DET0938 DET0941	Branched-chain amino acid (6);
	BAV1	DehaBAV1_0828 (97.43%); DehaBAV1_0829 (97.62%)	DehaBAV1_0826 (93.46%); DehaBAV1_0827 (95.81%)	DehaBAV1_0822 (93.33%); DehaBAV1_0825(97.34 %);	Branched-chain amino acid (5)
	VS	DeVSDRAFT_1243 (99.14%); DeVSDRAFT_1242 (99.66%)	DeVSDRAFT_1245 (94.86%); DeVSDRAFT_1244 (98.37%)	DeVSDRAFT_1249 (93.70%); DeVSDRAFT_1246 (97.34%)	Branched-chain amino acid (5)
	CBDB1	cbdb_A898 (97.43%); cbdb_A899 (97.62%)	cbdb_A896 (93.93%); cbdb_A897 (95.58%)	cbdb_A892 (93.33%); cbdb_A895 (97.72%)	Branched-chain amino acid (4)

Table S2. Sequences of the primers[#] used in this study and qPCR data validation.

Primer ID: 5'-Sequence-3'	Target gene (Accession no.)	Reference	R^{2*}	Amplification Efficiency[*]
Reference genes				
16S rRNA_F:GGTAATACGTAGGGAAGCAAGCG 16S rRNA_R:CCGGTTAAGCCGGGAAATT	16S rRNA gene DET_De16S (CP000027)	(1)	0.995	0.97
DET0435_F: ATTCTTTGATATGCCGGTAGACG DET0435_R: GGTCTTCCAGCTTCTTTTTCTTGA	*prs*, DET0435 (YP_181179.1)	This study	0.996	0.96
DET1606_F: AGCCGAAATAGAAGGCGACAT DET1606_R: CCCGATGGAAGCAGTCAGTT	*recA*, DET1606 (YP_182301.1)	This study	0.991	1.05
Genes of interest				
*tceA*_F: ATCCAGATTATGACCCTGGTGAA *tceA*_R:GCGGCATATATTAGGGCATCTT	*tceA* (AF228507)	(2)	0.995	0.97
DET0417_F: ACCCCCAGATAATGCTCTTTGA DET0417_R: TCACGTCCAGGACTTCTTTTATCA	DET0417 (YP_181162.1)	This study	0.993	1.06
DET0418_F: TTCGGGCCGATGATGAAT	DET0418	This study	0.994	0.90

4

DET0418_R: CATACCCCAGATAGCCATCACA	(YP_181163.1)			
DET0419_F: CATAACGACCTCACCGGTAAAAA DET0419_R: CCAGGCCGATTTCATCGTA	DET0419 (YP_181164.1)	This study	0.998	0.99
DET0941_F: TGGGTATAGTTCAGGCTCTGGAA DET0941_R: CGCCCCCACAATCAGATTT	DET0941 (YP_181666.1)	This study	0.996	1.09
DET0942_F: AGTTCACCATCACTTACCCATCCT DET0942_R: TGCATAAAAGCTACCCAGTCATCA	DET0942 (YP_181667.1)	This study	0.997	1.04
DET0944_F: GATAGCGTCCGGCATTATGG DET0944_R: AGTGATGGCAACCCGAACA	DET0944 (YP_181669.1);	This study	0.990	1.08
DET1547_F: TGTAGCGCGCCATAAATTCC DET1547_R: CTGATGGGAGGTGTCACTTTCA	DET1547 (YP_182245.1)	This study	0.995	1.04

[#]Primers were designed using the Primer Express 3.0 software (Applied Biosystems, Foster City, CA). The specificity of the primers was checked both *in silico* [Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/)] and *in vitro* [melt curves using serially diluted strain 195 gDNA standards on StepOnePlus™ real-time PCR system (Applied Biosystems, Foster City, CA)].

*qPCR analysis was valid when linear regression coefficient, R^2 was ≥0.99 and amplification

efficiency was within the range of 100±10%.

Table S3. Chemical characteristics of unambiguously imported amino acids in *D. ethenogenes*strain 195.

Amino acid	Dilution ratio	Polarity	Hydrophobicity	Charges
Phe	0.84		hydrophobic	neutral
Ile	0.46		hydrophobic	neutral
Leu	0.44		hydrophobic	neutral
Met	0.37		hydrophobic	neutral
His	0.27	Y	hydrophilic	Postive
Val	0.24		hydrophobic	neutral
Tyr	0.22	Y	hydrophobic	weakly acidic
Lys	0.18	Y	hydrophilic	Positive
Ala	0.11		hydrophobic	neutral
Gly	0.11		hydrophilic	neutral
Pro	0.10		hydrophobic	neutral
Thr	0.09	Y	hydrophilic	neutral
Ser	0.09	Y	hydrophilic	neutral

Table S4. The amino acid composition of *D. ethenogenes* strain 195 grown with and without amino acid supplementation (measurement errors below 0.2%, n=2).

Sample	AAC 0 Mole %	AAC 4 Mole %	AAC 20 Mole %
Ala	11.0	11.5	11.5
Asp/Asn	10.7	10.4	10.2
Glycine	9.5	9.9	9.7
Glu/Gln	9.0	9.3	9.2
Leucine	8.2	8.3	8.3
Valine	7.7	6.5	7.9
Lysine	7.3	7.6	7.2
Threonine	7.0	7.1	6.9
Isoleucine	6.2	6.3	6.2
Serine	5.8	5.8	5.6
Proline	5.0	5.1	4.9
Arginine	4.0	4.1	4
Phenylalanine	3.7	3.2	3.6
Tyrosine	1.5	1.7	1.6
Histidine	1.5	1.5	1.4
Methionine	1.0	1.0	1.2
Cysteine	1.0	0.7	0.6
Tryptophan	ND	ND	ND

ND: degraded during hydrolysis.

Supplemental references:

1. **Holmes, V. F., J. He, P. K. H. Lee, and L. Alvarez-Cohen.** 2006. Discrimination of Multiple Dehalococcoides Strains in a Trichloroethene Enrichment by Quantification of Their Reductive Dehalogenase Genes. Appl. Environ. Microbiol. **72:**5877-5883.
2. **Johnson, D. R., P. K. H. Lee, V. F. Holmes, A. C. Fortin, and L. Alvarez-Cohen.** 2005. Transcriptional Expression of the tceA Gene in a Dehalococcoides-Containing Microbial Enrichment. Appl. Environ. Microbiol. **71:**7145-7151.
3. **Kanehisa, M., and S. Goto.** 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research **28:**27-30.
4. **Kube, M., A. Beck, S. H. Zinder, H. Kuhl, R. Reinhardt, and L. Adrian.** 2005. Genome sequence of the chlorinated compound respiring bacterium Dehalococcoides species strain CBDB1. Nature Biotechnology **23:**1269-1273.
5. **McMurdie, P. J., S. F. Behrens, J. A. Muller, J. Goke, K. M. Ritalahti, R. Wagner, E. Goltsman, A. Lapidus, S. Holmes, F. E. Loffler, and A. M. Spormann.** 2009. Localized plasticity in the streamlined genomes of vinyl chloride respiring *Dehalococcoides*. PLoS Genet. **5:**e1000714.
6. **Seshadri, R., L. Adrian, D. E. Fouts, J. A. Eisen, A. M. Phillippy, B. A. Methe, N. L. Ward, W. C. Nelson, R. T. Deboy, H. M. Khouri, J. F. Kolonay, R. J. Dodson, S. C. Daugherty, L. M. Brinkac, S. A. Sullivan, R. Madupu, K. T. Nelson, K. H. Kang, M. Impraim, K. Tran, J. M. Robinson, H. A. Forberger, C. M. Fraser, S. H. Zinder, and J. F. Heidelberg.** 2005. Genome sequence of the PCE-dechlorinating bacterium Dehalococcoides ethenogenes. Science **307:**105-108.

Appendix 6

Central Metabolism in *Mycobacterium smegmatis* during the Transition from O_2-rich to O_2-poor Conditions as Studied by Isotopomer-assisted Metabolite Analysis

Biotechnol Lett (2009) 31:1233–1240
DOI 10.1007/s10529-009-9991-7

ORIGINAL RESEARCH PAPER

Central metabolism in *Mycobacterium smegmatis* during the transition from O_2-rich to O_2-poor conditions as studied by isotopomer-assisted metabolite analysis

Yinjie J. Tang · Wenqing Shui · Samuel Myers · Xueyang Feng · Carolyn Bertozzi · Jay D. Keasling

Received: 9 February 2009 / Revised: 17 March 2009 / Accepted: 19 March 2009 / Published online: 9 April 2009

Abstract Isotopomer-assisted metabolite analysis was used to investigate the central metabolism of *Mycobacterium smegmatis* and its transition from normal growth to a non-replicating state under a hypoxic environment. Tween 80 significantly promoted aerobic growth by improving O_2 transfer, while only small amount was degraded and metabolized via the TCA cycle for biomass synthesis. As the bacillus encountered hypoxic stress, isotopomer analysis suggested: (1) isocitrate lyase activity increased, which further induced glyoxylate pathway and glycine dehydrogenase for replenishing NAD^+; (2) the relative amount of acetyl-CoA entering the TCA cycle was doubled, whereas little entered the glycolytic and pentose phosphate pathways.

Yinjie J. Tang and Wenqing Shui contributed equally to this work.

Y. J. Tang (✉) · X. Feng
Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO 63130, USA
e-mail: yinjie.tang@wustl.edu

W. Shui · C. Bertozzi
Department of Chemistry, University of California, Berkeley, CA 94720, USA

S. Myers · J. D. Keasling (✉)
Department of Chemical Engineering, University of California, Berkeley, CA 94720, USA
e-mail: Keasling@berkeley.edu

Keywords Glycine · Glyoxylate · NAD^+ · Non-replicating · Tween 80

Introduction

Mycobacterium tuberculosis is a sophisticated pathogen that is able to persist in the human host for decades. It is estimated that up to one-third of the world's population is latently infected with this bacillus (Boshoff and Barry 2005), which enters a so-called non-replicating persistence (NRP) state. The NRP state is characterized by a lack of bacillus division and basic changes in metabolic activity. These non-replicating bacilli are difficult to eradicate because they are not reliant on the biosynthetic machinery targeted by conventional antibiotics (Höner zu Bentrup and Russell 2001). Thus, a deeper understanding of the characteristics of central metabolism and its changes during shifts from growth to the NRP state might open new avenues for pathogen clearance.

One of the key factors in the host environment responsible for inducing the NRP state is O_2 limitation in infected tissues and necrotic lesions (Wayne and Sohaskey 2001). Microarray and proteomic studies have shown many changes in gene expression as the bacillus undergoes hypoxic stress prior to metabolic dormancy (Boshoff and Barry 2005). In this study, isotopomer labeling was applied to

investigate mycobacterial central metabolism along with its change in response to hypoxia. Isotopomer-assisted metabolite analysis can quantitatively track metabolic pathways and identify enzyme activities in cells (Tang et al. 2009). When the carbon substrate (e.g. glycerol) is labeled in specific positions with a non-radioactive isotope (^{13}C), the resulting isotopic label pattern of cellular metabolites reveals the active metabolic pathways. GC-MS is a sensitive method to measure mass isotopomer distributions in metabolites. The most widely used GC-MS protocol is based on derivatizing amino acids from proteins with tert-butyldimethylsilyl (TBDMS) moieties (Wahl et al. 2004). Proteinogenic amino acids are predominantly synthesized during the growth phase; thus, the labeling pattern reflects cellular metabolism in the active growth phase. More reflective of the metabolism during hypoxic response would be the label pattern in free metabolites that are rapidly turned over (Toya et al. 2007). BSTFA [*N*,*O*-bis-(trimethylsilyl)trifluoroacetamide] is an attractive approach for isotopomer analysis of free metabolites in cancerous cells (Meadows et al. 2008). The label information in free metabolites can be used to capture the dynamic changes in metabolism when mycobacterium growth is under environmental perturbations.

Materials and methods

Growth and hypoxic stress of bacteria

Mycobacterium smegmatis was grown in Middlebrook 7H9 medium to middle growth-phase (OD about 1); then cells were transferred to a modified Sauton liquid medium (Dubos and Davis 1946) (inoculation volume 1:1,000) consisting of 99% labeled [2-^{13}C]glycerol (Cambridge Isotope, USA), 2.5 g/l; KH_2PO_4, 0.5 g/l; $MgSO_4$, 0.5 g/l; NH_4Cl, 3.24 g/l; 1% $ZnSO_4$, 0.1 ml/l and 20 mM PIPES at a final pH of 7.0. Two types of detergents, 0.05% Tween 80 or 0.025% non-hydrolyzable Tyloxapol (Sigma-Aldrich, neither labeled with ^{13}C), can be added into the initial medium to promote cell growth (Vandal et al. 2008). Cell density was monitored at 600 nm. Aerobic cultures were maintained at 37°C with shaking at 200 rpm in 250 ml flasks. To initiate a hypoxic stress, a 100 ml culture was grown under aerobic conditions to middle growth phase (OD_{600} ~1) and then transferred to a sealed glass bottle with a septum cap. The ratio of head-space volume to liquid volume was approximately 1:2. The hypoxic culture was grown at 37°C. The hypoxic culture was harvested 3.5 h after initiation of O_2 depletion. As a control, the aerobic culture of the same volume was collected from the shaking flask at the same time.

Extraction of metabolites

We modified a procedure for extracting free metabolites using methanol/chloroform (Winder et al. 2008) as follows. Cell suspensions (50 ml) from either hypoxic or aerobic cultures were plunged rapidly into an equal volume of 60% (v/v) methanol–water (pre-chilled on dry ice, pH = 7.5, 60 mM HEPES buffer), then centrifuged at 10,000×*g* (−10°C) for 5 min. The supernatant was quickly removed, and 5 ml cold 100% methanol (pre-chilled on dry ice) was added to the cell pellets. The cell/methanol mixture (setting on dry ice) was sonicated for 1.5 min with a 2 s on/3 s off cycle. The resulting cell lysate was added to a mixture of HPLC-grade chloroform (5 ml) and HPLC-grade water (2 ml), vortexed, and centrifuged for 2 min at 10,000×*g* at 4°C. The aqueous layer (top) was collected (~6 ml) and mixed with 7.5 ml HPLC-grade water. The solution was frozen with liquid N_2 and lyophilized overnight. We also prepared proteogenic amino acids from biomass grown under aerobic and hypoxic conditions. In brief, 10 ml cell culture was centrifuged, and the bacterial pellet was hydrolyzed (in 6 M HCl) at 100°C for 24 h before it was completely dried under air flow.

Metabolite derivatization and isotopomer analysis

To prepare bis(trimethylsilyl)trifluoroacetamide (Sigma-Aldrich) (BSTFA)-derivatized samples for GC-MS analysis, we pre-derivatized the dried samples with a solution (0.3 ml) of 2% (w/v) $NH_2OH \cdot HCl$ in pyridine (Sigma-Aldrich) (MOX solution) for 90 min at 40°C while shaking at 200 rpm. The MOX solution was then centrifuged for 2 min at 17,600×*g*, and 0.15 ml of the supernatant was derivatized by reaction with 0.25 ml BSTFA for 25 min at 40°C (shaking at 200 rpm). This derivatization step added trimethylsilyl groups to carboxyls and converted oxoacids (e.g. pyruvate) to oximes. One microliter

of the derivatized sample was injected into a gas chromatograph. The column was held at 60°C for 1 min after injection and then heated at 20°C/min to 130°C, 4°C/min to 150°C, and finally 40°C/min to 260°C, where it was held for 3 min (Meadows et al. 2008). The *N*-(tert-butyldimethylsilyl)-*N*-methyltrifluoroacetamide (TBDMS) was used to derivatize proteinogenic amino acids. The sample preparation for GC-MS analysis was performed as previously reported (Tang et al. 2007). All GC-MS measurements were carried out using a gas chromatograph (HP6890 series, Agilent Inc.) equipped with a DB5 column (J&W Scientific Inc.) and a mass spectrometer (5973 Network, Agilent Inc.). Because the natural abundance of isotopes alters the mass isotopomer spectrum in derivatized molecules, we corrected these changes using a published algorithm (Wahl et al. 2004).

Results and discussion

BSTFA metabolite derivatization

TBDMS-derivatized proteinogenic amino acids and fragment patterns have been studied extensively (Antoniewicz et al. 2007). On the other hand, BSTFA is a more sensitive method, which can also derivatize the functional groups –COO, –OH, –SH, and $–NH_2$.

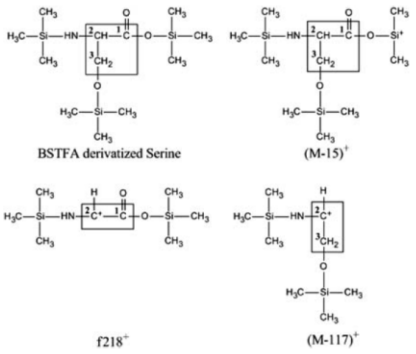


Fig. 1 BSTFA derivatized serine (serine C-skeleton was *boxed*) and its common ion fragments: fragments $(M - 15)^+$ ($–CH_3$ loss of TMS), $f218^+$ (loss of –R group from serine backbone) and $(M - 117)^+$ (loss of –COO and –TMS)

Thus, it can be used to determine intra-cellular free amino acids and organic acids as well as their labeling pattern (Meadows et al. 2008). Figure 1 shows an example of the structures of BSTFA-derivatized serine fragments in mass spectrum. Loss of a methyl group from the trimethylsilyl (TMS) group yields the fragment $(M - 15)^+$, which still contains the complete amino acid. The $(M - 117)^+$ ion originates from the loss of C(O)-O-TMS. The $f218^+$ results from cleavage of the bond between the R group and α-carbon atom of the amino acid. Using this BSTFA method, we identified 10 key metabolites in *M. smegmatis* extracts, including amino acids and several organic acids (Table 1).

Growth and metabolism of *M. smegmatis* under aerobic conditions

M. smegmatis is a fast-growing, non-virulent mycobacterium. When growing *M. smegmatis* in [^{13}C]-labeled glycerol minimal medium without any detergent (Tween 80 or Tyloxapol), cells formed large clumps. The doubling time was over 7 h, and the final OD_{600} was only 0.35 (Fig. 2). When the medium was supplemented with Tween 80 (or Tyloxapol), we observed its doubling time to be much faster (~3.5 h) and its final OD_{600} to be over 2 (for both Tween-grown or Tyloxapol-grown cultures). During aerobic growth, glycerol was completely utilized to produce the biomass, and no extracellular acetate or lactate was detected.

These findings indicate that *M. smegmatis* completely utilizes the majority of the carbon source (glycerol) for aerobic growth. It was noticed decades ago that *Mycobacterium* spp. all require the detergent supplemented in the medium for optimal growth (Mizuno and Tsukamura 1978). One reason for the faster growth is that oleic acid, the hydrolytic product of Tween 80 metabolism, can be converted into triacylglycerols as carbon sources for biosynthesis of lipid and other biomass building blocks (Saito et al. 1983). However, non-hydrolyzable Tyloxapol (Vandal et al. 2008) significantly enhanced growth under aerobic conditions with similar growth rate and final cell density compared to that of Tween 80 culture. Another possibility is that Tween 80 stimulates the growth by improving O_2 transfer to the cells rather than by providing extra nutrients or energy. When *M. smegmatis* grows in a detergent-free medium, the bacilli form large aggregates due to the extremely

Table 1 Labeling of free metabolites, derived with BSTFA, showing changes of labeling pattern under hypoxic condition ($n = 2$)

Synthesis from central pathways	Metabolites	Aerobic		Hypoxia for 3.5 h	
		$(M - 15)^+$ or $*f(218)^+$	$(M - 117)^+$	$(M - 15)^+$ or $f(218)^+$	$(M - 117)^+$
Glycolysis pathway	Ala[a]				
	M0	Overlapping peak	0.08 ± 0.02	Overlapping Peak	0.10 ± 0.01
	M1		0.89 ± 0.03		0.88 ± 0.01
	M2		0.02 ± 0.01		0.02 ± 0.01
	Pyruvate[b]				
	M0	0.02 ± 0.02	Overlapping peak	0.12 ± 0.02	Overlapping Peak
	M1	0.94 ± 0.03		0.82 ± 0.02	
	M2	0.04 ± 0.02		0.05 ± 0.01	
	M3	0.02 ± 0.01		0.01 ± 0.01	
	Gly				
	M0	0.10 ± 0.02	0.27 ± 0.01	0.26 ± 0.01	0.37 ± 0.01
	M1	0.86 ± 0.02	0.73 ± 0.01	0.71 ± 0.01	0.63 ± 0.01
	M2	0.03 ± 0.01	0	0.03 ± 0	0
	Ser[c]				
	M0	*0.09 ± 0.02**	0.09 ± 0.04	*0.15 ± 0.01**	0.17 ± 0.01
	M1	*0.88 ± 0.01**	0.83 ± 0.07	*0.81 ± 0.04**	0.69 ± 0.04
	M2	*0.03 ± 0.02**	0.08 ± 0.03	*0.04 ± 0.03**	0.13 ± 0.04
Pentose phosphate pathway	Phe[d]				
	M0	*0.10 ± .03**	0	*0.12 ± 0.03**	0
	M1	*0.85 ± 0.03**	0	*0.86 ± 0.04**	0
	M2	*0.04 ± 0.01**	0.06 ± 0.02	*0.01 ± 0.01**	0.08 ± 0.02
	M3	*0*	0.66 ± 0.04	*0*	0.64 ± 0.04
	M4	*0*	0.22 ± 0.03	*0*	0.21 ± 0.03
TCA cycle	Asp[e]				
	M0	Insufficient ions	0.11 ± 0.01	Insufficient ions	0.21 ± 0.02
	M1		0.63 ± 0.02		0.59 ± 0.01
	M2		0.25 ± 0.03		0.19 ± 0.02
	M3		0.01 ± 0.01		0.01 ± 0.01
	Glu				
	M0	0.01 ± 0	0.03 ± 0.01	0.03 ± 0.01	0.07 ± 0.02
	M1	0.15 ± 0.03	0.24 ± 0.04	0.23 ± 0.03	0.39 ± 0.04
	M2	0.63 ± 0.04	0.69 ± 0.05	0.52 ± 0.04	0.49 ± 0.04
	M3	0.21 ± 0.03	0.05 ± 0.02	0.18 ± 0.03	0.04 ± 0.02
	Citrate[f]				
	M0	Insufficient ions	0.38 ± 0.05	Insufficient ions	0.58 ± 0.07
	M1		0.21 ± 0.04		0.23 ± 0.06
	M2		0.29 ± 0.04		0.14 ± 0.02
	M3		0.11 ± 0.02		0.04 ± 0.01
	M4		0.01 ± 0		0.01 ± 0

Table 1 continued

Synthesis from central pathways	Metabolites	Aerobic		Hypoxia for 3.5 h	
		$(M - 15)^+$ or $*f(218)^+$	$(M - 117)^+$	$(M - 15)^+$ or $f(218)^+$	$(M - 117)^+$
TCA cycle	Succinate[g]				
	M0	0.05 ± 0.01	Not fragmented	0.12 ± 0.02	Not fragmented
	M1	0.22 ± 0.03		0.40 ± 0.04	
	M2	0.66 ± 0.04		0.41 ± 0.03	
	M3	0.05 ± 0.01		0.06 ± 0.03	

Note: The final isotopomer data were reported as *M0* unlabeled fraction, *M1* singly labeled fraction, *M2* fraction with two labeled carbons, *M3* fraction with labeled three carbons, etc. $*f(218)^+$ peak is used instead of $(M - 15)^+$. (1) Alanine's $(M - 15)^+$ peak overlaps the $f(218)^+$ peak. (2) Pyruvate's $(M - 117)^+$ peak overlaps other anomalous ion peaks. (3) Serine's $f(218)^+$ fragment is better for quantitative purposes than $(M - 15)^+$ due to high signal to noise ratio in the latter. (4) Phenylalanine's $f(218)^+$ fragment is better for qualitative purposes used for these amino acids than $(M - 15)^+$ due to high signal to noise ratio in the latter. (5) Aspartic acid's $(M - 15)^+$ and $f(218)^+$ peak have high background noise. Not recommended to be used. (6) Citrate $(M - 117)^+$ fragment is better for quantitative purposes than $(M - 15)^+$ due to high signal to noise ratio in the latter. (7) Succinate's $(M - 117)^+$ ion is not present, i.e., MS cannot fragment succinate

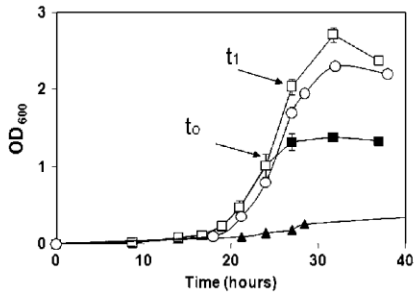


Fig. 2 *M. smegmatis* growth. The symbol (*filled triangle*): the growth curve without any detergent. The symbols (*filled square* and *empty squares*): the growth with Tween 80. The *empty squares*: growth under aerobic conditions. The *filled squares*: growth under hypoxic stress. The symbol (*empty circle*): the growth with detergent Tyloxapol. t_0: start of hypoxia, the culture was transferred to the serum bottle; t_1: biomass samples were taken for isotopomer analysis

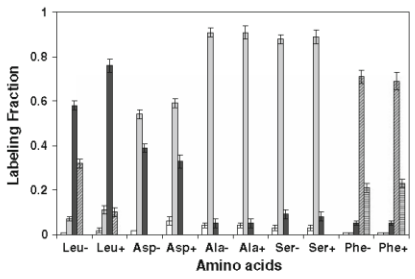


Fig. 3 Effect of Tween 80 on proteinogenic amino acid labeling. The *bars* from *left* to *right* represent isotope peaks of *M0* unlabeled fraction, *white bar*; *M1* singly labeled fraction, *gray bar*; *M2* fraction with two labeled carbons, *dark bar*; *M3* fraction with labeled three carbons, *diagonal hash*; *M4* fraction with labeled four carbons, *horizontal hash*. (−) Indicates the culture without any detergent; (+) indicates the culture with Tween 80. The $(M - 57)^+$ fragment was used for Asp, Ala, Ser, and Phe. The $(M - 159)^+$ fragment was used for Leu due to its overlapping peak with $(M - 57)^+$

waxy nature of the cell surface. Therefore, the O_2 availability in the center of such a clump may be low.

We compared the isotopomer labeling patterns of five key proteinogenic amino acids produced from cultures grown in Tween 80-containing and Tween 80-free media (Fig. 3). Specifically, precursors to alanine (pyruvate) and leucine (acetyl-CoA) are key metabolites bridging glycolysis and the TCA cycle, while the precursors to aspartic acid (oxaloacetate), serine (3-phosphoglycerate), and phenylalanine (phospho*enol*pyruvate and erythrose 4-phosphate) are derived from the TCA cycle, glycolysis, and pentose phosphate pathway, respectively. Therefore, if unlabeled Tween 80 was metabolized via β-oxidation (Muñoz-Elías and McKinney 2006), it would be incorporated into these amino acids and change their labeling patterns. As shown in Fig. 3, addition of Tween 80 caused measurable dilution of the label in

leucine (the fractions of the M3 isotopes in leucine were moderately deceased). Because acetyl-CoA and pyruvate (synthesis from glycerol) are precursors for leucine synthesis, the change in its label pattern confirmed previous report that unlabeled Tween 80 could be oxidized to acetyl-CoA (Höner zu Bentrup and Russell 2001; Schnappinger et al. 2003) and thus decreased the labeling concentrations in leucine. The label in aspartic acid was minimally diluted (up to 5%, just above the measurement noise), which indicated that only a small fraction of unlabeled acetyl-CoA from Tween 80 entered the TCA cycle. Amino acids derived from glycolysis and the pentose phosphate pathway (e.g., alanine, serine, and phenylalanine) did not show any changes in their label patterns. Thus, acetyl-CoA derived from Tween-80 did not enter the aforementioned pathways. These results show that only small amount of Tween 80 is utilized for synthesis of biomass building blocks under aerobic condition. Tween 80 or non-hydrolyzable Tyloxapol enhances mycobacterium aerobic growth mainly because it enhances O_2 transfer.

Growth and metabolism of *M. smegmatis* adapted to hypoxia

Previous studies showed that *M. smegmatis* can adapt to the non-replicating persistence state with growth kinetics and antibiotic resistance similar to *M. tuberculosis* under O_2-limited culture (Dick et al. 1998). In this study, the hypoxic condition was established by gradually depriving an actively Tween 80-growing culture of O_2 in a sealed glass bottle, and thus cell growth eventually stopped within 3 h (Fig. 2). Isotopomer analysis of proteinogenic amino acids derived from biomass harvested after 3 h of hypoxia did not show a significant difference in the labeling patterns of the hypoxic cells compared with those of proteins extracted from cells in the aerobic, active growth phase (data not shown), because proteinogenic amino acids from the O_2-deprived culture were derived primarily from the aerobic growth stage. To capture the transition in metabolism from aerobic to hypoxic conditions, we derivatized the free metabolites with BSTFA and analyzed their isotopic label patterns.

As shown in Table 1, the TCA cycle acids, i.e. citrate and succinate, had the highest label dilution (e.g., the M0 peak of citrate increased from 38% to 58%), indicating that a large fraction of unlabeled carbon entered central metabolism (compared to the aerobic condition) and caused a significant decrease in the isotope labeling of TCA cycle acids. These unlabeled metabolites in the TCA cycle originated from unlabeled hydrolysis products of Tween 80. Second, the isotopically labeled fractions of aspartate and glutamate from the hypoxic culture were lower compared with those from the aerobic culture (the unlabeled M0 peak of Asp increased from 11% to 21%, and that of Glu from 3% to 7%). Under hypoxic conditions, the two amino acids were either synthesized from oxaloacetate and 2-ketoglutarate, respectively, or from the degradation of protein grown under aerobic conditions, and as such, the change in labeling in amino acids was less significant than fast turnover TCA cycle intermediates (e.g., citrate). Lastly, we did not observe a dramatic change in the label patterns of free alanine and phenylalanine, which indicated that the Tween 80 catabolites were not used to synthesize their precursors (e.g., pyruvate, phospho*enol*pyruvate, and erythrose 4-phosphate) under hypoxia. Taken together, the isotopomer data of free metabolites indicate that catabolism of Tween 80 was relatively enhanced compared to that of glycerol so as to serve as a secondary source of carbon assimilated mainly via the TCA cycle under hypoxic stress. This finding confirmed the transcript studies showing that expression of genes encoding enzymes involved in fatty acid oxidation are up-regulated under O_2 limited conditions (Muñoz-Elías and McKinney 2006; Schnappinger et al. 2003).

Glyoxylate and glycine metabolism in *M. smegmatis*

The glyoxylate metabolism plays a major role in enabling mycobacteria to utilize carbon sources when TCA cycle is shut down, particularly in infected tissues where O_2 and nutrients are limited (Boshoff and Barry 2005). The activity of isocitrate lyase, the first enzyme in the glyoxylate pathway, has been shown to increase in hypoxic culture of *M. tuberculosis* (Muñoz-Elías and McKinney 2005; Muttucumaru et al. 2004). Interestingly, the activity of glycine dehydrogenase (GDH), which channels glyoxylate out of the cycle, was detected in the cell extracts of *M. tuberculosis* by biochemical assays (Boshoff and Barry 2005). GDH catalyzes the reductive conversion of glyoxylate to glycine (concomitantly oxidizing NADH to NAD^+) and was found to be

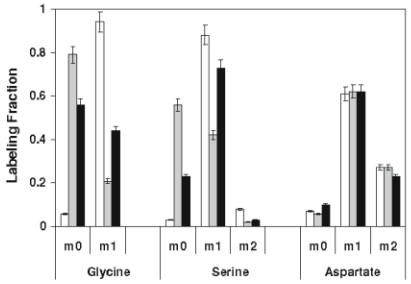


Fig. 4 Effect of addition of glycine and glyoxylate (100 mg/l) on the isotopomer distribution in proteinogenic amino acids. The $(M - 57)^+$ fragment was used. *White bar* (control), *gray bar* (addition of glycine), *black bar* (addition of glyoxylate)

strongly activated in microaerophilic cultures of *M. tuberculosis*.

Here we revisited the in vivo activity of GDH by using ^{13}C isotopomer labeling. In order to test whether GDH is active in aerobically grown bacilli, we supplemented the labeled culture medium with unlabeled glyoxylate or glycine (100 mg/l) and analyzed the label patterns of proteinogenic amino acids. A marked change in the labeling of glycine and serine was observed but not in any other amino acids measured (Fig. 4). After addition of unlabeled glyoxylate, the unlabeled fraction (M0) of proteinogenic glycine increased from 5% to 57%, which indicates that over 50% of glycine in the biomass was derived from the unlabeled glyoxylate. A dilution of the label in serine, which was less significant than that in glycine, suggested that glycine could be a precursor for serine synthesis. Therefore, the reactions glyoxylate → glycine → serine is an alternate pathway for glycine and serine biosynthesis in *M. smegmatis* regardless of the O_2 concentration. Under hypoxic conditions, glyoxylate synthesis is induced because isocitrate lyase activity (isocitrate → glyoxylate + succinate) is up-regulated (Muñoz-Elías and McKinney 2005). Thus, the flux through the pathway from glyoxylate to glycine and serine would be expected to have a concurrent increase. Indeed, the label in free glycine and serine were dramatically diluted (>10%, Table 1) under hypoxic conditions even without addition of unlabeled glyoxylate to the medium. Such dilution resulted primarily from a large fraction of acetyl-CoA (derived from unlabeled Tween 80)

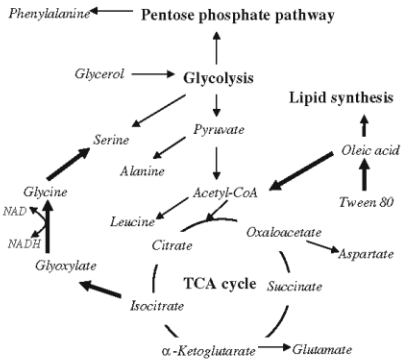


Fig. 5 Central metabolic pathways of *M. smegmatis* revealed by isotopomer analysis. The *thick arrows* indicate the pathways up-regulated under hypoxia. The metabolites are in *Italic*

entering the TCA cycle. Since the alternate pathway for glycine and serine synthesis was up-regulated (i.e., isocitrate → glyoxylate → glycine → serine), the unlabeled carbon was thus introduced into glycine and serine from the TCA cycle. Figure 5 summarizes the central metabolic pathways of *M. smegmatis* and highlights all the up-regulated pathways (thick arrows) under hypoxic stress. It is noteworthy that the glyoxylate → glycine → serine pathway, with concomitant oxidation of NADH to NAD^+, may be of particular importance for bacilli responding to O_2 shortage in the non-replicating persistence state by facilitating lipid utilization and maintaining the redox balance.

Conclusion

Isotopomer-assisted metabolite analysis revealed change of Tween 80 metabolism under O_2 rich and O_2 poor conditions. It also demonstrated an increase in the flux through the pathway that transforms glyoxylate to serine (glyoxylate → glycine → serine) under hypoxic conditions, which might contribute to the bacillus' survival and persistence in the O_2-limited host tissue. The impact of hypoxia on mycobacterial metabolism and growth has been studied using other molecular biology tools before. The isotopomer-assisted metabolite analysis described herein is an

alternate and complementary tool to these previously-used technologies, and thus it may have broad application in other complicated biological systems.

Acknowledgments We thank Yisheng Kang and Mario Ouellet for their advice. This work was funded in part by the Virtual Institute for Microbial Stress and Survival and in part of the Joint BioEnergy Institute. This work was also funded by I-CARES (International Center for Advanced Renewable Energy and Sustainability) and the Missouri Life Sciences Research Board at Washington University in St Louis.

References

Antoniewicz MR, Kelleher JK, Stephanopoulos G (2007) Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. Anal Biochem 79: 7554–7559

Boshoff HI, Barry CER (2005) Tuberculosis—metabolism and respiration in the absence of growth. Nat Rev Microbiol 3:70–80

Dick T, Lee BH, Murugasu-Oei B (1998) Oxygen depletion induced dormancy in *Mycobacterium smegmatis*. FEMS Microbiol Lett 163:159–164

Dubos RJ, Davis BD (1946) Factors affecting the growth of *Tubercle bacilli* in liquid media. J Exp Med 83:409–423

Höner zu Bentrup K, Russell DG (2001) Mycobacterial persistence: adaptation to a changing environment. Trends Microbiol 9:597–605

Meadows AL, Kong B, Berdichevsky M, Roy S, Rosiva R, Blanch HW, Clark DS (2008) Metabolic and morphological differences between rapidly proliferating cancerous and normal breast epithelial cells. Biotechnol Prog 24: 334–341

Mizuno S, Tsukamura M (1978) Utilization of tween 80 as carbon source for growth of slowly growing *mycobacteria*. Kekkaku 53:537–540

Muñoz-Elías EJ, McKinney JD (2005) *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. Nat Med 11:638–644

Muñoz-Elías EJ, McKinney JD (2006) Carbon metabolism of intracellular bacteria. Cell Microbiol 8:10–22

Muttucumaru DG, Roberts G, Hinds J, Stabler RA, Parish T (2004) Gene expression profile of *Mycobacterium tuberculosis* in a non-replicating state. Tuberculosis (Edinb) 84:239–246

Saito H, Tomioka H, Watanabe T, Yoneyama T (1983) Mycobacteriocins produced by rapidly growing mycobacteria are Tween-hydrolyzing esterases. J Bacteriol 153:1294–1300

Schnappinger D, Ehrt S, Voskuil MI, Liu Y, Mangan JA, Monahan IM, Dolganov G, Efron B, Butcher PD, Nathan C, Schoolnik GK (2003) Transcriptional adaptation of *Mycobacterium tuberculosis* within macrophages: insights into the phagosomal environment. J Exp Med 198:693–704

Tang YJ, Pingitore F, Mukhopadhyay A, Phan R, Hazen TC, Keasling JD (2007) Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using GC-MS and FT-ICR mass spectrometry. J Bacteriol 189:940–949

Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EK, Keasling JD (2009) Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. Mass Spectrom Rev 28:362–375

Toya Y, Ishii N, Hirasawa T, Naba M, Hirai K, Sugawara K, Igarashi S, Shimizu K, Tomita M, Soga T (2007) Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis. J Chromatogr A 1159:134–141

Vandal OH, Pierini LM, Schnappinger D, Nathan CF, Ehrt S (2008) A membrane protein preserves intrabacterial pH in intraphagosomal *Mycobacterium tuberculosis*. Nat Med 14:849–854

Wahl SA, Dauner M, Wiechert W (2004) New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnol Bioeng 85:259–268

Wayne LG, Sohaskey CD (2001) Nonreplicating persistence of *Mycobacterium tuberculosis*. Annu Rev Microbiol 55: 139–163

Winder CL, Dunn WB, Schuler S, Broadhurst D, Jarvis R, Stephens GM, Goodacre R (2008) Global metabolic profiling of *Escherichia coli* cultures: an evaluation of methods for quenching and extraction of intracellular metabolites. Anal Chem 80:2939–2948

Appendix 7

Correlation of Genomic and Physiological Traits of *Thermoanaerobacter* Species with Biofuel Yields

APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Nov. 2011, p. 7998–8008 Vol. 77, No. 22
0099-2240/11/$12.00 doi:10.1128/AEM.05677-11

Correlation of Genomic and Physiological Traits of *Thermoanaerobacter* Species with Biofuel Yields▽†

Christopher L. Hemme,[1] Matthew W. Fields,[2] Qiang He,[3,4] Ye Deng,[1] Lu Lin,[1,5] Qichao Tu,[1] Housna Mouttaki,[1,6] Aifen Zhou,[1] Xueyang Feng,[7] Zheng Zuo,[7] Bradley D. Ramsay,[2] Zhili He,[1] Liyou Wu,[1] Joy Van Nostrand,[1] Jian Xu,[5] Yinjie J. Tang,[6] Juergen Wiegel,[8] Tommy J. Phelps,[9] and Jizhong Zhou[1,10]*

Institute for Environmental Genomics, University of Oklahoma, Norman, Oklahoma[1]; Department of Microbiology, Montana State University, Bozeman, Montana[2]; Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, Tennessee[3]; Center for Environmental Biotechnology, University of Tennessee, Knoxville, Tennessee[4]; Qingdao Institute of Bioenergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, China[5]; Institute of Groundwater Ecology, Helmholtz Zentrum München, German Research Center for Environment and Health, Munich, Germany[6]; Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Missouri[7]; Department of Microbiology, University of Georgia, Athens, Georgia[8]; Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee[9]; and Earth Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California[10]

Received 18 May 2011/Accepted 15 September 2011

Thermophilic anaerobic noncellulolytic *Thermoanaerobacter* species are of great biotechnological importance in cellulosic ethanol production due to their ability to produce high ethanol yields by simultaneous fermentation of hexose and pentose. Understanding the genome structure of these species is critical to improving and implementing these bacteria for possible biotechnological use in consolidated bioprocessing schemes (CBP) for cellulosic ethanol production. Here we describe a comparative genome analysis of two ethanologenic bacteria, *Thermoanaerobacter* sp. X514 and *Thermoanaerobacter pseudethanolicus* 39E. Compared to 39E, X514 has several unique key characteristics important to cellulosic biotechnology, including additional alcohol dehydrogenases and xylose transporters, modifications to pentose metabolism, and a complete vitamin B_{12} biosynthesis pathway. Experimental results from growth, metabolic flux, and microarray gene expression analyses support genome sequencing-based predictions which help to explain the distinct differences in ethanol production between these strains. The availability of whole-genome sequence and comparative genomic analyses will aid in engineering and optimizing *Thermoanaerobacter* strains for viable CBP strategies.

Global energy demands will increase significantly in coming decades (16), and renewable energy sources such as biofuels have been proposed to help reduce dependence upon fossil energy (8, 23, 27, 49). Current efforts focus on biofuel production from renewable lignocellulosic feedstock (e.g., switchgrass), which constitutes ~50% of the world's biomass (2, 8, 23, 27, 49). Although intensive research and development have been performed on the effective utilization of lignocellulose, problems associated with practical use of this material have not been resolved fully (13). When enzymatic hydrolysis is adopted for cellulosic ethanol production, different levels of process integration can be visualized: (i) separate (or sequential) hydrolysis and fermentation (SHF), where the enzymes (cellulases) are used separately from fermentation tanks; (ii) simultaneous saccharification and fermentation (SSF), which consolidates enzymatic hydrolysis with fermentation of hexose or pentose; (iii) simultaneous saccharification and cofermentation (SSCF), which further combines the fermentation of hexose and pentose together; and (iv) consolidated bioprocessing (CBP), where all required enzymes and ethanol are produced in a single reactor (1, 13, 24, 29, 46, 51). While these production schemes represent increasing levels of simplification through process consolidation, consolidation of multiple steps often results in a loss of process efficiency. Thus, improving the efficiency of individual steps, such as cellulose hydrolysis and ethanol fermentation, remains an important task for the development of economically feasible cellulosic bioethanol.

Recent efforts have focused on metabolic engineering and (more recently) synthetic biology to produce strains or consortia capable of producing biofuels. Efforts to engineer ethanologenic strains optimized for specific process conditions require critical knowledge of the genetic and metabolic determinants of the divergent characteristics of ethanol production, which are best illustrated by cultures of thermophilic ethanologenic *Thermoanaerobacter* species and *Clostridium thermocellum*. *C. thermocellum* is a thermophilic, ethanol-producing bacterium capable of cellulose and hemicellulose decomposition into pentose and hexose, but it ferments hexose to ethanol with

* Corresponding author. Mailing address: Institute for Environmental Genomics, Department of Botany and Microbiology, University of Oklahoma, Norman, OK 73019. Phone: (405) 325-6073. Fax: (405) 325-3442. E-mail: jzhou@ou.edu.

† Supplemental material for this article may be found at http://aem.asm.org/.

▽ Published ahead of print on 23 September 2011.

only very low yields (12, 52). In contrast, *Thermoanaerobacter* species efficiently convert both pentose and hexose to ethanol at high yields but cannot degrade cellulose. Previous studies in our laboratory have shown that stable cellulolytic cocultures of *Thermoanaerobacter* and *C. thermocellum* can be established for more efficient simultaneous cellulose degradation and ethanol production (J. Zhou, unpublished results). Such cocultures are a possible solution for consortium-based CBP schemes (8, 18, 28). Interestingly, ethanol yields in these cocultures vary greatly depending on which *Thermoanaerobacter* species is present. Physiological analyses have shown that *Thermoanaerobacter* sp. X514 (33) produces more ethanol in cocultures with selected *C. thermocellum* strains than the corresponding *Thermoanaerobacter pseudethanolicus* 39E (30) cocultures, despite a very close phylogenetic relationship between the two strains (9, 18). To provide insight into the molecular mechanisms underlying differences in sugar utilization and ethanol production, physiological, comparative, and functional genomic analyses were conducted with X514 and 39E. The results indicate key differences in the mechanisms involved in vitamin B_{12} biosynthesis, xylose transport and metabolism, and ethanol production, which may explain the phenotypic differences in ethanol yields observed between these two strains. These results will aid efforts to engineer *Thermoanaerobacter* strains for efficient cellulosic biofuel production in coculture-based CBP schemes.

MATERIALS AND METHODS

Whole-genome sequencing and analysis. (i) Organism source and culturing. Strain 39E (ATCC 33223) was obtained from the American Type Culture Collection (ATCC). A laboratory culture of X514 was used for sequencing, and the culture has been deposited in the ATCC as strain ATCC BAA-938. *Thermoanaerobacter* strains were cultured as described previously, using DCB-1 medium (33).

(ii) Sequencing and annotation. 39E and X514 were sequenced and annotated as previously described (15). All general aspects of library construction and sequencing performed at the Joint Genome Institute (JGI) can be found at http://www.jgi.doe.gov/. Additional manual curation was conducted at the University of Oklahoma, based on literature searches and comparative sequence analysis.

(iii) Sequence analysis. The majority of sequence analysis, including metabolic pathway analysis and ortholog detection, was conducted using online tools at the Joint Genome Institute Integrated Microbial Genomes database (26; http://img.jgi.doe.gov/cgi-bin/w/main.cgi) or was done manually using *ad hoc* Perl scripts. Putative lateral gene transfer events were detected by SIGI-HMM analysis (47).

Growth analysis. (i) Cellulose measurement. Cellulose degradation was measured as previously described (45). Briefly, 1-ml samples were harvested at the stationary phase of growth by centrifugation at 5,000 × *g* for 15 min. The pellet was suspended in distilled water and heated at 100°C for 30 min to achieve cell lysis. The residual cellulose was washed with distilled water two times by centrifugation at 5,000 × *g* for 15 min and then hydrolyzed into soluble sugars with 65% H_2SO_4. The hydrolysis samples were diluted 200 times, and soluble sugars were assayed by using the phenol-sulfuric acid method, with glucose as a standard.

(ii) Phenol-sulfuric acid method for measurement of total sugars. Glucose (25, 50, 75, 100, and 150 mg/liter) was used as the standard. Samples (200 μl) were diluted to the appropriate concentrations and mixed with 200 μl 5% phenol and 1 ml of 98% H_2SO_4. The mixtures were incubated at room temperature for 30 min, and the absorbances of the standards and samples were measured at 490 nm.

(iii) Ethanol production measured by HPLC. One-milliliter samples were harvested at the stationary phase of growth by centrifugation at 5,000 × *g* for 15 min. The supernatant was used to analyze the concentration of ethanol by use of a high-performance liquid chromatography (HPLC) apparatus (Agilent Technologies, CA) operating at 55°C. The mobile phase consisted of 0.025% sulfuric acid at a flow rate of 0.6 ml/min. Ethanol yields with different concentrations of exogenous vitamin B_{12} were compared between the two strains, using the one-tailed paired *t* test (34).

Metabolic flux analysis. (i) Biomass preparation. *Thermoanaerobacter* sp. X514 and 39E were grown anaerobically at 60°C without shaking in DCB-1 medium (33). The strains were initially grown in a 50-ml standard culture medium with an unlabeled carbon source (xylose). At the mid-log phase of growth, a 3% inoculum was added to a 50-ml culture containing 2 g/liter of [1-^{13}C]xylose. At the mid-log phase of growth of this culture, a 3% inoculum from the first ^{13}C-labeled culture medium was used to inoculate a 50-ml subculture (with the same labeled carbon source) in order to reduce the effect of unlabeled carbon from the initial stock.

Biomass was harvested at the late log phase of growth by centrifugation at 8,000 × *g* for 15 min at 10°C. The concentrations of glucose, xylose, acetate, ethanol, and lactate were analyzed with a high-performance liquid chromatography apparatus (Agilent Technologies, CA) equipped with a variable-wavelength (190 to 600 nm) detector (VWD) (with UV absorption measured at 245 nm) and an ion-exclusion column (Aminex HPX-87H; 300 mm × 7.8 mm; Bio-Rad Laboratories, CA) operating at 55°C. The mobile phase consisted of 0.025% sulfuric acid at a flow rate of 0.6 ml/min.

(ii) Flux estimation. Isotopic analysis of proteogenic amino acids was performed (42, 43). Biomass was hydrolyzed in 6 M HCl at 100°C. The amino acid solution was dried and then derivatized in tetrahydrofuran and *N*-(*tert*-butyl dimethylsilyl)-*N*-methyl-trifluoroacetamide at 70°C for 1 h. A gas chromatograph (GC) (model 7890A; Hewlett-Packard, Agilent Technologies, CA) equipped with a DB5-MS column (J&W Scientific, Folsom, CA) and a mass spectrometer (MS) (model 5975C; Agilent Technologies, CA) was used to analyze amino acid labeling profiles. The $[M - 57]^+$ MS group (containing unfragmented amino acids) was used for flux calculations (48). Central metabolic pathway maps of X514 and 39E included the glycolysis pathway, the pentose phosphate pathway, the Entner-Doudoroff (ED) pathway, the citric acid cycle, and the anaplerotic pathway. The fluxes to extracellular metabolites (ethanol, acetate, and lactate) were measured using HPLC. The standard deviations of fluxes were based on errors in concentration measurements of extracellular metabolites. The fluxes toward biomass synthesis were estimated based on dry biomass weight and biomass compositions (assumed to be the same as those of *Escherichia coli*). A ^{13}C flux analysis model (42) was developed using MATLAB 7.0 (The Mathworks, Natick, MA). Metabolic fluxes were constrained by mass balance as follows: $Sv = 0$, where S was the stoichiometry matrix. Fluxes (v) were calculated using isotopomer information from amino acids. The following grid search algorithm was applied to find a set of flux distributions that minimized the difference (ε) between the experimentally observed and predicted isotopomer labeling patterns in five key amino acids (see Table S4 in the supplemental material):

$$\varepsilon(v_n) = \sqrt{\sum_{i=1}^{k} \left(\frac{M_i - N_i(v_n)}{\delta_i} \right)^2},$$

where v_n were the unknown fluxes to be optimized in the program, M_i were the measured MS data, N_i were the corresponding model-simulated MS data, and δ_i were the corresponding standard deviations in the GC-MS data.

Global transcriptomic analysis. (i) Oligonucleotide probe design and microarray construction. Microarrays for 39E and X514 were constructed with 70-mer nucleotide probes. The microarray for 39E contains 2,334 nucleotide probes, covering 2,261 of 2,362 annotated gene sequences. The microarray for X514 contains 2,315 probes, covering 2,322 of 2,475 annotated gene sequences. Gene-specific, inclusive and exclusive group-specific oligonucleotide probes were designed by a new version of the computer program CommOligo (22), based on the following criteria: (i) for gene-specific probes, ≤85% sequence identity, a ≤20-base continuous stretch, and ≥40 kcal/mol of free energy; and (ii) group-specific probes had to meet the above requirements for nontarget groups and also had to have ≥96% sequence identity, a ≥55-base continuous stretch, and ≤90 kcal/mol of free energy within the group.

All designed oligonucleotides were synthesized commercially without modification by MWG Biotech Inc. (High Point, NC). The concentration of oligonucleotides was adjusted to 100 pmol/μl. Oligonucleotide probes were prepared in 50% (vol/vol) dimethyl sulfoxide (Sigma) and spotted onto UltraGAPS glass slides by use of a BioRobotics Microgrid II microarrayer (Genomic Solutions, Ann Arbor, MI). Each oligonucleotide probe had two replicates on a single slide. In addition, six different concentrations (11, 22, 45, 90, 180, and 360 ng/μl) of genomic DNA were spotted (eight replicates for each of the six concentrations on a single slide) as additional positive controls. After the oligonucleotide probes were printed, they were fixed onto the slides by UV cross-linking (600 mJ of energy).

(ii) Culture growth for microarray analysis. X514 and 39E were grown at 60°C in defined anaerobic DCB-1 medium as described above. Cysteine-HCl (0.005 g/liter) was added as the reductant after the medium had been boiled and cooled to room temperature and the pH value adjusted to 7.0. Xylose at different concentrations (1 mM, 10 mM, 50 mM, and 200 mM) was subsequently added to the autoclaved medium from filter-sterilized anaerobic stock solutions. When X514 and 39E cultures were grown to the mid-log phase, cell samples of each culture were harvested by centrifugation (7,000 × *g*) for 10 min at 4°C. Cell pellets were then immediately frozen in liquid N_2 and stored at −80°C. All experiments were conducted in triplicate.

(iii) Total RNA extraction, purification, and labeling. Total cellular RNAs from 39E and X514 were isolated using TRIzol reagent (Invitrogen) according to the manufacturer's protocol. RNA samples were purified by use of an RNeasy minikit (Qiagen), and RNA samples were treated with on-column RNase-free DNase I (Qiagen) to digest genomic DNA. RNA samples were quantified by Nanodrop ND-1000 spectrophotometry at wavelengths of 260 and 280 nm. To generate labeled cDNA targets with reverse transcriptase, 10 μg purified total RNA was used in each reaction mix, using a previously described protocol (44). Briefly, 10 μg random primers (3.0 μg/μl; Invitrogen) was used for priming, and the fluorophore Cy5-dUTP was used for labeling. After labeling, cDNA was purified with a Qiagen PCR purification kit and concentrated with an SPD1010 Speedvac apparatus. The efficiency of labeling was measured by Nanodrop ND-1000 spectrophotometry.

(iv) Genomic DNA labeling. To facilitate data comparison, genomic DNAs from X514 and 39E were labeled separately with Cy3-dUTP, which was used as the control. Each 50-μl labeling reaction mixture contained 500 ng genomic DNA, 20 μl 2.5× random primers (Invitrogen), a 5 mM concentration (each) of dATP, dGTP, and dCTP, 2.5 mM dTTP, 2 μl Klenow polymerase, and 0.5 mM Cy3-dUTP. The labeling reaction proceeded for 3 h at 37°C, and the labeled DNA probe was purified by use of a Qiagen PCR purification kit and concentrated in an SPD1010 Speedvac apparatus (5).

(v) Microarray hybridization and scanning. To hybridize microarray glass slides, the Cy5-dUTP-labeled cDNA targets were mixed with the Cy3-dUTP-labeled genomic DNA and dissolved in a hybridization solution that contained 50% formamide, 5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 10% SDS, and 0.1 mg/ml herring sperm DNA. The labeled sample was incubated for 98°C for 3 min to denature the probe and then kept at 65°C and applied to microarray slides. Hybridization, washing, and drying were all carried out by a Tecan HS4800 Pro hybridization station (U.S. Tecan, Durham, NC). The slides were scanned for the fluorescence intensities of both the Cy5 and Cy3 fluorophores by a ScanArray Express microarray analysis system (Perkin Elmer, Boston, MA).

(vi) Image quantification and data analysis. To determine the signal intensity of fluorescence for each spot, 16-bit scanned TIFF images were analyzed by ImaGene 6.1 software (Biodiscovery Inc., El Segundo, CA) to quantify spot signal, spot quality, and background fluorescence intensities. Data preprocessing and analysis included the following major steps. (i) Empty spots (with a signal-to-noise ratio [SNR] of <2.0), poor spots, and negative spots were flagged according to the instructions of the software and were removed in the subsequent analysis. (ii) The net signal of each spot was calculated by subtracting the background signal and adding a pseudosignal of 100 to obtain a positive value. If the resulting net signal was less than 50, a value of 50 was used. (iii) The total signal intensity for all microarrays (slides) in this experiment was normalized with the genomic DNA signal (Cy3 signal), and a normalization factor was calculated for each slide. (iv) Both Cy3 and Cy5 signal intensities of each spot were adjusted by multiplying by the normalization factor calculated above, and the resultant Cy5 signal of each probe presents the normalized signal intensity for each gene. (v) A ratio (R) of Cy5/Cy3 signals was calculated for each spot on each array. (vi) For comparative genomic analysis, the control and treatment conditions were defined so that the ratios R_c and R_t were defined for each gene under control and treatment conditions, respectively. The $\log_2(R_t/R_c)$ value was then calculated for each gene. (vii) Finally, to assess the significance of gene expression with the treatment, a Z score was calculated by using the following equation: $Z = \log_2(R_t/R_c)/\sqrt{0.25 + \sum\text{variance}}$, where 0.25 is a pseudovariance term. Typically, a Z score cutoff of >2.0 was used for significant changes.

Accession numbers. The sequences and annotations of the complete *T. pseudethanolicus* 39E and *Thermoanaerobacter* sp. X514 genomes are available under GenBank accession numbers CP000924 and CP000923, respectively. (*Caldanaerobacter subterraneus* subsp. *tengcongensis* MB4 was previously deposited under the name *Thermoanaerobacter tengcongensis* MB4, under accession number AE008691, but is no longer considered the prototypical *Thermoanaerobacter* genome.) All sequences are also available at the JGI-IMG website (26; http://img.jgi.doe.gov/cgi-bin/w/main.cgi). Microarray results are available through the NCBI GEO site under series accession numbers GSE31521 and GSE31522.

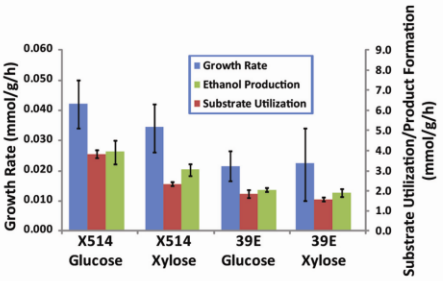


FIG. 1. Physiological comparison of X514 and 39E grown on glucose and xylose and of corresponding ethanol yields. Error bars represent errors for three replicates. The samples from which these data were derived were used for flux analyses as described in the main text.

RESULTS

Strain X514 produces more ethanol from xylose than does strain 39E. *T. pseudethanolicus* 39E was originally isolated from Octopus Spring in Yellowstone Park (30), a relatively nutritionally stable environment, whereas X514 was isolated from the presumably nutrient-poor deep subsurface (33) (see Table S1 in the supplemental material). Though the growth rates on xylose were not significantly different (0.022 ± 0.012 mmol/g/h for 39E and 0.034 ± 0.008 mmol/g/h for X514), X514 exhibited higher rates of substrate utilization (1.5 ± 0.1 mmol/g/h for 39E and 2.3 ± 0.1 mmol/g/h for X514) and higher ethanol yields (1.9 ± 0.2 mmol/g/h for 39E and 3.0 ± 0.3 mmol/g/h for X514) than 39E when the strains were grown on xylose (Fig. 1; see Fig. S1). Multiple combinations of cocultures using different *Thermoanaerobacter* and *C. thermocellum* strains showed that the combination of X514 and *C. thermocellum* LQRI gave the best results for cellulose degradation rates and ethanol yields. Average nucleotide identity (ANI) (19) values suggest that X514 and 39E are likely members of the same species (97.6% ANI; a 97% ANI cutoff is used to define species), and thus they are expected to show similar but not identical genome structures and metabolisms. Thus, the genomic differences exhibited by the two strains could potentially correlate with the phenotypes observed in mono- and coculture. The completed genomes of the two strains were thus compared to identify genetic differences which could manifest as the observed phenotypic differences in substrate utilization and product formation (15).

Energy metabolism and cofactor biosynthesis. Cellular energetics in *Thermoanaerobacter* appear to be based on the establishment of transmembrane Na^+ gradients, as observed in multiple *Clostridia* species (Fig. 2) (37, 40). *Thermoanaerobacter* species employ a vacuolar V-type ATPase which likely operates as a sodium efflux pump during fermentation, as well as ferredoxin-linked Na^+-translocating Rnf complexes (Fig. 2, reactions 1 and 2). Predicted Na^+-translocating decarboxylase enzymes (Fig. 2, reactions 3 and 4) and Na^+/H^+ antiporters

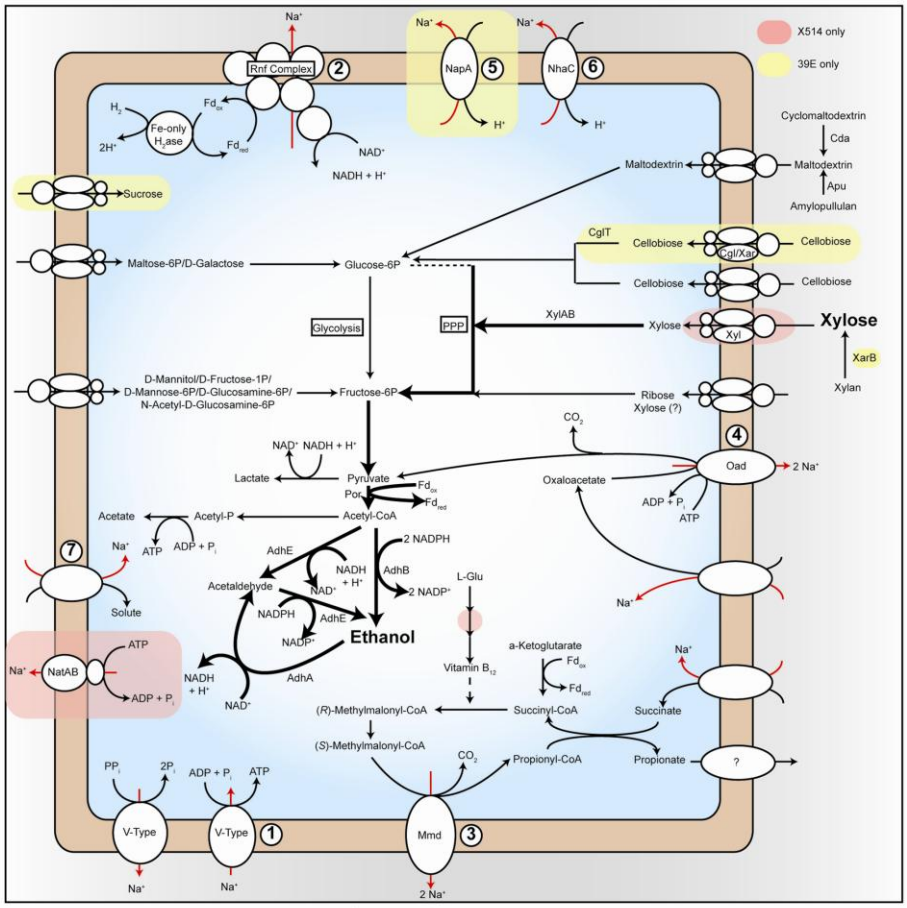


FIG. 2. Central carbon and energy metabolism of *Thermoanaerobacter* strains. Key pathways and enzymes related to ethanol production are indicated. Enzymes/pathways unique to a specific lineage are colored as follows: pink, X514 only; yellow, 39E only. The dotted line in the pentose phosphate pathway indicates an incomplete pathway, while the dotted line from vitamin B_{12} to methylmalonyl-CoA mutase indicates the requirement of this cofactor for enzyme activity.

(Fig. 2, reactions 5 and 6) likely serve to maintain H^+ and Na^+ gradients, while Na^+/solute symporters may allow for nonspecific influx of solutes, including sugars (Fig. 2, reaction 7). While numerous *Clostridia* species employ Na^+ energetics, one particular aspect of *Thermoanaerobacter* Na^+-based energetics was predicted to have a significant, albeit indirect, effect on ethanol yields.

The key to energetics in *Thermoanaerobacter* appears to be Na^+-translocating decarboxylation reactions, in particular those with methylmalonyl-coenzyme A (methylmalonyl-CoA) decarboxylase (Fig. 2, reaction 3). A key enzyme in methylmalonyl-CoA metabolism is methylmalonyl-CoA mutase, which requires vitamin B_{12} as a cofactor and feeds methylmalonyl-CoA to methylmalonyl-CoA decarboxylase. A total of 21 genes encoding putative vitamin B_{12}-dependent enzymes were identified in the two *Thermoanaerobacter* genomes, in-

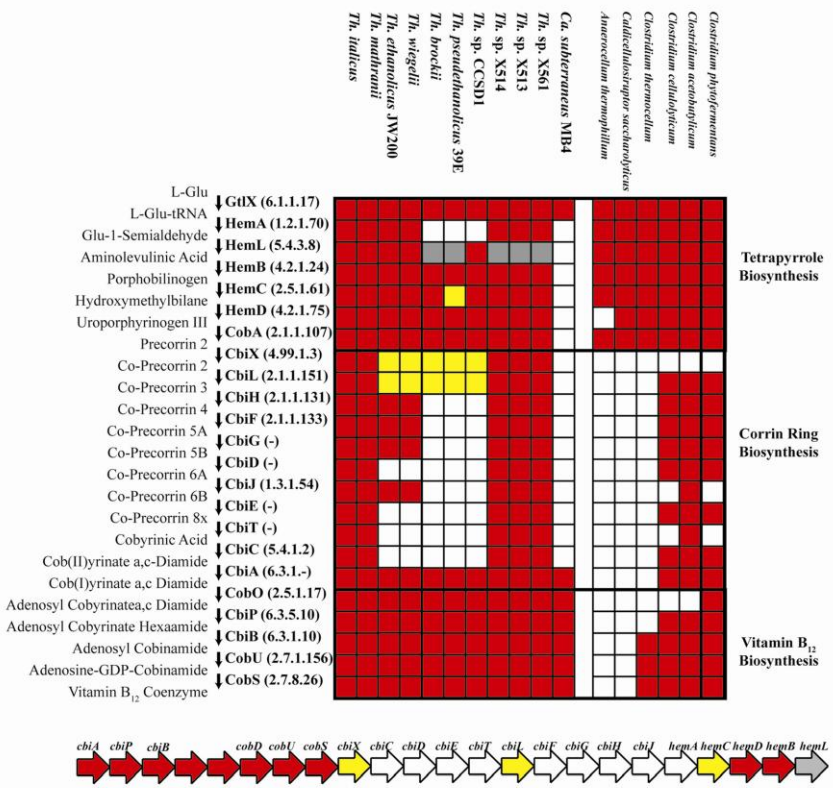


FIG. 3. Comparison of vitamin B_{12} biosynthesis genes in *Thermoanaerobacter* species. The grid is arranged with columns representing *Thermoanaerobacter* and outlier *Clostridia* species and rows representing protein-encoding genes of the vitamin B_{12} biosynthesis pathway. Each square on the grid thus represents a given vitamin B_{12} biosynthesis gene for a given species and is colored depending on whether the gene is present (red), present but degraded or truncated (yellow), presumed absent (white), or presumed to be part of a gene fusion (gray; in this case representing a bifunctional siroheme synthase/glutamate-1-semialdehyde aminotransferase). Below the grid is the complete vitamin B_{12} biosynthesis operon structure of X514, with genes colored as in the grid to indicate that the orthologous 39E gene is present, absent, or degraded. *Caldanaerobacter subterraneus* subsp. *tengcongensis* sp. MB4 was originally classified as *Thermoanaerobacter tengcongensis* MB4 when the genome was sequenced, but the strain has since been reclassified.

cluding methylmalonyl-CoA mutase, methionine synthase, and homocysteine *S*-methyltransferase genes (see Table S2 in the supplemental material). X514 carries a complete vitamin B_{12} biosynthesis operon, whereas 39E has lost genes encoding the anaerobic corrin ring biosynthesis portion of the pathway (*cbiCDT* and *cbiFGHJ-hemA*) (Fig. 3). Extension of this analysis to all currently available *Thermoanaerobacter* genomes suggests that the *de novo* vitamin B_{12} pathway is likely ancestral to the *Thermoanaerobacter* genus and that the anaerobic corrin ring biosynthesis portion of the pathway has been lost in the 39E/*Thermoanaerobacter brockii*/CCSD1 lineage (Fig. 3). This loss appears to have occurred in two steps, with the *cbiCDET* cluster first lost in the *Thermoanaerobacter ethanolicus*/*pseudethanolicus* lineage and the *cbiFGHJ-hemA* genes lost in the *Thermoanaerobacter pseudethanolicus* lineage (Fig. 3). Consistent with this hypothesis, *in silico* analysis of the sirohydrochlorin cobalt chelatase (*cbiX*) genes of the *Thermoanaerobacter* strains showed that in X514, the *cbiX* gene is adjacent to the precorrin-8X methylmutase (*cbiC*) gene, which is the first gene in the cluster missing from 39E (Fig. 3). In 39E, the *cbiX* gene lacks approximately 120 nucleotides corresponding to the N terminus of the expressed full-length protein, resulting in the

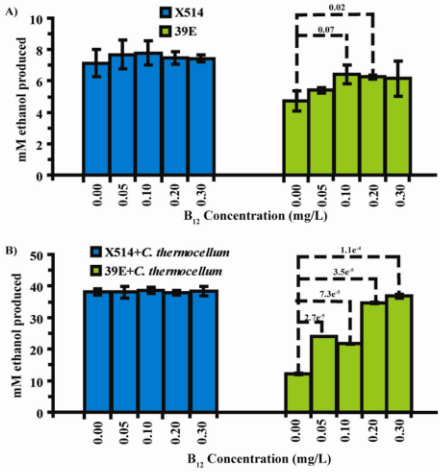


FIG. 4. Effects of exogenous vitamin B_{12} on ethanol yields for *Thermoanaerobacter* strains (X514 and 39E) grown on 50 mM xylose (A) and *Thermoanaerobacter-C. thermocellum* cocultures grown on 1% cellulose (B). Vitamin B_{12} concentrations in the media are indicated on the *x* axis (mg/liter). The data represent the means for three replicates, with error bars presenting standard deviations. Significant differences in mean ethanol yields were calculated using the one-tailed paired *t* test, and *P* values of <0.1 are shown above the dotted line.

gene being located adjacent to the truncated *cbiL* gene (Fig. 3). Interestingly, some strains show a putative gene fusion encoding siroheme synthase and glutamate-1-semialdehyde aminotransferase, whereas in other strains these functions appear to be encoded by different genes (Fig. 3). Thus, disruption of the Na^+ gradient in strain 39E by removing vitamin B_{12} from the growth medium would be expected to impair growth and ethanol yields, while X514 would be unaffected. Furthermore, it has been observed previously that increased ethanol yields from cellulose fermentation can be achieved in some strains of *C. thermocellum* by the addition of exogenous vitamin B_{12} (35). Thus, the ability to synthesize vitamin B_{12} *de novo* could potentially enhance the ability of X514 to maintain energy metabolism under nutrient-poor conditions and may provide a needed cofactor in cellulolytic cocultures with *C. thermocellum* which would enhance cellulose fermentation.

To test this hypothesis, X514 and 39E were grown individually on xylose and in coculture with *C. thermocellum* LQRI on cellulose, with differing concentrations of exogenous vitamin B_{12} added to the medium (Fig. 4). In *Thermoanaerobacter* monocultures grown on 50 mM xylose, ethanol yields in X514 were largely independent of vitamin B_{12} concentrations even when no vitamin B_{12} was present in the medium, while yields decreased significantly (~60% of maximum yield) in 39E when no exogenous vitamin B_{12} was present (Fig. 4A). This effect was much more pronounced when the strains were grown in coculture with *C. thermocellum* on 1% cellulose (Fig. 4B). Ethanol yields were strikingly lower (~25% of maximum yield) in the 39E coculture with *C. thermocellum* on 1% cellulose when no exogenous vitamin B_{12} was added, and yields approached those of the X514 coculture only when 2- to 3-fold more vitamin B_{12} was added to the growth medium (Fig. 4B). In contrast, ethanol yields remained unchanged in X514-*C. thermocellum* cocultures, even when no exogenous vitamin B_{12} was added to the growth medium (Fig. 4B). These results suggest that X514 synthesizes vitamin B_{12} *de novo*, as predicted, and that this ability has a significant effect on ethanol yields in both mono- and coculture.

Central carbon metabolism. Differences in energetics between the strains provide a clear, albeit indirect, link between differing genotypes and ethanol yields. However, other genotypic differences appear to provide a more direct correlation to ethanol yield phenotypes. An identical core set of Embden-Meyerhof-Parnas (EMP) pathway genes were identified in both strains (Fig. 2). 39E carries additional ED pathway genes, though neither strain appears to employ the modified ED pathway observed in many *Clostridia* species (6). Multiple pathways for metabolizing hexose and pentose sugars (glucose, fructose, xylose, maltose, ribose, and galactose) are present, and xylose metabolism appears to occur exclusively via the traditional xylose isomerase-xylulokinase-pentose phosphate pathway. Both strains also carry genes annotated as encoding pyruvate decarboxylase subunit B, a key enzyme in ethanol production in *Zymomonas* strains (3, 38), but sequence homology suggests that this enzyme is more likely an oxaloacetate decarboxylase and thus not expected to be involved directly in ethanol formation. In addition, both strains utilize multiple complex carbohydrates, including amylopectin, cyclodextrin, cellodextrin (cellobiose, cellotriose, cellotetraose, and cellopentaose), pectin, and pullulan, while only 39E is capable of metabolizing sucrose and degrading xylan (see below). Thus, the phenotypic difference in ethanol yields between the strains does not appear to be due to observable differences in carbon metabolism pathways between the strains.

A total of 46, 45, and 48 presumptive carbohydrate-active enzymes (7) were identified in 39E, X514, and MB4, respectively (see Table S3 in the supplemental material). While many of the relevant genes correspond to characterized carbohydrate metabolism enzymes (e.g., amylopullulanase, etc.), several remain uncharacterized. Whole-transcriptome analysis of each of the strains showed that many of these genes were actively expressed when cells were grown with 50 mM xylose (see Table S3). SIGI-HMM analysis (47) suggests that X514 has acquired (via lateral gene transfer) a lineage-specific, ~1-kb gene cluster encoding additional copies of xylulokinase, transketolase, transaldolase, alcohol dehydrogenase (ADH), aldehyde dehydrogenase, acetate kinase, and bacterial microcomponent shell components that could potentially facilitate carbon metabolism in X514 by alleviating metabolic bottlenecks between xylose and ethanol. However, microarray analysis suggests that these genes are not actively expressed when cells are grown on xylose and thus likely have no significant effect on ethanol production yields in X514.

A total of 7 and 9 alcohol dehydrogenase genes were identified in 39E and X514, respectively (Table 1). Both strains encode an NADPH-dependent secondary alcohol dehydrogenase (*adhB*) predicted to catalyze the terminal step in ethanol

TABLE 1. Alcohol dehydrogenase genes carried by strains 39E and X514[a]

adh gene product	39E		X514	
	Locus tag	GeneID	Locus tag	GeneID
Iron-containing ADH			Teth514_0145	167038817
	Teth39_1979	167038363	Teth514_0241	167038910
			Teth514_0564	167039223
Bifunctional secondary ADH/aldehyde dehydrogenase (*adhE*)	Teth39_0206	167036636	Teth514_0627	167039286
Bifunctional secondary ADH/acetyl-CoA thioesterase (*adhB*)	Teth39_0218	167036648	Teth514_0653	167039312
Primary ADH (*adhA*)	Teth39_0220	167036650	Teth514_0654	167039313
Short-chain ADH	Teth39_2190	167038570	Teth514_1808	167040441
ADH with GroES domain	Teth39_0878	167037292	Teth514_1882	167040513
Iron-containing ADH			Teth514_1935	167040564
	Teth39_1597	167037999		

[a] Orthologous genes are presented in rows.

production in *Thermoanaerobacteraceae* species (also found in other *Clostridia* species, such as *Thermosinus carboxidivorans*, *Clostridium botulinum*, and *Ethanoligenens harbinense*) and an NADPH-dependent primary alcohol dehydrogenase (*adhA*) involved in ethanol catabolism (3a, 4). In both strains, these genes are located adjacent to each other (in opposite orientation) and adjacent to a predicted $NADP^+$ reductase gene (Teth514_0652). Both strains also encode a bifunctional secondary alcohol dehydrogenase/aldehyde dehydrogenase (*adhE*) (Fig. 2; Table 1) implicated in ethanol production in *Thermoanaerobacter* (32) and in butanol production in *Clostridium acetobutylicum* (11). This complement of ADH enzymes is novel compared to that of most ethanol-producing bacteria, which typically employ a primary ADH (*adhA*) as the terminal step in ethanol production (4). X514 also carries three additional lineage-specific iron-only *adh* genes compared to 39E, which carries only one additional lineage-specific *adh* gene (Table 1). Microarray analysis of 39E and X514 grown on xylose showed that the three major *adh* genes (*adhA*, *adhB*, and *adhE*) were expressed in both strains at different xylose concentrations (see Fig. S2 in the supplemental material), a result consistent with previous analysis of the expression of these genes (4). *adhB* in particular appears to be expressed highly in both strains at 10 mM xylose, suggesting that AdhB is the key terminal enzyme involved in ethanol production (see Fig. S2). Interestingly, the *adhA* gene encoding the catabolic primary ADH is expressed much more highly in 39E than in X514, and the conserved ADH encoded by Teth514_0241/Teth39_1979 is expressed more highly in 39E under the same growth conditions (see Fig. S2). AdhB has been proposed as a biotechnological target for enhancing ethanol yields in *Thermoanaerobacter* strains, and the observation of differential expression of *adh* genes in response to xylose is intriguing, but no definitive evidence linking differences in *adh* expression to strain-specific ethanol yields was observed.

Flux analyses of carbon metabolism pathways. Genomic differences alone were insufficient to link the observed physiological phenotypes to genotypic traits. To extend the analyses, ^{13}C-assisted flux analysis was conducted for both strains to quantify the relative and absolute flux rates from xylose to ethanol for the two strains and to complement a previous metabolomic analysis of X514 (Fig. 5; see Table S4 in the supplemental material) (10). The measured extracellular fluxes (i.e., ethanol production, growth rates, and the consumption rates of glucose and xylose) in X514 were >50% higher than those in 39E (Fig. 1). Using glucose as the carbon source (see below) revealed that the oxidative pentose phosphate pathway (a common pathway associated with cellular NADPH production) in X514 exhibited very low activity for glucose metabolism (relative flux, <5% of total flux) (Fig. 3), consistent with the lack of a 6-phosphogluconolactonase (EC 3.1.1.31) gene in the *Thermoanaerobacter* genomes (Fig. 2 and 5). The flux through the oxidative pentose phosphate pathway was moderately higher in 39E (relative flux, ~12% of total flux), suggesting that X514 might be more dependent on alternate pathways for NADPH production. For example, based on sequence annotation, X514 could use pyruvate:ferredoxin oxidoreductase (Teth514_0781) to produce reduced ferredoxin and a ferredoxin-$NADP^+$ reductase (Teth514_0652) to reduce $NADP^+$.

As expected from physiological characterization, a substantially higher carbon overflow (2- to 3-fold for glucose culture and 5- to 10-fold for xylose culture) was channeled to ethanol than to acetate and lactate for both strains (e.g., 1 mol glucose can generate about 1 mol ethanol by X514) (Fig. 5). Furthermore, the intracellular carbon flux distributions after normalization by carbon substrate (i.e., glucose or xylose) consumption rates were not significantly different between these two strains (Fig. 5; see Table S4 in the supplemental material), though absolute growth rates and metabolite production in X514 were much higher than those in 39E under both glucose and xylose growth conditions. Such observations suggest that the two phylogenetically closely related strains share similar regulation of central metabolic pathways, but unique genotypes of X514 likely significantly enhance ethanol synthesis rates.

Carbon transporters. Since observed differences in carbon metabolism pathways were insufficient to explain increased absolute flux rates in X514, the complement of sugar transporters (for xylose in particular) was examined. *Thermoanaerobacter* strains carry a variety of genes for carbon transporters, encompassing primarily ABC-type (120 and 134 genes in 39E and X514, respectively) and phosphotransferase system (PTS) (28 and 29 genes in 39E and X514, respectively) transporters (Fig. 2). These transporters are specific for hexoses (glucose, galactose, fructose, and mannose), pentoses (xylose and ribose), disaccharides (sucrose, lactose, cellobiose, and maltose), oligosaccharides (maltodextrin), and unspecified sugar and polysaccharide transporters (Fig. 2). A major difference in

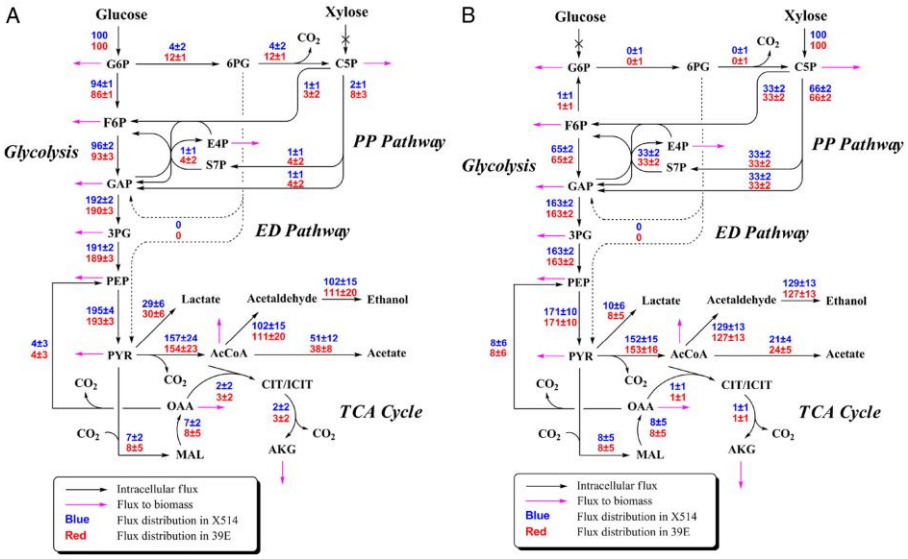


FIG. 5. Relative flux distributions for glucose (A) and xylose (B) metabolism. The actual carbon utilization mole rates were normalized to 100. The flux distributions describe the utilization of "100-mol" carbon sources by each pathway. The numbers in blue show the fluxes of X514, and the numbers in red show the fluxes of 39E. The fluxes to biomass are shown with pink arrows. Abbreviations: AcCoA, acetyl-coenzyme A; CIT, citrate; E4P, erythrose-4-phosphate; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; 6PG, 6-phosphogluconate; ICIT, isocitrate; MAL, malate; OAA, oxaloacetate; AKG, 2-oxoglutarate; PEP, phosphoenolpyruvate; 3PG, 3-phosphoglycerate; C5P, ribose-5-phosphate (or ribulose-5-phosphate) and xylulose-5-phosphate; S7P, sedoheptulose-7-phosphate; SUC, succinate; GAP, 3-P-glyceraldehyde; PYR, pyruvate.

carbon transport between these two strains is the complement of putative xylose transport genes in X514 (Fig. 6).

Genes encoding a bifunctional xylanosidase/arabinosidase (*xarB*), associated cellobiose transporters (*xarG*), and β-glucosidase (*cglT*) have previously been cloned and characterized for *Thermoanaerobacter ethanolicus* JW200 (25). The genomes of 39E and X514 were compared to genomes of eight additional *Thermoanaerobacter* species to determine if the variability of this region extends across multiple *Thermoanaerobacter* species and if this variability can be linked to observed physiological phenotypes (Fig. 6). The JW200 *xarBG-cglFT* gene cluster is conserved in 39E and is associated with genes of a ribose transport operon, with xylose metabolism (*xylAB*), and with mobile elements (Fig. 6). In contrast, X514 utilizes a xylose-specific XylFGH system but lacks the *xarB* xylanosidase gene (Fig. 6). *Thermoanaerobacter italicus* and *Thermoanaerobacter mathranii*, which represent a distinct lineage within *Thermoanaerobacter*, carry both operons as well as multiple lineage-specific xylanase genes (Fig. 6). Transcriptomic analysis revealed that all of the genes of the *xyl* operon in X514 were highly expressed when cells were grown with xylose at a low concentration (10 mM) compared to a high concentration (200 mM), but the *cgl-xar* operon in 39E was largely unresponsive to xylose, consistent with its presumed function as a cellobiose transport operon (see Fig. S4 and Table S5 in the supplemental material). The presence of a dedicated xylose transport system may explain why X514 can grow on xylose concentrations as low as 1 mM while 39E cannot, and it may also explain the observed increase in absolute flux rates from xylose to ethanol in X514 and raises the question of how xylose is taken up by 39E. A conserved ribose-specific ABC transport operon is located adjacent to the *xyl* and *xar* operons and could potentially display xylose uptake activity as well. However, transcriptomic analysis does not suggest that this operon is upregulated in response to xylose in 39E (see Table S5). A fructose/mannose/sorbose-specific ABC transport operon showed >2-fold increased expression in 39E cells grown on 10 mM xylose versus 200 mM xylose, suggesting that this operon could potentially have xylose uptake activity (see Table S5). Alternatively, 39E may employ less specific uptake systems, possibly via Na^+ gradient-mediated solute uptake.

DISCUSSION

The study of *Thermoanaerobacter* species has long been driven by interest in the ability to efficiently produce ethanol at high yields from both glucose and xylose (21, 32, 50) and in the potential applications in CBP, a concept offering the most

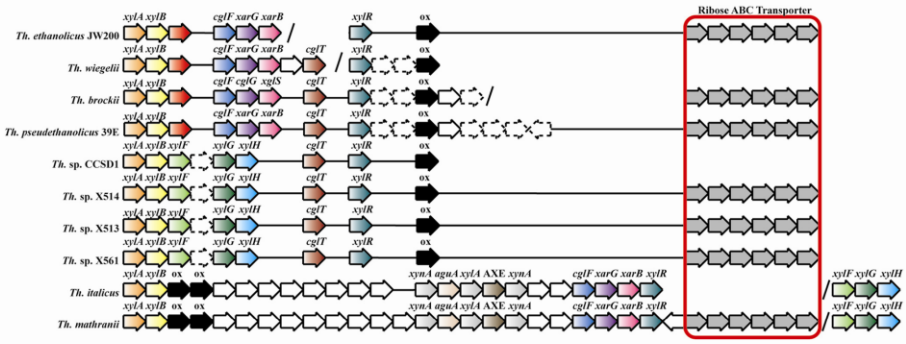


FIG. 6. Xylan degradation and xylose transport genes of *Thermoanaerobacter* species. For a given strain, genes are considered to be contiguous except where indicated by a slash (indicating a contig boundary or large genomic region). Orthologous genes are colored similarly and aligned vertically as much as possible. Open reading frames with dotted outlines indicate mobile elements such as transposons and integrase genes. The six-gene cluster colored gray represents a conserved ribokinase/ribose ABC transporter gene cluster. "ox" refers to a conserved oxidoreductase gene of unknown function. White arrows indicate hypothetical genes or miscellaneous genes not obviously relevant to xylose metabolism or transport.

promise in addressing central challenges to cellulosic bioethanol production (23). Several features in particular make *Thermoanaerobacter* species attractive for integration into CBP schemes, including (i) the ability to efficiently ferment both hexose and pentose sugars (particularly xylose) (21) to ethanol in their natural state, in contrast to non-pentose-fermenting strains such as *Zymomonas* (38); (ii) relatively high ethanol yields generated from sugar fermentation (14); (iii) simplification of nutrient requirements due to the capability for *de novo* cofactor (i.e., vitamin B_{12}) biosynthesis; (iv) a high growth temperature which is better suited for industrial processing of ethanol and minimization of microbial contamination than those for mesophilic strains such as engineered *E. coli*; (v) a high substrate affinity coupled with nearly complete substrate utilization; (vi) ease of growth in microbial consortia; and (vii) a unique terminal ethanol production pathway involving novel lineage-specific alcohol dehydrogenase enzymes which may positively affect fermentation balance and ethanol yields (3a, 4, 32). Although *Thermoanaerobacter* species have great biotechnological potential for cellulosic ethanol production, their genetics, biochemistry, and physiology are only partially understood. The sequencing of multiple *Thermoanaerobacter* genomes coupled to experimental analyses allowed for the first comprehensive attempt to link genotypic differences to phenotypic effects on substrate usage and product yields. This information in turn is vital to future attempts to engineer these strains for integration into CBP processes.

The use of *Thermoanaerobacter* species for ethanol production presents several major challenges. First, ethanol yields from *Thermoanaerobacter*, while high, do not currently meet desired industrial standards. Accumulation of alternative fermentation end products, such as lactate and acetate, reduces the efficiency of substrate conversion to ethanol, though *Thermoanaerobacter* strains are naturally more efficient at maintaining a high ratio of ethanol formed to other fermentation products than other strains commonly employed in industrial fermentation processes. Genetic engineering of strains by blocking metabolic pathways of acetate and lactate formation is a common mechanism for obtaining higher ethanol yields (17, 39, 41), though such efforts have proven difficult with *Thermoanaerobacter* species due to the lack of reliable genetic systems. Second, *Thermoanaerobacter* species efficiently convert sugars to ethanol at relatively low substrate concentrations (<10 g/liter), considerably lower than the desired, industry-scale substrate levels (>50 g carbohydrate/liter). Understanding how substrate concentrations affect ethanol yields and how they are regulated at the molecular level is important for the optimization of biotechnological processes. Third, pretreatment of lignocellulosic materials typically results in the formation of inhibitory compounds, including weak acids, furans, and phenolic compounds (31). Knowledge on how various inhibitors affect ethanol production at the molecular level is also critical to developing resistant strains for more efficient conversion of lignocellulosic materials to ethanol. Fourth, certain strains require growth factors (e.g., vitamins) susceptible to high temperatures prevalent under thermophilic conditions, which could hinder the efficiency of high-density thermophilic fermentation. Finally, rapid cell lysis has been observed in *Thermoanaerobacter* cultures reaching high densities, making it difficult to maintain the high process efficiencies required by industry-scale fermentation operations. The availability of whole-genome sequences of X514 and 39E provides the basis for further genetic and metabolic engineering for desired biotechnological applications.

Strains X514 and 39E are phylogenetically closely related, but significant differences in physiology have been observed with respect to ethanol production. For example, X514 produces higher yields of ethanol from xylose, and the stable coculture of *C. thermocellum* LQRI and X514 produces substantially more ethanol than the corresponding *C. thermocel-*

lum LQRI-39E coculture. The whole-genome sequence comparison and experimental analyses in this study have revealed key genotypic differences between these two strains which may explain these phenotypes. In contrast to 39E, X514 encodes a xylose-specific transporter, additional lineage-specific alcohol dehydrogenase enzymes, and a complete vitamin B_{12} *de novo* biosynthesis pathway, which is critical for maintenance of Na^+ gradient-based energetics. Additional genes encoding xylose metabolism and pentose phosphate enzymes in X514 are present but do not appear to significantly affect carbon flux in X514. While relative carbon flux from xylose appears to be the same in each strain, the observed greater absolute flux in X514 and the ability of X514 to grow at lower xylose concentrations than those for growth of 39E may be explained by the employment of xylose-specific Xyl transporters in X514. These transporters may allow for more efficient uptake and utilization of xylose in this strain than in 39E, which must presumably utilize a more generic xylose uptake system (i.e., Na^+ gradient-linked solute transporters, etc.). Furthermore, the capacity for *de novo* synthesis of vitamin B_{12} by X514 appears to be a key factor in maintaining high ethanol yields in cellulosic cocultures compared to those with 39E. Finally, transcriptomic analysis shows that alcohol dehydrogenase genes are expressed differentially in both strains under different growth conditions but that the AdhB enzyme is likely the key ADH enzyme involved in ethanol production and is thus a logical candidate for future experimental analysis and engineering efforts. Recent studies of the closely related *Thermoanaerobacteraceae* species *Thermoanaerobacterium saccharolyticum* have shown that metabolic and genetic engineering can be used successfully to enhance ethanol yields from sugar, to yields of 37 g/liter in the case of *T. saccharolyticum* (39). While similar results would be expected from engineered *Thermoanaerobacter* species, such efforts have been hampered by the relative genetic recalcitrance of strains. Further systematic in-depth functional genomic studies are needed to fully understand how these genotypic differences contribute to the observed phenotypic differences in sugar utilization and ethanol production. Such information will be essential for the metabolic engineering of *Thermoanaerobacter* strains for high ethanol production. In addition, the unique genes observed in *Thermoanaerobacter* (e.g., alcohol dehydrogenase and xylose transporter genes) could also be useful for genetically engineering non-xylose-utilizing microorganisms for xylose metabolism and/or improved ethanol yields (17, 20, 36, 39, 40, 41).

In conclusion, *Thermoanaerobacter* species are important in biotechnological applications of biofuel production. However, implementation of *Thermoanaerobacter* species into CBP schemes requires a detailed, systems-level understanding of the genetics, biochemistry, physiology, and ecology of this group of bacteria. The comparative whole-genome sequence analyses of X514 and 39E presented here provide valuable insights into the nature of carbon metabolism, energy metabolism, carbon transporters, ethanol production, and cofactor biosynthesis in these *Thermoanaerobacter* strains. Such information is important not only for implementing *Thermoanaerobacter* strains into industrially viable CBP schemes through further developing and/or optimizing more efficient strains but also for potentially engineering other, non-xylose-utilizing and/or ethanol-producing microorganisms.

ACKNOWLEDGMENTS

We thank D. Klingman at Oak Ridge National Laboratory for preparation of genomic DNA for sequencing.

This material is based on work supported by the Oklahoma Bioenergy Center (J.Z.), the U.S. Department of Energy Joint BioEnergy Institute (JBEI) (J.Z.), the Chinese Academy of Sciences (J.X.), and the National Science Foundation EPSCoR program under grant EPS-0814361 (J.Z.). The genome sequencing work was performed under the auspices of the U.S. Department of Energy's Office of Science, Biological and Environmental Research Program and by the University of California, Lawrence Berkeley National Laboratory, under contract DE-AC02-05CH11231, the Lawrence Livermore National Laboratory, under contract DE-AC52-07NA27344, and the Los Alamos National Laboratory, under contract DE-AC02-06NA25396, as previously described (15).

Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

J.Z. and M.W.F. developed the original concept. J.Z. and T.J.P. isolated strain X514. C.L.H. conducted the comparative computational analysis, designed experiments, and analyzed results. Y.D. and Q.T. aided in computational analyses. Z.H., L.W., and J.V.N. oversaw microarray design and construction. L.L. and H.M. conducted growth analysis of strains on xylose, and L.L. and A.Z. conducted microarray analysis of strains on xylose. M.W.F. and B.D.R. conducted growth analysis of strains on cellodextrins and other sugars. Y.J.T., X.F., and Z.Z. conducted flux analysis. J.Z. and C.L.H. prepared the manuscript, with contributions from Y.J.T., M.W.F., Q.H., Z.H., J.X., T.J.P., and J.W.

The authors declare that they have no competing interests.

REFERENCES

1. **Ballesteros, I., et al.** 1991. Selection of thermotolerant yeasts for simultaneous saccharification and fermentation (SSF) of cellulose to ethanol. Appl. Biochem. Biotechnol. **28–29:**307–316.
2. **Bayer, E. A., R. Lamed, and M. E. Himmel.** 2007. The potential of cellulases and cellulosomes for cellulosic waste management. Curr. Opin. Biotechnol. **18:**237–245.
3. **Bräu, B., and H. Sahm.** 1986. Cloning and expression of the structural gene for pyruvate decarboxylase of *Zymomonas mobilis* in *Escherichia coli*. Archives of gene encoding the *Thermoanaerobacter ethanolicus* 39E secondary-alcohol dehydrogenase and biochemical characterization of the enzyme. Biochem. J. **316:**115–122.

3a. **Burdette, D. S., C. Vielle, and J. G. Zeikus.** 1996. Cloning and expression of the gene encoding the *Thermoanaerobacter ethanolicus* 39E secondary-alcohol dehydrogenase and biochemical characterization of the enzyme. Biochem. J. **316:**115–122.

4. **Burdette, D., and J. G. Zeikus.** 1994. Purification of acetaldehyde dehydrogenase and alcohol dehydrogenases from *Thermoanaerobacter ethanolicus* 39E and characterization of the secondary-alcohol dehydrogenase as a bifunctional alcohol dehydrogenase–acetyl-CoA reductive thioesterase. Biochem. J. **302:**163–170.
5. **Chhabra, S. R., et al.** 2006. Global analysis of heat shock response in *Desulfovibrio vulgaris* Hildenborough. J. Bacteriol. **188:**1817–1828.
6. **Conway, T.** 1992. The Entner-Doudoroff pathway: history, physiology and molecular biology. FEMS Microbiol. Lett. **103:**1–28.
7. **Coutinho, P. M., and B. Henrissat.** 1999. Carbohydrate-active enzymes: an integrated database approach, p. 3–12. *In* H. J. Gilbert, G. Davies, B. Henrissat, and B. Svensson (ed.), Recent advances in carbohydrate bioengineering. The Royal Society of Chemistry, Cambridge, United Kingdom.
8. **Demain, A. L., M. Newcomb, and J. H. D. Wu.** 2005. Cellulase, clostridia, and ethanol. Microbiol. Mol. Biol. Rev. **69:**124–154.
9. **Fang, Z.** 2010. Enhanced role of the co-culture of thermophilic anaerobic bacteria on cellulosic ethanol. Huan Jing Ke Xue **4:**1059–1065.
10. **Feng, X., et al.** 2009. Characterization of the central metabolic pathways in *Thermoanaerobacter* sp. X514 via isotopomer-assisted metabolite analysis. Appl. Environ. Microbiol. **75:**5001–5008.
11. **Fontaine, L., et al.** 2002. Molecular characterization and transcriptional analysis of *adhE2*, the gene encoding the NADH-dependent aldehyde/alcohol dehydrogenase responsible for butanol production in alcohologenic cultures of *Clostridium acetobutylicum* ATCC 824. J. Bacteriol. **184:**821–830.
12. **Freier, D., C. P. Mothershed, and J. Wiegel.** 1988. Characterization of *Clostridium thermocellum* JW20. Appl. Environ. Microbiol. **54:**204–211.

13. **Hamelinck, C. N., G. V. Hooijdonk, and A. P. C. Faaij.** 2005. Ethanol from lignocellulosic biomass: techno-economic performance in short-, middle- and long-term. Biomass Bioenergy **28:**384–410.
14. **He, Q., P. M. Lokken, S. Chen, and J. Zhou.** 2009. Characterization of the impact of acetate and lactate on ethanolic fermentation by *Thermoanaerobacter ethanolicus*. Bioresour. Technol. **100:**5955–5965.
15. **Hemme, C. L., et al.** 2010. Sequencing of multiple *Clostridia* genomes related to biomass conversion and biofuels production. J. Bacteriol. **192:**6494–6496.
16. **Houghton, J., S. Wheatherwax, and J. Ferrell.** 2006. Breaking the biological barriers to cellulosic ethanol: a joint research agenda, p. 206. U.S. Department of Energy, Washington, DC.
17. **Ingram, L. O., et al.** 1998. Metabolic engineering of bacteria for ethanol production. Biotechnol. Bioeng. **58:**204–214.
18. **Jiang, H., Z. He, Z. Fang, and J. Zhou.** 2008. Abstr. 108th Gen. Meet. Am. Soc. Microbiol., Boston, MA, abstr. O–009.
19. **Konstantinidis, K. T., and J. M. Tiedje.** 2005. Genomic insights that advance the species definition for prokaryotes. Proc. Natl. Acad. Sci. U. S. A. **102:** 2567–2572.
20. **Kuyper, M., et al.** 2005. Metabolic engineering of a xylose-isomerase-expressing *Saccharomyces cerevisiae* strain for rapid anaerobic xylose fermentation. FEMS Yeast Res. **5:**399–409.
21. **Lacis, L. S., and H. G. Lawford.** 1988. Ethanol production from xylose by *Thermoanaerobacter ethanolicus* in batch and continuous culture. Arch. Microbiol. **150:**48–55.
22. **Li, X., Z. He, and J. Zhou.** 2005. Selection of optimal oligonucleotide probes for microarrays using multiple criteria, global alignment and parameter estimation. Nucleic Acids Res. **33:**6114–6123.
23. **Lynd, L. R., et al.** 2008. How biotech can transform biofuels. Nat. Biotechnol. **26:**169–172.
24. **Lynd, L. R., W. H. van Zyl, J. E. McBride, and M. Laser.** 2005. Consolidated bioprocessing of cellulosic biomass: an update. Curr. Opin. Biotechnol. **16:** 577–583.
25. **Mai, V., J. Wiegel, and W. W. Lorenz.** 2000. Cloning, sequencing, and characterization of the bifunctional xylosidase-arabinosidase from the anaerobic thermophile *Thermoanaerobacter ethanolicus*. Gene **247:**137–143.
26. **Markowitz, V. M., et al.** 2006. The integrated microbial genomes (IMG) system. Nucleic Acids Res. **34:**D344–D348.
27. **Mielenz, J. R.** 2001. Ethanol production from biomass: technology and commercialization status. Curr. Opin. Microbiol. **4:**324–329.
28. **Ng, T. K., A. Ben-Bassat, and J. G. Zeikus.** 1981. Ethanol production by thermophilic bacteria: fermentation of cellulosic substrates by cocultures of *Clostridium thermocellum* and *Clostridium thermohydrosulfuricum*. Appl. Environ. Microbiol. **41:**1337–1343.
29. **Olofsson, K., M. Bertilsson, and G. Liden.** 2008. A short review on SSF—an interesting process option for ethanol production from lignocellulosic feedstocks. Biotechnol. Biofuels **1:**7.
30. **Onyenwoke, R. U., V. V. Kevbrin, A. M. Lysenko, and J. Wiegel.** 2007. *Thermoanaerobacter pseudethanolicus* sp. nov., a thermophilic heterotrophic anaerobe from Yellowstone National Park. Int. J. Syst. Evol. Microbiol. **57:**2191–2193.
31. **Palmqvist, E., and B. Hahn-Hagerdal.** 2000. Fermentation of lignocellulosic hydrolysates. II. Inhibitors and mechanisms of inhibition. Bioresour. Technol. **74:**25–33.
32. **Peng, H., G. Wu, and W. Shao.** 2008. The aldehyde/alcohol dehydrogenase (AdhE) in relation to the ethanol formation in *Thermoanaerobacter ethanolicus* JW200. Anaerobe **14:**125–127.
33. **Roh, Y., et al.** 2002. Isolation and characterization of metal-reducing *Thermoanaerobacter* strains from deep subsurface environments of the Piceance Basin, Colorado. Appl. Environ. Microbiol. **68:**6013–6020.
34. **Rosner, B.** 2005. Fundamentals of biostatistics, 6th ed. Duxbury Press, Boston, MA.
35. **Sato, K., et al.** 1992. Effect of yeast extract and vitamin B_{12} on ethanol production from cellulose by *Clostridium thermocellum* I-1-B. Appl. Environ. Microbiol. **58:**734–736.
36. **Sauer, U.** 2001. Evolutionary engineering of industrially important microbial phenotypes. Adv. Biochem. Eng. Biotechnol. **73:**129–169.
37. **Seedorf, H., et al.** 2008. The genome of *Clostridium kluyveri*, a strict anaerobe with unique metabolic features. Proc. Natl. Acad. Sci. U. S. A. **105:**2128–2133.
38. **Seo, J.-S., et al.** 2005. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. Nat. Biotechnol. **23:**63–68.
39. **Shaw, A. J., et al.** 2008. Metabolic engineering of a thermophilic bacterium to produce ethanol at high yield. Proc. Natl. Acad. Sci. U. S. A. **105:**13769–13774.
40. **Speelmans, G., B. Poolman, T. Abee, and W. N. Konings.** 1994. The F- or V-type Na(+)-ATPase of the thermophilic bacterium *Clostridium fervidus*. J. Bacteriol. **176:**5160–5162.
41. **Stephanopoulos, G.** 2007. Challenges in engineering microbes for biofuels production. Science **315:**801–804.
42. **Tang, Y. J., J. S. Hwang, D. Wemmer, and J. D. Keasling.** 2007. The *Shewanella oneidensis* MR-1 fluxome under various oxygen conditions. Appl. Environ. Microbiol. **73:**718–729.
43. **Tang, Y. J., et al.** 2009. Analysis of metabolic pathways and fluxes in a newly discovered thermophilic and ethanol-tolerant *Geobacillus* strain. Biotechnol. Bioeng. **102:**1377–1386.
44. **Thompson, D. K., et al.** 2002. Transcriptional and proteomic analysis of a ferric uptake regulator (Fur) mutant of *Shewanella oneidensis*: possible involvement of Fur in energy metabolism, transcriptional regulation, and oxidative stress. Appl. Environ. Microbiol. **68:**881–892.
45. **Updegraff, D. M.** 1969. Semimicro determination of cellulose in biological materials. Anal. Biochem. **32:**420–424.
46. **van Zyl, W., L. Lynd, R. den Haan, and J. McBride.** 2007. Consolidated bioprocessing for bioethanol production using *Saccharomyces cerevisiae*. Adv. Biochem. Eng. Biotechnol. **108:**205–235.
47. **Waack, S., et al.** 2006. Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. BMC Bioinformatics **7:**142.
48. **Wahl, S. A., M. Dauner, and W. Wiechert.** 2004. New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnol. Bioeng. **85:** 259–268.
49. **Wiegel, J.** 1982. Ethanol from cellulose. Experientia (Basel) **38:**151–156.
50. **Wiegel, J.** 1980. Formation of ethanol by bacteria. A pledge for the use of extreme thermophilic anaerobic bacteria in industrial ethanol fermentation processes. Cell. Mol. Life Sci. **36:**1434–1446.
51. **Wingren, A., M. Galbe, and G. Zacchi.** 2003. Techno-economic evaluation of producing ethanol from softwood: comparison of SSF and SHF and identification of bottlenecks. Biotechnol. Prog. **19:**1109–1117.
52. **Zhang, Y.-H. P., and L. R. Lynd.** 2005. Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. Proc. Natl. Acad. Sci. U. S. A. **102:**7321–7325.

Supplementary Tables

Table S1. Physiological, ecological and genomic characteristics of *Thermoanaerobacter* strains analyzed in this study.

	Strain	39E	X514
Physiology and Ecology			
	Habitat	Algal-cyanobacterial mat (Octopus Spring, Yellowstone National Park, Wyoming, USA)	Deep subsurface rock (~2000 m below surface, Piceance Basin, Colorado, USA)
	Optimal growth Temp.	65°C	65°C
	Spores	Yes	Yes
	Motile	Yes	Yes
	Glucose Fermentation Products	Ethanol, lactate, acetate, H_2	Ethanol, lactate, acetate, H_2
	Cellulose Degradation	No	No
	Cellobiose Utilization	Yes	Yes
	Xylan Degradation	Yes	No
	Xylose Utilization	Yes	Yes
Genome Characteristics			
	Genome Size (Mb)	2,362,816	2,457,259
	%G+C	34.5	34.5
	Protein Coding Genes (total)	2243	2349
	Structural RNAs	75	73
	Regulatory Proteins (total)[a]	156	186
	Signal Transduction Proteins[a]	57	64
	Transporters (total)[b]	308	331
	Carbohydrate Active Enzymes[c]	46	42
	Alcohol Dehydrogenase	7	9

[a] Based on IMG COG assignment
[b] JGI-ORNL transporter classification
[c] Derived from CAZy database

Table S2. **Vitamin B_{12}-dependent enzymes of *Thermoanaerobacter*.** Methylmalonyl-CoA mutase catalyzes the conversion of succinyl-CoA to *(R)*-methylmalonyl-CoA which is subsequently propionyl-CoA by the Na^+-translocating methylmalonyl-CoA decarboxylase. Thus, the lack of exogenous B_{12} may impair this pathway forcing physiological changes in energy metabolism.

Gene Functions	**39E**		**X514**	
	Locus Tag	gi#	Locus Tag	gi#
Ribonucleoside-diphosphate reductase, adenosylcobalamine-dependent	Teth39_0329	167036756	Teth514_0821	167039475
Anaerobic ribonucleoside-triphosphate reductase	Teth39_0653	167037075	Teth514_1144	167037075
Homocysteine S-methyltransferase	Teth39_0697	167037117	Teth514_1211	167039856
Radical SAM domain protein	Teth39_0963	167037376	Teth514_1404	167040044
Vitamin B12-dependent methionine synthase, activation region	Teth39_0998	167037411	Teth514_1438	167040078
Cobalamin B12-binding domain protein	Teth39_1044	167037457	Teth514_1481	167040119
Methylaspartate mutase	Teth39_1045	167037458	Teth514_1482	167040120
Methylmalonyl-CoA mutase, C-terminal domain (cobalamin-binding)	Teth39_0903	167037317	Teth514_1854	167040485
Methylmalonyl-CoA mutase, N-terminal domain	Teth39_0902	167037316	Teth514_1855	167040486
Methionine synthase I, cobalamine-binding domain	Teth39_0902	167037316	Teth514_1855	167040485
Glycerol dehydratase	-	-	Teth514_1953	167040582

Table S3. Carbohydrate active enzymes of analyzed *Thermoanaerobacter* species. Normalized signal intensity is presented for each gene based on microarray analysis for each strain grown on 50 mM xylose. Low quality spots or signals with SNR < 2 are not shown.

CAZy Category[a]	Family[a]	39E		X514	
		Gene	SI[b]	Gene	SI[b]
Glycoside Hydrolase Family	1	Teth39_2054 (*cglT*)	2385.464	Teth514_0267	2416.898
		Teth39_1943	-		
		Teth39_1927	4870.096		
	2	Teth39_1928	1596.471	Teth514_1095	4483.051
		Teth39_0611	1466.611		
	3	Teth39_2055 (*xarB*)	4356.47	Teth514_0182	604.4459
	4	Teth39_1944	1411.264	Teth514_0266	632.8227
				Teth514_0394	-
	5	Teth39_2182	-	Teth514_1790	3897.125
	13	Teth39_0668	1868.204	Teth514_1781	1010.442
		Teth39_0676[c]	15885.95	Teth514_1180	2051.315
		Teth39_2173	1106.995	Teth514_1187	4318.655
		Teth39_0672	7972.532		
	15	Teth39_0265	-	Teth514_0758	2167.945
	18	Teth39_1477	-	Teth514_2159	-
				Teth514_1268	-
	23	Teth39_2229	-	Teth514_2149	25424.83
		Teth39_1467	-	Teth514_2154	3088.769
		Teth39_1472	-	Teth514_0717	1038.583
	31	Teth39_0006	-	Teth514_0006	9804.891
	32	Teth39_0628	-	Teth514_1118	-
	36	Teth39_0610	1222.864	Teth514_1094	12446.3
	38	Teth39_0612[c]	2774.335	Teth514_1096	-
	65	Teth39_1512	-	Teth514_2202	-
		Teth39_1521	-	Teth514_2193	-
	66	Teth39_0264	-	Teth514_0757	2950.148
		Teth39_0266	-	Teth514_0759	1733.18
	94	Teth39_0505	941.1781	Teth514_0980	1312.107
Glycosyl Transferase Family	2	Teth39_1631	11352.39	Teth514_1651	1266.353
		Teth39_1215	-	Teth514_1549	7173.967
		Teth39_0707	-		
		Teth39_1630	8444.277		
		Teth39_1112	2155.781		
	4	Teth39_1632	3346.071	Teth514_2283	3372.149
		Teth39_2001	-	Teth514_2281	15286.39
		Teth39_2178	-	Teth514_0212	4769.243
				Teth514_1786	952.5164
				Teth514_2282	4574.194
				Teth514_2278	14460.91
	28	Teth39_0813	-	Teth514_2009	3356.766
	35	Teth39_0695	2585.812	Teth514_1209	4565.471

	51	Teth39_1978	-	Teth514_0242	-
	84	Teth39_0505	941.1781	Teth514_0980	1312.107
Carbohydrate Esterase Family	4	Teth39_0532	-	Teth514_1016	4738.704
		Teth39_1202[c]	6955.108	Teth514_1638	895.811
	9	Teth39_1750	-	Teth514_0482	3773.147
		Teth39_2028	8851.411	Teth514_0184	13875.16
Carbohydrate-Binding Module Family	20	Teth39_2173	1106.995	Teth514_1781	336.8139
		Teth39_0672	7972.532		
	34	Teth39_0676[c]	15885.95	Teth514_1187	4318.655
		Teth39_2173	1106.995	Teth514_1781	1010.442

[a] Family designations and number of genes in a family are derived from the CAZy database

[b] Mean normalized background-deducted signal intensity (3 replicates)

[c] Group probes (see Materials and Methods for description)

Table S4. Measured (Exp) and the Best Fitted (Calc.) Isotopomer Labeling Patterns in A) [1-^{13}C] Glucose (n=2, the errors of GC-MS data were below 2%) and B) [1-^{13}C] Xylose experiments. M0, M1, M2, M3 are fractions of unlabeled, singly labeled, doubly and triply labeled amino acids respectively. The amino acids labeling reflects their precursors (in central metabolic pathways) labeling patterns and thus amino acid labeling information can be used to determine the fluxes through central metabolic pathways.

A)

[M-57] Fragment	M0	M1	M2	M3
Alanine_X514				
Exp.	0.53	0.46	0.00	0.00
Calc.	0.53	0.47	0.01	0.00
Alanine_39E				
Exp.	0.55	0.44	0.02	0.00
Calc.	0.54	0.45	0.01	0.00
Serine_X514				
Exp.	0.56	0.43	0.01	0.00
Calc.	0.53	0.47	0.01	0.00
Serine_39E				
Exp.	0.56	0.42	0.02	0.00
Calc.	0.54	0.45	0.01	0.00
Aspartate_X514				
Exp.	0.50	0.47	0.02	0.00
Calc.	0.50	0.47	0.02	0.00
Aspartate_39E				
Exp.	0.51	0.46	0.03	0.00
Calc.	0.51	0.46	0.03	0.00
Glutamate_X514				
Exp.	0.29	0.48	0.22	0.01
Calc.	0.33	0.48	0.18	0.01
Glutamate_39E				
Exp.	0.31	0.47	0.21	0.01
Calc.	0.35	0.48	0.17	0.01
Phenylalanine_X514				
Exp.	0.26	0.38	0.28	0.07
Calc.	0.27	0.49	0.23	0.01
Phenylalanine_39E				
Exp.	0.30	0.36	0.26	0.07
Calc.	0.29	0.48	0.21	0.01

B)

[M-57] Fragment	M0	M1	M2	M3
Alanine_X514				
Exp.	0.64	0.32	0.03	0.00
Calc.	0.61	0.21	0.18	0.00
Alanine_39E				
Exp.	0.60	0.36	0.04	0.00
Calc.	0.61	0.21	0.18	0.00
Serine_X514				
Exp.	0.67	0.24	0.09	0.01
Calc.	0.61	0.21	0.18	0.00
Serine_39E				
Exp.	0.59	0.29	0.11	0.00
Calc.	0.61	0.21	0.18	0.00
Aspartate_X514				
Exp.	0.58	0.32	0.10	0.00
Calc.	0.58	0.32	0.10	0.00
Aspartate_39E				
Exp.	0.57	0.32	0.11	0.01
Calc.	0.57	0.32	0.11	0.01
Glutamate_X514				
Exp.	0.40	0.45	0.15	0.00
Calc.	0.41	0.42	0.14	0.02
Glutamate_39E				
Exp.	0.38	0.45	0.16	0.01
Calc.	0.40	0.42	0.15	0.02
Phenylalanine_X514				
Exp.	0.37	0.27	0.20	0.11
Calc.	0.38	0.35	0.19	0.07
Phenylalanine_39E				
Exp.	0.30	0.31	0.24	0.11
Calc.	0.38	0.36	0.19	0.07

Table S5. A) Genes significantly upregulated (LogR > 2) in 39E (A) or X514 (B) grown on 10 mM xylose vs. 200 mM xylose. Yellow boxes indicate putative xylose metabolism/transport genes shown in Fig. 2.

A)

Gene Name[a]	Probe[b]	Annotation	LogR	Z-Score
Teth39_0103		UDP-N-acetylglucosamine 1-carboxyvinyltransferase	2.389603	2.376784
Teth39_0148	GR	Integrase catalytic subunit	2.12111	3.187191
Teth39_0193		ABC transporter related	3.45394	2.763824
Teth39_0194		ABC-2 type transporter	2.56742	3.160152
Teth39_0476	GR	Ku domain-containing protein	3.481872	2.020556
Teth39_0545	GR	Beta-lactamase domain-containing protein	2.101631	3.783784
Teth39_1184	GR	Aspartate aminotransferase	2.133548	3.840099
Teth39_1194	GR	MiaB-like tRNA modifying enzyme YliG	4.123493	5.907747
Teth39_1385		Integrase catalytic subunit	2.399756	3.293958
Teth39_1501	GR	Aspartate-semialdehyde dehydrogenase	2.052442	3.164596
Teth39_1623		Glycosyl transferase, group 1	2.03841	2.857166
Teth39_1654	GR	Hydroxyethylthiazole kinase	2.441081	4.241612
Teth39_1692	GR	Aspartyl/glutamyl-tRNA amidotransferase subunit A	2.116651	4.233301
Teth39_1914		Metal-dependent hydrolase with the TIM-barrel fold-like protein	2.18458	3.137472
Teth39_2053		ROK family protein (***xylR***)	1.864403	2.60898
Teth39_2054		Beta-glucosidase (***cglT***)	0.190124	0.356071
Teth39_2055		Glycoside hydrolase family 3 protein (***xarB***)	-0.12682	-0.23997
Teth39_2056		Binding-protein-dependent transport systems inner membrane component (***xarG***)	0.084884	0.159501
Teth39_2057		Binding-protein-dependent transport systems inner membrane component (***cglF***)	1.080483	1.772103
Teth39_2058		Extracellular solute-binding protein	1.385484	1.97418
Teth39_2059		Xylulokinase (***xylB***)	0.825173	0.995909
Teth5140101		Hypothetical protein	2.272177	2.593625
Teth5140132		PTS system fructose subfamily IIA component	2.882356	3.114153
Teth5140133		Phosphotransferase system PTS, sorbose-specific IIC subunit	2.640958	4.348187
Teth5140134		PTS system mannose/fructose/sorbose family IID component	2.630022	3.722983
Teth5140153		Xylose isomerase (***xylA***)	-1.20693	-1.33222
Teth5140154		Xylulokinase (***xylB***)	-0.64707	-0.84262
Teth5140160		Oxidoreductase domain-containing protein	1.238789	1.875806
Teth5140161		Periplasmic binding protein/LacI translational regulator	-0.54465	-1.08807
Teth5140162		Ribokinase	-0.51676	-0.9877
Teth5140163		D-ribose pyranase	-0.51217	-1.01599
Teth5140164		ABC transport related	-1.07528	-2.0779
Teth5140165		Monosaccharide-transporting ATPase	0.001387	0.002658
Teth5140166		Monosaccharide-transporting ATPase	1.722688	2.798406
Teth5140412		PTS system, N-acetylglucosamine-specific IIBC subunit	3.968167	4.518356
Teth5140653		Alcohol dehydrogenase (***adhB***)	3.415051	2.65966

Teth5140668		Ferritin, Dps family protein	2.418374	2.254949
Teth5140965		CopG family transcriptional regulator	2.056221	2.248755
Teth5141018		Glycerol-3-phosphate responsive antiterminator	4.72931	2.646766
Teth5141019		MIP family channel protein	4.793331	3.418584
Teth5141020		Glycerol kinase	3.601087	4.060061
Teth5141041		50S ribosomal protein L25/general stress protein Ctc	2.164301	1..079815
Teth5141110		Putative transaldolase	3.211802	2.648733
Teth5141306		Glyceraldehyde-3-phosphate dehydrogenase, type I	2.444676	2.019509
Teth5141309		Phosphoglyceromutase	2.145191	2.497524
Teth5141311		Preprotein translocase subunit SecG	2.848063	3.054025
Teth5141443		NUDIX hydrolase	2.655174	4.52064
Teth5141546		Response regulator receiver protein	2.466589	1.955731
Teth5141571		Sodium:dicarboxylate symporter	2.316901	3.515017
Teth5141718		Stage V sporulation protein S	2.988586	4.020257

B)

Gene Name[a]	Probe[b]	Annotation	LogR	Z-Score
Teth39_0031	GR	Phospholipase D/transphosphatidylase	3.109077	5.465829
Teth39_0076	GR	Peptidase A24A, prepilin type IV	2.515622	3.960918
Teth39_0144	GR	Alanine—glyoxylate transaminase	6.830311	3.744652
Teth39_0170	GR	PpiC-type peptidyl-prolyl cis-trans isomerase	2.133067	3.3066
Teth39_0217	GR	Ferredoxin-NADP(+) reductase subunit alpha	3.2686	6.032877
Teth39_0223	GR	N-acetyl-gamma-glutamyl-phosphate reductase	2.00148	3.449673
Teth39_0236	GR	Hypothetical protein	2.267868	3.838706
Teth39_0238	GR		2.334038	3.752496
Teth39_0318	GR	Lysyl-tRNA synthetase	2.29713	1.382165
Teth39_0318		Lysyl-tRNA synthetase	2.344591	3.80228
Teth39_0325	GR	Hypothetical protein	2.46839	2.47385
Teth39_0336	GR	Elongation factor G	3.85813	2.256987
Teth39_0348	GR	2-C-methyl-D-erythritol 2,4-cyclodiphosphate synthase	3.663349	5.500018
Teth39_0368	GR	Putative L7Ae-like ribosomal protein	3.593758	3.680345
Teth39_0430	GR	Hypothetical protein	2.506383	3.990389
Teth39_0454	GR	Osmosensitive K+ channel histidine kinase-like protein	2.832809	4.641398
Teth39_0476	GR	Ku domain-containing protein	2.249851	3.918101
Teth39_0484	GR	4'-phosphopantetheinyl transferase	2.878185	4.970924
Teth39_0512	GR	DNA polymerase beta subunit	2.133439	2.088801
Teth39_0540	GR	Phosphotransferase system, phosphocarrier protein Hpr	2.244277	3.331138
Teth39_0546	GR		2.67001	4.330387
Teth39_0555	GR	DNA polymerase beta subunit	2.532627	3.515997
Teth39_0732	GR	Phosphoglycerate kinase	2.464548	4.686554
Teth39_0772	GR	Phenylalanyl-tRNA synthetase subunit beta	3.067554	5.039841
Teth39_0775	GR	Recombination and DNA strand exchange inhibitor protein	2.419387	3.949334
Teth39_0784	GR	Hypothetical protein	4.702836	4.569688
Teth39_0856	GR	Binding-protein-dependent transport systems inner membrane component	2.201499	4.027396
Teth39_0917	GR	CRISPR-associated TM1812 family protein	4.16446	7.200551
Teth39_0922		Hypothetical protein	2.790735	3.201333
Teth39_0924		DEAD/DEAH box helicase domain-containing protein	2.431041	2.36841
Teth39_0933	GR	Hypothetical protein	2.525085	4.049361
Teth39_0984	GR	3-deoxy-7-phosphoheptulonate synthase	2.079062	3.915808
Teth39_0989	GR	Hypothetical protein	2.278271	3.351146
Teth39_0997	GR	Tranposase IS116/IS110/IS902 family protein	2.772267	4.821744
Teth39_0998	GR	Vitamin B_{12} dependent methionine synthase, activation region	4.150209	2.75148
Teth39_1014	GR	Holliday junction DNA helicase RuvB	2.761315	4.592518
Teth39_1056	GR	Ferric uptake regulator family protein	2.576854	4.638333
Teth39_1118	GR	Tyrosine recombinase XerD	3.778524	2.073802
Teth39_1171	GR	S-adenosylmethionine decarboxylase proenzyme	2.271755	3.088048
Teth39_1172	GR	MCP methyltransferase, CheR-type	2.215164	2.107381
Teth39_1210	GR	Ribosomal protein L7Ae/L30e/S12e/Gadd45	2.222454	2.042635
Teth39_1253	GR	Flagellar export protein FliJ	3.990691	2.360236
Teth39_1295	GR	Acetate kinase	2.200354	2.352667
Teth39_1308	GR	50S ribosomal protein L28	2.835086	1.820278
Teth39_1350	GR	Lipoprotein signal peptidase	2.645365	2.523519
Teth39_1377		Diguanylate cyclase with PAS/PAC sensor	3.735285	4.549558

Teth39_1407	GR	Beta-lactamase domain-containing protein	2.98157	4.28029
Teth39_1501	GR	Aspartate-semialdehyde dehydrogenase	3.86909	6.175262
Teth39_1654	GR	Hydroxyethylthiazole kinase	2.370444	4.048238
Teth39_1689	GR	Extracellular solute-binding protein	2.130517	4.019827
Teth39_1705	GR	Amidohydrolase	3.329449	2.259527
Teth39_1729	GR	Ferredoxin-dependent glutamate synthase	4.122489	7.222698
Teth39_1741	GR	D-isomer specific 2-hydroxyacid dehydrogenase, NAD-binding	3.77008	6.366967
Teth39_1754	GR	Peptide chain release factor 2	2.751738	4.606285
Teth39_1758	GR	3H domain-containing protein	2.491401	4.39875
Teth39_1786	GR	Hypothetical protein	2.896312	4.585538
Teth39_1826	GR	Pyrimidine-nucleoside phosphorylase	4.11471	2.518064
Teth39_1912	GR		2.776936	4.276916
Teth39_1927	GR	Beta-glucosidase	2.796181	2.683315
Teth39_1937		nifR3 family TIM-barrel protein	2.890973	5.781947
Teth39_1941	GR	Transcriptional antiterminator, BglG	2.058308	3.809609
Teth39_1955	GR	AzlC family protein	2.024577	2.595737
Teth39_2025	GR	FeoA family protein	2.085473	2.196582
Teth39_2046	GR		3.53747	4.040259
Teth39_2095	GR	Sporulation transcriptional regulator SpoIIID	2.101927	3.395
Teth39_2108	GR	Zinc/iron permease	2.279533	4.212248
Teth39_2177		Hypothetical protein	2.748521	3.301702
Teth39_2190		Short-chain dehydrogenase/reductase SDR	2.278128	3.927205
Teth39_2242		Excision promoter, Xis	4.983522	5.187201
Teth39_2270	GR	Phosphoesterase domain-containing protein	2.670977	3.415358
Teth39_2280	GR	Cyanophycinase	2.778993	4.952121
Teth5140003		RNA-binding S4 domain-containing protein	3.920171	5.571918
Teth5140009		DNA gyrase subunit B	2.514189	3.874428
Teth5140011		NADH:flavin oxidoreductase/NADH oxidase	2.279398	3.990275
Teth5140029		GDSL family lipase	2.220708	2.963159
Teth5140031		CMP/dCMP deaminase, zinc-binding	2.890645	3.882952
Teth5140076		Hypothetical protein	2.120932	1.525608
Teth5140079		Electron transport complex, RnfABCDGE, C subunit	2.275948	2.944447
Teth5140087		Response regulator receiver protein	2.986993	4.434851
Teth5140101		Hypothetical protein	3.522331	2.458532
Teth5140104		Alanine dehydrogenase	2.103456	3.495919
Teth5140116		Hypothetical protein	2.10213	2.698261
Teth5140118		Prevent-host-death family protein	3.246939	3.566221
Teth5140153		Xylose isomerase (***xylA***)	0.634055	0.763193
Teth5140154		Xylulokinase (***xylB***)	1.548282	2.024848
Teth5140155		D-xylose ABC transporter, periplasmic substrate-binding protein (***xylF***)	3.827646	2.613255
Teth5140156	GR	RNA-directed DNA polymerase (reverse transcriptase)	2.646491	4.693521
Teth5140157		ABC transporter related (***xylG***)	2.457269	3.387036
Teth5140158		Monosaccharide-transporting ATPase (***xylH***)	2.215879	3.655725
Teth5140159		ROK family protein (***xylR***)	3.381175	5.198676
Teth5140160		Oxidoreductase domain-containing protein	4.756193	3.735705
Teth5140164		ABC transport related	2.289943	4.12144
Teth5140165		Monosaccharide-transporting ATPase	2.667733	4.645616
Teth5140168		RpiR family transcriptional regulator	-0.31733	-0.57835
Teth5140170		Phosphotransferase system, EIIC	1.216543	2.309444
Teth5140177		Hypothetical protein	3.138433	5.791016
Teth5140187		FeoA family protein	2.443894	4.887799
Teth5140188		Ferrous iron transport protein B	3.531465	5.509271

Teth5140203		UvrD-like DNA helicase, C terminal	2.458478	4.320088
Teth5140207		Putative peroxiredoxin	2.339683	2.113802
Teth5140208		Ferric uptake regulator family protein	2.661185	1.412727
Teth5140209		Hypothetical protein	3.305432	3.29811
Teth5140220		Hypothetical protein	2.580702	2.519616
Teth5140242		1A family penicillin-binding protein	2.20474	3.482971
Teth5140245		Regulatory protein, ArsR	2.443214	4.063751
Teth5140247		Heterodisulfide reductase subunit A and related polyferredoxins-like protein	2.493076	4.307302
Teth5140272		5'-nucleotidase domain-containing protein	5.233916	7.454533
Teth5140280		ATPase	3.296485	5.264682
Teth5140282		Hypothetical protein	3.054945	4.971037
Teth5140330		ABC-2 type transporter	2.300277	3.580614
Teth5140332		Hypothetical protein	2.893568	4.151117
Teth5140344		Hypothetical protein	2.537784	2.562114
Teth5140348		Regulatory protein, LuxR	2.081791	3.275169
Teth5140365		Hypothetical protein	2.373943	3.778382
Teth5140380	GR	Hypothetical protein	2.45469	4.674641
Teth5140388		Hypothetical protein	2.202133	4.15525
Teth5140392		Hypothetical protein	2.402512	2.817032
Teth5140400		Formiminotransferase-cyclodeaminase	2.274455	3.61386
Teth5140401		Basic membrane lipoprotein	2.404819	1.263238
Teth5140402		ABC transporter related	2.444793	3.671429
Teth5140409		Pyrimidine-nucleoside phosphorylase	3.522346	5.188863
Teth5140413		PTS system, glucose subfamily, IIA subunit	2.380748	3.672342
Teth5140437		Cof-like hydrolase	2.061141	1.798512
Teth5140439		Two component transcriptional regulator	2.092619	3.161568
Teth5140441		Hypothetical protein	2.429625	2.967785
Teth5140444		Rubrerythrin	3.053965	1.727636
Teth5140450		Hypothetical protein	2.340935	3.356738
Teth5140465		Flagellar protein FliS	2.098255	3.119845
Teth5140466		Hypothetical protein	3.655159	5.061468
Teth5140468		Extracellular solute-binding protein	2.895463	4.424365
Teth5140471		Sigma 54 modulation protein/ribosomal protein S30EA	2.342673	2.845479
Teth5140484		Ornithine carbamoyltransferase	2.208528	3.349262
Teth5140485		Carbamate kinase	2.787146	4.688643
Teth5140493		ABC transporter related	2.794735	5.104146
Teth5140498		LytTR family two component transcriptional regulator	2.373377	4.371139
Teth5140499		Signal transduction histidine kinase regulating citrate/malate metabolism	3.128076	5.119333
Teth5140503			2.35156	3.63454
Teth5140525		Bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase/IMP cyclohydrolase	4.085959	3.289196
Teth5140536		TrpR like protein, YerC/YecD	5.869534	8.625729
Teth5140537		ATP-dependent DNA helicase PcrA	2.221503	3.567181
Teth5140552		RdgB/HAM1 family non-canonical purine NTP pyrophosphatase	2.904813	4.051509
Teth5140553		Phosphodiesterase	2.736572	3.948126
Teth5140567		Hydroxyethylthiazole kinase	2.396024	3.500347
Teth5140579		Hypothetical protein	2.037113	2.825981
Teth5140585		SpoVG family protein	3.885039	2.253802
Teth5140594		PpiC-type peptidyl cis-trans isomerase	2.260107	2.026076
Teth5140613		Zn-dependent hydrolase of the beta-lactamase fold-like protein	2.846523	4.015741

Teth5140625		Aldo/keto reductase	2.756659	4.224322
Teth5140631		ABC transporter related	2.387651	4.190554
Teth5140643		Hypothetical protein	2.268909	3.798924
Teth5140668		Ferritin, Dps family protein	2.437752	1.936761
Teth5140675		FMN-binding domain-containing protein	2.539393	1.702173
Teth5140680		Desulfoferredoxin	4.735118	3.458693
Teth5140682		Aspartate ammonia-lyase	2.323949	3.851918
Teth5140688		RNA-binding S4 domain-containing protein	2.413943	3.138948
Teth5140692		2,3-diketo-5-methylthio-1-phosphopentane phosphatase	3.234853	5.402425
Teth5140702		XRE family transcriptional regulator	2.099238	3.33589
Teth5140709		Hypothetical protein	2.615121	3.182382
Teth5140722		Type II secretion protein E	2.030904	3.511218
Teth5140777		Phage holing	2.121036	2.323423
Teth5140790		Phosphoprotein phosphatase	2.422778	4.417694
Teth5140795		Formate—tetrahydrofolate ligase	2.038704	3.068944
Teth5140796		Hypothetical protein	2.24891	4.090475
Teth5140809		Transcription elongation factor GreA	3.116054	4.069502
Teth5140822		Diguanylate cyclase	3.322257	5.367249
Teth5140824		Putative PTS IIA-like nitrogen-regulatory protein PtsN	3.37901	3.333275
Teth5140828		Elongation factor G	2.279207	4.082642
Teth5140831		Hypothetical protein	2.400754	3.54879
Teth5140834		ATP:guanido phosphotransferase	2.418872	3.989228
Teth5140838		PilT domain-containing protein	2.243331	1.703292
Teth5140841		2-C-methyl-D-erythritol 2,3-dicyclodiphosphate synthase	2.19213	3.644063
Teth5140877		Ribosomal protein L14	2.125523	3.233164
Teth5140902		Ribosomal protein S9	3.223183	3.264063
Teth5140906		Ferritin, Dps family protein	2.511413	1.023514
Teth5140908		Hypothetical protein	4.604925	5.230062
Teth5140910		Amino acid carrier protein	4.160458	6.395228
Teth5140911		Cation antiporter	2.149327	1.841152
Teth5140913		Monovalent cation/proton antiporter, MnhG/PhaG subunit	2.41259	3.453718
Teth5140926		Cell wall hydrolase/autolysin	2.007329	3.14039
Teth5140927		HAD family hydrolase	2.942865	5.885729
Teth5140928		ATPase	2.596623	4.15873
Teth5140929		Hypothetical protein	2.092767	2.30559
Teth5140931		Osmosensitive K+ channel histidine kinase-like protein	2.033945	4.06789
Teth5140934		Hypothetical protein	2.092767	2.30559
Teth5140936		Butyrate kinase	3.268833	4.868477
Teth5140939		Butyrate kinase	2.228817	3.209599
Teth5140942		2-ketoisovalerate ferredoxin reductase	3.036	4.930148
Teth5140943		Thiamine pyrophosphate binding domain-containing protein	2.35268	4.074368
Teth5140944		Pyruvate/ketoisovalerate oxidoreductase, gamma subunit	3.782455	4.922674
Teth5140950		Glucosamine—fructose-6-phosphate aminotransferase	2.05201	3.466556
Teth5140958		Helix-turn-helix domain-containing protein	2.2446	2.667075
Teth5140963		Outer membrane lipoprotein-sorting protein-like protein	2.221899	3.61729
Teth5140970		Ferredoxin domain-containing protein	2.573109	3.736709
Teth5140984		Hypothetical protein	3.859316	5.093197
Teth5140988		Oxidoreductase domain-containing protein	3.540951	5.728887
Teth5140994			2.045326	2.729542
Teth5141009		Membrane dipeptidase	2.051335	3.521363
Teth5141015		CRISPR-associated Cas4 family protein	2.984754	3.593296
Teth5141019		MIP family channel protein	2.727976	4.974765

Teth5141026		Dihydroxyacetone kinase, L subunit	2.126793	3.735766
Teth5141027		Dihydroxyacetone kinase, phosphotransfer subunit	2.200692	3.330512
Teth5141032		Hypothetical protein	2.56077	2.913144
Teth5141041		50S ribosomal protein L25/general stress protein Ctc	2.543907	1.45025
Teth5141044		ABC transporter related	3.374409	5.416184
Teth5141046		Cell division ATP-binding protein FtsE	2.073221	3.255604
Teth5141056		Peptidoglycan glycotransferase	2.063054	3.268801
Teth5141064		Hypothetical protein	3.268581	4.920468
Teth5141069		Hypothetical protein	4.232279	4.985553
Teth5141076		Xylose isomerase domain-containing protein	2.033276	3.528262
Teth5141084		Carbohydrate kinase, FGGY	2.166255	3.772659
Teth5141110		Putative transaldolase	2.581878	1.797656
Teth5141112		Peptidoglycan glycosyltransferase	2.29883	3.638873
Teth5141121		Hypothetical protein	2.276843	3.164057
Teth5141138		FeoA family protein	2.572091	2.909938
Teth5141139		FeoA family protein	2.149638	2.58462
Teth5141149			2.107863	3.376185
Teth5141151		Hypothetical protein	2.13525	3.55231
Teth5141175		Hypothetical protein	3.415368	2.54092
Teth5141176		UDP-N-acetylenolpyruvateglucosamine reductase	2.070078	3.333465
Teth5141201		Glycine betaine/L-proline ABC transporter, ATPase subunit	2.417176	4.099718
Teth5141216		RNA methyltransferase	2.070727	3.23289
Teth5141279		Dipeptidyl aminopeptidase/acylaminoacyl-peptidase-like protein	2.125236	3.223083
Teth5141283		Hypothetical protein	2.143298	2.727714
Teth5141284		Recombination factor protein RarA	2.545636	4.396441
Teth5141295		4Fe-4S ferredoxin iron-sulfur binding domain-containing protein	2.473042	1.319715
Teth5141313		DNA primase	2.289999	3.734526
Teth5141314		RNA polymerase sigma factor RpoD	2.937167	4.110723
Teth5141318		Purine nucleoside phosphorylase	3.461189	3.777737
Teth5141321		Hypothetical protein	2.257228	3.823035
Teth5141349		Signal-transduction protein	3.177698	3.736726
Teth5141351		Pyruvate phosphate dikinase	2.000114	1.741521
Teth5141353		Ribonuclease R	2.767974	2.886517
Teth5141354		SsrA-binding protein	2.737197	2.816526
Teth5141356		Helix-turn-helix domain-containing protein	2.5004	3.409252
Teth5141361		Hypothetical protein	4.524104	4.260518
Teth5141375		Hypothetical protein	2.20283	2.153518
Teth5141377		PfpI family intracellular peptidase	2.126401	3.483912
Teth5141385		Molybdopterin-guanine dinucleotide biosynthesis protein B	2.933208	4.694006
Teth5141386		Extracellular solute-binding protein	3.312897	4.64773
Teth5141391		Molybdonum cofactor synthesis domain-containing protein	2.69385	3.907138
Teth5141396		ABC-type sulfate transport system permease component-like protein	2.276033	3.23263
Teth5141398		ABC-type nitrate/sulfonate/bicarbonate transport system ATPase component-like protein	2.416034	3.299214
Teth5141400		Fis family transcriptional regulator	3.44147	6.882939
Teth5141401		Alpha/beta fold family hydrolase/acetyltransferase-like protein	3.283241	4.330729
Teth5141405		Glycerophosphodiester phosphodiesterase	3.060605	5.505382

Teth5141407		Appr-1-p processing domain-containing protein	2.306658	3.761112
Teth5141418		Cyclase family protein	2.060523	2.893529
Teth5141422		Oligoendopeptidase F	2.514417	4.038555
Teth5141430		Hypothetical protein	5.699281	6.121539
Teth5141432		ROK family protein	2.549402	1.734732
Teth5141435		PRC-barrel domain-containing protein	2.251503	3.519211
Teth5141440		SpoIID/LytB domain-containing protein	2.186389	2.979809
Teth5141443		NUDIX hydrolase	2.815032	4.844273
Teth5141446		Hypothetical protein	2.540772	4.556306
Teth5141447		Hypothetical protein	2.468358	3.779976
Teth5141450		Putative CheW protein	2.425842	2.384401
Teth5141457		Queuine tRNA-ribosyltransferase	2.968284	3.770184
Teth5141458		Preprotein translocase, YajC subunit	2.026017	1.382466
Teth5141464		2-polyprenylphenol 6-hydroxylase	2.46005	4.258077
Teth5141466		Adenine phosphoribosyltransferase	2.283977	2.8199
Teth5141475		Hypothetical protein	3.005261	4.517686
Teth5141476		Thioredoxin	2.093815	0.987545
Teth5141492		Hypothetical protein	2.328542	0.880019
Teth5141514		Tfp pilus assembly protein ATPase PilM-like protein	3.106608	5.039671
Teth5141524		Hypothetical protein	2.29052	2.025209
Teth5141544		DNA repair protein RecN	2.027438	3.337333
Teth5141546		Response regulator receiver protein	2.799524	1.811559
Teth5141548		Hypothetical protein	2.631803	4.501782
Teth5141555		Tyrosine recombinase XerD	2.2045	3.214272
Teth5141564		Polynucleotide adenylyltransferase region	2.444067	3.832156
Teth5141570		Hypothetical protein	2.56848	4.538884
Teth5141571		Sodium:dicarboxylate symporter	2.133193	3.124323
Teth5141586		4Fe-4S ferredoxin iron-sulfur binding domain0containing protein	3.534196	5.427275
Teth5141606		4-hydroxy-3-methylbut-2-enyl diphosphate reductase	2.944546	4.682649
Teth5141608		S-adenosylmethionine decarboxylase proenzyme	3.467849	2.984746
Teth5141612		DNA mismatch repair protein	2.3036	3.812741
Teth5141614		RNA chaperone Hfq	3.440702	3.629384
Teth5141616		Ribonuclease H	2.038216	3.017599
Teth5141623		Stage V sporulation protein S	2.080572	2.56289
Teth5141625		Phosphodiesterase	2.122743	1.41682
Teth5141630		MiaB-like tRNA modifying enzyme YliG	2.659754	4.459355
Teth5141644		Ribosome-binding factor A	2.088347	3.274192
Teth5141659		Uridylate kinase	2.135227	3.55959
Teth5141661		30S ribosomal protein S2	4.413536	4.84516
Teth5141665		Chemoreceptor glutamine deamidase CheD	2.08766	3.133233
Teth5141683		Flagellar FlbD family protein	2.600519	4.137133
Teth5141692		Flagellar motor switch protein G	2.084497	3.460961
Teth5141697		Transcriptional repressor CodY	2.327557	3.448666
Teth5141700		DNA topoisomerase I	2.258315	3.928035
Teth5141708		50S ribosomal protein L19	2.836612	3.360741
Teth5141709		tRNA (guanine-N1)-methyltransferase	2.913635	4.610771
Teth5141710		16S rRNA-processing protein RimM	2.54333	2.210682
Teth5141711		Hypothetical protein	2.476659	2.339879
Teth5141712		30S ribosomal protein S16	2.896695	3.25301
Teth5141718		Stage V sporulation protein S	3.186851	3.012012
Teth5141722		Acyl carrier protein	2.651926	3.117647
Teth5141727		Putative glycerol-3-phosphate acyltransferase PlsX	2.010254	3.178425
Teth5141728		Fatty acid biosynthesis transcriptional regulator	2.730844	3.982385

Teth5141734		Nucleoside recognition domain-containing protein	2.208076	3.538515
Teth5141749		Protein kinase	2.061772	2.487364
Teth5141759		Phosphopantothenoylcysteine decarboxylase/phosphopantothenate—cysteine ligase	2.282969	3.628668
Teth5141790		Glycoside hydrolase family protein	3.580817	6.53919
Teth5141791		Hypothetical protein	3.194939	5.222201
Teth5141792		Oligopeptide/dipeptide ABC transporter, ATPase subunit	2.854088	4.314276
Teth5141794		Binding-protein-dependent transport systems inner membrane component	2.680633	4.006314
Teth5141795		Binding-protein-dependent transport systems inner membrane component	3.109956	4.585416
Teth5141800		Hypothetical protein	2.298241	3.126508
Teth5141812		Peptidoglycan binding domain-containing protein	2.524918	3.815661
Teth5141824		Hypothetical protein	2.241168	3.554389
Teth5141835		Hypothetical protein	3.586377	6.304435
Teth5141840		Hypothetical protein	3.762125	4.983716
Teth5141861		Alkyl hydroperoxide reductase/Thiol specific antioxidant/ Mal allergen	4.405177	2.638748
Teth5141862		3-deoxy-7-phosphoheptulonate synthase	3.089054	5.385466
Teth5141864		Tryptophan synthase subunit beta	2.270949	3.741344
Teth5141895		Gluconate transporter	3.619863	5.473214
Teth5141896		Glycerate kinase	2.607584	3.863878
Teth5141897		Transcriptional regulator, CdaR	2.064158	3.309437
Teth5141944		Ethanolamine utilization protein EutN/carboxysome structural protein CcmL	2.022255	3.242498
Teth5141951		Dehydratase, small subunit	2.908228	2.826412
Teth5141973		SecC motif-containing protein	3.490712	4.953322
Teth5141979			2.191544	3.971047
Teth5141991		Phosphate uptake regulator, PhoU	2.392451	3.214507
Teth5141992		Phosphate ABC transporter, ATPase subunit	2.067244	1.453443
Teth5141997		Two component transcriptional regulator	2.114315	3.438774
Teth5142020		Peptidase M16 domain-containing protein	2.297656	3.503344
Teth5142022		GTP-dependent nucleic acid-binding protein EngD	2.117769	3.659775
Teth5142023		Hypothetical protein	2.536402	2.156856
Teth5142024		Thiamine biosynthesis protein ThiI	2.708455	4.601611
Teth5142045		Hypothetical protein	2.120517	3.622403
Teth5142046		Recombination and DNA strand exchange inhibitor protein	2.658792	4.231866
Teth5142054		Translation initiation factor IF-3	2.244771	2.739884
Teth5142058		Hypothetical protein	2.093757	2.340223
Teth5142062		Thioredoxin domain-containing protein	2.177353	3.299756
Teth5142067		NifU domain-containing protein	2.379058	1.59385
Teth5142068		Putative GAF sensor protein	2.675	1.924457
Teth5142070		Dipeptidase PepV	2.669946	3.983
Teth5142076		16S ribosomal RNA methyltransferase RsmE	2.931072	4.672839
Teth5142086		Hypothetical protein	2.813311	4.677185
Teth5142090		30S ribosomal protein S20	4.676364	2.169984
Teth5142095		Hypothetical protein	2.213789	3.15314
Teth5142115		50S ribosomal protein L27	2.428564	2.63023
Teth5142116		Hypothetical protein	2.273609	2.822519
Teth5142144		Hypothetical protein	2.180876	3.602593
Teth5142151		NLP/P60 protein	2.479942	3.561487
Teth5142152		Extracellular solute-binding protein	2.299702	3.872592
Teth5142169		Hypothetical protein	2.051738	3.424056

Teth5142170		Histidine ammonia-lyase	2.856186	4.409293
Teth5142187		ATP:corrinoid adenosyltransferase BtuR/CobO/CobP	3.487401	4.469198
Teth5142206		Hypothetical protein	3.718153	3.557741
Teth5142221		CDP-diacylglycerol—glycerol-3-phosphate 3-phosphatidyltransferase	2.070389	3.539808
Teth5142222		Hypothetical protein	3.785108	2.43132
Teth5142225		Hypothetical protein	2.557033	2.301737
Teth5142233		Cell envelope-related transcriptional attenuator	2.533995	3.951654
Teth5142234		Protein-tyrosine-phosphatase	2.540322	4.207084
Teth5142239		NusG antitermination factor	3.288117	5.017937
Teth5142243		Hypothetical protein	2.248267	2.407161
Teth5142244		DNA polymerase beta subunit	4.685732	5.786027
Teth5142247		Hypothetical protein	2.249784	3.569659
Teth5142248		Hypothetical protein	3.520229	6.111619
Teth5142253		Membrane-bound proton-translocating pyrophosphatase	2.222296	3.471442
Teth5142267		Hypothetical protein	3.158032	3.71786
Teth5142285		Hypothetical protein	2.034347	1.737883
Teth5142286		IS605 family transposase OrfB	4.498655	5.164699
Teth5142288		Site-specific integrase-resolvase-like protein	4.090045	5.530868
Teth5142289		Hypothetical protein	2.62397	4.331291
Teth5142291		Magnesium transporter	2.553319	4.318911
Teth5142318		CRISPR-associated Cas4 family protein	2.254152	3.499218
Teth5142328		NifU family SUF system FeS assembly protein	2.303673	3.444323
Teth5142344		Hypothetical protein	3.504614	4.193449
Teth5142345		Peptidase M23B	2.50955	3.799919
Teth5142348		Peptidase M1, membrane alanine aminopeptidase	2.053593	3.57074
Teth5142356		Methyl-accepting chemotaxis sensory transducer	2.5343	3.886583
Teth5142362		V-type ATP synthase subunit D	2.305604	3.03262
Teth5142372		Hypothetical protein	2.295187	3.80386
Teth5142393		Single-strand binding protein	3.510685	1.9131
Teth5142394		30S ribosomal protein S6	4.493078	3.256168
Teth5142403		parB-like partition protein	2.479312	4.14322
Teth5142405		parB-like partition protein	2.581513	4.067574
Teth5142410		Single-strand nucleic acid binding R3H domain-containing protein	2.248	3.116024

[a] Locus tag (39E, “Teth39_XXXX”; X514, “Teth514XXXX”)
[b] “GR” indicates group-specific probe

Supplementary Figures

Figure S1. Growth of 39E (●,○) and X514 (■,□) in defined medium at 60°C with 1.25 g/l xylose (●,■) and 7 g/l xylose (○,□).

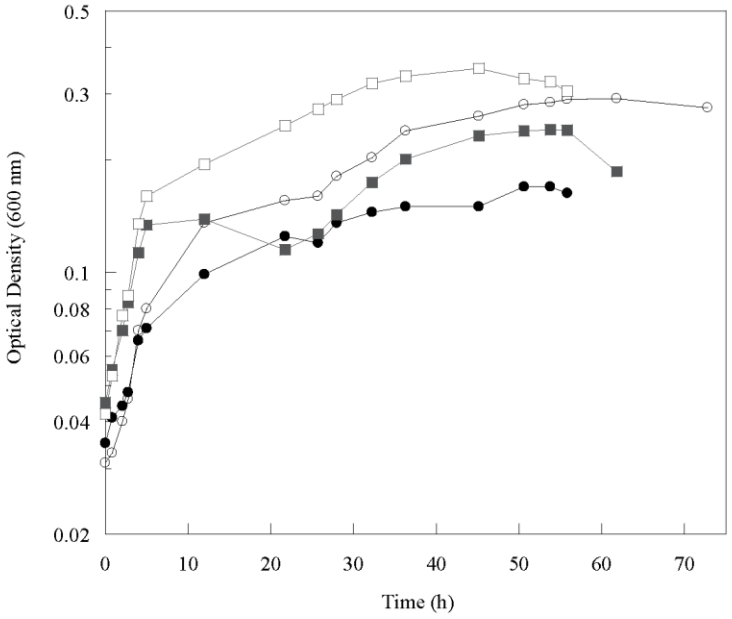


Figure S2. Microarray analysis of alcohol dehydrogenase genes of 39E and X514 grown on xylose. Concentrations used for growth were 1 mM (X514), 10 mM (both), and 200 mM (both). Lineage-specific genes include Teth514_0145 (no signal on array) and Teth39_1597. Starred bars indicate |z-score| > 2.0.

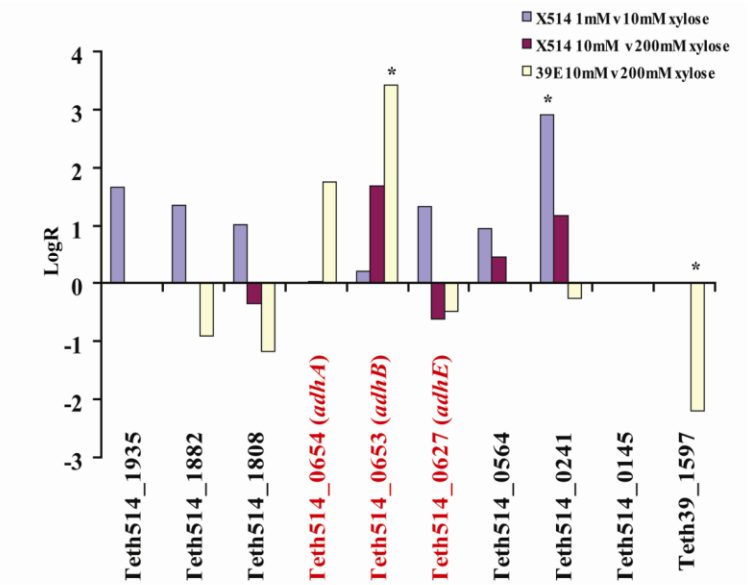
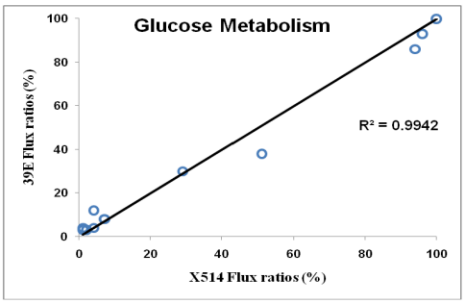


Figure S3. Correlation between flux distributions for 39E (y-axis) and X514 (x-axis) grown on (A) glucose and (B) xylose. The flux data are from Figure S3.

(A)



(B)

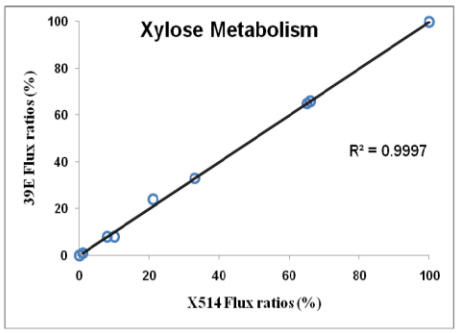
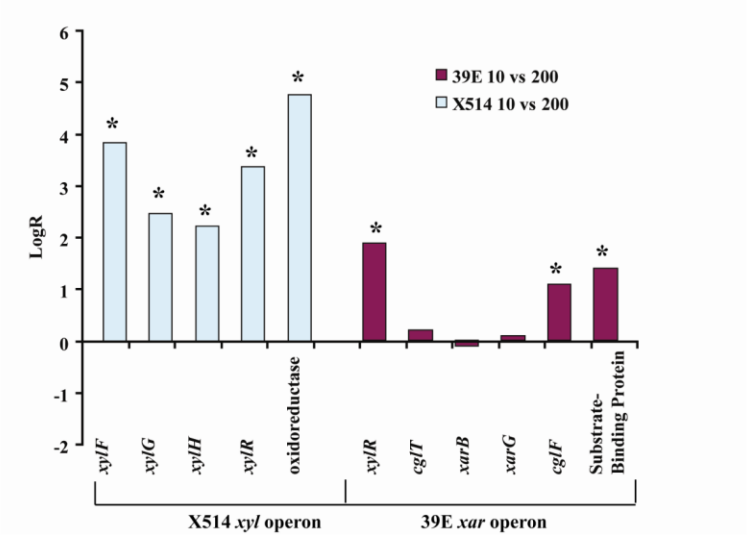


Figure S4. Microarray analysis of putative xylose transporters of 39E and X514 grown on xylose. Concentrations used for growth were 10 mM and 200 mM xylose. Asterisks indicate |Z-score| > 2.0. The putative *xylR* genes of the two strains are orthologs (90% amino acid identity) as are the associated oxidoreductase genes (95% amino acid identity) (see also, Fig. 2), though expression data is only available for the X514 oxidoreductase ortholog.



Appendix 8

Invariability of Central Metabolic Flux Distribution in *Shewanella oneidensis* MR-1 under Environmental or Genetic Perturbations

AIChE

Invariability of Central Metabolic Flux Distribution in *Shewanella oneidensis* MR-1 Under Environmental or Genetic Perturbations

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

Hector Garcia Martin
Fuels Synthesis Division, Joint BioEnergy Institute, Hollis, Emeryville, CA 94608

Adam Deutschbauer
Virtual Institute of Microbial Stress and Survival, Lawrence Berkeley National Laboratory, Berkeley, CA
Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA

Xueyang Feng and Rick Huang
Dept. of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO 63130

Xavier Llora
National Center for Supercomputing Applications, University of Illinois at Urbana-Champaign, Urbana, IL 61801

Adam Arkin
Virtual Institute of Microbial Stress and Survival, Lawrence Berkeley National Laboratory, Berkeley, CA
Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA
Dept. of Bioengineering, University of California, Berkeley, CA

Jay D. Keasling
Fuels Synthesis Division, Joint BioEnergy Institute, 5885 Hollis, Emeryville, CA 94608
Virtual Institute of Microbial Stress and Survival, Lawrence Berkeley National Laboratory, Berkeley, CA
Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA
Dept. of Bioengineering, University of California, Berkeley, CA
Dept. of Chemical Engineering, University of California, Berkeley, CA

DOI 10.1002/btpr.227
Published online July 16, 2009 in Wiley InterScience (www.interscience.wiley.com).

An environmentally important bacterium with versatile respiration, Shewanella oneidensis *MR-1, displayed significantly different growth rates under three culture conditions: minimal medium (doubling time ~3 h), salt stressed minimal medium (doubling time ~6 h), and minimal medium with amino acid supplementation (doubling time ~1.5 h). ^{13}C-based metabolic flux analysis indicated that fluxes of central metabolic reactions remained relatively constant under the three growth conditions, which is in stark contrast to the reported significant changes in the transcript and metabolite profiles under various growth conditions. Furthermore, 10 transposon mutants of* S. oneidensis *MR-1 were randomly chosen from a transposon library and their flux distributions through central metabolic pathways were revealed to be identical, even though such mutational processes altered the secondary metabolism, for example, glycine and C1 (5,10-Me-THF) metabolism.* © 2009 American Institute of Chemical Engineers *Biotechnol. Prog.,* 25: 1254–1259, 2009
Keywords: growth rate, ^{13}C-based, transposon mutants, transcript, metabolite profiles, secondary metabolism

Introduction

Shewanella oneidensis MR-1 has versatile respiration and can engage in cometabolic bioremediation of a diverse number of environmental contaminants, such as chromium, radionuclides, and halogenated organic compounds.[1–3] In addition, its ability to transfer electrons to solid metals

Additional Supporting Information may be found in the online version of this article.

Correspondence concerning this article should be addressed to J. D. Keasling at keasling@berkeley.edu.

Table 1. *Shewanella oneidensis* MR-1 Mutants Tested in This Study

Mutant	Growth Rate (h^{-1})	Gene	Description
MR-1	0.23 ± 0.04	Wild type	Minimal medium (no stress)
MR-1	0.10 ± 0.02	Wild type	Salt stress
1	0.20 ± 0.03	SO0781	Glycine cleavage system protein P; gcvP
2	*	Intergenic	None
3	*	SO3925	Biotin synthase family protein
4	*	SO_A0184	Conserved hypothetical protein
5	*	SO1420	Outer membrane porin, putative
6	*	SO2867	Permease PerM, putative
7	*	SO1143	Conserved hypothetical protein
8	*	SO3230	Flagellar regulatory protein C; flrC
9	*	SO4529	RNA methyltransferase, TrmH family, group 2
10	*	SO3106	Cold-active serine alkaline protease; aprE

* Growth rates for mutant 2 to 10 were same as the MR-1 wild type.

indicates its potential application in microbial fuel cells.[1] Previous studies on *S. oneidensisis* MR-1 have examined its transcript and metabolite profiles in response to various growth conditions.[4 7] However, cell physiology might not be accurately reflected by the annotated genome or by the transcript, protein, and metabolite profiles.[8 10] For example, *E. coli*'s transcript profile can have little relationship to the metabolic flux profile due to post-transcriptional regulation of protein synthesis and enzymes activity.[10,11] Since one of the most physiologically relevant descriptions of a cell's metabolism is the set of metabolic fluxes (a key determinant of cellular physiology),[8] we investigated *S. oneidensisis* MR-1's flux distributions in response to environmental and genetic perturbations, and thus its metabolic robustness (a recently recognized microbial phenotype[12 14]) in the face of environmental uncertainty and genetic perturbations. This study improves our understanding of *Shewanella* phenotypes and gene regulation attributed to the nature of adaptation in their environment. The variation of flux distributions in the MR-1 transposon mutants can reveal the possible impact of genetic modification on central metabolism, i.e., from an evolutionary robustness viewpoint,[15] whether the random nature of mutational processes (e.g., genetic drift or horizontal gene transfer) in the environment can have a significant effect on central flux distributions in *Shewanella* species.[16]

Materials and Methods

Culture conditions and analytical methods for metabolites

Shewanella oneidensis MR-1 was obtained from The American Type Culture Collection (ATCC 700550). Ten MR-1 mutants were provided by Adam Arkin's Group at University of California, Berkeley (Table 1), which were obtained via Tn5 transposon mutagenesis using plasmid pRL27.[17] In brief, mutagenesis was carried out by conjugation of the *E. coli* donor strain WM3064 harboring pRL27 with *Shewanella oneidensis* MR-1; transposon mutants were selected on Luria Broth (LB) medium with kanamycin (10 μg/mL); the insertion location of transposon mutants was mapped via a two-step degenerate PCR protocol as described previously.[18] From over 1000 mutants, 10 strains were randomly picked and investigated how natural mutation could affect the central flux distribution.

All MR-1 strains were grown in the defined [3-^{13}C] sodium L-lactate (98%, Cambridge Isotope, USA) minimal medium (30 mM, pH = 7, buffered with 20 mM PIPES) and duplicate experiments were performed (n = 2).[19] For salt stress experiments, 330 mM NaCl was added to the medium. To enhance growth, a standard amino acid mix (containing 17 unlabeled amino acids, without tryptophan, glutamine, and asparagine, Cat#AA-S-18, Sigma, USA) was added to the medium for a final concentration of 25 μM for each amino acid. The inoculum was prepared in LB medium and incubated overnight; the cultures were started with a 0.09% inoculation volume. All cultures (12 mL) were incubated in glass tubes at 30°C and 200 rpm shaking speed. Total biomass growth was monitored by measuring the OD_{600} and converting it to the corresponding dry weight (specifically, the harvested culture was centrifuged at 4,800*g* for 20 min and lyophilized overnight; then the dried biomass was weighed to obtain a correlation curve between dried biomass and its corresponding OD_{600}). The concentrations of lactate, acetate, and pyruvate in the medium were measured using enzyme kits (r-Biopharm, Darmstadt, Germany).

The GC-MS protocol for isotopomer measurement has been described previously.[9,20,21] In brief, the protein in biomass from the early exponential growth phase (OD_{600} = 0.3–0.4) was hydrolyzed to free amino acids in 6 M HCl at 100°C for 24 h. GC-MS samples were prepared in 100 μL tetrahydrofuran (THF) and 100 μL N-(tert-butyldimethylsilyl)-N-methyl-trifluoroacetamide (Sigma-Aldrich, USA) and baked at 65–80°C for one hour. GC-MS analysis was carried out using a gas chromatograph (DB5 column, HP6890 series, Agilent Inc, USA) equipped with a mass spectrometer (5973 Network, Agilent Inc, USA). Two types of positively charged species were used in flux calculation: unfragmented amino acids, $[M\text{-}57]^+$, and fragmented amino acids without α carboxyl group, $[M\text{-}159]^+$.[22]

Algorithm for isotopomer analysis and flux calculation

The metabolic network used to calculate the metabolic flux profile of *S. oneidensisis* MR-1 included the tricarboxylic acid (TCA) cycle (including the glyoxylate shunt), C1 metabolism, the Entner-Doudoroff (ED) pathway, gluconeogenesis, and the pentose phosphate (PP) pathway (total 43 reactions, Supporting Information Table 1). The algorithm for metabolic flux calculation has been described in detail in our previous papers.[15,23] In brief, the flux profile was represented through an independent fluxes vector $\boldsymbol{v}_{\mathbf{ind}}$, which comprised 16 independent fluxes plus the unlabeled fractions of CO_2 and C1, for a total of 18 independent variables. The carbon labeling expected from a flux profile $\boldsymbol{v}_{\mathbf{ind}}$ was calculated via the cumomer method[24] and the best fit for the data was calculated by minimizing the error function (i.e. average difference between the calculated and the experimental data)

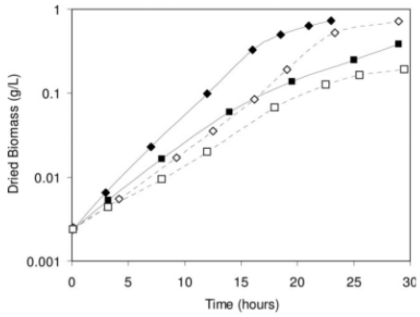


Figure 1. Growth kinetics of *Shewanella oneidensis* MR-1 under salt stress condition and with amino acid supplements ($n = 2$).

◆, minimal medium with amino acids supplement, no salt; ◇, minimal medium, no salt; ■, minimal medium with salt (0.33 M NaCl supplemented into minimal medium) and amino acids (17 amino acids, 25 μM each), □, minimal medium with salt (0.33 M NaCl supplementation into minimal medium), no amino acids.

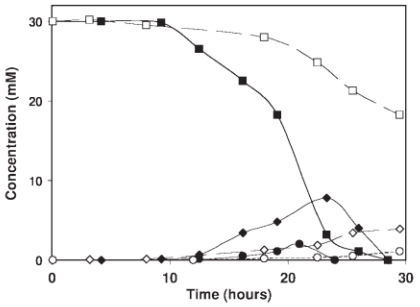


Figure 2. Lactate consumption and metabolite production by *Shewanella oneidensis* MR-1 under normal growth or salt stress conditions in minimal medium (n = 2).

Normal growth conditions: ■, lactate; ◆, acetate; ●, pyruvate. Salt stress (0.33 M NaCl supplemented): □, minimal medium; ◇, acetate; ○, pyruvate.

through the use of genetic algorithms.[25] The independent fluxes to biomass pools were not fixed by measured values, but were loosely constrained by MR-1 biomass compositions,[15,23] which were optimized by the model using isotopomer information. The lower and upper limits for the fluxes to biomass synthesis in the model calculations are listed in Supporting Information Table 1. Confidence intervals were obtained by using a Monte Carlo Method[26]: the GC-MS data were changed randomly within the measurement error, and simulated annealing was performed[27] until the error function did not decrease any further.

Amino acid isotopomer noise correction

The incorporation of unlabeled amino acids from the rich medium into proteins affects the isotopomer distribution determined for each amino acid and therefore the flux calculation. A mathematical algorithm was used to correct for noise in the amino acid measurements by assuming that the fraction of completely unlabeled proteinogenic amino acids was only determined by uptake of unlabeled amino acids supplemented in the medium and unlabeled lactate. This assumption was based on the fact that lactate was labeled in the third position, and this labeled carbon was not lost before it entered TCA cycle and gluconeogenesis (note: the first carbon of lactate is lost as CO_2 in the step pyruvate→ acetylCoA). In the algorithm for correcting the isotopomer distributions, M0 of the GC-MS data (completely unlabeled proteinogenic amino acids) was thus assumed to derive only from unlabeled amino acids and unlabeled lactate (see Supporting Information for details):

$$M'_o = 0.02;\ M'_i = \frac{0.98 \cdot M_i}{1 - M_o} \forall i \neq 0 \qquad (1)$$

where M is the original GC-MS data, M' is the corrected GC-MS data; i is the number of labeled carbon atoms (0, 1, 2, 3…); and 0.02 refers to 2% nonlabeled carbon substrate (i.e., 2% of lactate is not labeled) in the medium.

Results and Discussion

Shewanella oneidensis. MR-1 had very different exponential phase growth rates in the various culture media: in minimal medium, the doubling time was ∼3 h, whereas in the salt stress medium, the doubling time was ∼6 h (Figure 1). Under salt stress, the lactate consumption and metabolite production (pyruvate and acetate) rates were significantly slower compared to those rates measured from cultures grown under normal conditions (Figure 2). The flux distributions of the normal growth and stressed cultures were calculated based on the fitting of isotopomer data (Supporting Information Figures 1–3). Despite significant change in the growth rate under salt stress (half the rate), the calculated relative intracellular fluxes in the early exponential phase were very similar between salt stressed and normal growth conditions (within the standard deviation) (Figure 3). In the exponential growth phase, MR-1 had limited flux through the TCA cycle (36–37% of lactate uptake), resulting in lactate not being fully oxidized and acetate and pyruvate accumulating in the medium due to overflow metabolism (up to ∼50% of lactate uptake).[9] Gluconeogenesis, the pentose phosphate pathway, and the ED pathway were mainly used for biomass production, and as such their fluxes were very small (<3% of lactate uptake). Two anapleurotic reactions (pyruvate to malate; oxaloacetate to phosphoenolpyruvate) and the glyoxylate shunt appeared to be active (10–20% of the lactate uptake) in both culture conditions. Although salt stress reduced the specific growth rate of cells, it did not change the relative flux distribution for most pathways. This observation does not correlate with microarray data obtained from elevated salt conditions,[5] which indicated that the expressions of most of the TCA cycle-related genes were up-regulated under salt stress. A similar lack of correlation between fluxome and transcriptome data had been recently reported for *E. coli* grown on lactate.[11]

Addition of amino acids to minimal medium significantly enhanced the growth rate and biomass production (without amino acids the doubling time was ∼3 h; with amino acids the doubling was ∼1.5 h) (Figure 1). Similarly, addition of amino acids improved MR-1's growth rate under salt stress conditions (without amino acids the doubling time was

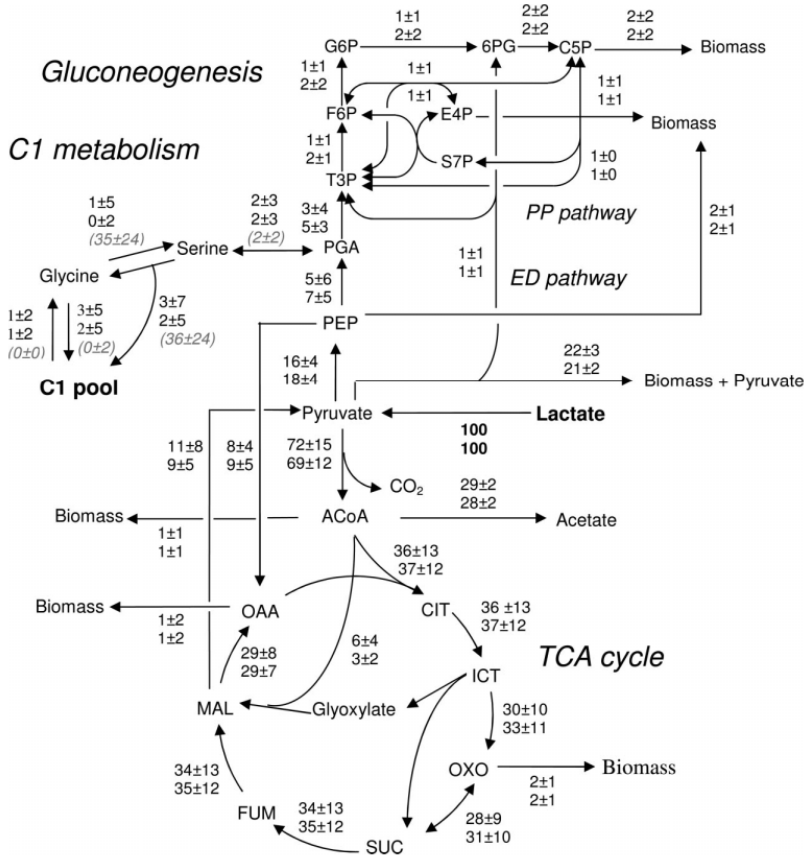


Figure 3. Fluxes in central metabolic reactions of *S. oneidensis* MR-1 under normal growth conditions (upper number) and salt stress conditions (bottom number).

The flux distributions through central metabolic pathways in ten selected mutants were identical to those under normal growth conditions. The estimated fluxes for serine and glycine metabolism in mutant SO0781 are in parentheses. Abbreviations: ACoA, acetyl-coenzyme A; CIT, citrate; E4P, erythrose-4-phosphate; C1, 5,10-Me-THF; F6P, fructose-6-phosphate; G6P, glucose-6-phosphate; 6PG, 6-phosphogluconate; ICT, isocitrate; MAL, malate; OAA, oxaloacetate; OXO, 2-oxoglutarate; PEP, phosphoenolpyruvate; PGA, 3-phosphoglycerate; C5P, ribose-5-phosphate (or ribulose-5-phosphate or xylulose-5-phosphate); S7P, sedoheptulose-7-phosphate; SUC, succinate; T3P, triose-3-phosphate.

~6 hrs; with amino acids the doubling was ~3 h) (Figure 1). The percentages of unlabeled proteogenic amino acids provided information on the relative ratios of endogenous and exogenous amino acid utilization.[28] In a defined rich medium containing an amino acid mix (17 amino acids, 25 μM each), the labeling patterns of Ala, Asp, Glu, Gly, Ser (group 1◆, Figure 4a) did not change (<2%); those of Val, Leu, Iso, Pro, Thr, and Lys (group 2□) changed moderately (<10%); and those of Met, Phe, His, Tyr (group 3○) changed significantly (>10%). The three groups of amino acids correlated with an increasing number of biochemical steps away from central metabolism and the amount of energy required for synthesis (e.g., group 1 is closest to central metabolism and requires the least energy for production). Hence, the first group was mostly endogenously synthesized, whereas the last group was mainly imported from the medium. After a suitable correction using Eq. 1, the labeling data from the amino acid supplemented medium was very close to the labeling data from a completely minimal medium (Figure 4b). This suggested that the relative flux distribution through the central pathways did not change significantly even though the addition of certain amount of amino acids doubled the growth rate, i.e., the additive effect on biomass synthesis with extra building blocks did not

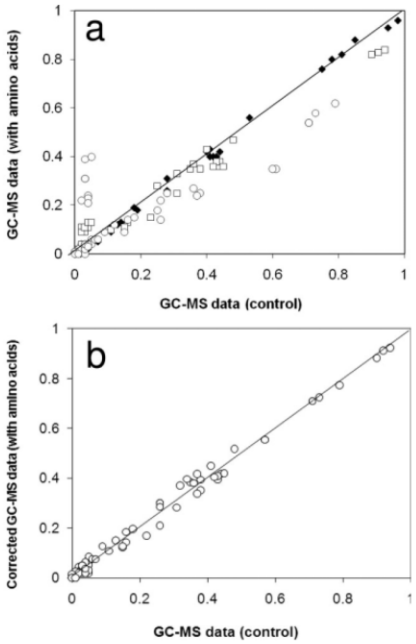


Figure 4. Effect of unlabeled amino acids on central metabolism in MR-1 (illustrated by a change in GC-MS data in key amino acids).

(a) Comparison of GC-MS data for proteinogenic amino acids from cells grown on minimal medium (control) and those grown on rich medium (addition of amino acid mixtures, 17 amino acids, 25 μM each). ◆: Ala, Asp, Glu, Gly, Ser; □: Val, Leu, Iso, Pro, Thr, Lys; ○: Met, Phe, His, Tyr. (b) GC-MS data for the amino acid-supplemented medium were corrected for incorporation of unlabeled amino acids into protein using Eq. 1.

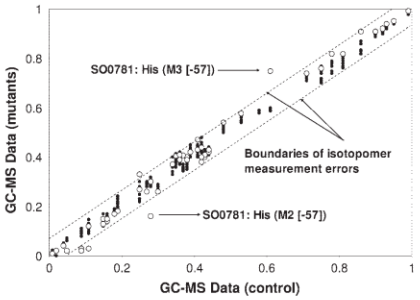


Figure 5. Effect of transposon mutagenesis on central metabolism.

The GC-MS data were from 10 different mutants (Table 1). Nine key amino acids were included (Ala, Asp, Glu, Gly, Ser, Val, Leu, Phe, His). The open dots were GC-MS data from SO0781 (lack of glycine cleavage system P protein). The outliers M2[-57] and M3[-57] are histidine labeling data. The increased ^{13}C abundance in histidine (M3[-57] was above the diagnal line) was due to the knockout of the glycine dehydrogenase (see main text). The measurement noise for isotopic data from independent tracer experiments should be below 5%.

dramatically change the reaction ratios of overall enzymatic reactions in central pathways.

The robustness of flux distributions to genetic perturbations was also investigated by comparing isotopomer distributions in amino acids between mutants and the wild-type strain. Ten mutants were randomly picked from the MR-1 mutant bank (>1000 strains) to study the possible change in central metabolic flux distribution under genetic perturbations. The growth rates of the 10 mutants (Table 1) are similar to the growth rate of the wild type in the minimal growth medium (within the measurement noise). The isotopomer data for all 10 mutants were mostly identical to those for the wild type, and very few of the mutated genes appeared to have a specific impact on the relative flux through central carbon metabolism (Figure 5). However, perturbation of certain amino acid metabolisms was observed, i.e., the SO0781 mutant (lacking glycine dehydrogenase: glycine→C1+CO_2) showed increased ^{13}C abundance in histidine. This was due to the fact that the pool of C1 used in histidine biosynthesis is derived mainly from two reactions: serine degradation and glycine degradation (Figure 3).[23] Since glycine displayed a relatively low level of ^{13}C labeling compared to serine, eliminating the glycine degradation pathway (i.e., altered C1 metabolism) increased the amount of histidine produced from serine degradation and resulted in a higher abundance of ^{13}C in histidine. The observed invariability in the relative flux distributions to genetic perturbations agrees with previous reports on the robustness of central metabolism in *Bacillus subtilis* and *E. coli*.[12,14] However, it must be mentioned that the ten randomly chosen mutants in this study had no obvious phenotypic change. Based on the "neutral mutation theory,"[16] the vast majority of single-nucleotide changes (such as genetic drift) have little or no biological effect. Through the transposon mutants, this study improves our understanding of the general effect of molecular evolution on the robustness of central metabolic flux distributions in *Shewanella*. Indeed, the mutants with the most perturbed central metabolism would have grown very poorly compared to the wild type strain in the same environment. Such deleterious MR-1 mutants, of course, would be rejected through natural selection.

Conclusion

In spite of versatile respiration metabolism, the relative flux distribution for *Shewanella oneidensis* MR-1 in the early exponential phase is robust with respect to amino acid addition, salt stress, and genetic perturbation. The rigidity of MR-1 metabolism under the studied conditions provides further evidence that microbial metabolism is not solely geared towards growth rate maximization when carbon sources are sufficient.[12] Since metabolic fluxes are the final functional output of the interaction of all molecular machinery, such rigidity of the metabolic flux distributions in microorganisms indicates that the functional output of combined regulatory activity (e.g. transcription, translation, post-translational regulation, etc.) leads to a stable relative flux profile despite the large differences in growth rates. Although it is easy to

change a cell's secondary metabolism or biosynthesis pathways (e.g., MR-1 glycine dehydrogenase mutant), the relative fluxes through central metabolic reactions are not easily altered and this may be the key mechanism for organism to survive and evolve in the environments.

Acknowledgments

This work is part of the Virtual Institute for Microbial Stress and Survival (http://VIMSS.lbl.gov) supported by the U.S. Department of Energy, Office of Science, Office of Biological and Environmental Research, Genomics:GTL Program through contract DE-AC02-05CH11231 between the Lawrence Berkeley National Laboratory and the US Department of Energy. This work is also supported by the Joint BioEnergy Institute to JDK, and I-CARES (International Center for Advanced Renewable Energy and Sustainability) at Washington University in St Louis to YJT.

Literature Cited

1. Li F, Hagemeier CH, Seedorf H, Gottschalk G, Thauer RK. Re-citrate synthase from *Clostridium kluyveri* is phylogenetically related to homocitrate synthase and isopropylmalate synthase rather than to Si-citrate synthase. *J Bacteriol*. 2007;189:4299–4304.
2. Tiedje JM. *Shewanella*- the environmentally versatile genome. *Nat Biotechnol*. 2002;20:1093–1094.
3. Venkateswaran K, Moser DP, Dollhopf ME, Lies DP, Saffarini DA, MacGregor BJ, Ringelberg DB, White DC, Nishijima M, Sano H, Burghardt J, Stackebrandt E, Nealson KH. Polyphasic taxonomy of the genus *Shewanella* and description of *Shewanella oneidensis* sp. *Int J Syst Bacteriol*. 1999;49 (Part 2):705–724.
4. Leblanc L, Gouffi K, Leroi F, Hartke A, Blanco C, Auffray Y, Pichereau V. Uptake of choline from salmon flesh and its conversion to glycine betaine in response to salt stress in *Shewanella putrefaciens*. *Int J Food Microbiol*. 2001;65:93–103.
5. Liu Y, Gao W, Wang Y, Wu L, Liu X, Yan T, Alm E, Arkin A, Thompson DK, Fields MW, Zhou J. Transcriptome analysis of *Shewanella oneidensis* MR-1 in response to elevated salt conditions. *J Bacteriol*. 2005;187:2501–2507.
6. Nichols DS, Olley J, Garda H, Brenner RR, McMeekin TA. Effect of temperature and salinity stress on growth and lipid composition of *Shewanella gelidimarina*. *Appl Environ Microbiol*. 2000;66:2422–2429.
7. Tang YJ, Meadows AL, Keasling JD. A kinetic model describing *Shewanella oneidensis* MR-1 growth, substrate consumption, and product secretion. *Biotechnol Bioeng*. 2007;189:894–901.
8. Sauer U. High-throughput phenomics: experimental methods for mapping fluxomes. *Curr Opin Biotechnol*. 2004;15:58–63.
9. Tang YJ, Pingitore F, Mukhopadhyay A, Phan R, Hazen TC, Keasling, JD. Pathway confirmation and flux analysis of central metabolic pathways in *Desulfovibrio vulgaris* Hildenborough using GC-MS and FT-ICR mass spectrometry. *J Bacteriol*. 2007;189:940–949.
10. Fong SS, Nanchen A, Palsson BO, Sauer, U. Latent pathway activation and increased pathway capacity enable *Escherichia coli* adaptation to loss of key metabolic enzymes. *J Biol Chem*. 2006;281:8024–8033.
11. Hua Q, Joyce AR, Palsson BO, Fong, SS. Metabolic characterization of *Escherichia coli* adapted to growth on lactate. *Appl Environ Microbiol*. 2007;73:4639–4647.
12. Fischer E, Sauer U. Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nat Genet*. 2005;37:636–640.
13. Blank LM, Kuepfer L, Sauer U. Large-scale 13C-flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome Biol*. 2005;6:R49.
14. Sauer U, Lasko DR, Fiaux J, Hochuli M, Glaser R, Szyperski T, Wuthrich K, Bailey JE. Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J Bacteriol*. 1999;181:6679–6688.
15. Tang YJ, Martin HG, Dehal PS, Deutschbauer A, Llora X, Meadows A, Arkin A, Keasling JD. Metabolic flux analysis of *Shewanella spp.* reveals evolutionary robustness in central carbon metabolism. *Biotechnol Bioeng*. 2009;102:1161–1169.
16. Voet D, Voet JG, Pratt CW. *Fundamentals of Biochemistry: Life at the Molecular Level*, 3 ed. Hoboken, NJ: Wiley; 2008: p 116.
17. Larsen RA, Wilson MM, Guss AM, Metcalf WW. Genetic analysis of pigment biosynthesis in *Xanthobacter autotrophicus* Py2 using a new, highly efficient transposon mutagenesis system that is functional in a wide variety of bacteria. *Arch Microbiol*. 2002;178:193–201.
18. Jacobs MA, Alwood A, Thaipisuttikul I, Spencer D, Haugen E, Ernst S, Will O, Kaul R, Raymond C, Levy R, Chun-Rong L, Guenthner D, Bovee D, Olson MV, Manoil C. Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proc Natl Acad Sci USA*. 2003;100:14339–14344.
19. Tang YJ, Laidlaw D, Gani K, Keasling JD. Evaluation of the effects of various culture conditions on Cr(VI) reduction by *Shewanella oneidensis* MR-1 in a novel high-throughput mini-bioreactor. *Biotechnol Bioeng*. 2006;95:176–184.
20. Tang YJ, Meadows AL, Keasling JD. A kinetic model describing *Shewanella oneidensis* MR-1 growth, substrate consumption, and product secretion. *Biotechnol Bioeng*. 2007;96:125–133.
21. Tang YJ, Ashcroft M, Chen D, Min G, Kim C, Murkhejee B, Larabell C, Keasling JD, Chen FF. Charge-associated effects of fullerene derivatives on microbial structural integrity and central metabolism. *Nano Lett*. 2007;7:754–760.
22. Wahl SA, Dauner M, Wiechert, W. New tools for mass isotopomer data evaluation in ^{13}C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. *Biotechnol Bioeng*. 2004;85:259–268.
23. Tang YJ, Hwang JS, Wemmer D, Keasling JD. The *Shewanella oneidensis* MR-1 fluxome under various oxygen conditions. *Appl Environ Microbiol*. 2007;73:718–729.
24. Wiechert W, Möllney M, Isermann N, Wurzel M, de Graaf AA. Bidirectional reaction steps in metabolic networks. III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol Bioeng*. 1999;66:69–85.
25. Goldberg DE. *Genetic Algorithms in Search, Optimization, and Machine Learning Reading*. Massachusetts: Addison-Wesley; 1989.
26. Zhao J, Shimizu K. Metabolic flux analysis of *Escherichia coli* K12 grown on 13C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *J Biotechnol*. 2003;101: 101–117.
27. Press WH, Teukolsky SA, Vetterling WT, Flannery BP. *Numerical Recipes in Fortran*, 2nd ed. Cambridge: Cambridge University Press; 1992: 387–448.
28. Christiansen T, Christensen B, Nielsen J. Metabolic network analysis of *Bacillus claussi* on minimal and semirich medium using ^{13}C-Labeled glucose. *Metab Eng*. 2002;4:159–169.

Manuscript received Nov. 4, 2008, and revision received Mar. 2, 2009.

Supplementary material

Amino acid isotopomer noise correction. Let us assume that, for a given amino acid of c carbons, the number of molecules with i ($=0...c$) labeled carbons is X_i, for a given *Shewanella* in the labeling experiments. The addition of unlabeled amino acids from the medium supplements to the feed would change only the unlabeled fraction by a factor k:

$$X_0 = X_0' + k$$

$$X_i = X_i' \ (i=1..c)$$

The labeled fractions of amino acids are then:

$$M_i = X_i \Big/ \sum_{i=0}^{c} X_i$$

$$M'_i = X_i' \Big/ \sum_{i=0}^{c} X'_i$$

Assuming the fraction $M_0=0.02$ to be the same as the unlabeled fraction in the feed, we get:

$$\sum_{i=1}^{c} M'_i = 1 - 0.02 = 0.98 = \frac{\sum_{i=1}^{c} X'_i}{\sum_{i=0}^{c} X'_i} = \frac{\sum_{i=1}^{c} X_i}{k + \sum_{i=0}^{c} X_i}$$

Whence the relationship between M_i and M'_i can be derived:

$$M'_i = \frac{X_i}{k + \sum_{i=0}^{c} X_i} = 0.98 \frac{X_i}{\sum_{i=1}^{c} X_i} = 0.98 \frac{X_i}{\sum_{i=0}^{c} X_i - X_0} = \frac{0.98}{1 - M_0} \frac{X_i}{\sum_{i=0}^{c} X_i} = \frac{0.98}{1 - M_0} M_i$$

where the definition of M_0 has been used:

$$M_0 = \frac{X_0}{\sum_{i=0}^{c} X_i}$$

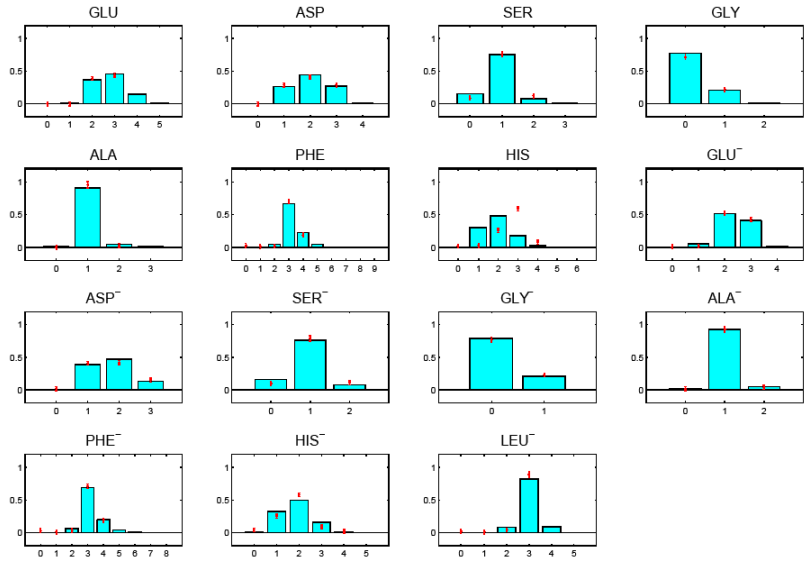
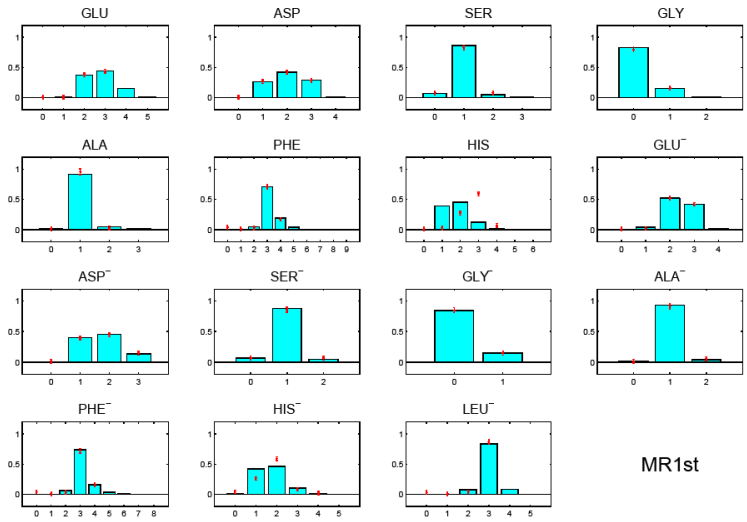


Figure S-1: GC-MS data of amino acids (green columns are measured data) and model fittings (red dots are fitted data) for MR-1 normal growth conditions. The y-axis is the labeling fractions. The x-axis is: 0(M0, unlabeled molecules); 1 (M1, singly labeled molecules); 2 (M2, doubly labeled molecules); 3 (M3, triple-labeled molecules), etc… Two type of GC-MS data are used in the model, M-57 ions and M-159 ions (marked by sup script (-) sign near the amino acid abbreviation).



MR-1

Figure S-2: GC-MS data of amino acids (green columns are measured data) and model fittings (red dots are fitted data) for MR-1 growth under salt stress conditions. The y-axis is the labeling fractions. The x-axis is: 0(M0, unlabeled molecules); 1 (M1, singly labeled molecules); 2 (M2, doubly labeled molecules); 3 (M3, triple-labeled molecules), etc… Two type of GC-MS data are used in the model, M-57 ions and M-159 ions (marked by sup script (-) sign near the amino acid abbreviation).

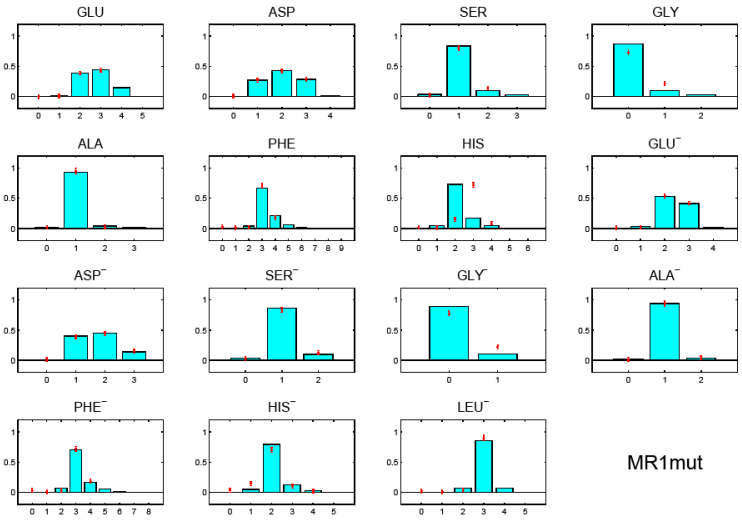


Figure S-3: GC-MS data of amino acids (green columns are measured data) and model fittings (red dots are fitted data) for MR-1 mutant (SO0781) growth under normal conditions. The y-axis is the labeling fractions. The x-axis is: 0(M0, unlabeled molecules); 1 (M1, singly labeled molecules); 2 (M2, doubly labeled molecules); 3 (M3, triple-labeled molecules), etc… Two type of GC-MS data are used in the model, M-57 ions and M-159 ions (marked by sup script (-) sign near the amino acid abbreviation).

Supplementary Table S-1: reaction lists, carbon transition routes and lower / upper limits allowed for the fluxes. The limits for extracellular acetate and pyruvate were obtained from enzymatic measurements of the concentration of these metabolites and carbon source (lactate).

% Lactate to pyruvate	**flux range = [0 100]**
1: ABC -> ABC	
% Pyruvate to Acetyl-CoA	**flux range = [0 200]**
2: ABC -> BC + A	
% Oxaloacetate to citrate	**flux range = [0 100]**
3: AB + abcd -> dcbBAa	
% Citrate to isocitrate	**flux range = [0 100]**
4: ABCDEF -> ABCDEF	
% Isocitrate to 2-oxoglutarate	**flux range = [0 100]**
5: ABCDEF-> ABCDE + F	
% 2-oxoglutarate to succinyl-CoA	**flux range = [0 100]**
6: ABCDE -> BCDE + A :: ABCDE -> EDCB+ A	
% Succinate to Malate	**flux range = [0 100]**
7: ABCD -> ABCD	
% Fumarate -> malate	**flux range = [0 100]**
8: ABCD-> ABCD	
% Malate to oxaloacetate	**flux range = [0 100]**
9: ABCD-> ABCD	
% Isocitrate to succinate	**flux range = [0 100]**
10: ABCDEF-> FCDE+ AB:: ABCDEF-> EDCF+ AB	
% Glyoxylate to malate	**flux range = [0 100]**
11: AB + CD -> ABCD	
% Serine production, serine→serine(biomass)	**flux range = [1 10]**
12: ABC -> ABC	
% Glycine formation from serine	**flux range = [0 100]**
13: ABC ->AB + C	
% Glycine to C-1 group	**flux range = [0 100]**
14: AB -> A + B	
% Oxaloacetate to phosphoenolpyruvate	**flux range = [0 100]**
15: ABCD -> ABC + D	
% Pyruvate to malate	**flux range = [0 100]**
16: ABC + D-> ABCD	
% Reaction glycine and C1 to serine	**flux range = [0 100]**
17: AB + C -> ABC	

```
% CO2 + C1 to glycine                              flux range = [0 100]
18: A + B -> AB

% Phosphoenolpyruvate to Oxaloacetate              flux range = [0 100]
19: ABC + D -> ABCD

% Malate to pyruvate                               flux range = [0 100]
20: ABCD -> ABC + D

% Pyruvate to phosphoenolpyruvate                  flux range = [0 100]
21: ABC -> ABC

% Phosphoenolpyrutate to 3-phospho-D-glycerate     flux range = [0 20]
22: ABC -> ABC

% 3-phosphoglycerate to triose-3-phosphate         flux range = [0 100]
23: ABC -> ABC

% triose-3-phosphate to dihydroxyacetone phosphate flux range = [0 100]
24: ABC -> ABC

% triose-3-phosphate to fructose-6-phosphate       flux range = [0 100]
25: ABC + abc -> CBAabc

% fructose-6-phosphate to glucose-6-phosphate      flux range = [0 100]
26: ABCDEF -> ABCDEF

% glucose-6-phosphate to 6-phosphogluconate        flux range = [0 100]
27: ABCDEF -> ABCDEF

% ED pathway:
%pyruvate + triose-3-phosphate→6-phosphogluconate flux range = [0 100]
28: ABC + DEF -> ABCDEF

% 6-phosphogluconate to Ribulose 5-phosphate       flux range = [0 100]
29: ABCDEF -> A + BCDEF

% Transketolase: triose-3-phosphate + fructose 6-P -> xylulose 5-P +
epythrose 4-P                                     flux range = [0 100]
30: ABC + abcdef -> abABC + cdef

% fructose-6-P + epythrose 4-P -> triose-3-phosphate + D-sedoheptulose
7-P                                               flux range = [0 100]
31: ABC + abcdefg ->  defg + abcABC

% Tranketolase: D-sedoheptulose 7-P + triose-3-phosphate -> D-Xylulose
5-P + D-Ribose 5-P                                flux range = [0 100]
32: ABCDE + abcde -> CDE + ABabcde

% Ribulose 5-P Epimerase:
%D-Ribulose 5-P->D-Ribose 5-P                     flux range = [0 100]
33: ABCDE -> ABCDE

% Ribulose-P 3-Epimerase:
%D-Xylulose 5-P -> D-Ribulose 5-P                 flux range = [0 100]
34: ABCDE -> ABCDE
```

```
% Leu synthesis (step 1): acetylCoA + pyruvate → acetolactate (carbon
2-4)+ CO2                                          flux range = [1 10]
35: AB + CDE -> ABDE + C

% Glu synthesis: 2-oxoglutarate to glutamate       flux range = [1 10]
36: ABCDE -> ABCDE

% Alanine synthesis and extracelluar pyruvate:
pyruvate consumption                               flux range = [20 30]
37: ABC -> ABC

% Asp synthesis:
%oxaloacetate to aspartate and other biomass       flux range = [1 10]
38: ABCD -> ABCD

% Leu synthesis(step 2):
%acetolactate(carbon 2-4)+acetylCoA                flux range = [1 10]
39: ABDE + GH -> ABDGHE

% Histidine biosynthesis:
Ribose-5-P and C1 to histidine                     flux range = [0 10]
40: ABCDE+ a -> ABCDEa

% Phenyalanine precursor synthesis: E4P+PEP→DAHP   flux range = [0 10]
41: ABCD + abc-> abcABCD

% Phenylalanine synthesis
%PEP + DAHP → CO2+ Phenylalanine                   flux range = [0 10]
42: def + abcABCD -> a + bcABCDfed

% Acetate efflux: acetylCoA→acetate                flux range = [20 40]
43: AB -> AB
```

For MR1 mutant reaction glycine → C1 + CO2 was eliminated and the reaction was substituted by:

```
% Glycine biomass: glycine → glycine               flux range = [0 10]
14: AB -> AB
```

Appendix 9

Construction of a Parsimonious Kinetic Model to Capture Microbial Dynamics via Parameter Estimation

Running title: Model reconstruction via parameter estimation

Xueyang Feng[1*], Yinjie J. Tang[1], Kirk D. Dolan[2,3*]

1. Department of Energy, Environmental and Chemical Engineering, Washington University, MO 63130

2. Department of Food Science & Human Nutrition, Michigan State University, East Lansing, MI 48824.

3. Department of Biosystems & Agricultural Engineering, Michigan State University, East Lansing, MI 48824

*Correspondence to:

Xueyang Feng, Tel: 314-935-6125, Fax: 314-935-7211; E-mail: fengx@seas.wustl.edu

Kirk D. Dolan, Tel: 517-355-8474, Fax: 517-353-8963; E-mail: dolank@msu.edu

Abstract

Understanding microbial kinetic behavior is important for bioprocessing engineering, such as chemical synthesis and bioremediation. However, development of proper models to capture complicated microbial kinetics is a challenging task. In this study, we demonstrate an example for rationally modeling the growth of *Shewanella oneidensis* MR-1 in a batch culture. Based on a series of analyses, including residual analysis, scaled sensitivity coefficient analysis, parameter correlation analysis, and the *F*-test, we estimated model parameters to construct a parsimonious Monod-based model that used the fewest parameters for best simulation of *Shewanella* growth using different carbon substrates. Sequential analysis was also applied to identify the time window for estimating each parameter in the kinetic model. This study shows that statistics-based parameter estimation is an efficient method to successively reconstruct and fine tune kinetic models for complex biological systems.

Key words: Monod, *Shewanella*, parsimonious model, sensitivity, *F*-test, sequential analysis

Introduction

The relation between microbial growth rates and substrate concentrations (often referred to as microbial growth kinetics) has been studied and substantiated for over a century [1]. Numerous models, including the well-known Monod equation [2], have been established to describe the growth kinetics. In spite of the wide applications of the Monod model, it does not always lead to a good description of a given bioprocess. One criticism of the Monod model is that the two parameters in the model, i.e., the maximum specific growth rate (μ_{max}) and substrate saturation constant (K_s), can be correlated in estimation, which results in reported values that differ by several orders of magnitude [1]. In addition, the complicated culture conditions (e.g., growth with mixed substrates rather than a single substrate) make the original Monod model insufficient to reflect every aspect of growth kinetics. To overcome this problem, the model has to be refined by introducing additional parameters to account for the complexity of growth kinetics. In doing so, however, there is a risk of over-parameterization [3].

In this study, we applied parameter estimation techniques [4,5] in reconstructing the Monod model to decipher growth kinetics of *Shewanella oneidensis* MR-1. MR-1 is an environmentally important bacterium that plays important roles in bioremediation, microbial fuel cells, and ecological carbon cycling [6,7,8,9]. As revealed by our previous kinetic study, *S. oneidensis* MR-1 utilizes lactate as the favorable carbon source and produces waste metabolites

(acetate and pyruvate) which are consumed during the late growth stage [6]. However, the previous model has not rigorously considered the optimal selection of model parameters to improve the quality of data fitting. To build a parsimonious model for capturing MR-1 growth dynamics (i.e., a model with the best simulation and fewest parameters), we applied a series of tools that were originally used in nonlinear parameter estimation to reconstruct the Monod model successively until it satisfied all the statistical assumptions used in parameter estimation. During the model reconstruction, several underlying metabolic features of MR-1 were uncovered. Finally, the impact of each experimental measurement on estimating the updated kinetic parameters was also evaluated by sequential parameter estimation [10,11,12,13] of the parsimonious model.

Methods

Cell culture and analytical methods

S. oneidensis MR-1 was first grown in LB medium in shake flasks overnight. A 0.1% inoculum volume was then cultured into the modified MR-1 defined medium [14] in shaking flasks at 150 rpm and 30°C. The initial carbon source was 30 mM lactate. The growth curve was monitored by dried biomass weight. The concentrations of lactate and acetate in the medium were measured using enzyme kits (r-Biopharm, Darmstadt, Germany). The concentration of

pyruvate in the medium was measured with the enzyme assay developed by Marbach and Weli, 1967. For each sampling point, we made triplicate measurements.

Seed model for growth kinetics

To find a parsimonious model describing MR-1 growth with the fewest parameters, a seed model was built first and then successively reconstructed (Table 1). In the seed model, a multiple-substrate Monod model was developed to simulate the cell growth, lactate consumption, and acetate and pyruvate secretion and reuse.

$$\frac{\mathrm{d}LACT}{\mathrm{dt}} = (-\frac{X \cdot \mu_L}{Y_{X/L}} - r_{P,L} - r_{A,L}) \quad (1)$$

$$\frac{\mathrm{d}ACT}{\mathrm{dt}} = (r_{A,L} + r_{A,P} - \frac{X \cdot \mu_A}{Y_{X/A}}) \quad (2)$$

$$\frac{\mathrm{d}PYR}{\mathrm{dt}} = (r_{P,L} - \frac{X \cdot \mu_{\mathrm{P}}}{Y_{X/P}} - r_{A,P}) \quad (3)$$

$$\frac{\mathrm{d}X}{\mathrm{dt}} = X \cdot (\mu - k_e) \quad (4)$$

The growth rates of MR-1 on different carbon substrates were assumed to be additive:

$$\mu = \mu_L + \mu_A + \mu_P \quad (5)$$

$$\mu_L = \frac{\mu_{max,L} \cdot LACT}{K_{s,l} + LACT} \quad (6)$$

$$\mu_A = \frac{\mu_{max,A} \cdot ACT}{K_{s,a} + ACT} \tag{7}$$

$$\mu_P = \frac{\mu_{max,P} \cdot PYR}{K_{s,p} + PYR} \tag{8}$$

In the above model, X is biomass (g dry cell/L); *LACT*, *ACT*, and *PYR* are lactate, acetate, and pyruvate concentrations (mmol/L), respectively; μ is the overall specific growth rate (h^{-1}); k_e is the endogenous metabolism rate constant (h^{-1}); $Y_{X/L}$, $Y_{X/A}$, and $Y_{X/P}$ are the biomass yield coefficients (g dry cell /mmol substrate) of lactate, acetate, and pyruvate, respectively; $r_{P,L}$ and $r_{A,L}$ are the production rate (mmol/L/h) of pyruvate and acetate from lactate, respectively; $r_{A,P}$ is the production rate (mmol/L/h) of acetate from pyruvate. The biomass yield coefficients were the reported values: $Y_{X/L}$ = 0.0212 (g dry cell /mmol lactate), $Y_{X/A}$ = 0.0132 (g dry cell /mmol acetate) and $Y_{X/P}$ = 0.0195 (g dry cell /mmol pyruvate) [15]. μ_L, μ_A, and μ_P are the specific growth rates (h^{-1}) on lactate, acetate, and pyruvate, respectively; $\mu_{max,L}$, $\mu_{max,A}$, and $\mu_{max,P}$ are the maximal growth rates (h^{-1}) for fully aerobic growth on lactate, acetate, and pyruvate, respectively; and $K_{s,l}$, $K_{s,a}$, and $K_{s,p}$ are Monod constants (mmol/L) for lactate, acetate, and pyruvate, respectively.

The acetate and pyruvate secretions were assumed to follow first order kinetic forms in the seed model:

$$r_{A,L} = k_{al} \cdot LACT \cdot X \quad (9)$$

$$r_{P,L} = k_{pl} \cdot LACT \cdot X \quad (10)$$

$$r_{A,P} = k_{ap} \cdot PYR \cdot X \quad (11)$$

where k_{al} and k_{pl} are rate constants of acetate and pyruvate production from lactate, respectively (L· (h·g DCW)$^{-1}$). k_{ap} is rate constant of acetate production from pyruvate (L· (h·g DCW)$^{-1}$). The seed model was then reconstructed following the flowchart in Figure 1. In general, parameter estimation was accomplished for each reconstructed model, and then it guided the addition or removal of parameters. Ten different models (Table 1) were built and compared until the parsimonious model was capable of satisfying all the statistical assumptions.

Nonlinear parameter estimation

To estimate the parameters in the seed model and the reconstructed models, an ordinary least squares (OLS) method [4] was applied to solve the inverse problems. In general, OLS aims to minimize the residual sum of the squares (R) between model simulations and experimental measurements:

$$R = [\mathrm{Y} - \eta\ (t\ ;\ \boldsymbol{\beta})]^{\mathrm{T}}\ [\mathrm{Y} - \eta\ (t\ ;\ \boldsymbol{\beta})\ ] \quad (12)$$

where t is the time of the experimental measurements, η represents four dependent variables (i.e., concentration of biomass, lactate, acetate and pyruvate) simulated by the kinetic model; β represent the vector of the parameters to be estimated; and

Y is the vector of the experimentally measured value of the dependent variables. Our experiment had a total of n measurements for each of the four dependent variables, so η and Y were vectors with a length of $4\times n$. The scales of the four dependent variables had order of magnitude differences (e.g., the scale of the biomass measurement was < 1 g/L, while the lactate measurement was > 10 mmol/L), and the direct application of OLS would overemphasize the fitting of dependent variables with large scales. To overcome the estimation bias, each simulated dependent variable η_j ($n\times 1$ vector) and the corresponding experimental measurements Y_j ($n\times 1$ vector) were normalized by dividing by the maximum concentrations observed in the n experimental measurements (i.e., $Y_j' = Y_j / max(Y_j)$, $\eta_j' = \eta_j / max(Y_j)$, $j = 1\ldots4$). The weighted OLS method was

$$R = [Y' - \eta'(t;\boldsymbol{\beta})]^T [Y' - \eta'(t;\boldsymbol{\beta})] \qquad (13)$$

To check the simulation quality, we applied a series of analyses, including residual analysis, scaled sensitivity coefficient analysis, parameter correlation analysis, and the F-test. The mean of the residuals in the model simulation (i.e., $\varepsilon = Y' - \eta'(t;\beta)$ in each model) should be close to zero (i.e., $E(\varepsilon)$ should be close to zero, where E is the expectation function), should be normally distributed (i.e., the histogram of ε should follow Gaussian distribution), and should be uncorrelated (i.e. $E\{[\varepsilon_i - E(\varepsilon_i)][\varepsilon_j - E(\varepsilon_j)]\} = 0$). A simple check for uncorrelated residuals was to count the number of runs (i.e., number of changes in signs of the residuals) [4]. In this study, n measurements (triplicates) were taken

at m time points (i.e., $n = 3 \cdot m$). The means of the triplicates at each time point were used to check the residual correlations. For m independent, zero-mean, random model fitting residuals, the minimum value for non-correlated residuals should be [4]

$$r = (m+1)/2 \tag{14}$$

If the number of runs in the model simulation was smaller than r, the model fitting residuals were correlated with each other, and vice versa.

The impact of parameters on the model simulation was identified by formulating and analyzing a scaled sensitivity (SS) matrix. The scaled sensitivity matrix is the normalization for the commonly-used Jacobian matrix, which can be applied to compare the impact of the parameters on model simulation:

$$SS = \begin{bmatrix} \beta_1 \cdot \dfrac{\partial \eta_1'}{\partial \beta_1} & \cdots & \beta_p \cdot \dfrac{\partial \eta_1'}{\partial \beta_p} \\ \vdots & \ddots & \vdots \\ \beta_1 \cdot \dfrac{\partial \eta_n'}{\partial \beta_1} & \cdots & \beta_p \cdot \dfrac{\partial \eta_n'}{\partial \beta_p} \end{bmatrix} \tag{15}$$

where η_i' is the simulation in ith time point, β_j is the jth parameter, n is the number of time points in experimental measurements, and p is the number of parameters. The units of the scaled sensitivity coefficients (SSCs) are dimensionless since all the dependant variables are normalized in η'. Each column of the SS matrix reflects the dynamic impact of different parameters on the model

simulation. The SSCs of the two most influential parameters are plotted in Figure S1, for both the seed model and the parsimonious model. From a comparison of the scale of each column (i.e., the maximum of the absolute values in the column), the impact generated by each parameter could be compared and used to guide the model reconstruction. For each dependent variable, the SSCs versus time could be plotted for each parameter. For better parameter estimation that would result in the smallest standard error for that parameter, we desired SSCs to be large and uncorrelated.

The correlations among parameters were characterized by a correlation matrix *Corr*, which yielded the correlation coefficient of the two parameters β_i and β_j, by using the covariance matrix *C*:

$$\mathrm{Corr(i,\,j)} = \frac{\mathrm{C(i,i)}}{\sqrt{\mathrm{C(i,\,j)}\cdot\mathrm{C(i,\,j)}}} \quad \in [-1.0\ 1.0] \tag{16}$$

A higher absolute value of *Corr* (i, j) in the correlation matrix indicated a strong correlation between the two parameters and hence more difficult estimation for both parameters.

The "ode23" command in MATLAB (2009a) was used to solve differential equations, and the "fmincon" command was used to estimate the parameters. Since many local solutions existed around the real parameter settings, the likely global solution of parameters was derived by randomly choosing the

initial guesses within a known range in the nonlinear optimization and running for 100 times. Next, a bootstrap method was used to find the confidence intervals of each parameter by following the approach described before [4,16]. In general, the residuals in the parsimonious model were randomly re-sampled to obtain a new bootstrap set Y', which constructed a new parameter estimation problem that was solved by the same procedure. The 95% confidence intervals of the parameters were derived by running the bootstrap method 999 times [4], which added to the original estimate gave 1,000 values for each parameter. For each parameter, the 1,000 values were sorted from lowest to highest, and the confidence interval was found by choosing the 25^{th} and 975^{th} value. The standard deviation of each parameter was calculated from the 1,000 values.

The F-test for the reconstructed models

Introducing more parameters in the reconstructed model usually led to a better (or at least equally good) simulation of the microbial growth kinetics. However, it was unclear whether the improvement in simulation was statistically significant. In this study, the *F*-test was applied to address this question. For the original model η_i and the reconstructed model η_j, the critical *F* value was calculated as

$$F = \frac{\left(\frac{R_i - R_j}{p_j - p_i} \right)}{\left(\frac{R_j}{n - p_j} \right)} \qquad (17)$$

where R_i and R_i are the residual sum of the squares of models η_i and η_j, respectively; n is the number of experimental measurements; and p_i and p_j are the number of parameters in models η_i and η_j, respectively. If the F value calculated from the reconstructed model was larger than the critical F value for some desired false-rejection probability (e.g., 0.05 in this study), the improvement of the reconstructed model was significant. Otherwise, the new model could not provide a significantly better fit than the original model. The F-test can check not only whether introducing new parameters leads to a significant improvement in model simulation, but also can be used in a reversible way to test whether some parameters can be removed from the original model. If the reconstructed model with fewer parameters has a smaller F value than the critical F value, then by removing redundant parameters, the simulation in the reconstructed model is not degraded (i.e., the original model is over-parameterized).

Sequential analysis of the parsimonious model

To estimate the effect of experimental measurements on model parameters, we applied a sequential analysis [10,11,12,13] by successively adding experimental data. Each experimental data point was treated as one step in the

parameter estimation. There were four dependent variables and 18 time-course samples (triplicates), so the total steps were 4×18×3 = 216. At each step, the parameters in the parsimonious model were estimated by the sequential procedure which was developed based on a matrix inversion lemma [4]. A series of equations was used iteratively to find the parameter at each step:

$$
\begin{cases}
\text{for j=1:m} \\
\begin{cases}
\mathbf{e} = Y' - \eta(t, \mathbf{b}) \\
\mathbf{J} = \begin{bmatrix} \beta_1 \cdot \dfrac{\partial \eta_1'}{\partial \beta_1} & \cdots & \beta_p \cdot \dfrac{\partial \eta_1'}{\partial \beta_p} \\ \vdots & \ddots & \vdots \\ \beta_1 \cdot \dfrac{\partial \eta_n'}{\partial \beta_1} & \cdots & \beta_p \cdot \dfrac{\partial \eta_n'}{\partial \beta_p} \end{bmatrix} \\
\text{for i=1:n-1} \\
\begin{cases}
\mathbf{A}(i+1) = \mathbf{P}(i)\mathbf{J}^T(i+1) \\
\Delta(i+1) = \Phi(i+1)^2 + \mathbf{J}(i+1)\mathbf{A}(i+1) \\
\mathbf{K}(i+1) = \mathbf{A}(i+1) \setminus \Delta(i+1) \\
\mathbf{P}(i+1) = \mathbf{P}(i) - \mathbf{K}(i+1)\mathbf{X}(i+1)\mathbf{P}(i) \\
\mathbf{B}(i+1) = \mathbf{B}(i) + \mathbf{K}(i+1)\{\mathbf{e}(i+1) - \mathbf{X}(i+1)[\mathbf{B}(i) - \mathbf{b}]\}
\end{cases} \\
\mathbf{b\_new} = \mathbf{B}(n); \\
\text{if } (\mathbf{b\_new} - \mathbf{b}) < \text{tol break; else } \mathbf{b} = \mathbf{b\_new};
\end{cases}
\end{cases}
\tag{18}
$$

where **e** is the vector for normalized residuals; **b** is the vector for parameters estimated by sequential analysis; **J** is the scaled sensitivity matrix, **P** was the parameter covariance matrix; **Φ** is the vector for variance of measurement; **A**, Δ, **K**, **B,** and **b_new** are matrices used in the sequential analysis; n is the number of

steps in sequential analysis (i.e., 216 in this study), m is the number of runs of the sequential analysis; tol is the convergence tolerance in the sequential analysis.

In general, the parameter vector **b** and the parameter covariance matrix **P** were initialized. The residual vector **e** and the scaled sensitivity matrix **J** were calculated based on **b**, and they initiated one round of sequential analysis, as shown in the inner for-loop. The sequential analysis then updated **b**, which was used as the new initial parameter vector in the next round of sequential analysis. This procedure was repeated until the difference between **b** and **B** converged below the tolerance, 0.3% in this study.

Results and discussion

Simulation of growth kinetics by the seed model

As a starting point to investigate the growth kinetics of *S. oneidensis* MR-1, a seed model (Model 0, eq.1-11) assumed that the growth rate on different carbon substrates was additive and that first-order kinetics in the carbon substrate switch from lactate to acetate and pyruvate. The seed model showed gaps between the model simulation and experimental measurements. For example, the concentrations of lactate were always overestimated, while the biomass productions were underestimated by the model prediction at the early growth stage (Figure 2A). The differences between the simulation and experimental

measurements made the residuals of the seed model highly correlated (Figure 2B) and of large variance as indicated in the width of histogram (Figure 2C).

In the seed model, the largest SSCs for the seed model are those related to the growth rates of lactate ($\mu_{max,L}$) and of acetate ($\mu_{max,A}$). These two SSCs v.s. time are plotted for each dependent variable (i.e., biomass, lactate, acetate and pyruvate) in Figure S1A, C, E & G, respectively. For best estimation results, we desire SSCs to be large and uncorrelated. The maximum absolute value of each SSC is shown by its corresponding length in the bar charts in Figure S2 A-F, where each chart is for a different model. For example, the maximum absolute value of SSC for $\mu_{max,L}$ in acetate simulation was 5.57 (Figure S1E). The corresponding bar runs from 5.78 to 11.35 (Figure S2A, green bar). In this way, Figure S2 A-F could readily show which parameter is the most or least influential for each model. Figure S2 usefully summarizes the large number of SSCs and avoid clutter in Figure S1.

In addition, the scaled sensitivity analysis (Figure S2A) of the seed model indicated that the impacts of $\mu_{max,P}$ and K_{sp} (the two parameters related to MR-1 growth with pyruvate) on simulation of growth kinetics were 7~8 magnitudes lower than the other parameters. The correlation analysis among parameters (Text S1) showed that the two parameters $\mu_{max,L}$ and K_{sl} (the two parameters related to *Shewanella* growth with lactate) were correlated with each other (correlation

coefficient was 0.885). Based on the fitting analysis of the seed model, a series of model reconstructions were made following the procedure in Figure 1.

Monod model reconstructions

The imperfection of the growth kinetics simulation in the seed model initiated the successive model reconstructions to find the parsimonious model (Figure 1). In general, based on the statistical analysis of the seed model, the model was reconstructed by either adding or removing parameters in the growth kinetics. The reconstructed model with the hypothesized kinetics was analyzed by a series of statistical methods, including scaled sensitivity analysis (Figure S2), residual analysis (Figure S3), parameter correlation analysis (Text S1), and residual correlation analysis (Text S2). The reconstructed model was accepted as the new seed model if it passed the *F*-test. The procedure was repeated until a parsimonious model was found to satisfy all the statistical assumptions.

In this study, ten reconstructed models were built and improved (labeled as Model 0-Model 3.1 in Table 2). The entire reconstruction procedure was composed of two steps. The first step in model reconstruction (Model 0-Model 1.3) was to remove the parameters in Model 0 that were not necessary in describing the growth rates with multiple carbon substrates. Since the scaled sensitivity coefficients of the parameters $\mu_{max,P}$ and $K_{s,p}$ related to *Shewanella* growth with pyruvate were 7~8 magnitudes lower than for the other parameters (Figure S2A), the pyruvate term had a trivial contribution to MR-1growth.

Therefore, we built Model 1.1 (Table 1) by removing $\mu_{max,P}$ and $K_{s,p}$ from the seed model. By using the *F*-test in reverse, it was found that the reconstructed model did not lead to a significantly worse simulation (i.e. $F < F\_critical$ in Table 2), and thus Model 1.1 was accepted for further reconstruction. To build Model 1.2, the Monod constant K_{sa} was then removed because it had the lowest maximum value of scaled sensitivity (Figure S2B). This assumption was again proved to be reasonable by the *F*-test. Next, it was found that $\mu_{max,L}$ and K_{sl} could be mildly correlated (the correlation coefficient is 0.776) in Model 1.2, which indicated that the value of the Monod constant K_{sl} was much more influential than lactate concentrations in the experimental measurements, and hence first-order kinetics would be sufficient to describe the growth kinetics with lactate. Model 1.3 was accordingly built by assuming that the biomass production from lactate followed first order kinetics. Although such an assumption affected the kinetic model simulation (Figure S3), the smaller *F* (F = 2.308) than the critical *F* (F = 4.043) value in Table 2 suggested that the reconstructed model avoided over-parameterization.

Reconstruction from Model 0 to Model 1.3 aimed at a more concise description of the growth kinetics. However, owing to the improper assumption of first-order kinetics in the substrate switch, the simulated growth kinetics needed to be improved. Therefore, the second step in model reconstruction (Model 2.1-Model 2.3.3) was to increase the complexity of products secretion kinetics in the

reconstructed model. From Model 2.1 to 2.3.3 (Table 1), five different kinetics were constructed to describe the acetate and pyruvate production. We used the rational function to empirically simulate the product secretion kinetics. The denominator function was set as a first-order polynomial, while the numerator function was gradually made more complicated from Model 2.1 to Model 2.3.3 (Table 1) by increasing the degrees of the polynomial function. The *F*-test was applied to each reconstructed model to test whether the rational function was sufficient to describe the growth kinetics. As identified in the *F*-test (Table 2), first-order kinetics was sufficient for acetate production from lactate or pyruvate. However, the kinetics of pyruvate production from lactate was more complicated than first-order or Monod kinetics. Therefore, we described pyruvate production from lactate by adding a second order polynomial term in the numerator of the Monod equation (i.e., Model 2.2, Table 1). The reconstructed model, Model 2.2, dramatically improved the fitting of biomass, acetate, and pyruvate concentrations by decreasing both the scale and variances of residuals (Figure S3 D1 and D2, vs. E1 and E2). However, correlated residuals were still found in the model simulation. So we added a hypothetical kinetic term to describe additional lactate loss during the cultivation or measurement processes (Model 3.1). Based on the *F*-test, the introduction of the hypothetical lactate loss coefficient did not significantly reduce the residual sum of squares for the entire model. However, it did improve the overall fitting by eliminating the correlated residuals (Text S2) in the biomass, lactate, and acetate simulations. Accordingly, the lactate loss term

was retained in the parsimonious model to account for the consumption of lactate not considered by the original Monod kinetics (Model 3.1, Table 3).

Compared to Model 0, the residuals in the parsimonious model were smaller (a root mean square error of 0.15, compared to 0.26 in Model 0, Figure 2D). The residuals were also non-correlated in biomass, lactate, and acetate simulations (Figure 2E), and normally distributed with a smaller variance (0.004, compared to 0.012 in Model 0, Figure 2F). All parameters were confirmed to be uncorrelated and to influence the growth kinetics simulation. Therefore, the statistical assumptions were satisfied in the parsimonious model. The largest SSCs for the parsimonious model were those related to the growth rate of lactate (β_2) and of acetate (β_1), as shown in Figure S1 B, D, F, H and Figure 3. The parameter values in the parsimonious model and 95% confidence intervals derived from the bootstrap method are listed in Table 4.

Sequential analysis of the parsimonious model

In designing experiments to reveal microbial growth kinetics, it is necessary to know the best time window for estimating each parameter. By gradually including more experimental measurements in parameter estimation, we applied a sequential analysis to investigate the impact of experimental measurements on parameter estimation in the parsimonious model. The sequentially estimated parameters initially fluctuated as expected with the addition of more experimental measurements (Figure 4), but eventually converged

to results similar to those from nonlinear parameter estimation (Text S3). However, the measurement steps in sequential analysis at which the parameters no longer varied were not the same. For example, parameter k_e could be accurately estimated within 100 measurement steps (i.e., ~18 h), while k_{ap} could not be reasonably estimated until the 150th measurement step (i.e., ~25 h). Some parameters (i.e., β_{11}) could not smoothly converge, indicating that current experimental measurements may not be effective in estimating such parameters, or that a longer experiment may be necessary for estimating the parameter. All other parameters converged to a constant well before the experiment ended, strengthening the conclusion that the model is appropriately reconstructed to simulate the growth kinetics of MR-1.

Concluding remarks

The Monod model was successively reconstructed in this study until a parsimonious model emerged that described the growth kinetics of *Shewanella oneidensis* MR-1. From the model reconstructions, we found that pyruvate made a trivial contribution to biomass production; rather, it served as a transient metabolite in producing acetate. The growth rate with lactate and acetate could be sufficiently described by first-order and zero-order kinetics, respectively. The kinetics of carbon substrate utilization, especially for pyruvate production by lactate, was more complicated than first-order reactions, and it suggest that unidentified lactate loss may be involved in the growth of MR-1. Finally, by using

sequential analysis in the parsimonious model, the time windows necessary for parameter estimation were provided, which can be further used for guiding the experimental designs in microbial growth kinetics studies. The general modeling approach in this research can also be potentially used to study other complex biological systems.

Acknowledgements

The authors would like to thank Dr. James Beck for consultation on parameter estimation. This study was partially supported by an NSF Career Grant (MCB0954016) to YJT. We also thank Jerry Brusher from Mathworks for providing us with MATLAB software through a Mathworks education grant. We sincerely appreciate James Ballard in the Engineering Communication Center of Washington University for editing the manuscript.

References

Beck, J. V., Arnold, K. J., 1977. Parameter estimation. John Wiley & Sons.

Djurić, P. M., et al., 2002. Sequential Parameter Estimation of Time-Varying Non-Gaussian Autoregressive Processes. EURASIP Journal on Applied Signal Processing. 2002**,** 865-875.

Faber, R., et al., 2003. Sequential Parameter Estimation for Large-Scale Systems with Multiple Data Sets. 1. Computational Framework. Ind. Eng. Chem. Res. 42**,** 5850-5860.

Faber, R., et al., 2004. Sequential Parameter Estimation for Large-Scale Systems with Multiple Data Sets. 2. Application to an Industrial Coke-Oven-Gas Purification Process. Ind. Eng. Chem. Res. 43**,** 4350-4362.

Ju, P., et al., 1996. Sequential parameter estimation of a simplified induction motor load model. IEEE Transactions on Power Systems. 11.

Kovárová-Kovar, K., Egli, T., 1998. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. Microbiol Mol Biol Rev. 62**,** 646-66.

Lillacci, G., Khammash, M., 2010. Parameter estimation and model selection in computational biology. PLoS Comput Biol. . 6(3):e1000696.

Logan, B. E., et al., 2005. Electricity generation from cysteine in a microbial fuel cell. Water Research. 39**,** 942-952.

Mailier J, et al., 2011. Parametric sensitivity analysis and reduction of a detailed nutritional model of plant cell cultures. Biotechnol Bioeng. . 108**,** 1108-18.

Monod, J., 1949. The growth of bacterial cultures. Annu. Rev. Microbiol. 3**,** 371-394.

Myers, C. R., Nealson, K. H., 1988. Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. Science. 240**,** 1319-1321.

Pinchuk, G. E., et al., 2010. Constraint-based model of Shewanella oneidensis MR-1 metabolism: a tool for data analysis and hypothesis generation. PLoS Comput Biol. 6**,** e1000822.

Ritz, C., Streibig, J. C. Eds.), 2008. Nonlinear Regression with R, p 96-97. Springer.

Tang, Y. J., et al., 2006. Evaluation of the effects of various culture conditions on Cr(VI) reduction by *Shewanella oneidensis* MR-1 in a novel high-throughput mini-bioreactor. Biotechnology and Bioengineering. 95**,** 176-184.

Tang, Y. J., et al., 2007. A kinetic model describing *Shewanella oneidensis* MR-1 growth, substrate consumption, and product secretion. Biotechnol Bioeng. . 189**,** 894-901.

Tiedje, J. M., 2002. *Shewanella*- the environmentally versatile genome. Nature Biotechnology. 20**,** 1093-1094.

Figure Legends

Figure 1. Flowchart of Monod model reconstruction.

Figure 2. Growth kinetics of *Shewanella oneidensis* MR-1. Left (A, B, and C), simulation by seed model; right (D, E, and F), simulation by the reconstructed parsimonious model. In the fitting plots (A, D), the experimental measurements are plotted with markers (○: lactate concentration, mmol/L; ◊: measured acetate concentration, mmol/L; ×: pyruvate concentration, mmol/L; □: biomass, g/L) and the model predictions were plotted in solid lines. In the residual plots (B, E), the normalized residuals were plotted with different markers (○: normalized lactate residuals; ◊: normalized acetate residuals; ×: normalized pyruvate residuals; □: normalized biomass residuals). The residual histograms (C, F) are plotted for all the normalized residuals.

Figure 3. The maxima of scaled sensitivity for parameters in the parsimonious model. Black bars: the maximum scaled sensitivity for biomass simulation; red bars: maximum scaled sensitivity for lactate simulation; green bars: the maximum scaled sensitivity for acetate simulation yellow bars: the maximum scaled sensitivity for pyruvate simulation.

Figure 4. Sequential analysis for the reconstructed Monod model.

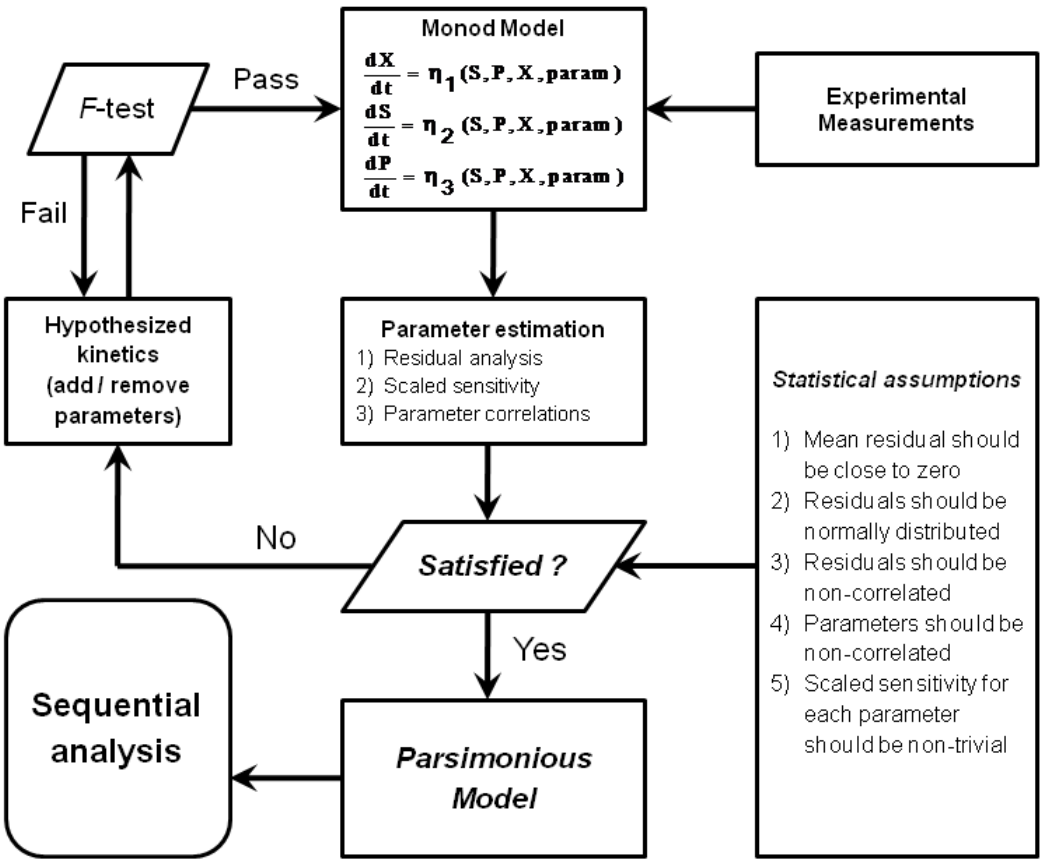

Monod Model
$\frac{dX}{dt} = \eta_1(S,P,X,param)$
$\frac{dS}{dt} = \eta_2(S,P,X,param)$
$\frac{dP}{dt} = \eta_3(S,P,X,param)$
F-test
Pass
Fail
Experimental Measurements
Hypothesized kinetics (add / remove parameters)
Parameter estimation
1) Residual analysis
2) Scaled sensitivity
3) Parameter correlations
Statistical assumptions
1) Mean residual should be close to zero
2) Residuals should be normally distributed
3) Residuals should be non-correlated
4) Parameters should be non-correlated
5) Scaled sensitivity for each parameter should be non-trivial
No
Satisfied ?
Yes
Sequential analysis
Parsimonious Model

Fig.1

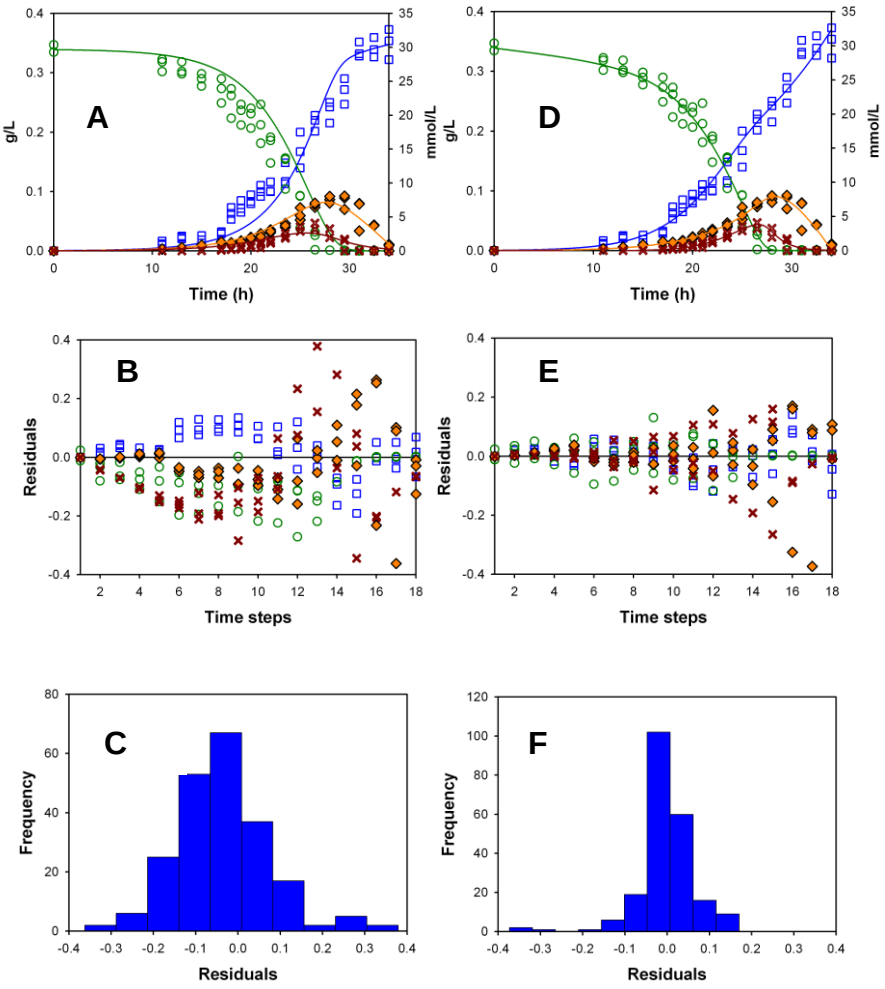


Fig.2

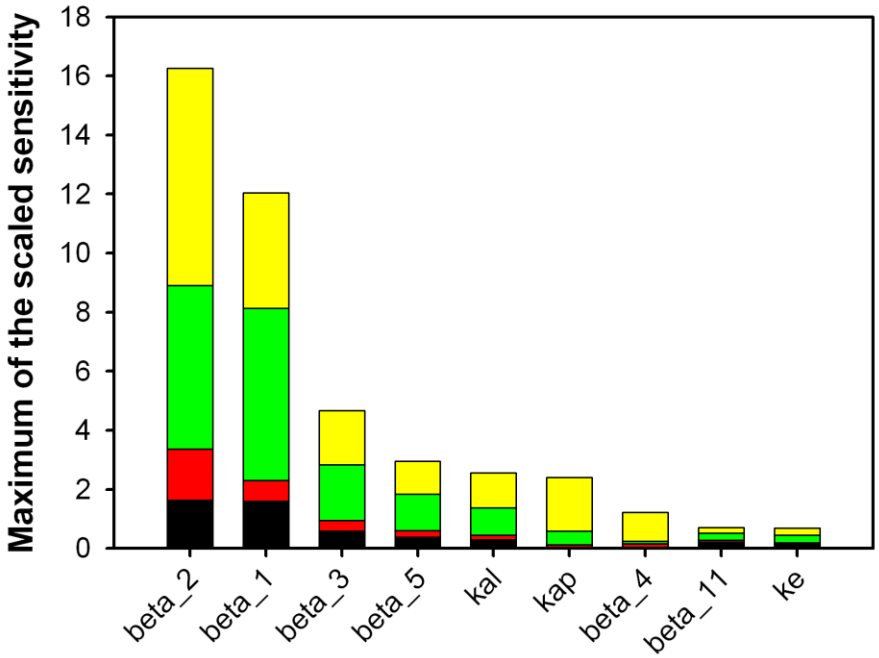

Maximum of the scaled sensitivity
18
16
14
12
10
8
6
4
2
0
beta_2
beta_1
beta_3
beta_5
kal
kap
beta_4
beta_11
ke

Fig.3

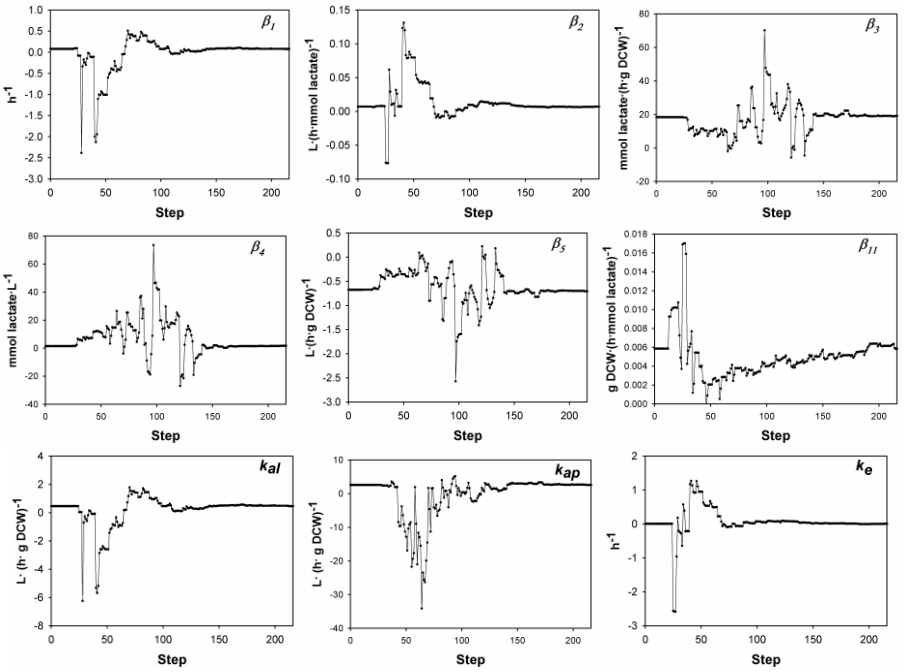

β_1
h^{-1}
β_2
L·(h·mmol lactate)$^{-1}$
β_3
mmol lactate·(h·g DCW)$^{-1}$
β_4
mmol lactate·L^{-1}
β_5
L·(h·g DCW)$^{-1}$
β_{11}
g DCW·(h·mmol lactate)$^{-1}$
k_{al}
L· (h· g DCW)$^{-1}$
k_{ap}
L· (h· g DCW)$^{-1}$
k_e
h^{-1}
Step

Fig.4

Table 1. Models used in the *F*-test

Model	Assumptions
0	Seed model as developed in eq. (1) - (11)
1.1	Extracellular pyruvate was neglected for biomass production: $\mu = \mu_L + \mu_A$ in eq. (5)
1.2	Biomass production from acetate was assumed to follow zero order kinetics: $\mu_A = \beta_1$ in eq. (7)
1.3	Biomass production from lactate was assumed to follow first order kinetics: $\mu_L = \beta_2 \cdot LACT$ in eq. (6)
2.1	Pyruvate production from lactate was assumed to follow Monod kinetics: $r_{P,L} = \frac{\beta_3 \cdot LACT}{\beta_4 + LACT}$ in eq. (10)
2.2	Pyruvate production from lactate was assumed to follow the kinetics: $r_{P,L} = \frac{\beta_5 \cdot LACT^2 + \beta_3 \cdot LACT}{\beta_4 + LACT}$ in eq. (10)
2.3.1	Pyruvate production from lactate was assumed to follow the kinetics: $r_{P,L} = \frac{\beta_6 \cdot LACT^3 + \beta_5 \cdot LACT^2 + \beta_3 \cdot LACT}{\beta_4 + LACT}$ in eq. (10)
2.3.2	Acetate production from lactate was assumed to follow Monod kinetics: $r_{A,L} = \frac{\beta_7 \cdot LACT}{\beta_8 + LACT}$ in eq. (9)
2.3.3	Acetate production from pyruvate was assumed to follow Monod kinetics: $r_{A,P} = \frac{\beta_9 \cdot PYR}{\beta_{10} + PYR}$ in eq. (11)
3.1	Lactate was assumed to have non-growth related degradation: $\frac{dLACT}{dt} = (-\frac{X \cdot \mu_L}{Y_{X/L}} - r_{P,L} - r_{A,L} - \beta_{11} \cdot LACT)$ in eq. (1)

Table 2. *F*-test for different models

Model No.	No. of Parameters, *p*	Degree of freedom, *n-p*	Sum of squares, *R*	Mean square, $s^2=R/(n-p)$	Models compared	ΔR	$F=\frac{\Delta R}{s^2}$	*F*_ critical	Best Model No.
0	10	44	2.9174	0.0663					
1.1	8	46	2.9174	0.0634	Model 0 v.s. 1.1	0.0000	0.000	F(0.95,2,46)=3.200	1.1
1.2	7	47	2.9717	0.0632	Model 1.1 v.s. 1.2	0.0543	0.215	F(0.95,1,47)=4.047	1.2
1.3	6	48	3.6794	0.0767	Model 1.2 v.s. 1.3	0.7077	2.308	F(0.95,1,48)=4.043	1.3
2.1	7	47	2.4463	0.0520	Model 1.3 v.s. 2.1	1.2331	5.923	F(0.95,1,47)=4.047	2.1
2.2	8	46	1.1603	0.0252	Model 2.1 v.s. 2.2	1.2860	12.746	F(0.95,1,46)=4.052	2.2
2.3.1	9	45	1.1603	0.0258	Model 2.2 v.s. 2.3.1	0.0000	0.000	F(0.95,1,45)=4.057	2.2
2.3.2	9	45	1.1603	0.0258	Model 2.2 v.s. 2.3.2	0.0000	0.000	F(0.95,1,45)=4.057	2.2
2.3.3	9	45	1.1603	0.0258	Model 2.2 v.s. 2.3.3	0.0000	0.000	F(0.95,1,45)=4.057	2.2
3.1	9	45	0.9650	0.0214	Model 2.2 v.s. 3.1	0.1953	2.277	F(0.95,1,45)=4.057	3.1 *

* Based on the *F*-test, Model 3.1 cannot significantly improve the sum of squares, R. However, Model 3.1 can eliminate the correlation among residuals and hence improve the overall fitting.

Table 3. A parsimonious model for *S. oneidensis* MR-1 growth

No.	Reaction	Equations
1	Bacteria growth	$\frac{dX}{dt} = (\mu_L + \mu_A - k_e) \cdot X$
2	Lactate consumption	$\frac{dLACT}{dt} = (-\frac{X \cdot \mu_L}{Y_{X/L}} - r_{P,L} - r_{A,L} - \beta_{11} \cdot LACT)$
3	Acetate production and reuse	$\frac{dACT}{dt} = (r_{A,L} + r_{A,P} - \frac{X \cdot \mu_A}{Y_{X/A}})$
4	Pyruvate production and reuse	$\frac{dPYR}{dt} = (r_{P,L} - r_{A,P})$
6	Specific growth rate with lactate	$\mu_L = \beta_2 \cdot LACT$
7	Specific growth rate with acetate	$\mu_A = \beta_1$
8	Acetate production from lactate	$r_{A,L} = k_{al} \cdot LACT \cdot X$
9	Acetate production from pyruvate	$r_{A,P} = k_{ap} \cdot PYR \cdot X$
10	Pyruvate production from lactate	$r_{P,L} = \frac{\beta_5 \cdot LACT^2 + \beta_3 \cdot LACT}{\beta_4 + LACT}$

Table 4. Parameters of Monod model for *S. oneidensis* MR-1 growth

Symbols	Notation	Unit	Value	SD (%)	95% Confidence Intervals
β_1	Zero order growth coefficient from acetate	h^{-1}	0.0773[6][16][16](Tang et al., 2007)	0.33	[0.0768 0.0778]
β_2	First order growth coefficient from lactate	$L\cdot(h\cdot mmol\ lactate)^{-1}$	0.0070	0.37	[0.0070 0.0071]
β_3	Coefficient for pyruvate production from lactate	$mmol\ lactate\cdot(h\cdot g\ DCW)^{-1}$	18.8010	0.35	[18.6442 18.9666]
β_4	Coefficient for pyruvate production from lactate	$mmol\ lactate\cdot L^{-1}$	1.6223	7.36	[1.4174 1.9163]
β_5	Coefficient for pyruvate production from lactate	$L\cdot(h\cdot g\ DCW)^{-1}$	-0.6916	2.64	[-0.7315 -0.6630]
β_{11}	Non-growth related lactate	$g\ DCW\cdot(h\cdot mmol\ lactate)^{-1}$	0.0057	0.36	[0.0057 0.0058]

	degradation coefficient				
k_{al}	Acetate production coefficient from lactate	$L \cdot (h \cdot g\ DCW)^{-1}$	0.4672	4.86	[0.4378 0.5334]
k_{ap}	Acetate production coefficient from pyruvate	$L \cdot (h \cdot g\ DCW)^{-1}$	2.5487	3.36	[2.3986 2.7851]
k_e	Endogenous metabolism rate	h^{-1}	0.0063	5.46	[0.0061 0.0074]

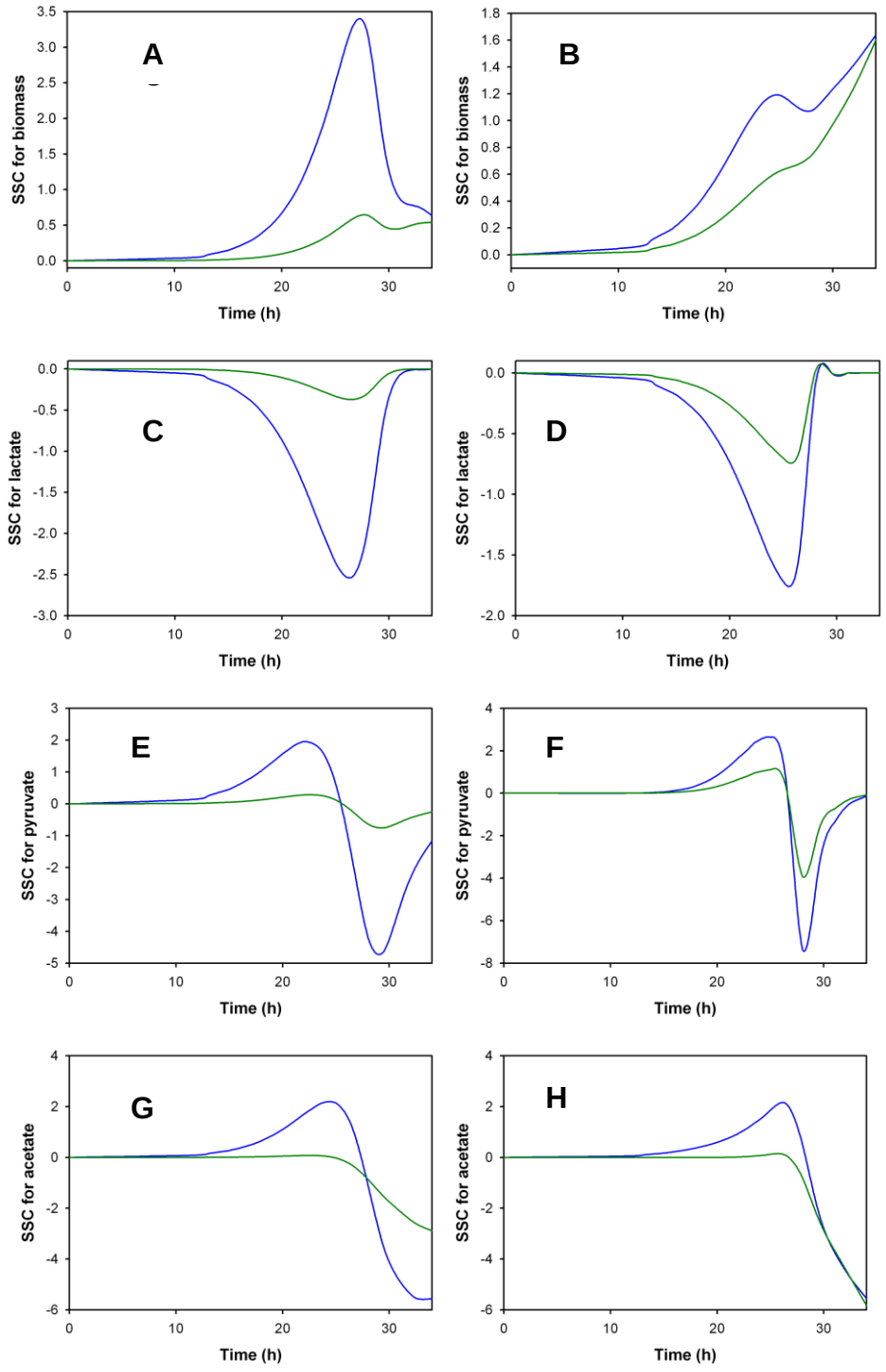

A
SSC for biomass
Time (h)
B
SSC for biomass
Time (h)
C
SSC for lactate
Time (h)
D
SSC for lactate
Time (h)
E
SSC for pyruvate
Time (h)
F
SSC for pyruvate
Time (h)
G
SSC for acetate
Time (h)
H
SSC for acetate
Time (h)

Figure S1. Scaled sensitivity coefficients (SSCs) in the seed model and the parsimonious model. Left column (A, C, E and G): the scaled sensitivity coefficients of $\mu_{max,L}$ (blue line) and $\mu_{max,A}$ (green line) in the seed model. These two parameters were chosen as examples to plot because they had the largest values. Right column (B, D, F and H): the SSCs of β_1 (blue line) and β_2 (green line) in the parsimonious model. These two parameters were chosen because they had the largest value. The maximums of the scaled sensitivity coefficients plotted in Figure 3 were the maximums of the absolute values in the corresponding scaled sensitivity plots above. We desire SSCs to be large and uncorrelated for best parameter estimation.

Note: SSCs for triplicate measurements were the same.

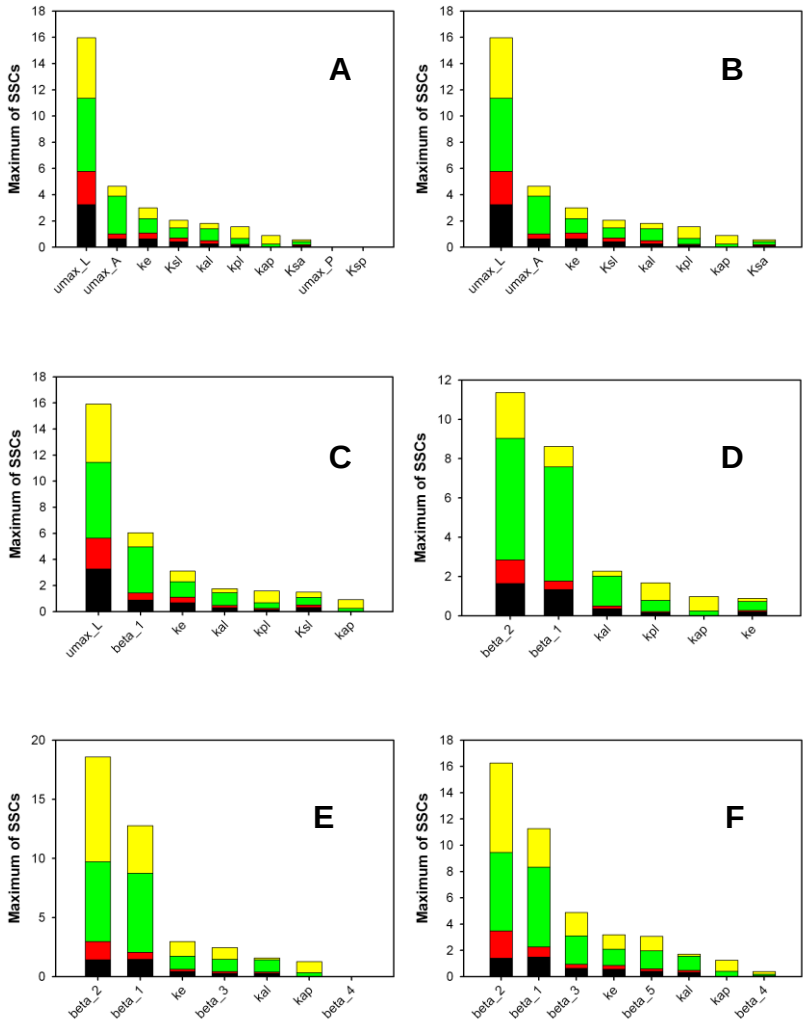
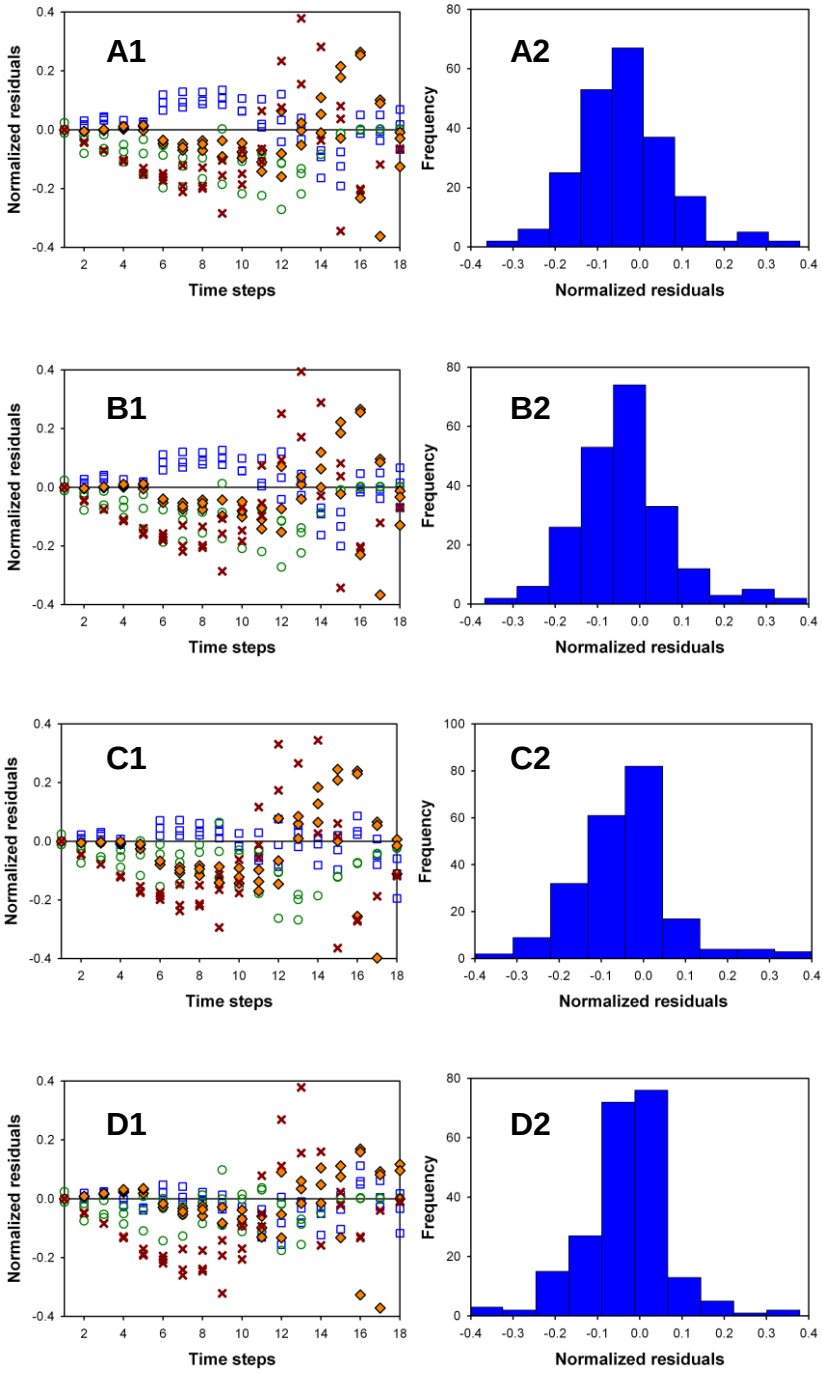


Figure S2 The maximal scaled sensitivity coefficients (SSCs) of parameters in reconstructed models. **A**: Model 0; **B**: Model 1.1; **C**: Model 1.2; **D**: Model 1.3; **E**: Model 2.1; **F**: Model 2.2. Black bars: scaled sensitivity on biomass simulation; red bars: scaled sensitivity on lactate simulation; green bars: scaled sensitivity on acetate simulation; yellow bars: scaled sensitivity on pyruvate simulation. The length of each colored bar came from the absolute maximum value of the corresponding scaled sensitivity coefficient plots such as those in Figure S1.

Note: the maximal SSCs of certain parameters (e.g. $\mu_{max,P}$ and $K_{s,p}$ in Figure S2A) were too small to see.


A1
A2
B1
B2
C1
C2
D1
D2
Normalized residuals
Time steps
Frequency
Normalized residuals

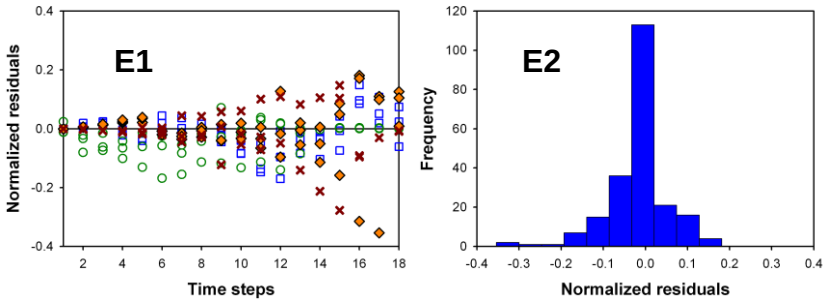


Figure S3 Residual analysis for reconstructed models. **A1 and A2**: Model 1.1; **B1 and B2**: Model 1.2; **C1 and C2**: Model 1.3; **D1 and D2**: Model 2.1; **E1 and E2**: Model 2.2. Green dots: normalized lactate residuals; orange diamonds: normalized acetate residuals; red crosses: normalized pyruvate residuals; blue squares: normalized biomass residuals.

Text S1 Correlation matrix in reconstructed models.

Correlation matrix in Model 0

	umax_L	umax_A	umax_P	Ksl	Ksa	Ksp	kpl	kal	kap	ke
umax_L	1.000	0.700	0.027	0.885	0.931	0.056	0.653	-0.187	0.373	-0.300
umax_A	0.700	1.000	-0.058	0.433	0.854	0.040	0.359	-0.456	0.888	-0.500
umax_P	0.027	-0.058	1.000	0.046	0.046	0.993	0.357	0.331	-0.169	0.019
Ksl	0.885	0.433	0.046	1.000	0.683	0.036	0.603	0.010	0.064	-0.303
Ksa	0.931	0.854	0.046	0.683	1.000	0.105	0.627	-0.278	0.593	-0.387
Ksp	0.056	0.040	0.993	0.036	0.105	1.000	0.341	0.246	-0.052	-0.043
kpl	0.653	0.359	0.357	0.603	0.627	0.341	1.000	0.341	0.001	0.070
kal	-0.187	-0.456	0.331	0.010	-0.278	0.246	0.341	1.000	-0.713	0.497
kap	0.373	0.888	-0.169	0.064	0.593	-0.052	0.001	-0.713	1.000	-0.519
ke	-0.300	-0.500	0.019	-0.303	-0.387	-0.043	0.070	0.497	-0.519	1.000

Correlation matrix in Model 1.1

	umax_L	umax_A	Ksl	Ksa	kpl	kal	kap	ke
umax_L	1.000	0.887	0.937	0.962	0.746	-0.046	0.680	-0.211
umax_A	0.887	1.000	0.858	0.899	0.859	0.313	0.797	-0.166
Ksl	0.937	0.858	1.000	0.834	0.625	-0.097	0.696	-0.405
Ksa	0.962	0.899	0.834	1.000	0.825	0.084	0.666	-0.179
kpl	0.746	0.859	0.625	0.825	1.000	0.246	0.829	0.009
kal	-0.046	0.313	-0.097	0.084	0.246	1.000	0.031	0.247
kap	0.680	0.797	0.696	0.666	0.829	0.031	1.000	-0.118
ke	-0.211	-0.166	-0.405	-0.179	0.009	0.247	-0.118	1.000

Correlation matrix in Model 1.2

	umax_L	beta_1	Ksl	kpl	kal	kap	ke
umax_L	1.000	-0.149	0.776	-0.406	-0.593	0.031	-0.048
beta_1	-0.149	1.000	0.294	0.477	0.551	0.582	0.014
Ksl	0.776	0.294	1.000	-0.259	-0.412	0.300	-0.500
kpl	-0.406	0.477	-0.259	1.000	0.279	0.655	0.287
kal	-0.593	0.551	-0.412	0.279	1.000	-0.082	0.292

kap	0.031	0.582	0.300	0.655	-0.082	1.000	-0.005
ke	-0.048	0.014	-0.500	0.287	0.292	-0.005	1.000

Correlation matrix in Model 1.3

	beta_2	**beta_1**	**kpl**	**kal**	**kap**	**ke**
beta_2	1.000	-0.762	-0.558	-0.620	-0.514	0.535
beta_1	-0.762	1.000	0.687	0.863	0.601	0.120
kpl	-0.558	0.687	1.000	0.365	0.865	0.050
kal	-0.620	0.863	0.365	1.000	0.216	0.104
kap	-0.514	0.601	0.865	0.216	1.000	0.049
ke	0.535	0.120	0.050	0.104	0.049	1.000

Correlation matrix in Model 2.1

	beta_2	**beta_1**	**kal**	**kap**	**ke**	**beta_3**	**beta_4**
beta_2	1.000	-0.704	-0.478	-0.431	0.411	-0.517	0.187
beta_1	-0.704	1.000	0.643	0.624	0.342	0.689	-0.141
kal	-0.478	0.643	1.000	-0.058	0.110	0.003	-0.214
kap	-0.431	0.624	-0.058	1.000	0.292	0.849	0.112
ke	0.411	0.342	0.110	0.292	1.000	0.231	0.026
beta_3	-0.517	0.689	0.003	0.849	0.231	1.000	0.217
beta_4	0.187	-0.141	-0.214	0.112	0.026	0.217	1.000

Correlation matrix in Model 2.2

	beta_2	**beta_1**	**kal**	**kap**	**ke**	**beta_3**	**beta_4**	**beta_5**
beta_2	1.000	-0.690	-0.549	-0.243	0.466	-0.600	-0.271	0.461
beta_1	-0.690	1.000	0.695	0.514	0.301	0.392	-0.098	-0.273
kal	-0.549	0.695	1.000	-0.029	0.069	0.311	0.104	-0.361
kap	-0.243	0.514	-0.029	1.000	0.351	0.011	-0.361	0.130
ke	0.466	0.301	0.069	0.351	1.000	-0.246	-0.430	0.213
beta_3	-0.600	0.392	0.311	0.011	-0.246	1.000	0.672	-0.936
beta_4	-0.271	-0.098	0.104	-0.361	-0.430	0.672	1.000	-0.588
beta_5	0.461	-0.273	-0.361	0.130	0.213	-0.936	-0.588	1.000

Correlation matrix in Model 3.1

	beta_2	**beta_1**	**kal**	**kap**	**ke**	**beta_3**	**beta_4**	**beta_5**	**beta_11**
beta_2	1.000	-0.486	-0.586	-0.234	0.621	-0.417	0.245	0.275	-0.432
beta_1	-0.486	1.000	0.552	0.398	0.360	0.293	-0.515	0.003	-0.271
kal	-0.586	0.552	1.000	-0.042	-0.209	0.096	-0.338	-0.210	0.295
kap	-0.234	0.398	-0.042	1.000	0.075	0.308	-0.763	-0.152	0.129
ke	0.621	0.360	-0.209	0.075	1.000	-0.174	-0.101	0.331	-0.806
beta_3	-0.417	0.293	0.096	0.308	-0.174	1.000	-0.457	-0.852	0.187
beta_4	0.245	-0.515	-0.338	-0.763	-0.101	-0.457	1.000	0.437	-0.204
beta_5	0.275	0.003	-0.210	-0.152	0.331	-0.852	0.437	1.000	-0.432
beta_11	-0.432	-0.271	0.295	0.129	-0.806	0.187	-0.204	-0.432	1.000

Text S2 Residual correlations in reconstructed models

Model No.	Residuals of biomass simulation	Residuals of lactate simulation	Residuals of acetate simulation	Residuals of pyruvate simulation
0	Correlated	Correlated	Correlated	Correlated
1.1	Correlated	Correlated	Correlated	Correlated
1.2	Correlated	Correlated	Correlated	Correlated
1.3	Non-correlated	Correlated	Correlated	Correlated
2.1	Correlated	Correlated	Correlated	Correlated
2.2	Correlated	Correlated	Correlated	Correlated
3.1	Non-correlated	Non-correlated	Non-correlated	Correlated

Note: the criteria for checking residuals correlations were following the counting-number-of-runs method as described in the main content. The residuals were treated as correlated if the number of runs was smaller than value given by Eq 14 in the main content.

Text S3 Parameters estimated by sequential analysis in the Parsimonious model

Symbols	Unit	Value from OLS	Value from Sequential analysis
β_1	h^{-1}	0.0773[6][16][16](Tang et al., 2007)	0.0768[6][16][16](Tang et al., 2007)
β_2	$L\cdot(h\cdot mmol\ lactate)^{-1}$	0.0070	0.0071
β_3	$mmol\ lactate\cdot (h\cdot g\ DCW)^{-1}$	18.8010	19.1150
β_4	$mmol\ lactate\cdot L^{-1}$	1.6223	1.7057
β_5	$L\cdot (h\cdot g\ DCW)^{-1}$	-0.6916	-0.7078
β_{11}	$g\ DCW\cdot(h\cdot mmol\ lactate)^{-1}$	0.0057	0.0058
k_{al}	$L\cdot (h\cdot g\ DCW)^{-1}$	0.4672	0.4631
k_{ap}	$L\cdot (h\cdot g\ DCW)^{-1}$	2.5487	2.5951
k_e	h^{-1}	0.0063	0.0064

Appendix 10

Experimental Analysis and Kinetic Modeling of Isobutanol Fermentation by a Eecombinant *Escherichia coli* Strain

Yi Xiao†, Xueyang Feng†, Huifeng Yu, Arul M. Varman, Yinjie J. Tang

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

Running title: kinetic model for isobutanol fermentation

Corresponding author: Dr. Yinjie J. Tang, Department of Energy, Environmental and Chemical Engineering; Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130, USA

E-mail.: yinjie.tang@seas.wustl.edu;

Tel.: 314-935-3441; Fax: 314-935-7211

†:Yi Xiao and Xueyang Feng contributed equally to this work

Abstract

This study constructed an *E. coli* strain for isobutanol fermentation via the Ehrlich pathway, which included two genes from *Lactococcus lactis*: a *kivd* gene encoding 2-ketoisovalerate decarboxylase and an *adhA* gene encoding aldehyde reductase. We performed fermentation under aerobic and oxygen-limited conditions, and then built an empirical Monod model with 19 parameters to simulate glucose consumption, biomass growth, and product secretion and loss. Based on fermentation kinetics and ^{13}C-isotopic experiments to analyze nutrient utilization for isobutanol and biomass synthesis, we had following findings. 1) Oxygen-limited conditions showed the highest isobutanol titer (0.95 g/L). 2) Isobutanol production took place during both growth and stationary phases (described by mixed-growth-associated product formation kinetics). 3) Aerobic condition had better isobutanol yield from glucose than oxygen-limited conditions. 4) Nutrient supplement (yeast extract) provided building blocks for both biomass and isobutanol synthesis, but it also enhanced the waste product secretion (such as acetate and lactate) and thus reduced glucose-based isobutanol yield. This study offered both metabolic insights and bioprocess guidelines for designing scaled-up isobutanol fermentation using engineered *E. coli* strains.

Key words: ^{13}C-isotopic experiment, Ehrlich pathway, Mixed-growth-associated, scaled-up bioprocess, yeast extract

Introduction

Bioethanol produced by yeast fermentation is a common biofuel used as a gasoline additive. Other biofuels, such as butanol and biodiesels, have higher energy density and lower water solubility than ethanol, and thus receive extensive study. Among these biofuels, butanol is compatible with automotive internal combustion engines. Acetone-butanol-ethanol (ABE) fermentation is a bioprocess that uses *Clostridium acetobutylicum* to produce *n*-butanol [17], but such a process is restrained by a relatively low alcohol production rate during the anaerobic fermentation. To overcome this disadvantage, the *n*-butanol pathway derived from *Clostridium* has been reconstructed in fast-growing *E. coli* or yeast strains [18,19,20,21]. Butanol biosynthesis via *Clostridium* pathway has limitations including low product titer, accumulation of toxic metabolites, and the requirement of large nutrient supplements (such as yeast extract). Another novel approach is via the non-fermentative pathway to produce low-toxicity isobutanol [22], where the amino acid biosynthesis pathways and the Ehrlich pathway [23,24] are utilized for alcohols synthesis. The non-fermentative pathway was previously found in yeast and lactic acid bacteria, which form fuel alcohols from intermediates in amino acids pathways [23]. A non-fermentative pathway, with ketoisovalerate as the central intermediate, has shown promising industrial potential for effective isobutanol production.

Diverse butanol production strategies have been reported (Table 1), including overexpression of targeted pathway, elimination of competing pathways, systems redesign of host metabolism, and *in situ* removal of inhibitory products using gas stripping. However, very few studies have developed process models to study the kinetics of engineered microbial platform for butanol fermentation. To apply a newly developed host in the biofuel industry, a kinetic-based model is of practical importance not only for designing optimal scaled-up fermentation, but also for understanding the internal metabolic features of microbial hosts in responses to various nutrient sources and cultivation conditions. To fulfill the gap between microbial strain reconstruction and industrial bioprocess engineering, we have first created an *E. coli* mutant that is able to produce isobutanol via the standard non-fermentative pathway. The model strain provided isobutanol fermentation data from a well-controlled bioreactor under different cultivation conditions. Based on the experimental data, we developed an empirical model to simulate the key bioprocess variables and investigate the influential factors affecting isobutanol fermentation. To provide metabolic insights into the kinetic model, ^{13}C-experiments were also performed to investigate the nutrient utilization for biomass and isobutanol synthesis.

Materials and Methods

Pathway construction

The non-fermentative isobutanol pathway was introduced into *E. coli* BL21 (DE3) (Fig. 1), in which glucose was converted to 2-ketoisovalerate via glycolysis and valine biosynthesis pathway, and then transformed to isobutanol by the Ehrlich pathway. Via heterologous expression of Kivd (2-ketoisovalerate decarboxylase) and AdhA (aldehyde reductase), the *E. coli* strain can produce isobutanol as well as other higher alcohols. In this study, *kivd* and *adhA* were amplified from *Lactococcus lactis* by PCR with high fidelity DNA polymerase Pfx (Invitrogen).

Primers for *kivd*:

5'-GACACTCGAGTAATGTATACAGTAGGAGATTAC-3';

5'- TGCGGGTACCTTATGATTTATTTTGTTC-3'.

Primers for *adhA*:

5'-

TCAACTAGTGGTACCAGGAGATATAATATGAAAGCAGCAGTAGTAAG

AC-3';

5'- ATTTGCGGCCGCGCATGCTTATTTAGTAAAATCAATGAC-3'.

The genes *kivd* (treated with XhoI / KpnI) and *adhA* (treated with KpnI / SphI) were cloned into pTAC-MAT-Tag-2 Expression Vector (Sigma-Aldrich) using XhoI / SphI to create the plasmid pTAC-KA, and then transformed into *E.*

coli BL21 (DE3). To confirm the expression of Kivd and AdhA, we performed SDS-PAGE analysis of the recombinant strain and observed the protein bands of Kivd (60.9 kDa) and AdhA (30.8 kDa). Moreover, the measurement also showed that the strain synthesized isobutanol, acetate, lactate, ethanol, and a small amount of *n*-propanol and methyl-butanol. Therefore, the host had similar product profiles to other isobutanol producing *E.coli* strains [22], and thus could be used as a model to study the general kinetic behaviors of engineered *E.coli* for isobutanol fermentation.

Fermentation conditions

Fermentations were performed in a 1 L New Bruswick Bioflo 110 fermentor with dissolved oxygen (DO) electrode, temperature electrode, and pH meter. The 100% DO value was defined as the point where the cell-free medium was saturated and achieved by purging air (~2 L/min) for 15 minutes. The culture reached 0% DO in the oxygen limited condition (air rate 0 L/min) after the cell had consumed the residual oxygen in the culture. Two culture media were used in this study: A minimal medium contained 2% glucose, M9 salts (Difco), 10 mg/L vitamin B1, and 50 mg/L ampicillin; a rich medium contained minimal medium with 5 g/L yeast extract. The culture (400 mL) was inoculated with 5 ml overnight LB culture (OD_{600}~2) of the recombinant *E. coli* strain. The pH value, measured by the pH meter, was maintained around 7.0 by addition of 2 mol/L NaOH via an auto-pump. The temperature was held at 30 °C and the stirring speed was kept at

200 rpm. Cultures in the bioreactor were first grown for 5~7 hours under aerobic conditions (OD_{600} 0.3~0.5; DO>50%), before adding 0.2 mM IPTG (Isopropyl β-D-1-thiogalactopyranoside) to induce the isobutanol pathway.

After the *E. coli* entered the isobutanol production phase, we imposed three different fermentation conditions. The first fermentation (F1: aerobic condition and minimal medium) was carried out in aerobic condition. Air (airflow rate: ~1 L/min) was bubbled in the bioreactor to provide oxygen and to remove isobutanol (i.e., gas stripping) from bioreactor. In the second fermentation (F2: oxygen-limited condition and minimal medium), air was turned off and the DO was maintained around zero throughout the fermentation in a closed bioreactor. In the third fermentation (F3: oxygen-limited condition and rich medium containing 5 g/L yeast extract), air was turned off and the DO was maintained around zero for the duration of the fermentation in a closed bioreactor.

Analytical methods for biomass and metabolites

The biomass growth was monitored based on optical density (OD_{600}). There was a linear relationship between the dry weight of *E. coli* cells and OD_{600}. To measure the dry biomass weight, biomass samples were harvested by centrifugation and washed with DI water and dried at 100°C until their weight remained constant. The concentrations of glucose, ethanol, acetate, and lactate were measured using enzyme kits (R-Biopharm). The alcohols could be detected using GC (Hewlett Packard model 7890A, Agilent Technologies, equipped with a

DB5-MS column, J&W Scientific) and a mass spectrometer (5975C, Agilent Technologies). The GC-MS spectrum revealed that alcohols in the culture samples were mainly ethanol and isobutanol. We also detected small amounts of propanol and methyl-butanols. Isobutanol concentration was determined by a modified GC-MS method [19]. Briefly, 400 µl of supernatant was extracted with 400 µl of toluene (Sigma-Aldrich) by 2-min vortex, followed by high-speed centrifugation (16000×g). The organic layer was taken for GC-MS analysis under the following program: hold at 70 °C for 2 min, ramp to 230 °C at 20 °C min^{-1}, and then hold at 300 °C for 6 min. The carrier gas was helium. The MS scan mode was from m/z 20 to 200. Samples were quantified relative to a standard curve of 16, 32, 64, 125, 250, 500, and 1000 mg/L isobutanol for MS detection, and methanol was taken as an internal standard.

^{13}C-experiments for analyzing carbon substrates

In the ^{13}C-experiments, the minimal medium with 2% fully labeled glucose (U-^{13}C, Cambridge Isotope Laboratories) was supplemented with either 0.1% or 0.5% yeast extract (Bacto). By measuring ^{13}C-abundance in key metabolites, we could estimate the contribution of yeast extract (non-labeled) to biomass and isobutanol synthesis in the ^{13}C-glucose medium. Specifically, 5 mL cultures (with ^{13}C-glcuose and yeast extract) were inoculated with 5 µl of overnight LB culture of engineered strain in a 50 mL falcon tube with a closed cap to prevent product vaporization (shaking at 200 rpm, 30°C). The cultures

were induced by 0.2 mM IPTG (at t=7 hours), and the samples were taken (at t=24 hours, middle-log growth phase) for isotopomer analysis. The isobutanol content was determined using the previously described GC-MS procedure [19]. The ratio of two mass-to-charge peaks (m/z=74 for unlabeled isobutanol and m/z=78 for labeled isobutanol) corresponded to the ratio of isobutanol synthesized from unlabeled yeast extract vs. labeled glucose. Concurrently, we also performed isotopic analysis of proteinogenic amino acids to identify the incorporation of unlabeled carbon from yeast extract into biomass (i.e., protein). The measurements were based on a previously published GC-MS based protocol [25], using TBDMS (N-(tert-butyldimethylsilyl)-N-methyl-trifluoroacetamide, Sigma-Aldrich) to derivatize hydrolyzed amino acids from the biomass. The m/z ions $[M-57]^+$ from unfragmented amino acids were used for analysis [26] except leucine and isoleucine. Because the $[M-57]^+$ in leucine and isoleucine overlap with other ions, the $[M-159]^+$ group was used to obtain the isotopomer labeling information of leucine and isoleucine.

Model formulation

We developed a Monod-based kinetic model, with six ordinary differential equations, to describe three different batch fermentation processes. Here, we should emphasize that the model only describe the fermentation data after IPTG induction (i.e., the *E. coli* strain entered into the butanol production phase).

$$\frac{dX}{dt} = R_X - k_d \cdot X + R_{X,YE} \quad (1)$$

$$\frac{dACT}{dt} = R_A + Y_{AL} \cdot k_{act} \cdot LACT \cdot X \quad (2)$$

$$\frac{dLACT}{dt} = R_L - k_{act} \cdot LACT \cdot X \quad (3)$$

$$\frac{dEtOH}{dt} = R_E \quad (4)$$

$$\frac{dIB}{dt} = R_{IB} + \beta \cdot X - k_{IB} \cdot IB \quad (5)$$

$$\frac{dGlu}{dt} = -\frac{R_X}{Y_{XG}} - \frac{R_A}{Y_{AG}} - \frac{R_E}{Y_{EG}} - \frac{R_L}{Y_{LG}} - \frac{R_{IB}}{Y_{IBG}} \quad (6)$$

The model contained six time-dependent process variables: *X*, *ACT*, *LACT*, *EtOH*, *IB*, and *Glu*, which represented the concentrations (g/L) of biomass, acetate, lactate, ethanol, isobutanol, and glucose, respectively. In the above equations, k_d was the cell death rate; Y_{AL} was the acetate yield from lactate (equal to 0.67 g *ACT*/g *LACT*, based on a 1:1 mol ratio); Y_{XG}, Y_{AG}, Y_{EG}, Y_{LG}, and Y_{IG} were the glucose-associated yields of biomass, acetate, ethanol, lactate, and isobutanol, respectively. k_{IB} was the first-order removal rate of isobutanol due to gas stripping under aerobic fermentation F1. The isobutanol loss was minimal in a closed bioreactor during oxygen-limited fermentations (k_{IB} was equal to zero). Under oxygen-limited conditions (i.e. F2 and F3), we used a first-order kinetic parameter (k_{act}) to describe acetate production from lactate.

$$R_X = \frac{\mu_{\max,app} \cdot Glu}{K_S + Glu} \cdot \frac{1}{1 + \frac{ACT}{K_{iA}}} \cdot X \quad (7)$$

$$R_A = \alpha_{AX} \cdot R_X \quad (8)$$

$$R_E = \alpha_{EX} \cdot R_X \quad (9)$$

$$R_L = \alpha_{LX} \cdot R_X \quad (10)$$

$$R_{IB} = \alpha_{IBX} \cdot R_X \quad (11)$$

$$R_{X,\,YE} = \mu_{\max,YE} \cdot e^{-k_{YE} \cdot t} X \quad (12)$$

Equations 7 represented a Monod-based growth kinetics. Since acetate inhibited *E. coli* growth by decreasing the intracellular pH and interfering with enzyme reactions, a non-competitive inhibition K_{iA} (assumed to be constant in F1, F2 and F3) was included in the model [27]. The dependence of the growth rate on oxygen necessitates (i.e., aerobic growth vs. anaerobic growth) was implicitly included in the calculation of the apparent maximal growth rate, $\mu_{max,app}$, (i.e., the oxygen conditions affected $\mu_{max,app}$ in different fermentation process). The Monod constant K_S was set to be constant for all three cultures (F1, F2 and F3). Equations 8~11 simulated the extracellular metabolite production associated with biomass growth. R_X, R_A, R_E, R_L, and R_{IB} were the production rates of biomass, acetate, ethanol, lactate, and isobutanol from glucose, respectively; α_{AX}, α_{EX}, α_{LX}, and α_{IBX} were the growth-associated coefficients of acetate, ethanol, lactate, and isobutanol productions, respectively. β was the non-growth associated isobutanol production coefficient. For experiments using rich medium (F3), the isotopic experiments indicated that the yeast extract was quickly consumed to support biomass growth at the early stage of the fermentation. Since precise measurement of yeast extract concentrations was difficult, we included a yeast-extract-associated biomass

growth rate ($R_{X,YE}$) using a two-parameter exponential decay function (Equation 12). Table 2 summarized model parameters and their units.

For each batch culture, 18 unknown model parameters were determined by minimizing the sum of the squares of the differences between the model's predictions and the experimentally observed growth and metabolite profiles (data after IPTG induction) [28]. The "ode23" command in MATLAB (R2009a, Mathworks) solved differential equations, while the "fmincon" command searched suitable values of parameters. To avoid having local solutions during the nonlinear parameter estimation, we perturbed the initial guesses for 30 times within the range of possible values (as observed experimentally) to identify the global solution. Meanwhile, to narrow the search-range for reasonable parameters, we assumed that the values of each parameter acquired from different fermentations differed by less than one order of magnitude.

To evaluate the quality of the parameter estimates, a statistical analysis checked the reliability and the sensitivity of the estimated parameters to the measurement inaccuracies. Fifty simulated fermentation data sets (including both biomass and metabolite data) were generated by addition of normally distributed noise to the measured fermentation data set (i.e., randomly perturbed the measured data by 30%). The same data-fitting algorithm found new sets of parameters. From the probability distribution of these parameter distributions, confidence limits of model-fitted parameters were estimated.

Results and discussion

Isobutanol fermentation results

We engineered an *E. coli* strain which was used as a model to study isobutanol fermentation kinetics. The isobutanol pathway of the strain was in Fig.1. The key precursor, 2-ketoisovalerate (KIV), is also an intermediate of valine biosynthesis and degradation pathways. When *E. coli* converts one mole of glucose to two moles of pyruvate through glycolysis, it generates two moles of NADH, whereas the isobutanol synthesis consumes one mole of NADPH by keto-acid reductoisomerase and one mole of NADH by aldehyde reductase. To keep balance, the cell metabolism has to remove the redundant NADH by O_2 oxidization (aerobic condition) or by synthesis of reduced metabolites (leading to lactate and ethanol secretion). In the F1 (Fig. 2), ethanol and lactate were barely detected. Isobutanol concentration only reached (0.2 g/L), and the *in situ* removal of isobutanol was considerable since the airflow carried isobutanol out of the fermentor. Such gas stripping has been reported as an effective strategy to avoid the isobutanol accumulation in the bioreactor to cause inhibitory effect on alcohol production [29].

In the oxygen-limited conditions (Fig. 3 & 4), the F2 generated 0.95 g/L isobutanol, 1.5 g/L ethanol, 2 g/L acetate, and 5 g/L lactate, while the lactate was reused in the late fermentation stage (stationary growth phase). With the addition

of yeast extract, the F3 had fast biomass growth (Fig. 4). The cell density reached the peak (2 g DCW/L biomass) after seven hours of IPTG induction, and glucose was consumed within ~12 hours (compared to ~40 hours in the F1 and F2). The high rates for biomass growth and glucose uptake promoted isobutanol production rate. It took the F3 15 hours to generate 0.6 g/L isobutanol (it took F2 40 hours to generate same amount of isobutanol). The F3 produced 0.7 g/L isobutanol and 2.0 g/L ethanol. The addition of yeast extract also resulted in a large amount of growth-associated organic acids (6.0 g/L lactate and 3.6 g/L acetate).

Kinetic modeling of isobutanol fermentation

The kinetic model simulated biomass growth, isobutanol production, and metabolic byproduct (lactate, ethanol, and acetate) secretion and reuse in all three culture conditions. Here, we should emphasize that all data points in Fig. 2~4 were after IPTG induction (i.e., cell entered isobutanol production metabolic status). Table 2 listed the kinetic parameters, where the maximal glucose-based cell growth rate $\mu_{max,app}$ (0.007~0.026 h^{-1}) in the anaerobic culture conditions was lower than that in the aerobic culture condition (0.051 h^{-1}). To simulate the influence of yeast extract on alcohol fermentation, the Monod growth equation included a yeast-extract associated biomass growth rate ($\mu_{max,YE}$=0.44 hr^{-1}) that was more than one order of magnitude higher than glucose-associated growth rates.

As for the intrinsic biomass yield Y_{XG}, Y_{XG} (0.18) under aerobic condition was higher than that in the anaerobic condition F2 (0.11). Aerobic fermentation produced less extracellular waste metabolites, while the citric acid cycle in cell metabolism was highly active to provide energy and building blocks for biomass and butanol synthesis. Isobutanol was produced in both aerobic and oxygen-limited conditions. The growth (α_{IBX}) and non-growth (β) associated isobutanol production coefficients were both positive, indicating the isobutanol production took place during both growth and stationary phases. Although cell metabolism shared the same pathway for synthesizing both biomass building block (valine) and isobutanol, this pathway was still functional after biomass growth stopped. The theoretical isobutanol yield from glucose is 0.41 g isobutanol/g glucose. In this study, coefficient Y_{IG} under aerobic condition was 0.25 g isobutanol/g glucose, which was higher than anaerobic conditions. For anaerobic condition F2, non-growth associated isobutanol production β was ~0.11 (g isobutanol/g biomass · h), while aerobic condition F1 showed much smaller non-growth associated productivity (β ~0.002 g isobutanol/g biomass · h). For anaerobic condition F3, the addition of yeast extract significantly improved the growth associated isobutanol production (α_{IBX} = 0.91 g isobutanol/g biomass), but it also generated significantly more acetate, lactate and ethanol. The yield coefficients associated with waste products (Y_{AG}, Y_{EG}, Y_{LG}) were all the highest in the F3. Although the addition of nutrients improved the biomass growth and cell energy (such as NADH) generation, the NADPH-dependent valine synthesis pathway could not

utilize NADH to produce isobutanol precursors (Figure 1). The excessive NADH from fast cell metabolism had to be redirected to waste extracellular metabolites. To alleviate such co-factor imbalance, an NADH-dependent keto-acid reductoisomerase or transhydrogenase *PntAB* could be used to consume NADH for isobutanol synthesis [30].

Analysis of the role of yeast extract for isobutanol synthesis

Nutrient supplements often plays important role in improving fermentation performance. Rich media have been commonly used for butanol fermentations [19,21,22,31]. In addition to providing the building blocks for biomass growth, it has been reported that *E. coli* can also utilize the Ehrlich pathway to convert protein hydrolysates to higher alcohols by engineering nitrogen flux [32]. To reveal the contribution of nutrient to biomass and isobutanol production during the cell growth stage, we used ^{13}C-experiemnts to determine the ratio of carbon utilization from two different sources (nonlabeled yeast extract vs. fully labeled ^{13}C-glucose) under oxygen limited condition via GC-MS analysis (Fig. 5). When yeast extract was minimal (1 g/L), isobutanol mostly came from labeled glucose in the medium (isobutanol was all labeled with four carbons, m/z=78), and yeast extract mainly supported biomass growth. In the more nutrient-rich medium, yeast extract (5 g/L, commonly used for biobutanol production) provided major building blocks for cell growth. Based on the labeling data, proteinogenic amino acids (histidine, leucine, isoleucine, valine, lysine, and proline) in the biomass

protein were highly imported from exogenous amino acids (>95%, corresponding to the ^{12}C-dilutions), while other proteinogenic amino acids (alanine and serine) were largely synthesized from ^{13}C-glucose (50~60%). Furthermore, high concentration of yeast extract (5g/L) provided the precursors (i.e., keto-acids from exogenous amino acid degradation pathway) for production of alcohols (non-labeled isobutanol ~ 50%). A similar observation was also found in the glucose-labeled medium supplemented with Lysogeny broth (LB) powder (data not shown). These results implied that the nutrient supplements provided building blocks for both biomass and isobutanol synthesis in the glucose-based culture medium. However, nutrient addition may not enhance butanol yields from its main carbon source (glucose) due to higher waste metabolite production.

Concluding remarks

This study developed an empirical Monod model for capturing the influential process factors during isobutanol fermentation using an engineered *E. coli* strain. The model strain shared common kinetic features with other reported isobutanol producers. These features included similar biomass growth behavior and generation of multiple-metabolites, such as acetate, ethanol, and isobutanol. The model results implied that byproducts were deleterious to isobutanol synthesis (which could be the genetic targets for future improvement). Nutrients (such as yeast extract) increased the production rates for both alcohols and organic acids, but reduced isobutanol yields. The kinetic model may not only

provide guidelines for controlling isobutanol fermentation conditions, but also serve as a springboard to develop useful bioprocess models for higher alcohols fermentations in the biotechnology industry.

Acknowledgement

This study was supported by an NSF Career Grant (MCB0954016) and the Clean Coal Consortium at Washington University. We thank Chunlei Mei for technical assistance. The authors have declared no conflict of interest.

References

1. Kovárová-Kovar K, Egli T (1998) Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. Microbiol Mol Biol Rev 62: 646-666.
2. Monod J (1949) The growth of bacterial cultures. Annu Rev Microbiol 3: 371-394.
3. Mailier J, Delmotte A, Cloutier M, Jolicoeur M, Vande Wouwer A (2011) Parametric sensitivity analysis and reduction of a detailed nutritional model of plant cell cultures. Biotechnol Bioeng 108: 1108-1118.
4. Beck JV, Arnold KJ (1977) Parameter estimation: John Wiley & Sons.
5. Lillacci G, Khammash M (2010) Parameter estimation and model selection in computational biology. PLoS Comput Biol 6(3):e1000696.
6. Tang YJ, Meadows AL, Keasling JD (2007) A kinetic model describing *Shewanella oneidensis* MR-1 growth, substrate consumption, and product secretion. Biotechnol Bioeng 189: 894-901.
7. Myers CR, Nealson KH (1988) Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. Science 240: 1319-1321.
8. Tiedje JM (2002) *Shewanella*- the environmentally versatile genome. Nature Biotechnology 20: 1093-1094.
9. Logan BE, Murano C, Scott K, Gray ND, Head IM (2005) Electricity generation from cysteine in a microbial fuel cell. Water Research 39: 942-952.
10. Ju P, Handschin E, Wei ZN, Schlucking U (1996) Sequential parameter estimation of a simplified induction motor load model. IEEE Transactions on Power Systems 11.

11. Djurić PM, Kotecha JH, Esteve F, Perret E (2002) Sequential Parameter Estimation of Time-Varying Non-Gaussian Autoregressive Processes. EURASIP Journal on Applied Signal Processing 2002: 865-875.
12. Faber R, Li P, Wozny G (2003) Sequential Parameter Estimation for Large-Scale Systems with Multiple Data Sets. 1. Computational Framework. Ind Eng Chem Res 42: 5850-5860.
13. Faber R, Li P, Wozny G (2004) Sequential Parameter Estimation for Large-Scale Systems with Multiple Data Sets. 2. Application to an Industrial Coke-Oven-Gas Purification Process. Ind Eng Chem Res 43: 4350-4362.
14. Tang YJ, Laidlaw D, Gani K, Keasling JD (2006) Evaluation of the effects of various culture conditions on Cr(VI) reduction by *Shewanella oneidensis* MR-1 in a novel high-throughput mini-bioreactor. Biotechnology and Bioengineering 95: 176-184.
15. Pinchuk GE, Hill EA, Geydebrekht OV, Ingeniis JD, Zhang X, et al. (2010) Constraint-based model of Shewanella oneidensis MR-1 metabolism: a tool for data analysis and hypothesis generation. PLoS Comput Biol 6: e1000822.
16. Ritz C, Streibig JC, editors (2008) Nonlinear Regression with R, p 96-97: Springer.
17. Jones DT, Woods DR (1986) Acetone-butanol fermentation revisited. Microbiol Rev 50: 484-524.
18. Atsumi S, Cann AF, Connor MR, Shen CR, Smith KM, et al. (2008) Metabolic engineering of Escherichia coli for 1-butanol production. Metab Eng 10: 305-311.
19. Bond-Watts BB, Bellerose RJ, Chang MC (2011) Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. Nat Chem Biol 7: 222-227.
20. Shen CR, Lan EI, Dekishima Y, Baez A, Cho KM, et al. (2011) Driving Forces Enable High-Titer Anaerobic 1-Butanol Synthesis in Escherichia coli. Appl Environ Microbiol 77: 2905-2915.
21. Steen EJ, Chan R, Prasad N, Myers S, Petzold CJ, et al. (2008) Metabolic engineering of Saccharomyces cerevisiae for the production of n-butanol. Microb Cell Fact 7: 36.
22. Atsumi S, Hanai T, Liao JC (2008) Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451: 86-89.
23. Hazelwood LA, Daran JM, van Maris AJ, Pronk JT, Dickinson JR (2008) The Ehrlich pathway for fusel alcohol production: a century of research on Saccharomyces cerevisiae metabolism. Appl Environ Microbiol 74: 2259-2266.
24. Sentheshanmuganathan S, Elsden SR (1958) The mechanism of the formation of tyrosol by Saccharomyces cerevisiae. Biochem J 69: 210-218.

25. Feng X, Bandyopadhyay A, Berla B, Page L, Wu B, et al. (2010) Mixotrophic and photoheterotrophic metabolism in Cyanothece sp. ATCC 51142 under continuous light. Microbiology 156: 2566-2574.
26. Wahl SA, Dauner M, Wiechert W (2004) New tools for mass isotopomer data evaluation in (13)C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. Biotechnol Bioeng 85: 259-268.
27. Luli GW, Strohl WR (1990) Comparison of growth, acetate production, and acetate inhibition of Escherichia coli strains in batch and fed-batch fermentations. Appl Environ Microbiol 56: 1004-1011.
28. Tang YJ, Meadows AL, Keasling JD (2007) A kinetic model describing Shewanella oneidensis MR-1 growth, substrate consumption, and product secretion. Biotechnol Bioeng 96: 125-133.
29. Baez A, Cho K-M, Liao J (2011) High-flux isobutanol production using engineered Escherichia coli: a bioreactor study with in situ product removal. Applied Microbiology and Biotechnology: 1-10.
30. Bastian S, Liu X, Meyerowitz JT, Snow CD, Chen MM, et al. (2011) Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-methylpropan-1-ol production at theoretical yield in Escherichia coli. Metab Eng 13: 345-352.
31. Nielsen DR, Leonard E, Yoon SH, Tseng HC, Yuan C, et al. (2009) Engineering alternative butanol production platforms in heterologous bacteria. Metab Eng 11: 262-273.
32. Huo YX, Cho KM, Rivera JG, Monte E, Shen CR, et al. (2011) Conversion of proteins into biofuels by engineering nitrogen flux. Nat Biotechnol 29: 346-351.
33. Atsumi S, Higashide W, Liao JC (2009) Direct photosynthetic recycling of carbon dioxide to isobutyraldehyde. Nat Biotechnol 27: 1177-1180.
34. Higashide W, Li Y, Yang Y, Liao JC (2011) Metabolic engineering of Clostridium cellulolyticum for production of isobutanol from cellulose. Appl Environ Microbiol 77: 2727-2733.
35. Trinh CT, Li J, Blanch HW, Clark DS (2011) Redesigning Escherichia coli metabolism for anaerobic production of isobutanol. Appl Environ Microbiol 77: 4894-4904.
36. Shen CR, Liao JC (2008) Metabolic engineering of Escherichia coli for 1-butanol and 1-propanol production via the keto-acid pathways. Metab Eng 10: 312-320.
37. Lan EI, Liao JC (2011) Metabolic engineering of cyanobacteria for 1-butanol production from carbon dioxide. Metab Eng 13: 353-363.

Table 1. **Recent studies on butanol production by genetically engineered microorganisms**

Products	Substrate	Host cell	Titer	Research Highlights	Ref
Isobutanol	Glucose	*E. coli*	22 g/L	Introduction of a non-fermentative pathway to produce isobutanol; elimination of competing pathways to reduce waste metabolite secretion	[22]
Isobutanol	Glucose	*E. coli*	50 g/L	*In situ* isobutanol removal from the bioreactor using gas stripping	[29]
Isobutanal	CO_2	*Synechococcus elongatus*	1.1 g/L	Overexpression of both non-fermentative pathway and *Rubisco* for autotrophic isobutanol production	[33]
Isobutanol	Cellulose	*Clostridium cellulolyticum*	0.66 g/L	Direct conversion of cellulose to isobutanol using engineered cellulolytic bacterial species	[34]
Isobutanol	Glucose	*E. coli*	1.7 g/L	A strain optimized for isobutanol production via elementary mode analysis	[35]
Isobutanol	Glucose	*E. coli*	13.4 g/L	Utilization of the NADH-dependent enzyme to enable anaerobic isobutanol production	[30]
Isobutanol	Amino acids	*E. coli*	~1g/L	Utilization of protein hydrolysates for C4 and C5 alcohols synthesis by introducing enzymes for exogenous transamination and deamination cycles to biobutanol producing strains	[32]
n-Butanol	Glucose	*E. coli*	1 g/L	A strain engineered for 1-butanol and 1-propanol production via isoleucine biosynthesis pathway	[36]
n-Butanol	Galactose	*Saccharomyces cerevisiae*	2.5 mg/L	Overexpression of *n*-butanol pathway derived from *Clostridium*	[21]
n-Butanol	Glucose	*E. coli*	4.6 g/L	Increase of the barrier for the reverse reaction of butyryl-CoA to crotonyl-CoA via trans-enoyl-CoA reductase	[19]
n-Butanol	Gluocse	*E. coli*	30 g/L	Construction of butanol pathway by employing trans-enoyl-CoA reductase and creating NADH & acetyl-CoA driving forces	[20]
n-Butanol	CO_2	*Synechococcus elongatus* 7942	14.5 mg/L	Anaerobic production of 1-butanol from CO_2 using CoA-dependent butanol pathway	[37]

Table 2. Parameters of Monod kinetic model for *E. coli* isobutanol fermentation

Symbol	Notations	Units	Aerobic, condition F1	Anaerobic, condition F2	Anaerobic, condition F3
K_S	Monod parameter for Glu	g/L		0.30±0.08 [a]	
K_{iA}	Inhibition of acetate for growth	g/L		26±5 [a]	
$\mu_{max,app}$	Maximal growth rate with glucose	/h	0.051±0.005	0.0070±0.0020	0.026±0.004
Y_{XG}	Intrinsic biomass yield from Glu	g biomass/g glucose	0.18±0.03	0.11±0.01	0.23±0.03
Y_{AG}	Acetate yield from Glu	g acetate/g glucose	0.08±0.01	0.08±0.01	0.39±0.03
Y_{EG}	Ethanol yield from Glu	g ethanol/g glucose	NA	0.27±0.01	0.46±0.03
Y_{LG}	Lactate yield from Glu	g lactate/g glucose	NA	0.60±0.02	0.85±0.06
Y_{IG}	Isobutanol yield from Glu	g isobutanol/ g glucose	0.25±0.05	0.03±0.00	0.18±0.02
α_{AX}	Growth associated acetate synthesis	g acetate/g biomass	0.62±0.02	0.29±0.03	2.8±0.3
α_{EX}	Growth associated ethanol synthesis	g ethanol/g biomass	NA	6.7±1.0	3.2±0.3
α_{LX}	Growth associated lactate synthesis	g lactate/g biomass	NA	26±4	11±2
α_{IBX}	Growth associated isobutanol synthesis	g isobutanol/ g biomass	0.57±0.04	0.10±0.01	0.91±0.12
Y_{AL}	Acetate yield from lactate	g acetate/g lactate	NA	0.67	0.67
k_d	Cell death rate	/h	0.010±0.002	0.0010±0.0005	0.0087±0.0007
k_{IB}	gas stripping rate coefficient for isobutanol	/h	0.11±0.02	NA	NA
k_{act}	First-order acetate production from lactate	$(h \cdot g\ biomass/L)^{-1}$	NA	0.012±0.001	0.0022±0.0003
k_{YE}	Yeast extract first-order decay rate	/h	NA	NA	0.55±0.04
$\mu_{max,YE}$	Apparent maximal growth rate with yeast extract	/h	NA	NA	0.44±0.05
β	Non-growth associated isobutanol production rate	›butanol)/(g omass·h)	0.0021±0.0017	0.011±0.001	0.0038±0.0006

a): assuming same value for all three fermentations.

Glu: glucose; NA: not applicable.

Figure legend

Figure 1. Pathways and kinetics for isobutanol and metabolites production in the recombinant *E. coli* BL21 (DE3)/pTAC-KA. R_X, $R_{X,YE}$, R_A, R_E, R_L, and R_{IB} were shown in the Equation 1~12. Abbreviations: G6P, glucose 6-phosphate; F6P, fructose 6-phosphate; G3P, glyceraldehydes 3-phosphate; PYR, pyruvate; ACA, acetyl-CoA; ALA, 2-acetolactate; DHI, 2,3-dihydroxy-isovalerate; KIV, 2-keto-isovalerate; Val, valine; IBA, isobutanal; NADH, reduced nicotinamide adenine dinucleotide; NADPH, reduced nicotinamide adenine dinucleotide phosphate; PP pathway, pentose phosphate pathway; TCA cycle, tricarboxylic acid cycle; Kivd, 2-keto-isovalerate decarboxylase; AdhA, aldehyde reductase.

Figure 2: Growth kinetics of aerobic culture after IPTG induction (condition F1). The dots were experimental measurements, and the solid lines are simulations from the Monod kinetic model.

Figure 3: Growth kinetics of anaerobic culture after IPTG induction (condition F2). The dots are experimental measurements, and the solid lines are simulations from the Monod kinetic model.

Figure 4: Growth kinetics of anaerobic culture with 0.5% yeast extract after IPTG induction (the F3 includes two replicated cultivation experiments). The dots were experimental measurements, and the solid lines were simulations from the Monod kinetic model.

Figure 5. The fraction of ^{13}C carbon (^{13}C carbons/total carbons) in each metabolite. The fraction of ^{13}C carbon represents the contribution of labeled glucose to metabolite synthesis. The cells were grown on M9 medium (Difco), containing 20 g/L fully labeled ^{13}C-glucose, with 1 g/L (black bar) or 5 g/L (gray bar) nonlabeled yeast extract. All samples were taken at 24 hrs (n=2, GC-MS standard errors < 2%).

Fig. 1

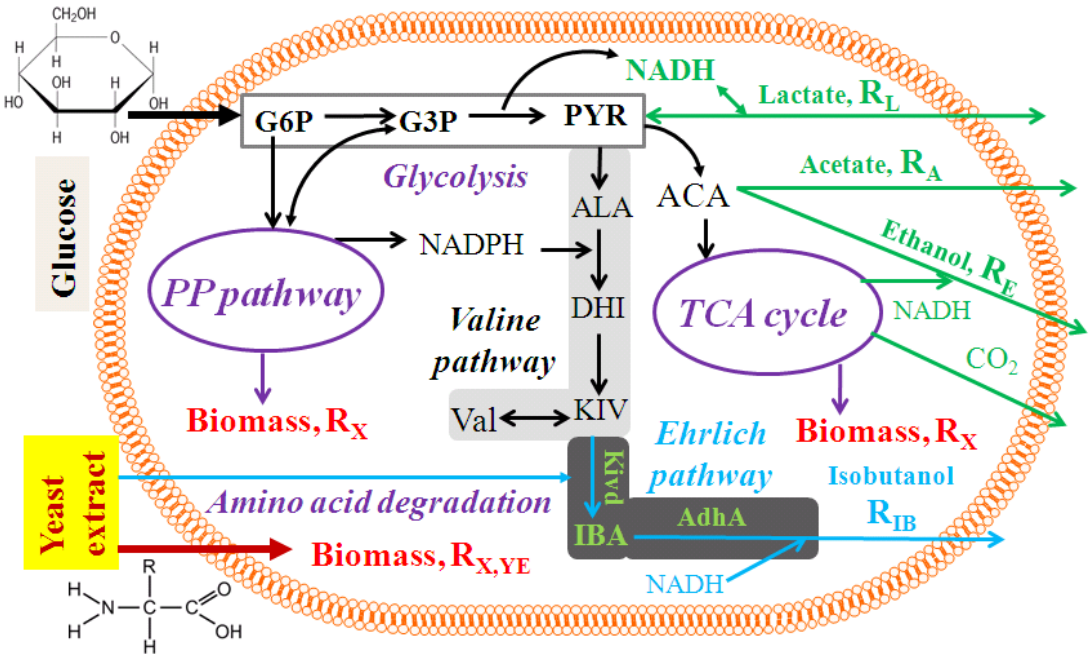

Glucose
G6P
G3P
PYR
NADH
Lactate, R_L
Glycolysis
Acetate, R_A
ALA
ACA
Ethanol, R_E
PP pathway
NADPH
Valine pathway
DHI
TCA cycle
NADH
CO_2
Biomass, R_X
Val
KIV
Ehrlich pathway
Biomass, R_X
Yeast extract
Amino acid degradation
KivD
Isobutanol
R_{IB}
IBA
AdhA
Biomass, $R_{X,YE}$
NADH

Fig. 2

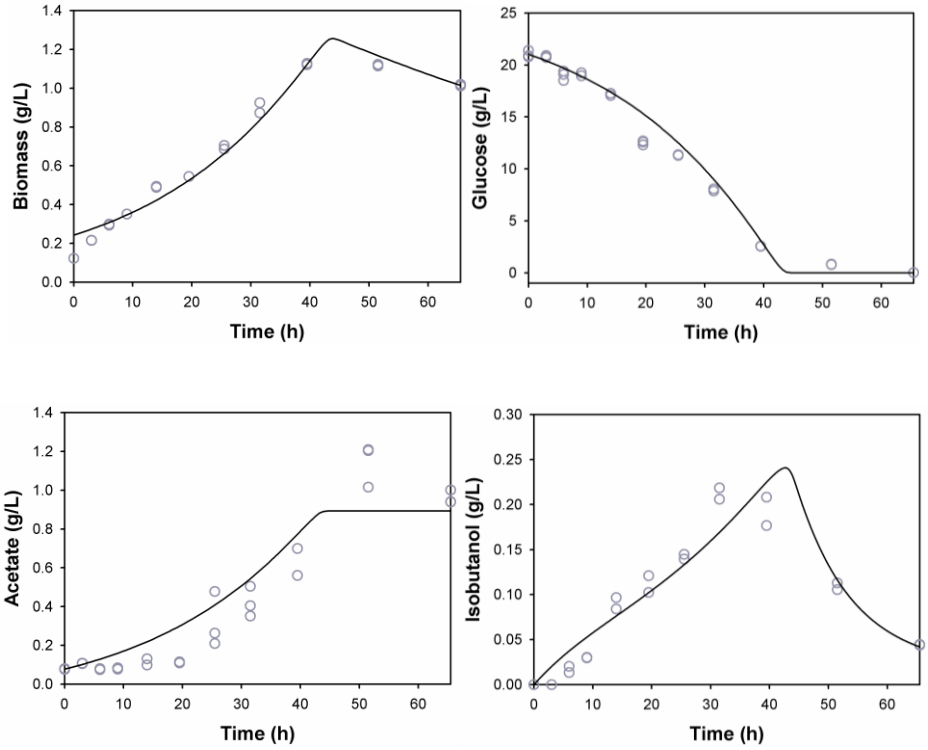

Biomass (g/L)
Time (h)
Glucose (g/L)
Time (h)
Acetate (g/L)
Time (h)
Isobutanol (g/L)
Time (h)

Fig. 3

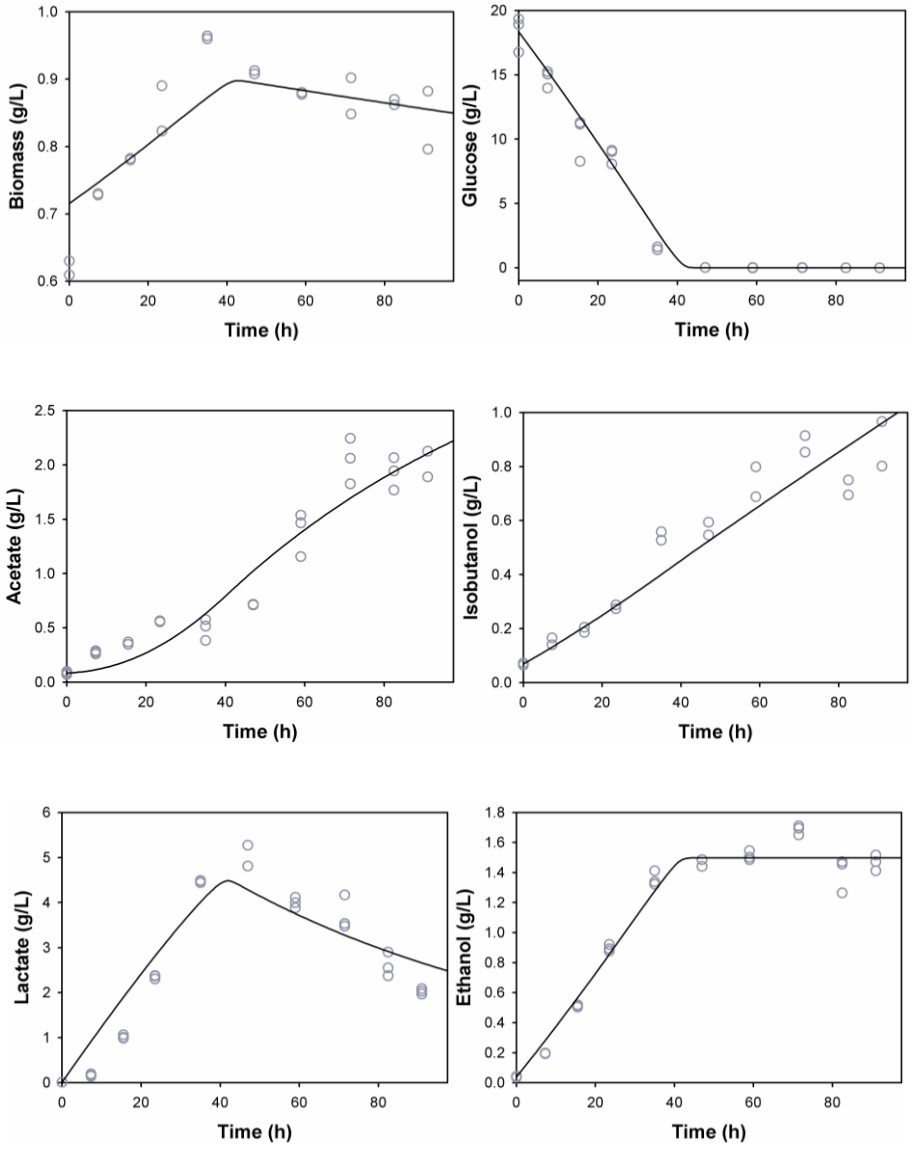

Biomass (g/L)
Time (h)
Glucose (g/L)
Time (h)
Acetate (g/L)
Time (h)
Isobutanol (g/L)
Time (h)
Lactate (g/L)
Time (h)
Ethanol (g/L)
Time (h)

Fig. 4

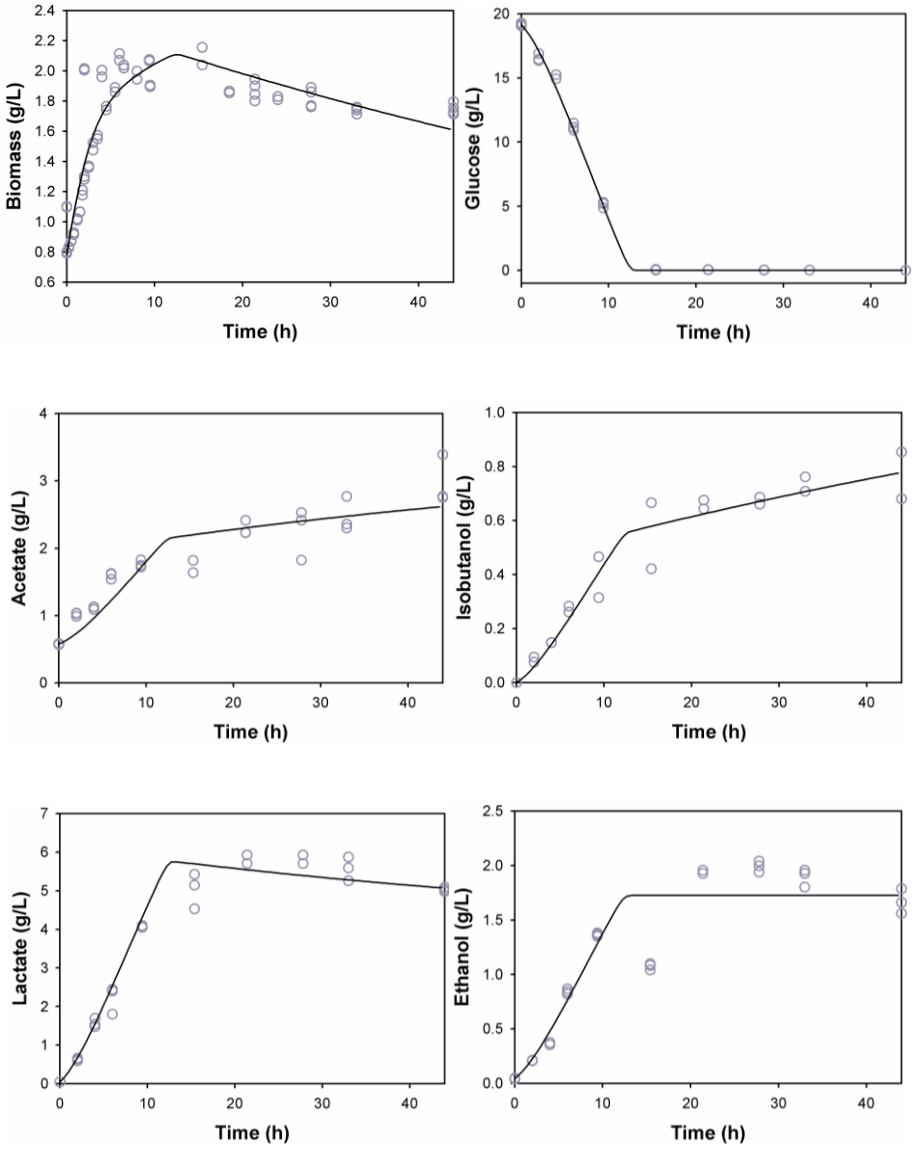
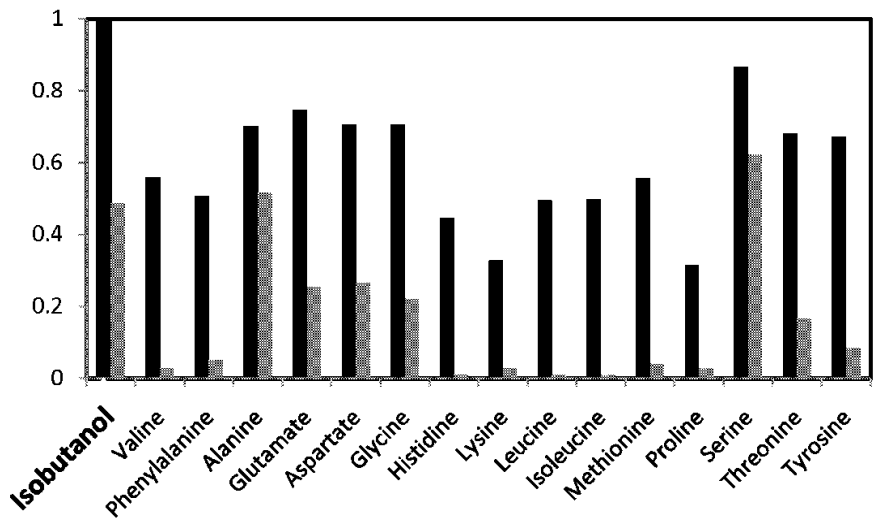

Biomass (g/L)
Time (h)
Glucose (g/L)
Time (h)
Acetate (g/L)
Time (h)
Isobutanol (g/L)
Time (h)
Lactate (g/L)
Time (h)
Ethanol (g/L)
Time (h)

Fig. 5


1
0.8
0.6
0.4
0.2
0
Isobutanol
Valine
Phenylalanine
Alanine
Glutamate
Aspartate
Glycine
Histidine
Lysine
Leucine
Isoleucine
Methionine
Proline
Serine
Threonine
Tyrosine

Appendix 11

Evaluating Factors That Influence Microbial Synthesis Yields by Linear Regression with Numerical and Ordinal Variables

ARTICLE

BIOTECHNOLOGY and BIOENGINEERING

Evaluating Factors That Influence Microbial Synthesis Yields by Linear Regression with Numerical and Ordinal Variables

Peter F. Colletti, Yogesh Goyal, Arul M. Varman, Xueyang Feng, Bing Wu, Yinjie J. Tang

Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, Missouri 63130; telephone: 314-935-3441; fax: 314-935-7211; e-mail: yinjie.tang@seas.wustl.edu

Received 3 August 2010; revision received 18 October 2010; accepted 20 October 2010

Published online 12 November 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/bit.22996

ABSTRACT: In the production of chemicals via microbial fermentation, achieving a high yield is one of the most important objectives. We developed a statistical model to analyze influential factors that determine product yield by compiling data obtained from engineered *Escherichia coli* developed within last 10 years. Using both numerical and ordinal variables (e.g., enzymatic steps, cultivation conditions, and genetic modifications) as input parameters, our model revealed that cultivation modes, nutrient supplementation, and oxygen conditions were the three significant factors for improving product yield. Generally, the model showed that product yield decreases as the number of enzymatic steps in the biosynthesis pathway increases (7–9% loss of yield per enzymatic step). Moreover, overexpression of enzymes or removal of competitive pathways (e.g., knockout) does not necessarily result in an amplification of product yield (*P*-value >0.1), possibly because of limited capacity in the biosynthesis pathway to accommodate an increase in flux. The model not only provides general guidelines for metabolic engineering and fermentation processes, but also allows a priori estimation and comparison of product yields under designed cultivation conditions.

Biotechnol. Bioeng. 2011;108: 893–901.

KEYWORDS: enzymatic steps; *Escherichia coli*; flux; nutrients; overexpression; *P*-value

Introduction

In light of rising interests and investments in green biotechnology, numerous efforts have focused on the metabolic engineering of microbial hosts to synthesize pharmaceuticals and biofuels (Jarboe et al., 2010; Stephanopoulos et al., 1998). For economical production of a target chemical compound, it is crucial for a microbial process to achieve high yield, titer, and productivity. Therefore, an array of engineering strategies has been routinely used for improving production, including introduction of heterologous genes, modulation of gene expression, protein modification, and high-throughput phenotype screen methods (Clomburg and Gonzalez, 2010). Optimizations of fermentation characteristics such as culture medium, pH control, temperature, and oxygen conditions are also equally important.

However, it is generally difficult to identify rate-limiting components in a complex cellular metabolism in order to improve microbial performance (Stephanopoulos et al., 2004). To this end, systems biology-based models have been developed to provide useful information for rationally engineering microbial hosts with the desired phenotype as well as to design optimal fermentation conditions (Blazeck and Alper, 2010; Boghigian et al., 2010; Feist et al., 2010; Meadows et al., 2010). Cell-wide metabolic analysis via fluxomics and metabolic control theories are often used to estimate metabolic potential, product yield, nutrient limitations, and gene targets for metabolic engineering (Feist et al., 2010; Wildermuth, 2000). For example, to advance the industrial application of *Escherichia coli*, genome-scale flux balance analysis integrates a series of physical, chemical, and biological characteristics (e.g., thermodynamic directionality and energy balance) to evaluate theoretical yields for multiple native products from different feed stocks. Nevertheless, it is still difficult to reliably predict the yield produced by a culture from a given set of genetic changes and cultivation parameters. To address this, the present study proposes an empirical model based on recently published microbial production data to provide insight into the important parameters for yield optimization. We construct a linear regression model which accounts for both numerical and ordinal variables to

Peter F. Colletti and Yogesh Goyal contributed equally to this work.

Yogesh Goyal's present address is Department of Chemical Engineering, Indian Institute of Technology, Gandhinagar, Gujarat 382424, India.

Correspondence to: Yinjie J. Tang

Contract grant sponsor: NSF Career Grant

Contract grant number: MCB0954016

Additional Supporting Information may be found in the online version of this article.

investigate their effects on metabolic yield. Without a priori knowledge of enzyme kinetics, biomass growth, or the stoichiometry of metabolic reactions, the model is able to estimate and compare microbial "chassis" for product synthesis. Our model provides general guidelines for pathway modifications, bioprocess optimization, and cost-estimation in industrial biotechnology.

Model Development

The construction of the empirical model focuses on *E. coli* because it is one of the most common "industrial workhorses." It is capable of utilizing a variety of carbon sources (such as glucose, xylose, glycerol, and fatty acids) and is easily manipulated by current recombinant DNA technology. Table I lists the key influential factors that may control product yields from *E. coli*. Product biosynthesis requires a "common" route to generate metabolic intermediates such as pyruvate and acetyl-CoA via central metabolism (i.e., glycolysis, TCA cycle, and pentose phosphate pathway). A "specialized" pathway then converts the common precursors into the compounds of interest. In our model, we first divide metabolic steps into two separate numerical variables. The first variable (primary pathway) is the number of steps from the chosen carbon source to the appropriate common metabolic precursor in the central metabolism; the second variable (secondary pathway) counts the number of enzymatic steps from the common precursors to the final product of interest (Fig. 1). The choice to use two distinct variables is based on the fact that cellular central metabolisms often show high fluxes and a robust nature for synthesizing central metabolites (e.g., pyruvate and TCA acids), while the metabolic fluxes through secondary pathways to more complex chemicals are relative small. If multiple carbon sources are used in the medium, we count the number of enzymatic steps from the carbon source which is the fewest number of steps away from the desired product.

To further develop the model, the remaining factors listed in Table I are described as ordinal-dependent variables. Each ordinal variable has either two or three categories with an intrinsic ordering to the categories. For example, nutrient conditions are divided into two categories (low and high nutrient supplements). Table II shows the categories for each ordinal-dependent variable. The relationship between model predicted yield (on a logarithmic scale) and the independent variables (regressors) can be correlated using linear regression (Tang et al., 2005), such that:

$$\begin{aligned}\log_{10} Y = {} & \beta_0 + \beta_{\mathrm{PRI}}\mathrm{PRI} + \beta_{\mathrm{SEC}}\mathrm{SEC} \\ & + \beta_{\mathrm{MOD,C2}}\mathrm{MOD_{C2}} + \beta_{\mathrm{MOD,C3}}\mathrm{MOD_{C3}} \\ & + \beta_{\mathrm{KNO,C2}}\mathrm{KNO_{C2}} + \beta_{\mathrm{NUT,C2}}\mathrm{NUT_{C2}} \\ & + \beta_{\mathrm{CUL,C2}}\mathrm{CUL_{C2}} + \beta_{\mathrm{OXY,C2}}\mathrm{OXY_{C2}} \\ & + \beta_{\mathrm{TMP,C2}}\mathrm{TMP_{C2}} \qquad (1)\end{aligned}$$

where $\log_{10} Y$ is the response variable. Y represents the yield (mol C in product/mol C in primary substrate), given the set of independent variables i (i.e., influential factors). β_i, with

Table I. Review of influential factors affecting final product yield by *Escherichia coli*.

Factors	Mechanisms and implications	Refs.
Number of primary and secondary metabolic steps	The overall pathways for the microbial biosynthesis can be separated into two elements. The first element encompasses the pathways required to generate key metabolic intermediates from a carbon source, while the second element includes the pathways which convert these intermediates to the product	Clomburg and Gonzalez (2010)
Extent of pathway improvement	Genetic engineering, such as overexpression of native or introduction of non-native enzymes, can increase the efficiency of product biosynthesis and allow production of new compounds	Jarboe et al. (2010) and Stephanopoulos et al. (1998)
Removal of competitive pathways	Strategic deletion of specific genes or down-regulation of competitive pathways can re-direct fluxes to targeted pathways and reduce the accumulation of toxic metabolites	Bailey (1991) and Clark and Blanch (1997)
Nutrients	Cell growth requires various nutrients. Addition of extra carbon sources, intermediates or precursor compounds can improve the productivity	Pelczar et al. (1993)
Cultivation modes	Shaking flasks do not provide controlled pH and temperature conditions; a bioreactor regulates these parameters. In addition, a bioreactor provides good agitation and mixing conditions. Fed-batch operation prevents substrate inhibition which may occur if all of the nutrients are added simultaneously	Clark and Blanch (1997)
Oxygen conditions	Respiration supports high biomass growth rate but not necessarily high yield. Some products have to be fermented under low oxygen conditions. Fermentative conditions often result in the production of mixed acids and alcohol	Clark and Blanch (1997) and Pelczar et al. (1993)
Temperatures	Temperature affects enzyme activity by the Arrhenius law. An increase in temperature may also result in a decrease in dissolved oxygen content in the culture	Pelczar et al. (1993)

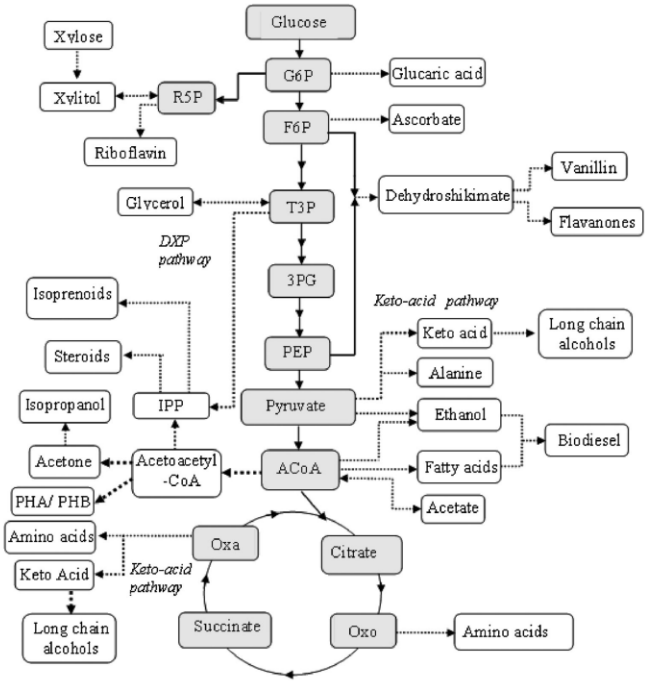


Figure 1. Simplified metabolic pathways for product synthesis. Solid arrows (→) represents primary pathways; dotted arrows represents secondary pathways (check Supplemental Fig. 1 for a more detailed pathway map). 3PG, 3-phosphoglycerate; ACoA, acetyl-CoA; F6P, fructose 6-phosphate; G6P, glucose 6-phosphate; T3P, glyceraldehyde 3-phosphate; IPP, isopentenyl pyrophosphate; Oxa, oxaloacetate; Oxo, 2-oxoglutarate; PEP, phosphoenolpyruvate; PHA, polyhydroxy acids; PHB, poly(3-hydroxybutyrate); R5P, ribose 5-phosphate.

the exception of β_0, is the regression coefficient of a variable i (numerical or ordinal variable), which reflects the contribution of the corresponding variable to the final product yield. The ordinal variables are assigned a value of 1 if and only if the condition fits the category in Table II; otherwise, that variable is assigned a value of 0 (Weisberg, 1985). β_0 is the intercept, which represents the combined contribution of the Category 1 of all ordinal variables. β_0 is given by Equation (2):

$$\beta_0 = \beta_{MOD,C1} + \beta_{KNO,C1} + \beta_{NUT,C1} + \beta_{CUL,C1} + \beta_{OXY,C1} + \beta_{TMP,C1} \quad (2)$$

In the final step, we compile a comprehensive database for various compounds (including isoprenoids, biofuels, flavanones, amino acids, etc.) that have been produced using *E. coli* from 40 recently published articles (Supplementary Table I). Included in this table are the categories assigned to the experimental conditions and the reported product yield. The software package R (Team, 2009) is used to find the regression coefficients and *P*-values for Equation (1). A particular coefficient is considered statistically significant (90% probability) if the *P*-value is below 0.1.

Results

We analyze the contributions of both numerical and ordinal parameters to product yield by examining the regression model derived from the compiled data (Supplementary Table I). In general, we only select the best yields achieved by each article for the regression model (i.e., choose "the

Table II. Variables used in the linear regression model.

Notation	Category 1 (subscript C1)	Category 2 (subscript C2)	Category 3 (subscript C3)
Categorical variables			
MOD: pathway improvement (i.e., overexpression)	No genetic modification of biosynthetic pathway	Moderate pathway modification (<3 genes)	Extensive pathway modification (≥3 genes)
KNO: removal of competitive pathways (i.e., gene knockout)	No competitive pathway is removed	At least one competitive pathway is inhibited or removed	
NUT: nutrient sources	Fermentation occurs in defined medium (only includes trace amounts of amino acids or vitamins)	Fermentation occurs in a rich medium (nutrient carbon source >10% of primary carbon source)	
CUL: cultivation modes	Fermentation occurs in a shaking flask	Fermentation occurs in a well-controlled bioreactor	
OXY: oxygen conditions	Fermentation occurs in aerobic conditions	Fermentation occurs under oxygen-limited conditions	
TMP: fermentation temperatures	Fermentation occurs at 34°C or cooler	Fermentation occurs above 34°C	

Notation—numerical variables: steps from substrate to product.
PRI: number of reaction steps in central metabolic pathways (solid arrow in Fig. 1).
SEC: number of reaction steps in secondary metabolic pathways (dotted arrow in Fig. 1).

highest yield'' under each unique production condition). The regression is based on 155 yield data (in logarithmic scale) of 36 chemicals from different carbon substrates (e.g., glucose, glycerol, etc.). From these data points, we obtain the following coefficients for Equation (1):

$$\log_{10} Y = -0.88 - 0.031\mathrm{PRI} - 0.041\mathrm{SEC} - 0.22\mathrm{MOD}_{C2} - 0.32\mathrm{MOD}_{C3} - 0.018\mathrm{KNO}_{C2} + 0.28\mathrm{NUT}_{C2} + 0.46\mathrm{CUL}_{C2} + 0.35\mathrm{OXY}_{C2} + 0.018\mathrm{TMP}_{C2} \quad (3)$$

Figure 2 shows a plot of the product yields obtained experimentally against those predicted by the model for the corresponding conditions. The correlation of this model to the 155-point dataset is fair, with an R^2 value of 0.46 that represents the uncertainty attributed to the sample size and distribution. The magnitude of each coefficient indicates the extent to which the engineering of the organism and the experimental conditions affect the yield of desired product. The coefficient values must be interpreted in conjunction with the standard error and the *P*-value. Here we have used a *P*-value of 0.1 (>90% probability) as the limit below which we have considered the results to be significant (du Prel et al., 2009). Out of the nine variables used, we find that six of them have *P*-value of ≤0.1 (Table III). To ensure that the generated model is statistically sound, we have examined the

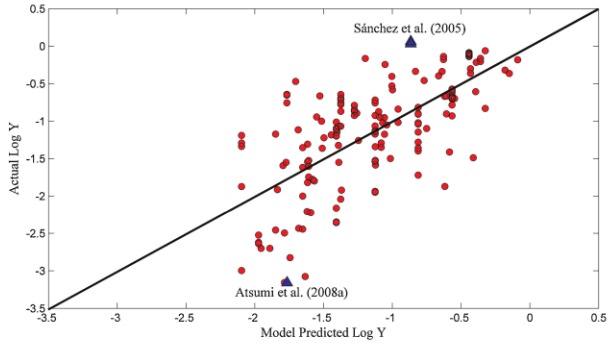


Figure 2. Distribution of actual $\log_{10} Y$ versus model generated $\log_{10} Y$ for *Escherichia coli*. The line drawn is one-to-one and passes through the origin. The R^2 value for the model is 0.46. (▲) Corresponds to the two articles discussed in the main text. [Color figure can be seen in the online version of this article, available at http://wileyonlinelibrary.com/bit]

Table III. Regression coefficients for the linear model comprising of eight factors (six factors are ordinal and two factors are numerical).

	Category 2			Category 3		
	Coefficient	Standard error	*P*-value	Coefficient	Standard error	*P*-value
Ordinal variables						
MOD	−0.22	0.15	0.16	−0.32	0.16	0.040
KNO	−0.018	0.11	0.88			
NUT	0.28	0.10	0.0054			
CUL	0.46	0.12	0.0036			
OXY	0.35	0.11	0.0027			
TMP	0.018	0.12	0.88			
Numerical variables						
PRI	−0.031	0.017	0.063			
SEC	−0.041	0.015	0.007			

Intercept: −0.88; standard error: 0.24; *P*-value: 0.00036; R^2: 0.46.

standardized residuals of the data. The standardized residual is given by the quotient of the residual of a value and its standard deviation. We found that 96% of the standardized residuals of the dataset lie between −2 and 2. Therefore, the multiple linear regression model is statistically significant and can be a valuable tool to predict yield (Anderson et al., 2009; Montgomery et al., 2006). A plot of the standardized residual values can be found in the supplementary data (Supplementary Fig. 2).

The obtained value of β_0 (-0.88 ± 0.24) corresponds to a product yield when all of the ordinal variables for Category 2 and Category 3 are set equal to zero. The intercept β_0 reflects the yield obtained from these base conditions. To calculate product yield under actual conditions, we have to assign variables for influential factors (Supplementary Table I). For example, to predict ethanol synthesis by a wild-type strain in an anaerobic bioreactor, we can choose the following variables (Table II): PRI = 10; SEC = 2; MOD = KNO = NUT = Category 1; CUL = OXY = TMP = Category 2. The model predicts:

$$\text{Yield} = 10^{-0.88+(-0.031\times 10)+(-0.041\times 2)+0.46+0.35+0.018} = 0.36 \quad (4)$$

In this example, the result suggests that the majority of carbon supplied to the culture is used for the production of biomass and waste metabolites.

Six factors fit our criterion for significance: PRI, SEC, CUL_{C2}, OXY_{C2}, MOD_{C3}, and NUT_{C2}. As expected, the values of PRI (−0.031) and SEC (−0.041) are negative. The coefficients of these numerical variables represent the amount of decrease in log-yield for every additional enzymatic step: each enzyme may reduce the carbon yield by 7% $(=10^{-0.031})$ in the primary pathway and by 9% $(=10^{-0.041})$ in the secondary pathway. The two coefficients represent the loss of product yield that can accumulate as the metabolic pathway gets longer. Therefore, products from primary pathways such as pyruvate and succinate may have a yield close to the theoretical limit, but the yield of products requiring many enzyme steps will deviate significantly from the theoretical output because of imperfect enzyme specificity and activity. We note that fermentation using a well-controlled bioreactor (reducing substrate inhibition, pH perturbation, oxygen maintenance, etc.) can improve the yield by 2.9 times $(=10^{0.47})$ when compared to a shake-flask mode. Meanwhile, fed-batch culture allows the use of high concentrations of substrates so that the product titer can be also improved. Under anaerobic or microaerobic conditions, the yield is increased by a factor of 2.2 $(=10^{0.35})$ compared to aerobic conditions. These low oxygen conditions, however, limit biomass growth and metabolic reaction rates.

On the other hand, the effect of temperature on the product yield is not significant because most culture temperatures are between 30 and 37°C. In this range, the difference in yield between Category 1 (≤34°C) and Category 2 (>34°C) is minimal $(4\% = 10^{-0.018})$. Furthermore, the coefficients for MOD_{C2} and MOD_{C3} are both negative, suggesting that genetic modification of pathways may result in a decreased yield compared to the base condition where the pathway native to the host is employed. Since the *P*-value for MOD_{C2} is 0.16, there is a probability that a moderate genetic modification may still result in an increased yield. The *P*-value for MOD_{C3} is 0.04, which indicates that extensive pathway engineering will possibly result in a low yield (>95%) in *E. coli* even for carefully designed genetic strategies. There are a few explanations for this observation. Many studies involve the importation and expression of multiple heterologous genes for biosynthesis, and non-native pathways may be incompatible or inferior to the native pathway in the host, which may lead to lower biosynthetic efficiency. Second, pathway engineering often induces metabolic imbalance and accumulation of toxic metabolites (Atsumi et al., 2008b). Third, the initial rate-limiting enzymes in the pathway may shift to other enzymes after overexpression because of inherent low enzyme activities through the entire pathway, that is, limited capacity in the biosynthesis pathway cannot accommodate flux increase (Leonard et al., 2010).

Without a doubt, gene knockout is a useful tool to improve product yield (especially for biosynthesis of metabolites in the central metabolism, e.g., succinate) (Sánchez et al., 2006). However, we also note that the coefficient of KNO_{C2} is slightly negative (−0.018) which means that removal of competitive pathways (i.e., genetic knockout) may not successfully improve the biosynthesis yield. While this conclusion seems highly uncertain (P-value = 0.88; standard error = 0.11), it has been shown via in silico analysis that knockout strategies cannot ensure the improvement for product yield even though it is expected to channel more carbon to the final product (Blazeck and Alper, 2010; Boghigian et al., 2010; Feist et al., 2010; Meadows et al., 2010). This inconsistency can be attributed to unfavorable metabolite accumulation and low capacity of biosynthesis pathways for flux amplification after removal of competitive pathways. For example, deletion of the acetate biosynthesis pathway did not improve the productivity of acetyl-CoA (an intermediate for many bioproducts), but rather resulted in an accumulation of pyruvate (Tomar et al., 2003). Furthermore, it has been shown that invariability of metabolic flux under mutagenic genotypes is an important feature in biological systems, and down-regulation of certain pathways may inhibit cell growth and cause imbalance in the energy metabolism. Such self-regulation of metabolism minimizes the performance of knockout-strains and makes the outcome from knockout strategies uncertain.

Discussion

The regression model presented in this article permits a priori calculation of the potential product yield from engineered *E. coli*. The model is developed through a mathematical and conceptual understanding of metabolic processes. Every variable, except for the primary and secondary number of metabolic steps, is categorized to account for the vast number of reported experimental conditions. The coefficients from the multiple linear regressions suggest several guidelines for improving final product yield, such as choosing the carbon substrate which has the shortest pathway to the final products, adding a proper nutrient source, improving cultivation modes, and utilizing pathways native to the host. Although application of these strategies can improve the conversion of carbon sources to product, reduce waste metabolites, and alleviate inhibitions due to toxic metabolites, such strategies have to be balanced with economical considerations. Furthermore, current genetic modification technologies may have unpredictable results when applied for the improvement of microbial performance. For example, knockout of competitive pathways (Bailey, 1991; Lee et al., 2005, 2007; Park et al., 2007; Trinh et al., 2006, 2008) or expression of heterologous genes (Bailey, 1991; Choi et al., 2010; Park et al., 2007; Zhang et al., 2007) are expected to shift metabolic fluxes to the desired product by removing a rate-limiting step or permitting the use of an alternative carbon source. However, the actual behavior of a designed biosynthetic pathway can vary significantly, especially because the optimal impact of a heterologous enzyme on production improvement is not guaranteed due to its potential "side effects" such as metabolic imbalance, cytotoxicity, or incompatibility with the host organism (Atsumi et al., 2008b). Therefore, if we are to improve current metabolic engineering strategies, it is important to develop systems biology in order to understand unpredictable metabolic regulations and control behavior.

Although this empirical model provides a prediction of product yield under different conditions, the model may only be approximate. First, the model cannot determine if the final product is toxic or if the imported non-native pathway is compatible to the host. For example, the yield of 1-butanol production is low (yield = 0.0007, 20-fold below the model predicted yield) for engineered *E. coli* with a synthetic pathway from *Clostridium* (Atsumi et al., 2008a). As a comparison, production of isobutanol (less toxic than 1-butanol) via a native keto-acid pathway shows a yield higher than the model prediction (Atsumi et al., 2008b). Second, the yield calculation does not include all carbon sources such as nutrient supplements and CO_2. For example, very high succinate yield (above 100%) from glucose was achieved because an anaplerotic pathway could fix CO_2 as a secondary carbon source (Sánchez et al., 2005) (Fig. 2). Furthermore, the category defining nutrient conditions is imprecise since most experiments involve the addition of nutrient-rich supplements (e.g., yeast extract) in varying amounts. A few studies involved the addition of intermediate compounds or precursors to improve biosynthesis fluxes. In order to reduce the potential complexity of the model, the variable NUT_{C2} categorizes all such nutrient conditions in a generalized manner (Table II). Therefore, the model is not able to specifically account for the contribution of special nutrients or the addition of precursors to the final yield. For example, Leonard et al. (2007) engineered pathways for flavonoid production using a minimal medium including both glucose and trace phenylpropanoic acids. The regression model predicts a calculated yield (average yield = 0.008) 2–3 times lower than the experimental yield (average yield = 0.020). Third, the model cannot predict the efficiency of a heterogeneous enzyme when employed in metabolic engineering experiments. For example, Hanai et al. (2007) incorporated two different alcohol dehydrogenases in *E. coli* strains for isopropanol production. Although the two enzymes encode for the same function, the *adh* from *Clostridium beijerinckii* produced twice as much isopropanol as the gene from *Thermoanaerobacter brockii*. Finally, other factors may also influence the model accuracy, including insufficient information to precisely calculate yield, complicated metabolic behavior and pathway regression, variation of experimental conditions and analytic methods, unequal research quality and merit between different articles, and simplification of model factors using categorical numbers.

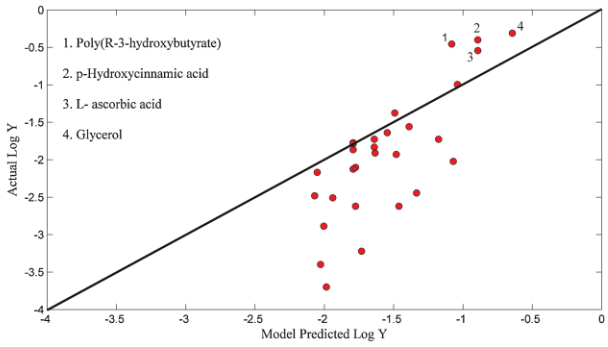


Figure 3. Distribution of actual $\log_{10} Y$ versus model generated $\log_{10} Y$ for *Saccharomyces cerevisiae*. The diagonal line drawn is one-to-one and passes through the origin. The numbered points are from the following articles: (1) (Carlson and Srienc, 2006); (2) (Vannelli et al., 2007); (3) (Sauer et al., 2004); and (4) (Overkamp et al., 2002). [Color figure can be seen in the online version of this article, available at http://wileyonlinelibrary.com/bit]

In an effort to examine and compare bio-production in another system, we use our model (Eq. 3, same coefficients) to analyze the yields of ~20 products achieved by *Saccharomyces cerevisiae* based on several recent articles. A table of the collected data for *S. cerevisiae* is available in the supplementary data (Supplementary Table II). Figure 3 shows that the product yields for *S. cerevisiae* are mostly below the predicted yield calculated from the *E. coli* model, especially for the biosynthesis of "low-yield compounds." Notably, several data points in Figure 3 are greater than those of model predictions due to unique biosynthesis methods. First, the production of *p*-hydroxycinnamic acid from glucose (20 g/L) was unusually high (Vannelli et al., 2007). This very high yield was achieved by supplementing 10 g/L phenylalanine to enhance product synthesis. Second, a study on ascorbic acid production from galactose incorporated a two-stage process: using glucose for biomass growth, then adding galactose for product synthesis in the stationary phase (Sauer et al., 2004). Such a cultivation process could improve the yield of ascorbic acid because small carbon fluxes are directed to biomass synthesis during the product production phase. Third, a study on poly[(R)-3-hydroxybutyrate] (PHB) synthesis supplemented the culture with pantothenate (a precursor of final product) to enhance PHB production (Carlson and Srienc, 2006). This indicates that addition of precursor or intermediate compounds can be significantly more efficient for promoting yield than addition of undefined nutrient sources (e.g., yeast extract) because it only requires a small number of enzymatic steps to convert the precursor or intermediate to the final product. Fourth, a study on glycerol production using glucose produced a higher yield than the model prediction (Overkamp et al., 2002). This can be attributed to the inclusion of ethanol along with glucose in chemostat cultivation.

Furthermore, we also compared chemical production from different microbial hosts (e.g., *S. cerevisiae*, *Schizosaccharomyces pombe*, *Bacillus subtilis*, *Corynebacterium glutamicum*, *Klebsiella pnumoniae*, *Clostridium diolis*, and *Clostridium acetobutylicum*) (Fig. 4). Without requiring extensive genetic modifications, some microorganisms can synthesize certain metabolic products using their native pathways (Alper and Stephanopoulos, 2009). Data points from these environmental microbes are distributed about the one-to-one line in the figure, showing their biosynthesis capability is comparable to that of *E. coli*. The model also demonstrates that other species may possess a significant disadvantage in the production of a particular product. For example, butanol synthesis by *S. cerevisiae* is 1–2 orders of magnitude lower than the model predictions because yeast shows an inability to utilize the engineered heterogeneous genes efficiently and a low tolerance for butanol (Fig. 4), while *C. acetobutylicum* has a very robust butanol production comparable to engineered *E. coli* strains (Atsumi et al., 2008a; Sillers et al., 2008; Steen et al., 2008).

Finally, the model can evaluate new methods that are being developed to increase final product yields. An example is protein scaffolding, which can avoid channeling intermediates away from the desired product and significantly improve the product yields (Dueber et al., 2009). By plotting the actual yield and the predicted yield (Supplementary Table III), the sample point (glucaric acid) lies above the one-to-one line and the sample point (mevalonate) lies close to the one-to-one line after scaffolding (Fig. 4). This implies that protein scaffolding can be useful for maximizing biosynthetic potential in *E. coli*. Another example of a new method is the application of both metabolic and protein engineering for levopimaradiene synthesis (Leonard et al., 2010). The experimental yield (~0.03) is fairly close to the

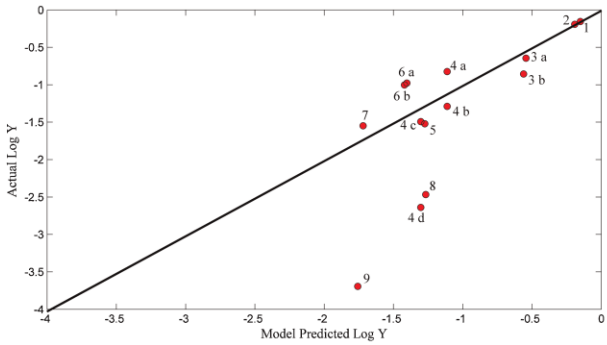


Figure 4. Distribution of actual $\log_{10} Y$ versus model generated $\log_{10} Y$ for various non-model species as well as protein scaffolding in *Escherichia coli*. The line drawn is one-to-one and passes through the origin. The numbers represents: (1) *Klebsiella pneumoniae*; 1,3-propanediol (Zhang et al., 2006), (2) *Clostridium diolis*; 1,3-propanediol (Otte et al., 2009), (3) *Clostridium acetobutylicum*; butanol without knockout (3a); butanol with knockout (3b) (Sillers et al., 2008), (4) *Escherichia coli*; glumaric acid with scaffold (4a); glumaric acid without scaffold (4b); mevalonate with scaffold (4c); mevalonate without scaffold (4d) (Dueber et al., 2009), (5) *Escherichia coli*; levopimaradiene (Leonard et al., 2010), (6) *Corynebacterium glutamicum*; isobutanol with knockout (6a); isobutanol without knockout (6b) (Smith et al., 2010), (7) *Bacillus subtilis*; riboflavin (Li et al., 2006), (8) *Schizosaccharomyces pombe*; vanillin (Hansen et al., 2009), and (9) *Saccharomyces cerevisiae*; butanol (Steen et al., 2008). [Color figure can be seen in the online version of this article, available at http://wileyonlinelibrary.com/bit]

predicted yield (~0.05), which indicates the effectiveness of their pathway engineering strategies.

Conclusions

The statistic-based model permits a priori estimation of final product yield of a metabolically produced compound based upon the genetic design of the microbial host system and the conditions under which the fermentation process occurs. This empirical model reflects the capability of current biotechnology to engineer *E. coli* to achieve the maximal biosynthetic yield. Such a model can provide information complementary to the conventional flux balance model which reflects the productivity of the "optimal" cellular metabolism. The model equation can be used to economically optimize the conditions for the production of a specific compound, providing its users with an estimate of the yield without performing actual experiments. The model can also be applied to other microbial chassis and used to compare their biosynthesis productivity to *E. coli*. The predictive power of our model, however, is limited by variations in experimental conditions and uncertainties in biological behaviors. For example, the model cannot evaluate the effect of precursor metabolites or toxic intermediates on final product yields. In general, this statistically based model provides a new method towards developing a comprehensive and quantitative understanding of experimental approaches for metabolic engineering purposes. The model can be widely useful to plan wet lab experiments across a wide range of applications in both scientific and industrial communities.

This study was supported by an NSF Career Grant (MCB0954016) and an international undergraduate research program at the Indian Institute of Technology (directed by Dr. Sudhir K. Jain). The authors are grateful to Jean Zhang in the Division of Biostatistics at Washington University Medical School for her beneficial discussions.

References

Alper H, Stephanopoulos G. 2009. Engineering for biofuels: Exploiting innate microbial capacity or importing biosynthetic potential? Nat Rev Microbiol 7(10):715–723.

Anderson DR, Sweeney DJ, Williams TA. 2009. Essentials of statistics for business and economics, 5th edn. Mason, OH: South-Western, Cengage Learning.

Atsumi S, Cann AF, Connor MR, Shen CR, Smith KM, Brynildsen MP, Chou KJ, Hanai T, Liao JC. 2008a. Metabolic engineering of *Escherichia coli* for 1-butanol production. Metab Eng 10(6):305–311.

Atsumi S, Hanai T, Liao JC. 2008b. Non-fermentative pathways for synthesis of branched-chain higher alcohols as biofuels. Nature 451(7174):86–89.

Bailey JE. 1991. Toward a science of metabolic engineering. Science 252(5013):1668–1675.

Blazeck J, Alper H. 2010. Systems metabolic engineering: Genome-scale models and beyond. Biotechnol J 5(7):647–659.

Boghigian BA, Seth G, Kiss R, Pfeifer BA. 2010. Metabolic flux analysis and pharmaceutical production. Metab Eng 12(2):81–95.

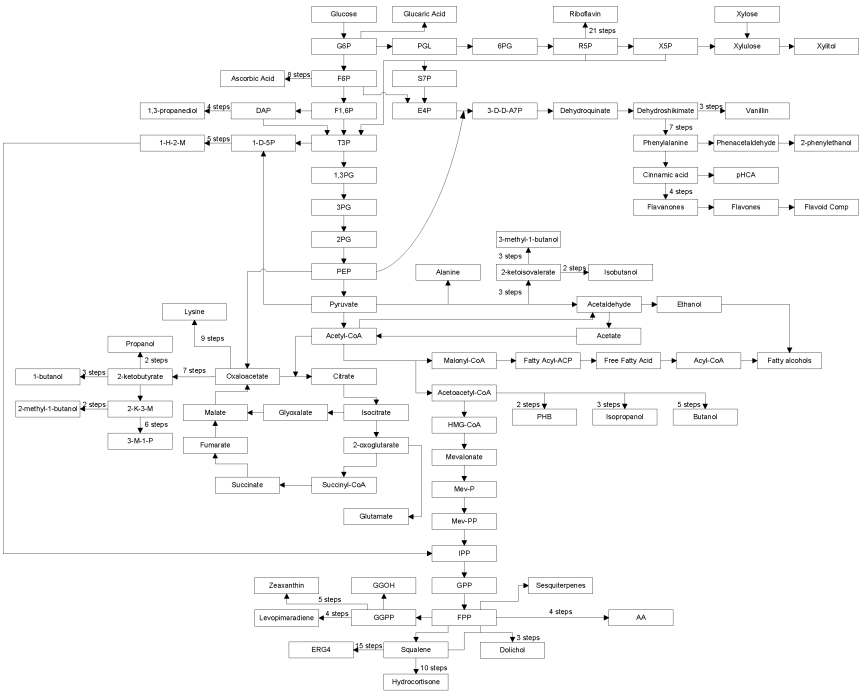
Carlson R, Srienc F. 2006. Effects of recombinant precursor pathway variations on poly[(R)-3-hydroxybutyrate] synthesis in *Saccharomyces cerevisiae*. J Biotechnol 124(3):561–573.

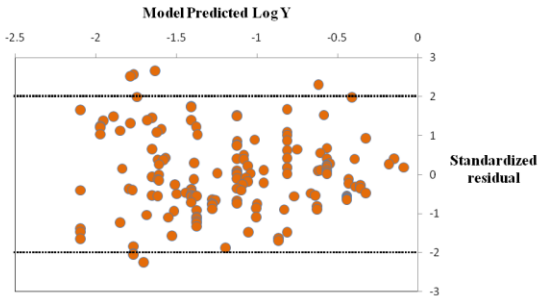
Choi HS, Lee SY, Kim TY, Woo HM. 2010. *In silico* identification of gene amplification targets for improvement of lycopene production. Appl Environ Microbiol 76(10):3097–3105.

Clark DS, Blanch HW. 1997. Biochemical engineering. New York: Marcel Dekker, Inc.
Clomburg JM, Gonzalez R. 2010. Biofuel production in *Escherichia coli*: The role of metabolic engineering and synthetic biology. Appl Microbiol Biotechnol 86(2):419–434.
du Prel JB, Hommel G, Rohrig B, Blettner M. 2009. Confidence interval or *p*-value?: Part 4 of a series on evaluation of scientific publications. Dtsch Arztebl Int 106(19):335–339.
Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, Prather KL, Keasling JD. 2009. Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol 27(8):753–759.
Feist AM, Zielinski DC, Orth JD, Schellenberger J, Herrgard MJ, Palsson BO. 2010. Model-driven evaluation of the production potential for growth-coupled products of *Escherichia coli*. Metab Eng 12(3):173–186.
Hanai T, Atsumi S, Liao JC. 2007. Engineered synthetic pathway for isopropanol production in *Escherichia coli*. Appl Environ Microbiol 73(24):7814–7818.
Hansen EH, Moller BL, Kock GR, Bunner CM, Kristensen C, Jensen OR, Okkels FT, Olsen CE, Motawia MS, Hansen J. 2009. De novo biosynthesis of vanillin in fission yeast (*Schizosaccharomyces pombe*) and baker's yeast (*Saccharomyces cerevisiae*). Appl Environ Microbiol 75(9):2765–2774.
Jarboe LR, Zhang X, Wang X, Moore JC, Shanmugam KT, Ingram LO. 2010. Metabolic engineering for production of biorenewable fuels and chemicals: Contributions of synthetic biology. J Biomed Biotechnol 2010: 761042.
Lee SJ, Lee DY, Kim TY, Kim BH, Lee J, Lee SY. 2005. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and *in silico* gene knockout simulation. Appl Environ Microbiol 71(12):7880–7887.
Lee KH, Park JH, Kim TY, Kim HU, Lee SY. 2007. Systems metabolic engineering of *Escherichia coli* for L-threonine production. Mol Syst Biol 3:149.
Leonard E, Lim KH, Saw PN, Koffas MA. 2007. Engineering central metabolic pathways for high-level flavonoid production in *Escherichia coli*. Appl Environ Microbiol 73(12):3877–3886.
Leonard E, Ajikumar PK, Thayer K, Xiao WH, Mo JD, Tidor B, Stephanopoulos G, Prather KL. 2010. Combining metabolic and protein engineering of a terpenoid biosynthetic pathway for overproduction and selectivity control. Proc Natl Acad Sci USA 107(31):13654–13659.
Li XJ, Chen T, Chen X, Zhao XM. 2006. Redirection electron flow to high coupling efficiency of terminal oxidase to enhance riboflavin biosynthesis. Appl Microbiol Biotechnol 73(2):374–383.
Meadows AL, Karnik R, Lam H, Forestell S, Snedecor B. 2010. Application of dynamic flux balance analysis to an industrial *Escherichia coli* fermentation. Metab Eng 12(2):150–160.
Montgomery DC, Peck EA, Vining GG. 2006. Introduction to linear regression analysis. New York: John Wiley & Sons.
Otte B, Grunwaldt E, Mahmoud O, Jennewein S. 2009. Genome shuffling in *Clostridium diolis* DSM 15410 for improved 1,3-propanediol production. Appl Environ Microbiol 75(24):7610–7616.
Overkamp KM, Bakker BM, Kotter P, Luttik MA, Van Dijken JP, Pronk JT. 2002. Metabolic engineering of glycerol production in *Saccharomyces cerevisiae*. Appl Environ Microbiol 68(6):2814–2821.
Park JH, Lee KH, Kim TY, Lee SY. 2007. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and *in silico* gene knockout simulation. Proc Natl Acad Sci USA 104(19):7797–7802.
Pelczar MJ, Jr., Chan ECS, Krieg NR. 1993. Microbiology: Concepts and applications. New York: McGraw-Hall Inc.
Sánchez AM, Bennett GN, San KY. 2005. Novel pathway engineering design of the anaerobic central metabolic pathway in *Escherichia coli* to increase succinate yield and productivity. Metab Eng 7(3):229–239.
Sánchez AM, Bennett GN, San KY. 2006. Batch culture characterization and metabolic flux analysis of succinate-producing *Escherichia coli* strains. Metab Eng 8(3):209–226.
Sauer M, Branduardi P, Valli M, Porro D. 2004. Production of L-ascorbic acid by metabolically engineered *Saccharomyces cerevisiae* and *Zygosaccharomyces bailii*. Appl Environ Microbiol 70(10):6086–6091.
Sillers R, Chow A, Tracy B, Papoutsakis ET. 2008. Metabolic engineering of the non-sporulating, non-solventogenic *Clostridium acetobutylicum* strain M5 to produce butanol without acetone demonstrate the robustness of the acid-formation pathways and the importance of the electron balance. Metab Eng 10(6):321–332.
Smith KM, Cho KM, Liao JC. 2010. Engineering *Corynebacterium glutamicum* for isobutanol production. Appl Microbiol Biotechnol 87(3): 1045–1055.
Steen EJ, Chan R, Prasad N, Myers S, Petzold CJ, Redding A, Ouellet M, Keasling JD. 2008. Metabolic engineering of *Saccharomyces cerevisiae* for the production of n-butanol. Microb Cell Fact 7:36.
Stephanopoulos GN, Aristidou AA, Nielsen J. 1998. Metabolic engineering: Principles and methodologies. San Diego: Academic Press.
Stephanopoulos G, Alper H, Moxley J. 2004. Exploiting biological complexity for strain improvement through systems biology. Nat Biotechnol 22(10):1261–1267.
Tang YJ, Qi L, Krieger-Brockett B. 2005. Evaluating factors that influence microbial phenanthrene biodegradation rates by regression with categorical variables. Chemosphere 59(5):729–741.
Team RDC. 2009. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna. Austria: ISBN 3-900051-07-0, URL. http://www.R-project.org.
Tomar A, Eiteman MA, Altman E. 2003. The effect of acetate pathway mutations on the production of pyruvate in *Escherichia coli*. Appl Microbiol Biotechnol 62(1):76–82.
Trinh CT, Carlson R, Wlaschin A, Srienc F. 2006. Design, construction and performance of the most efficient biomass producing *E. coli* bacterium. Metab Eng 8(6):628–638.
Trinh CT, Unrean P, Srienc F. 2008. Minimal *Escherichia coli* cell for the most efficient production of ethanol from hexoses and pentoses. Appl Environ Microbiol 74(12):3634–3643.
Vannelli T, Wei Qi W, Sweigard J, Gatenby AA, Sariaslani FS. 2007. Production of p-hydroxycinnamic acid from glucose in *Saccharomyces cerevisiae* and *Escherichia coli* by expression of heterologous genes from plants and fungi. Metab Eng 9(2):142–151.
Weisberg S. 1985. Applied linear regression. New York: John Wiley & Sons.
Wildermuth MC. 2000. Metabolic control analysis: Biological applications and insights. Genome Biol 1(6):reviews1031.
Zhang Y, Li Y, Du C, Liu M, Cao Z. 2006. Inactivation of aldehyde dehydrogenase: A key factor for engineering 1,3-propanediol production by *Klebsiella pneumoniae*. Metab Eng 8(6):578–586.
Zhang X, Jantama K, Moore JC, Shanmugam KT, Ingram LO. 2007. Production of L-alanine by metabolically engineed *Escherichia coli*. Appl Microbiol Biotechnol 77(2):355–366.

Supplementary Data

Supplementary Figure 1: A complete pathway map for biosynthesis (Abbreviations are in Supplementary Table 4).





Supplementary Figure 2: A plot between the square root of the standardized residuals and the fitted values for the regression model generated by R.

Supplementary Tables are in a separate spreadsheet (please download the file named "SupplementaryData_Spreadsheet_Final.xls" from the website:

http://users.seas.wustl.edu/tang/SupplementaryData_Spreadsheet_Final.xls):

Supplementary Table 1: A comprehensive list of data used for generating the regression model for *E. coli* using R.

Supplementary Table 2: A list of selected data used for comparing the model predicted $\log_{10}$ Y to the actual $\log_{10}$ Y by using the regression model (generated by *E. coli*) for *S. cerevisiae* yields.

Supplementary Table 3: A list of selected data used for comparing the model predicted log Y to the actual log Y by using the regression model for various environmental microorganisms.

Supplementary Table 4: A table denoting the nomenclature used in Supplementary Figure 1.

Appendix 12

Bacterial Responses to Cu-doped TiO_2 Nanoparticles

Science of the Total Environment 408 (2010) 1755–1758

Contents lists available at ScienceDirect

Science of the Total Environment

journal homepage: www.elsevier.com/locate/scitotenv

Bacterial responses to Cu-doped TiO_2 nanoparticles

Bing Wu, Rick Huang, Manoranjan Sahu, Xueyang Feng, Pratim Biswas *, Yinjie J. Tang *

Department of Energy, Environmental and Chemical Engineering, One Brookings Drive, Campus Box 1180, Washington University in St. Louis, St. Louis, MO 63130, United States

ARTICLE INFO

Article history:
Received 6 August 2009
Received in revised form 2 November 2009
Accepted 2 November 2009
Available online 20 November 2009

Keywords:
Mycobacterium smegmatis
Shewanella oneidensis
EDTA
EPS
Remediate

ABSTRACT

The toxicity of Cu-doped TiO_2 nanoparticles (NPs, 20 nm), synthesized by a flame aerosol reactor, to *Mycobacterium smegmatis* and *Shewanella oneidensis* MR-1, is the primary focus of this study. Both doped and non-doped TiO_2 NPs (20 nm) tended to agglomerate in the medium solution, and therefore did not penetrate into the cell and damage cellular structures. TiO_2 particles (<100 mg/L) did not apparently interfere with the growth of the two species in aqueous cultures. Cu-doped TiO_2 NPs (20 mg/L) significantly reduced the *M. smegmatis* growth rate by three fold, but did not affect *S. oneidensis* MR-1 growth. The toxicity of Cu-doped TiO_2 NPs was driven by the release of Cu^{2+} from the parent NPs. Compared to equivalent amounts of Cu^{2+}, Cu-doped TiO_2 NPs exhibited higher levels of toxicity to *M. smegmatis* (P-value<0.1). Addition of EDTA in the culture appeared to significantly decrease the anti-mycobacterium activity of Cu-doped TiO_2 NPs. *S. oneidensis* MR-1 produced a large amount of extracellular polymeric substances (EPS) under NP stress, especially extracellular protein. Therefore, *S. oneidensis* MR-1 was able to tolerate a much higher concentration of Cu^{2+} or Cu-doped TiO_2 NPs. *S. oneidensis* MR-1 also adsorbed NPs on cell surface and enzymatically reduced ionic copper in culture medium with a remediating rate of 61 μg/(liter·OD_{600}·hour) during its early exponential growth phase. Since the metal reducing *Shewanella* species can efficiently "clean" metal-oxide NPs, the activities of such environmentally relevant bacteria may be an important consideration for evaluating the ecological risk of metal-oxide NPs.

1. Introduction

Due to the significant applications of nanotechnology, the environmental and ecological effects of nanomaterials have to be considered. Evaluation of nanomaterials will not only help ensure the safety of nanotechnological applications, but also help design functional materials that have minimal adverse effects (Biswas and Wu, 2005; Handy et al., 2008; Klaine et al., 2008; Wiesner et al., 2006). Titanium dioxide (TiO_2) has been widely used in many fields (Biswas and Wu, 1998; Hoffmann et al., 1995; Honda and Fujishima, 1972; Sahle-Demessie et al., 2000). To enhance the functional properties and applicability of titanium dioxide, doped versions of TiO_2 are being synthesized to enhance catalytic activity for light harvesting applications (Colón et al., 2006; Namiki et al., 2005; Wang et al., 2008).

Many researchers have conducted studies to evaluate if nano-scale titanium dioxide would have biological impacts (Adams et al., 2006; Jiang et al., 2007, 2009; Oberdörster et al., 2005; Sayes et al., 2006; Warheit et al., 2007). In general, TiO_2 NPs have a relatively low toxicity, although the presence of light can generate reactive oxygen species (ROS) from TiO_2 NPs and thus enhance their antimicrobial activities. On the other hand, copper NPs appear to have higher cytotoxicity than copper ions because they may penetrate the cell membrane and release copper ions inside the cell (Karlsson et al., 2008; Yoon et al., 2007). However, it is still not clear whether there is synergistic effect when TiO_2 NPs are doped with CuO. Also, very few studies have examined the natural remediation of toxic metal NPs from the environment (Lewinski et al., 2008), which can be another important consideration of NPs' ecological impact. This study employed two model bacterial species: *Mycobacterium smegmatis*, a gram-positive bacterium and a model pathogenic strain for the study of *Tuberculosis*; and *Shewanella oneidensis* MR-1, a gram-negative environmentally relevant bacterium. The objectives of this study are: 1) to determine the toxicity of Cu-doped TiO_2 NPs; and 2) to investigate bacterial responses to NPs.

2. Materials and methods

A flame aerosol reactor (FLAR) with a three-port co-flow diffusion burner was used to synthesize Cu-doped TiO_2 NPs (20 nm) and TiO_2 NPs (20 nm) (Jiang et al., 2009). Titanium tetra-isopropoxide (TTIP, 97%, Aldrich-Sigma, USA) and copper (II) ethyl hexanoate (Aldrich-Sigma, USA) were used as the precursors for the synthesis of NPs. Copper (II) ethyl hexanoate was dissolved in xylene and atomized by a stainless steel nebulizer in the high temperature zone. Nitrogen gas at 1.5 liter per minute (lpm) was bubbled through TTIP in a bubbler maintained at 88 °C. Doping percentages were varied by feeding different molar ratios of the precursors into the high temperature

* Corresponding authors. Tang is to be contacted at Tel.: +1 314 935 3441. Biswas, Tel.: +1 314 935 5482.
E-mail addresses: Pratim.biswas@seas.wustl.edu (P. Biswas), yinjie.tang@seas.wustl.edu (Y.J. Tang).

0048-9697/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.scitotenv.2009.11.004

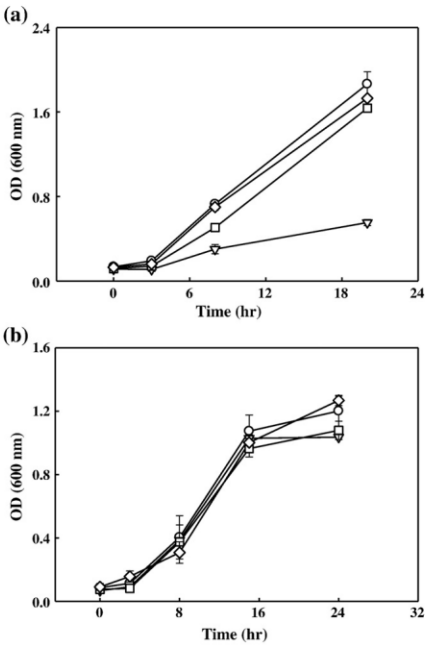


Fig. 1. Responses of *M. smegmatis* (a) and *S. oneidensis* MR-1 (b) to 0.02 g/L NPs ($n=2$): (○) control; (▽) 1.8%-Cu-doped TiO_2 NPs; (□) 0.6%-Cu-doped TiO_2 NPs; (◇) TiO_2 NPs.

combustion zone of the diffusion flame. Oxygen and methane at 7 lpm, and 1.6 lpm, respectively, were introduced into the FLAR system. The temperature and residence time of the NPs in the combustion zone were controlled to obtain the desired size and composition of the NPs. For antibacterial tests, 1.8%-Cu-doped TiO_2 (20 nm, containing 1.8% Cu by weight), 0.6%-Cu-doped TiO_2 (20 nm, containing 0.6% Cu by weight), TiO_2, and CuO NPs (40 nm, Aldrich-Sigma, USA) were suspended in sterilized DI water to make the stock solutions (1 g/L). To reduce agglomeration, the solution was sonicated for 2 min (pulse 20 s on and 20 s off) using a sonicator (Misonix S-4000, USA). The zeta potential of the NPs in water suspensions was measured by a Zeta-sizer (Nanoseries ZS, Malvern, UK).

Two types of bacteria were grown in the presence of NPs: *M. smegmatis* was grown in modified Sauton liquid medium at 37 °C (Tang et al., 2009), and *S. oneidensis* MR-1 was grown in minimal MR-1 medium at 30 °C (Tang et al., 2007). The initial cell density in both cultures was equal to $OD_{600}\sim0.1$, and the volume of each culture was 5 mL (placed in a 50 mL falcon tube). All culture tubes were shaken at a speed of 200 rpm in the dark. Cell density was monitored by a UV spectrometer (Genesys, Thermo Scientific, USA) at a wavelength of 600 nm. To monitor microbial remediation of Cu-doped TiO_2 NPs, the cultures with Cu-doped TiO_2 NPs were filtered using a 0.22 μm membrane filter (Nylon, Millipore, USA) after centrifugation at 19,000 $\times g$ (10 min) to remove cells and large agglomerated NPs. The cell-free supernatant was treated with 2% HNO_3 and 0.5% HCl, and then dissolved copper concentration was determined by Inductively Coupled Plasma Mass Spectrometry (ICP-MS, Agilent, USA). The interaction between NPs and cells in the culture was observed by a scanning electron microscope (SEM) and a transmission electron microscope (TEM). All samples for imaging were prepared by the Histology & Microscopy Core Facility of the Washington University School of Medicine.

The effect of NPs on enzyme activity (cell-free) was tested based on NADPH production rates from glucose in the presence of the enzymes hexokinase (HK) and glucose-6-phosphate dehydrogenase (G6PD). The two enzymes convert sugar, ATP, and cofactor NADP to gluconate-6-phosphate, ADP, and NADPH (all chemicals were obtained from R-Biopharm, Germany). In the test, samples with known amount of NPs (or $CuCl_2$) were added to the cell-free assay solutions (containing a standard amount of NADP, ATP, the two enzymes, and 0.5 g/L glucose). The reaction was carried out in 2 mL cuvettes. NADPH production rate was calculated based on the change of the UV absorbance at a wavelength of 340 nm. Since NADPH is essential for energy metabolism, measurement of the rate of NADPH production provides a simple approach to test the effect of NPs or ionic copper on general enzyme activities. In addition, the extracellular polymeric substances (EPS) were measured under NP stress. Polysaccharides and proteins were dominant components in the EPS. The polysaccharide concentration was determined according to the phenol-sulfuric acid method using glucose as a standard (Dubois et al., 1956). The protein concentration was determined by Bradford reagent using bovine serum albumin (BSA) as a standard (Bradford, 1976).

3. Results and discussion

Antibacterial properties of the four kinds of NPs (1.8%-Cu-doped TiO_2, 0.6%-Cu-doped TiO_2, TiO_2, and CuO) were tested using *M. smegmatis* or *S. oneidensis* MR-1 culture. When a low concentration of NPs (0.005 g/L) was added, 1.8%-Cu-doped TiO_2, 0.6%-Cu-doped TiO_2 and TiO_2 NPs had no apparent effect on microbial growth (data not shown). With the addition of 0.02 g/L NPs, the growth of *M. smegmatis* was significantly inhibited by 1.8%-Cu-TiO_2 NPs (0.02 g/L NPs contained 0.36 mg/L total copper), and moderately negatively-affected by 0.6%-Cu-doped TiO_2 NPs (0.02 g/L NPs contained 0.12 mg/L total copper) (Fig. 1). *S. oneidensis* MR-1 growth was only slightly affected by 1.8%- and 0.6%-Cu-doped TiO_2 NPs (0.02 g/L), i.e., final cell density was reduced by 10–20%. TiO_2 NPs did not show any effect on

Table 1
Percentage of bacterial growth inhibition after exposure to NPs.

	TiO_2 (20 mg/L)	0.6%-Cu-TiO_2 NPs(20 mg/L)	1.8%-Cu-TiO_2 NPs (20 mg/L)	CuO (20 mg/L)	Cu^{2+} (0.36 mg/L)	TiO_2 (20 mg/L) + Cu^{2+} (0.36 mg/L)
S. oneidensis ($n=4$)	7±6	8±2	15±8	13±4	11±2	7±5
M. smegmatis ($n=4$)	5±7	19±3	55±10	98±5	47±4	51±7
Enzyme[a] ($n=2$)	1±1	22±4	24±5	16±4	4±5	4±5

Inhibition rate (%) $=(1-OD_{600\ (sample)}/OD_{600\ (control)})\times100\%$; the OD_{600} was measured during the middle exponential growth phase (9–10 h).
[a] Enzymatic assay was conducted in the presence of NPs:

$$Inhibition\ rate\,(\%)=\left[1-\frac{NADPH\ production\ rate\ (\text{with NPs})}{NADPH\ production\ rate\ (\text{control})}\right]\times100\%.$$

M. smegmatis (or *S. oneidensis* MR-1) growth even at higher concentrations (up to 0.1 g/L). CuO NPs (0.02 g/L) fully inhibited *M. smegmatis* growth, while *S. oneidensis* MR-1 displayed the similar growth rate as that without CuO NPs (Table 1). The SEM and TEM images (Fig. 2) show that NPs were not inside the *M. smegmatis* or *S. oneidensis* MR-1 cells and there was no apparent physical damage to cellular structures. The TEM images indicate the agglomeration of the NPs in the aqueous culture, which may decrease the capability of NPs to penetrate cellular structure. A similar observation was also reported by Franklin et al. (2007): nano-sized ZnO particles agglomerated and formed micro-scale particles in aqueous solution, so they showed no size-related toxic effect.

The antibacterial property of Cu-doped TiO_2 NPs was apparently associated with copper and the antibacterial effect increased with the copper content in NPs. To reveal ionic copper stresses to bacterial strains, *M. smegmatis* and *S. oneidensis* MR-1 were both exposed to Cu^{2+} (0.36 mg/L). Table 1 indicates that the percentage of inhibition to *M. smegmatis* growth was around 47%. *S. oneidensis* MR-1 had much higher tolerance to copper ions and the percentage of inhibition was only about 11%. Cu-doped TiO_2 NPs also slightly interfered with enzyme activities. NADPH production rate in a cell-free solution (with HK and G6PD) were decreased by ~20% in the presence of Cu-doped TiO_2 NPs, while TiO_2 NPs or Cu^{2+} (0.36 mg/L) had a weak influence on enzyme activity (Table 1).

Our observation indicated that Cu^{2+} and NPs might have a synergistic effect on *M. smegmatis*' growth. Our experiments showed that all TiO_2 based NPs were negatively charged (indicated by negative zeta potential: −9.79 to −7.73 mV), and its surface could adsorb positively charged Cu^{2+}, and thus Cu^{2+} concentration along the NP surface might increase: TiO_2 NPs served as a dissolved ionic copper carrier/concentrator (Lundqvist et al., 2008; Navarro et al., 2008). Therefore, the presence of low-toxic TiO_2 NPs could enhance Cu^{2+} toxicity to *M. smegmatis*. With the equal amount of total copper, the statistical analysis of such synergistic toxicity showed: the percentage of inhibition increased from 47 ± 4% (under stress by 0.36 mg/L Cu^{2+}) to 55±10% (under stress by 20 mg/L 1.8%-Cu-doped-TiO_2 NP) (the *P*-values were about 0.06 by examining a Student *T*-test respectively, which reveals that this difference was statistically significant when the confidence interval is 90%). The combined effect from NPs and toxic ions had also been reported by Navarro et al. (2008), where they observed that toxicity of ionic silver to *Chlamydomonas reinhardtii* was enhanced in the presence of NPs.

The antibacterial level of Cu-doped TiO_2 NPs can be alleviated by higher cell density or the presence of chelating agents in the culture (Fig. 3), which reduced the interaction of NPs with the individual cells. For example, addition of nutrients (Middlebrook 7H9 nutrients) (Tang et al., 2009) into the culture could promote cellular reproduction and recover the growth of *M. smegmatis* in the presence of Cu-doped TiO_2 NPs. Furthermore, EDTA is a chelating agent that sequesters di- and trivalent metal ions (Iijima et al., 2007). After addition of 0.02 g/L EDTA to a *M. smegmatis* culture solution, cells were able to maintain normal growth under the stress of Cu-doped TiO_2 NPs.

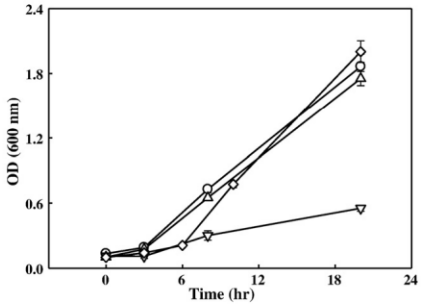


Fig. 3. The effect of EDTA and nutrient addition on antibacterial properties of NPs (*M. smegmatis*) ($n=2$); (○) control (without NPs and EDTA); (▽) 1.8%-Cu-doped TiO_2 NPs (0.02 g/L); (◇) 1.8%-Cu-doped TiO_2 NPs (0.02 g/L with EDTA); (△) 1.8%-Cu-doped TiO_2 NPs (0.02 g/L with Middlebrook 7H9 nutrients).

S. oneidensis MR-1 has the potential to remediate toxic metal NPs. When *S. oneidensis* MR-1 grew in the media containing 1.8%-Cu-doped TiO_2 NPs (0.02 g/L), the total "free" copper remaining in the solution after centrifugation and micro-filtration of cell culture was measured by ICP-MS at different time points (0, 3, 6 and 20 h). The free copper concentrations (mainly Cu^{2+} and Cu-doped TiO_2 NPs not adsorbed on bacterial surface) in the filtered MR-1 culture solution continuously decreased over time (Table 2). In contrast, *M. smegmatis* had much less NP removing activity. During the exponential growth period from 3 to 6 h, the average rate for *S. oneidensis* MR-1 to remove the total "free" copper from the culture medium was ~61 μg/(liter·OD_{600}·hour). During the 20 h growth period, ~70% of total ionic copper was removed from culture medium. This was due to two unique cellular properties of *S. oneidensis* MR-1: 1) it adsorbed NPs on the cell surface and immobilized free NPs in the solution (Fig. 2b); and 2) it could convert ionic copper to insoluble metal copper via its versatile respiratory metabolisms (Gorby et al., 2006; Tang et al., 2007; Tiedje, 2002; Toes et al., 2008). In addition, *S. oneidensis* MR-1 had a capability to produce more EPS, especially more extracellular protein than those of *M. smegmatis*. For example, in the early exponential phase ($t=3$ h), *S. oneidensis* MR-1 contained 43.2 mg/L extracellular protein (Table 2), which may act as a

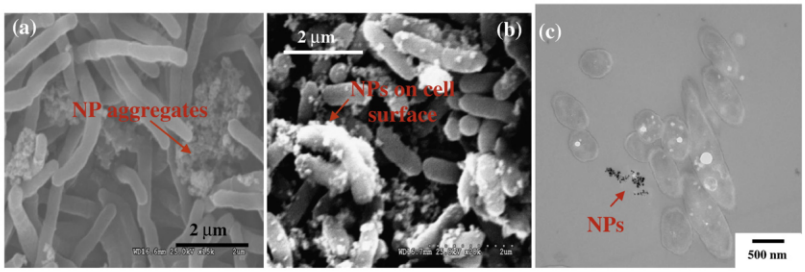


Fig. 2. SEM and TEM images of cells with 1.8%-Cu-doped TiO_2 NPs: (a) SEM of *M. smegmatis*; (b) SEM of *S. oneidensis* MR-1; (c) TEM of *M. smegmatis*.

Table 2
"Free" copper reduction and EPS production ($n=2$).

1.8%-Cu-TiO_2 NPs (~20 mg/L)		*M. smegmatis*	*S. oneidensis*
Total copper in initial cell-free medium (μg/L)		**320 ± 35**	**361 ± 18**
0 h	OD_{600}	0.13 ± 0.01	0.19 ± 0.03
	Free [copper] (μg/L)	**290 ± 14**	**317 ± 16**
	Protein (mg/L)	<2	25.2 ± 1.3
	Polysaccharides (mg/L)	11.7 ± 0.6	15.7 ± 0.8
3 h	OD_{600}	0.16 ± 0.01	0.27 ± 0.01
	Free [copper] (μg/L)	**260 ± 22**	**268 ± 4**
	Protein (mg/L)	<2	43.2 ± 2.2
	Polysaccharides (mg/L)	14.2 ± 0.7	18.0 ± 0.9
6 h	OD_{600}	0.24 ± 0.01	0.35 ± 0.03
	Free [copper] (μg/L)	**262 ± 13**	**232 ± 29**
	Protein (mg/L)	<2	50.4 ± 2.5
	Polysaccharides (mg/L)	22.6 ± 1.1	18.4 ± 0.9
20 h	OD_{600}	Not measured	0.80 ± 0.05
	Free [copper] (μg/L)		**113 ± 15**
Free copper reduction rate	Δ [copper]/($OD_{600\ ave}$ × time) = μg/(liter·OD_{600}·hour)	~0[a]	~61[a]

The "free" copper refers to the total copper in the culture solutions (including copper doped in NPs and dissolved Cu^{2+}) which could not be removed by centrifugation and micro-size filtration. The data in bold format describe the free copper concentrations.
[a] Rate was calculated based on the data collecting at two sampling time points: $t=3$ h and $t=6$ h.

protective barrier to enhance their resistance to Cu-doped TiO_2 NPs (Teitzel and Parsek, 2003). However, Gram-positive *M. smegmatis* did not produce much extracellular protein, and it appeared to be more sensitive to Cu-TiO_2 NP stress.

In summary, *S. oneidensis* MR-1 displayed an excellent resistance to metal NPs. Such metal reducing bacteria are widely distributed in the aquatic environments. Natural bioremediation of toxic metal NPs may significantly alleviate the NPs' impact on the ecological system. The behaviors of these environmentally relevant bacteria may be an important consideration for evaluating the risk of nanomaterials in the environment.

Acknowledgements

This study was supported by MAGEEP at Washington University, St. Louis Institute of Nanomedicine Pilot Grant Program and the DOD-MURI Grant (FA9550-04-1-0430). The authors thank Patricia Wurm in the Department of Energy, Environmental and Chemical Engineering and Howard Wynder in the Histology & Microscopy Core Facility in Washington University for helping with the experiments.

References

Adams LK, Lyon DY, Alvarez PJJ. Comparative toxicity of nano-scale TiO_2, SiO_2 and ZnO water suspensions. Wat Res 2006;40:3527–32.
Biswas P, Wu CY. Control of toxic metal emissions from combustors using sorbents: a review. J Air Waste Manage Assoc 1998;48:113–27.
Biswas P, Wu CY. Nanoparticles and the environment. J Air Waste Manage Assoc 2005;55:708–46.
Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem 1976;72:248–54.
Colón G, Maicu M, Hidalgo MC, Navio JA. Cu-doped TiO_2 systems with improved photocatalytic activity. Appl Catal B Environ 2006;67:41–51.
Dubois M, Gilles KA, Hamilton JK, Rebers PA, Smith F. Colorimetric method for determination of sugars and related substances. Anal Chem 1956;28:350–6.
Franklin NM, Rogers NJ, Apte SC, Batley GE, Gadd GE, Casey PS. Comparative toxicity of nanoparticulate ZnO, Bulk ZnO, and $ZnCl_2$ to a freshwater microalga (*Pseudokirchneriella subcapitata*): the importance of particle solubility. Environ Sci Technol 2007;41:8484–90.
Gorby YA, Yanina S, Malean JS, Rosso KM, Moyles D, Dohnalkova A, et al. Electrically conductive bacterial nanowires produced by *Shewanella oneidensis* strain MR-1 and other microorganisms. PNAS 2006;103(30):11358–63.
Handy RD, von der Kammer F, Lead JR, Hassello M, Owen R, Mark C. The ecotoxicology and chemistry of manufactured nanoparticles. Ecotoxicology 2008:287–314.
Hoffmann MR, Martin ST, Choi W, Bahnemann DW. Environmental applications of semiconductor photocatalysis. Chem Rev 1995;95:69–96.
Honda K, Fujishima A. Photolysis-decomposition of water at surface of an irradiated semiconduction. Nature 1972;238:37.
Iijima M, Sato N, Tsukada M, Kamiya H. Dispersion behavior of barium titanate nanoparticles prepared by using various polycarboxylic dispersants. J Am Ceram Soc 2007;90:2741–6.
Jiang J, Chen DR, Biswas P. Synthesis of nanoparticles in a flame aerosol reactor with independent and strict control of their size, crystal phase, and morphology. Nanotechnology 2007;18:285603.
Jiang J, Oberdörster G, Biswas P. Characterization of size, surface charge, and agglomeration state of nanoparticle dispersions for toxicological studies. J Nanopart Res 2009;11:77–89.
Karlsson H, Cronholm P, Gustafsson J, Möller L. Copper oxide nanoparticles are highly toxic: a comparison between metal oxide nanoparticles and carbon nanotubes. Chem Res Toxicol 2008;21:1726–32.
Klaine SJ, Alvarez PJJ, Batley GE, Fernandes TF, Handy RD, Lyon DY, et al. Nanomaterials in the environment: behavior, fate, bioavailability, and effects. Environ Toxicol Chem 2008;27:1825–51.
Lewinski N, Colvin V, Drezek R. Cytotoxicity of nanoparticles. Small 2008;4:26–49.
Lundqvist M, Stigler J, Elia G, Lynch I, Cedervall T, Dawson KA. Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts. PNAS 2008;105:14265–70.
Namiki N, Cho K, Fraundorf P, Biswas P. Tubular reactor synthesis of doped nanostructured titanium dioxide and its enhanced activation by coronas and soft X-rays. Ind Eng Chem Res 2005;44:5213–20.
Navarro E, Piccapietra F, Wagner B, Marconi F, Kaegi R, Odzak N, et al. Toxicity of silver nanoparticles to *Chlamydonas reinhardtii*. Environ Sci Technol 2008;42:8959–64.
Oberdörster G, Oberdörster E. Nanotoxicology: an emerging discipline evolving from studies of ultrafine particles. Environ Health Persp 2005;113:823–39.
Sahle-Demessie E, Gonzalez MA, Enriquez J, Zhao Q. Selective oxidation in supercritical carbon dioxide using clean oxidants. Ind Eng Chem Res 2000;39:4858–64.
Sayes CM, Wahi R, Kurian PA, Liu YP, West JL, Ausman KD. Correlating nanoscale titania structure with toxicity: a cytotoxicity and inflammatory response study with human dermal fibroblasts and human lung epithelial cells. Toxicol Sci 2006;92:174–85.
Tang Y, Shui W, Myers S, Feng X, Bertozzi C, Keasling J. Isotopomer analysis of both free metabolites and proteinogenic amino acids to investigate aerobic metabolism and hypoxic response of *Mycobacterium smegmatis*. Biotechnol Lett 2009;31:1233–40.
Tang YJ, Meadows AL, Keasling JD. A kinetic model describing *Shewanella oneidensis* MR-1 growth, substrate consumption, and product secretion. Biotechnol Bioeng 2007;96:125–33.
Teitzel GM, Parsek MR. Heavy metal resistance of biofilm and planktonic *Pseudomonas aeruginosa*. Appl Environ Microbiol 2003;69:2313–20.
Tiedje JM. *Shewanella*—the environmentally versatile genome. Nat Biotechnol 2002;20:1093–4.
Toes ACM, Daleke MH, Kuenen JG, Muyzer G. Expression of copA and cusA in *Shewanella* during copper stress. Microbiology 2008;154:2709–18.
Wang S, Lu W, Tovmachenko O, Rai US, Yu H, Ray PC. Challenge in understanding size and shape dependent toxicity of gold nanomaterials in human skin keratinocytes. Chem Phys Lett 2008;463:145–9.
Warheit DB, Hoke RA, Finlay C, Donner EM, Reed KL, Sayes CM. Development of a base set of toxicity tests using ultrafine TiO_2 particles as component of nanoparticle risk management. Toxicol Lett 2007;171:99–110.
Wiesner MR, Lowry GV, Alvarez P, Dionysiou D, Biswas P. Assessing the risks of manufactured nanomaterials. Environ Sci Technol 2006;40:4336–45.
Yoon K-Y, Byeon JH, Park J-H, Hwang J. Susceptibility constants of *Escherichia coli* and *Bacillus subtilis* to silver and copper nanoparticles. Sci Total Environ 2007;373:572–5.