



FLUXOMICS AND METABOLIC ANALYSIS IN SYSTEMS MICROBIOLOGY

EDITED BY: Wei Xiong, Yinjie Tang and Lars Keld Nielsen
PUBLISHED IN: Frontiers in Microbiology



frontiers

Frontiers Copyright Statement

© Copyright 2007-2019 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88963-069-1

DOI 10.3389/978-2-88963-069-1

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

FLUXOMICS AND METABOLIC ANALYSIS IN SYSTEMS MICROBIOLOGY

Topic Editors:

Wei Xiong, National Renewable Energy Laboratory (DOE), United States

Yinjie Tang, Washington University in St. Louis, United States

Lars Keld Nielsen, University of Queensland, Australia

Citation: Xiong, W., Tang, Y., Nielsen, L. K., eds. (2019). Fluxomics and Metabolic Analysis in Systems Microbiology. Lausanne: Frontiers Media.
doi: 10.3389/978-2-88963-069-1

Table of Contents

- 04** *Membrane-Inlet Mass Spectrometry Enables a Quantitative Understanding of Inorganic Carbon Uptake Flux and Carbon Concentrating Mechanisms in Metabolically Engineered Cyanobacteria*
Damien Douchi, Feiyan Liang, Melissa Cano, Wei Xiong, Bo Wang, Pin-Ching Maness, Peter Lindblad and Jianping Yu
- 17** *The Design of FluxML: A Universal Modeling Language for ¹³C Metabolic Flux Analysis*
Martin Beyß, Salah Azzouzi, Michael Weitzel, Wolfgang Wiechert and Katharina Nöh
- 41** *¹³C-Metabolic Flux Analysis Reveals Effect of Phenol on Central Carbon Metabolism in Escherichia coli*
Sayaka Kitamura, Yoshihiro Toya and Hiroshi Shimizu
- 49** *EMUlator: An Elementary Metabolite Unit (EMU) Based Isotope Simulator Enabled by Adjacency Matrix*
Chao Wu, Chia-hsin Chen, Jonathan Lo, William Michener, PinChing Maness and Wei Xiong
- 61** *Dietary Energy Level Promotes Rumen Microbial Protein Synthesis by Improving the Energy Productivity of the Ruminal Microbiome*
Zhongyan Lu, Zhihui Xu, Zanming Shen, Yuanchun Tian and Hong Shen
- 75** *Approaches to Computational Strain Design in the Multiomics Era*
Peter C. St. John and Yannick J. Bomble
- 82** *¹³C-Metabolic Flux Analysis Reveals the Metabolic Flux Redistribution for Enhanced Production of Poly-γ-Glutamic Acid in dlt Over-Expressed Bacillus licheniformis*
Penghui He, Ni Wan, Dongbo Cai, Shiyong Hu, Yaozhong Chen, Shunyi Li and Shouwen Chen
- 93** *Tandem Mass Spectrometry for ¹³C Metabolic Flux Analysis: Methods and Algorithms Based on EMU Framework*
Jungik Choi and Maciek R. Antoniewicz
- 101** *Flux Connections Between Gluconate Pathway, Glycolysis, and Pentose–Phosphate Pathway During Carbohydrate Metabolism in Bacillus megaterium QM B1551*
Julie A. Wushensky, Tracy Youngster, Carroll M. Mendonca and Ludmilla Aristilde
- 114** *Bacterial Metabolism During Biofilm Growth Investigated by ¹³C Tracing*
Ni Wan, Hao Wang, Chun Kiat Ng, Manisha Mukherjee, Dacheng Ren, Bin Cao and Yinjie J. Tang
- 123** *Isotope-Assisted Metabolite Analysis Sheds Light on Central Carbon Metabolism of a Model Cellulolytic Bacterium Clostridium thermocellum*
Wei Xiong, Jonathan Lo, Katherine J. Chou, Chao Wu, Lauren Magnusson, Tao Dong and PinChing Maness



Membrane-Inlet Mass Spectrometry Enables a Quantitative Understanding of Inorganic Carbon Uptake Flux and Carbon Concentrating Mechanisms in Metabolically Engineered Cyanobacteria

Damien Douchi¹, Feiyan Liang², Melissa Cano¹, Wei Xiong¹, Bo Wang¹, Pin-Ching Maness¹, Peter Lindblad² and Jianping Yu^{1*}

¹Biosciences Center, National Renewable Energy Laboratory, Golden, CO, United States, ²Microbial Chemistry, Department of Chemistry-Ångström, Uppsala University, Uppsala, Sweden

OPEN ACCESS

Edited by:

Marc Strous,
University of Calgary, Canada

Reviewed by:

Kathleen Scott,
University of South Florida,
United States
Xuefeng Lu,
Qingdao Institute of Bioenergy and
Bioprocess Technology (CAS), China

*Correspondence:

Jianping Yu
jianping.yu@nrel.gov

Specialty section:

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

Received: 31 October 2018

Accepted: 31 May 2019

Published: 25 June 2019

Citation:

Douchi D, Liang F, Cano M, Xiong W,
Wang B, Maness P-C, Lindblad P
and Yu J (2019) Membrane-Inlet
Mass Spectrometry Enables a
Quantitative Understanding of
Inorganic Carbon Uptake
Flux and Carbon Concentrating
Mechanisms in Metabolically
Engineered Cyanobacteria.
Front. Microbiol. 10:1356.
doi: 10.3389/fmicb.2019.01356

Photosynthesis uses solar energy to drive inorganic carbon (Ci) uptake, fixation, and biomass formation. In cyanobacteria, Ci uptake is assisted by carbon concentrating mechanisms (CCM), and CO₂ fixation is catalyzed by RubisCO in the Calvin-Benson-Bassham (CBB) cycle. Understanding the regulation that governs CCM and CBB cycle activities in natural and engineered strains requires methods and parameters that quantify these activities. Here, we used membrane-inlet mass spectrometry (MIMS) to simultaneously quantify Ci concentrating and fixation processes in the cyanobacterium *Synechocystis* 6803. By comparing cultures acclimated to ambient air conditions to cultures transitioning to high Ci conditions, we show that acclimation to high Ci involves a concurrent decline of Ci uptake and fixation parameters. By varying light input, we show that both CCM and CBB reactions become energy limited under low light conditions. A strain over-expressing the gene for the CBB cycle enzyme fructose-bisphosphate aldolase showed higher CCM and carbon fixation capabilities, suggesting a regulatory link between CBB metabolites and CCM capacity. While the engineering of an ethanol production pathway had no effect on CCM or carbon fixation parameters, additional fructose-bisphosphate aldolase gene over-expression enhanced both activities while simultaneously increasing ethanol productivity. These observations show that MIMS can be a useful tool to study the extracellular Ci flux and how CBB metabolites regulate Ci uptake and fixation.

Keywords: MIMS, carbon uptake rate, cyanobacteria, FbaA, carbon fixation

INTRODUCTION

Photosynthesis has been responsible for decreasing CO₂ in the atmosphere from 30–35% to 0.04% over the last 3 billion years or so (Blankenship, 2010) and is the major physicochemical process that generates organic molecules and, as such, supports most of the life on Earth. The main enzyme responsible for this CO₂ fixation, ribulose-1,5-bisphosphate carboxylase/oxygenase

(RubisCO) appeared 2.5 billion years ago and has evolved to adapt to the decreasing CO₂ concentrations and concurrently increasing O₂ levels (Whitney et al., 2011). CO₂ and O₂ are competitive substrates for RubisCO, leading to either CO₂ fixation (carboxylation) or photorespiration (oxygenation). As a result, photosynthetic organisms have evolved various strategies to favor carboxylation over oxygenation by increasing carbon availability for RubisCO.

In cyanobacteria, the inorganic carbon (Ci) is concentrated in the cytoplasm to levels over 100 times the external concentration in air-saturated condition (Woodger et al., 2005) *via* carbon concentrating mechanisms (CCM; **Figure 1**). At least five different uptake proteins or complexes are involved in this process, each with different affinities for Ci. Among them are the BicA and SbtA, sodium/bicarbonate symporters, powered by a sodium gradient across the plasma membrane. NdhD3 and D4 were proposed to be involved in regenerating this sodium gradient, powered by NADPH or Ferredoxin (Wang et al., 2004; Woodger et al., 2007). The involvement of plasma membrane sodium/proton antiporters and ATPase was also hypothesized (Kamennaya et al., 2015). Another bicarbonate transporter is the Bct1 complex, which has its own ATPase activity. In addition, the CO₂ uptake systems NDH1-3 and NDH1-4 directly convert the CO₂ to bicarbonate in the cytoplasm using energy from photosynthetic or respiratory thylakoid electron flow (Battchikova et al., 2011; Artier et al., 2018), locking incoming CO₂ which diffuses freely across the membrane and limiting Ci leakages.

RubisCO is confined to a bacterial micro-compartment, the carboxysome, along with carbonic anhydrase enzymes, which converts HCO₃⁻ into CO₂. Together the membrane transporters, cytoplasm and carboxysome carbonic anhydrases

form a Ci conduit from the medium to the carboxysome, so that the concentration of CO₂ within the carboxysome is up to 4,000-fold higher than it is externally (Sültemeyer et al., 1995; Price et al., 1998; Kaplan and Reinhold, 1999; Woodger et al., 2005). RubisCO converts one molecule of ribulose bis-phosphate (RuBP) and one molecule of CO₂ to two molecules of 3-phosphoglycerate (3PG), which enter the Calvin-Benson-Bassham (CBB) cycle. While RuBP is regenerated, other intermediates are formed and connect to the central carbon metabolism and anabolic pathways, feeding the production of cellular constituents.

An in-depth understanding of photosynthetic mechanism, such as the CCM, requires a more comprehensive, systems-level approach to measure photosynthetic carbon flux within a biochemical network. With respect to photosynthetic organisms, the intracellular carbon flux can be determined through a new ¹³CO₂/NaH¹³CO₃ labeling approach: isotopically nonstationary metabolic flux analysis (INST-MFA) (Adebiyi et al., 2015). This method allows the estimation of relative photosynthesis and photorespiration fluxes yielding sugar phosphates, organic acids, and other intracellular metabolites in a model phototroph such as *Synechocystis* sp. PCC 6803 (hereafter *Synechocystis*). To obtain an absolute quantification of fluxomic values, an isotope tracer experiment must be coupled to fundamental determinants of *in vivo* cell physiology. For example, measurements of cell specific rates of nutrient uptake and product formation (i.e., normalized to cell density) allow for intracellular flux calculations using INST-MFA. The measured Ci fixation kinetics is a key input to these methods, because they constrain the solution space of feasible intracellular fluxes. Therefore, an accurate estimation of Ci fixation kinetics and their associated uncertainties is an essential task in the construction of accurate metabolic flux maps for phototrophs.

In vivo measurement of Ci utilization rates in a photosynthetic system is challenging. In aqueous solution, dissolved CO₂ exists in equilibrium with bicarbonate ions, and both forms can be taken up into photosynthetic cells. Ci uptake can be measured by various methods. We have previously used sealed tubes and gas chromatography to measure the difference of Ci concentration over time (Xiong et al., 2015). This method provides the averaged Ci uptake rate over a longer time (hours) but cannot distinguish Ci uptake from fixation kinetics nor measure real time performance at specific Ci concentrations. The isotope labeling method has also been widely used. However, it has a higher leakage rate and suffers from interference caused by non-labeled Ci brought into the growth medium by some of its constituents (Eichner et al., 2015). An alternative technique is to measure gas phase CO₂ by infrared absorption (Oakley et al., 2012). This technique requires higher cell density, which causes self-shading that affects the reliability of experiments. It also requires large culture volumes and is dependent on the equilibrium of CO₂ between the gas phase and the liquid phase, which can differ under various conditions including temperature, pressure, medium composition, and pH. Additionally, water vapor has a high absorption coefficient in the infrared region, making

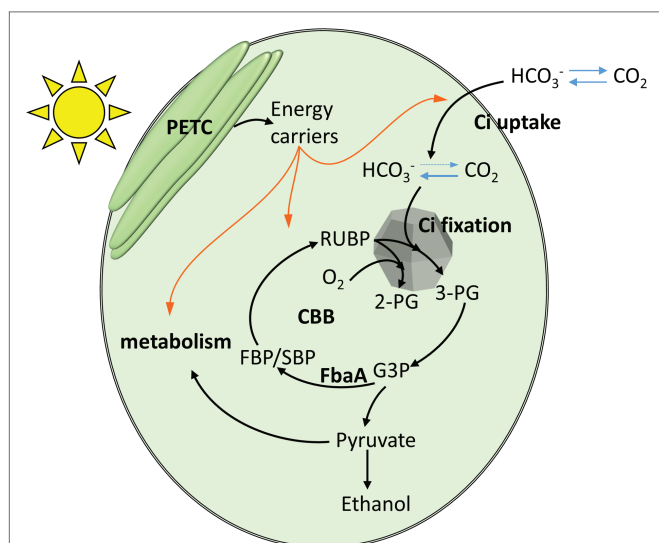


FIGURE 1 | A schematic representation of the relationship between photosynthetically harvested energy by the photosynthetic electron transfer chain (PETC), inorganic carbon (Ci) uptake and fixation, and the anabolic metabolism. Orange arrows indicate chemical energy fluxes. Black arrows indicate carbon flux.

the efficiency of the desiccant or compensation mechanisms critical for reliable measurements, especially at very low Ci concentrations. The measurement of O_2 evolution at different Ci concentrations is also a common method to determine the overall affinity of cells for Ci (Woodger et al., 2005). However, this is an indirect measurement and is complicated by the existence of multiple O_2 consumption pathways in cyanobacteria and other phototrophs.

In the 1960s, a technique was developed based on the permeation of gases through a silicone membrane which separate the culture media from the high vacuum line that leads to the detector of a mass spectrometer (Hoch and Kok, 1963). This experimental setup, now called MIMS (Membrane-Inlet Mass Spectrometry), was used to measure algal Ci consumption directly and in real time in the growth medium (Radmer and Kok, 1976). This instrument measures only small molecules that can pass through the silicone membrane. While CO_2 is detectable, carbonates and bicarbonates are not permeable. Therefore, the measure of total inorganic carbon relies on the existing equilibrium that results from the fast interconversion of $\text{CO}_2 \leftrightarrow \text{HCO}_3^- \leftrightarrow \text{CO}_3^{2-}$, which is highly sensitive to the medium pH.

While MIMS was used to study cyanobacterial CCM in the past (Espie et al., 1988; Miller et al., 1988; Tchernov et al., 2003), it has not been widely used in recent years. We adapted this method to measure Ci uptake flux in engineered cyanobacteria quantitatively by determining the extracellular Ci consumption rate per Ci concentration in real time. The kinetic parameters at various Ci concentrations are dependent on intracellular physiology and used to differentiate Ci uptake by CCM and Ci consumption by the carbon fixation machinery. We found that CBB engineering influences the regulation of the Ci uptake as well as fixation activities. We compared Ci uptake flux in wild-type (WT) *Synechocystis* and engineered strains over-expressing fructose-bisphosphate aldolase gene (*fbaA*) either in a WT background or in an ethanol-producing strain (Liang and Lindblad, 2016, 2017). *fbaA* over-expression affects the carbon metabolites in CBB cycle and likely Ci fluxes. Interestingly, the results indicate a positive relationship between the activity of the CBB cycle and the kinetics of CCM, which will help inform genetic engineering strategies of cyanobacterial central carbon metabolism for enhanced carbon utilization.

MATERIALS AND METHODS

Growth of Cyanobacteria

The *Synechocystis* strains used in this study were previously reported (Liang and Lindblad, 2016, 2017). We compared here the WT strain (empty vector control strain with kanamycin resistance, named Km in the below Liang et al. papers) to over-expressing strains generated *via* the same vector. They were grown on a modified BG-11, without added carbonate or bicarbonate, and supplemented with 50 mM NaCl and 20 mM TES, 50 mg/L kanamycin, and filter sterilized. The pH was adjusted to 7.4. Plates were kept under 5% CO_2 ,

30°C, 50 $\mu\text{E m}^{-2} \text{s}^{-1}$ provided by cool white fluorescent light tubes. The physiological tests were carried out after at least 48 h of incubation in 100 ml of BG-11 in 250 ml baffled Erlenmeyer flasks, continuously air bubbled at 100 ml min^{-1} , 30°C, 100 $\mu\text{E m}^{-2} \text{s}^{-1}$ of white LED light (4,500 K). The OD_{730} was always kept under 1, and the culture was diluted about 16 h prior to the experiment and harvested at an OD_{730} of about 0.4.

The MIMS set up, operation, and data processing are detailed in the section “Results.”

Doubling Time

Cultures with an $\text{OD}_{730} < 1$ pre-acclimated for at least 48 h at 100 $\mu\text{E m}^{-2} \text{s}^{-1}$ with air bubbling were diluted to an OD_{730} of about 0.050 in pre-warmed 30°C BG-11. The first density measurement was taken 1 h after the dilution. A second measurement was carried out after 16 h, when the OD_{730} was still under 0.5. These two measurements were used to calculate the doubling time.

Net O_2 Production

Cultures grown overnight with OD_{730} lower than 0.4 were spun down and resuspended in fresh BG-11 medium to OD_{730} of 0.7–0.8. Samples were incubated in 50 ml falcon tubes in the dark under agitation for 30 min. About 9 ml aliquots of the sample were transferred to the measurement cuvette in the dark and were stirred at 700 rpm and air bubbled for 10 min with 0.0022% antifoam 204, kept at 30°C by water bath. Measurement by MIMS started with this 10-min dark incubation. The air bubbling lead to an air saturated sample and used to calibrate at 278 $\mu\text{M O}_2$. When the incubation was completed, the bubbling was turned off, and after 10–20 s, the light was turned to 150 $\mu\text{E m}^{-2} \text{s}^{-1}$ (LED, 6,000 K) for 4 min to measure oxygen production. The measurement was maintained for 4 more minutes in the dark to measure respiration. Finally, the sample was bubbled with N_2 to a steady state to calibrate at 0 $\mu\text{M O}_2$.

High Ci Acclimation

WT strain was cultured in BG-11 medium (with 50 mM NaHCO_3 replacing the NaCl) for at least 3 days prior the measurement (re-diluted every 24 h), without air bubbling, but under the same stirring, temperature, and light conditions. The cells were harvested by centrifugation with the same procedure described in the Ci uptake method, except that the cells went through an extra step of washing with a medium without HCO_3^- .

Chlorophyll Content Measurement

WT strain and the *fbaA* OE strain were grown overnight from a healthy low Ci -acclimated culture. About 14 ml of culture with $\text{OD}_{730} < 0.4$ were spun down at 2,500 g for 5 min at room temperature, and the pellet was resuspended in 2 ml of -80°C methanol and left overnight at -80°C . After another spinning, the absorbance of the supernatant was measured at 665 and 720 nm. The chlorophyll content

was calculated using an equation of $\mu\text{g Chla/ml} = 12.9447 * (\text{OD}_{665} - \text{OD}_{720})$ (Ritchie, 2006).

RESULTS

MIMS Experimental Setup and the Biological Significance of the Measurement

It is widely accepted that CO_2 diffuses through biological membranes by using aquaporins and a number of other ways

(Tchernov et al., 2001). Under high Ci levels at pH 7–8, bicarbonate transporters and other constituents of the CCM are bypassed by the saturating flow of CO_2 entering the cell. In this work, Ci uptake was studied under conditions where all CCM components are actively involved. Air bubbling into the cultures was used to maintain a low but constant Ci supply while avoiding over accumulation of the O_2 generated by the photosynthetic electron transport chain (PETC). Cultures were maintained in BG-11 medium below an OD_{730} of 0.4, where self-shading is minimal, and nutrients are replete. Cells were harvested by centrifugation and re-suspended in fresh

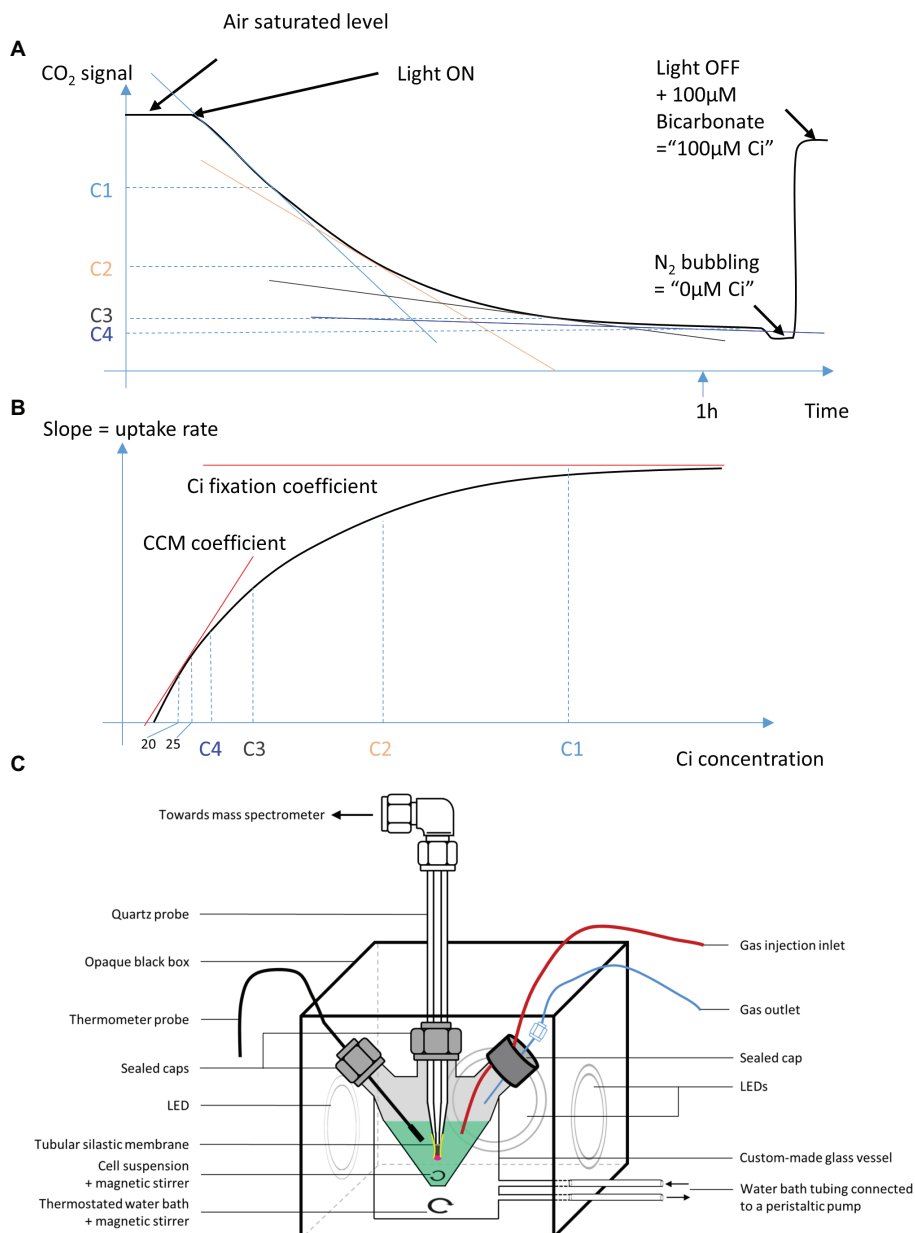


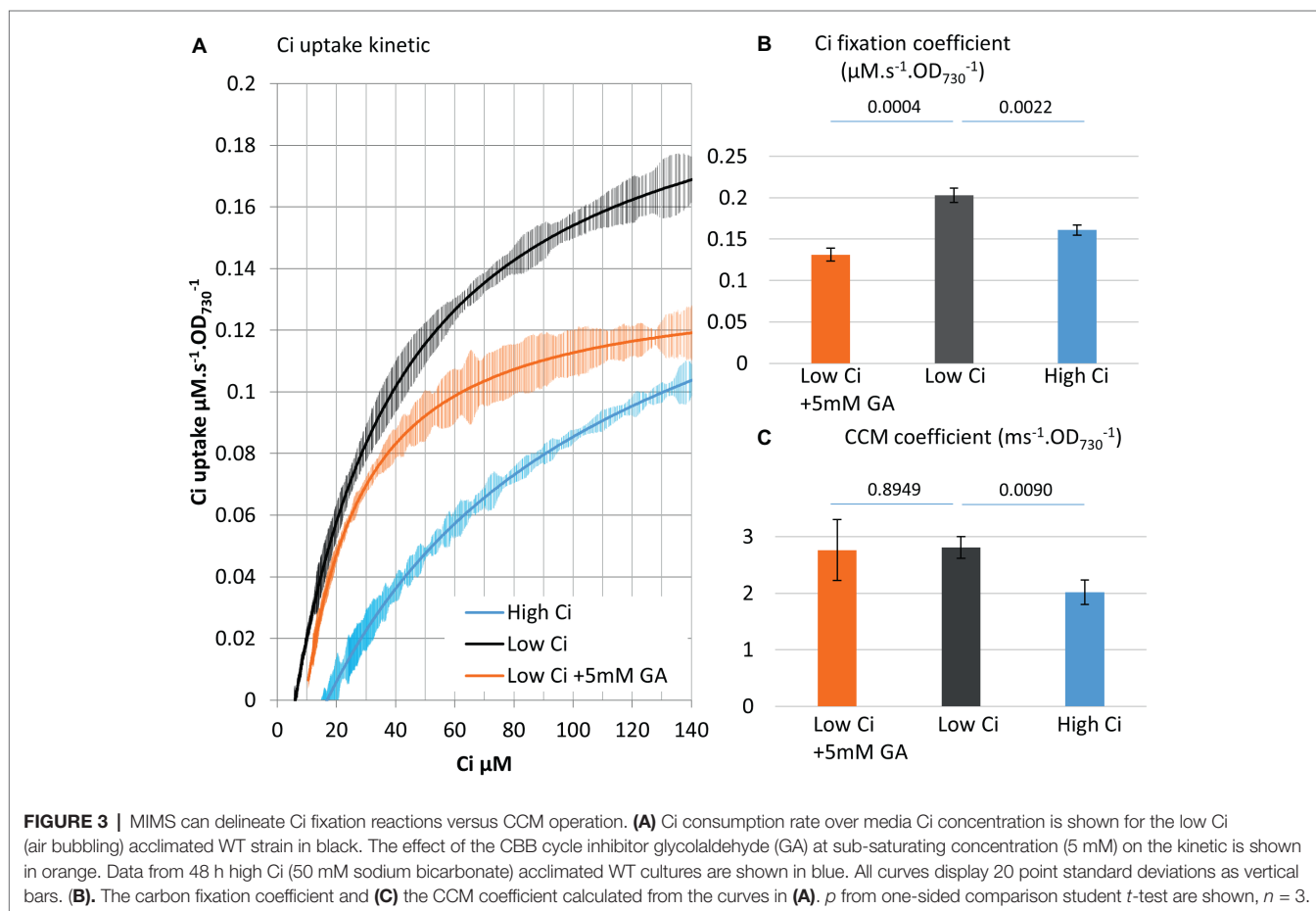
FIGURE 2 | A schematic representation of the MIMS set up and data processing. **(A)** A raw curve typically obtained on the MIMS. C1, C2, C3, and C4 are arbitrary Ci concentrations used to illustrate data processing. **(B)** A plot of the slopes calculated from curve A, over the Ci concentration, as displayed for all Ci uptake assays in this paper. **(C)** The MIMS experimental setup used in this study.

BG-11 media to an OD_{730} of 0.5. For consistent results in this measurement, the same media batch was used for all compared samples and replicas. About 9 ml aliquots of this suspension were subjected to CO_2 concentration monitoring by MIMS (HIDEN HAS-301-1503A) using the SEM detector (settle time 400 ms, dwell time 2,750 ms). The sample was injected into a 20 ml measurement cuvette, which was custom designed and crafted (Allen Glass, Boulder, Colorado; **Figure 2**). In such conditions, buildup of photosynthetic oxygen was minor, which was confirmed by MIMS measurements. The pH was observed to be constant during the course of the experiment due to the presence of 20 mM TES in BG11 media, compared to the 100 μM of Ci consumed. The suspension was completed with 0.0022% antifoam 204 (Sigma), stirred at 700 rpm, and bubbled with air for 8 min 30 s in 10 $\mu E m^{-2} s^{-1}$ (room light). The MS measurement was then commenced, and the bubbling stopped (T0). The measurement vessel was sealed, and the light was turned on to about 200 $\mu E m^{-2} s^{-1}$ (white LED at 6,000 K). Dissolved CO_2 was measured over time, from T0 when the sample is air saturated, to T-final when CO_2 consumption becomes undetectable. A calibration was performed after each sample, where zero Ci was set by bubbling the sample with N_2 for 10 min, and 100 μM Ci was set by an injection of freshly prepared $NaHCO_3$ at pH 7.4. The

concentration of dissolved CO_2 was considered a readout of the total Ci .

While it was hypothesized that in some cases, at the RubisCO vicinity, the consumption of CO_2 can drive an imbalance in the $CO_2 \rightleftharpoons HCO_3^-$ equilibrium (Sun et al., 2019), it is compensated intracellularly by carbonic anhydrases. The measurement of extracellular CO_2 could also possibly be interfered by an imbalanced Ci hydration equilibrium. However, in our conditions, it was shown that *Synechocystis* uptakes more HCO_3^- than CO_2 (Benschop et al., 2003), yet we observed an immediate decrease in CO_2 readings upon illumination as illustrated in **Figure 2A** (data not shown), indicating equilibrium of the Ci species. This is also supported by the absence of growth delay in cytoplasmic carbonic anhydrase mutant, indicating that the natural homeostasis between the two Ci species is able to support growth in laboratory conditions (So et al., 1998). In addition, the conditions we use are moderate in terms of pH, light, and cell density, thus unlikely to drive extracellular Ci out of equilibrium. Based on (Dreybrodt et al., 1996) and our calculation, at the time scale of 1 h experiment, we consider that the CO_2 hydration is not limiting.

Because *Synechocystis* acclimation to low Ci concentration begins after about 2 h (Benschop et al., 2003), our method was designed so that each individual measurement did not last more than an hour.



The C_i uptake rate was calculated from the slope of the C_i concentration versus time over a 20 timepoint window, and this rate was plotted against the C_i concentration of the first point used to calculate the slope. A fitting curve was calculated using a five-parameter Hill modified equation.

$$Y = Y_0 + \frac{a(X - X_0)^b}{c^b + (X - X_0)^b}$$

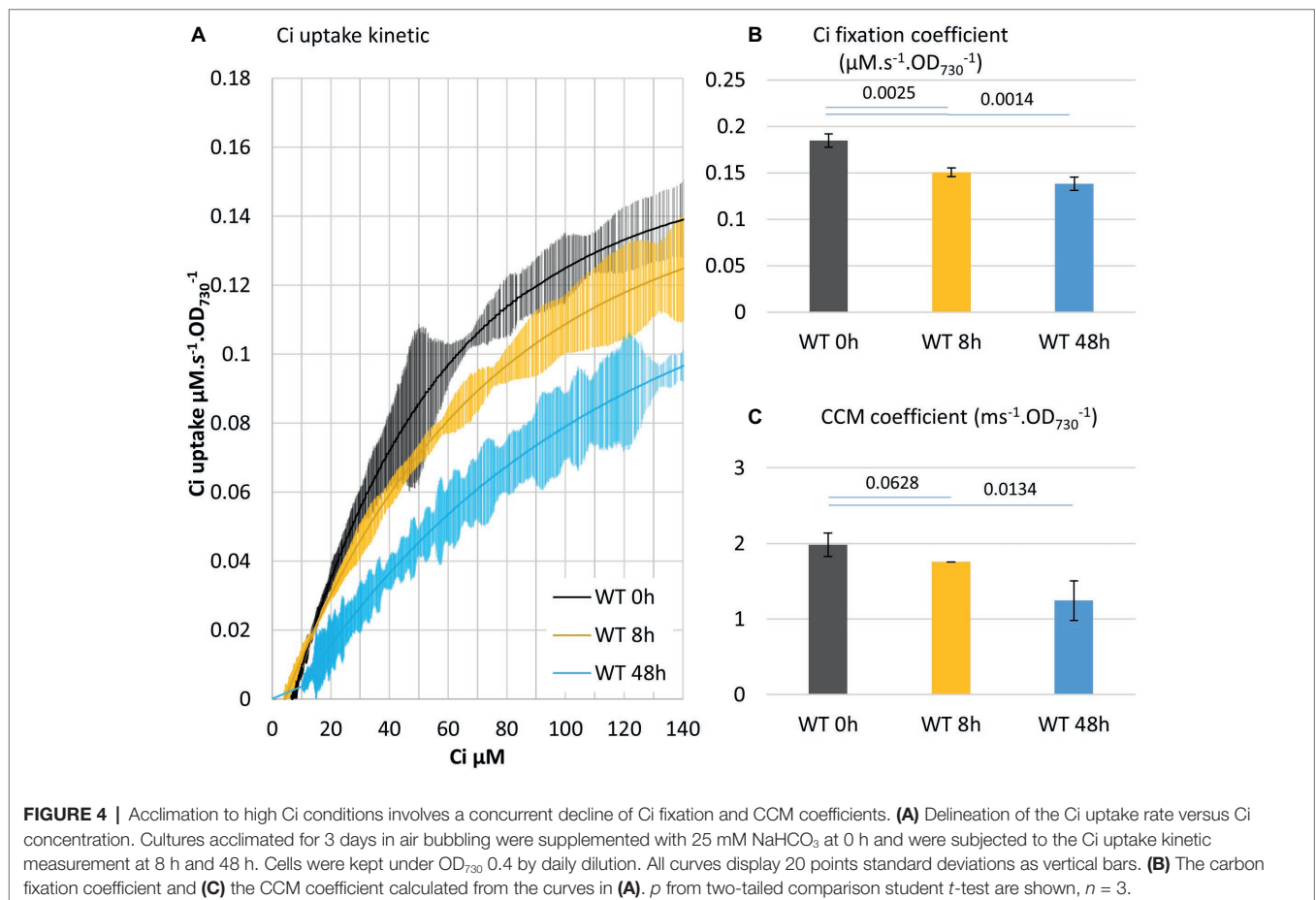
Y is the slope, plotted in ordinate; X is the C_i concentration plotted in abscissa; a , b , c , Y_0 , and X_0 are variables used to fit the curve to experimental data.

As shown in **Figures 2, 3**, the MIMS method provides real-time C_i uptake rate over a range of external C_i concentrations. At higher external C_i concentrations (air bubbling), C_i import into cells is considered not limiting, and the measured C_i uptake rates are limited by carbon fixation reactions. In contrast, as external C_i concentrations drop toward zero, CCM activity should become the rate limiting step for the measured C_i uptake rates. Thus, the fitting curve can provide quantitative measurement of both CCM activity and C_i fixation (CBB cycle) activity. On the low C_i end, we set 20 μM C_i as the reliable detection limit; thus, the region between 20 and 25 μM C_i on these fitting curves was used to calculate an initial slope, designated as the *CCM coefficient*.

On the high C_i end, we designate the *C_i fixation coefficient* as the maximum of the fitting sigmoid, representing C_i fixation rate at C_i sufficient conditions. The kinetic observed here is non-Michaelian, common for multicomponent reactions. Thus, the parameters described are not V_{\max} and K_m . Other parameters may be calculated from the curve as our understanding of C_i uptake advances.

Designation of the two coefficients allows quantitative differentiation between C_i fixation and CCM activities. An inhibitor study was performed to support this differentiation. The partial inhibition of the CBB cycle with 5 mM of glycolaldehyde, which targets the phosphoribulokinase (Miller and Canvin, 1989), does not affect CCM (Rotatore et al., 1992). When this inhibitor was used in the MIMS method, we observed partial inhibition of C_i uptake in the upper phase (~35%) but not the lower phase of the curve as predicted (**Figure 3**). At the lower phase, C_i becomes very low, and the C_i supply to the C_i fixation reactions *via* CCM activity becomes limiting. The results obtained here indicate that the lower C_i phase of the uptake curve can serve as an indicator of CCM.

We found that the growth phase and cell density of the culture were critical factors for data reproducibility. We also observed that WT strains from different laboratories do not always behave similarly; some do not induce complete CCM when placed in air bubbling.



Acclimation to High Ci Conditions Involves a Concurrent Decline of the CO₂ Fixation and CCM Coefficients

The acclimation of *Synechocystis* to high Ci concentrations, from bicarbonate addition or bubbling with CO₂ enriched air, triggers a decrease in the level of active transporters in CCM (Burnap et al., 2015; Holland et al., 2016), and their apparent affinities for Ci (Skleryk et al., 2002; Benschop et al., 2003). We attempted to verify the previous observation of a delayed acclimation of the CCM machinery to high Ci (Benschop et al., 2003) using the MIMS method and parameters.

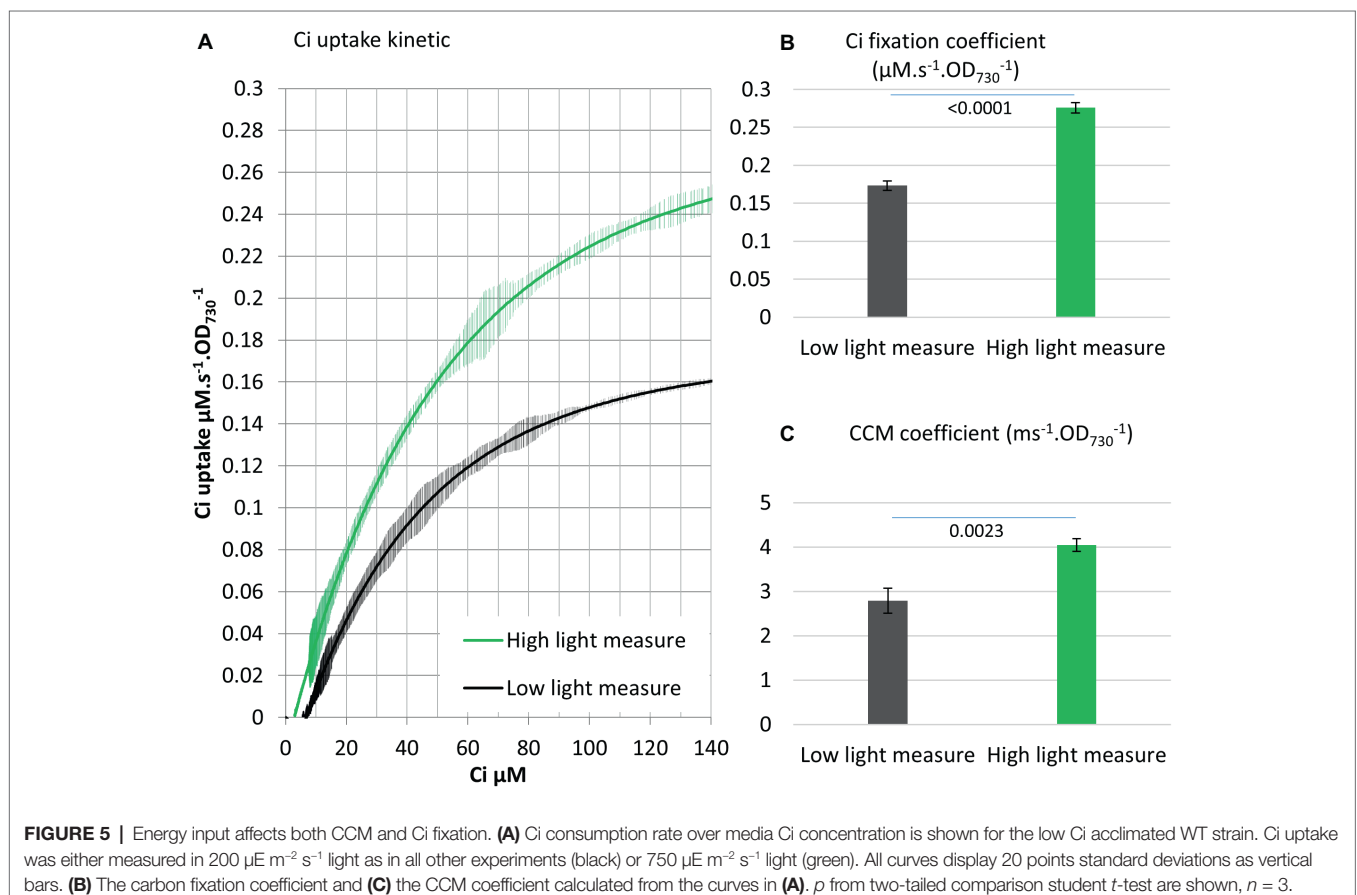
A WT culture adapted to low Ci was subjected to the addition of a saturating level of bicarbonate (50 mM) and sampled for MIMS measurement at 0, 8, or 48 h. Acclimation to high Ci conditions affected both the CCM coefficient and Ci fixation coefficient, but with different timing: the Ci fixation coefficient decreased by 19% in 8 h, while the CCM coefficient just started to decrease (note higher *p*). This is followed by a 25% decrease in the Ci fixation coefficient in 48 h, when the CCM coefficient decreased by 37% (Figure 4). These observations are in agreement with prior results (Benschop et al., 2003) and suggest that synthesis/activity of CCM and CBB enzymes are both regulated during adaptation to high Ci conditions. The regulation interestingly appears to be shifted in time. They also further support the assignation of the two phases of the curve to CCM and carbon fixation, respectively.

Energy Input Affects Both CCM and Ci Fixation as Exhibited by MIMS Measurement

Both the concentrating of Ci and the fixation of Ci *via* the CBB cycle requires energy input. The energy needed for those processes is provided by PETC, which harvests light energy and carries out a series of redox reactions leading to the production of energy carriers, mainly NADPH and ATP (Figure 1). We tested the influence of a higher light intensity (200 vs. 750 $\mu\text{E m}^{-2} \text{s}^{-1}$) on the Ci uptake kinetics and observed enhanced Ci fixation (+60%) and CCM (+45%) coefficients (Figure 5). While Ci uptake increased with increasing light intensity as previously reported (Kranz et al., 2010), our method further indicates that energy input can be a rate limiting step for both Ci concentrating and Ci fixation activities in *Synechocystis*.

Over-Expression of the CBB Cycle Enzyme *fbaA* Gene Impacts CCM and Ci Fixation

Central carbon metabolites, such as 2-phosphoglycerate (2PG), RuBP, 3PG, and alpha-ketoglutarate (AKG), as well as the electron carrier NAD(P)⁺ and the size of the intracellular Ci pool play an important role in the transcriptional regulation of cyanobacterial Ci uptake activity (Burnap et al., 2015). Based on these observations, we postulated that substrate regeneration by the CBB cycle could be engineered to enhance



Ci uptake and fixation. We tested this hypothesis using engineered strains, including an ethanol-producer, and strains over-expressing the fructose biphosphate aldolase gene (*fbaA* OE) alone or in combination with the ethanol-producer (Liang and Lindblad, 2016; Liang et al., 2018). MIMS was used to evaluate the effects of FbaA on carbon uptake and fixation machinery and to further characterize the effects of *fbaA* OE in carbon accumulation. FbaA is involved in two reactions within the CBB cycle. First, it catalyzes the combination of erythrose-4-phosphate (E4P) and dihydroxyacetone phosphate (DHAP) to form sedoheptulose-1,7-bisphosphate (SBP). Second, it also combines glyceraldehyde-3-phosphate (G3P) with DHAP to form fructose-1,6-bisphosphate (FBP). The *fbaA* OE strain grew faster (6.0 vs. 6.5 h⁻¹), had a higher chlorophyll content (+20%), and had a higher net O₂ evolution rate (+21%) compared to the WT (Figure 6). Both the Ci fixation and the CCM coefficients were about 14% higher in *fbaA* OE than WT (Figure 7).

fbaA Over-Expression Enhances Ci Uptake and Fixation in an Ethanol-Producing Strain

Many cyanobacterial strain engineering efforts involve the production of a target metabolite that is volatile or excreted into the medium, thus expanding the cell's metabolic sinks.

A number of such strains are reported to show enhanced carbon fixation although there is limited knowledge at the molecular level on how this enhancement is achieved (Nozzi and Atsumi, 2015; Xiong et al., 2015; Gao et al., 2016; Zhou et al., 2016). As a test case, we measured Ci uptake parameters in an ethanol-producing strain over-expressing pyruvate decarboxylase and alcohol dehydrogenase from *Zymomonas mobilis* (Luan et al., 2015; Liang and Lindblad, 2016, 2017; Liang et al., 2018). Over-expression of these two ethanol pathway genes and ethanol production had no effect on Ci uptake and fixation kinetics compared to WT (Figure 8). The additional *fbaA* OE feature in this strain nearly doubled ethanol productivity (Liang et al., 2018). We compared the growth rates of the WT and the two ethanol-producing strains. The introduction of the ethanol pathway increased the doubling time (8.6 h⁻¹) versus WT (6.4 h⁻¹), while the *fbaA* over-expression in the ethanol-producer restored WT like growth (7.1 h⁻¹) (Figure 6D). The doubling time of the *fbaA* OE in a WT background, on the other hand, was about 10% shorter than the WT (Figure 6C). While expressing ethanol production genes alone had no effect on Ci uptake and fixation kinetics, their co-expression with *fbaA* OE enhanced both CCM and Ci fixation parameters (10 and 7%, respectively; Figure 8), comparable to the enhancement observed in WT background (Figure 6).

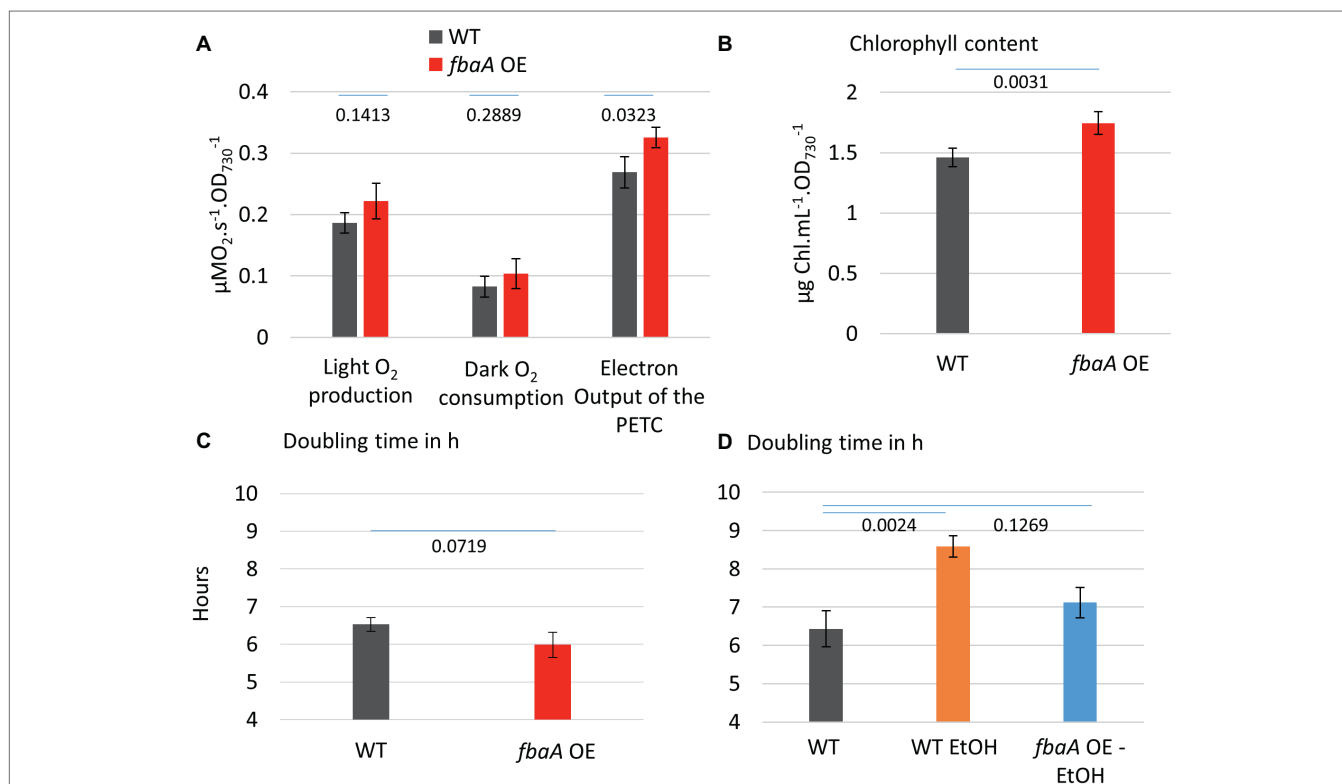


FIGURE 6 | *fbaA* over-expressing strain outperforms WT in multiple aspects (A) O₂ evolution and dark respiration of the WT and *fbaA* OE. Net oxygen evolution rate and dark respiration rate in a closed cuvette are used for the estimation of PETC output at 200 μE m⁻² s⁻¹ light, 30°C, and prior air bubbling. (B) Chlorophyll content relative to the OD₇₃₀. (C) Doubling time in the growth conditions including 100 μE m⁻² s⁻¹ light, 30°C, and air bubbling. (D) Doubling time of ethanol-producing strains with or without simultaneous over-expression of *fbaA*. *p* from one-sided comparison student *t*-test are shown, *n* = 3.

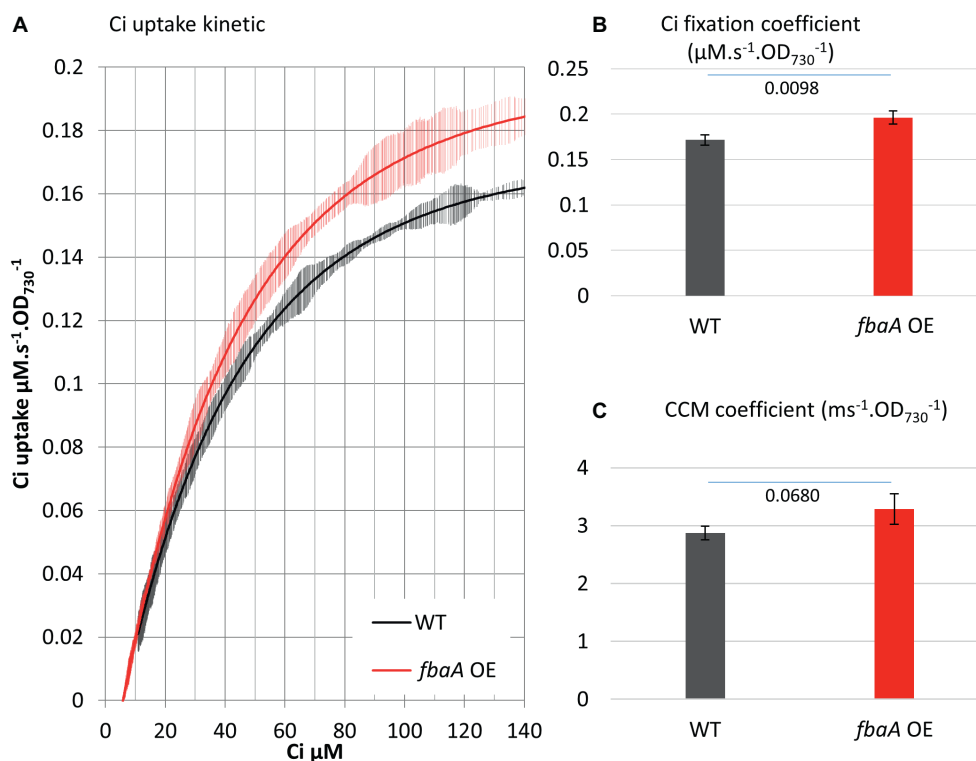


FIGURE 7 | *fbaA* over-expression impacts CCM and Ci fixation. **(A)** Ci consumption rate over media Ci concentration is shown for the low Ci acclimated WT (black) or *fbaA* OE (red). All curves display 20 points standard deviations as vertical bars. **(B)** The carbon fixation coefficient and **(C)** the CCM coefficient calculated from the curves in **(A)**. *p* from one-sided comparison student *t*-test are shown, *n* = 4.

DISCUSSION

The MIMS Method Can Quantify Both Ci Concentrating and Fixation Activities

Photosynthesis is nature's primary Ci utilization pathway and is the basis for developing phototrophs biotechnology for enhanced Ci utilization and the production of fuels and chemicals. Toward this end, a standard and simple method is needed to measure the Ci concentrating and fixing activities *in vivo*. Accurate measurement of Ci uptake rate is also essential for metabolic flux analysis. In this study, a MIMS method that is capable to measure Ci uptake in real time in declining external Ci concentrations is used to quantify CCM and Ci fixation kinetics separately. The results from three different conditions – inhibitor of the CBB, acclimation to high Ci, and increase of the light intensity during the measurement – were in agreement with literature and further support the dual phase readings we report here (Rotatore et al., 1992; Benschop et al., 2003; Kranz et al., 2010; Burnap et al., 2015; Holland et al., 2016).

The MIMS data yielded Ci uptake rates that were then fitted to a plot against the Ci concentration using the Hill equation. The fit allowed calculation of the Ci fixation coefficient reflecting the organism's maximum Ci uptake rate in air bubbling conditions, as well as the CCM coefficient reflecting the organism's carbon concentrating capability. The distinction between these

two parameters was confirmed by experiments using a sub-inhibitory concentration of the CBB cycle inhibitor glycolaldehyde. Future studies may assign specific CCM mechanisms to various regions along the curve based on Ci affinity, enabling more detailed studies of the individual mechanisms. The MIMS method can also potentially accommodate variations in light intensity and other physiological parameters during measurement.

Genetic Modification of the CBB Cycle Can Stimulate Both Ci Fixation and Uptake Reactions

Genetic modification of cyanobacterial central carbon metabolism can positively impact culture growth and carbon fixation. However, there has been limited understanding of how the enhancement of carbon utilization is achieved. Here, we tested Ci uptake parameters in a strain over-expressing the *fbaA* gene and in ethanol-producing strains with and without concomitant over-expression of the *fbaA* gene. While it is known that metabolites from the central carbon metabolism directly regulate the expression of CCM genes (Burnap et al., 2015; Orf et al., 2016), it has not been shown *in vivo* whether modifying the metabolic flux in CBB cycle impacts CCM activity. Our observation that *fbaA* OE increased both the Ci fixation coefficient and the CCM coefficient (**Figure 7**) provides *in vivo* evidence of a regulatory link between CBB

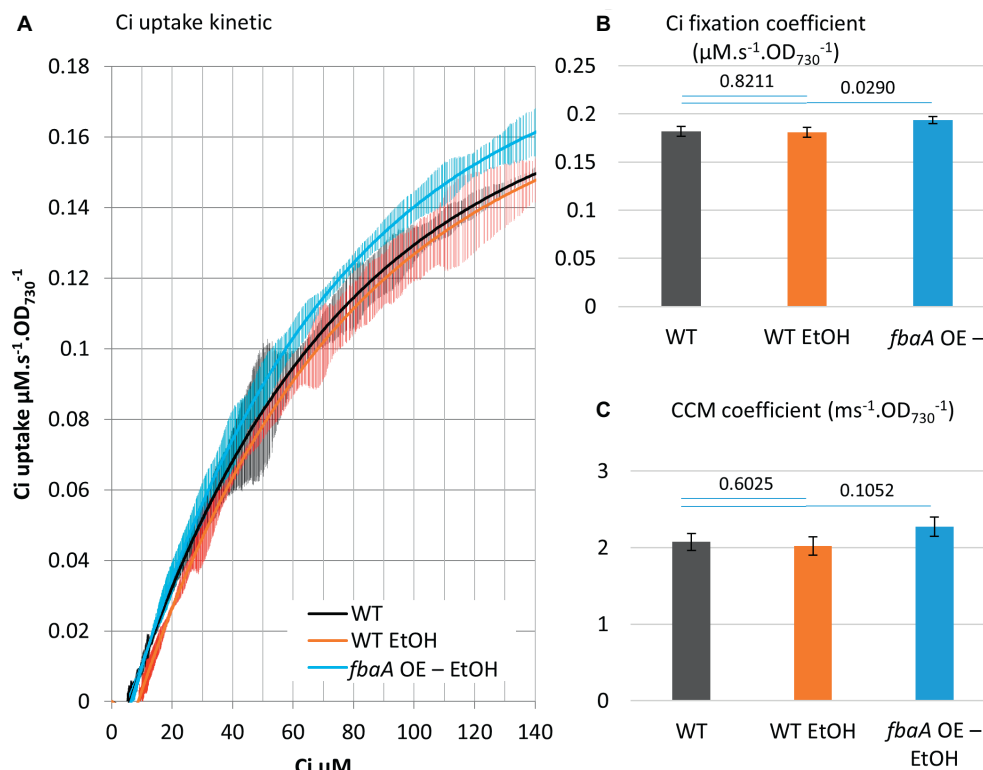


FIGURE 8 | *fbaA* over-expression enhances Ci uptake and fixation in an ethanol-producing strain. **(A)** Ci consumption rate over media Ci concentration is shown for the low Ci acclimated WT (black) and the ethanol-producing strain in WT background (WT EtOH, orange) or a *fbaA* OE background (*fbaA* OE – EtOH, blue). All curves display 20 points standard deviations as vertical bars. **(B)** The carbon fixation coefficient and **(C)** the CCM coefficient calculated from the curves in **(A)**. p from one-sided comparison student t -test are shown, $n = 3$.

metabolites and CCM enhancement. The *fbaA* OE strain is expected to have a modified metabolic flux toward the regeneration of RuBP and the depletion of 3PG. Our data indicate that the enhancement of RuBP regeneration in the CBB cycle likely improves the overall Ci concentrating and fixation processes. The increased fluxes of Ci uptake and CBB cycle demand more energy, which is provided by an increased photosynthetic machinery activity and light harvesting. This was evidenced by about 20% increase in O_2 evolution rate and chlorophyll content (Figures 6A,B). The stimulation of photosynthesis, including CCM, by CBB cycle modification represents another example of metabolic plasticity in cyanobacteria (Xiong et al., 2017).

The mechanism by which *fbaA* OE regulates CBB activity has been studied in transgenic plants. It was observed that FbaA activity affected photosynthetic efficiency, and its activity can be linked to the CBB metabolic balance and carbon partitioning in potatoes (Haake et al., 1998). Changes in CBB metabolite flux or concentrations are thought to trigger CBB enzymes regulation, likely through gene expression (Henkes et al., 2001; Cai et al., 2016; Simkin et al., 2017). Although contradictory results were found in other plants (Uematsu et al., 2012), all studies agree that the enhancement of photosynthetic activity is likely limited by 3PG depletion and RuBP regeneration. The enhancement of the overall carbon

fixation by *fbaA* OE is even stronger at higher Ci conditions (Uematsu et al., 2012), where the rate of Ci fixation by RubisCO depends on the ability of the CBB cycle to regenerate RuBP (Raines, 2003; Ma et al., 2005; Farazdaghi, 2011), assuming that RubisCO quantity itself is not limiting under high Ci conditions (Kanno et al., 2017). Whether the increase in the Ci fixation coefficient in *fbaA* OE cyanobacterium is due to CBB metabolite levels, changes and/or enzyme rearrangement remain to be studied.

In contrast, the mechanism by which *fbaA* OE regulates CCM may be suggested from prior studies of cyanobacterial CCM gene expression in response to metabolite signaling (Burnap et al., 2015). Metabolites, including 2PG, RuBP, 3PG, and AKG, as well as cofactor NAD(P) directly affect major LysR-type transcriptional regulators such as the NAD(P)H dehydrogenase regulator (NdhR), the *cmp* operon regulator (CmpR), and the *rubisCO* operon regulator (RbcR) (Burnap et al., 2015). The internal concentration of Ci itself and the Ci/O_2 ratio may also be a major signal for gene regulation (Woodger et al., 2005). Further studies of our strains, including quantification of CBB metabolites and transcriptome/proteome analyses, will be needed to clarify the regulatory mechanisms involved in photosynthetic enhancement by *fbaA* OE. In general, the positive effects on Ci utilization in some engineered strains could be attributed to metabolites/transcription

regulation pathways as described above and may also depend on biochemical constraints, abundance of transporters, affinity, and the half-lives of the proteins involved in the process. The MIMS method can determine *in vivo* how engineering affects RubisCO Ci fixation kinetics and/or carbon concentrating kinetics and thus help delineate these regulatory mechanisms.

Our observations in *fbaA* OE cyanobacterium extend to engineered strains producing ethanol. The production of ethanol alone negatively impacted the growth rate, suggesting that the diversion of pyruvate toward the ethanol pathway depleted carbon flux toward anabolic metabolism. This negative effect was compensated by the simultaneous *fbaA* OE in the combined strain, in which ethanol production doubled and the growth rate was restored close to WT levels, while both CCM and Ci fixation coefficients increased (Figure 8). It appears that the loss of carbon to ethanol production is compensated in the combined strain by an increase in Ci uptake and metabolic fluxes within the CBB cycle, therefore providing enough 3PG to feed both ethanol production and cell growth. These observations suggest that both growth and ethanol production are limited by Ci uptake and fixation in the ethanol-producing strain, while *fbaA* OE stimulated Ci uptake and fixation, thus improving both ethanol production and growth rate.

As discussed above, *fbaA* OE may lead to a modified balance between metabolites that may enhance the regeneration phase of the CBB cycle and therefore speeds up the RubisCO reactions. A similar hypothesis was made about the over-expression of another CBB enzyme gene, *fbp*, in the green alga *Chlorella* (Yang et al., 2017). Indeed, it is likely that the increased accumulation of other enzymes favoring the recycling of RuBP would also enhance Ci fixation. The benefit of increased growth potential is then balanced by the cost of producing additional proteins. It could be postulated that the metabolic changes within the CBB cycle would be different by over-expressing one gene versus another. If a step in the regulation of the CCM is modified by intermediate products of the CBB cycle, it is expected that the phenotype would be different among different overexpressers. Preliminary data show that a *Synechocystis* strain over-expressing the fructose/sedoheptulose

bisphosphatase gene displays an increase in Ci fixation and a decrease in Ci uptake (data not shown). Thus, over-expression of genes for different central metabolism enzymes could provide more complete information on the regulation of Ci uptake, fixation, and carbon partitioning. Investigating how such changes in the metabolite balance of engineered strains improves or impairs Ci uptake machinery and Ci fixation could guide future strain development for enhanced carbon utilization toward production of fuels and chemicals.

The dynamics of Ci consumption from genetically modified strains and cultures adapted to different growth conditions can be precisely measured using a simple reactor system equipped with MIMS. Such measurements can provide additional information or constraints for model-based ^{13}C -metabolic flux analysis and thus contribute to systematic studies of photosynthetic phenotype and fluxome.

AUTHOR CONTRIBUTIONS

DD and JY conceived the study and drafted the manuscript. DD conducted the experiments. FL and PL provided engineered strains. FL, MC, BW, WX, P-CM, PL, and JY assisted with experimental design, data interpretation, and troubleshooting. All authors edited and approved the manuscript. We thank Ms. Sunnyjoy Dupuis for language editing.

FUNDING

This work was authored in part by Alliance for Sustainable Energy, LLC, the manager and operator of the National Renewable Energy Laboratory for the U.S. Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. Funding provided by DOE Office of Energy Efficiency and Renewable Energy BioEnergy Technologies Office (WX, MC, BW, P-CM, JY). DD is supported by a Swiss National Science Foundation Postdoc Mobility Fellowship number: P2GEP3-168265. FL and PL acknowledge funding support from the NordForsk NCoE program “NordAqua” (project # 82845) and from the Swedish Energy Agency (project CyanoFuels, # P46607-1).

REFERENCES

- Adebisi, A. O., Jazmin, L. J., and Young, J. D. (2015). ^{13}C flux analysis of cyanobacterial metabolism. *Photosynth. Res.* 126, 19–32. doi: 10.1007/s11120-014-0045-1
- Artier, J., Holland, S. C., Miller, N. T., Zhang, M., and Burnap, R. L. (2018). Synthetic DNA system for structure-function studies of the high affinity CO_2 uptake NDH-13 protein complex in cyanobacteria. *Biochim. Biophys. Acta Bioenerg.* 1859, 1108–1118. doi: 10.1016/j.bbabo.2018.06.015
- Batchkikova, N., Eisenhut, M., and Aro, E.-M. (2011). Cyanobacterial NDH-1 complexes: novel insights and remaining puzzles. *Biochim. Biophys. Acta* 1807, 935–944. doi: 10.1016/j.bbabo.2010.10.017
- Benschop, J. J., Badger, M. R., and Dean Price, G. (2003). Characterisation of CO_2 and HCO_3^- uptake in the cyanobacterium *Synechocystis* sp. PCC6803. *Photosynth. Res.* 77, 117–126. doi: 10.1023/A:1025850230977
- Blankenship, R. E. (2010). Early evolution of photosynthesis. *Plant Physiol.* 154, 434–438. doi: 10.1104/pp.110.161687
- Burnap, R. L., Hagemann, M., and Kaplan, A. (2015). Regulation of CO_2 concentrating mechanism in cyanobacteria. *Life Basel Switz.* 5, 348–371. doi: 10.3390/life5010348
- Cai, B., Li, Q., Xu, Y., Yang, L., Bi, H., and Ai, X. (2016). Genome-wide analysis of the fructose 1,6-bisphosphate aldolase (FBA) gene family and functional characterization of FBA7 in tomato. *Plant Physiol. Biochem. PPB* 108, 251–265. doi: 10.1016/j.plaphy.2016.07.019
- Dreybrodt, W., Lauckner, J., Zaihua, L., Svensson, U., and Buhmann, D. (1996). The kinetics of the reaction $\text{CO}_2 + \text{H}_2\text{O} \rightarrow \text{H}^+ + \text{HCO}_3^-$ as one of the rate limiting steps for the dissolution of calcite in the system $\text{H}_2\text{O}-\text{CO}_2-\text{CaCO}_3$. *Geochim. Cosmochim. Acta* 60, 3375–3381. doi: 10.1016/0016-7037(96)00181-0
- Eichner, M., Thoms, S., Kranz, S. A., and Rost, B. (2015). Cellular inorganic carbon fluxes in *Trichodesmium*: a combined approach using measurements and modelling. *J. Exp. Bot.* 66, 749–759. doi: 10.1093/jxb/eru427

- Espie, G. S., Miller, A. G., Birch, D. G., and Canvin, D. T. (1988). Simultaneous transport of CO₂ and HCO₃⁻ by the cyanobacterium *Synechococcus* UTEX 625 1. *Plant Physiol.* 87, 551–554. doi: 10.1104/pp.87.3.551
- Farazdaghi, H. (2011). The single-process biochemical reaction of Rubisco: a unified theory and model with the effects of irradiance, CO₂ and rate-limiting step on the kinetics of C₃ and C₄ photosynthesis from gas exchange. *Biosystems* 103, 265–284. doi: 10.1016/j.biosystems.2010.11.004
- Gao, X., Gao, F., Liu, D., Zhang, H., Nie, X., and Yang, C. (2016). Engineering the methylerythritol phosphate pathway in cyanobacteria for photosynthetic isoprene production from CO₂. *Energy Environ. Sci.* 9, 1400–1411. doi: 10.1039/C5EE03102H
- Haake, V., Zrenner, R., Sonnewald, U., and Stitt, M. (1998). A moderate decrease of plastid aldolase activity inhibits photosynthesis, alters the levels of sugars and starch, and inhibits growth of potato plants. *Plant J. Cell Mol. Biol.* 14, 147–157.
- Henkes, S., Sonnewald, U., Badur, R., Flachmann, R., and Stitt, M. (2001). A small decrease of plastid transketolase activity in antisense tobacco transformants has dramatic effects on photosynthesis and phenylpropanoid metabolism. *Plant Cell* 13, 535–552. doi: 10.1105/tpc.13.3.535
- Hoch, G., and Kok, B. (1963). A mass spectrometer inlet system for sampling gases dissolved in liquid phases. *Arch. Biochem. Biophys.* 101, 160–170. doi: 10.1016/0003-9861(63)90546-0
- Holland, S. C., Artier, J., Miller, N. T., Cano, M., Yu, J., Ghirardi, M. L., et al. (2016). Impacts of genetically engineered alterations in carbon sink pathways on photosynthetic performance. *Algal Res.* 20, 87–99. doi: 10.1016/j.algal.2016.09.021
- Kamennaya, N. A., Ahn, S., Park, H., Bartal, R., Sasaki, K. A., Holman, H.-Y., et al. (2015). Installing extra bicarbonate transporters in the cyanobacterium *Synechocystis* sp. PCC6803 enhances biomass production. *Metab. Eng.* 29, 76–85. doi: 10.1016/j.ymben.2015.03.002
- Kanno, K., Suzuki, Y., and Makino, A. (2017). A small decrease in Rubisco content by individual suppression of RBCS genes leads to improvement of photosynthesis and greater biomass production in rice under conditions of elevated CO₂. *Plant Cell Physiol.* 58, 635–642. doi: 10.1093/pcp/pcx018
- Kaplan, A., and Reinhold, L. (1999). CO₂ concentrating mechanisms in photosynthetic microorganisms. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 50, 539–570. doi: 10.1146/annurev.arplant.50.1.539
- Kranz, S. A., Levitan, O., Richter, K.-U., Prásil, O., Berman-Frank, I., and Rost, B. (2010). Combined effects of CO₂ and light on the N₂-fixing cyanobacterium *Trichodesmium* IMS101: physiological responses. *Plant Physiol.* 154, 334–345. doi: 10.1104/pp.110.159145
- Liang, F., Englund, E., Lindberg, P., and Lindblad, P. (2018). Engineered cyanobacteria with enhanced growth show increased ethanol production and higher biofuel to biomass ratio. *Metab. Eng.* 46, 51–59. doi: 10.1016/j.ymben.2018.02.006
- Liang, F., and Lindblad, P. (2016). Effects of overexpressing photosynthetic carbon flux control enzymes in the cyanobacterium *Synechocystis* PCC 6803. *Metab. Eng.* 38, 56–64. doi: 10.1016/j.ymben.2016.06.005
- Liang, F., and Lindblad, P. (2017). *Synechocystis* PCC 6803 overexpressing RuBisCO grow faster with increased photosynthesis. *Metab. Eng. Commun.* 4, 29–36. doi: 10.1016/j.meteno.2017.02.002
- Luan, G., Qi, Y., Wang, M., Li, Z., Duan, Y., Tan, X., et al. (2015). Combinatory strategy for characterizing and understanding the ethanol synthesis pathway in cyanobacteria cell factories. *Biotechnol. Biofuels* 8:184. doi: 10.1186/s13068-015-0367-z
- Ma, W., Shi, D., Wang, Q., Wei, L., and Chen, H. (2005). Exogenous expression of the wheat chloroplastic fructose-1,6-bisphosphatase gene enhances photosynthesis in the transgenic cyanobacterium, *Anabaena* PCC7120. *J. Appl. Phycol.* 17, 273–280. doi: 10.1007/s10811-005-4850-y
- Miller, A. G., and Canvin, D. T. (1989). Glycolaldehyde inhibits CO₂ fixation in the cyanobacterium *Synechococcus* UTEX 625 without inhibiting the accumulation of inorganic carbon or the associated quenching of chlorophyll a fluorescence. *Plant Physiol.* 91, 1044–1049. doi: 10.1104/pp.91.3.1044
- Miller, A. G., Espie, G. S., and Canvin, D. T. (1988). Active transport of CO₂ by the cyanobacterium *Synechococcus* UTEX 625 1. *Plant Physiol.* 86, 677–683. doi: 10.1104/pp.86.3.677
- Nozzi, N. E., and Atsumi, S. (2015). Genome engineering of the 2,3-butanediol biosynthetic pathway for tight regulation in cyanobacteria. *ACS Synth. Biol.* 4, 1197–1204. doi: 10.1021/acssynbio.5b00057
- Oakley, C. A., Hopkinson, B. M., and Schmidt, G. W. (2012). A modular system for the measurement of CO₂ and O₂ gas flux and photosynthetic electron transport in microalgae. *Limnol. Oceanogr. Methods* 10, 968–977. doi: 10.4319/lom.2012.10.968
- Orf, I., Schwarz, D., Kaplan, A., Kopka, J., Hess, W. R., Hagemann, M., et al. (2016). CyAbrB2 contributes to the transcriptional regulation of low CO₂ acclimation in *Synechocystis* sp. PCC 6803. *Plant Cell Physiol.* 57, 2232–2243. doi: 10.1093/pcp/pcw146
- Price, G. D., Sültemeyer, D., Klughammer, B., Ludwig, M., and Badger, M. R. (1998). The functioning of the CO₂ concentrating mechanism in several cyanobacterial strains: a review of general physiological characteristics, genes, proteins, and recent advances. *Can. J. Bot.* 76, 973–1002. doi: 10.1139/b98-081
- Radmer, R. J., and Kok, B. (1976). Photoreduction of O₂ primes and replaces CO₂ assimilation. *Plant Physiol.* 58, 336–340. doi: 10.1104/pp.58.3.336
- Raines, C. A. (2003). The Calvin cycle revisited. *Photosynth. Res.* 75, 1–10. doi: 10.1023/A:1022421515027
- Ritchie, R. J. (2006). Consistent sets of spectrophotometric chlorophyll equations for acetone, methanol and ethanol solvents. *Photosynth. Res.* 89, 27–41. doi: 10.1007/s11120-006-9065-9
- Rotatore, C., Lew, R. R., and Colman, B. (1992). Active uptake of CO₂ during photosynthesis in the green alga *Eremosphaera viridis* is mediated by a CO₂-ATPase. *Planta* 188, 539–545. doi: 10.1007/BF00197046
- Simkin, A. J., Lopez-Calcano, P. E., Davey, P. A., Headland, L. R., Lawson, T., Timm, S., et al. (2017). Simultaneous stimulation of sedoheptulose 1,7-bisphosphatase, fructose 1,6-bisphosphate aldolase and the photorespiratory glycine decarboxylase-H protein increases CO₂ assimilation, vegetative biomass and seed yield in Arabidopsis. *Plant Biotechnol. J.* 15, 805–816. doi: 10.1111/pbi.12676
- Skleryk, R. S., So, A. K. C., and Espie, G. S. (2002). Effects of carbon nutrition on the physiological expression of HCO₃⁻-transport and the CO₂-concentrating mechanism in the cyanobacterium *Chlorococcopsis* sp. ATCC 27193. *Planta* 214, 572–583. doi: 10.1007/s004250100640
- So, A. K., Van Spall, H. G., Coleman, J. R., and Espie, G. S. (1998). Catalytic exchange of 18O from 13C18O-labelled CO₂ by wild-type cells and ecaA, ecaB, and ccaA mutants of the cyanobacteria *Synechococcus* PCC7942 and *Synechocystis* PCC6803. *Can. J. Bot.* 76, 1153–1160. doi: 10.1139/b98-063
- Sültemeyer, D., Price, G. D., Yu, J.-W., and Badger, M. R. (1995). Characterisation of carbon dioxide and bicarbonate transport during steady-state photosynthesis in the marine cyanobacterium *Synechococcus* strain PCC7002. *Planta* 197, 597–607. doi: 10.1007/BF00191566
- Sun, N., Han, X., Xu, M., Kaplan, A., Espie, G. S., and Mi, H. (2019). A thylakoid-located carbonic anhydrase regulates CO₂ uptake in the cyanobacterium *Synechocystis* sp. PCC 6803. *New Phytol.* 222, 206–217. doi: 10.1111/nph.15575
- Tchernov, D., Helman, Y., Keren, N., Luz, B., Ohad, I., Reinhold, L., et al. (2001). Passive entry of CO₂ and its energy-dependent intracellular conversion to HCO₃⁻ in cyanobacteria are driven by a photosystem I-generated delta muH⁺. *J. Biol. Chem.* 276, 23450–23455. doi: 10.1074/jbc.M101973200
- Tchernov, D., Silverman, J., Luz, B., Reinhold, L., and Kaplan, A. (2003). Massive light-dependent cycling of inorganic carbon between oxygenic photosynthetic microorganisms and their surroundings. *Photosynth. Res.* 77, 95–103. doi: 10.1023/A:1025869600935
- Uematsu, K., Suzuki, N., Iwamae, T., Inui, M., and Yukawa, H. (2012). Increased fructose 1,6-bisphosphate aldolase in plastids enhances growth and photosynthesis of tobacco plants. *J. Exp. Bot.* 63, 3001–3009. doi: 10.1093/jxb/ers004
- Wang, H.-L., Postier, B. L., and Burnap, R. L. (2004). Alterations in global patterns of gene expression in *Synechocystis* sp. PCC 6803 in response to inorganic carbon limitation and the inactivation of ndhR, a LysR family regulator. *J. Biol. Chem.* 279, 5739–5751. doi: 10.1074/jbc.M311336200
- Whitney, S. M., Houtz, R. L., and Alonso, H. (2011). Advancing our understanding and capacity to engineer nature's CO₂-sequestering enzyme, Rubisco. *Plant Physiol.* 155, 27–35. doi: 10.1104/pp.110.164814
- Woodger, F. J., Badger, M. R., and Price, G. D. (2005). Sensing of inorganic carbon limitation in *Synechococcus* PCC7942 is correlated with the size of

- the internal inorganic carbon pool and involves oxygen. *Plant Physiol.* 139, 1959–1969. doi: 10.1104/pp.105.069146
- Woodger, F. J., Bryant, D. A., and Price, G. D. (2007). Transcriptional regulation of the CO₂-concentrating mechanism in a euryhaline, coastal marine cyanobacterium, *Synechococcus* sp. strain PCC 7002: role of NdhR/CcmR. *J. Bacteriol.* 189, 3335–3347. doi: 10.1128/JB.01745-06
- Xiong, W., Cano, M., Wang, B., Douchi, D., and Yu, J. (2017). The plasticity of cyanobacterial carbon metabolism. *Curr. Opin. Chem. Biol.* 41, 12–19. doi: 10.1016/j.cbpa.2017.09.004
- Xiong, W., Morgan, J. A., Ungerer, J., Wang, B., Maness, P.-C., and Yu, J. (2015). The plasticity of cyanobacterial metabolism supports direct CO₂ conversion to ethylene. *Nat. Plants* 1:15053. doi: 10.1038/NPLANTS.2015.53
- Yang, B., Liu, J., Ma, X., Guo, B., Liu, B., Wu, T., et al. (2017). Genetic engineering of the Calvin cycle toward enhanced photosynthetic CO₂ fixation in microalgae. *Biotechnol. Biofuels* 10:229. doi: 10.1186/s13068-017-0916-8
- Zhou, J., Zhang, F., Meng, H., Zhang, Y., and Li, Y. (2016). Introducing extra NADPH consumption ability significantly increases the photosynthetic efficiency and biomass production of cyanobacteria. *Metab. Eng.* 38, 217–227. doi: 10.1016/j.ymben.2016.08.002
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.
- The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government. The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a nonexclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.*



The Design of FluxML: A Universal Modeling Language for ^{13}C Metabolic Flux Analysis

Martin Beyß^{1†}, Salah Azzouzi^{1†}, Michael Weitzel^{1†}, Wolfgang Wiechert^{1,2‡} and Katharina Nöh^{1*‡}

¹ Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, Jülich, Germany,

² Computational Systems Biotechnology (AVT.CSB), RWTH Aachen University, Aachen, Germany

OPEN ACCESS

Edited by:

Lars Keld Nielsen,
University of Queensland, Australia

Reviewed by:

Sonia Cortassa,
National Institutes of Health (NIH),
United States
Hiroshi Shimizu,
Osaka University, Japan
Maciek R. Antoniewicz,
University of Delaware, United States
Fumio Matsuda,
Osaka University, Japan

*Correspondence:

Katharina Nöh
k.noeh@fz-juelich.de

[†]These authors have contributed
equally to this work

[‡]These authors have contributed
equally to this work

Specialty section:

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

Received: 16 October 2018

Accepted: 24 April 2019

Published: 24 May 2019

Citation:

Beyß M, Azzouzi S, Weitzel M,
Wiechert W and Nöh K (2019) The
Design of FluxML: A Universal
Modeling Language for
 ^{13}C Metabolic Flux Analysis.
Front. Microbiol. 10:1022.
doi: 10.3389/fmicb.2019.01022

^{13}C metabolic flux analysis (MFA) is the method of choice when a detailed inference of intracellular metabolic fluxes in living organisms under metabolic quasi-steady state conditions is desired. Being continuously developed since two decades, the technology made major contributions to the quantitative characterization of organisms in all fields of biotechnology and health-related research. ^{13}C MFA, however, stands out from other “-omics sciences,” in that it requires not only experimental-analytical data, but also mathematical models and a computational toolset to infer the quantities of interest, i.e., the metabolic fluxes. At present, these models cannot be conveniently exchanged between different labs. Here, we present the implementation-independent model description language FluxML for specifying ^{13}C MFA models. The core of FluxML captures the metabolic reaction network together with atom mappings, constraints on the model parameters, and the wealth of data configurations. In particular, we describe the governing design processes that shaped the FluxML language. We demonstrate the utility of FluxML to represent many contemporary experimental-analytical requirements in the field of ^{13}C MFA. The major aim of FluxML is to offer a sound, open, and future-proof language to unambiguously express and conserve all the necessary information for model re-use, exchange, and comparison. Along with FluxML, several powerful computational tools are supplied for easy handling, but also to maintain a maximum of flexibility. Altogether, the FluxML collection is an “all-around carefree package” for ^{13}C MFA modelers. We believe that FluxML improves scientific productivity as well as transparency and therewith contributes to the efficiency and reproducibility of computational modeling efforts in the field of ^{13}C MFA.

Keywords: ^{13}C metabolic flux analysis, FluxML, machine-readable format, model specification language, computational modeling, reproducible science, data models, model exchange

INTRODUCTION

Systems Biology combines high-throughput experimentation with quantitative analysis and computational modeling to approach an understanding on how cellular phenotypes emerge from molecular interactions (Wolkenhauer, 2001; Westerhoff and Hofmeyr, 2005). To this end, a comprehensive set of “omics” techniques has been developed ranging from transcriptomics, proteomics, metabolomics to fluxomics, the quantification of metabolic reaction rates (fluxes)

in vivo (Nielsen, 2003). In the field of fluxomics, metabolic flux analysis (MFA) with stable isotope tracers, typically a ^{13}C labeled carbon source, is being considered as the “gold standard” for flux quantification under metabolic quasi-steady state conditions (Wiechert, 2001; Sauer, 2006). Being systematically developed in the mid-1990s (Marx et al., 1996; Christensen and Nielsen, 1999), ^{13}C MFA has been applied to a wide variety of organisms (microbes, plants, mammalian cell lines), cultivated under different conditions (chemostat, batch, fed-batch), in single, co-culture and host-pathogen systems (Beste et al., 2013; Ghosh et al., 2014; Gebreselassie and Antoniewicz, 2015), and probed with diverse labeling strategies (e.g., $^{13}/^{14}\text{C}$, ^2H , ^{15}N)¹ within isotopic transient or steady-state regimes (Zamboni et al., 2009; Niedenführ et al., 2015; Allen, 2016; Schwechheimer et al., 2018). For introductory texts on ^{13}C MFA, the reader is referred to the literature (Zamboni et al., 2009; Wiechert et al., 2015; Dai and Locasale, 2017).

Direct procedures to measure fluxes exist solely for extracellular rates, i.e., uptake and secretion fluxes. The determination of intracellular fluxes *in vivo* requires two additional ingredients: First, the measurement of the labeling incorporation into the intracellular metabolites. To this end, various analytical techniques such as homo- or heteronuclear, scalar- or multi-dimensional nuclear magnetic resonance (NMR) as well as single or tandem mass spectrometry (MS) are nowadays applied (Wittmann and Heinzle, 2001; Luo et al., 2007; Lane et al., 2008; Yuan et al., 2010; Giraudeau et al., 2011; Blank et al., 2012; Chu et al., 2015; Kappelmann et al., 2017). Second, and in contrast to other omics technologies, a powerful computational machinery is mandatory for data evaluation and flux inference. This means that the measured information, i.e., the isotopic data of intracellular metabolites together with the extracellular rates, does not directly uncover the desired flux information. The relation between isotopic enrichments and the fluxes is captured in a mathematical model which predicts the emerging fractional labeling patterns from given flux values. Clearly, this model has to be operated in the inverse direction to infer the, in reality, unknown fluxes from the observed data. These fluxes are then determined in an iterative fitting procedure in which the log-likelihood function, expressing the discrepancies between the model-predicted and measured quantities, is minimized. Finally, statistical measures estimate the confidence with which the fluxes are inferred from the data in view of their precision (Wiechert et al., 1997; Theorell et al., 2017).

As a consequence of this procedure, the results of any ^{13}C MFA intimately depend on the metabolic network model used.

Metabolic networks for ^{13}C MFA heavily vary in size, from focused representations consisting of only a few tens of reaction steps (Zamboni et al., 2009) to comprehensive descriptions with hundreds of reactions (Gopalakrishnan and Maranas, 2015; McCloskey et al., 2016b). Since the flux estimation procedure with such networks is computationally demanding, a number of algorithms have been proposed over the last two decades to speed up the core computation steps (Wiechert et al., 1999; Zamboni et al., 2005; Antoniewicz et al., 2007; Weitzel et al., 2007; Tepper and Shlomi, 2015). Unsurprisingly, these developments have led to the emergence of a variety of software tools that are almost as diverse as the experimental scenarios of ^{13}C MFA (see **Supplementary S1 Table 1.1**).

More on the ^{13}C MFA methodology and the assortment of flux analysis methods being applied is found elsewhere in the literature (Zamboni et al., 2009; Niedenführ et al., 2015; Wiechert et al., 2015). For the following considerations, it is sufficient to recognize that ^{13}C MFA in practice means a combinatorial variety of possible experimental, analytical, and computational configurations as well as model incarnations. The pros and cons of these different frameworks should not be scrutinized here. However, one aspect has to be emphasized: Despite of the heterogeneity of use cases, there is little debate about the principal conditions under which a ^{13}C MFA experiment must be conducted (i.e., metabolic pseudo-stationarity, homogeneous cell populations), and the input required for setting up the computational model (e.g., the structural description of the biochemical network underlying the model, specification of tracers, and measurements). Consequently, the precise configuration for an individual case comprises lots of specific details about the experimental-analytical setup and is, as evidenced in Section FluxML IN A NUTSHELL, rather complex.

Why Is There a Need for a Standardized Model Exchange Format in ^{13}C MFA?

Bundling all the aspects specific to an individual ^{13}C MFA study in a standardized document is undoubtedly of tremendous value for the community. This has already been proven by the success of the Systems Biology Markup Language (SBML, Hucka et al., 2003), which is today used as *lingua franca* to handle model-exchange between hundreds of different computational systems biology tools, as well as various other established modeling languages such as CellML (Lloyd et al., 2004) and NeuroML (Gleeson et al., 2010). Transferred to the ^{13}C MFA domain, this means to have a flux document formulated in a universal, i.e., network, algorithm-, tool-, and measurement-independent modeling language that is governed by controlled vocabularies and covers all current application cases.

A universal ^{13}C MFA modeling language allows sharing and publishing models in a complete, unambiguous, and reusable way. At present, this is only wishful thinking as existing guidelines (Crown and Antoniewicz, 2013) are, as we argue, not sufficiently strict. As a result, published papers do almost never supply all the information required to enable full reproduction of the model(s) used in the study. Partly, this incompleteness is due to the configuration processes that are too complex for

¹The use of ^{13}C labeled tracers is the mainstream scenario for ^{13}C MFA. The utilization of other stable isotope labeling strategies such as ^{15}N or simultaneous hetero-isotopic tracer combinations such as ^{13}C - ^{15}N is conceptually equivalent. Since it is the established notion in the field, the term “ ^{13}C MFA” is used throughout, although we extend its meaning to all of these alternative labeling strategies.

Abbreviations: FluxML, Flux Markup Language; FTBL, Flux TaBuLar format; ILE, isotope labeling experiment; INST, isotopically non-stationary; MFA, metabolic flux analysis; MID, mass isotopomer distribution; MS, mass spectrometry; NMR, nuclear magnetic resonance; OED, optimal experimental design; (S)LOC, (single) lines of code.

full reproduction in a paper. But also, implicit assumptions made in the modeling process—either by the modeler or hidden in the encoding of the software tool—remain undocumented, maybe unintentionally. In this sense, a standardized ^{13}C MFA modeling language provides a rule set to scientists for reporting re-usable models.

In a wider context, model exchange formats are an essential component for the reproduction of simulation results within the complex computational pipelines (Ebert et al., 2012; Dalman et al., 2016). As a practical benefit, a ^{13}C MFA modeling language empowers modelers to concentrate on the specification of the underlying network model, independent of the specific implementation in a software tool (cf. **Figure 1**). Such an “Esperanto” format is, thus, the central component for serving the FAIR Data Principles (Wilkinson et al., 2016). To put it straight, a standardized model exchange format fills the void and resolves many, if not all, of the current deficiencies. In addition, it paves the way for enhancing the models’ shelf lives and increases the efficiency of modeling efforts.

In this work we discuss the question: How should a universal model specification look like that digitally codifies all data required to carry out a ^{13}C MFA? By expanding on our former work (Wiechert et al., 2001), we motivate the benefits of a modern computer readable markup language for ^{13}C MFA, called Flux Markup Language (FluxML), and describe the governing principles of its design. To this end, we work out the required content that constitutes a model, formally known as *syntax standard*. Here, special focus is given to the model-data integration and extensibility aspects to keep pace with ongoing experimental-analytical developments. Clearly, to be adopted, such a general model representation effort must be accompanied by a set of supporting tools facilitating validation and modification tasks. We supply several computational tools with the modeling language, making the FluxML collection an “all-around carefree package” for modelers. The collection is illustrated with typical ^{13}C MFA examples at hand: First, we demonstrate that the FluxML model format unlocks the comparability of state-of-the-art simulators, an aspect that is dearly missing, even 20 years after advent of the first simulators. Secondly, we illustrate how easy the configuration task of parallel labeling experiments is with FluxML.

A CONTENT STANDARD FOR THE EXCHANGE OF MODELS AND DATA IN ^{13}C MFA

A model exchange document has to encapsulate all (necessary and optional) components, their interconnections, and the parameter information from which the computational model is built. In addition, since ^{13}C MFA is an experimental method, experimental data descriptors have to be defined from which the fluxes are to be inferred. To illustrate this in more detail, the specification of an isotope labeling experiment (ILE) and the corresponding measurement data includes the following elements (Wiechert, 2001; Wiechert et al., 2001):

1. Reaction stoichiometry, i.e., the structure of the metabolic model that defines the scope of the flux analysis. Depending on the desired resolution, metabolic pathways are formulated in full detail or as simplified, lumped reaction chains.
2. For modeling the flow of the relevant isotope tracer through the metabolic pathways, atom transitions for each reaction step, e.g., carbon atom transitions in case of ^{13}C labeling or carbon and nitrogen transitions in case of simultaneous ^{13}C - ^{15}N tracers. The atom transitions specify the precise mapping between the atoms of the reactions’ substrates and products. For reactants with rotation-symmetrical molecular structure multiple combinatorial atom mapping possibilities ensue (*scrambling reactions*).
3. Settings for metabolic fluxes as, for example, reaction directionality, range restrictions to keep flux values in physiologically sensible or desired limits, or to constrain believed-equal fluxes through scrambling reactions.
4. Tracer composition of the substrate or substrate mixture used for the ILE. This includes the exchange of intracellular metabolites with isotopically unlabeled external ones (such as rich media components, CO_2 etc.).
5. The types of measurements that are attained from the ILE (or, at the stage of a priori experimental design, are envisioned to be attainable), i.e., the *measurement configuration*:
 - a. Extracellular rates (or external fluxes), as derived from concentration profiles of exometabolites or bioprocess models.
 - b. Fractional labeling enrichments obtained by analytical instruments (e.g., positional labeling, mass isotopomer fragments, multiplets, etc.). For isotopically steady state conditions one set of labeling data is specified, while under isotopically non-stationary conditions (INST ^{13}C MFA) a time series of such sets is to be integrated representing the transient incorporation of label.
 - c. Intracellular pool sizes (i.e., concentrations) are key determinants of the labeling incorporation velocity and should, thus, be specified in INST ^{13}C MFA if experimentally accessible (Wiechert and Nöh, 2005; Nöh et al., 2006).
6. A set of variables that parametrizes the underlying computational model and enables its execution, e.g., the set of free fluxes (Wiechert and de Graaf, 1996).

This list can be regarded as minimal content standard for ^{13}C MFA models in the notion of Minimum Information Requested In the Annotation of biochemical Models (MIRIAM) (Le Novère et al., 2005). However, to define a language, a *syntax standard* (or *format*) is needed that provides structures for formatting the information laid down in the content standard. In addition, terminology and rules to specify valid models have to be declared, therewith enabling the semantic interpretation of the model descriptions.

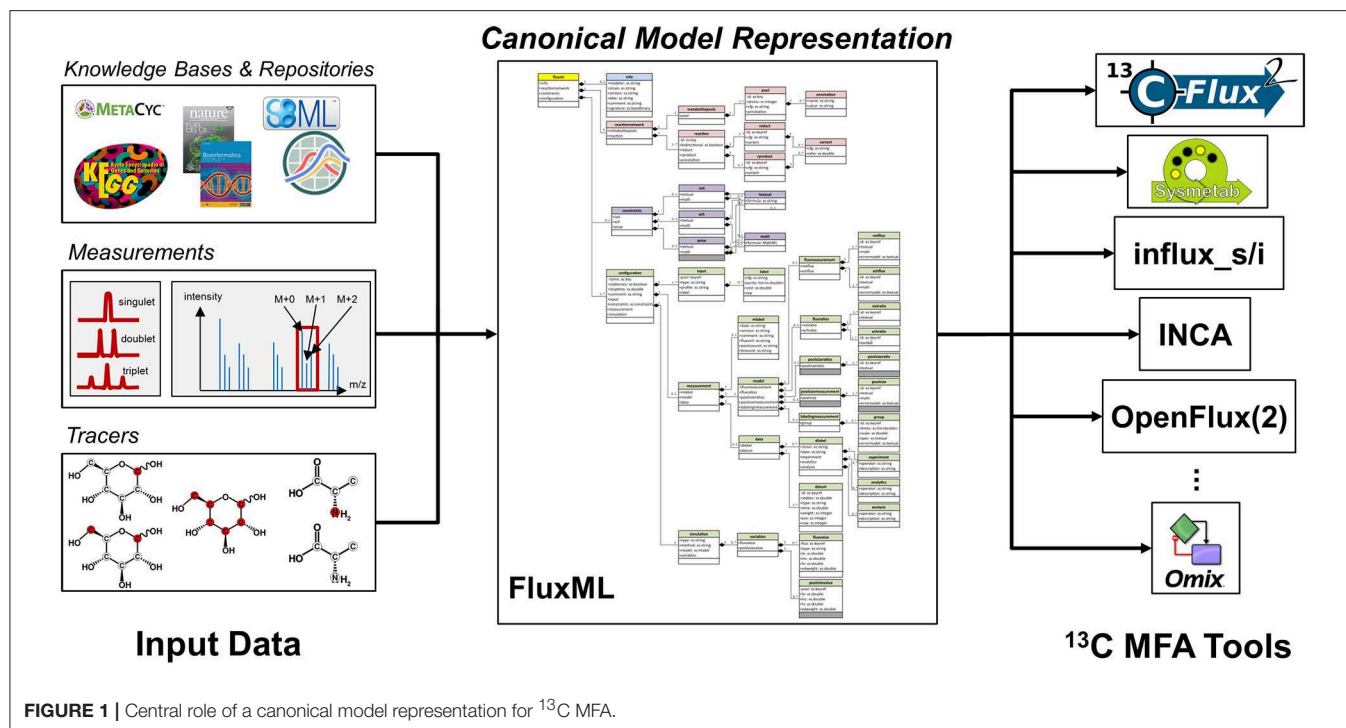


FIGURE 1 | Central role of a canonical model representation for ^{13}C MFA.

The requirement to deal with the broad diversity of ^{13}C MFA options renders the design of a modeling language a challenging endeavor. Before discussing the design decisions for FluxML in detail, former developments in the computational field should be briefly reviewed.

A Short History of ^{13}C MFA Modeling

Software systems developed in the past have used different approaches to supply the information needed to execute ^{13}C MFA. Several of the first generation flux analysis tools developed in the 90ies did not rely on dedicated specification formats but rather formulate the network and associated measurements by a set of matrices: atom mapping matrices to describe atom transitions (Zupke and Stephanopoulos, 1994) or isotopomer mapping matrices that unfold the system of isotopomer balance equations (Schmidt et al., 1997). One obvious problem with this matrix-centered approach is that it is prone to introduce specification errors which are hardly detectable afterwards.

To overcome this weakness, many second generation tools such as FiatFlux (Zamboni et al., 2005), tcaSIM/tcaCALC (Sherry et al., 2004), Metran (Young et al., 2008), INCA (Young, 2014), or WuFlux (He et al., 2016) have been equipped with graphical user interfaces (GUI) for a convenient model formulation (cf. **Supplementary S1 Table 1.1**). Such solutions are designed with having the end-user, typically an experimentalist, in mind who does not want to care about too many technical details. While the user-friendliness of these GUI-based tools is unraveled, they come at the price of a substantially restricted modeling flexibility: the abilities to change the reaction network or to formulate different measurement configurations are rather limited.

The first software framework for ^{13}C MFA that was able to deal with any isotopically stationary experimental setup in a freely configurable manner was 13CFLUX (Wiechert et al., 2001). Owing to the popularity of spreadsheets among experimentalists, 13CFLUX relies on tabulator-delimited text files for model and data specification, the FTBL (*Flux TaBuLaR*) format. FTBLs' concept to divide the required information into several contextual sections has been adopted by many software packages such as OPENFLUX(2) (Quek et al., 2009; Shupletsov et al., 2014), FIA (Srouf et al., 2011), and influx_s(i) (Sokol et al., 2012).

Despite the widespread use of FTBL, recent trends for automated lab experimentation and computational analysis pipelines (Dalman et al., 2010, 2016; Heux et al., 2017) call for contemporary model specification formats that are computationally easier to access and better verifiable than spreadsheets. Consequently, with our second generation ^{13}C MFA software 13CFLUX2 (Weitzel et al., 2013) an update to FTBL was proposed: the Flux Markup Language FluxML. FluxML exploits the powerful eXtensible Markup Language (XML) framework which has been designed to ease the computational processing of structured text documents. However, at the time of its publication, FluxML supported exclusively the formulation of isotopic stationary ^{13}C MFA models.

DECISIONS ON THE DESIGN OF FLUXML

Universal ^{13}C MFA Model Exchange Formats—Why an Update Is Needed

^{13}C MFA has been developed rapidly in the last decade. These developments have been impelled, in particular, by advances

in analytical measurement technologies where MS and NMR based approaches have been extended in scope and optimized in speed, resolution, precision, and accuracy (Moseley et al., 2011; Choi et al., 2012; Giraudeau et al., 2012; McCloskey et al., 2016a; Nilsson and Jain, 2016; Borkum et al., 2017; Kappelmann et al., 2017; Mairinger and Hann, 2017; Su et al., 2017). In turn, these developments triggered the setup of more comprehensive network models (Gopalakrishnan and Maranas, 2015; McCloskey et al., 2016b; Nilsson and Jain, 2016). Also INST ^{13}C MFA application scenarios have become more commonplace (Niedenführ et al., 2015; Cheah and Young, 2018; Delp et al., 2018; Gopalakrishnan et al., 2018). In view of these developments, existing formats have several limitations making a revision necessary.

Two decades of experiences with planning, modeling and analyzing ILEs and the continuous exchange with the 13CFLUX(2) user community have led to the specification of the updated FluxML format, which we present in this work. FluxML now covers isotopically stationary and non-stationary ILEs and is fully universal in terms of network, atom transition, measurement (error), and constraint formulation, including the use of multiple isotopes as tracers. It should be noted that the involved design processes, which we discuss in the following, were driven by the pragmatism to support modelers. Nevertheless, the FluxML format aims at a canonical model representation and follows the recommendations provided by the COMBINE (COMputational Modeling in BIOlogy NETwork, <http://co.mbine.org/>) initiative.

Design Decision 1—Scope: Data Pre-processing Is Not Part of FluxML

Measurement instruments generate raw data that first must be processed to be utilizable for ^{13}C MFA. For example, fractional labeling patterns must be extracted from NMR or MS spectra. This includes the identification of target fragments followed by the determination of their abundance by peak integration. For INST ^{13}C MFA, in addition, absolute intracellular pool sizes are to be determined. Here, special care has to be taken to correct for known biases in the sampling procedure (e.g., quenching, cell separation, and metabolite extraction). For example, the loss of intracellular metabolites during quenching (known as *leakage effect*) has to be counteracted by application of advanced protocols (Noack and Wiechert, 2014). For both quantities, the labels and pool sizes, standardization and modeling the propagation of the measurement error throughout the analytical processing pipelines is becoming best practice (Tillack et al., 2012; Mairinger et al., 2018).

On the other hand, most software systems for ^{13}C MFA emulate metabolite backbones rather than the analytically observed molecules. This means, that the data derived from the raw mass spectra must be corrected for “artificial” and/or “natural” isotope labeling contributions before conforming with ^{13}C MFA (Lee et al., 1991; Fernandez et al., 1996; Wahl et al., 2004; Jungreuthmayer et al., 2016; Niedenführ et al., 2016; Su et al., 2017). Also, the specific chemical nature of the analyte mixture and the analysis technique employed might lead to

distorted observations, such as proton-loss/gain, which require correction prior to model integration (Poskar et al., 2012). In addition, non-negligible inoculation residues or preliminary labeling sampling times may bias the interpretation of labeling enrichments in the classical case and need, thus, to be corrected (van Winden et al., 2001; Wiechert and Nöh, 2005). Finally, cell-specific external rates and their errors are calculated from cultivation data (concentration time courses of extracellular metabolites, off-gas analysis, biomass composition etc.) by means of simple regression (Murphy and Young, 2013), differentiation after smoothing (Llaneras and Picó, 2007), stochastic filtering (Cinquemani et al., 2017), or tailored bioprocess models (Noack et al., 2011).

That said, it becomes clear that such pre-processing procedures are extremely eclectic and heterogeneous, require a high degree of expertise, and underlie continuous change due to changing experimental setups, instrumentation, vendor formats, and analytical method developments. Recently, the metabolomics community got sensitized about their needs for reporting standards. Data formats and repositories are now under development, to report and store raw data along with its meta-information (Kale et al., 2016; Rocca-Serra et al., 2016). To avoid duplication, FluxML includes only those details about the evaluation procedures that contain the necessary key information about the measurement data that is actually used for producing the flux map (i.e., the *use data*, s. a. Section Experimental Data). The decision to not incorporate data pre-processing is also reasonable from a computer science perspective, since encapsulating complex designs in compact, orthogonal modules limits the overall complexity of the specification and eases future developments.

Design Decision 2—Technical Considerations: An XML Format for ^{13}C MFA

Generally, a modeling language must have a clearly defined syntax and succinct and precise specification of its semantics, for the computer but also for the human if necessary. Re-employing language concepts that are accepted by the target audience help to reduce learning hurdles. With SBML, a XML dialect is already available that is familiar to systems biologists. Technically, the design of FluxML was influenced by SBML as well as the following general considerations:

- ^{13}C MFA is embedded in workflows consisting of raw data acquisition, customized pre- and post-data evaluation, visualization, computational experimental design and further processes interfacing digital data. Repeated and iterative tool application is commonplace (Dalman et al., 2016). Therefore, it must be possible to use a ^{13}C MFA tool on a distributed computing platform, at best as a web service. In this scenario, XML is by far the most ubiquitous information exchange format worldwide. Nowadays, XML formats are commonly used for the structured information exchange in computational biology.
- For large-scale networks and complex experimental setups the specific software configuration tasks, i.e., how the

computational model is actually created and mapped to the internal data structures of a simulator, are error-prone. For this reason, all required data structures must be generated automatically by some kind of model compiler. For dealing with XML files, hundreds of off-the-shelf parsing, verification and transformation tools are available. This eases the writing of processing software for developers.

- Clearly, XML is not designed for a human reader, risking low acceptance among biologists. However, as the SBML success story exemplifies, this argument becomes invalid as soon as convenient, at best graphical, tools for model editing, validation, and formatted export are available. In the scenario of large-scale modeling and proofreading, diagnosis of inconsistencies in the model formulation is vitally important. Using structured XML entails the capability to benefit from powerful validation mechanisms that allow for the precise diagnosis of syntactical and semantical errors.
- XML combines the flexibility of full configurability with the user-friendliness of lowering complexity. It is very easy to extract partial information from XML files or to extend XML formats with additional information. With this, XML allows model pre-configuration to present only those parts to the experimentalist that are relevant, e.g., because they change over a series of experiments.
- A frequently discussed issue in simulation technology is the separation of model structures, model parameters, and measurement data. For instance, it is desirable to use the same model structure for identically configured ILEs, with parameters and data changed. Using XML, this poses no problem as long as model, parameters, and data are deposited in different branches of the XML tree. Furthermore, XML provides mechanisms to store associated model structures and data in separate files.

Design Decision 3—FluxML a Domain-Specific Language

One of the key design objectives of FluxML was to allow for automated model interpretation (analysis and code generation) for large-scale isotope labeling networks without forcing the modeler to resort to text-based specifications of low-level model description languages. Here, it could be argued that the flexibility of general description languages like CellML, offering a low-level description of the mathematical equations, is unraveled when new experimental or analytical paradigms become available. However, the generality comes at the price of readability and clearly challenges the proofreading capabilities of the modeler.

On the other hand, isotope labeling networks share many aspects with stoichiometric metabolic network models. For this reason, FluxML and SBML have a common subset of information that contains the metabolite and reaction names as well as the network stoichiometry and flux constraints. While reaction kinetic information is currently not in the scope of ^{13}C MFA, atom transitions, tracer mixtures, as well as experimental data are not part of SBML. Thus, the set of common features is not that large. Recently, an attempt has been made to encode the surplus information required for ^{13}C MFA in the SBML notation (Birkel

et al., 2017). Here, the construct *notes* (extending *reaction* and *species* in the notion of SBML) has been utilized to express carbon atom mappings and measurement data. However, because atom transitions and measurement specifications are vital for generating the essential mathematical system (Weitzel et al., 2007), it is clear that specifying this information in optional add-on elements, such as *notes*, complicates validation and consistency checking enormously. Hence, such a solution is not recommended by the SBML designers².

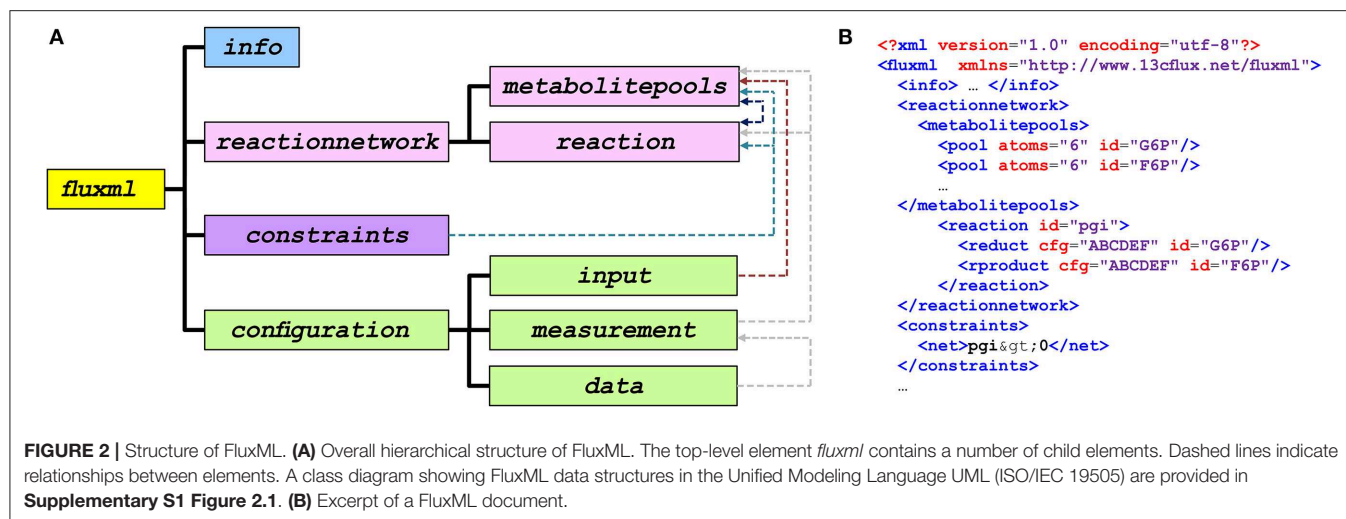
Taken together, these reasons speak in favor of the domain-specific standalone XML-based language. We followed the example of SBML and adopted those parts belonging to the common language subset with only minor changes to FluxML. The common subset is then extended by the information necessary to specify ILEs. Firstly, this way the entry level for a newcomer already familiar with SBML is lowered. Secondly, extracting the common information from a FluxML file and generating a rudimentary SBML document, or vice versa, is fairly straightforward.

FLUXML IN A NUTSHELL

FluxML development branches are organized in major *Levels* and minor *Versions*. Level 1 is dedicated to isotopically stationary ^{13}C MFA (Weitzel et al., 2013) while Level 2 covers both, the isotopically stationary and non-stationary cases. During language design special care has been taken to keep Level 2 backward compatible to Level 1, meaning that existing simulation tools designed for using the published FluxML version (Weitzel et al., 2013) do not need adaption when being used with Level 2 files. This helps third-party software developers using Level 1-models as input in keeping their versions stable. Lastly, FluxML Level 3 has been developed which extends Level 2 to the general case of multiple isotopically labeled elements. Here, for obvious reasons, backward compatibility could no longer be maintained.

The general hierarchical structure of FluxML documents that are common for all Levels is shown in **Figure 2**. **Figure 2A** overviews the main elements of the FluxML language while **Figure 2B** shows a code excerpt from the serialization of a model. The top-level element *fluxml* contains the elements *info*, for providing basic information about the model, and *reactionnetwork* containing metabolites and reactions which, together with the *constraints* element define the isotope network structure. An important key concept of FluxML, which is not present in SBML, is that of *configurations*. *configurations* entail the convenient possibility to connect the same model structure with different experimental or simulation settings. In this way, instances that, for example, differ in the selection of the tracer mixture, flux parametrization, and/or measurement configuration can be stored in different *configurations* sections within one model file. Another core concept of FluxML, distinguishing it from SBML, is the incorporation of experimental data. Here, the measurement

²"In particular, it is critical that data essential to a model definition ... is not stored in annotations" [http://sbml.org/special/specifications/sbml-level-3/version-2/core/release-1-rc1/sbml-level-3-version-2-core.pdf].



data declaration is separated from the data specification by the *measurement* sub-elements *model* and *data*, respectively. Finally, the *simulation* element contains details about the model parameterizations in terms of free model parameters as well as their values.

Instead of an exhaustive language description, only the major features of each FluxML section should be highlighted in the following, in particular those that eliminate limitations of the FTBL format and represent novel developments in the field.

Metabolites, Reaction Network Structure, and Atom Mappings

Clearly, biochemical reaction steps and their atom transitions constitute the core of any ^{13}C MFA model. The section *reactionnetwork* defines the metabolite pools, the reactions interconnecting them, and atom transitions which, altogether, give rise to the network structure of the ^{13}C MFA model. Each metabolite and reaction is labeled with a unique identifier (*id*) which assures its consistent usage throughout the FluxML document (cf. **Figure 2B**). Here, the atom enumerations are of particular importance not only for tracking the atoms, but also the correct association of the measured labeling fractions with the reactants.

Before going into specification details, it is appropriate to briefly summarize the most important facts about the network and atom transition compilation. Although there are plenty of ways to retrieve information from reaction databases (KEGG [http://www.genome.jp/kegg/], BioCyc [https://biocyc.org/], MetRxn Kumar et al., 2012), fluxomics collections (Zhang et al., 2014) [http://www.cecafdb.org], model repositories such as Biocompare [https://www.ebi.ac.uk/biocompare-main/] and BIGG [http://bigg.ucsd.edu/], as well as algorithmic approaches (Kumar and Maranas, 2014; Hadadi et al., 2017), there is currently no “one” curated source containing all the structural information needed for setting up a ^{13}C MFA model. In this context it is worthwhile to remember that solid biochemical knowledge beyond simple net reaction stoichiometry is needed. One prominent example is the

transketolase- and transaldolase-catalyzed reaction complex in the pentose phosphate pathway (PPP) where the kinetic enzyme mechanism impacts the formulation of the associated carbon atom transitions (van Winden et al., 2001; Kleijn et al., 2005). Considering this, it is fallacious to solely rely on information available in biochemistry textbooks and reaction databases. Further study-specific factors to be considered are reaction reversibilities, transamination reactions, isoenzymes showing evidence for differences in substrate affinity and activity, and (micro-)compartmentalization due to metabolite channeling or metabolically inactive pools (van Winden et al., 2001). All these factors may influence flux inference from available labeling distributions. On the other hand, it is common to simplify reaction networks, e.g., by lumping “linear” reaction chains into one surrogate reaction when the labeling distribution (and incorporation speed in case of INST) is not affected.

The mentioned considerations imply that the ^{13}C MFA model compilation procedure is hardly automatable, at least for non-standard cases. Currently the best way to build and verify the network model from scratch is to use various information sources and a visual tool for specification and proofreading purposes (Nöh et al., 2015). Having a list of relevant reactions and metabolites at hand, different naming conventions exist for representing the associated atom transitions. Traditionally, case-sensitive characters have been used to specify carbon transitions, as exemplified by the Fructose-bisphosphate aldolase reaction in glycolysis (in biochemical enumeration and FTBL notation):

```

emp4: FBP > GAP + DHAP
#abcdef > #cba + #def
  
```

Although this notation is convenient for an end-user, and still used by many software tools, it obviously does not fulfill the aforementioned requirements of a universal language. For this reason, atom transitions are specified in FluxML as follows:

```

<reaction id="emp4">
  <reduct cfg="abcdef" id="FBP"/>
  <rproduct cfg="cba" id="GAP"/>
  <rproduct cfg="def" id="DHAP"/>
</reaction>
  
```

This way, a reaction (*reaction*) can accommodate an arbitrary number of educts (*reduct*) and products (*rproduct*). These refer to unique metabolite names that are declared in the *metabolitepools* section of the FluxML document along with the definition of label-carrying atom types and numbers (cf. **Figure 3B**).

An implicit assumption underlying both *emp4* representations is the use of the IUPAC recommendation for coding the carbon atom-character relation of the metabolites. Herein, the lettering starts with the highest oxidized group of a molecule following the main carbon chain etc. For instance, following the biochemical enumeration the first carbon atom of glyceraldehyde 3-phosphate (GAP) is the one in proximity of the phosphate group (cf. **Figure 3A**). Due to its popularity among biochemists the IUPAC “biochemical enumeration scheme” has settled as pseudo-standard.

However, having genome-sized networks and multi-element ILEs in mind, this enumeration practice becomes questionable. In this situation, a veritable alternative is the International Chemical Identifier (InChI) (Heller et al., 2015). The InChI identifier is a computer-generated unique character string for encoding molecular structures that is widely accepted in the chemical community. The InChI identifier does not only facilitate database/web-search and information exchange in the field of metabolomics, it also comes with an outstanding merit for ¹³C MFA model exchange: InChI gives an identifier and canonical ordering to each atom of a metabolite (except for hydrogen). Thereby, employing InChI strings for metabolite declaration and atom enumeration makes network descriptions self-contained and exchangeable.

As an example, **Figure 3** shows the atom numbering as provided by the InChI software [<http://www.inchi-trust.org/>]. Accordingly, the carbon atom transitions for the aldolase reaction *emp4* in FluxML notation is:

```
<reaction id="emp4">
  <reduct cfg="C#1@1 C#2@1 C#3@1 C#4@1 C#5@1
    C#6@1" id="FBP"/>
  <rproduct cfg="C#4@1 C#1@1 C#3@1"
    id="GAP"/>
  <rproduct cfg="C#5@1 C#2@1 C#6@1"
    id="DHAP"/>
</reaction>
```

Herein, the atoms of the educt *FBP* are represented by white-space separated list of entries of the form *element#canonical_atom_index@educt_index* which are mapped to the respective atom positions in the products. Of course, the mapping can still be expressed by letters³. However, the use of the more complicated #@ notation pays off immediately when ILEs with multiple isotopic tracers are considered. Using the InChI notation, generalization of transitions is straightforward, without losing readability, as exemplified with the glutamate dehydrogenase *gdhA* converting α-ketoglutarate (AKG) and ammonia (NH₃) to L-glutamate (GLU):

```
<reaction bidirectional="false" id="gdhA">
  <reduct cfg="C#1@1 C#2@1 C#3@1 C#4@1
    C#5@1" id="AKG"/>
  <reduct cfg="N#1@2" id="NH3"/>
  <reduct id="NADPH"/>
  <reduct id="H"/>
  <rproduct cfg="C#1@1 C#2@1 C#3@1 C#4@1
    C#5@1 N#1@2" id="GLU"/>
  <rproduct id="H2O"/>
  <rproduct id="NADP"/>
</reaction>
```

Herein co-factors *NADPH*, *NADP*, *H*, and *H₂O* (i.e., metabolites that do not carry labeled material in the scope of the model) are explicitly specified as reaction partners, a feature that helps to keep FluxML and SBML reaction network representations consistent.

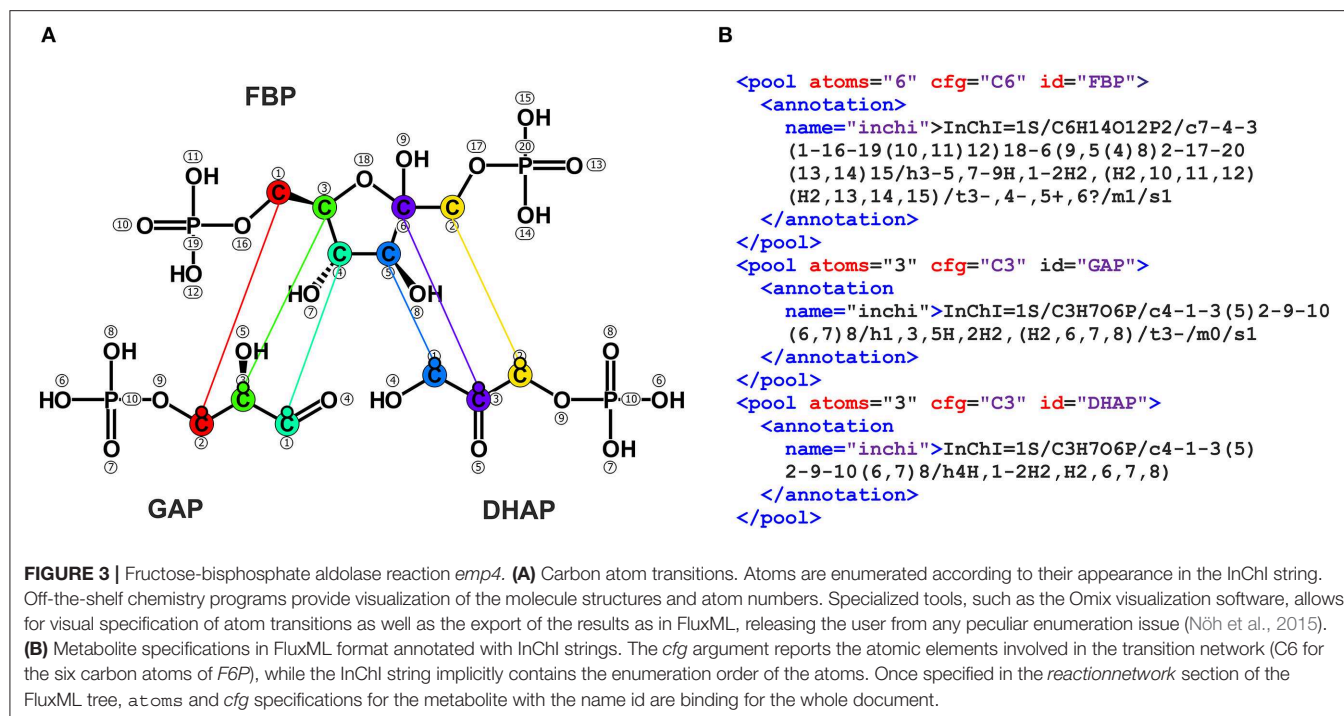
Stoichiometric Constraints

Constraints on the fluxes that impose bounds on the reaction rates on top of the stoichiometric mass balances are important components of any flux model. Typically, such constraints express principled condition-dependent biological or simulation settings. Unfortunately, these equality or inequality relations remain undocumented in ¹³C MFA publications and are, in our experience, a frequent reason why the reproduction of published flux maps fails. Hence, it is vitally important to bundle the complete constraint set together with the model.

An aspect which is conceptually closely related to flux constraints is that of reaction directionality. Here, it often depends on the actual *in vivo* conditions whether a reversible reaction operates in forward and backward direction (*bidirectional*) or only in one of the directions (*unidirectional*). In ¹³C MFA this setting must be carefully considered since it impacts flux inferences. Technically, in purely stoichiometric models bidirectional reactions are split into non-negative forward and backward parts. In ¹³C MFA, however, it is common to use an alternative description for bidirectional reactions, i.e., that of net and exchange fluxes (Wiechert and de Graaf, 1997). Exchange fluxes are net-neutral intracellular material exchanges between reactants (not to be confused with extracellular rates). An advantage of the net/exchange flux system over the backward/forward formulation is that it leads to a “decoupling” of the underlying mathematical equation system for the two flux types, making it easier to express assumptions on both of them.

In FluxML, reaction directionalities are set with the Boolean attribute *bidirectional* = “true” or *bidirectional* = “false” (cf. *gdhA* reaction above). Since net fluxes can take positive and negative values (n.b., exchange fluxes are always non-negative), typical assumptions on net fluxes are “sign” constraints (e.g., $v^{net} \geq 0$) indicating known net flux directions owing to thermodynamic reasoning, upper limits to individual fluxes from enzyme capacity measurements ($v^{net} \leq v^{net}_{max}$), or specific flux ranges ($v^{net}_{min} \leq v^{net} \leq v^{net}_{max}$). Similarly, upper boundaries for exchange fluxes may be applicable for thermodynamic (Wiechert, 2007) or numerical reasons (Theorell et al., 2017). Finally, net and exchange fluxes, respectively, can be related through equality and inequality relations to express further specific relationships

³Notice that a simple permutation transforms the atom positions in the biochemical lettering and the InChI-based canonical enumeration into one another.



such as the rate equalities of scrambling reactions (cf. Section Symmetric (Scrambling) Reactions). The following excerpt gives a typical example:

```

<constraints>
  <net>
    <!-- BM-coeff: [mumol/gCDW], fluxes
                  [mumol/gCDW/s] -->
    <textual>
      Glc_upt=2.38;
      TCA7_v26_1=TCA7_v26_2;
      0=488*mu_v-Ala_bm;
    </textual>
  </net>
  <xch>
    <apply>
      <!-- MathML -->
      <eq/><ci>TCA7_v26_1</ci>
      <ci>TCA7_v26_2</ci>
    </apply>
  </xch>
  <psize>
    <!-- Poolsizes [mumol/gCDW] -->
    <textual>
      ALA<gt;=0.2327;
      ALA<lt;=1.49286;
    </textual>
  </psize>
</constraints>

```

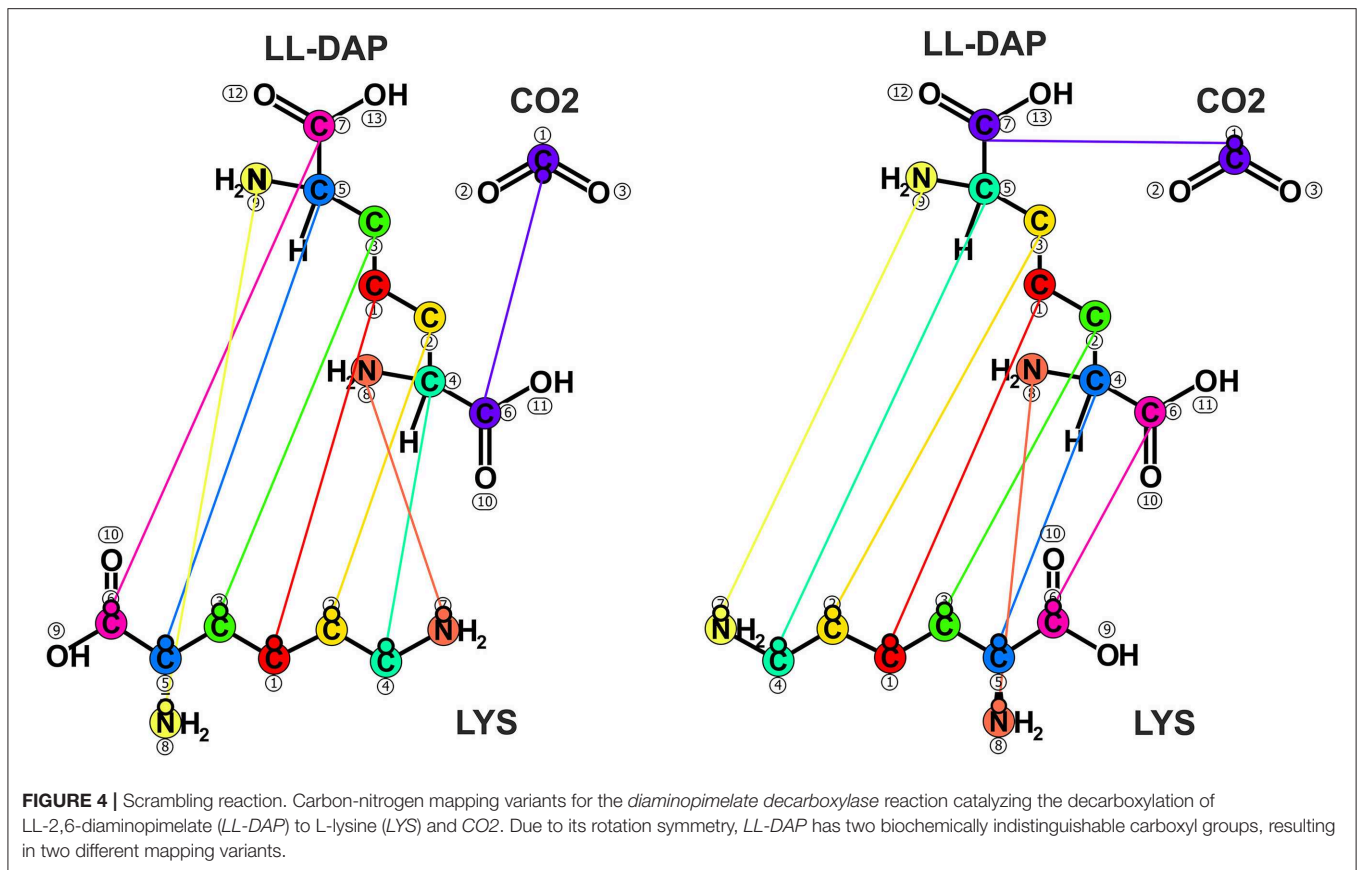
Here the glucose uptake rate (*Glc_upt*) is assigned to a value of 2.38 [μmol/gCDW/s] and net as well as exchange fluxes of the two succinate dehydrogenase reaction variants (*TCA7_v26_1,2*) converting succinate to fumarate are equalized. The third entry encodes a biomass efflux (*Ala_bm*) that is proportional to the cell growth flux (*mu_v*). Importantly, mathematical relations can be expressed in human-friendly text-string representation as well as in Content-MathML [<https://www.w3.org/TR/MathML3/chapter4.html>] (cf. **Supplementary S1 Section 4.1** for an example). Besides the fluxes, pool sizes may also be subject to restrictions. For alanine (*ALA*) a lower and upper boundary is

specified, indicated by the XML entities *>*; (*>*) and *<*; (*<*), respectively.

Symmetric (Scrambling) Reactions

Scrambling reactions constitute a special class of reactions that involve symmetric molecules, i.e., molecules that are biochemically indistinguishable due to their rotational symmetry. For instance, the metabolite LL-2,6-diaminopimelate (LL-DAP), an intermediate of the lysine biosynthesis pathway, contains a rotation axis which gives two symmetric groups (cf. **Figure 4**). In the general case of *n* symmetric groups, *n!* different mapping variants exist, which all have to be specified to describe the emerging labeling patterns correctly.

Technically, any scrambling reaction can be specified as a set of reaction variants, implementing the alternative atom mappings. Here, it is typically assumed that the catalyzing enzyme treats all biochemically indistinguishable isotopomers equally, resulting in identical fluxes of each of the mapping variants. In turn, the associated fluxes are set equal by formulating appropriate equality constraints. Depending on the symmetry level this approach can lead to numerous “virtual” reactions that have to be handled appropriately, also in the post-processing of the results, e.g., the visualization of the flux map. To alleviate the specification process, specific elements (*variant*) and attributes (*ratio*) for modeling scrambling reactions have been introduced to FluxML. The following listing showcases the specification the *diaminopimelate decarboxylase* scrambling reaction *AA13_v49* by means of the variant notation (cf. **Figure 4** and **Supplementary S1 Section 4.2** for the traditional specification):



```
<reaction id="AA13_v49_1 AA13_v49_2">
  <annotation name="pathway">
    Lysine Biosynthesis
  </annotation>
  <annotation name="name">
    AA13_v49
  </annotation>
  <reduct id="LL DAP">
    <variant cfg="C#1@1 C#2@1 C#3@1 C#4@1
      C#5@1 C#6@1 C#7@1 N#8@1 N#9@1"
      ratio="0.5"/>
    <variant cfg="C#1@1 C#3@1 C#2@1 C#5@1
      C#4@1 C#7@1 C#6@1 N#9@1 N#8@1"
      ratio="0.5"/>
  </reduct>
  <rproduct cfg="C#6@1" id="CO2"/>
  <rproduct cfg="C#1@1 C#2@1 C#3@1 C#4@1 C#5@1
    C#7@1 N#8@1 N#9@1" id="LYS"/>
</reaction>
```

Herein, two reaction variants *AA13_v49_1* and *AA13_v49_2* are specified, induced by the symmetry of the educt *LL-DAP*, having fixed equal fluxes (*ratio*="0.5"). Furthermore, the FluxML excerpt shows how elements can be enriched with additional information, e.g., associating the reaction variants to their superordinate reaction (*AA13_v49*) and the pathway names.

Configurations

Experience shows that after an initial set-up phase, ^{13}C MFA evaluation workflows are accompanied by a series of minor model modifications. Here, the majority of differences lie in the settings of constraints, parameter sets and values,

and the composition of data sets, while the model structure itself remains largely untouched. Configurations are created having these experiences in mind. In a *configuration* branch of the FluxML tree, *input*-, *constraints*-, *measurement*-, and *simulation*-settings are bundled, each specific to one ILE or simulation experiment. A FluxML document can then contain an arbitrary number of such configurations.

```
<configuration name="config_0">
  <comment>
    some comment about config_0
  </comment>
  <input pool="input_pool_0"
    type="isotopomer"> ... </input>
  <constraints> ... </constraints>
  <measurement> ... </measurement>
  <simulation method="auto" type="auto">
    <variables>
      <fluxvalue flux="flux_0"
        type="net">2.3</fluxvalue>
      <fluxvalue flux="flux_1"
        type="xch">70.1</fluxvalue>
      <poolvalue pool="pool_0">
        0.36725</poolvalue>
      ...
    </variables>
  </simulation>
</configuration>
<configuration name="config_1"> ...
</configuration>
```

Combined with the reaction network, each single configuration constitutes a complete ^{13}C MFA model. Consequently,

the use of configurations releases the modeler from the necessity to duplicate files beyond necessity and, thus, makes model management more transparent and less error-prone. A typical application scenario where this is enormously useful, are so called parallel ILEs (cf. Section Parallel Labeling Experiments for a worked example). Therewith, configurations are one of the most powerful paradigms of FluxML, as compared to its predecessor FTBL and other modeling languages such as SBML. In the following, the single *configuration* elements are briefly overviewed.

Input Mixture Specification

A broad variety of labeled substrates has been used in ^{13}C MFA, individually or in mixtures, to elucidate metabolic fluxes (Crown et al., 2015; Nöh et al., 2018). Optimal experimental design (OED) heuristics give guidance on the selection of the tracer mixture to maximize the chance of the ILE to be informative about the fluxes. How to select the labeled species for a specific question under study, rather than taking a standard experimental design, is a computational question par excellence (see Section Special Settings for ILE Design). As such, the composition of the substrate pools in terms of labeled species has been subject of various design studies and the OED of ILEs has become a built-in feature of contemporary software systems.

In FluxML, the composition of a substrate labeling is specified in the *input* section by supplying the fractions of the input species present in the substrate pool(s), usually in form of isotopomers. Here, it must be taken into account that neither “unlabeled” nor “labeled” proportions are 100% pure in practice: the abundance of ^{12}C and ^{13}C isotopes (0.9893 and 0.0107, respectively) leads to a natural variation in the isotopomer compositions. In case of naturally labeled substrates, it is sufficient to correct for the variation in each single atom position while neglecting occurrences of combinations of two or more labeled positions (the error due to the occurrence of multiple labeled molecules is below $1.1 \cdot 10^{-4}$ and decreases rapidly with increasing number of labeled positions). As an example, the formulation for ^{12}C glucose is:

```
<input pool="GLC_ext" type="isotopomer">
  <!-- the set has to sum up to 1.0 -->
  <label cfg="000000">0.9375</label>
  <label cfg="000001">0.0107</label>
  <label cfg="000010">0.0107</label>
  <label cfg="000100">0.0107</label>
  <label cfg="001000">0.0107</label>
  <label cfg="010000">0.0107</label>
  <label cfg="100000">0.0107</label>
</input>
```

Commercially available isotopic tracers vary in their isotopic purity in a cost-dependent manner, implying that not only the natural abundance impacts the fractions of the single labeled species, but also the manufacturing and purification quality. In FluxML, the attributes *purity* and *costs* have been created to precisely express these contributions. As an example, a glucose mixture consisting of 77% $[\text{U-}^{13}\text{C}]$ -, 20.5% $[\text{U-}^{13}\text{C}]$ -, and 2.5% $[\text{U-}^{13}\text{C}]$ -glucose is specified in the following succinct way:

```
<input pool="GLC_ext" type="isotopomer">
  <!-- InChI numbering; EUR/g -->
  <label cfg="000001" purity="0.996"
    cost="147.0">0.770</label>
  <label cfg="111111" purity="0.995"
    cost="134.0">0.205</label>
  <label cfg="000000" purity="0.989"
    cost="0.3">0.025</label>
</input>
```

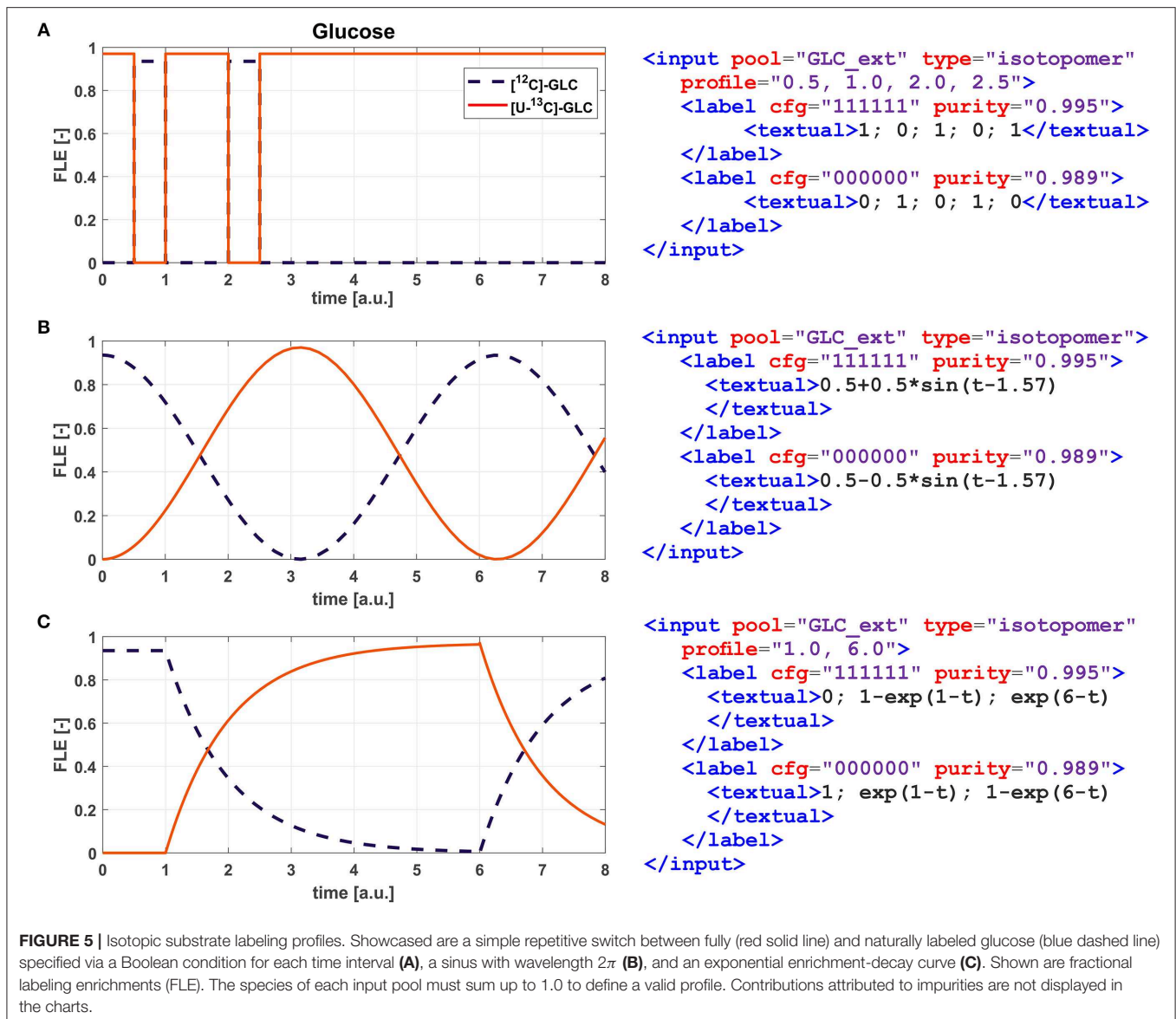
The extension to the multiple-element input substrate specification is then straightforward (cf. **Supplementary S1 Section 4.3**).

For designing ILEs, different substrate sources are mixed with the aim to determine those tracer proportions that are optimally informative about the fluxes. Arbitrary mixtures of labeled substrates are modeled in FluxML by specifying one uptake flux per tracer in the metabolic network. All these uptake fluxes then amount to the total uptake rate of the corresponding substrate which is specified in the constraint section of a FluxML document, for example:

```
<!-- tracer specification -->
<input pool="GLC_ext_12C" type="isotopomer">
  <label cfg="000000" purity="0.989"
    cost="0.04">1</label>
</input>
<input pool="GLC_ext_13C1" type="isotopomer">
  <label cfg="000001" purity="0.996"
    cost="93.0">1</label>
</input>
<input pool="GLC_ext_U13C" type="isotopomer">
  <label cfg="111111" purity="0.995"
    cost="188.0">1</label>
</input>
...
<constraints>
  <net>
    <textual>
      Glc_upt=Glc_upt_12C+
      Glc_upt_13C1+Glc_upt_U13C
    </textual>
  </net>
</constraints>
```

where *Glc_upt* is the total uptake rate and *Glc_upt_12C*, *Glc_upt_13C1*, *Glc_upt_U13C* are the individual uptake rates of naturally $[\text{U-}^{12}\text{C}]$ -, $[\text{U-}^{13}\text{C}]$ -, and fully $[\text{U-}^{13}\text{C}]$ -labeled glucoses, respectively. Uptake fluxes are canonically unidirectional, usually with an extracellular rate assigned (cf. Section Extracellular Rates). For the case that intra- and extracellular metabolites are exchanged, a specification example is given in **Supplementary S1 Section 4.4**.

Principally, the labeling states of the intracellular metabolites depend on the input labeling composition which is usually constantly administered. For INST ILEs such kind of restriction is no longer mandatory, therewith paving the way for the targeted exploitation of dynamic labeling profiles to design highly informative ILEs. Certainly, the most simplistic form of labeling profiles is a repetitive switch between two isotopomer species of a substrate. But also more sophisticated profiles, such as sinusoidal and pulse-width modulated waveforms, have been considered theoretically (Sokol and Portais, 2015). Another scenario, where profiles are of practical value, is when ILEs are conducted under cultivation conditions where the administered carbon source is present in excess. In



FluxML, such labeling profile functions can be flexibly specified (cf. Figure 5).

Specific Constraints

Besides constraints that are inherently linked to the network structure irrespective of the experimental conditions (i.e., globally valid constraints, cf. Section Stoichiometry Constraints), FluxML configurations allow to specify additional *specific* constraints, i.e., those that may only be valid in the context of a concrete experimental setting. For instance, the flux solution space can be tightened by such specific constraints in the context of simulation experiments. Both types of constraints are syntactically equivalent.

Experimental Data

Measurements are an integral part of ^{13}C MFA models, being the basis of flux inference. But also, ^{13}C MFA

codes are tuned for specific measurement types (mostly MS, cf. **Supplementary S1 Table 1.1**). The reason is that the labeling system that is actually needed to describe the sub-set of observable labeling states can be tremendously smaller than the labeling network describing all intracellular labeling states. For the reduction of the high-dimensional labeling systems powerful graph theoretic algorithms have been developed (Weitzel et al., 2007), which are implemented in the high-performance code 13CFLUX2. Consequently, the resulting reduced labeling systems intimately rely on the specific measurement configuration. Notably, the reduction crucially impacts the computational efficiency of flux fitting, rather than the final flux map. Before explaining how the specific measurement setup of an ILE is specified in FluxML, some general remarks on the present measurement equipment are appropriate.

Modeling Data

Measurement models provide a link between the models' state variables and parameters (fluxes, pool sizes in the case of INST) and the observables (extracellular rates, pool sizes, labeling measurements). These three data models are essentially linear, which is trivial to see for the first two types. Therefore, we concentrate on the modeling of the labeling patterns. Consider a metabolite fragment M with n atom positions. Each atom can be present in one of k labeling states ($\{0,1\}$ for ^{12}C , ^{13}C , and ^{14}N , ^{15}N , $\{0,1,2\}$ for ^{16}O , ^{17}O , ^{18}O etc.). For the isotopomer fractions of M then it holds:

$$0 \leq m_{k_1, k_2, \dots, k_n} \leq 1.0 \quad \text{with} \quad \sum_{k_1, k_2, \dots, k_n} m_{k_1, k_2, \dots, k_n} = 1.0 \quad (1)$$

With the isotopomer fractions any labeling measurement is formulated based on the following criteria, which should be obeyed by any well-calibrated measurement procedure:

- Each single isotopomer of M contributes to the spectrum (including a zero contribution).
- All isotopomer contributions superpose linearly.
- The signal intensities scale proportionally with the total amount of the specific isotopomer in a sample.
- This superposition of contributions results in number of distinguishable peaks.
- These peaks can be properly identified.
- Signal intensities are quantified, usually by integrating the respective peak areas.

To make these considerations more concrete, the case of a mass isotopomer distribution (MID) generated in MS is discussed. The MID of an analyte is the vector of fractional labeling enrichments that are derived from the contribution of the single peak areas relative to the sum of all peak areas of the respective analyte. Apart from aspects of pre-processing (cf. Section Design Decision 1—Scope: Data Pre-processing Is Not Part of FluxML), an ideal MS ion chromatogram of a metabolite fragment M with three carbon atoms, contains four distinguishable peaks ($m.0$, $m.1$, $m.2$, $m.3$) to which in total $2^3 = 8$ isotopomers contribute. Precisely, the $M000$ isotopomer contributes to the $m.0$, $M001$, $M010$, $M100$ isotopomers to the $m.1$, $M011$, $M101$, $M110$ isotopomers to the $m.2$, and the $M111$ isotopomer to the $m.3$ peak, respectively. The relation between isotopomers (\mathbf{x}_M) and the MID of M (\mathbf{y}_M) can be represented using matrix notation:

$$\underbrace{\begin{pmatrix} y_M^{m.0} \\ y_M^{m.1} \\ y_M^{m.2} \\ y_M^{m.3} \end{pmatrix}}_{\mathbf{y}_M} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}}_{\mathbf{M}_M} \cdot \underbrace{\begin{pmatrix} m_{000} \\ m_{001} \\ m_{010} \\ m_{011} \\ m_{100} \\ m_{101} \\ m_{110} \\ m_{111} \end{pmatrix}}_{\mathbf{x}_M} \quad (2)$$

with \mathbf{M}_M the MS *measurement matrix* of metabolite fragment M . In principle, this measurement matrix scheme holds true for

other analytical techniques (with different sparsity pattern of the measurement matrix \mathbf{M}_M), as long as the measurements obey the common criteria of good analytical practice described before. Because in Equation (2) appropriate re-scaling of the intensities by (unknown) group-specific scaling factors ω_M may be required to match the simulated enrichments (Wiechert et al., 1999), the general *measurement models* read:

$$\mathbf{y}_M = \omega_M \cdot \mathbf{M}_M \cdot \mathbf{x}_M, \quad \mathbf{y}_M(t_i) = \omega_{M,t_i} \cdot \mathbf{M}_{M,t_i} \cdot \mathbf{x}_M(t_i) \quad (3)$$

for the isotopically stationary and non-stationary cases, respectively. It should be remarked that isotopomer fractions are not the only systematic that can be used for expressing labeling states. Alternatives are cumomers (Möllney et al., 1999), EMUs (Antoniewicz et al., 2007), or tandemers (Tepper and Shlomi, 2015). Since all three labeling systematics can be linearly transformed into isotopomer fractions, the general measurement model formulations given in Equation (3) are equally valid for these alternate frameworks.

Measurement Specification in FluxML

Experimental data are located in the *measurement* branch of the FluxML tree. By design, we distinguish between the *declaration* of the measurements (`<model>`) and the *specification* of the quantitative data (`<data>`):

```
<measurement>
  <mlabel>some notes on the measurements
</mlabel>
<model>
  <fluxmeasurement> ...
</fluxmeasurement>
  <labelingmeasurement> ...
</labelingmeasurement>
  <poolsize measurement> ...
</poolsize measurement>
</model>
<data> ... </data>
</measurement>
```

As in the case of reactions and metabolite pools, each measurement group must be accompanied with a unique identifier (*id*) to unambiguously crosslink the declared reactions and metabolite pools with the specified measured entities (cf. Figure 2).

Extracellular rates (<fluxmeasurement> section, Level 1+)

Flux measurements are essential to any network-wide ^{13}C MFA study. Uptake and secretion fluxes are net rates, specified one-by-one with the following notation:

```
<netflux id="fm_0">
  <textual>glc_upt</textual>
</netflux>
```

On the other hand, FluxML also allows for the formulation of functional relations between model parameters and to equip these with measurements. This feature can be used to incorporate flux ratios, e.g., obtained using FiatFlux or SUMOFLUX (Kogadeeva and Zamboni, 2016):


```
<netratio id="frm_0">
  <!-- flux ratio between glycolysis and
  PPP -->
  <textual>
    emp2/(emp2+ppp1)
  </textual>
</netratio>
```

Isotopic labeling (<labelingmeasurement> section, Level 1+)

The remarks on measurement models above make clear that in practice only one approach works for a universal specification language: The user should be enabled to compose specific measurement configurations from predefined basic expressions (*primitives*) with which more complex measurement specifications can be expressed. These primitives describe (real or envisioned) measurements with concise code fragments. Consequently, in FluxML labeling spectra are composed by linear combinations of measured signals:

1. The most basic primitive specifies a single isotopomer fraction:

```
M#010
```

This means that the isotopomer M010, which carries a labeled atom only at its second atom position, contributes to the measurement matrix. As an extension of the isotopomer notation, a positional atom entry can be marked by an “x” expressing that no information is available for this position or, with other words, any labeling state is allowed. For example,

```
M#01x
```

denotes the set of isotopomers {M010, M011} (if the third atom position of *M* codes for an element with two possible isotopic labeling states). In terms of the measurement models Equation (3) this means that all isotopomers of the set contribute a “1” to the row of the measurement matrix while all other isotopomers lead to a zero entry in \mathbf{M}_M . If only the symbols “1” and “x” are used, the notation coincides with the cumomer notation (Wiechert et al., 1999). Labeling patterns of fragments are identified by the associated atom numbers given in squared brackets, e.g. $M[1-2]\#$. This way, the seven EMUs (moieties comprising any distinct subset of the compound's atoms Antoniewicz et al., 2007) of *M* are represented by $M[1]\#$, $M[2]\#$, $M[3]\#$, $M[1-2]\#$, $M[1,3]\#$, $M[2-3]\#$, and $M[1-3]\#$ (or $M\#$)

2. Apart from these primitives, FluxML contains convenient short-notations for expressing measured signals for a plethora of measurement techniques:

- a. One-dimensional ¹H-NMR generate positional enrichment information:

```
<group id="NMR1H_Ala_23">
  <textual>ALA#P2,3</textual>
</group>
```

Here, the positions $P = 2$ and 3 of the metabolite Alanine (ALA) are specified, coding for isotopomer fractions that are ¹³C labeled at position *P*. Since the two sets of positional isotopomers interfere, they are combined to one measurement group, named *NMR1H_Ala_23*.

- b. Beyond positional observations, two-dimensional ¹³C-NMR can discriminate between certain labeling positions in the direct neighborhood of a ¹³C-labeled position, giving rise to multiplets: peak singlets (*S*) occur, when the focused position is surrounded by unlabeled atoms. Right or left doublets (*DR*, *DL*) emerge if exactly one of the adjacent carbon atoms is labeled with ¹³C. If two surrounding positions are occupied with ¹³C isotopes, double doublets (*DD*) or triplets (*T*) may be obtained. In the following FluxML snippet, two measurement groups of ALA are listed, targeting the second and third carbon position, respectively:

```
<group id="NMR13C_Ala_2">
  <textual>ALA#S2,DL2,DR2,DD2</textual>
</group>
<group id="NMR13C_Ala_3">
  <textual>ALA#S3,DL3</textual>
</group>
```

- c. In MS measurements all isotopomers with the same number of labeled atoms are pooled, resulting in MIDs as exemplified for the C9 metabolite phenylalanine (*PHE*):

```
<group id="MS_Phe" scale="one">
  <textual>
    PHE[1-9]#M0,1,2,3,4,5,6,7,8,9
  </textual>
</group>
```

This measurement group specifies 10 mass isotopomers (m.0,..., m.9) which share a common scaling factor ω , as represented by the *scale* attribute (Möllney et al., 1999). The scale factor is a nuisance parameter that translates between the simulated ([0,1]) and measured ranges of the enrichment data (cf. Equation (3)). This attribute is either *one*, all measurement values of the *PHE* measurement group are taken as specified, or *auto*, meaning that the scale factor is to be determined within the fitting procedure. With FluxML Level 3, also MS data from multiple-isotope tracer experiments can be conveniently specified (cf. **Supplementary S1 Section 4.5** for an example).

- d. Beyond simple MS, tandem MS has proved to be very informative about fluxes, since it can deliver positional information. In FluxML, tandem MS measurements of *PHE* are specified as follows:

```
<group id="MSMS_Phe_2_9" scale="auto">
  <textual>
    PHE[1-9:2-9]#M(0,0),(1,0),
    (1,1),(2,1),(2,2),(3,2),(3,3),
    (4,3),(4,4),(5,4),(5,5),(6,5),
    (6,6),(7,6),(7,7),(8,7),(8,8),
    (9,8)
  </textual>
</group>
```

Here, the first atom range (1–9) refers to the precursor ion, whereas the second range (2–9) relates to the product ion (i.e., the first carbon atom is filtered). The tuples specify the tandem mass isotopomers defined by the precursor and product ion, respectively.

Aside from such shortcuts to specify measurement configurations, the FluxML notation is fully universal because any possible linear measurement combination can be described.

This way, arbitrary setups can be expressed, for instance, a ¹³C-NMR measurement of valine (VAL, cf. **Figure 6**). The flexibility of the FluxML syntax is further demonstrated with the formulation of the *summed fractional labeling*, the sum of the fractional labeling of the atoms contained in a molecule (fragment) (Christensen et al., 2002). The summed fractional labeling can be specified by either using the generalized isotopomer notation:

```
<group id="gen_SFL_Ala_1" scale="auto">
  <textual>
    (ALA#1xx+ALA#x1x+ALA#xx1)/3
  </textual>
</group>
```

or, alternatively, in terms of MS measurements:

```
<group id="gen_SFL_Ala_0" scale="auto">
  <textual>
    (0*ALA#M0+1*ALA#M1+2*ALA#M2+3*ALA#M3)/3
  </textual>
</group>
```

The isotopically non-stationary case (Level 2+): In contrast to classical, isotopically stationary ¹³C MFA where labeling data sets consist of one single labeling measurement vector, in the INST case labeling measurements are time series data. In FluxML, with Level 2 upwards, the measurement time points are introduced as attributes of the measurement groups. This way, time resolved MID of ALA can be formulated as follows:

```
<group id="MS_Ala_0" scale="auto",
  times="0.0,0.1,0.5,1.0,INF">
  <textual>ALA#M0,1,2,3</textual>
</group>
```

expressing that MID of ALA are available at five time points (0.0, 0.1, 0.5, 1.0, ∞). This notation enables joining isotopically stationary and non-stationary data in a single measurement group.

A consideration, which becomes especially important in the INST case, is that fluxes, pool sizes, and time need to be formulated in a coherent physical unit system to produce meaningful results⁴. Metabolic fluxes (and extracellular rates) are amounts of substance transported per time unit. Practically, this means that with the choice of the flux unit, the units of the pool sizes and time are implicitly determined as well. Fluxes are reported in diverse units, e.g., [mmol/g_{CDW}/h], [mmol/L_{Cell}/s], or [nmol/10⁶ cells/h]. Due to this variety, FluxML does not enforce a specific unit system. However, modelers are strongly advised to document the applied units in the FluxML elements *fluxunit*, *poolsizeunit*, and *timeunit*.

Pool sizes (<poolsize measurement> section, Level 2+)

Likewise important for flux inference in the INST case are intercellular pool size data. In FluxML they are specified as follows:

```
<poolsize id="psm_0">
  <textual>ALA</textual>
</poolsize>
<poolsize id="psm_1">
  <textual>RU5P+R5P</textual>
</poolsize>
```

for single and pooled measurements, respectively. Due to the metabolic steady state, pool sizes and extracellular rates remain constant throughout the ILE. Thus, specification of one measurement per entity is appropriate in both cases.

Specification of Experimental Data

To decouple the model structure from the corresponding measurement data is good practice in model-based data evaluation. In FluxML the network formulation and the data descriptors are located in different sub-branches of the document tree (cf. **Figure 2**). This way, model specification and data can be combined in one single document or, alternatively, in two separate files. Principally all measured quantities have to be supplied together with a (strictly positive) measurement error. This measurement uncertainty may refer to data precision or accuracy and cover solely technical or biological uncertainty. Although standard deviations are often based on experiences, such kinds of assumptions are better explicitly documented. In FluxML the description of the measurements and their errors is located within the *data*-branch:

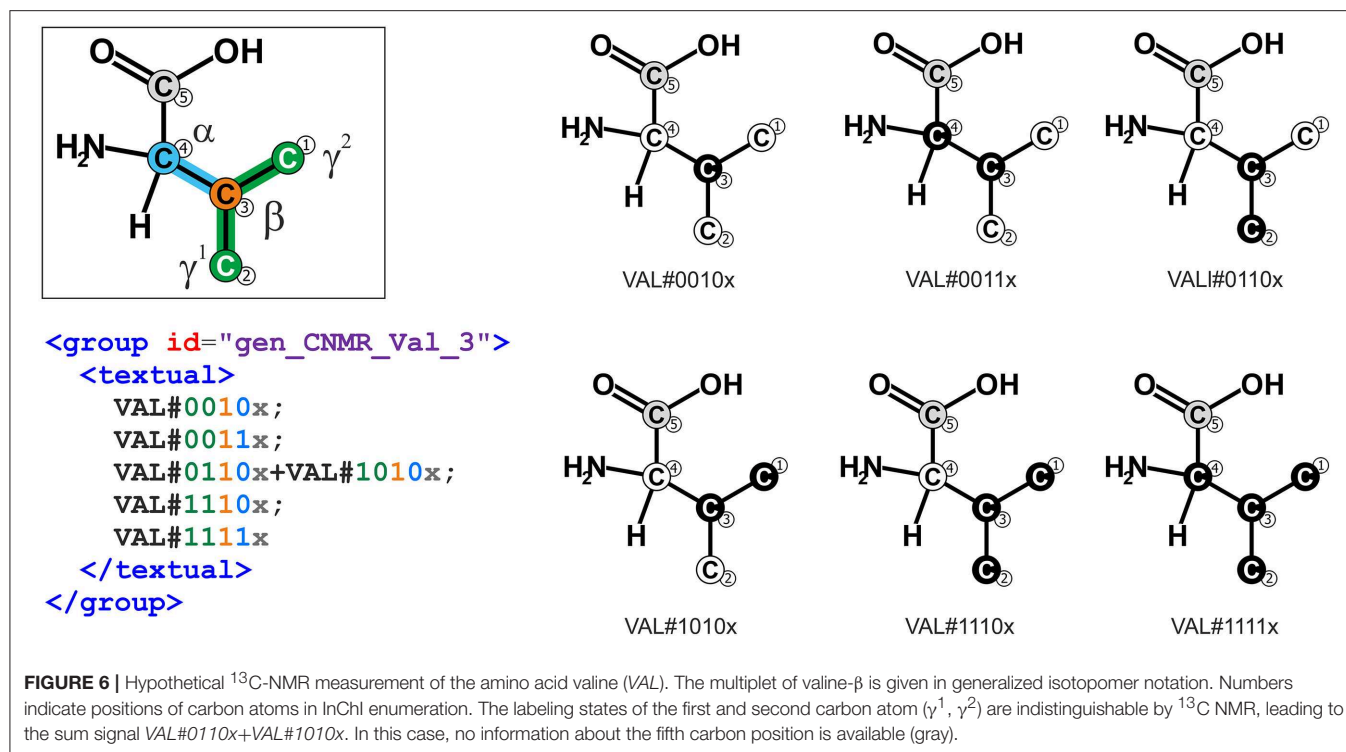
```
<data>
  <dlabel>
    <date>2018-08-30</start>
    <strain>E. coli K12</strain>
    <experiment>Experiment 2018-09-01
  </dlabel>
  ...
  <datum id="MS_Ala_0" stddev="0.01"
    weight="0" time="10.0">0.40669
  </datum>
  <datum id="MS_Ala_0" stddev="0.01"
    weight="1" time="10.0">0.03156
  </datum>
  <datum id="MS_Ala_0" stddev="0.01"
    weight="2" time="10.0">0.30904
  </datum>
  <datum id="MS_Ala_0" stddev="0.01"
    weight="3" time="10.0">0.03579
  </datum>
  ...
  <datum id="fm_0" stddev="0.12">2.39
  </datum>
  ...
  <datum id="psm_0" stddev="0.26">0.78
  </datum>
  ...
</data>
```

Here, the identifier *MS_Ala_0* refers to the time-resolved MID of ALA, distinguished by the number of labeled ions contained (*weight*). Each *datum* element specifies exactly one measurement value along with its associated standard deviation (*stddev*) and sampling time point. The measurements for the uptake rate *Glc_{upt}* (*fm_0*) and the pool size of ALA (*psm_0*) are specified similarly.

Simulation

Whether fluxes and pool sizes are specified as free parameters or being constraint to fixed values impacts flux estimation and

⁴Note that in classical ¹³C MFA the fluxes can be scaled arbitrarily, as long as they share the same units.



the statistical assessment of the final flux map (Heise et al., 2015; Theorell et al., 2017). Thus, the parametrization of the model should also be part of a model. To this end, in FluxML the *variables* element within the simulation branch collects the models' variables (free fluxes and, in case of INST, pool sizes) and their values as well as the minimum information to connect the model description with the simulation framework of choice:

```

<simulation method="auto" type="auto">
  <!-- example: method = cumomer, type = INST -->
  <variables>
    <fluxvalue flux="Glc upt"
      type="net">2.234</fluxvalue>
    <fluxvalue flux="emp6_v6"
      type="xch">1383.3</fluxvalue>
    <poolsizevalue pool="ALA"
      edweight="0.1">0.46</poolsizevalue>
    ...
  </variables>
</simulation>

```

Being designed as a simulator-independent language, details about specific simulation scenarios and settings (solver parametrization, integration times etc.) are intentionally not part of FluxML. For this purpose, scientific workflow, and provenance data description languages have been developed such as CWL [www.commonwl.org/].

Special Settings for ILE Design

OED of ILEs aims at customizing the experimental settings in a way that the ILE's information gain is maximized. As such, optimal ILE design has become an integral part of ¹³C MFA workflows. Many contemporary software systems provide

decision metrics for selecting "informative" tracer mixtures (Möllney et al., 1999; Weitzel et al., 2013; Millard et al., 2014; Shupletsov et al., 2014; Young, 2014). In optimal ILE design, the information gain of an ILE is tested *in silico* by assuming hypothetical experimental-analytical settings. In this context it is important to recognize that OED strategies require not only the measurement model of the envisioned (but physically not yet available) data sets, but also an estimation of their associated standard deviations. Literature mining indicates that errors of labeling enrichments can be heteroscedastic, rather than obeying constant absolute or relative variances (Nöh et al., 2018). For a universal modeling language this implies the need to formulate arbitrary functional dependencies between the "envisioned" measurements and their errors to overcome a lack of real data. How this is solved in FluxML, is exemplified for a tandem-MS measurement group of ALA:

```

<group id="LCMSMS_Ala_3_2">
  <errormodel>
    <!-- one model for all tandem mass
      isotopomers -->
    <textual>
      0.01650*meas_sim+0.0017
    </textual>
  </errormodel>
  <textual>
    ALA[1-3:1-2]#M(0,0),(1,0),(1,1),
    (2,1),(2,2),(3,2)
  </textual>
</group>

```

Within the *errormodel* construct, functional expressions derived from analytical expert knowledge relate the simulated measurement values (*meas_sim*), to their associated errors. In

an analogous manner, error models for extracellular rate and pool size measurements can be formulated.

Besides a full experimental design, scenarios can be envisioned in which some parameters are given a higher importance than others. The importance of a parameter can be specified by the *edweight* attribute in the *variables* section (cf. listing in Section Simulation, where the pool size of *ALA* is given minor importance compared to the two flux values). In this way, partial experimental designs can be realized (Möllney et al., 1999).

Housekeeping: Enriching FluxML With Annotations

Developing a model requires documentation which puts the model into the context of the analysis scenario it is built for. FluxML has various dedicated fields to deposit such kind of information. For instance, the top-level *info* element contains the necessary information to achieve MIRIAM-compliance:

```
<info>
  <date>2018-08-28</date>
  <modeler>info@13cflux.net</modeler>
  <strain>Escherichia coli K12</strain>
  <version>2.0</version>
  <comment>
    Version 1.0 from
    10.1016/j.jymben.2015.01.001
    Version 2.0: N transitions added
    Document license: CC BY-SA 4.0
  </comment>
</info>
```

Owing to the configuration concept, also the *data*-branch contains dedicated elements to carry information about the experiment, analytics and data, such as units. Furthermore, annotation elements can be added to any FluxML element, in which XML-compliant content can be stored. For instance, pathway information is helpful to structure comprehensive models or, associating a reactant with its InChI code enables metabolite identification and database matching.

FLUXML COLLECTION AND SUPPORTING TOOLS

Although human-readable, FluxML documents are not made for direct editing by modelers. Additional software tools are necessary to verify, read, write, and edit the information contained in a document, to display its contents in a digestible form, and to check the documents' syntax and semantics. These tools fall into three categories:

1. The FluxML language definition.
2. A FluxML parser to analyze model files according to the rules laid down in the language definition and to check their syntactic and semantic validity, as well as for completeness.
3. Converters and utility tools providing facilities for convenient access of FluxML files.

FluxML Schema

A language is commonly defined by a formal syntax description (the *grammar*). Accordingly, for each released FluxML Level,

the formal syntax is defined in W3C XML Schemas [http://www.13cflux.net/fluxml]. Each XML Schema Definition (XSD) describes the structure of a FluxML document and defines strict syntax rules for the elements and attributes contained. This grammar definition constitutes the essential basis for checking the well-formedness of model files and, therefore, any further FluxML processing procedure. The checking procedure itself is the task of the FluxML parser.

The FluxML Parser *fmlint*

The parser *fmlint* is an error-detection oriented software tool that analyzes the syntactical and semantical validity of FluxML model files according to the rules defined in the associated FluxML Schema. The parser loads a specified FluxML document, traverses through the tree structure and turns the textual representation into a set of objects, the in-memory Document Object Model (DOM) tree [www.w3.org/DOM]. To this end, *fmlint* uses the capabilities of the DOM XML parser Xerces-C [www.xerces.apache.org] to perform strict validation according to the XSD file. To facilitate precise semantic model validation, in addition to the grammar, an extensive set of semantic rules is implemented in *fmlint*. Thus, with parsing, existing document structure inconsistencies and context-sensitive issues are detected and expressive error messages and warnings are reported, mostly along with specific correction suggestions.

Figure 7 gives an example where the metabolite pool *F*, which participates in the reaction *w*, has been forgotten to be specified. Here, during the parsing process, *fmlint* detects the missing metabolite *F* and reports an error. The error message provides precise information such as the error location (row and column number) which helps to quickly fix the issue. Several examples, typical for erroneous ¹³C MFA models, are given in the **Supplementary S1 Section 3**. Some of the most important validity checks of *fmlint*, specific to ¹³C MFA models are:

- Validation of reaction network, atom transitions and the labeling sources:
 - o Missing labeling sources or effluxes
 - o Dead-end and disconnected metabolites
 - o Traps and isles in the metabolic and atom transition networks
 - o Missing metabolite/reaction declarations or duplicates
 - o Invalid and elementally imbalanced atom transitions
 - o Infeasible/inconsistent input (mixture) specification or purities
- Validation of stoichiometric balances and constraints:
 - o Too few/many equality constraints leading to under- or overdetermined stoichiometry
 - o Duplicate or linearly dependent equality constraints
 - o Infeasible inequality constraints
 - o Too few/many free parameters
 - o Infeasible parameter values violating the set of constraints
- Validation of the measurements:
 - o Missing measurement declarations or duplicates
 - o Invalid measurement specification or duplicates

```

<metabolitepools>
  <pool atoms="2" id="A"/>
  <pool atoms="2" id="B"/>
  <pool atoms="4" id="C"/>
  <pool atoms="3" id="D"/>
  <pool atoms="2" id="E"/>
  <pool atoms="2" id="H"/>
  <pool atoms="1" id="G"/>
</metabolitepools>
...
<reaction bidirectional="false" id="w">
  <reduct cfg="C#1@1 C#2@1 C#3@1 C#4@1" id="C"/>
  <rproduct cfg="C#1@1 C#2@1 C#3@1" id="D"/>
  <rproduct cfg="C#4@1" id="F"/>
</reaction>

```

```
>> fmllint -i network.fml
```

```

XML exception: XML parser (error): ID attribute 'F' is referenced
but was never declared in /home/juflux/fluxml/examples/network.fml;
row: 118, column: 10

```

```

  location: #document/fluxml
node details: #<fluxml xmlns="http://www.13cflux.net/fluxml"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://www.13cflux.net/fluxml
http://www.13cflux.net/fluxml"/>

```

```

XML exception: element "rproduct": metabolite "F" is not specified
in "metabolitepools"

```

```

  location: #document/fluxml/reactionnetwork/reaction/rproduct
node details: #<rproduct cfg="C#4@1" id="F"/>
XML exception: [element "rproduct": metabolite "F" is not
specified in "metabolitepools"] in reaction "w"

```

```

FluxML parsing error: [element "rproduct": metabolite "F" is not
specified in "metabolitepools"] in reaction "w"

```

FIGURE 7 | Error backtracking provided by the FluxML parser *fmlint*. In the FluxML model network.fml, the metabolite pool *F*, which participates in the reaction *w* is not declared in the metabolitepools section. *fmlint* provides an expressive error log, pointing to the origin of the error.

- Missing and infeasible values for labeling fractions, pool sizes or measurement times
- Validation of the FluxML structure:
 - Missing or invalid XML namespace
 - Well-formedness of textual- or MathML notations
 - Missing or invalid element nodes
 - Invalid attributes or attribute combinations

Currently, in total more than 500 different errors are detected. This number emphasizes the complexity of model specification and the critical importance for having a concise and clean language standard definition. It also demonstrates the complexity and power of the *fmlint* parser which is written in ANSI/ISO C/C++ and consists at the time of writing of more than 70k LOCs (lines of code). The implementation solely depends on standard ANSI C libraries only and is, thus, highly portable.

Converters and Utilities

Tools for Model Reuse

To enable the effective reuse of models was the main driver for developing the FluxML language. To support this goal, the following language translators are supplied:

- **ftbl2fml**: The FTBL-to-FluxML converter conveniently transfers the tabular-separated FTBL format to FluxML Level 1. The converter is implemented in Python and uses the C++ program *expr2mm1* to analyze the equality- and inequality-constraints. By employing the **ftbl2fml** converter, only minimal steps are needed to transform models built for the ¹³C MFA tools such as *influx_s* and *OpenFLUX* into FluxML.
- **fmlupdate**: Modeling languages evolve over time. Therefore, it is important to support modelers with handling changes in language constructs. The FluxML update tool **fmlupdate** transforms FluxML documents to new versions. This makes model reuse convenient for end-users when updates of simulation tools inquire newer FluxML versions.
- **sbml2fml**, **fml2sbml**: Due to their shared language subset, part of the metabolic network and flux constraints can be translated between FluxML and SBML. For easy translation of such network structures, the Python-based converters **sbml2fml** and **fml2sbml** have been developed.

Auxiliary Tools for Everyday Operations

From a users' perspective graphical tools for model building and configuration are preferable. To this end, the comprehensive Fluxomix modeling suite has been developed as plugin-suite for the visualization software Omix (Nöh et al., 2015). However, for integrating model configuration procedures into computational evaluation workflows, programmatic model access is much more convenient than visual modeling. To support commonly performed steps, tools that have been initially released with the 13CFLUX2 software suite, are lifted as standalone Python tools:

- **fmlstats**: Summarizes the most important information about the model structure
- **setinputs**: Tool for manipulation of the mixture composition of input pools
- **setparameters**: Tool to transfer fluxes and/or pool sizes from CSV into FluxML files
- **setmeasurements**: Tool for transferring labeling and flux measurements from CSV documents into FluxML files

A Software Library for FluxML Tool Developers

The software library **libFluxML** is a library for reading, writing and altering FluxML documents. The library provides a rich application programming interface (API) enabling full access to the FluxML language content and a range of functions that facilitate the creation, validation, and manipulation of FluxML documents. **libFluxML** offers helper functions for processing and manipulating mathematical formulas in both, human-readable textual notation and machine-readable Content-MathML format, as well as the ability to interconvert mathematical expressions between these forms. Many higher-level convenience features are included, such as for obtaining the number of reactions or constructing the stoichiometric matrix

of the reaction network. The library is written in standard ANSI/ISO C/C++ and uses the FluxML parser **fmlint** for parsing and validity checking.

Availability

The FluxML collection consisting of the formal schema definitions, the **fmlint** parser, versatile tools and the core library **libFluxML** represents an all-inclusive suite to validate and manipulate FluxML documents. Schema files are located at <http://www.13cflux.net/fluxml>. The source codes of the FluxML parser **fmlint**, the **libFluxML** library, and the auxiliary tools are available at the github repository <https://github.com/modsim/FluxML/> with full built instructions, comprehensive documentation and usage examples. In addition, precompiled binary distributions for Linux and Mac OS X are provided. The FluxML collection is licensed under the open-source Creative Commons Attribution-ShareAlike (CC BY-SA 4.0)⁵ and MIT⁶ licenses. In addition, for model checking without installation, a web-based FluxML validator is available at <http://www.13cflux.net/fluxml/validator/>. Altogether, this collection provides a set of tools for interfacing and validating FluxML documents and, as such, provides a solid tool base for future developments of the modeling language FluxML.

HARNESSING THE BENEFITS OF FLUXML

Finally, we give two examples for the utility and usability of the FluxML language. We first illustrate how using one single model, formulated in FluxML, can be used with different ¹³C MFA tools to facilitate the comparison of results. Secondly, we demonstrate how parallel ILEs are efficiently modeled starting from a single ILE setup.

FluxML for Simulator Comparisons

From a users' perspective, the lack of abilities to compare and validate numerical results generated by different ¹³C MFA tools is unsatisfactory. Clearly, a precise and unambiguous representation of a model provides the basis for any of these tasks. Extracting the encoding of a model formulated for one piece of software and transferring it to another format is a step prone to errors that should be subjected to converters. Here, we exemplify a simulator comparison, taking the deterministic forward simulation step with 13CFLUX2 (v2.0) and Sysmetab (v5.1, Mottelet et al., 2017) as representative test case. The comparison is done with a central metabolism model of *E. coli* contained in the Sysmetab distribution, precisely, a isotopically stationary and non-stationary variant mimicking ILEs with a 3:7 [U-¹³C]: [1-¹³C]-glucose mixture. The **fmlstats** tool reports that the network consists of 51 metabolites and 86 reactions. In total 9 MS measurement groups and one extracellular flux measurement are contained.

In the classic isotopically stationary case, the corresponding Sysmetab FluxML was conform with the FluxML Level 1 definition. Both simulators were invoked and simulated labeling

⁵<https://creativecommons.org/licenses/by-sa/4.0/legalcode>

⁶<https://opensource.org/licenses/mit-license.php>

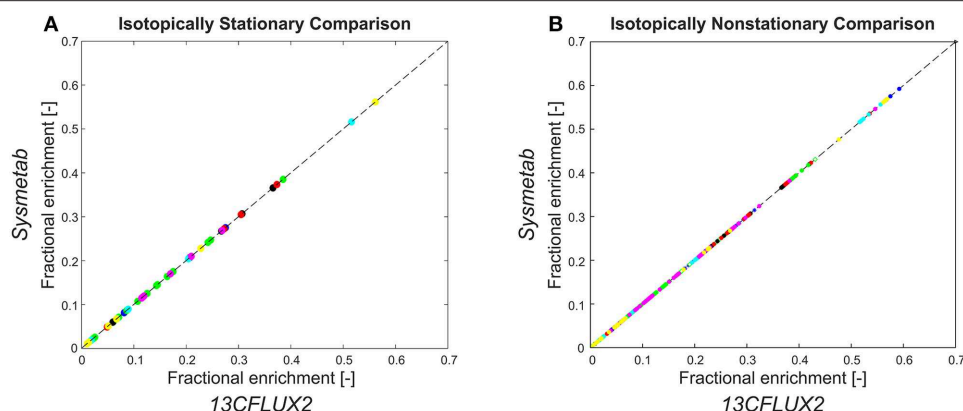


FIGURE 8 | Parity plots showing simulation results generated by 13CFLUX2 and Sysmetab. Classical, isotopically stationary approach **(A)** and INST ^{13}C MFA approach **(B)**. One color is used per measurement group. For comparability, both simulators were configured to use the same labeling framework, i.e., the reduced cumomers (Weitzel et al., 2007). Both simulators use sparse LU factorization for solving the cascaded cumomer systems. The differential equation systems were solved using Sundial's CVODE solver [https://computation.llnl.gov/projects/sundials/cvode/] (13CFLUX2) and a 4th order singly diagonally implicit Runge-Kutta method (Sysmetab). Both solvers used adaptive stepsize control with error tolerance (10^{-6}). Simulation results and scripts for generating the parity plots are available in the **Supplementary Data Sheet 2**.

patterns extracted from the tools' output. The comparison of the simulated fractional enrichments shows perfect agreement (cf. **Figure 8A**). For the isotopically non-stationary case, it turned out that the model shipped with Sysmetab lacks pool sizes and is, thus, syntactically invalid with respect to the Level 2 specification for INST ^{13}C MFA. Since Sysmetab internally allocates positive random values to the pool sizes in the simulation step, these model parameters had to be extracted from the simulation output. After updating the INST FluxML model to conform with the Level 2 Schema definition using the `fmlupdate` tool, the pool sizes were incorporated into the file utilizing the `setparameters` command. Lastly, the 13CFLUX2 simulator was called to execute the simulation step. Again, the simulated MIDFs for the nine measurement groups produced by the two simulators were extracted from the output files and plotted in a parity plot, showing excellent agreement (cf. **Figure 8B**). This example shows the importance of syntactic model validation in view of reporting standards. Besides a clear and complete language definition, appropriate converters, and auxiliary tools are needed to tame the zoo of available model files, often developed specific to ^{13}C MFA software systems.

Parallel Labeling Experiments

Experimental design has been an essential part of the general flux analysis workflow since the beginnings of ^{13}C MFA. As such, numerous studies investigated how specific experimental configurations, predominantly the input tracers, and substrate mixtures, but also the number, type and quality of measurements influence the statistical quality of the flux estimates. To increase the information gain about fluxes, it has been suggested to use multiple experiments operated under exactly the same physiological conditions, each with a different tracer (Schwender

et al., 2006; Antoniewicz, 2015). The evaluation of such, so called, parallel ILEs (pILE) requires the modeler to merge all data sets in one measurement specification. By expressing pILEs in the FluxML language, their simulation can be readily achieved by employing standard ^{13}C MFA tools. In particular, we show that the evaluation of pILEs becomes a special case of the traditional single ILE-based ^{13}C MFA.

The general principle is in fact simple: When N different experiments are performed (in practice usually up to tens), one option is that the modeler supplies the original network formulation in N multiple copies. In each of the copies all metabolites and reactions are multiplied (practically by appending their identifiers with an additional suffix relating them to the experiment to which their associated measurement sets). In addition, for each flux (and pool size, in the case of INST) an additional constraint must be specified in the FluxML document which assures that the values of the model parameters are the same for all network copies⁷. Clearly, to perform these operations manually is laborious and means to pay painstaking attention to details.

By using the available FluxML capabilities, automation of this operation is straightforward. To this end, the program `multiply_fml` was implemented with only 400 single LOCs (SLOC) of Python code. `multiply_fml` expects an FluxML file with N configurations each equipped with the input specification and a corresponding measurement set for one ILE. The duplication process is showcased with 14 different ILEs with the setting reported by Crown et al. (2015). First, the different input mixtures are specified one-at-a-time in 14 different configurations of the reference network by invoking the `setinputs` function:

⁷Another option is to double the metabolites and add the pool copies to the reactions, along with the formulation of pool size constraints in case of INST.

```
setinputs -i crown.fml -c config_01 -C
input_01.csv -o crown.fml
```

etc. Secondly, the measurement data sets are sequentially incorporated by using the **setmeasurements** tool:

```
setmeasurements -i crown.fml -c config_01
-C data_01.csv -o crown.fml
```

Finally, the network duplication step is performed using the **multiply_fml** program:

```
multiply_fml -i
crown.fml -o crown_multiplied.fml
```

The resulting model file (*crown_multiplied.fml*) consists of one network description and a single configuration comprising all 14 ILE data sets (all FluxML and CSV files used in this showcase are available in the **Supplementary Data Sheet 3**). With this model at hand, all ¹³C MFA tools can be invoked. For example, optimal tracer design for a series of pILEs is possible, rephrased as the choice of the best substrates per experiment. This makes application of experimental design tools of ¹³C MFA software straightforward.

CONCLUSION

¹³C MFA is the primary experimental technique for measuring intracellular fluxes at metabolic pseudo-steady state conditions. After two decades of active research there is consensus about the minimal information set needed to specify a computational ¹³C model and its associated data. However, this consensus has not yet found its way into a model format that contains the complete information set of an ILE configuration in a well-structured manner. Most importantly, implicit assumptions made in the modeling process are rarely included in publications because they are considered to be common sense or of purely technical nature. This makes it essentially impossible to reproduce many published flux analysis results.

On the one hand, the complexity and depth of ILE specifications should not hinder experimentalists to deliver complete ¹³C MFA models. In this context, it is of great advantage that tailored model templates can be configured (often only once for an organism or strain) whereas experiment specific data is fed into these templates using preconfigured scripts. For power users, on the other hand, computational model components should

be programmatically accessible, so that they are embeddable in computational pipelines.

Following these two guiding principles, here we describe the Flux Markup Language FluxML along with its design. The major aim of FluxML is to offer a sound universal, open source, simulator-independent, and future-proof platform that conserves all the necessary and optional information for model description, reuse, exchange and comparison. Specifically, FluxML enables practitioners to describe valid isotopically stationary and non-stationary models, while the format is fully universal in term of network, atom mapping, measurement (error) and constraint formulation, including the use of homo- and hetero-isotopic tracers. With the language, the FluxML collection is supplied which contains the powerful FluxML parser **fmlint** for model (in)validation and several auxiliary tools for easy handling, but also to allow for a maximum of flexibility. We believe that FluxML improves scientific productivity, efficiency as well as transparency and contributes to the reproducibility of computational modeling efforts in the field of ¹³C MFA.

AUTHOR CONTRIBUTIONS

WW and KN conceived the work. MW developed FluxML Level 1. SA and MB developed Levels 2/3 and performed the computational analyses. WW and KN wrote the manuscript to which SA and MB contributed. All authors approved the content of the manuscript.

FUNDING

MW was funded by Deutsche Forschungsgemeinschaft (DFG) Grant WI 1705/12-1.

ACKNOWLEDGMENTS

The authors thank Peter Droste for insightful discussions.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.01022/full#supplementary-material>

REFERENCES

- Allen, D. K. (2016). Assessing compartmentalized flux in lipid metabolism with isotopes. *Biochim. Biophys. Acta Mol. Cell Biol. Lipids* 1861, 1226–1242. doi: 10.1016/j.bbalip.2016.03.017
- Antoniewicz, M. R. (2015). Parallel labeling experiments for pathway elucidation and ¹³C metabolic flux analysis. *Curr. Opin. Biotechnol.* 36, 91–97. doi: 10.1016/j.copbio.2015.08.014
- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2007). Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* 9, 68–86. doi: 10.1016/j.ymben.2006.09.001
- Beste, D., Nöh, K., Niedenführ, S., Mendum, T., Hawkins, N., Ward, J., et al. (2013). ¹³C-flux spectral analysis of host-pathogen metabolism reveals a mixed diet for intracellular *Mycobacterium tuberculosis*. *Chem. Biol.* 20, 1012–1021. doi: 10.1016/j.chembiol.2013.06.012
- Birkel, G. W., Ghosh, A., Kumar, V. S., Weaver, D., Ando, D., Backman, T. W. H., et al. (2017). The JBEI quantitative metabolic modeling library (jQMM): a python library for modeling microbial metabolism. *BMC Bioinformatics* 18:205. doi: 10.1186/s12859-017-1615-y
- Blank, L. M., Desphande, R. R., Schmid, A., and Hayen, H. (2012). Analysis of carbon and nitrogen co-metabolism in yeast by ultrahigh-resolution mass spectrometry applying ¹³C- and ¹⁵N-labeled substrates simultaneously. *Anal. Bioanal. Chem.* 403, 2291–2305. doi: 10.1007/s00216-012-6009-4
- Borkum, M. I., Reardon, P. N., Taylor, R. C., and Isern, N. G. (2017). Modeling framework for isotopic labeling of heteronuclear moieties. *J. Cheminform.* 9:14. doi: 10.1186/s13321-017-0201-7

- Cheah, Y. E., and Young, J. D. (2018). Isotopically nonstationary metabolic flux analysis (INST-MFA): putting theory into practice. *Curr. Opin. Biotechnol.* 54, 80–87. doi: 10.1016/j.copbio.2018.02.013
- Choi, J., Grossbach, M. T., and Antoniewicz, M. R. (2012). Measuring complete isotopomer distribution of aspartate using gas chromatography/tandem mass spectrometry. *Anal. Chem.* 84, 4628–4632. doi: 10.1021/ac300611n
- Christensen, B., Gombert, A. K., and Nielsen, J. (2002). Analysis of flux estimates based on ¹³C-labelling experiments. *Eur. J. Biochem.* 269, 2795–2800. doi: 10.1046/j.1432-1033.2002.02959.x
- Christensen, B., and Nielsen, J. (1999). Isotopomer analysis using GC-MS. *Metab. Eng.* 1, 282–290. doi: 10.1006/mben.1999.0117
- Chu, D. B., Troyer, C., Mairinger, T., Ortmayr, K., Neubauer, S., Koellensperger, G., et al. (2015). Isotopologue analysis of sugar phosphates in yeast cell extracts by gas chromatography chemical ionization time-of-flight mass spectrometry. *Anal. Bioanal. Chem.* 407, 2865–2875. doi: 10.1007/s00216-015-8521-9
- Cinquemani, E., Laroute, V., Coccagn-Bousquet, M., de Jong, H., and Ropers, D. (2017). Estimation of time-varying growth, uptake and excretion rates from dynamic metabolomics data. *Bioinformatics* 33, i301–i310. doi: 10.1093/bioinformatics/btx250
- Crown, S. B., and Antoniewicz, M. R. (2013). Publishing ¹³C metabolic flux analysis studies: a review and future perspectives. *Metab. Eng.* 20, 42–48. doi: 10.1016/j.ymben.2013.08.005
- Crown, S. B., Long, C. P., and Antoniewicz, M. R. (2015). Integrated ¹³C-metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. *Metab. Eng.* 28, 151–158. doi: 10.1016/j.ymben.2015.01.001
- Dai, Z., and Locasale, J. W. (2017). Understanding metabolism with flux analysis: from theory to application. *Metab. Eng.* 43, 94–102. doi: 10.1016/j.ymben.2016.09.005
- Dalman, T., Doernemann, T., Juhnke, E., Weitzel, M., Smith, M., Wiechert, W., et al. (2010). “Metabolic flux analysis in the cloud,” in *2010 IEEE 6th International Conference on e-Science* (Los Alamitos, CA: IEEE), 57–64. doi: 10.1109/eScience.2010.20
- Dalman, T., Wiechert, W., and Nöh, K. (2016). A scientific workflow framework for ¹³C metabolic flux analysis. *J. Biotechnol.* 232, 12–24. doi: 10.1016/j.jbiotec.2015.12.032
- Delp, J., Gutbier, S., Cerff, M., Zasada, C., Niedenführ, S., Zhao, L., et al. (2018). Stage-specific metabolic features of differentiating neurons: implications for toxicant sensitivity. *Toxicol. Appl. Pharmacol.* 354, 64–80. doi: 10.1016/j.taap.2017.12.013
- Ebert, B. E., Lamprecht, A. L., Steffen, B., and Blank, L. M. (2012). Flux-P: automating metabolic flux analysis. *Metabolites* 2, 872–890. doi: 10.3390/metabo2040872
- Fernandez, C. A., Des Rosiers, C., Previs, S. F., David, F., and Brunengraber, H. (1996). Correction of ¹³C mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.* 31, 255–262. doi: 10.1002/(SICI)1096-9888(199603)31:3<255::AID-JMS290>3.0.CO;2-3
- Gebreselassie, N. A., and Antoniewicz, M. R. (2015). ¹³C-metabolic flux analysis of co-cultures: a novel approach. *Metab. Eng.* 31, 132–139. doi: 10.1016/j.ymben.2015.07.005
- Ghosh, A., Nilmeier, J., Weaver, D., Adams, P. D., Keasling, J. D., Mukhopadhyay, A., et al. (2014). A peptide-based method for ¹³C metabolic flux analysis in microbial communities. *PLoS Comput. Biol.* 10:e1003827. doi: 10.1371/journal.pcbi.1003827
- Giraudeau, P., Cahoreau, E., Massou, S., Pathan, M., Portais, J.-C., and Akoka, S. (2012). UFJCOSEY: a fast 3D NMR method for measuring isotopic enrichments in complex samples. *Chemphyschem* 13, 3098–3101. doi: 10.1002/cphc.201200255
- Giraudeau, P., Massou, S., Robin, Y., Cahoreau, E., Portais, J. C., and Akoka, S. (2011). Ultrafast quantitative 2D NMR: an efficient tool for the measurement of specific isotopic enrichments in complex biological mixtures. *Anal. Chem.* 83, 3112–3119. doi: 10.1021/ac200007p
- Gleeson, P., Crook, S., Cannon, R. C., Hines, M. L., Billings, G. O., Farinella, M., et al. (2010). NeuroML: a language for describing data driven models of neurons and networks with a high degree of biological detail. *PLoS Comput. Biol.* 6:e1000815. doi: 10.1371/journal.pcbi.1000815
- Gopalakrishnan, S., and Maranas, C. D. (2015). ¹³C metabolic flux analysis at a genome-scale. *Metab. Eng.* 32, 12–22. doi: 10.1016/j.ymben.2015.08.006
- Gopalakrishnan, S., Pakrasi, H. B., and Maranas, C. D. (2018). Elucidation of photoautotrophic carbon flux topology in *Synechocystis* PCC 6803 using genome-scale carbon mapping models. *Metab. Eng.* 47, 190–199. doi: 10.1016/j.ymben.2018.03.008
- Hadadi, N., Hafner, J., Soh, K. C., and Hatzimanikatis, V. (2017). Reconstruction of biological pathways and metabolic networks from *in silico* labeled metabolites. *Biotechnol. J.* 12:1600464. doi: 10.1002/biot.2016.00464
- He, L., Wu, S. G., Zhang, M., Chen, Y., and Tang, Y. J. (2016). WUFlux: an open-source platform for ¹³C metabolic flux analysis of bacterial metabolism. *BMC Bioinformatics* 17:444. doi: 10.1186/s12859-016-1314-0
- Heise, R., Fernie, A. R., Stitt, M., and Nikoloski, Z. (2015). Pool size measurements facilitate the determination of fluxes at branching points in non-stationary metabolic flux analysis: the case of *Arabidopsis thaliana*. *Front. Plant Sci.* 6:386. doi: 10.3389/fpls.2015.00386
- Heller, S. R., McNaught, A., Pletnev, I., Stein, S., and Tchekhovskoi, D. (2015). InChI, the IUPAC International Chemical Identifier. *J. Cheminform.* 7:23. doi: 10.1186/s13321-015-0068-4
- Heux, S., Bergès, C., Millard, P., Portais, J.-C., and Létisse, F. (2017). Recent advances in high-throughput ¹³C-fluxomics. *Curr. Opin. Biotechnol.* 43, 104–109. doi: 10.1016/j.copbio.2016.10.010
- Hucka, M., Finney, A., Sauro, H. M., Bolouri, H., Doyle, J. C., Kitano, H., et al. (2003). The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models. *Bioinformatics* 19, 524–531. doi: 10.1093/bioinformatics/btg015
- Jungreuthmayer, C., Neubauer, S., Mairinger, T., Zanghellini, J., and Hann, S. (2016). ICT: isotope correction toolbox. *Bioinformatics* 32, 154–156. doi: 10.1093/bioinformatics/btv514
- Kale, N. S., Haug, K., Conesa, P., Jayseelan, K., Moreno, P., Rocca-Serra, P., et al. (2016). MetaboLights: an open-access database repository for metabolomics data. *Curr. Protoc. Bioinforma.* 2016, 14.13.1–14.13.18. doi: 10.1002/0471250953.bi1413s53
- Kappelmann, J., Klein, B., Geilenkirchen, P., and Noack, S. (2017). Comprehensive and accurate tracking of carbon origin of LC-tandem mass spectrometry collisional fragments for ¹³C-MFA. *Anal. Bioanal. Chem.* 409, 2309–2326. doi: 10.1007/s00216-016-0174-9
- Kleijn, R. J., van Winden, W. A., van Gulik, W. M., and Heijnen, J. J. (2005). Revisiting the ¹³C-label distribution of the non-oxidative branch of the pentose phosphate pathway based upon kinetic and genetic evidence. *FEBS J.* 272, 4970–4982. doi: 10.1111/j.1742-4658.2005.04907.x
- Kogadeeva, M., and Zamboni, N. (2016). SUMOFLUX: a generalized method for targeted ¹³C metabolic flux ratio analysis. *PLoS Comput. Biol.* 12:e1005109. doi: 10.1371/journal.pcbi.1005109
- Kumar, A., and Maranas, C. D. (2014). CLCA: maximum common molecular substructure queries within the MetRxn database. *J. Chem. Inf. Model.* 54, 3417–3438. doi: 10.1021/ci5003922
- Kumar, A., Suthers, P. F., and Maranas, C. D. (2012). MetRxn: a knowledgebase of metabolites and reactions spanning metabolic models and databases. *BMC Bioinformatics* 13:6. doi: 10.1186/1471-2105-13-6
- Lane, A. N., Fan, T. W., and Higashi, R. M. (2008). Isotopomer-based metabolomic analysis by NMR and mass spectrometry. *Methods Cell Biol.* 84, 541–588. doi: 10.1016/S0091-679X(07)84018-0
- Le Novère, N., Finney, A., Hucka, M., Bhalla, U. S., Campagne, F., Collado-Vides, J., et al. (2005). Minimum information requested in the annotation of biochemical models (MIRIAM). *Nat. Biotechnol.* 23, 1509–1515. doi: 10.1038/nbt1156
- Lee, W. N., Byerley, L. O., Bergner, E. A., and Edmond, J. (1991). Mass isotopomer analysis: theoretical and practical considerations. *Biol. Mass Spectrom.* 20, 451–458. doi: 10.1002/bms.1200200804
- Llaneras, F., and Picó, J. (2007). A procedure for the estimation over time of metabolic fluxes in scenarios where measurements are uncertain and/or insufficient. *BMC Bioinformatics* 8:421. doi: 10.1186/1471-2105-8-421
- Lloyd, C. M., Halstead, M. D. B., and Nielsen, P. F. (2004). CellML: its future, present and past. *Prog. Biophys. Mol. Biol.* 85, 433–450. doi: 10.1016/j.pbiomolbio.2004.01.004
- Luo, B., Groenke, K., Takors, R., Wandrey, C., and Oldiges, M. (2007). Simultaneous determination of multiple intracellular metabolites in glycolysis, pentose phosphate pathway and tricarboxylic acid cycle by

- liquid chromatography-mass spectrometry. *J. Chromatogr. A* 1147, 153–164. doi: 10.1016/j.chroma.2007.02.034
- Mairinger, T., and Hann, S. (2017). Implementation of data-dependent isotopologue fragmentation in ^{13}C -based metabolic flux analysis. *Anal. Bioanal. Chem.* 409, 3713–3718. doi: 10.1007/s00216-017-0339-1
- Mairinger, T., Wegscheider, W., Peña, D. A., Steiger, M. G., Koellensperger, G., Zanghellini, J., et al. (2018). Comprehensive assessment of measurement uncertainty in ^{13}C -based metabolic flux experiments. *Anal. Bioanal. Chem.* 410, 3337–3348. doi: 10.1007/s00216-018-1017-7
- Marx, A., de Graaf, A. A., Wiechert, W., Eggeling, L., and Sahm, H. (1996). Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. *Biotechnol. Bioeng.* 49, 111–129. doi: 10.1002/(SICI)1097-0290(19960120)49:2<111::AID-BIT1>3.0.CO;2-T
- McCloskey, D., Young, J. D., Xu, S., Palsson, B. O., and Feist, A. M. (2016a). MID Max: LC-MS/MS method for measuring the precursor and product mass isotopomer distributions of metabolic intermediates and cofactors for metabolic flux analysis applications. *Anal. Chem.* 88, 1362–1370. doi: 10.1021/acs.analchem.5b03887
- McCloskey, D., Young, J. D., Xu, S., Palsson, B. O., and Feist, A. M. (2016b). Modeling method for increased precision and scope of directly measurable fluxes at a genome-scale. *Anal. Chem.* 88, 3844–3852. doi: 10.1021/acs.analchem.5b04914
- Millard, P., Sokol, S., Letisse, F., and Portais, J.-C. (2014). IsoDesign: a software for optimizing the design of ^{13}C -metabolic flux analysis experiments. *Biotechnol. Bioeng.* 111, 202–208. doi: 10.1002/bit.24997
- Möllney, M., Wiechert, W., Kownatzki, D., and de Graaf, A. A. (1999). Bidirectional reaction steps in metabolic networks. IV. Optimal design of isotopomer labeling experiments. *Biotechnol. Bioeng.* 66, 86–103. doi: 10.1002/(SICI)1097-0290(1999)66:2<86::AID-BIT2>3.0.CO;2-A
- Moseley, H. N. B., Lane, A. N., Belshoff, A. C., Higashi, R. M., and Fan, T. W. M. (2011). A novel deconvolution method for modeling UDP-N-acetyl-D-glucosamine biosynthetic pathways based on ^{13}C mass isotopologue profiles under non-steady-state conditions. *BMC Biol.* 9:37. doi: 10.1186/1741-7007-9-37
- Mottelet, S., Gaullier, G., and Sadaka, G. (2017). Metabolic flux analysis in isotope labeling experiments using the adjoint approach. *IEEE ACM Trans. Comput. Biol. Bioinform.* 14, 491–497. doi: 10.1109/TCBB.2016.2544299
- Murphy, T. A., and Young, J. D. (2013). ETA: robust software for determination of cell specific rates from extracellular time courses. *Biotechnol. Bioeng.* 110, 1748–1758. doi: 10.1002/bit.24836
- Niedenführ, S., ten Pierick, A., van Dam, P. T. N., Suarez-Mendez, C. A., Nöh, K., and Wahl, S. A. (2016). Natural isotope correction of MS/MS measurements for metabolomics and ^{13}C fluxomics. *Biotechnol. Bioeng.* 113, 1137–1147. doi: 10.1002/bit.25859
- Niedenführ, S., Wiechert, W., and Nöh, K. (2015). How to measure metabolic fluxes: a taxonomic guide for ^{13}C fluxomics. *Curr. Opin. Biotechnol.* 34, 82–90. doi: 10.1016/j.copbio.2014.12.003
- Nielsen, J. (2003). It is all about metabolic fluxes. *J. Bacteriol.* 185, 7031–7035. doi: 10.1128/JB.185.24.7031-7035.2003
- Nilsson, R., and Jain, M. (2016). Simultaneous tracing of carbon and nitrogen isotopes in human cells. *Mol. Biosyst.* 12, 1929–1937. doi: 10.1039/C6MB00009F
- Noack, S., Nöh, K., Moch, M., Oldiges, M., and Wiechert, W. (2011). Stationary versus non-stationary ^{13}C -MFA: a comparison using a consistent dataset. *J. Biotechnol.* 154, 179–190. doi: 10.1016/j.jbiotec.2010.07.008
- Noack, S., and Wiechert, W. (2014). Quantitative metabolomics: a phantom? *Trends Biotechnol.* 32, 238–244. doi: 10.1016/j.tibtech.2014.03.006
- Nöh, K., Droste, P., and Wiechert, W. (2015). Visual workflows for ^{13}C -metabolic flux analysis. *Bioinformatics* 31, 346–354. doi: 10.1093/bioinformatics/btu585
- Nöh, K., Niedenführ, S., Beyß, M., and Wiechert, W. (2018). A Pareto approach to resolve the conflict between information gain and experimental costs: multiple-criteria design of carbon labeling experiments. *PLOS Comput. Biol.* 14:e1006533. doi: 10.1371/journal.pcbi.1006533
- Nöh, K., Wahl, A., and Wiechert, W. (2006). Computational tools for isotopically instationary ^{13}C labeling experiments under metabolic steady state conditions. *Metab. Eng.* 8, 554–577. doi: 10.1016/j.ymben.2006.05.006
- Poskar, C. H., Huege, J., Krach, C., Franke, M., Shachar-Hill, Y., and Junker, B. H. (2012). iMS2Flux - A high-throughput processing tool for stable isotope labeled mass spectrometric data used for metabolic flux analysis. *BMC Bioinformatics* 13:295. doi: 10.1186/1471-2105-13-295
- Quek, L.-E., Wittmann, C., Nielsen, L. K., and Krömer, J. O. (2009). OpenFLUX: efficient modelling software for ^{13}C -based metabolic flux analysis. *Microb. Cell Fact.* 8:25. doi: 10.1186/1475-2859-8-25
- Rocca-Serra, P., Salek, R. M., Arita, M., Correa, E., Dayalan, S., Gonzalez-Beltran, A., et al. (2016). Data standards can boost metabolomics research, and if there is a will, there is a way. *Metabolomics* 12, 1–13. doi: 10.1007/s11306-015-0879-3
- Sauer, U. (2006). Metabolic networks in motion: ^{13}C -based flux analysis. *Mol. Syst. Biol.* 2:62. doi: 10.1038/msb4100109
- Schmidt, K., Carlsen, M., Nielsen, J., and Villadsen, J. (1997). Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol. Bioeng.* 55, 831–840.
- Schwechheimer, S. K., Becker, J., and Wittmann, C. (2018). Towards better understanding of industrial cell factories: novel approaches for ^{13}C metabolic flux analysis in complex nutrient environments. *Curr. Opin. Biotechnol.* 54, 128–137. doi: 10.1016/j.copbio.2018.07.001
- Schwender, J., Shachar-Hill, Y., and Ohlrogge, J. B. (2006). Mitochondrial metabolism in developing embryos of *Brassica napus*. *J. Biol. Chem.* 281, 34040–34047. doi: 10.1074/jbc.M606266200
- Sherry, A. D., Jeffrey, F. M. H., and Malloy, C. R. (2004). Analytical solutions for ^{13}C isotopomer analysis of complex metabolic conditions: substrate oxidation, multiple pyruvate cycles, and gluconeogenesis. *Metab. Eng.* 6, 12–24. doi: 10.1016/j.ymben.2003.10.007
- Shupletsov, M. S., Golubeva, L. I., Rubina, S. S., Podvyaznikov, D. A., Iwatani, S., and Mashko, S. V. (2014). OpenFLUX2: ^{13}C -MFA modeling software package adjusted for the comprehensive analysis of single and parallel labeling experiments. *Microb. Cell Fact.* 13:152. doi: 10.1186/s12934-014-0152-x
- Sokol, S., Millard, P., and Portais, J. C. (2012). influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments. *Bioinformatics* 28, 687–693. doi: 10.1093/bioinformatics/btr716
- Sokol, S., and Portais, J. (2015). Theoretical basis for dynamic label propagation in stationary metabolic networks under step and periodic inputs. *PLoS ONE* 10:e0144652. doi: 10.1371/journal.pone.0144652
- Srour, O., Young, J. D., and Eldar, Y. C. (2011). Fluxomers: a new approach for ^{13}C metabolic flux analysis. *BMC Syst. Biol.* 5:129. doi: 10.1186/1752-0509-5-129
- Su, X., Lu, W., and Rabinowitz, J. D. (2017). Metabolite spectral accuracy on orbitraps. *Anal. Chem.* 89, 5940–5948. doi: 10.1021/acs.analchem.7b00396
- Tepper, N., and Shlomi, T. (2015). Efficient modeling of MS/MS data for metabolic flux analysis. *PLoS ONE* 10:e0130213. doi: 10.1371/journal.pone.0130213
- Theorell, A., Leweke, S., Wiechert, W., and Nöh, K. (2017). To be certain about the uncertainty: Bayesian statistics for ^{13}C metabolic flux analysis. *Biotechnol. Bioeng.* 114, 2668–2684. doi: 10.1002/bit.26379
- Tillack, J., Paczia, N., Nöh, K., Wiechert, W., and Noack, S. (2012). Error propagation analysis for quantitative intracellular metabolomics. *Metabolites* 2, 1012–1030. doi: 10.3390/metabo2041012
- van Winden, W., Verheijen, P., and Heijnen, S. (2001). Possible pitfalls of flux calculations based on ^{13}C -labeling. *Metab. Eng.* 3, 151–162. doi: 10.1006/mben.2000.0174
- Wahl, S. A., Dauner, M., and 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. doi: 10.1002/bit.10909
- Weitzel, M., Nöh, K., Dalman, T., Niedenführ, S., Stute, B., and Wiechert, W. (2013). 13CFLUX2 - High-performance software suite for ^{13}C -metabolic flux analysis. *Bioinformatics* 29, 143–145. doi: 10.1093/bioinformatics/bts646
- Weitzel, M., Wiechert, W., and Nöh, K. (2007). The topology of metabolic isotope labeling networks. *BMC Bioinformatics* 8:315. doi: 10.1186/1471-2105-8-315
- Westerhoff, H. V., and Hofmeyr, J.-H. S. (2005). “What is systems biology? From genes to function and back,” in *Systems Biology*, eds L. Alberghina and H. V. Westerhoff (Berlin: Springer-Verlag), 119–141. doi: 10.1007/b137122
- Wiechert, W. (2001). ^{13}C metabolic flux analysis. *Metab. Eng.* 3, 195–206. doi: 10.1006/mben.2001.0187

- Wiechert, W. (2007). The thermodynamic meaning of metabolic exchange fluxes. *Biophys. J.* 93, 2255–2264. doi: 10.1529/biophysj.106.099895
- Wiechert, W., and de Graaf, A. A. (1996). *In vivo* stationary flux analysis by ^{13}C labeling experiments. *Adv. Biochem. Eng. Biotechnol.* 54, 109–154. doi: 10.1007/BFb0102334
- Wiechert, W., and de Graaf, A. A. (1997). Bidirectional reaction steps in metabolic networks. Part I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnol. Bioeng.* 55, 101–117. doi: 10.1002/(SICI)1097-0290(19970705)55:1<101::AID-BIT12>3.0.CO;2-P
- Wiechert, W., Möllney, M., Isermann, N., Wurzel, M., and de Graaf, A. A. (1999). Bidirectional reaction steps in metabolic networks. Part III: explicit solution and analysis of isotopomer labeling systems. *Biotechnol. Bioeng.* 66, 69–85.
- Wiechert, W., Möllney, M., Petersen, S., and de Graaf, A. A. (2001). A universal framework for ^{13}C metabolic flux analysis. *Metab. Engineering* 3, 265–283. doi: 10.1006/mben.2001.0188
- Wiechert, W., Nidenführ, S., and Nöh, K. (2015). “A Primer to ^{13}C metabolic flux analysis,” in *Fundamental Bioengineering*, ed J. Villadsen (Weinheim: Wiley-VCH Verlag GmbH and Co. KGaA), 97–142. doi: 10.1002/9783527697441.ch05
- Wiechert, W., and Nöh, K. (2005). From stationary to instationary metabolic flux analysis. *Adv. Biochem. Eng. Biotechnol.* 92, 145–172. doi: 10.1007/b98921
- Wiechert, W., Siefke, C., de Graaf, A. A., and Marx, A. (1997). Bidirectional reaction steps in metabolic networks. Part II: Flux estimation and statistical analysis. *Biotechnol. Bioeng.* 55, 118–135. doi: 10.1002/(SICI)1097-0290(19970705)55:1<118::AID-BIT13>3.0.CO;2-I
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3:160018. doi: 10.1038/sdata.2016.18
- Wittmann, C., and Heinze, E. (2001). Application of MALDI-TOF MS to lysine-producing *Corynebacterium glutamicum*: a novel approach for metabolic flux analysis. *Eur. J. Biochem.* 268, 2441–2455. doi: 10.1046/j.1432-1327.2001.02129.x
- Wolkenhauer, O. (2001). Systems biology: the reincarnation of systems theory applied in biology? *Brief. Bioinformatics* 2, 258–270. doi: 10.1093/bib/2.3.258
- Young, J. D. (2014). INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* 30, 1333–1335. doi: 10.1093/bioinformatics/btu015
- Young, J. D., Walther, J. L., Antoniewicz, M. R., Yoo, H., and Stephanopoulos, G. (2008). An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnol. Bioeng.* 99, 686–699. doi: 10.1002/bit.21632
- Yuan, Y., Yang, T. H., and Heinze, E. (2010). ^{13}C metabolic flux analysis for larger scale cultivation using gas chromatography-combustion-isotope ratio mass spectrometry. *Metab. Eng.* 12, 392–400. doi: 10.1016/j.ymben.2010.02.001
- Zamboni, N., Fendt, S.-M., Rühl, M., and Sauer, U. (2009). ^{13}C -based metabolic flux analysis. *Nat. Protoc.* 4, 878–892. doi: 10.1038/nprot.2009.58
- Zamboni, N., Fischer, E., and Sauer, U. (2005). FiatFlux - A software for metabolic flux analysis from ^{13}C -glucose experiments. *BMC Bioinformatics* 6:209. doi: 10.1186/1471-2105-6-209
- Zhang, Z., Shen, T., Rui, B., Zhou, W., Zhou, X., Shang, C., et al. (2014). CeCaFDB: a curated database for the documentation, visualization and comparative analysis of central carbon metabolic flux distributions explored by ^{13}C -fluxomics. *Nucleic Acids Res.* 43, D549–D557. doi: 10.1093/nar/gku1137
- Zupke, C., and Stephanopoulos, G. (1994). Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices. *Biotechnol. Prog.* 10, 489–498. doi: 10.1021/bp00029a006

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

Copyright © 2019 Beyß, Azzouzi, Weitzel, Wiechert and Nöh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



¹³C-Metabolic Flux Analysis Reveals Effect of Phenol on Central Carbon Metabolism in *Escherichia coli*

Sayaka Kitamura, Yoshihiro Toya and Hiroshi Shimizu*

Department of Bioinformatic Engineering, Graduate School of Information Science and Technology, Osaka University, Osaka, Japan

OPEN ACCESS

Edited by:

Wei Xiong,
National Renewable Energy
Laboratory (DOE), United States

Reviewed by:

Volker F. Wendisch,
Bielefeld University, Germany
Chen Shouwen,
Hubei University, China

*Correspondence:

Hiroshi Shimizu
shimizu@ist.osaka-u.ac.jp

Specialty section:

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

Received: 30 October 2018

Accepted: 23 April 2019

Published: 07 May 2019

Citation:

Kitamura S, Toya Y and Shimizu H
(2019) ¹³C-Metabolic Flux Analysis
Reveals Effect of Phenol on Central
Carbon Metabolism in *Escherichia*
coli. *Front. Microbiol.* 10:1010.
doi: 10.3389/fmicb.2019.01010

Phenol is an important chemical product that can be used in a wide variety of applications, and it is currently produced from fossil resources. Fermentation production of phenol from renewable biomass resources by microorganisms is highly desirable for sustainable development. However, phenol toxicity hampers phenol production in industrial microorganisms such as *Escherichia coli*. In the present study, it was revealed that culturing *E. coli* in the presence of phenol not only decreased growth rate, but also biomass yield. This suggests that phenol affects the carbon flow of the metabolism, but the mechanism is unknown. To investigate the effect of phenol on the flux distribution of central carbon metabolism, ¹³C-metabolic flux analysis (¹³C-MFA) was performed on cells grown under different phenol concentrations (0, 0.1, and 0.15%). ¹³C-MFA revealed that the TCA cycle flux reduced by 25% increased acetate production from acetyl-CoA by 30% in the presence of 0.1% phenol. This trend of flux changes was emphasized at a phenol concentration of 0.15%. Although the expression level of citrate synthase, which catalyzes the first reaction of the TCA cycle, does not change regardless of phenol concentrations, the *in vitro* enzyme activity assay shows that the reaction was inhibited by phenol. These results suggest that the TCA cycle flux decreased due to phenol inhibition of citrate synthase; therefore, ATP could not be sufficiently produced by respiration, and growth rate decreased. Furthermore, since carbon was lost as acetate due to overflow metabolism, the biomass yield became low in the presence of phenol.

Keywords: phenol toxicity, *Escherichia coli*, ¹³C-metabolic flux analysis, enzymatic assay, citrate synthase, acetate overflow metabolism

INTRODUCTION

Phenol is a valuable compound used in various applications, such as in the production of polycarbonate, epoxy resin, phenol resin, and aniline (Schmidt, 2005). Although it is mainly produced by the cumene method using benzene derived from fossil resources at present (Wierckx et al., 2005), fermentation production using microorganisms from renewable biomass resources

is required to achieve sustainable development and solve environmental problems. A tyrosine phenol-lyase (TPL), which has been discovered in some microorganisms such as *Bacterium coli phenologenes*, *Pantoea agglomerans*, and *Clostridium tetanomorphum*, catalyzes the conversion between tyrosine and phenol (Brot et al., 1965; Kumagai et al., 1970). Microbial phenol production has been reported in *Pseudomonas putida* by overexpressing the heterologous TPL of *P. agglomerans* (Wierckx et al., 2005). More recently, engineered strains of *Escherichia coli*, widely used as industrial bio-production hosts, can produce phenol from glucose by introducing the TPL from *P. agglomerans* (Kim et al., 2014; Thompson et al., 2016). However, productivity is still low for industrial applications.

Organic solvents, including phenol, are toxic toward a wide range of microorganisms, including *E. coli*, and reduce their growth (Ramos et al., 2002). In the phenol-producing *E. coli* strain, there is a challenge of low phenol productivity due to toxicity (Kim et al., 2014). Phenol tolerance varies among microorganisms, with some species including *P. putida*, used for bioremediation, being able to tolerate the presence of phenol (Ramos et al., 2002). In the same gram-negative bacteria, however, *E. coli* is weak against phenol and stops growing under a 1.2 g/L, corresponding to 0.11%(v/v), phenol condition (Kim et al., 2014). Despite this disadvantage, *E. coli* is an excellent host for industrial applications because of its rapid growth rate, easy genetic manipulation, and abundant biological knowledge. Elucidation of the reasons why *E. coli* is weak against phenol would make possible to overcome this weakness and enhance phenol production.

Various organic solvents are toxic to *E. coli*, and this toxicity correlates with a logarithm of the partition coefficient in *n*-octanol and water ($\log P_{ow}$) (Aono et al., 1994). The $\log P_{ow}$ of phenol is 1.46 (Verschuere, 1983), and is very severe among organic solvents. In *E. coli*, some efflux pumps such as AcrAB-TolC and AcrEF-TolC are involved in solvent tolerance (Ramos et al., 2002). Furthermore, it has been reported that the composition of the cell membrane changes in the presence of phenol (Keweloh et al., 1990), and phenol tolerance increases by alterations of the fatty acid composition of membranes (Keweloh et al., 1991). Because its respiratory activity is not reduced in the presence of phenol (Keweloh et al., 1990), the cause of growth reduction would not be respiratory chain inhibition due to membrane damage. We found in this work that culturing *E. coli* in the presence of phenol not only decreased growth rate, but also biomass yield. This suggests that phenol affects the carbon flow of central carbon metabolism, but the mechanism is unknown. ^{13}C -metabolic flux analysis (^{13}C -MFA) is an effective approach to investigate the carbon flux distribution on central carbon metabolism (Wittmann, 2007; Zamboni et al., 2009).

In the present study, we cultured wild type *E. coli* under different phenol concentrations (0, 0.1, and 0.15%), and compared these flux distributions to identify the effect of phenol on the metabolism. To investigate the cause of the flux changes, the *in vitro* enzyme assay was performed and revealed that citrate synthase is strongly inhibited by phenol.

MATERIALS AND METHODS

Strains and Culture Conditions

The *E. coli* strains used in this study are shown in **Supplementary Table S1**. In preculture, wild type, *gltA*+, and Δ *pta* strains were aerobically grown at 37°C overnight using 5 mL of M9 medium containing 4 g/L glucose in a test tube. For the *gltA*+ strain, 40 mg/mL histidine and 10 mg/mL Thiamine hydrochloride was supplemented to the M9 media in accordance with the previous report (Kitagawa et al., 2005).

To evaluate phenol toxicity using an automated optical density monitoring system, the preculture was inoculated into 5 mL of the same M9 medium in an L-shaped test tube at an initial optical density at 660 nm (OD_{660}) of 0.05. Phenol was added to be 0, 0.04, 0.08, 0.1, 0.12, and 0.16% (v/v). The cultures were incubated at 37°C at 70 rpm using a TVS062CA incubator (Advantec, Tokyo, Japan). The OD_{660} was measured every 5 min.

For evaluating the fermentation profiles in flask scale, the preculture of wild type and Δ *pta* strains were inoculated into 50 mL of M9 medium containing 4 g/L glucose as the sole carbon source in a 200 mL baffled flask at an initial OD_{660} of 0.05, and incubated at 37°C at 200 rpm using a BR-43FL incubator (TAITEC, Saitama, Japan). Phenol was added to be 0 and 0.15% (v/v). For ^{13}C -MFA, the cultures were performed using the same condition except that the glucose was replaced to [$1\text{-}^{13}\text{C}$] glucose. Phenol was added to be 0, 0.1, and 0.15% (v/v). All cultures were performed in triplicate.

Measurement of Cell Concentration and Extracellular Metabolites

The OD_{660} was measured using an UVmini-1240 UV-VIS spectrophotometer (Shimadzu, Kyoto, Japan). The dry cell weight (DCW) was calculated using a conversion coefficient of 0.3 g/L/ OD_{660} based on a previous report (Soini et al., 2008). Concentrations of glucose, acetate, formate, ethanol, lactate, and succinate in the culture were measured using a high-performance liquid chromatography system (Shimadzu, Kyoto, Japan) with an Aminex HPX-87H column (Bio-Rad, Hercules, CA, United States). The detailed method was described in Okahashi et al. (2017). The detection limits were 5 mM for ethanol and 1 mM for organic acids including lactate, acetate, formate, and succinate.

Measurement of Proteinogenic Amino Acids

During the mid-log phase, an appropriate amount of cells (0.0015 gDCW) was collected by centrifugation. After hydrolyzation, the proteinogenic amino acids were derivatized with tert-butyldimethylsilyl (tBDMS). The mass isotopomer distributions of amino acids were measured using a gas chromatography/mass spectrometer (GC/MS) (Agilent Technologies, United States) with a DB-5MS column (Agilent Technologies, United States). The mass isotopomer distributions were corrected for the presence of naturally occurring isotopes (van Winden et al., 2002). The detailed

protocols for sample preparation and measurement were described in Okahashi et al. (2017).

¹³C-Metabolic Flux Analysis

The central carbon metabolism including glycolysis, the pentose phosphate pathway, the TCA cycle, the Entner–Doudoroff (ED) pathway, the anaplerotic pathway, and the glyoxylate shunt, was considered for ¹³C-MFA based on previous analyses of *E. coli*. The elementary metabolites unit framework was used for modeling the carbon atom transitions (Antoniewicz et al., 2007). All reactions and carbon atom transitions are summarized in **Supplementary Table S2**. Flux distribution was optimized to minimize a residual sum of square (RSS) between calculated and measured amino acid mass isotopomer distributions. The fitness function was described below:

$$\text{Minimize RSS} = \sum_{i=1}^n \left(\frac{MID_i^{\text{measured}} - MID_i^{\text{simulated}}}{SD_i} \right)^2 + \sum_{j=1}^m \left(\frac{r_j^{\text{measured}} - r_j^{\text{simulated}}}{SD_j} \right)^2$$

where MID_i^{measured} and $MID_i^{\text{simulated}}$ are measured and simulated mass isotopomer distribution of *i*th amino acid, respectively. The standard deviation (SD) was set to 0.01 in GC/MS measurements (Wada et al., 2017). The r_j^{measured} and $r_j^{\text{simulated}}$ are measured and simulated flux of *j*th reaction, respectively. The difference between the calculated and measured values of the mass isotopomer distribution was statistically evaluated by the χ^2 test. The 95% confidence intervals of each flux were evaluated by the grid search method (Antoniewicz et al., 2006). All calculations were performed using the OpenMebius software (Kajihata et al., 2014) with MATLAB 2018a (MathWorks, Natick, MA, United States).

Measurement of Enzyme Activities

To measure the citrate synthase activity of the *E. coli* cells cultured at the different phenol concentrations, the cells obtained from the 1 mL culture were suspended in 0.1 M of Tris–HCl pH 8.2, and were disrupted by sonication repeating on/off 10 times at out level at 50% using a UD-100 ultrasonic homogenizer (Tomy Seiko, Tokyo, Japan). After removing the debris by centrifugation (20,000 × *g*, 4°C, 30 min), the crude extract was used for citrate synthase activity measurement. The protein concentration of the crude extract was measured using a nanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, United States). Citrate synthase activity was measured by Ellmann's method, that detects the thiol group of CoA by a colorimetric reaction using 5',5'-Dithiobis(2-nitrobenzoic acid) (DTNB) (Srere, 1969). DTNB was measured at 412 nm by UVmini-1240 UV-VIS spectrophotometer (Shimadzu, Kyoto, Japan). After incubating a mixture containing 10 μL of crude extract, 190 μL of MilliQ water, 12.5 μL of 10 mM acetyl-CoA, and 25 μL of 1 mM DTNB in 0.1 M Tris–HCl (pH 8.2) at 25°C for 5 min, 12.5 μL

of 10 mM oxaloacetate in 0.1 M Tris–HCl (pH 8.2) was added to start the reaction. The absorbance at 412 nm was measured every 10 s.

Phenol inhibition of the *E. coli* citrate synthase *in vitro* reaction was examined using a sample obtained from a *gltA* overexpressing strain, JW0710-AM, from the ASKA library (Kitagawa et al., 2005). In preculture, a single colony of this mutant was inoculated in 5 mL of Luria-Bertani (LB) medium, and was incubated overnight at 37°C at 150 rpm. The preculture was inoculated into 50 mL of LB medium at an initial OD₆₆₀ of 0.05, and was incubated at 37°C at 150 rpm using a BR-43FL incubator (TAITEC, Saitama, Japan). At OD₆₀₀ of approximately 0.3, 0.1 mM of isopropyl β-D-1-thiogalactopyranoside was added to induce gene expression. After 2 h, the cells were collected by centrifugation (3000 rpm, 4°C, 10 min), and were washed by 5 mL of filter sterilized buffer containing 50 mM sodium phosphate, 200 mM NaCl, and a protease inhibitor tablet (cOmplete tablets, Roche). To keep the plasmid overexpressing the *gltA* gene, 25 mg/L of chloramphenicol were added to the media. The obtained cell pellet was extracted by sonication and used for *in vitro* enzymatic assays in the same manner described above.

Phenol inhibition of glucose 6-phosphate isomerase (PGI) *in vitro* reaction was also evaluated using a sample obtained from wild type *E. coli*. PGI activity was measured as NADPH formation through a coupled reaction with glucose 6 phosphate dehydrogenase (G6PDH) (Salas et al., 1965; Peng and Shimizu, 2003). After addition of an appropriate amount of crude cell extract to a reaction mixture containing 0.1 M Tris–HCl (pH 7.8), 10 mM MgCl₂, 0.5 mM NADP⁺, 1 U G6PDH (Sigma-Aldrich, Merck, Darmstadt, Germany), and 2 mM D-Fructose 6 phosphate, NADPH formation was measured by absorption at 340 nm. One unit (U) is defined as the amount of enzyme required to convert 1 μmol of substrate into product per minute per milligram protein.

RESULTS

Evaluation of Phenol Toxicity in *E. coli*

The wild type *E. coli* BW25113 was cultured using a synthetic M9 medium with different phenol concentrations (0, 0.04, 0.08, 0.1, 0.12, and 0.16%) in L-shaped test tubes. **Figure 1** shows the specific growth rate and the maximum OD₆₆₀ at each phenol concentration. The specific growth rate decreased by 40% under 0.1% phenol, and by 84% under 0.16% phenol as compared to that without addition of phenol. A previous study has also reported that an increase in the phenol concentration of the culture led to a decrease in the growth rate of *E. coli*, and eventually cell growth was completely inhibited (Kim et al., 2014). These phenotypes were consistent with our result. Furthermore, the value of maximum OD₆₆₀ also decreased by 16% under 0.1% phenol and by 92% under 0.16% phenol. The drastic decrease of biomass yield suggests that phenol affects the carbon flux of the central carbon metabolism.

Growth Characteristics of *E. coli* Under Different Phenol Concentrations

In order to reveal the influence of phenol on the carbon flow of the metabolism, the cells were cultured in 50 mL of M9 medium using a baffled flask with three different phenol concentrations (0, 0.1, and 0.15%). The time courses of cell, glucose, and acetate concentrations in the culture broth are shown in **Figure 2**.

At 0% phenol, the OD₆₀₀ reached the maximum value of 4.7 at 9.3 h. Glucose consumption was the fastest among the three conditions, and was depleted by 8.3 h. Acetate concentration reached a maximum of 12.4 mM at 8.3 h, and was quickly consumed after glucose depletion. At 0.1% phenol, the maximum OD₆₀₀ was 4.4 at 24.2 h. Glucose was consumed more slowly and was depleted by 10.4 h. Acetate concentration reached a maximum of 20.1 mM (1.6 times of that at 0% phenol) at 10.4 h, and then it was gently consumed. At 0.15% phenol, the maximum OD₆₀₀ was significantly decreased to 2.6. Glucose consumption was the slowest and was depleted at 14 h. Acetate production was further enhanced to 34.5 mM at 24 h. No acetate consumption was observed after depletion of glucose. Furthermore, no ethanol, lactate, succinate, and formate productions were detected throughout the culture period under any conditions. The specific rates during the exponential

TABLE 1 | Specific growth, glucose consumption, and acetate production rates of wild type *E. coli* under different phenol concentrations.

Phenol concentration %	Cell growth h ⁻¹	Glucose consumption mmol gDCW ⁻¹ h ⁻¹	Acetate production mmol gDCW ⁻¹ h ⁻¹
0.00	0.59 ± 0.002	9.43 ± 0.58	6.05 ± 0.16
0.10	0.46 ± 0.001	7.68 ± 0.17	7.31 ± 0.19
0.15	0.34 ± 0.001	7.10 ± 0.22	9.53 ± 0.23

growth phase are summarized in **Table 1**. Compared to the phenol 0% condition, the specific growth rate and specific glucose consumption rate at 0.15% phenol decreased by 42 and 24%, respectively, whereas the specific acetate production rate increased by 56%.

To further investigate the effect of acetate overflow on the phenol tolerance, a Δpta strain which does not have phosphotransacetylase for acetate formation was cultured in the absence and presence of phenol (**Figure 3**). Interestingly, the growth profile of the Δpta strain was similar to that of the wild type in the absence of phenol, whereas the strain did not grow in the presence of 0.15% phenol. The result suggests that the acetate overflow is related to *E. coli* growth in the presence of phenol.

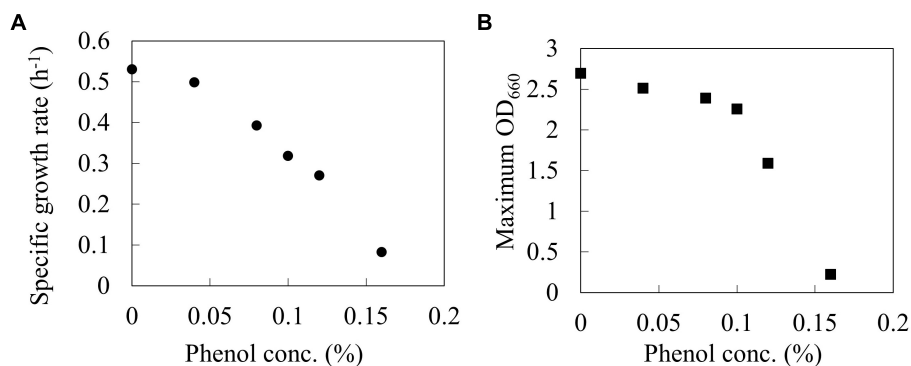


FIGURE 1 | Relationships between phenol concentration and Specific growth rate (A), and Maximum OD₆₆₀ (B) on wild type *E. coli* BW25113. Closed circles and squares represent the maximum specific growth rate (h⁻¹) and the maximum OD₆₆₀ during the 24 h period.

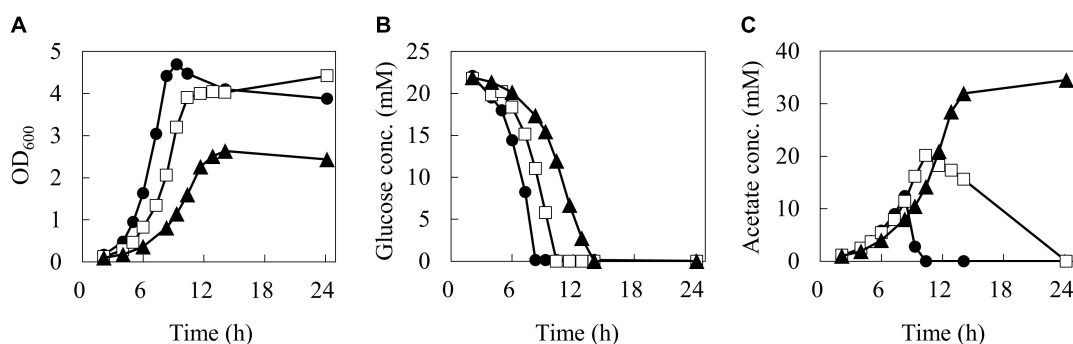
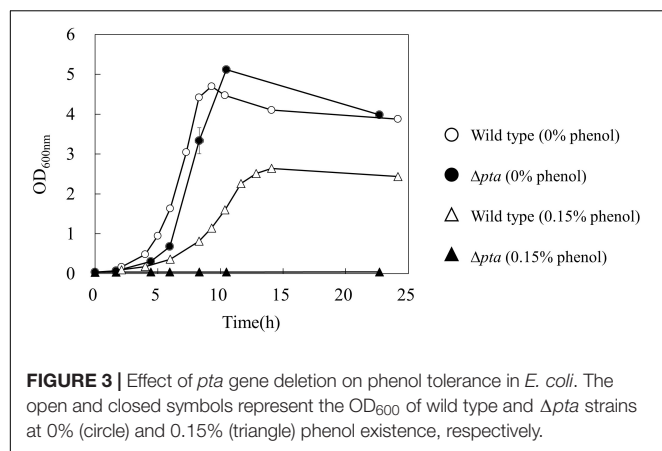


FIGURE 2 | Batch growth profiles of wild type *E. coli* under different phenol concentrations. Time courses of biomass (A), glucose (B), and acetate (C) concentrations. Closed circles, open squares, and closed triangles represent measurement data series during 24 h by 0, 0.1, and 0.15% of phenol concentrations, respectively.



Flux Distributions Under Different Phenol Concentrations by ^{13}C -Metabolic-Flux Analysis (^{13}C -MFA)

To identify the reaction in which the flux was changed depending on phenol concentrations, a ^{13}C -MFA was performed for each condition. The cells were cultured with $[1-^{13}\text{C}]$ glucose as a sole carbon source. The mass isotopomer distributions of proteinogenic amino acids at the exponential growth phase were measured by GC/MS. The mass isotopomer distributions of amino acids that have been corrected for natural isotope abundances are shown in **Supplementary Table S3**. The flux distributions were optimized to minimize the RSS between calculated and measured amino acid mass isotopomer distributions. The threshold of the χ^2 test for the goodness of fit was 106.4 (104 independent data and 18 degrees of freedom for the model). As a result of flux optimization, the RSS values of conditions with 0, 0.1, and 0.15% phenol were 92.1, 104.5, and 103.3, respectively (passed χ^2 test, $p < 0.05$). The 95% confidence interval of each flux, to explain the measured mass isotopomer distributions of amino acid, was also calculated. The best fitted flux distribution and the 95% confidence interval are shown in **Supplementary Table S4**.

Although no significant change was observed in the upstream pathways including glycolysis, the pentose phosphate pathway, and the ED pathway by phenol addition, a large change occurred in the downstream pathways including pyruvate metabolism and the TCA cycle. The flux distributions focused on the downstream pathways are shown in **Figure 4**. The flux of the TCA cycle excluding malate dehydrogenase was significantly decreased in the presence of phenol. At the 0% phenol condition, the flux of citrate synthase reaction, the first reaction of the TCA cycle, was $3.6 \text{ mmol gDCW}^{-1} \text{ h}^{-1}$. However, the fluxes of citrate synthase at 0.1 and 0.15% phenol were decreased to 0.5 and $0.4 \text{ mmol gDCW}^{-1} \text{ h}^{-1}$, respectively. This suggests that acetyl-CoA does not flow into the TCA cycle, therefore, increasing overflow from acetyl-CoA to acetate.

Although the upper bound of the estimated flux of glyoxylate shunt seems to be decreased as the phenol concentration increases, it is difficult to conclude the flux changes because the

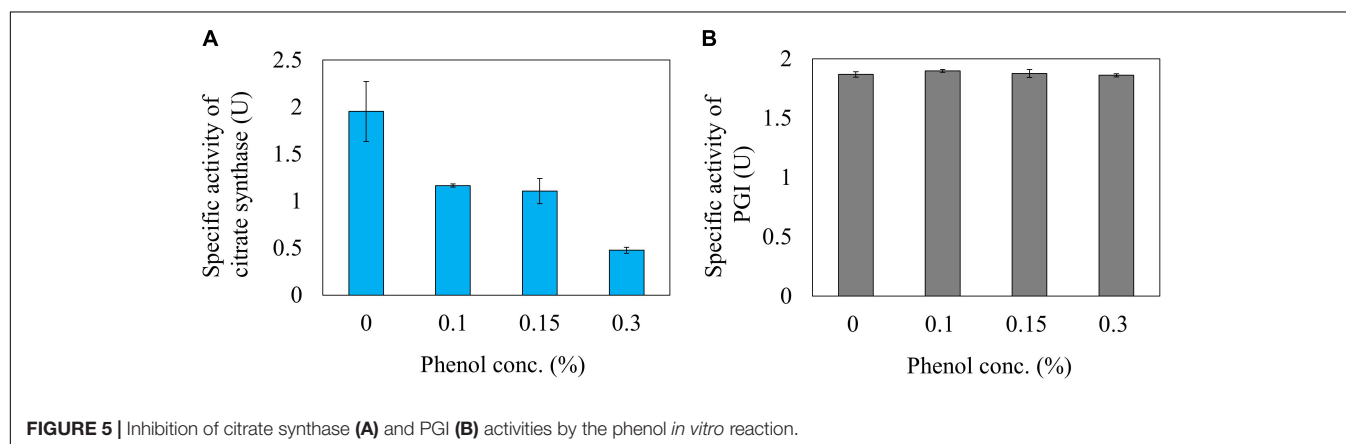
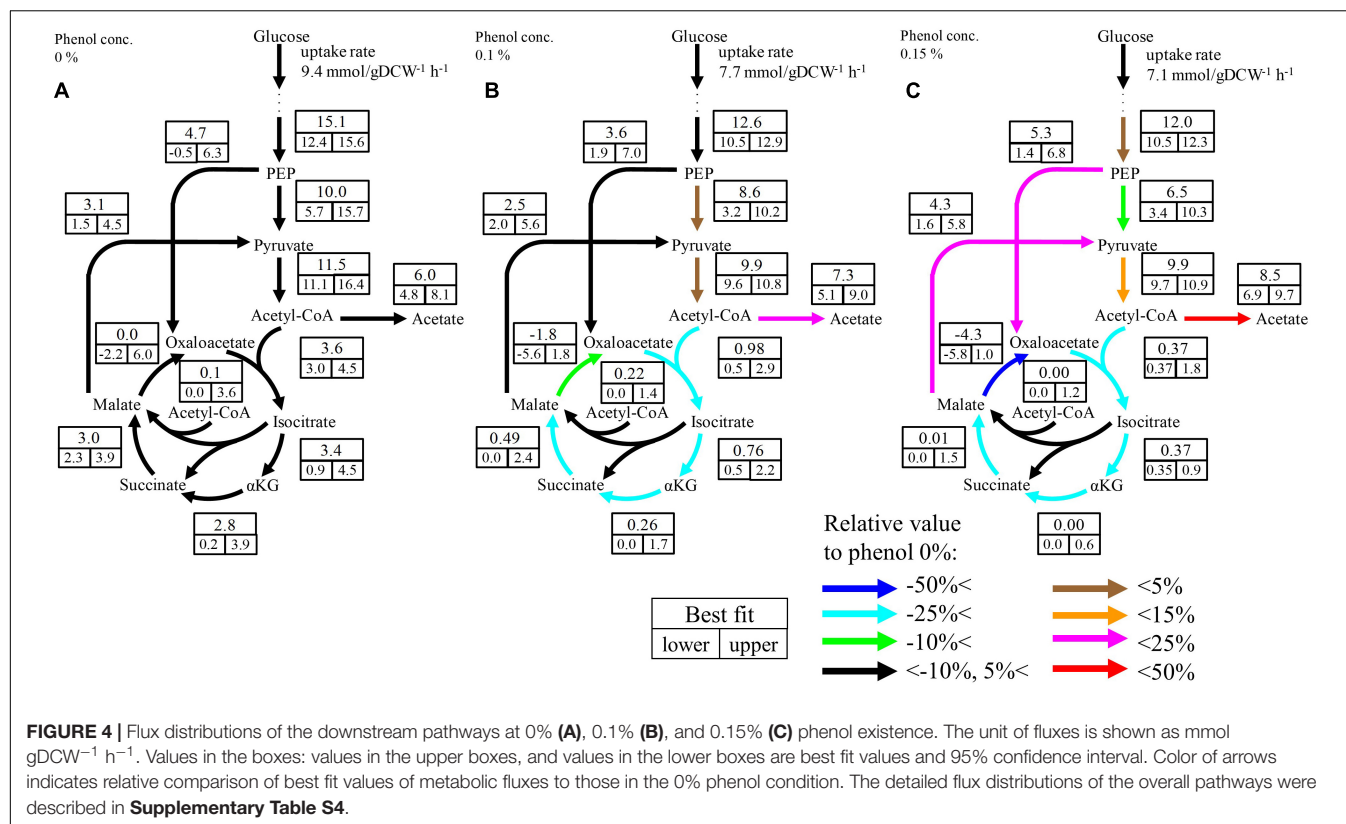
confidence interval of this flux at the low phenol concentration. The flux of this pathway is usually low in *E. coli* under glucose carbon source without phenol addition (Leighty and Antoniewicz, 2012). Therefore, it is considered that this pathway did not work so much, regardless of the phenol concentration.

Effect of Phenol on Enzyme Activities

^{13}C -metabolic flux analysis revealed that the TCA cycle flux decreased in the presence of phenol in *E. coli*. Because no leakage of the TCA cycle intermediates, such as citrate and α -ketoglutarate, was detected in the culture broths, we assumed that phenol inhibits a citrate synthase reaction, at the entrance of the TCA cycle. In order to evaluate the changes in the protein expression level of citrate synthase, the specific activities of the crude extracts obtained from the cells cultured with three different phenol concentrations (0, 0.1, and 0.15%) were measured using an *in vitro* enzymatic assay. The specific activities of the crude cell extracts obtained from the cultures with 0, 0.1, and 0.15% phenol were 0.095 ± 0.007 , 0.087 ± 0.009 , and $0.100 \pm 0.009 \text{ U}$, respectively. No significant difference was observed. This result suggests that the expression level of citrate synthase does not change among the culture conditions. Assuming 50% protein in the DCW, the citrate synthase activity (0.095 U) under the 0% phenol was calculated as 2.9 mmol/gDCW/h . In the ^{13}C -MFA, the 95% confidence interval of this reaction was 3 to $4.5 \text{ mmol gDCW}^{-1} \text{ h}^{-1}$. Since the citrate synthase activity is almost consistent with the range of the flux, the *in vitro* enzyme assay results are reasonable.

Next, the effect of phenol on the reaction rate of citrate synthase was investigated using a crude cell extract obtained from a *gltA* overexpressing strain. The relative activities of citrate synthase to a phenol free condition were 59.8, 56.6, and 24.6% at phenol concentrations in the reaction mixture of 0.1, 0.15, and 0.3%, respectively (**Figure 5A**). This result is consistent with the ^{13}C -MFA results showing that the flux of the reaction from acetyl-CoA and oxaloacetate to citrate was reduced from 3.6 to $0.37 \text{ mmol gDCW}^{-1} \text{ h}^{-1}$ by addition of 0.15% phenol. Since high concentration of phenol is involved in protein denaturation, the decreased citrate synthase activity by phenol addition may be due to denaturation of the enzyme. To remove the suspicion, the effect of phenol on other enzyme activity was examined. As the representative enzyme, phenol inhibition of PGI was investigated (**Figure 5B**), because little flux change in the reaction was observed in the ^{13}C metabolic flux analysis in response to phenol addition. No change in the PGI activity was observed by addition of phenol in the reaction mixture. This result supports that the cause of citrate synthase activity decrease in the presence of phenol is not denaturation of the protein.

If the inhibition of citrate synthase by phenol is the cause of growth reduction, the overexpression would restore the growth in the presence of phenol. To evaluate the effect of citrate synthase overexpression on phenol tolerance, the citrate synthase overexpressing (*gltA*+) strain was cultured using the M9 medium with different phenol concentrations. Because the expression of citrate synthase is regulated by an IPTG inducible T5-*lac* promoter (Kitagawa et al., 2005), 50 μM IPTG was added to the cultures. **Supplementary Figure S1** shows the specific growth



rate and the maximum OD₆₆₀ at each phenol concentration. The decrease of cell growth was also observed in this overexpressing strain in the presence of phenol. Because the overexpression of citrate synthase caused a severe growth defect without phenol condition, fine-tuning of the expression level would be needed to alleviate the phenol stress.

DISCUSSION

In the present study, we investigated in detail the cause of phenol toxicity in *E. coli* from the viewpoint of metabolism. ¹³C-MFA showed an enhanced overflow of acetyl-CoA to acetate

at phenol concentrations of 0.1 and 0.15%, indicating that the TCA cycle flux was significantly decreased. Further analysis based on measuring the specific enzyme activity *in vitro* revealed that citrate synthase, at the entrance reaction of the TCA cycle, was inhibited by phenol and led to the flux repression of this reaction.

¹³C-metabolic flux analysis can reveal how the flux changed on the metabolic pathways, but it does not provide the mechanism that caused this flux change. In addition, since it assumes a steady state condition, the flux change in a sequential pathway is found; however, it cannot identify which reaction in the pathway caused the flux change. The mechanism of flux changes must be discussed together with the knowledge about cellular

systems, such as enzyme expression and reaction inhibition. For instance, the flux information alone cannot conclude whether the enhanced acetate production led to the decrease of the TCA cycle flux, or the decrease of the TCA cycle flux led to the enhanced acetate production. In *E. coli*, it is known that acetate synthesis occurs due to overflow metabolism (Castaño-Cerezo et al., 2009). Therefore, based on the citrate synthase activity assays *in vitro*, it is considered that the overflow to acetate increased because the TCA cycle flux was repressed by phenol.

How did the decrease of the TCA cycle flux occur in the presence of phenol? Because respiration and the TCA cycle are linked with NADH, damage to the respiratory chain may cause a decrease in the TCA cycle flux. If so, however, NADH will accumulate and lactate and ethanol will be produced in order to recycle NADH produced by glycolysis; like in anaerobic fermentation, these fermentation products were not detected in the culture broth analysis by HPLC. Furthermore, inhibition of the respiratory chain increases uptake of glucose in *E. coli* (Kihira et al., 2012), but the specific glucose uptake rate decreased in the presence of phenol (Table 1). If α -ketoglutarate dehydrogenase and aconitase are inhibited by phenol, α -ketoglutarate and citrate should overflow, respectively, but these compounds were not detected either by the culture broth analysis. Based on these findings, it is considered that the reaction stopped at the point of citrate synthase, which catalyzes the first step of the TCA cycle. In addition, the culture profiles in Figure 2 show that acetate was consumed promptly after depletion of glucose in the absence of phenol. On the other hand, consumption of acetate after glucose depletion was remarkably decreased at 0.1% phenol, and was completely stopped at 0.15% phenol. In *E. coli* grown with acetate as a carbon source, acetate is converted to acetyl-CoA and is catabolized via the glyoxylate shunt into the TCA cycle to avoid CO₂ emission (Zhao and Shimizu, 2003). Therefore, the reaction in which phenol was directly inhibited is a common reaction for acetate catabolic pathways. The acetate uptake and the glyoxylate shunt do not work during the glucose consumption phase. Because reactions from acetyl-CoA to isocitrate are common, we focused on citrate synthase, which catalyzes this step.

Metabolic flux changes are caused by several factors, such as enzyme expression level, and reaction inhibition. Citrate synthase is an important enzyme that is regulated at the gene expression level by various transcription factors such as Cra, ArcAB, and Fnr (Matsuoka and Shimizu, 2011). First, in order to investigate the change in enzyme expression, the specific activity of the crude extract of the cells cultured at different phenol conditions was evaluated, but no activity change was observed. Next, in order to investigate whether phenol affected the citrate synthase reaction in *E. coli*, activity was evaluated by adding phenol to the *in vitro* reaction solution. As shown in Figure 5A, the enzyme activity decreased to 56.6% by addition of 0.15% phenol. The compound is highly permeable to the cell membrane due to its hydrophobicity. Assuming equal concentration of phenol inside and outside the cell, the examined phenol concentration affects the citrate synthase activity *in vivo*. This enzyme assay result suggests that phenol affects the enzyme reaction level, and not the gene expression level. In the ¹³C-MFA results, since the citrate

synthase flux is more severely inhibited to 11% at 0.15% phenol, it suggests that other factors are remained to decrease the TCA cycle flux. A comprehensive analysis of gene expression profiles might provide clues to identify the factors.

For decreasing the acetate overflow in the presence of phenol, a gene knockout of *pta* encoding phosphotransacetylase which catalyzes the acetate formation pathway was investigated (Figure 3). The phenol tolerance of this mutant was lower than that of wild type strain. No growth of the Δ *pta* strain was observed in the presence of 0.15% phenol. Since the knockout mutant cannot discard the end product of glycolysis as acetate in the presence of phenol, excess intermediates would accumulate in the cell and inhibit the metabolic flow in glycolysis. As another hypothesis, post-translational modification of proteins by acetylation may have affected the metabolic behavior. Acetyl-phosphate, an intermediate in the acetate synthesis pathway, is an acetyl group donor for lysine acetylation of various proteins. It has been reported that the acetylation of isocitrate lyase affects the activity of glyoxylate shunt and transcription factor RcsB which related to an acid stress susceptibility (Castaño-Cerezo et al., 2014).

In the present study, ¹³C-MFA and enzyme assays reveal that the TCA cycle flux decreased due to phenol inhibition of citrate synthase; therefore, ATP could not be sufficiently produced by respiration, and growth rate decreased. Furthermore, since the carbon was lost as acetate due to overflow metabolism, the biomass yield became low in the presence of phenol. In the central carbon metabolism of *E. coli*, acetyl-CoA node is a branch point between the TCA cycle and acetate synthesis. The inhibition of citrate synthase by phenol blocks carbon flux into the TCA cycle and enhances the overflow into the acetate synthesis. Furthermore, since the acetyl-phosphate is an intermediate of acetate synthesis pathway, the accumulation may affect the TCA cycle flux through changes in enzyme activity and gene expression via protein acetylation. Investigations of the gene expression profiles and post-translational modifications on phenol stress would be necessary for further understanding the metabolic regulation. However, flux of upstream pathways such as glycolysis and the pentose phosphate pathway did not decrease. Phenol is synthesized via the aromatic amino acid synthesis pathway from phosphoenolpyruvate of glycolysis and erythrose 4-phosphate of the pentose phosphate pathway as precursors. So, although phenol has a negative effect on *E. coli* growth, phenol would not inhibit its own biosynthesis from glucose. Regarding phenol production by *E. coli*, then, it would be effective to separate the growth phase from the production phases, because the TCA cycle is not involved in phenol synthesis. The reaction for phenol synthesis should be suppressed during the growth phase, and should be induced after the cells are sufficiently grown.

AUTHOR CONTRIBUTIONS

SK, YT, and HS conceived the study. SK designed the study and performed all the experiments and computational analysis. YT designed the study and helped to do the experiments and

computational analysis. HS designed and supervised the study. SK and YT drafted the manuscript. All authors contributed to preparing the final version of the manuscript and approved the manuscript to submit to this journal.

FUNDING

This work was supported by Grant-in-Aid for Scientific Research (B) No.16H04576, and Grant-in-Aid for Scientific Research (C) No. 18K04850.

REFERENCES

- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2006). Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab. Eng.* 8, 324–337. doi: 10.1016/j.ymben.2006.01.004
- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2007). Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* 9, 68–86. doi: 10.1016/j.ymben.2006.09.001
- Aono, R., Kobayashi, H., Joblin, K. N., and Horikoshi, K. (1994). Effects of organic solvents on growth of *Escherichia coli* K-12. *Biosci. Biotech. Biochem.* 58, 2009–2014.
- Brot, N., Smit, Z., and Weissbach, H. (1965). Conversion of L-tyrosine to phenol by clostridium tetanomorphum. *Arch. Biochem. Biophys.* 112, 1–6. doi: 10.1016/0003-9861(65)90002-0
- Castañero-Cerezo, S., Bernal, V., Post, H., Fuhrer, T., Cappadona, S., Sánchez-Díaz, N. C., et al. (2014). Protein acetylation affects acetate metabolism, motility and acid stress response in *Escherichia coli*. *Mol. Syst. Biol.* 10:762. doi: 10.15252/msb.20145227
- Castañero-Cerezo, S., Pastor, J. M., Renilla, S., Bernal, V., Iborra, J. L., and Cánovas, M. (2009). An insight into the role of phosphotransacetylase (*pta*) and the acetate/acetyl-CoA node in *Escherichia coli*. *Microb. Cell Fact.* 8:54. doi: 10.1186/1475-2859-8-54
- Kajihata, S., Furusawa, C., Matsuda, F., and Shimizu, H. (2014). OpenMebius: an open source software for isotopically nonstationary ^{13}C -based metabolic flux analysis. *Biomed. Res. Int.* 627014. doi:10.1155/2014/627014
- Keweloh, H., Diefenbach, R., and Rehm, H. J. (1991). Increase of phenol tolerance of *Escherichia coli* by alterations of the fatty acid composition of the membrane lipids. *Arch. Microbiol.* 157, 49–53.
- Keweloh, H., Weyrauch, G., and Rehm, H. J. (1990). Phenol-induced membrane changes in free and immobilized *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 33, 66–71.
- Kihira, C., Hayashi, Y., Azuma, N., Noda, S., Maeda, S., Fukui, S., et al. (2012). Alterations of glucose metabolism in *Escherichia coli* mutants defective in respiratory-chain enzymes. *J. Biotechnol.* 158, 215–223. doi: 10.1016/j.jbiotec.2011.06.029
- Kim, B., Park, H., Na, D., and Lee, S. Y. (2014). Metabolic engineering of *Escherichia coli* for the production of phenol from glucose. *Biotechnol. J.* 9, 621–629. doi:10.1002/biot.201300263
- Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., et al. (2005). Complete set of ORF clones of *Escherichia coli* ASKA library (a complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res.* 12, 291–299. doi: 10.1093/dnares/dsi012
- Kumagai, H., Yamada, H., Matsui, H., Ohkishi, H., and Ogata, K. (1970). Tyrosine phenol lyase. I. Purification, crystallization, and properties. *J. Biol. Chem.* 245, 1767–1772.
- Leighty, R. W., and Antoniewicz, M. R. (2012). Parallel labeling experiments with $[\text{U-}^{13}\text{C}]$ glucose validate *E. coli* metabolic network model for ^{13}C metabolic flux analysis. *Metab. Eng.* 14, 533–541. doi: 10.1016/j.ymben.2012.06.003
- Matsuoka, Y., and Shimizu, K. (2011). Metabolic regulation in *Escherichia coli* in response to culture environments via global regulators. *Biotechnol. J.* 6, 1330–1341. doi: 10.1002/biot.201000447
- Okahashi, N., Matsuda, F., Yoshikawa, K., Shirai, T., Matsumoto, Y., Wada, M., et al. (2017). Metabolic engineering of isopropyl alcohol-producing *Escherichia*

ACKNOWLEDGMENTS

We thank Prof. Fumio Matsuda (Osaka University, Japan) for his helpful comments.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.01010/full#supplementary-material>

- coli* strains with ^{13}C -metabolic flux analysis. *Biotechnol. Bioeng.* 114, 2782–2793. doi: 10.1002/bit.26390
- Peng, L., and Shimizu, K. (2003). Global metabolic regulation analysis for *Escherichia coli* K12 based on protein expression by 2-dimensional electrophoresis and enzyme activity measurement. *Appl. Microbiol. Biotechnol.* 61, 163–178. doi: 10.1007/s00253-002-1202-6
- Ramos, J. L., Duque, E., Gallegos, M. T., Godoy, P., Ramos-Gonzalez, M. I., Rojas, A., et al. (2002). Mechanisms of solvent tolerance in gram-negative bacteria. *Annu. Rev. Microbiol.* 56, 743–768.
- Salas, M., Vinuela, E., and Sols, A. (1965). Spontaneous and enzymatically catalyzed anomerization of glucose-6-phosphate and anomeric specificity of related enzymes. *J. Biol. Chem.* 240, 561–568.
- Schmidt, R. J. (2005). Industrial catalytic processes-phenol production. *Appl. Catal. A Gen.* 280, 89–103. doi: 10.1016/j.apcata.2004.08.030
- Soini, J., Ukkonen, K., and Neubauer, P. (2008). High cell density media for *Escherichia coli* are generally designed for aerobic cultivations – consequences for large-scale bioprocesses and shake flask cultures. *Microb. Cell Fact.* 7:26. doi:10.1186/1475-2859-7-26
- Srere, P. A. (1969). Citrate synthase. *Methods Enzymol.* 13, 3–11.
- Thompson, B., Machas, M., and Nielsen, D. R. (2016). Engineering and comparison of non-natural pathways for microbial phenol production. *Biotechnol. Bioeng.* 113, 1745–1754. doi: 10.1002/bit.25942
- van Winden, W. A., Wittmann, C., Heinzel, E., and Heijnen, J. J. (2002). Correcting mass isotopomer distributions for naturally occurring isotopes. *Biotechnol. Bioeng.* 80, 477–479. doi: 10.1002/bit.10393
- Verschueren, K. (1983). *Handbook of Environmental Data on Organic Chemicals*, 2nd Edn. New York, NY: Van Nostrand Reinhold.
- Wada, K., Toya, Y., Banno, S., Yoshikawa, K., Matsuda, F., Shimizu, H., et al. (2017). ^{13}C -metabolic flux analysis for mevalonate-producing strain of *Escherichia coli*. *J. Biosci. Bioeng.* 123, 177–182. doi: 10.1016/j.jbiosc.2016.08.001
- Wierckx, N. J., Ballerstedt, H., de Bont, J. A., and Wery, J. (2005). Engineering of solvent-tolerant *Pseudomonas putida* S12 for bioproduction of phenol from glucose. *Appl. Environ. Microbiol.* 71, 8221–8227. doi: 10.1128/aem.71.12.8221-8227.2005
- Wittmann, C. (2007). Fluxome analysis using GC-MS. *Microb. Cell Fact.* 6:6. doi: 10.1186/1475-2859-6-6
- Zamboni, N., Fendt, S. M., Rühl, M., and Sauer, U. (2009). ^{13}C -based metabolic flux analysis. *Nat. Protoc.* 4, 878–892. doi: 10.1038/nprot.2009.58
- Zhao, J., and Shimizu, K. (2003). Metabolic flux analysis of *Escherichia coli* K12 grown on ^{13}C -labeled acetate and glucose using GC-MS and powerful flux calculation method. *J. Biotechnol.* 101, 101–117. doi: 10.1016/s0168-1656(02)00316-4

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

Copyright © 2019 Kitamura, Toya and Shimizu. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



EMUlator: An Elementary Metabolite Unit (EMU) Based Isotope Simulator Enabled by Adjacency Matrix

Chao Wu¹, Chia-hsin Chen², Jonathan Lo¹, William Michener¹, PinChing Maness¹ and Wei Xiong^{1*}

¹ National Renewable Energy Laboratory, Golden, CO, United States, ² Institute of Nuclear Energy Research, Taoyuan, Taiwan

OPEN ACCESS

Edited by:

Manuel Kleiner,
North Carolina State University,
United States

Reviewed by:

Lian He,
University of Washington,
United States
Joshua Chan,
Colorado State University,
United States

*Correspondence:

Wei Xiong
wei.xiong@nrel.gov

Specialty section:

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

Received: 09 February 2019

Accepted: 11 April 2019

Published: 30 April 2019

Citation:

Wu C, Chen C, Lo J, Michener W,
Maness P and Xiong W (2019)
EMUlator: An Elementary Metabolite
Unit (EMU) Based Isotope Simulator
Enabled by Adjacency Matrix.
Front. Microbiol. 10:922.
doi: 10.3389/fmicb.2019.00922

Stable isotope based metabolic flux analysis is currently the unique methodology that allows the experimental study of the integrated responses of metabolic networks. This method primarily relies on isotope labeling and modeling, which could be a challenge in both experimental and computational biology. In particular, the algorithm implementation for isotope simulation is a critical step, limiting extensive usage of this powerful approach. Here, we introduce *EMUlator* a Python-based isotope simulator which is developed on Elementary Metabolite Unit (EMU) algorithm, an efficient and powerful algorithm for isotope modeling. We propose a novel adjacency matrix method to implement EMU modeling and exemplify it stepwise. This method is intuitively straightforward and can be conveniently mastered for various customized purposes. We apply this arithmetic pipeline to understand the phosphoketolase flux in the metabolic network of an industrial microbe *Clostridium acetobutylicum*. The resulting design enables a high-throughput and non-invasive approach for estimating phosphoketolase flux *in vivo*. Our computational insights allow the systematic design and prediction of isotope-based metabolic models and yield a comprehensive understanding of their limitations and potentials.

Keywords: adjacency matrix, elementary metabolite unit (EMU), fractional labeling (FL), *Clostridium acetobutylicum*, phosphoketolase

INTRODUCTION

¹³C metabolic flux analysis (MFA) is currently the only experimental methodology to quantitatively understand intracellular biochemical networks by means of stable isotope tracing, labeling pattern analysis and indispensably, metabolic modeling which is based on mass and isotope balancing (Stephanopoulos, 1999; Sauer, 2006). Importantly, isotope modeling can provide additional key information that further defines the system, enabling quantification of the fluxes in parallel or cyclic pathways that cannot be estimated reliably by mass balance only. A variety of mathematical models were developed to establish relationship between isotope distribution and metabolic flux, including isotopomers (Schmidt et al., 1997), cumomers (Wiechert et al., 1999), and bondomers (Van Winden et al., 2002). However, previous methods suffer from the computational challenge in resolving realistic and large-scale metabolic network, as a large number of isotopomer equations need to be solved.

To address this limitation, a creative computational framework based on Elementary Metabolite Unit (EMU) was proposed (Antoniewicz et al., 2007a). The EMUs of a metabolite are defined as the non-empty subsets of all that compound's atoms (usually carbon atom). EMU can be cataloged by size, i.e., the number of atoms in it. With atom transition map, this framework will trace and identify the minimal relevant metabolic information needed to simulate isotope patterns and solve the optimization problem, and therefore greatly reduces the number of balance equations and computation burden, i.e., 95% reduction of the variables needed for the simulation in an *E. coli* model without any loss of information (Antoniewicz et al., 2007b). To date, EMU algorithm has garnered increasing attention in metabolic analysis (Quek et al., 2009; Sokol et al., 2012; Weitzel et al., 2013; Kajihata et al., 2014; Shupletsov et al., 2014; Young, 2014). Better leveraging this approach in understanding cell metabolism will require the development of novel computational packages which are easy to program and can fit new and broad application scenarios. However, to date, computational approaches that can straightforwardly and efficiently implement the EMU algorithm are still insufficient.

Here, we develop a new computational toolbox for steady state metabolic modeling analysis, the *EMUlator*, which accomplishes EMU modeling through an adjacency matrix-based approach. In graph theory, an adjacency matrix is used to quantitatively represent a graph, of which the elements indicate the connectivity of vertex pairs in rows and columns. Essentially, metabolic network is a directed graph with branches, thus it can be transformed into adjacency matrix for the ease of programming. Utilizing adjacency matrix, the *EMUlator* can efficiently simulate isotope distributions for ^{13}C -MFA. To demonstrate its functionality, we decompose the EMUs for isotope simulation in a tricarboxylic acid (TCA) cycle which represents a realistic metabolic network. Furthermore, we applied this newly developed software in modeling and analyzing the flux of phosphoketolase pathway in *Clostridium acetobutylicum* xylose catabolism. By decomposing the network and simulating metabolite isotopomer patterns, we found a good correlation between phosphoketolase flux and the fractional labeling of acetate, which has never been characterized in an isotope tracer experiment. Coupled with GC-MS analysis of acetate, this *EMUlator*-enabled analysis leads to a novel and high-throughput methodology for quantitatively understanding the phosphoketolase pathway in response to environmental and genetic perturbation. As exemplified, *EMUlator* aims to be a universal and powerful tool for isotope tracer modeling and for gaining quantitative understanding of cell metabolism. The software and its instruction are available at **Supplementary Information**.

RESULTS

Overview of the *EMUlator* Pipeline

The *EMUlator* pipeline is designed in Python, capable of performing a complete isotope simulation and prediction of a metabolic network for ^{13}C -MFA. Previous tools, such as Metran (Antoniewicz et al., 2006, 2007a), OpenFlux (Quek

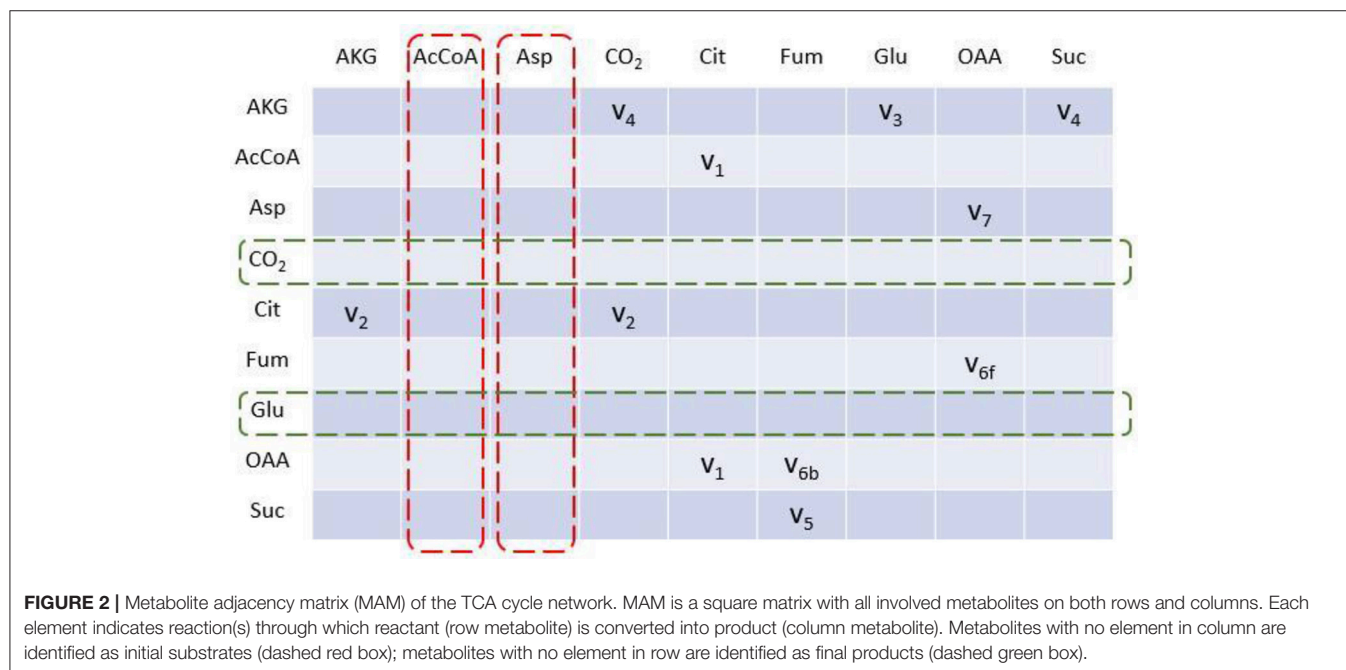
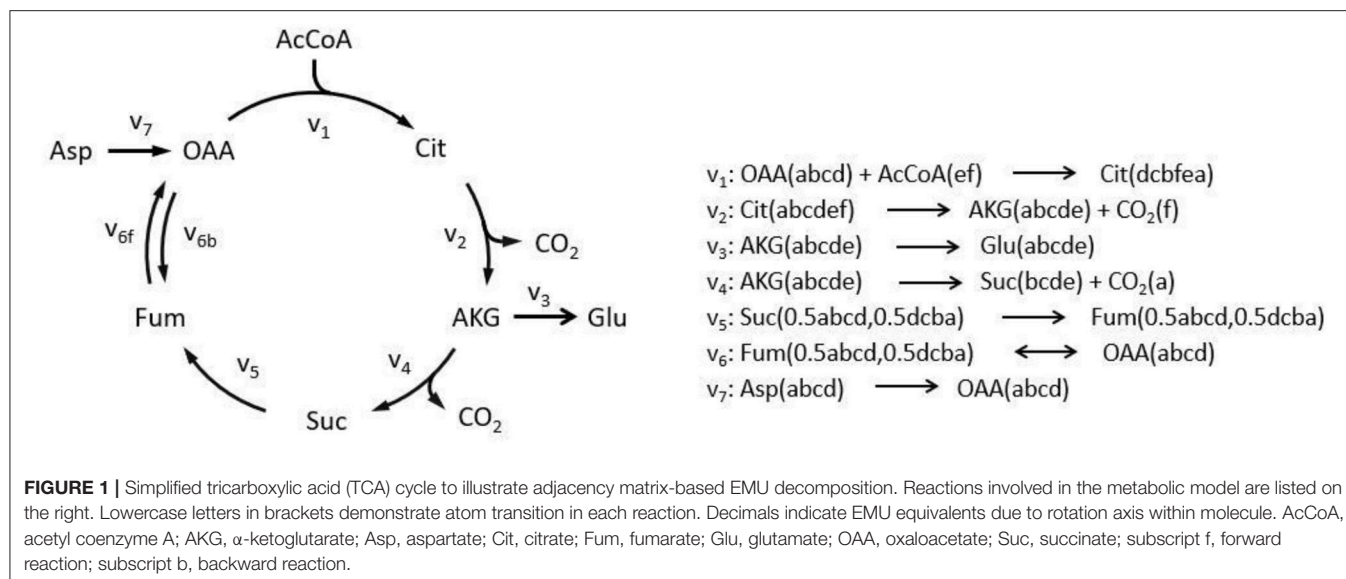
et al., 2009; Shupletsov et al., 2014) and INCA (Young, 2014), are able to perform such modeling, however based on Matlab platform which is not an open and free computing environment. In addition, previous EMU modeling was substantially less transparent than the one we present here. A key distinguishing feature of *EMUlator* is the usage of adjacency matrix. This ensures a graphic expression of the algorithm which can be understood intuitively and implemented iteratively. In particular, *EMUlator* provides a more detailed and principled procedure of EMU modeling, which decomposes metabolic network into EMU reactions, sets up EMU balances and simulates labeling distribution. Most importantly, the EMU deconstruction results in the reduction of the metabolic network model leading to a smaller set of EMU reactions which preserves all the information contained in it but decreases running time significantly.

EMUlator Transforms Metabolic Network Into Adjacency Matrix

To illustrate algorithm of the program comparably, we implement TCA cycle model as an example, as this representative metabolic network was also used in the original EMU work (Antoniewicz et al., 2007a). In this network, aspartate and acetyl coenzyme A (AcCoA) are the substrates, while CO_2 and glutamate are final products. Reactions with carbon atom transitions are listed in **Figure 1**. As a directed graph, any metabolic network with branches (due to cleavage and condensation reactions) can be transformed into a metabolite adjacency matrix (MAM). All metabolites are grouped in both row and column coordinates, thus forming a square matrix. Row metabolites appear as reactants while column metabolites are products of each reactions. Elements determined by row and column coordinates are connecting reactions for reactants and products. Reaction in element may not be unique because identical reactant and product could be involved in different reactions. As such, inputs and outputs of the network are easy to identify. Herein, columns without element are identified as substrates since they have no precursors (i.e., columns for AcCoA and Aspartate in dashed red boxes, **Figure 2**), while rows without element are identified as final products (i.e., rows for CO_2 and glutamate in dashed green boxes, **Figure 2**). Overall, MAM reflects the connectivity of metabolites in a network.

EMUlator Decomposes MAM Into EMU Adjacency Matrix (EAM)

EMU decomposition of a metabolic network can start at the size of EMU(s) that need to be simulated. In this example, we simulate the Mass Distribution Vector (MDV, the fractional abundance of each isotopolog normalized to the sum of all possible isotopologues.) (Nanchen et al., 2007) of glutamate Glu_{12345} (size 5). All EMU reactions that are needed for this simulation can be identified iteratively via MAM. Glu_{12345} as a product can be found in the column of EAM (size 5) (see **Figure 3A**), its precursor AKG_{12345} is illustrated in the corresponding row through the reaction V_3 . Similarly, for the product AKG_{12345} , we can locate its precursor Cit_{12345} via v_2 . Lastly for the product (in column) Cit_{12345} , we identify reaction



v_1 , in which both OAA_{234} and AcCoA_{12} are the reactants. Since EMUs of smaller size are identified in condensation reactions, they will be used as new start points for searching. Therefore, we can follow the OAA_{234} (AcCoA_{12} is identified as an EMU of the substrate, and thus the searching stops), and search all other EMUs at size 3 (Figure 3B). All precursor EMUs for multiple precursors [e.g., due to equivalent EMUs (Antoniewicz et al., 2007a)], can be identified through breadth-first search. As such, adjacent matrix provides a straightforward and iterative path allowing us to trace back EMUs of smaller sizes until the EMUs of network substrates are identified (i.e., Aspartate and AcCoA in this example). Once all set of the EMUs are obtained, EMUs

can be arranged into different EAMs per size. Similar to MAM, row and column coordinates of EAM correspond to reactants and products of each reaction, respectively, with the difference that EMUs subject to convolution also appear in rows of EAM, and the coefficients of an element equal to the stoichiometric coefficients of corresponding reactant and product. Complete EAMs after EMU decomposition of the example are shown in Figure 3.

EMULATOR Significantly Reduces the Size of EAMs

EMULATOR can also reduce the scale of EAM. In steady state isotope modeling, labeling pattern can only be modified by

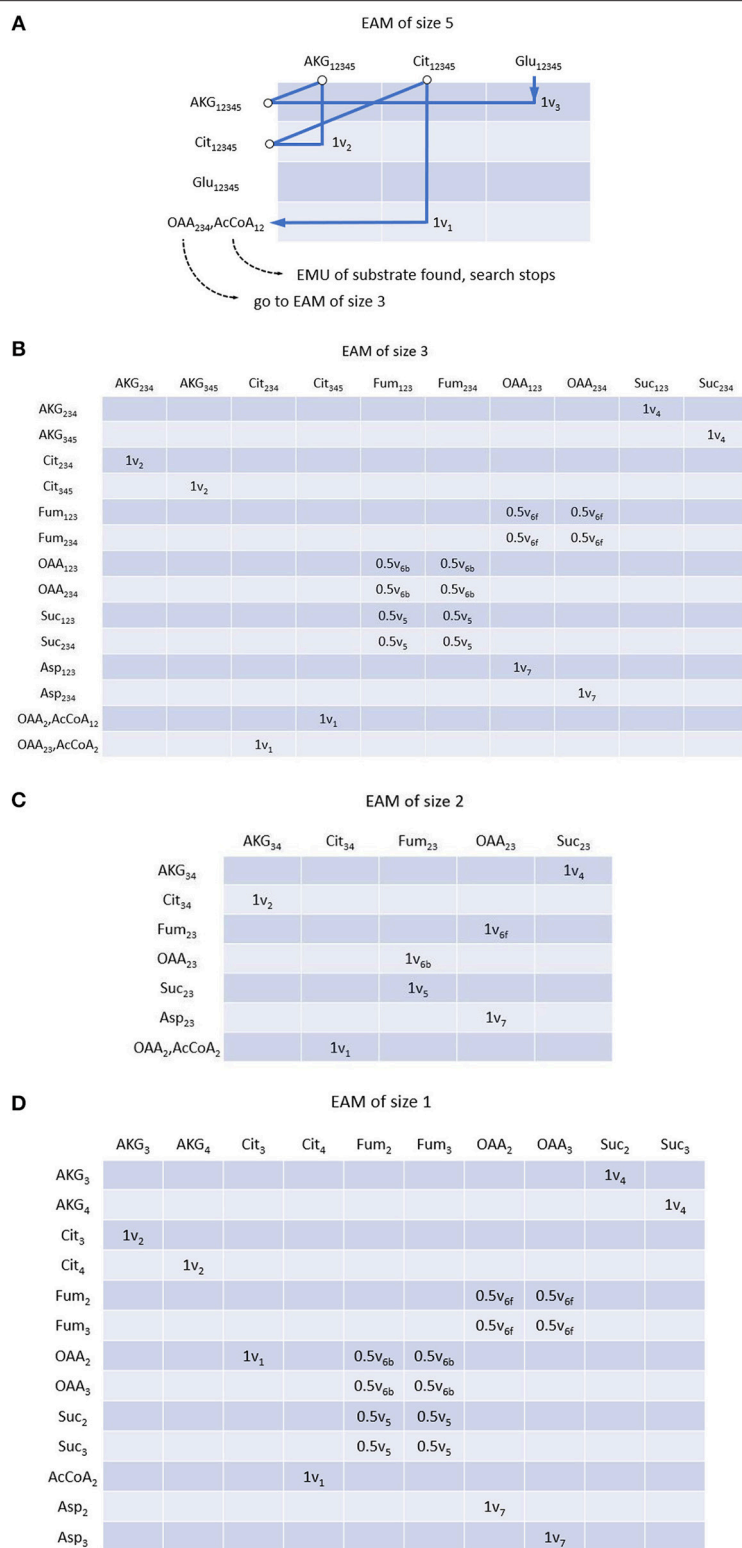


FIGURE 3 | EMU adjacency matrix (EAM) of size 5 (**A**), size 3 (**B**), size 2 (**C**) and size 1 (**D**). Each element indicates reaction(s) through which reactant (row metabolite) is converted into product (column metabolite) as well as reaction coefficients. Construction of EAM starts from the EMU to be simulated, i.e., Glu₁₂₃₄₅ in this example. All precursors of corresponding EMU can be found with MAM and atom transition using breadth-first search. Precursor searching continues until initial substrates. If condensation reactions are encountered, precursors of smaller size are considered as new start points, respectively. EMUs of same size are located in the same EAM.

convolution of two or more metabolites and unimolecular reactions may be lumped without affecting simulations which helps to reduce the scale of EAM. Unimolecular reactions can be easily identified in EAM as those with solo element in a column, which means its corresponding product only has a single source. Those columns will be deleted with identical metabolites in rows all renamed by their precursors. For example, in EAM of size 1, after AKG₃ column is eliminated, AKG₃ in row will be renamed with Cit₃ which produces AKG₃ by reaction v_2 (Figure 4). Moreover, since Cit₃ is still from a unimolecular reaction (v_1), we eliminate the column as well and Cit₃ in row will be renamed by its precursor OAA₂ via v_1 . The deletion of the column and renaming of corresponding row continue until there is no solo-element column in EAM. Finally, multiple rows with identical metabolite name will be combined, indicating the identical reactant and product are connected by different reactions.

EMUlator Identifies and Combines Equivalent EMUs

Rotationally symmetric molecules (i.e., fumaric acid and succinic acid) will give rise to equivalent EMUs (Antoniewicz et al., 2007a), which are undistinguishable for enzyme, and react identically in the reactions. EMUlator can combine those EMUs as they will have the same probability to get certain labeling pattern. Metabolites which could generate equivalent EMUs are indicated with fractional carbon atoms in Figure 1. Fum₂ and Fum₃ are equivalent EMUs of size 1 in this example (Figure 5). Element coefficients of row equivalent EMUs are combined, while coefficients of column equivalent EMUs are combined and divided by the number of equivalents. Eventually, EMU variables were reduced from 24 of initial EMU model to 9 after lumping unimolecular reactions and combining equivalent EMU, yielding the same results as the original EMU work (Antoniewicz et al., 2007a).

EMUlator Establishes EMU Balances From EAMs

To simulate the labeling pattern of a given metabolite, EMU balances are established from EMU reaction subnetworks of different size, and MDVs can be calculated according to:

$$X_i = A_i^{-1} \cdot B_i \cdot Y_i$$

Where A_i and B_i are matrix function of flux variable of size i . A_i is square matrix whose shape is dependent on the number of EMU balance m of current size. B_i 's shape is $m \times n$ where n is the number of available EMU variables (Antoniewicz et al., 2007a). Simulation starts from size 1, and Y_1 are MDVs of network substrates. Other MDVs are calculated and then used in Y of larger size. A_i and B_i can be easily deduced from EAM as demonstrated in Figure 6. Diagonal elements of EAM are first set to be the negative sum of other elements of current column. Then the transposed upper square submatrix will be A_1 , and the lower submatrix will be B_1 after transposition and all elements negated. The transformation is made according to the isotopomer balance which states that the sum of all influx to an EMU multiplied by its MDV ($\sum v_i \cdot MDV$) is equal to the sum of the individual

product of each influx v_i multiplied by MDV _{i} ($\sum(v_i \cdot MDV_i)$). Apparently, diagonal element represents the total influx of corresponding EMU. Multiplied by MDV of balanced EMU, the product is equal to the sum of all labeling pattern sources, including those unknown (X_i) and known (Y_i) EMU variables. EMU balances of larger size can be established likewise until the Glu₁₂₃₄₅ are eventually simulated. EMU balances of all sizes are shown below:

EMU balance of size 1

$$\begin{bmatrix} -v_{6b} - v_5 & 0.5v_{6b} + 0.5v_5 & 0.5v_{6b} \\ v_{6f} & -v_{6f} - v_7 & 0 \\ v_{6f} & 0 & -v_{6f} - v_7 \end{bmatrix} \cdot \begin{bmatrix} MDV_{Fum_2} \\ MDV_{OAA_2} \\ MDV_{OAA_3} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -0.5v_5 \\ -v_7 & 0 & 0 \\ 0 & -v_7 & 0 \end{bmatrix} \cdot \begin{bmatrix} MDV_{Asp_2} \\ MDV_{Asp_3} \\ MDV_{AcCoA_2} \end{bmatrix}$$

EMU balance of size 2

$$\begin{bmatrix} -v_{6b} - v_5 & v_{6b} \\ v_{6f} & -v_{6f} - v_7 \end{bmatrix} \cdot \begin{bmatrix} MDV_{Fum_{23}} \\ MDV_{OAA_{23}} \end{bmatrix} = \begin{bmatrix} 0 & v_5 \\ -v_7 & 0 \end{bmatrix} \cdot \begin{bmatrix} MDV_{Asp_{23}} \\ MDV_{OAA_2} \times MDV_{AcCoA_2} \end{bmatrix}$$

EMU balance of size 3

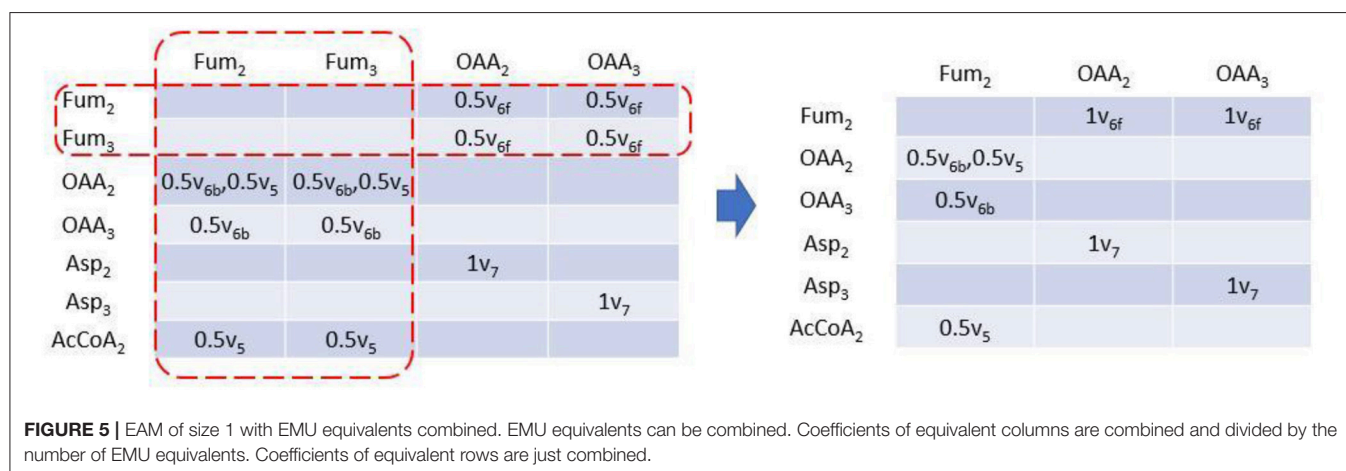
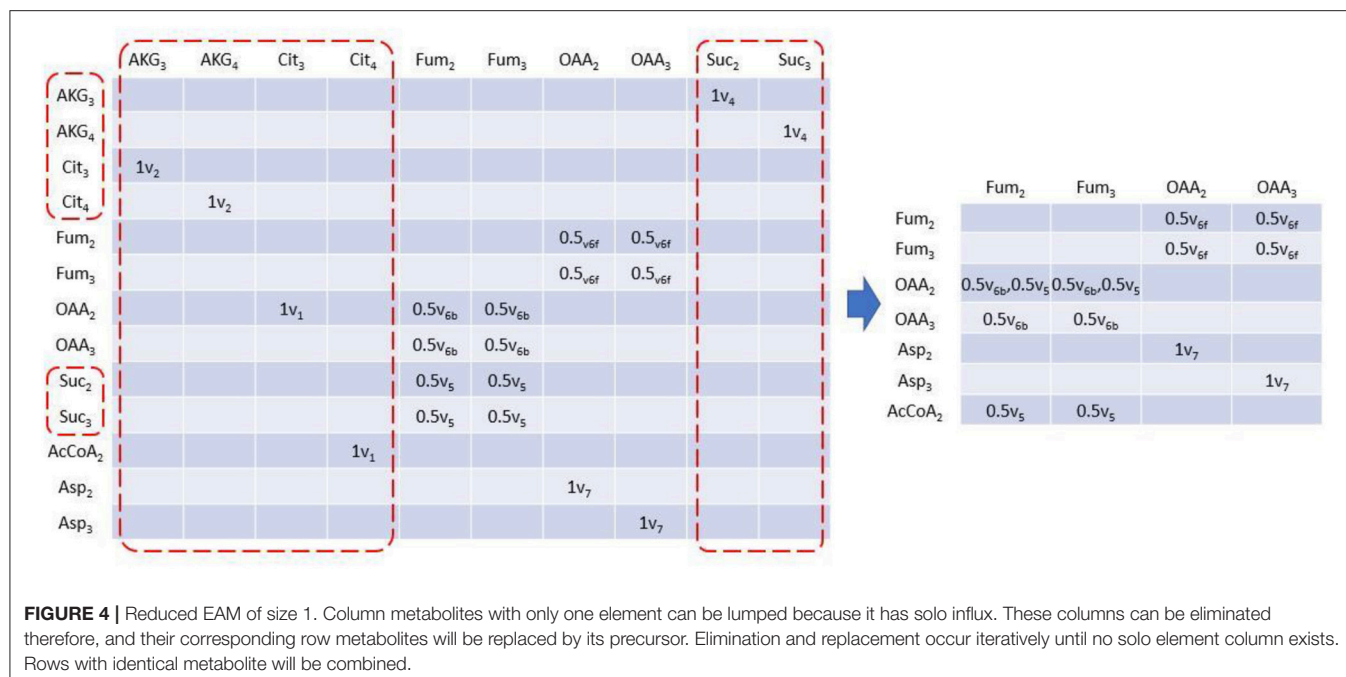
$$\begin{bmatrix} -v_{6b} - v_5 & 0.5v_{6b} & 0.5v_{6b} \\ v_{6f} & -v_{6f} - v_7 & 0 \\ v_{6f} & 0 & -v_{6f} - v_7 \end{bmatrix} \cdot \begin{bmatrix} MDV_{Fum_{123}} \\ MDV_{OAA_{123}} \\ MDV_{OAA_{234}} \end{bmatrix} = \begin{bmatrix} -0.5v_5 & -0.5v_5 & 0 & 0 \\ 0 & 0 & -v_7 & 0 \\ 0 & 0 & 0 & -v_7 \end{bmatrix} \cdot \begin{bmatrix} MDV_{OAA_2} \times MDV_{AcCoA_{12}} \\ MDV_{OAA_{23}} \times MDV_{AcCoA_2} \\ MDV_{Asp_{123}} \\ MDV_{Asp_{234}} \end{bmatrix}$$

EMU balance of size 4

$$[-v_3] \cdot [MDV_{Glu_{12345}}] = [-v_3] \cdot [MDV_{OAA_{123}} \times MDV_{AcCoA_{12}}]$$

EMUlator Designs Optimal ¹³C-Tracer Experiment for Quantifying Metabolic Flux of Interest

To demonstrate the performance of EMUlator in a large-scale setting that reflects the complexity of realistic cell metabolism, we next performed simulation of isotope distributions in a larger network model of *C. acetobutylicum*. *C. acetobutylicum* is a solventogenic clostridium and represents a promising chassis microbe capable of utilizing lignocellulose-derived pentose sugar (i.e., xylose) for biofuels production (Mitchell, 1998; Gu et al., 2011). In this case, pentose catabolism through phosphoketolase pathway (Grimmler et al., 2010; Servinsky et al., 2010; Liu et al., 2012) is of special interest in that this pathway has been recognized as a key target for constructing synthetic pathways (e.g., Non-oxidative glycolysis) that bypass CO₂ loss via pyruvate decarboxylase and thus enhance carbon yield in final products



(Bogorad et al., 2013). Here we used the adjacency matrix-based *EMULATOR* to simulate labeling patterns of metabolites and show how it facilitates the selection of best isotope substrates and readouts quantifying the *in vivo* phosphoketolase activity.

First, a biochemical network for xylose metabolism of *C. acetobutylicum* was constructed, based on the genome information (Nölling et al., 2001; Bao et al., 2011). As shown in **Figure 7A**, after phosphorylation, xylose can be metabolized either through the non-oxidative pentose phosphate pathway or cleaved by phosphoketolase to form acetyl-phosphate and glyceraldehyde-3-phosphate. Acetyl-phosphate is further directed to generate extracellular fermentative products (i.e., acetate, ethanol, acetone, butanol, butyrate) and glyceraldehyde-3-phosphate can enter pentose phosphate pathway in which reactions are highly reversible due to the nature of isomerase, epimerase, transketolase, and transaldolase. The oxidative

pentose phosphate pathway is not considered which was verified to be inactive in *C. acetobutylicum* (Au et al., 2014). The TCA cycle is not included as it does not influence the labeling patterns of the upstream metabolites.

We selected acetate (AC), ethanol, 3-phosphoglycerate, erythrose-4-phosphate and ribose-5-phosphate as candidate readouts for reflecting phosphoketolase activity since MDVs of these metabolites can be experimentally obtained either from direct determination or derivation from amino acid MDVs (Nanchen et al., 2007). Meanwhile, we tested all commercially available xylose tracers: 1-¹³C xylose, 2-¹³C xylose, 3-¹³C xylose, 4-¹³C xylose, 5-¹³C xylose, 1,2-¹³C xylose, U-¹³C xylose. MDVs were simulated using *EMULATOR* in all possible combinations of these candidate readouts and tracers. Goodness of correlation between Fractional Labeling (FL) (defined in Materials and Methods) and flux ratio through phosphoketolase (i.e., v_2/v_1 in

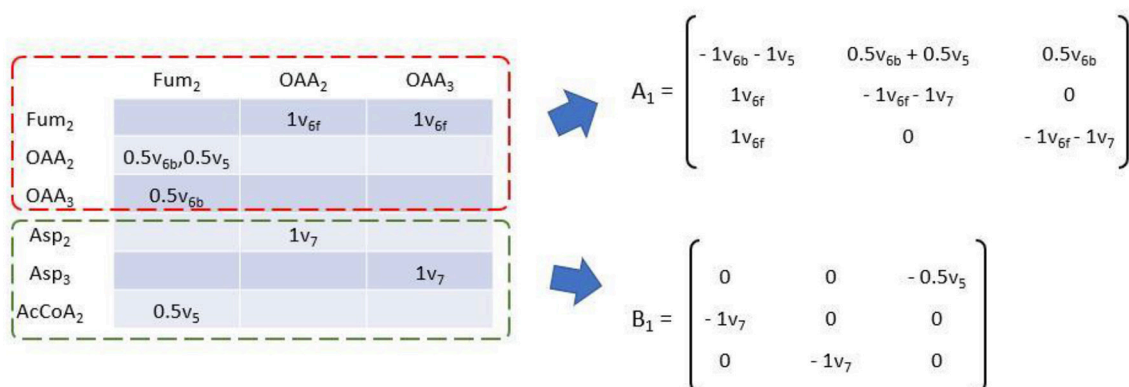


FIGURE 6 | Transformation of EAM to establish EMU balance of size 1. Matrix function A of flux variables can be obtained by transposition of the upper square submatrix (dashed red box) with negative sums of each column of original EAM as diagonal elements. Matrix function B can be obtained by transposition of the lower submatrix (dashed green box) with all elements taken opposite sign.

Figure 7A) and range of effective FL are used as the selection criteria. The modeling results are shown in **Figure 8**.

Among various xylose tracers, 1,2-¹³C xylose yields EMU AC₁₂ with best correlation between its FL and phosphoketolase flux ratio (Spearman correlation coefficient = 0.7). The widest effective range of FL (slope of regression line = 0.43) indicates a good sensitivity on the flux ratio of phosphoketolase (**Figure 8A**). EMU AC₁ and AC₂ show identical correlation between FL and flux ratio with that of AC₁₂ (**Figures 8B,C**) because the carbons of acetate originate from C2 and C1 of xylose which are both labeled (probably also from C4 and C5 of xylose converted from AcCoA) and behave equivalently in the atom transition. FL of G3P₁₂₃ could have a good correlation with the phosphoketolase flux ratio, which is, however, disturbed by many randomly distributed points (**Figure 8E**). This is probably due to the reversibility of the reactions in glycolytic and pentose phosphate pathways. EtOH₁₂, E4P₁₂₃₄ and R5P₁₂₃₄₅ all show poor correlation, and thus cannot be used as the indicators for phosphoketolase flux (**Figures 8D,F,G**). As for other tracers, 1-¹³C xylose and 2-¹³C xylose labeling result in poor correlation in AC₁ and AC₂, respectively (**Supplementary Figures 1, 2**). AC₁₂ and EtOH₁₂ will be totally unlabeled using 3-¹³C xylose tracer as no C3 of xylose will fractionate into these metabolites. G3P₁₂₃ shows a good correlation with phosphoketolase flux while the range of effective FL is too small to determine phosphoketolase activity (**Supplementary Figure 3**). If 4-¹³C xylose and 5-¹³C xylose are used as substrate, correlation between FL and flux ratio are inverted for most of the EMUs. In addition, AC₂ and AC₁ are totally unlabeled per 4-¹³C xylose and 5-¹³C xylose, respectively. (**Supplementary Figures 4, 5**). As a control, FLs of all MDVs are constant when fed with a mixture of U-¹³C labeled and unlabeled xylose (**Supplementary Figure 6**), which is, therefore excluded from the isotope tracer selection. In comparison of all tracer/readout combination, 1,2-¹³C-xylose/ AC₁₂ performed the best and this selection paves the way to the experimental measurement of flux in phosphoketolase pathway.

Guided by the above simulations, *C. acetobutylicum* was thereafter cultivated in 5 g L⁻¹ 1,2-¹³C xylose in following experiment. Fermentation kinetics including cell growth and the production of cell products over time are as shown in **Supplementary Figure 7**. The flux ratio from phosphoketolase pathway was then quantified by harvesting the supernatant and determining the isotope pattern of AC with GC-MS. MDVs of AC₁ and AC₁₂ were measured, and MDVs of AC₂ can be deduced. Accordingly, FL of AC₁, AC₂ and AC₁₂ were calculated to be 0.462 ± 0.018 , 0.469 ± 0.012 and 0.465 ± 0.015 , respectively, which is consistent with our prediction that FLs of all fragments' EMUs should be identical. A distribution of phosphoketolase flux ratio was obtained from simulated sample points with corresponding FL of AC₁, AC₂ and AC₁₂ falling into the measured ranges. The average and 95% confidence interval of the flux ratio through phosphoketolase were estimated as 22.8% and 14.3–36.6%, respectively under current conditions (**Figure 7B**).

DISCUSSION

Although equivalent to isotopomer models, EMU method is able to significantly reduce the variables needed to simulate labeling patterns of metabolites (90% reduction in our *C. acetobutylicum* case) without any loss of information, which therefore greatly facilitates ¹³C flux modeling of realistic and large-scaled metabolic network and dynamic systems (Young et al., 2008). Here we introduce a computational tool *EMUlator* to the metabolic research community. This engine utilizes adjacency matrix-based approach which is intuitively straightforward and easy to program. Metabolic network can be decomposed into EAMs of different size which can be further reduced by lumping unimolecular reactions and combining equivalent EMUs. For MDV simulation, matrix multiplication starts from EAM of the smallest size, and iteratively continues to larger size until the required EMU is simulated. Overall, the computation time for

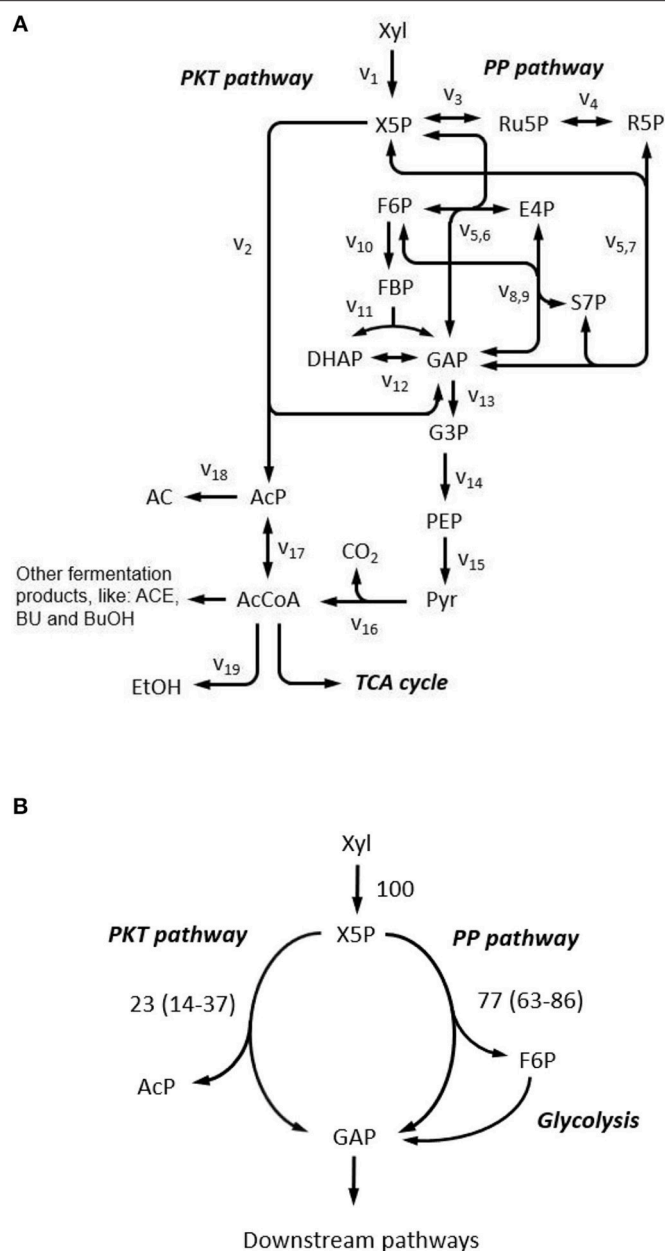


FIGURE 7 | Xylose metabolism in *C. acetobutylicum*. **(A)** Reaction network constructed for xylose metabolism. Complete reaction list is provided in **Supplementary Table 1**. AC, acetate; ACE, acetone; AcP, Acetyl phosphate; BU, butyrate; BuOH, butanol; DHAP, dihydroxyacetone phosphate; E4P, erythrose-4-phosphate; EtOH, ethanol; F6P, fructose-6-phosphate; FBP, fructose-1,6-bisphosphate; G3P, 3-phosphoglycerate; GAP, glyceraldehyde-3-phosphate; PEP, phosphoenolpyruvate; Pyr, pyruvate; R5P, ribose-5-phosphate; Ru5P, ribulose-5-phosphate; S7P, sedoheptulose-7-phosphate; X5P, xylulose-5-phosphate; Xyl, xylose; PKT pathway, phosphoketolase pathway; PP pathway, pentose phosphate pathway; TCA cycle, tricarboxylic acid cycle. **(B)** Metabolic flux ratio through phosphoketolase at 5 g L⁻¹ xylose. Flux ratios are relative to total xylose uptake rate, and values are presented as the mean of two replicates and 95% confidence intervals are provided in the parentheses.

EMULATOR to perform MDV simulation depends on the network complexity, or connectivity which can be represented by the number of non-zero elements in EAMs. For a realistic and moderate-sized network shown in the above examples, the time complexity is roughly $O(n^3)$, where n is the number of EMUs.

The EMULATOR allowed large-scale and efficient isotope modeling. To exemplify its capability, we applied it in

quantitative understanding of the phosphoketolase pathway in the central carbon metabolism of an industrial model microbe (*C. acetobutylicum*). We simulated labeling pattern of both intracellular and secretory metabolites using different xylose tracers, and identified the best tracer/readout combination reflecting phosphoketolase activity. ¹³C-flux measurement of the phosphoketolase pathway was enabled recently (Liu et al.,

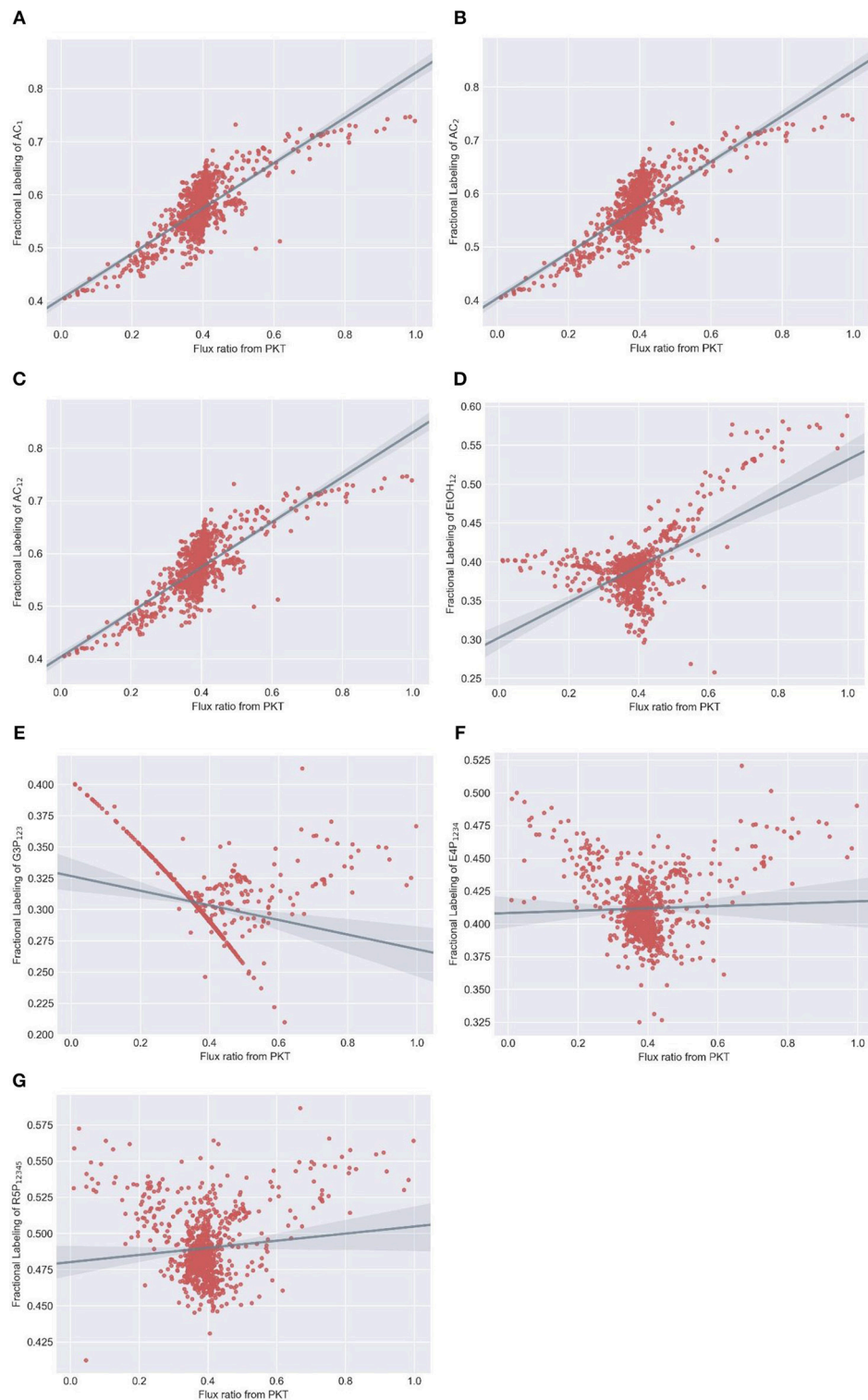


FIGURE 8 | Simulated fractional labeling (FL) of metabolites at different flux ratio from phosphoketolase with 100% 1,2-¹³C xylose as substrate. Random fluxes are generated 1000 times subjecting to xylose metabolism network. MDVs of (A) AC₁, (B) AC₂, (C) AC₁₂, (D) EtOH₁₂, (E) G3P₁₂₃, (F) E4P₁₂₃₄ and (G) R5P₁₂₃₄₅ are simulated using adjacency matrix-based EMU decomposition method proposed in this work. Metabolite FLs and flux ratio from PKT are subsequently calculated and plotted correspondingly. Regression line and 95% confidence intervals are also plotted.

2012), in which 3-phosphoglycerate was used as the indicator for phosphoketolase flux measurement in a 1- ^{13}C -xylose labeling experiment. This design is reasonable as the FL of 3-phosphoglycerate is monotonically reduced with the increased flux through phosphoketolase. The limitation is that 3-phosphoglycerate is not a direct product of phosphoketolase, therefore the flux estimation is largely dependent on the assumption that the labeling pattern of 3-phosphoglycerate and glyceraldehyde 3-phosphate, the product of phosphoketolase are identical. In our case, we compared all available xylose tracers and readouts with a more realistic metabolic network which also takes reversible reactions into account. Our results demonstrated that acetate (with EMUs of AC_1 , AC_2 , and AC_{12}) can be a better readout due to the strong correlations with phosphoketolase flux and a wide effective range of FL. Indeed, acetate can be exclusively derived from acetyl phosphate which is directly produced from phosphoketolase. More importantly, acetate is an extracellular metabolite and its isotope pattern can be easily measured by GC-MS without derivatization. The convenience of measurement without breaking the cells provides a high-throughput and non-invasive method for prompt ^{13}C -flux estimation, which, to our knowledge, was never developed previously. This estimation is based on the numerical relationship between fractional labeling of acetate and flux ratio through phosphoketolase which are positively correlated, even though it may not be strictly linear. It should be noticed that in the FL range 0.5–0.65, multiple flux ratio values are possible for a given value of FL. The jitter could be due to the reversibility of biochemical reactions and partial dependency of the EMU basis vectors (Crown and Antoniewicz, 2012), which cannot make all free fluxes absolutely solvable using only acetate labeling data. To further obtain a more precise prediction of phosphoketolase activity, advanced multiple regression method could be applied such as machine learning using FLs of other relevant EMUs as the training features.

We believe that the EMU modeling based on adjacency matrix approach opens a number of new possibilities in metabolic network analysis and ^{13}C -MFA. First, as exemplified, *EMUlator* can be used to select metabolites as readouts reflecting directly *in vivo* enzyme activities, and can also be used to do tracer simulations (Metallo et al., 2009; Young, 2014) which predict the labeling results of metabolic network models ahead of “wet” experiments, leading to further refinement. More promisingly, *EMUlator* is developed toward solving the inverse problem to estimate intracellular fluxes through an optimization search that minimize the sum-of-squared residuals between computationally simulated and experimentally determined measurements. With the EMU method, metabolic flux estimation can be further extended to computation-intensive scenarios as genome scale network (Gopalakrishnan and Maranas, 2015) and transient labeling process (Hendry et al., 2019). This task cannot be accomplished by other isotope modeling methods due to a tremendous computational burden. Currently, we are engaged in the development of an updated version focusing on the *de novo* and complete solution of ^{13}C -MFA, in which a global flux distribution can be estimated either from steady-state labeling or kinetic labeling experiments. It is our hope

that *EMUlator* will benefit the community and fuel metabolic research as the basis for innovative development of metabolic analysis tools.

MATERIALS AND METHODS

Implementation of EMU Algorithm in *EMUlator*

The TCA cycle example in the results illustrates the adjacency matrix approach for implementing EMU algorithm. The software and its instruction and are detailed in **Supplementary Information**.

Strain, Culture Conditions, and Medium

Clostridium acetobutylicum ATCC 824 was used in all experiments. For growth studies and biochemical characterization, *C. acetobutylicum* cells were grown anaerobically in 37°C in CTFUD defined medium (Olson and Lynd, 2012) which contains 3 g L^{-1} $\text{Na}_3\text{C}_6\text{H}_5\text{O}_7 \cdot 2\text{H}_2\text{O}$, 1.3 g L^{-1} $(\text{NH}_4)_2\text{SO}_4$, 1.5 g L^{-1} KH_2PO_4 , 0.13 g L^{-1} $\text{CaCl}_2 \cdot 2\text{H}_2\text{O}$, 0.5 g L^{-1} L-Cysteine-HCl, 11.56 g L^{-1} MOPS sodium salt, 2.6 g L^{-1} $\text{MgCl}_2 \cdot 6\text{H}_2\text{O}$, 0.001 g L^{-1} $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, 0.5 mL L^{-1} Resazurin 0.2% (w/v), supplemented with Wolfe's Vitamin solution (ATCC). D-xylose was supplied at concentration of 5 g L^{-1} as a carbon source. The cultures were started with the optical density at 600 nm ($\text{OD}_{600} = 0.05\text{--}0.08$) and performed in mid-log-phase. For labeling experiments, 1,2- ^{13}C -labeled xylose (99% pure; Cambridge Isotope Laboratories, Tewksbury, MA) was added to media at the concentration of 5 g L^{-1} . *C. acetobutylicum* strains were kept by freezing log-phase cultures at -80°C with 10% glycerol.

Quantitative Analysis of Fermentation Products

Cell growth was monitored by measuring the absorbance at OD_{600} with a Spectronic 21D UV-Visible Spectrophotometer (Milton Roy, Houston, TX). To analyze extracellular metabolites, cell samples were harvested by centrifugation at 13,000 g for 10 min. After filtration with 0.2 μm filter, the supernatant was analyzed by Agilent 1200 high pressure liquid chromatography (HPLC) (Agilent Technologies, Santa Clara, CA) and injected into a Bio-Rad Aminex HPLX-87H column with a Micro Guard Cation H Cartridge. 4 M H_2SO_4 was used as mobile phase at a flow rate of 0.6 mL/min. The column temperature was set to 55°C . Metabolites were detected by refractive index detector and UV/VIS detector.

Isotope Analysis

Labeling pattern of proteinogenic amino acids from cell mass were analyzed by Gas Chromatograph-mass spectrometry (GC-MS) as detailed in Xiong et al. (2018). The labeling pattern of acetate in the supernatant was directly analyzed by GC-MS without derivatization. Analysis of samples was performed on an Agilent 6890N GC equipped with a 5973 MS Detector (Agilent Technologies, Palo Alto, CA). Samples were injected

at a volume of 1 uL in splitless mode, and the analyte of interest was separated on a Restek Stabilwax-DA column (Restek Corporation, Bellefonte, PA). A flow of 1 mL min⁻¹ was held constant throughout the run with the following temperature profile: 35°C, hold for 3 min; ramped at 10 °C min⁻¹ to 225°C, hold for 1 min; ramped at 15°C min⁻¹ to 250°C, hold for 5 min.

Isotope Modeling From the Metabolic Network of *C. acetobutylicum*

Simulations were repeated 1,000 times with metabolic fluxes randomly generated subjecting to mass balances determined by reactions listed in **Supplementary Table 1**. Fractional labeling (FL) was calculated to indicate labeling status of metabolites according to:

$$FL = \frac{\sum_{i=0}^n i \cdot m_i}{n}$$

where n is the number of carbon in a EMU, and m_i represents components of MDV (Nanchen et al., 2007).

REFERENCES

- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2006). Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab. Eng.* 8, 324–337. doi: 10.1016/j.ymben.2006.01.004
- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2007a). Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* 9, 68–86. doi: 10.1016/j.ymben.2006.09.001
- Antoniewicz, M. R., Kraynie, D. F., Laffend, L. A., González-Lergier, J., Kelleher, J. K., and Stephanopoulos, G. (2007b). 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. doi: 10.1016/j.ymben.2007.01.003
- Au, J., Choi, J., Jones, S. W., Venkataramanan, K. P., and Antoniewicz, M. R. (2014). Parallel labeling experiments validate *Clostridium acetobutylicum* metabolic network model for (13)C metabolic flux analysis. *Metab. Eng.* 26, 23–33. doi: 10.1016/j.ymben.2014.08.002
- Bao, G., Wang, R., Zhu, Y., Dong, H., Mao, S., Zhang, Y., et al. (2011). Complete genome sequence of *Clostridium acetobutylicum* DSM 1731, a solvent-producing strain with multireplicon genome architecture. *J. Bacteriol.* 193, 5007–5008. doi: 10.1128/JB.05596-11
- Bogorad, I. W., Lin, T. S., and Liao, J. C. (2013). Synthetic non-oxidative glycolysis enables complete carbon conservation. *Nature* 502, 693–697. doi: 10.1038/nature12575
- Crown, S. B., and Antoniewicz, M. R. (2012). Selection of tracers for 13C-metabolic flux analysis using elementary metabolite units (EMU) basis vector methodology. *Metab. Eng.* 14, 150–161. doi: 10.1016/j.ymben.2011.12.005
- Gopalakrishnan, S., and Maranas, C. D. (2015). 13C metabolic flux analysis at a genome-scale. *Metab. Eng.* 32, 12–22. doi: 10.1016/j.ymben.2015.08.006
- Grimmler, C., Held, C., Liebl, W., and Ehrenreich, A. (2010). Transcriptional analysis of catabolite repression in *Clostridium acetobutylicum* growing on mixtures of D-glucose and D-xylose. *J. Biotechnol.* 150, 315–323. doi: 10.1016/j.jbiotec.2010.09.938
- Gu, Y., Jiang, Y., Wu, H., Liu, X., Li, Z., Li, J., et al. (2011). Economical challenges to microbial producers of butanol: feedstock, butanol ratio and titer. *Biotechnol. J.* 6, 1348–1357. doi: 10.1002/biot.201100046
- Hendry, J. I., Gopalakrishnan, S., Ungerer, J., Pakrasi, H. B., Tang, Y. J., and Maranas, C. D. (2019). Genome-scale fluxome of *Synechococcus elongatus*

AUTHOR CONTRIBUTIONS

CW developed the software and analyzed data. WX and CW designed the experiment. CC, JL, and WM conducted the experiment. WX proposed the idea and research questions, guided all stages of the research. CW and WX prepared the manuscript with editing from PM, JL, and CC.

ACKNOWLEDGMENTS

We appreciate Prof. Maciek R. Antoniewicz at the University of Delaware for his helpful suggestion on this work. This work was supported by the National Renewable Energy Laboratory LDRD project 0600.10001.18.42.01 (CW, JL, PM, and WX), the DOE Bioenergy Technology Office project under the agreement number 34715 (CW and WX), and Institute of Nuclear Energy Research in Taiwan ROC (CC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00922/full#supplementary-material>

- UTEX 2973 Using Transient ¹³C-labeling data. *Plant Physiol.* 179, 761–769. doi: 10.1104/pp.18.01357
- Kajihata, S., Furusawa, C., Matsuda, F., and Shimizu, H. (2014). OpenMebius: an open source software for isotopically nonstationary 13C-based metabolic flux analysis. *Biomed. Res. Int.* 2014:627014. doi: 10.1155/2014/627014
- Liu, L., Zhang, L., Tang, W., Gu, Y., Hua, Q., Yang, S., et al. (2012). Phosphoketolase pathway for xylose catabolism in *Clostridium acetobutylicum* revealed by 13C metabolic flux analysis. *J. Bacteriol.* 194, 5413–5422. doi: 10.1128/JB.00713-12
- Metallo, C. M., Walther, J. L., and Stephanopoulos, G. (2009). Evaluation of 13C isotopic tracers for metabolic flux analysis in mammalian cells. *J. Biotechnol.* 144, 167–174. doi: 10.1016/j.jbiotec.2009.07.010
- Mitchell, W. J. (1998). Physiology of carbohydrate to solvent conversion by clostridia. *Adv. Microb. Physiol.* 39, 31–130. doi: 10.1016/S0065-2911(08)60015-6
- Nanchen, A., Fuhrer, T., and Sauer, U. (2007). Determination of metabolic flux ratios from 13C-experiments and gas chromatography-mass spectrometry data: protocol and principles. *Methods Mol. Biol.* 358, 177–197. doi: 10.1007/978-1-59745-244-1_11
- Nölling, J., Breton, G., Omelchenko, M. V., Makarova, K. S., Zeng, Q., Gibson, R., et al. (2001). Genome sequence and comparative analysis of the solvent-producing bacterium *Clostridium acetobutylicum*. *J. Bacteriol.* 183, 4823–4838. doi: 10.1128/JB.183.16.4823-4838.2001
- Olson, D. G., and Lynd, L. R. (2012). Transformation of *Clostridium thermocellum* by electroporation. *Methods Enzymol.* 510, 317–330. doi: 10.1016/B978-0-12-415931-0.00017-3
- Quek, L. E., Wittmann, C., Nielsen, L. K., and Krömer, J. O. (2009). OpenFLUX: efficient modelling software for 13C-based metabolic flux analysis. *Microb. Cell Fact.* 8, 25. doi: 10.1186/1475-2859-8-25
- Sauer, U. (2006). Metabolic networks in motion: 13C-based flux analysis. *Mol. Syst. Biol.* 2, 62. doi: 10.1038/msb4100109
- Schmidt, K., Carlsen, M., Nielsen, J., and Villadsen, J. (1997). Modeling isotopomer distributions in biochemical networks using isotopomer mapping matrices. *Biotechnol. Bioeng.* 55, 831–840.
- Servinsky, M. D., Kiel, J. T., Dupuy, N. F., and Sund, C. J. (2010). Transcriptional analysis of differential carbohydrate utilization by *Clostridium acetobutylicum*. *Microbiology* 156, 3478–3491. doi: 10.1099/mic.0.037085-0
- Shupletsov, M. S., Golubeva, L. I., Rubina, S. S., Podvyaznikov, D. A., Iwatani, S., and Mashko, S. V. (2014). OpenFLUX2: (13)C-MFA

- modeling software package adjusted for the comprehensive analysis of single and parallel labeling experiments. *Microb. Cell. Fact.* 13, 152. doi: 10.1186/PREACCEPT-1256381938128538
- Sokol, S., Millard, P., and Portais, J. C. (2012). influx_s: increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments. *Bioinformatics* 28, 687–693. doi: 10.1093/bioinformatics/btr716
- Stephanopoulos, G. (1999). Metabolic fluxes and metabolic engineering. *Metab. Eng.* 1, 1–11.
- Van Winden, W. A., Heijnen, J. J., and Verheijen, P. J. (2002). Cumulative bondomers: a new concept in flux analysis from 2D [¹³C,1H] COSY NMR data. *Biotechnol. Bioeng.* 80, 731–745. doi: 10.1002/bit.10429
- Weitzel, M., Nöh, K., Dalman, T., Niedenführ, S., Stute, B., and Wiechert, W. (2013). 13CFLUX2—high-performance software suite for (13)C-metabolic flux analysis. *Bioinformatics* 29, 143–145. doi: 10.1093/bioinformatics/bts646
- Wiechert, W., Möllney, M., Isermann, N., Wurzel, M., and De Graaf, A. A. (1999). Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnol. Bioeng.* 66, 69–85.
- Xiong, W., Lo, J., Chou, K. J., Wu, C., Magnusson, L., Dong, T., et al. (2018). Isotope-assisted metabolite analysis sheds light on central carbon metabolism of a model cellulolytic bacterium *Clostridium thermocellum*. *Front. Microbiol.* 9:1947. doi: 10.3389/fmicb.2018.01947
- Young, J. D. (2014). INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* 30, 1333–1335. doi: 10.1093/bioinformatics/btu015
- Young, J. D., Walther, J. L., Antoniewicz, M. R., Yoo, H., and Stephanopoulos, G. (2008). An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnol. Bioeng.* 99, 686–699. doi: 10.1002/bit.21632
- Conflict of Interest Statement:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2019 Wu, Chen, Lo, Michener, Maness and Xiong. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Dietary Energy Level Promotes Rumen Microbial Protein Synthesis by Improving the Energy Productivity of the Ruminal Microbiome

Zhongyan Lu^{1†}, Zhihui Xu^{2,3†}, Zanning Shen¹, Yuanchun Tian⁴ and Hong Shen^{2,3*}

¹ The Key Laboratory of Animal Physiology and Biochemistry, College of Veterinary Medicine, Nanjing Agricultural University, Nanjing, China, ² College of Life Science, Nanjing Agricultural University, Nanjing, China, ³ Bioinformatics Center, Nanjing Agricultural University, Nanjing, China, ⁴ College of Agriculture, Nanjing Agricultural University, Nanjing, China

OPEN ACCESS

Edited by:

Garret Suen,
University of Wisconsin-Madison,
United States

Reviewed by:

Shengguo Zhao,
Institute of Animal Sciences (CAAS),
China
Luciano Takeshi Kishi,
São Paulo State University, Brazil
Hilario C. Mantovani,
Universidade Federal de Viçosa, Brazil

*Correspondence:

Hong Shen
hongshen@njau.edu.cn

[†]These authors have contributed
equally to this work

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 29 November 2018

Accepted: 02 April 2019

Published: 17 April 2019

Citation:

Lu Z, Xu Z, Shen Z, Tian Y and
Shen H (2019) Dietary Energy Level
Promotes Rumen Microbial Protein
Synthesis by Improving the Energy
Productivity of the Ruminal
Microbiome. *Front. Microbiol.* 10:847.
doi: 10.3389/fmicb.2019.00847

Improving the yield of rumen microbial protein (MCP) has significant importance in the promotion of animal performance and the reduction of protein feed waste. The amount of energy supplied to rumen microorganisms is an important factor affecting the amount of protein nitrogen incorporated into rumen MCP. Substrate-level phosphorylation (SLP) and electron transport phosphorylation (ETP) are two major mechanisms of energy generation within microbial cells. However, the way that energy and protein levels in the diet impact the energy productivity of the ruminal microbiome and, thereafter, rumen MCP yields is not known yet. In present study, we have investigated, by animal experiments and metagenome shotgun sequencing, the effects of energy-rich and protein-rich diets on rumen MCP yields, as well as SLP-coupled and ETP-coupled energy productivity of the ruminal microbiome. We have found that an energy-rich diet induces a significant increase in rumen MCP yield, whereas a protein-rich diet has no significant impacts on it. Based on 10 reconstructed pathways related to the energy metabolism of the ruminal microbiome, we have determined that the energy-rich diet induces significant increases in the total abundance of SLP enzymes coupled to the nicotinamide adenine dinucleotide (NADH) oxidation in the glucose fermentation and F-type ATPase of the electron transporter chain, whereas the protein-rich diet has no significant impact in the abundance of these enzymes. At the species level, the energy-rich diet induces significant increases in the total abundance of 15 ETP-related genera and 40 genera that have SLP-coupled fermentation pathways, whereas the protein-rich diet has no significant impact on the total abundance of these genera. Our results suggest that an increase in dietary energy levels promotes rumen energy productivity and MCP yield by improving levels of ETP and SLP coupled to glucose fermentation in the ruminal microbiome. But, an increase in dietary protein level has no such effects.

Keywords: rumen microbiome, energy productivity, substrate-level phosphorylation, electron transport phosphorylation, microbial protein synthesis, dietary modulation

INTRODUCTION

Dietary protein for ruminants includes nitrogen (N) occurring in true protein and non-protein. In the rumen, the true protein is degraded into amino acid (AA) and ammonia and then utilized by ruminal microorganisms to synthesize microbial protein (MCP). In the small intestine, more than 80% of rumen MCP is digested, accounting for 50–80% of the total absorbable protein contained there (Tas et al., 1981; Storm et al., 1983). In dairy cows, rumen MCP provides all of the AAs needed for milk protein synthesis (Virtanen, 1966). Because of the high digestibility and good AA composition of MCP, increasing the MCP yield in the rumen is of important significance for the promotion of animal performance. Moreover, increasing the MCP yield is the most effective strategy to reduce the protein feed waste in livestock, since the dietary protein that exceeds the requirement of ruminal microorganisms is degraded to ammonia in the rumen, metabolized to urea in the liver, and lost in the urine.

Previous studies have showed that the sources and amounts of fed carbohydrate are the major factors affecting the energy, in the form of adenosine triphosphate (ATP), available for rumen microbial growth (synthesis of MCP in particular), and that of fed protein affects the production of microbial dry matter (DM) per unit of carbohydrate fermented (Maeng et al., 1976; Hoover and Stokes, 1991). In the dairy cow, 12–13% protein content is needed to maximize the ruminal synthesis of MCP (Satter and Slyter, 1974; Satter and Roffler, 1975). More protein N is incorporated into rumen MCP only if more non-fiber carbohydrate (NFC), known to be the major energy substrate for ruminal microorganisms, are fed to the animals (Schwab et al., 2005). Microbiological studies have revealed that, once the extracellular AA is transported inside microbial cells, the fate of the absorbed AA will depend on the availability of ATPs within the microbial cells. If ATPs are available, AA will be transaminated or used directly for MCP synthesis. However, if ATPs are limited, AA will be deaminated into ammonia and excreted from the cytoplasm (Tamminga, 1979). Accordingly, when dietary protein meets the requirement of ruminal microorganisms, increasing the content of dietary NFC promotes rumen MCP yields via improving the amount of energy supplied to the ruminal microorganisms, whereas increasing the content of dietary protein may have no benefit on rumen MCP yields, since a protein-rich diet might not meet the energy requirement of ruminal microorganisms. So far, much attention has been paid to the ratio of dietary NFC/protein that is optimum for MCP yields (Casper and Schingoethe, 1989; Cameron et al., 1991; Henning et al., 1991). However, the mechanism that the ruminal microbiome applies to producing the energy, and the way that the NFC-rich diet or protein-rich diet impacts the energy productivity of the ruminal microbiota, compared with a basal diet, have not been studied as yet.

It is known that ATP is produced in the microbial metabolism via two major mechanisms: (1) substrate-level phosphorylation (SLP), and (2) electron transport phosphorylation (ETP). In the former, the high-energy phosphate in the substrate molecule is directly donated to ADP to produce ATP. In the latter, electrons that are obtained from energy sources are moved

by the electron transport chain (ETC) to reduce oxygen (aerobic respiration) or other oxidized components (anaerobic respiration). Meanwhile, the transmembrane electrochemical potential ($\Delta\mu_{H^+}/\Delta\mu_{Na^+}$) generated by the movement of electrons drives the synthesis of ATP in the cell. In the rumen, SLP majorly occurs in the glycolysis and short-chain fatty acids (SCFAs) production. During these processes, the electron generated in the degradation of glucose to pyruvate is transferred from nicotinamide adenine dinucleotide (NADH) to produce lactate, ethanol, butanol and malate (this is referred to as fermentation). ETP (anaerobic respiration) is known to be coupled only with the generation of propionate following succinate pathway and methanogenesis (Tamminga, 1979). Based on the comparison of incubation results using carbohydrate (cellobiose and maltose) and pyruvate as the substrates, ETP was shown to take a minor part in ATP yields, accounting for 20% of the total ATP yields in the carbohydrate fermentation (Demeyer and Van Nevel, 1986). However, recent studies have detected several new ETP-coupled pathways under anaerobic conditions, for example, acetogenesis via the reductive acetyl-CoA pathway (Wood-Ljungdahl pathway) (Schuchmann and Muller, 2014), and butyrate formation via the reduction of crotonyl-CoA (Herrmann et al., 2008). Accordingly, the importance of ETP to the ATP productivity in the ruminal microbiome, as well as the effects of fed NFC/protein on the ETP- and SLP-coupled ATPs productivity, needs to be re-evaluated.

Here, we present the first *in vivo* investigation of the microbial species, pathways and enzymes related to energy metabolism in the rumen, plus a systematic comparison of the energy productivity of the ruminal microbiome in goats receiving a protein-rich or a NFC-rich diet with that of goats receiving a basal diet. We first collected the ruminal microbiota from the goats receiving the three treatments, and compared the rumen MCP yields and fermentation parameters between the groups. Subsequently, by applying metagenome shotgun sequencing and bioinformatics analysis, we reconstructed the pathways related to SLP and the pathways related to ETP in the rumen. Next, we examined the enzymes that catalyze the SLP in these pathways and the enzymes that belong to the ETC and picked out the species whose genome encoded those specific enzyme. Finally, we compared the abundance of the enzymes and the relative abundance of the target species between the groups. By means of the above steps, we aimed to determine the way that the dietary protein or NFC impacted the energy productivity of the ruminal microbiome and, concomitantly to reevaluate the importance of ETP in the ATP productivities of the ruminal microbiome.

MATERIALS AND METHODS

Ethics Statement

This study was approved by the Animal Care and Use Committee of Nanjing Agricultural University, in compliance with the Regulations for the Administration of Affairs Concerning Experimental Animals (No. 588 Document of the State Council of China, 2011).

Animals

Twenty-four goats (Boer × Yangtze River Delta White, aged 4 months) were randomly allocated into three groups and received a NFC-rich diet (G group, $n = 8$), a protein-rich diet (P group, $n = 8$), or a basal diet (B group, $n = 8$). The ingredients and chemical compositions of the experimental diets are listed in **Table 1**.

Goats were housed in individual tie-stall barns and had free access to water. To avoid the selection of dietary components and to maintain the desired ratio, a total mixed ratio (TMR) was offered at 0800 and 1700 daily for the 42-day experimental period with the first 14 days being dedicated to adaptation. Feed intake and refusal of individual goats were measured daily during the experiment. The amount of diet offered during days 15–42 was adjusted on a weekly basis to allow at about 10% ort. Feeds were sampled at the beginning and end of the experiment. The DM, ash, crude fat, and crude protein contents of samples were analyzed according to the procedures of AOAC (Cunniff and AOAC International, 1997). The acid detergent fiber (ADF) and neutral detergent fiber (NDF) values of the samples were analyzed according to the procedures of Van Soest et al. (1991).

Sample Collection, Microbial DNA Extraction, and Metagenome Shotgun Sequencing

Urine samples were collected from 1200 on day 41 to 1200 on day 42 by indwelling catheters that were connected to containers containing 50% HCl, and then, stored at -20°C for the later analysis of rumen MCP. All goats were killed at the local slaughterhouse at 6 h after receiving their morning feed on day 43. Ruminal contents were strained through a 4-layer cheesecloth and immediately subjected to pH measurement. An aliquot (10 mL) of ruminal fluid was immediately stored at -20°C for

the extraction of metagenomic DNA. Another aliquot (10 mL) of ruminal fluid was added by 1 mL of 5% HgCl_2 solution to inactivate the microbial proteases, and subsequently, stored at -20°C for the determination of the volatile fatty acids (VFAs) concentration. An aliquot (5 mL) of ruminal fluid was stored at -20°C for the determination of ammonia N.

The metagenomic DNA was extracted from the ruminal fluid by using a Bacterial DNA Kit, following the instruction of the supplier (Omega, Shanghai, China). The DNA concentration was determined in a Nanodrop 1000 (Thermo Fisher Scientific, Wilmington, DE, United States). The integrity of the microbial DNA was evaluated on a 1.0% agarose gel. Metagenomic DNA libraries were constructed by using the TruSeq DNA Sample Prep kit (Illumina, San Diego, CA, United States). Libraries were sequenced via paired-end chemistry (PE150) on an Illumina HiSeq X Ten platform (Illumina, San Diego, CA, United States) at Biomarker Technologies, Beijing, China.

Rumen MCP and Fermentation Parameters Determination and Analysis

The amount of rumen MCP was calculated from purine derivative excretion in the urine by using the method described by Chen and Gomes (1995) on a spectrophotometer (721, INESA analytical instrument Co., LTD, Shanghai, China). The concentration of ruminal VFAs was determined by using a gas chromatograph (HP6890N, Agilent Technologies, Wilmington, DE, United States) as described by Yang et al. (2012). The ammonia N was determined by using a colorimetric method (Weatherburn, 1967) on a spectrophotometer.

The two-tailed t -test was used in the analysis of the rumen MCP amount and rumen fermentation parameters (VFAs concentration, ruminal pH and ammonia N concentration). Differences were considered significant when $p < 0.05$. These analyses were performed by using SPSS software package (SPSS Inc., Chicago, IL, United States).

Metagenome Shotgun Sequencing Analysis

Genome Assembly

Raw reads were first filtered by using FastX v0.0.13 (Gordon and Hannon, unpublished), with a quality cutoff of 20. Reads shorter than 30 bp were discarded from the sample. The reads that were likely to originate from the host and feeds were removed by using DeconSeq v0.4.3 (Schmieder and Edwards, 2011), with the NCBI goat, corn, soybean, and grass genome sequences as references. The remaining high-quality reads of all samples were taken together and then assembled into scaffolds by using IDBA-UD 1.1.1 (Peng et al., 2012) with the standard parameters.

Non-redundant Gene Set Construction

Genes were predicted from the scaffolds by using FragGeneScan 1.31 (Rho et al., 2010). Predicted genes from all samples were gathered together to form a large gene set. BLAT v35 (Kent, 2002) was used to construct the non-redundant gene set. Any two genes with more than 95% identity and more than 90% coverage of the

TABLE 1 | Ingredient and chemical composition of the experimental diets in the present study.

Item	B ¹	G ¹	P ¹
Ingredient, % of DM			
Guinea grass ²	90.0	65.0	75.0
Corn flour	0.0	25.0	0.0
Soybean meal	8.0	8.0	23.0
Additive ³	2.0	2.0	2.0
Chemical composition			
DM, %	90.6	89.5	89.7
Crude protein, %DM	9.6	10.0	15.6
Ether extract, %DM	3.0	3.0	2.8
Crude ash, %DM	6.0	5.2	6.1
NFC, %DM ⁴	14.1	28.3	16.3
NDF, %DM	67.3	53.6	59.3
ADF, %DM	40.1	29.7	6.0

DM, dry matter; NFC, non-fiber carbohydrate; NDF, neutral detergent fiber; ADF, acid detergent fiber. ¹The values are means \pm SE. ²*Aneurolepidimu chinense*.

³The additive was composed of calcium phosphate, limestone, trace mineral salt, and vitamin premix (vitamins A, D, and E). ⁴ $\text{NFC} = 100 - (\text{NDF} + \text{CP} + \text{ether extract} + \text{ash})$.

shorter gene were picked out, and subsequently, the shorter one was removed from the large gene set.

Gene Abundance Calculation

High-quality reads of each sample were mapped to the non-redundant gene set by using Bowtie2 v2.3.4 (Langmead and Salzberg, 2012) with default parameters. MarkDuplicates in the Picard toolkits version 2.0.1 was used to remove the PCR duplicates in the reads, and then, genomeCoverageBed in BEDTools 2.27.0 was employed to calculate the gene coverage. The reads per kilobase of exon model per million mapped reads (RPKM) of the gene, calculated by $[\text{gene coverage} \times 10^6 / (\text{total mapped reads} \times \text{gene length})]$, was used to normalize the gene abundance between the treatments.

Gene Function Annotation and Pathway Construction

The predicted genes were annotated to the Kyoto Encyclopedia of Genes and Genomes (KEGG) orthology databases via the KEGG Automatic Annotation Server (KAAS) (Moriya et al., 2007), and subsequently mapped to KEGG pathways (**Supplementary Table 1**). The pathways related to SLP and ETP (for details, see Results) were constructed by using Photoshop CS 8.01 on the background of KEGG pathways. Finally, the enzymes related to SLP, ETC components (details see below), and the enzymes catalyzing the oxidation of NADH during glucose fermentation were picked out according to KEGG Orthology (KO) annotation.

Taxonomic Annotation and Relative Abundance Calculation

The high-quality reads of each sample were mapped back to the scaffolds by using Bowtie2. PCR duplicates were removed by means of MarkDuplicates. The coverage of scaffolds in each sample was computed by using genomeCoverageBed and normalized to the relative abundance of 1 000 000. The composition-based rank-flexible classifier Epsilon-NB in FCP v1.0.7 (Parks et al., 2011) was used to assign the taxonomy of the scaffolds, with reference to the in-house reference catalog of 12 607 bacterial and 21 537 archaeal complete genomes downloaded from the NCBI RefSeq database¹. The relative abundance and taxonomic annotation of each genus are listed in **Supplementary Table 2**.

Metagenome Binning

Metagenome bins were recovered from the scaffolds by using CONCOCT package version 0.4.1 (Alneberg et al., 2014) with 300 as the number of the cluster. The coverage of the received bins was calculated by using ClusterMeanCov.pl in the CONCOCT package. CheckM (Parks et al., 2015) was used to estimate the completeness, contamination, strain heterogeneity and genome size of the received metagenome bins. Unfortunately, only 3/208 bins showed more than 90% completeness and less than 10% contamination, and 22/208 bins showed more than 70% completeness and less than 20% contamination in this study (**Supplementary Table 3**). Accordingly, scaffolds were used to represent the species in the present study.

¹ <ftp://ftp.ncbi.nlm.nih.gov/genomes/refseq/>

Gene and Species Abundance Comparison

The abundance of the enzymes and species were both compared by using the Wilcoxon test in R, respectively. Differences were considered significant when $p < 0.05$ and $|\log_2(G/B)|$ or $|\log_2(P/B)| > 1$.

Detection of the Major Contributors to SLP Enzymes and the Specific Fermenters in the Ruminal Microbiome

Based on the gene locations and the taxonomic annotation, the species whose genome encoded the SLP enzymes or the enzymes catalyzing the oxidation of NADH during glucose fermentation were picked out and then summarized at the genus level. In the present study, any genus whose relative abundance was more than 10 in at least one sample was defined as a major contributor to the specific SLP enzyme. The major contributor to all of the NADH oxidases of the specific fermentation pathway, as well as the specific SLP enzyme coupled to the specific fermentation pathway, was considered to be the fermenter to the corresponding product (details see **Figure 1**).

Detection of the Major Contributors to the ETC Components in the Ruminal Microbiome

The ETC is composed of a series of transmembrane ion pumps and F-type ATPase. By checking available publications, we have found the following enzymes that have been reported as ETC components under anaerobic conditions: FADH₂-NAD oxidoreductase Rnf; membrane-bound hydrogenases Ech and Mbh; methyltransferase Mtr; NADH dehydrogenases Nuo, Ndh, and Hox; cytochrome c reductase Pet and TorC; cytochrome bd oxidase Cyd; sulfate reductase Apr; nitrite reductase NrfA; nitrate reductases Nar and NapA; fumarate reductase Sdh; and cytochrome o oxidase Cyo (Bongaerts et al., 1995; Meurer et al., 1999; Leahy et al., 2010; Fu et al., 2014; De Rosa et al., 2015; Kracke et al., 2015). Except for Ech, Mbh, and Cyo, we detected 13 kinds of enzymes and all subunits of F-type ATPase in the sequences (**Table 2**). Based on the gene locations and the taxonomic annotation, the species whose genome encoded the ETC components was picked out. Since most enzymes consist of several subunits, only those species whose genome encoded more than half of the subunits of the specific enzyme were used in the subsequent analysis. Finally, the species whose genome encoded the ETC components was summarized at the genus level. The genus whose relative abundance was more than 10 in at least one sample was defined as the major contributor to the specific ETC component in the microbiome.

Result Visualization

The comparisons of the relative abundances of the species were visualized by using R program ggplot2.

Data Submission

The metagenome sequences are submitted to the NCBI under BioProject PRJNA492173.

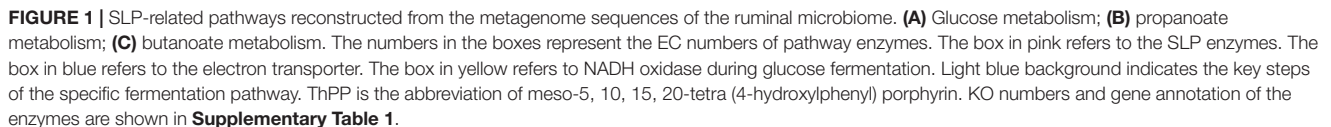


TABLE 2 | Comparisons of the abundances of SLP enzymes, ETC components, and NADH oxidases during glucose fermentation and of the relative abundance of corresponding species between groups.

PATHWAY	ENZYME	GENE ABUNDANCE				SPECIES RELATIVE ABUNDANCE					
		G ¹	P ¹	B ¹	G/B ²	P/B ²	G ¹	P ¹	B ¹	G/B ²	P/B ²
SLP enzymes	EC 2.7.2.3; pgk; phosphoglycerate kinase	1490.58 ± 76.36	1097.76 ± 26.08	1158.74 ± 67.61	0.36	-0.08	1308.5 ± 112.3	1276.4 ± 93.1	1823.2 ± 115.4	-0.48	-0.51
	EC 2.7.1.40; pyk; pyruvate kinase	589.59 ± 25.46	480.02 ± 39.72	436.98 ± 43.85	0.43	0.14	846.5 ± 57.9	751.5 ± 35.9	718.2 ± 37.2	0.24	0.07
Pyruvate/Propanoate metabolism	EC 2.7.2.1; ackA; acetate kinase	910.36 ± 120.20	625.56 ± 57.83	749.60 ± 63.99	0.28	-0.26	2163.9 ± 185.2	1326.4 ± 111.9	1669.2 ± 137.2	0.37	-0.33
	EC 2.7.2.7; buk; butyrate kinase	980.68 ± 41.05	410.10 ± 64.37	264.02 ± 130.54	1.89*	0.64	1320.1 ± 99.3	799.1 ± 74.4	518.3 ± 46.7	1.35*	0.62
Nitrogen metabolism	EC 2.7.2.2; arcC; carbamate kinase	202.21 ± 10.26	174.72 ± 46.94	152.81 ± 28.83	0.40	0.19	479.8 ± 24.5	448.3 ± 32.0	431.3 ± 32.1	0.15	0.06
	EC 1.2.1.10; adhE; acetaldehyde dehydrogenase	663.37 ± 20.90	208.92 ± 20.27	227.83 ± 10.88	1.54*	-0.13	1.7 ± 0.5	0.3 ± 0.2	1.1 ± 0.1	0.68	-1.79*
NADH oxidases in glucose fermentation	EC 1.1.1.1; adhP; alcohol dehydrogenase	14.74 ± 2.87	4.75 ± 1.14	4.29 ± 1.33	1.78*	0.15	436.7 ± 22.7	460.2 ± 34.1	632.7 ± 68.3	-0.53	-0.46
	EC 1.1.1.1; adh; alcohol dehydrogenase	8.15 ± 1.77	5.66 ± 1.65	3.87 ± 1.31	1.07*	0.55	32.8 ± 4.5	28.2 ± 3.1	9.9 ± 1.1	1.73*	1.51*
	EC 1.1.1.1; adhC; alcohol dehydrogenase	16.55 ± 6.07	2.01 ± 1.40	6.49 ± 1.48	1.35*	-1.69*	8.4 ± 1.1	4.5 ± 0.6	8.4 ± 0.4	-0.01	-0.89
	EC 1.1.1.1; yiaY; alcohol dehydrogenase	121.40 ± 10.44	32.83 ± 5.67	38.95 ± 7.32	1.64*	-0.25	16.8 ± 1.2	7.0 ± 1.1	11.0 ± 0.5	0.61	-0.67
	EC 1.1.1.2; adh; alcohol dehydrogenase (NADP+)	79.88 ± 11.90	81.79 ± 17.51	78.46 ± 18.96	0.03	0.06	53.6 ± 4.2	94.3 ± 9.7	79.0 ± 4.1	-0.56	0.26
	EutG; alcohol dehydrogenase	40.05 ± 5.12	13.59 ± 3.35	12.94 ± 3.21	1.63*	0.07	116.1 ± 6.8	164.8 ± 9.5	217.6 ± 5.8	-0.91	-0.40
Homolactic fermentation	EC 1.1.1.27; ldh; L-lactate dehydrogenase	118.76 ± 33.54	125.58 ± 27.77	86.46 ± 15.35	0.46	0.54	125.6 ± 5.2	128.1 ± 5.9	140.9 ± 7.7	-0.17	-0.14
	EC 1.3.1.44; ter; trans-2-enoyl-CoA reductase (NAD+)	0.62 ± 0.01	0.62 ± 0.01	0.97 ± 0.00	-0.64	-0.64	725.3 ± 54.7	962.8 ± 68.5	703.4 ± 55.6	0.04	0.45
Butanol fermentation	EC 1.1.1.-; bdhAB; butanol dehydrogenase	833.56 ± 48.40	414.17 ± 20.41	451.46 ± 20.64	0.88	-0.12	436.7 ± 22.7	460.2 ± 34.1	632.7 ± 68.3	-0.53	-0.46
	EC 1.2.1.10; adhE; acetaldehyde dehydrogenase	663.37 ± 20.90	208.92 ± 20.27	227.83 ± 10.88	1.54*	-0.13	1.7 ± 0.5	0.3 ± 0.2	1.1 ± 0.1	0.68	-1.79
Propionate fermentation	EC 1.1.1.37; mdh; malate dehydrogenase	612.69 ± 54.98	415.07 ± 96.21	383.46 ± 55.67	0.68	0.11	62.9 ± 3.8	33.6 ± 3.6	34.5 ± 3.2	0.86	-0.04

(Continued)

TABLE 2 | Continued

PATHWAY	ENZYME	GENE ABUNDANCE				SPECIES RELATIVE ABUNDANCE					
		G ¹	P ¹	B ¹	G/B ²	P/B ²	G ¹	P ¹	B ¹	G/B ²	P/B ²
ETC components	ATPase	768.75 ± 76.10	382.62 ± 57.24	353.58 ± 41.08	1.12*	0.11	1926.0 ± 197.8	1629.1 ± 187.1	1601.6 ± 192.2	0.27	0.02
	FADH ₂ -NAD oxidoreductase ³	910.04 ± 66.10	417.25 ± 67.79	443.12 ± 51.49	1.04*	-0.09	2295.6 ± 191.3	1574.6 ± 149.0	1424.0 ± 120.8	0.69	0.15
	Methanogenesis Mtr; tetrahydromethanopterin S-methyltransferase	2.20 ± 0.74	1.96 ± 0.29	2.09 ± 0.46	0.07	-0.09	7.0 ± 0.8	4.3 ± 0.5	6.8 ± 1.3	0.04	-0.65
	NADH dehydrogenase ³	230.44 ± 22.47	120.09 ± 20.18	113.48 ± 11.59	1.02*	0.08	2253.7 ± 176.9	1378.7 ± 155.5	1014.8 ± 73.2	1.15*	0.44
	Ndh; NAD(P)H-quinone oxidoreductase	13.78 ± 3.49	12.97 ± 3.49	10.58 ± 2.85	0.38	0.29	20.0 ± 2.1	8.2 ± 1.2	12.2 ± 0.6	0.71	-0.57
	Hox; bidirectional [NiFe] hydrogenase diaphorase	13.74 ± 3.20	7.52 ± 1.14	10.89 ± 2.14	0.34	-0.53	5.2 ± 0.7	2.1 ± 0.3	5.8 ± 0.6	-0.16	-1.45*
	Cytochrome complex ³	0.71 ± 0.21	0.89 ± 0.14	0.63 ± 0.11	0.16	0.51	0	0	0	0	0
	TorC; trimethylamine-N-oxide reductase	0.47 ± 0.03	0.83 ± 0.04	0.65 ± 0.07	-0.45	0.36	5.4 ± 1.1	2.8 ± 0.6	3.8 ± 0.4	0.50	-0.47
	Cyd; cytochrome bd ubiquinol oxidase	412.15 ± 47.13	539.51 ± 33.78	366.26 ± 50.13	0.17	0.56	2035.6 ± 212.7	1044.0 ± 116.0	911.6 ± 68.2	1.16*	0.2
Sulfur dissimilation	Apr; adenylylsulfate reductase	0.64 ± 0.06	0.59 ± 0.05	0.78 ± 0.31	-0.28	-0.39	2.8 ± 0.4	19.9 ± 2.1	24.6 ± 2.0	-3.13*	-0.31
	Nitrogen dissimilation	403.53 ± 65.93	202.95 ± 29.52	176.38 ± 24.09	1.19*	0.20	541.3 ± 45.7	337.5 ± 40.5	259.7 ± 22.2	1.06*	0.38
	Nar; nitrate reductase / nitrite oxidoreductase	0.55 ± 0.09	1.60 ± 0.15	1.72 ± 0.22	-1.63*	-0.10	0.2 ± 0.1	8.5 ± 0.9	3.1 ± 0.2	-3.90*	1.46*
	NapA; periplasmic nitrate reductase NapA	0.82 ± 0.03	1.89 ± 0.12	0.83 ± 0.31	-0.02	1.20*	0.34 ± 0.13	4.62 ± 1.08	1.56 ± 0.28	-3.76*	1.57*
Fumarate respiration	Sdh; succinate dehydrogenase	1210.06 ± 130.02	578.79 ± 32.16	500.64 ± 64.78	1.27*	0.21	1422.9 ± 114.3	1179.1 ± 95.8	976.6 ± 79.6	0.54	0.27

¹ mean value ± standard error; the unit of gene results is RPKM; the relative abundance of species results was normalized to 1,000,000 for each sample. ² the numbers were calculated from log₂(G/B) or log₂(P/B). ³ the electron transporters were reported in many kinds of ETCs. *p < 0.05 and |value| > 1 in the wilcox test between groups.

TABLE 3 | Comparisons of rumen pH, ammonia N, MCP, and VFAs concentrations between groups.

Item	B ¹	G ¹	P ¹
pH	6.74 ± 0.04	6.38 ± 0.05 ^a	6.47 ± 0.03 ^b
Ammonia N, mM	3.04 ± 0.17	8.73 ± 0.73 ^a	12.51 ± 0.87 ^b
MCP, g/d	8.47 ± 0.86	11.49 ± 1.22 ^a	9.31 ± 1.07
Total VFAs, mM ²	53.01 ± 2.00	78.67 ± 2.93 ^a	67.66 ± 3.57 ^b
VFA, molar proportions			
Acetate	67.7 ± 2.2	58.8 ± 2.7 ^a	70.6 ± 2.9 ^b
Propionate	21.1 ± 1.4	29.3 ± 2.4 ^a	18.8 ± 2.6
Butyrate	11.2 ± 1.5	11.9 ± 1.4 ^a	10.6 ± 1.4

¹Values are mean ± standard error (SE). ²Total VFA = Acetate + Propionate + Butyrate. ^a $p < 0.05$ in the two-tailed T-test of B group and G group.

^b $p < 0.05$ in the two-tailed T-test of B group and P group.

enzymes were high in the gene pools of the ruminal microbiome (153–1491). The total abundance of SLP enzymes was 2762, 4173, and 2788 in B, G, and P group, respectively. According to the wilcoxon test, it showed the significant increase ($p < 0.05$) when the diet shifted from B to G and no significant change when the diet shifted from B to P. For the individual SLP enzymes, only the abundance of butyrate kinase (buk; EC 2.7.2.7), which catalyzes the reaction of butanoyl-P and ADP to form butanoate and ATP, was significantly increased when the diet shifted from B to G ($p < 0.05$).

Within the ruminal microbiome, 21 of the 68 detected archaeal genera and 258 of the 549 detected bacterial genera included the species whose genome encoded the SLP enzymes (in addition to unclassified genera). However, the changes in the relative abundance of these microorganisms did not correspond to the changes in the total abundance of the SLP enzymes (Table 2), which showed a slight increase when the diet shifted from B to G (0.52–0.61%), and a significant decrease when the diet shifted from B to P (0.52–0.46%, $p < 0.05$). Although at least 8 G of high-quality data was received for each sample, around 90% metagenome bins had less than 70% completeness in the threshold of less than 20% contamination in the present study, possibly because the diversity and abundance of the ruminal microbiome are greater than we are aware. The diet-induced changes in the relative abundance of the major contributors to the individual SLP enzymes are shown in Supplementary Figure 1.

Comparison of Abundance of Enzymes Catalyzing the Oxidation of NADH in the Glucose Fermentation and the Specific Fermenters Between the Groups

Under anaerobic conditions, NADH produced by glycolysis is oxidized back to NAD⁺ by using pyruvate or one of its derivatives as an electron acceptor. In the present study, we analyzed the enzymes that catalyzed the oxidation of NADH during glucose fermentation between the groups. As a result, 11 enzymes belonging to 4 types of fermentation (heterolactic fermentation, homolactic fermentation, butanol fermentation, and propionate fermentation) were detected (Table 4). Among them, the acetaldehyde dehydrogenase (adhE;

EC 1.2.1.10) was the common enzyme used in the heterolactic fermentation and butanol fermentation. In a comparison of their abundances, five alcohol dehydrogenases (EC 1.1.1.1 and EutG) and adhE were significantly increased when the diet shifted from B to G. On the other hand, the abundance of one alcohol dehydrogenase (EC 1.1.1.1) was significantly decreased when the diet shifted from B to P (Table 2).

Within the ruminal microbiome, other than the unclassified genera, we detected 10 butanol fermenters, 11 heterolactic fermenters, 2 homolactic fermenters, and 6 propionate fermenters in the G group. Among them, the relative abundance of 5 butanol fermenters, 7 heterolactic fermenters, two homolactic fermenters, three propionate fermenters, as well as the total abundance of all fermenters, was significantly increased, whereas the relative abundance of one butanol fermenter and one heterolactic fermenter was significantly decreased when the diet shifted from B to G (Table 4). On the other hand, other than the unclassified genera, 10 butanol fermenters, 8 heterolactic fermenters, one homolactic fermenter, and 6 propionate fermenters were detected in the P group. Among them, the relative abundance of four butanol fermenters, four heterolactic fermenters, one homolactic fermenter, and one propionate fermenters were significantly increased, whereas the relative abundance of two butanol fermenters and two propionate fermenters was significantly decreased when the diet shifted from B to P. The total abundance of all fermenters had no significant change when the diet shifted from B to P.

Comparison of Abundance of ETC Components and Their Major Contributors Between the Groups

As shown in Table 2, the mean abundances of subunits of Rnf, Nuo, NrfA, and Sdh were significantly increased, whereas the mean abundance of subunits of Nar was significantly decreased when the diet shifted from B to G. On the other hand, only the mean abundance of NapA subunits showed a significant increase, and the others exhibited no significant changes when the diet shifted from B to P. Being the final step of ETC, F-type ATPase is known to be the most widely used ATP synthase in both prokaryotes and eukaryotes. In the present study, all subunits of F0-F1 ATPase were detected in the groups. The mean abundance of its subunits was 354, 769, and 383 in the B, G, and P group, respectively. Furthermore, it was significantly increased when the diet shifted from B to G, whereas no significant changes were seen when the diet shifted from B to P.

Within the ruminal microbiome, other than the unclassified genera, 17 major contributors to Rnf, 13 major contributors to Cyd, 20 major contributors to Sdh, 7 major contributors to Nuo, 7 major contributors to NrfA, and 11 major contributors to ATPase were detected in the G group (Supplementary Table 4 and Figure 3). Among them, the abundance of 5/5/9/6/5/3 major contributors to Rnf/Cyd/Sdh/Nuo/NrfA/ATPase was significantly increased, whereas the abundance of 8/5/6/2 major contributors to Rnf/Cyd/Sdh/ATPase was significantly decreased when the diet shifted from B to G. On the other hand, other than the unclassified genera, 16 major

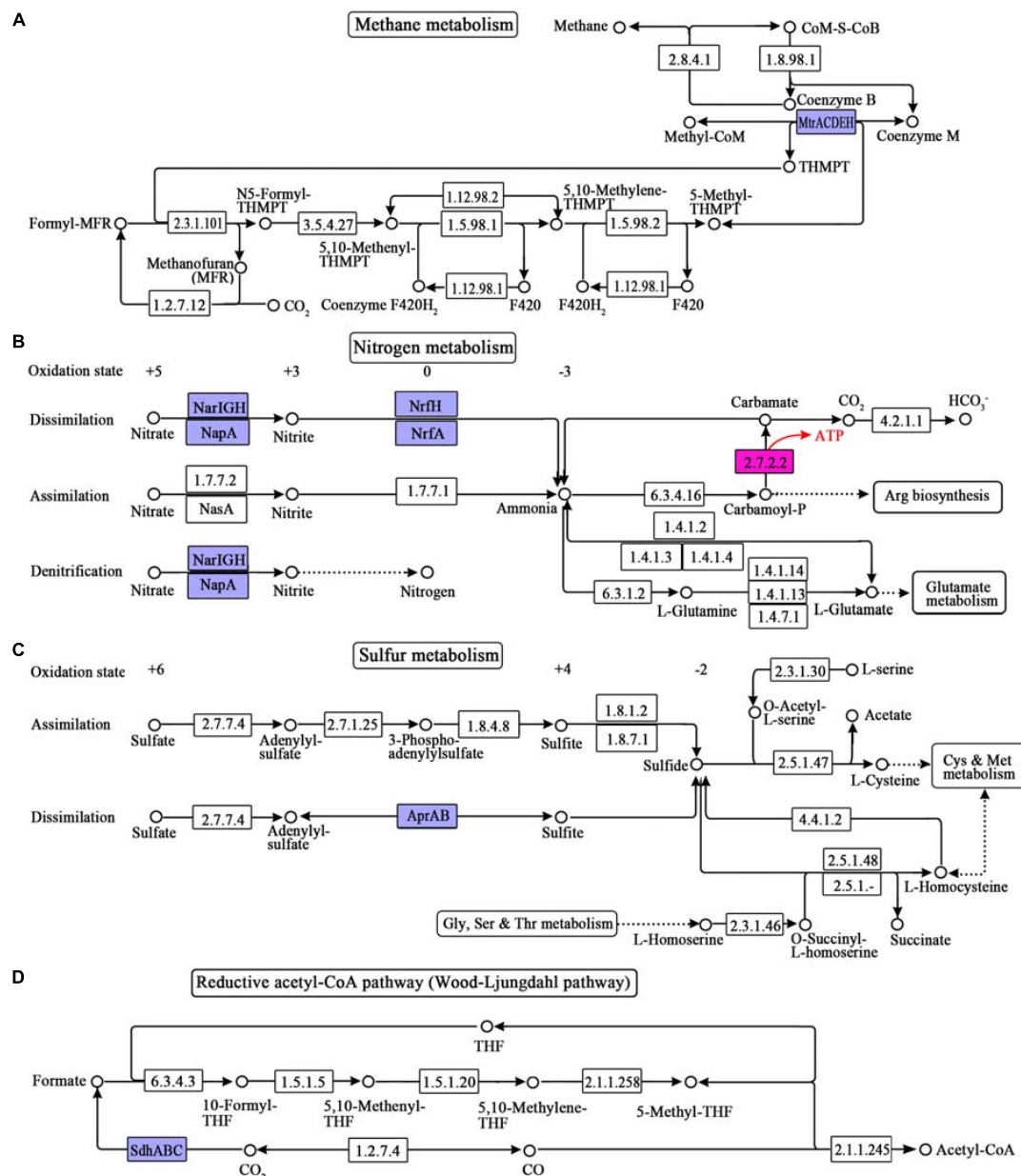


FIGURE 2 | ETP-related pathways reconstructed from the metagenome sequences of the ruminal microbiome. **(A)** Methane metabolism; **(B)** nitrogen metabolism; **(C)** sulfur metabolism; **(D)** reductive acetyl-CoA pathway (Wood-Ljungdahl pathway). The numbers in the boxes represent the EC numbers of pathway enzymes. The box in pink refers to the SLP enzymes. The box in blue refers to the electron transporter. THF is the abbreviation of tetrahydrofuran. THMPT is the abbreviation of tetrahydromethanopterin. KO numbers and gene annotation of the enzymes are shown in **Supplementary Table 1**.

contributors to Rnf, 10 major contributors to Cyd, 22 major contributors to Sdh, 5 major contributors to Nuo, 2 major contributors to NrfA, and 12 major contributors to ATPase were detected in the P group (**Supplementary Table 4** and **Figure 3**). Among them, the abundance of 1/4/12/1/5 major contributor(s) of Rnf/Cyd/Sdh/Nuo/ATPase was significantly increased, whereas the abundance of 3/2/2/2 major contributor(s) of Rnf/Cyd/Sdh/ATPase was significantly decreased when the diet shifted from B to P.

DISCUSSION

Effects of Dietary NFC and Protein on Rumen MCP Synthesis

In the present study, the significant increase of the rumen ammonia N was associated with the significant increase of MCP yields when the diet shifted from B to G. This is in accordance with the study of Kljak et al. (2017) and

TABLE 4 | Comparisons of the relative abundance of specific fermenters between groups.

Product	Fermenter	B ¹	G ¹	P ¹	G/B ²	P/B ²
Butanol	Prevotella	295.2 ± 27.1	859.3 ± 68.9	439.3 ± 44.8	1.54*	0.57
	Syntrophus	10.6 ± 1.1	52.7 ± 2.1	0	2.32*	
	Granulicella	33.3 ± 0.9	44.2 ± 5.4	0	0.41	
	Clostridium	51.5 ± 5.0	38.2 ± 2.3	27.3 ± 1.1	-0.43	-0.92
	Paenibacillus	26.2 ± 2.7	27.7 ± 3.4	26.2 ± 2.0	0.08	0.00
	Selenomonas	8.6 ± 1.0	19.9 ± 1.9	0	1.22*	
	Eggerthella	16.4 ± 1.8	17.8 ± 1.5	30.0 ± 2.2	0.12	0.87
	Slackia	1.7 ± 0.1	15.8 ± 2.5	0	3.23*	
	Ethanoligenens	0.9 ± 0.1	11.7 ± 1.1	0	3.74*	
	Geobacter	26.3 ± 2.6	11.0 ± 0.8	10.5 ± 0.9	-1.26*	-1.32*
	Bacteroides	12.0 ± 0.8	0	47.0 ± 3.8		1.97*
	Syntrophobacter	1.4 ± 0.2	0	15.2 ± 1.6		3.46*
	Bacillus	3.5 ± 0.4	0	14.0 ± 1.3		1.99*
	Staphylococcus	26.6 ± 3.3	0	10.7 ± 0.4		-1.32*
	Butyrivibrio	0.7 ± 0.1	0	10.2 ± 1.3		3.78*
Heterolactate	Clostridium	75.9 ± 6.7	82.8 ± 4.5	78.4 ± 5.5	0.13	0.05
	Paenibacillus	18.6 ± 1.2	60.9 ± 2.5	0	1.71*	
	Lactobacillus	1.9 ± 0.1	26.3 ± 1.4	0	3.81*	
	Oscillibacter	2.7 ± 0.2	22.8 ± 1.4	0	3.08*	
	Sinorhizobium	3.7 ± 0.3	20.0 ± 3.3	0	2.44*	
	Syntrophus	20.0 ± 2.1	18.3 ± 0.4	0	-0.13	
	Desulfovibrio	3.4 ± 0.6	17.1 ± 1.5	0	2.33*	
	Marinobacter	1.4 ± 0.1	14.9 ± 2.0	0	3.36*	
	Gluconobacter	2.1 ± 0.3	13.4 ± 1.2	0	2.69*	
	Prevotella	51.7 ± 3.1	11.3 ± 0.8	57.4 ± 6.7	-2.20*	0.15
	Syntrophobacter	10.8 ± 1.0	10.3 ± 0.5	0	-0.08	
	Bacillus	3.0 ± 0.1	0	44.9 ± 3.8		3.89*
	Geobacter	28.8 ± 3.1	0	24.2 ± 2.5		-0.25
	Treponema	3.3 ± 0.4	0	20.5 ± 2.6		2.63*
	Bacteroides	7.8 ± 1.1	0	14.0 ± 1.3		0.83
Homolactate	Sphaerochaeta	3.8 ± 0.3	0	13.2 ± 1.3		1.81*
	Ethanoligenens	5.1 ± 0.4	0	12.0 ± 0.8		1.23*
	Paenibacillus	7.2 ± 0.7	28.0 ± 1.4	0	1.95*	
	Ruminococcus	2.6 ± 0.6	22.2 ± 0.2	0	3.07*	
	Clostridium	8.8 ± 0.7	0	54.4 ± 2.7		2.63*
Propionate	Granulicella	30.1 ± 2.2	177.3 ± 14.7	12.5 ± 1.1	2.56*	-1.27*
	Prevotella	166.0 ± 13.7	119.8 ± 4.9	341.3 ± 31.5	-0.47	1.04*
	Desulfomicrobium	28.8 ± 1.9	89.8 ± 8.6	12.9 ± 1.4	1.64*	-1.16*
	Acidovorax	6.4 ± 0.6	52.4 ± 5.2	0	3.04*	
	Bacteroides	20.2 ± 1.8	39.6 ± 1.9	30.2 ± 2.9	0.97	0.58
	Geobacter	46.4 ± 3.0	25.2 ± 3.0	57.7 ± 5.5	-0.88	0.31
	Clostridium	15.1 ± 1.1	0	28.6 ± 2.7		0.92

¹ mean value ± standard error; the relative abundance was normalized to 1,000,000 for the sample. ² the numbers were calculated from $\log_2(G/B)$ or $\log_2(P/B)$. * $p < 0.05$ and $|value| > 1$ in the wilcox test between groups.

indicates that the G diet provides sufficient N and energy to the ruminal microbiome and therefore, has advantages for MCP synthesis compared with the B diet. On the other hand, rumen ammonia N was significantly increased, whereas ruminal MCP yields exhibited no significant change when the diet shifted from B to P. This result suggests that the P diet does not meet the energy requirement of ruminal microorganisms and therefore has no advantages for rumen

MCP synthesis compared with the B diet. Secondly, total VFAs was significantly increased in both of G and P group compared with the B group, indicating the increase of rumen fermentation in these groups. However, the fermentation substrates that induced the increase of rumen VFAs were different in these groups. G diet majorly induced the increase of glucose fermentation, whereas P diet majorly induced the increase of protein fermentation.

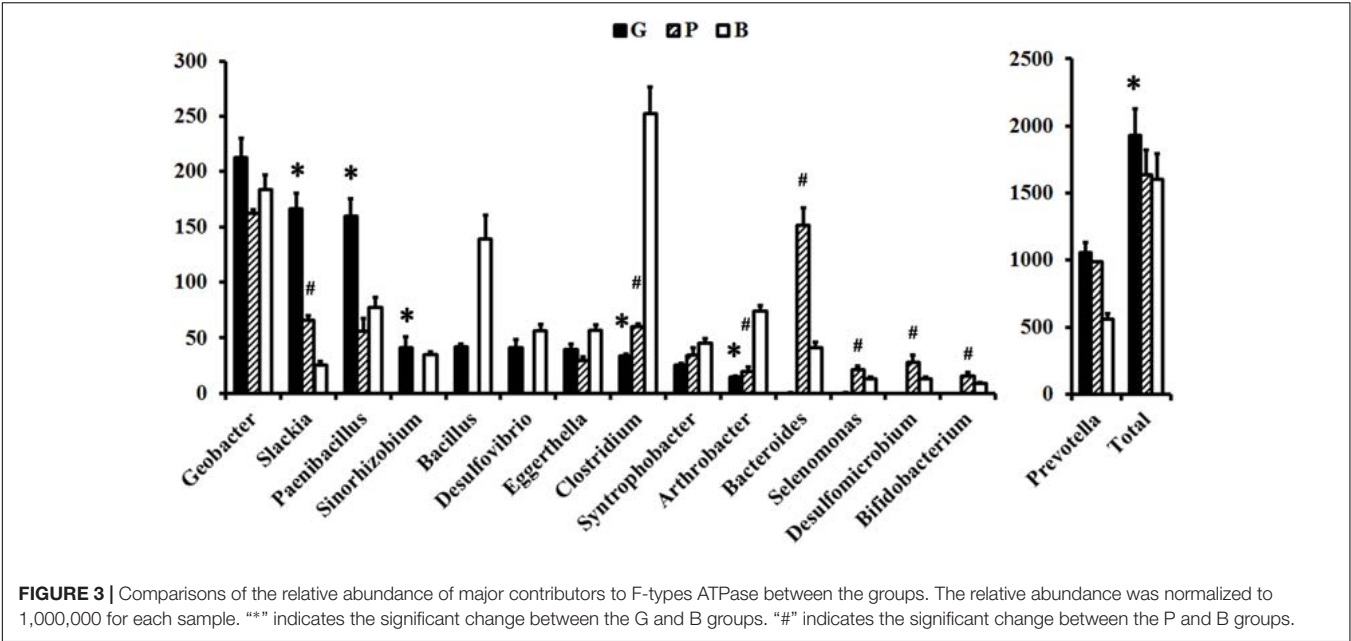


TABLE 5 | Detected types of electron transporters in ETP-related genera.

Major contributor	Rnf	Nuo	Ndh	NrfA	Mtr	Sdh	Cyd	ATPase
<i>Bacillus</i>	Rnf							ATPase
<i>Selenomonas</i>	Rnf					Sdh		ATPase
<i>Sinorhizobium</i>	Rnf					Sdh		ATPase
<i>Bacteroides</i>	Rnf	Nuo				Sdh	Cyd	ATPase
<i>Arthrobacter</i>	Rnf						Cyd	ATPase
<i>Geobacter</i>	Rnf	Nuo				Sdh	Cyd	ATPase
<i>Paenibacillus</i>	Rnf	Nuo		NrfA		Sdh	Cyd	ATPase
<i>Desulfomicrobium</i>						Sdh	Cyd	ATPase
<i>Desulfovibrio</i>	Rnf			NrfA		Sdh		ATPase
<i>Clostridium</i>	Rnf	Nuo	Ndh			Sdh		ATPase
<i>Eggerthella</i>	Rnf	Nuo				Sdh	Cyd	ATPase
<i>Syntrophobacter</i>				NrfA		Sdh	Cyd	ATPase
<i>Bifidobacterium</i>	Rnf					Sdh	Cyd	ATPase
<i>Slackia</i>	Rnf	Nuo				Sdh	Cyd	ATPase
<i>Prevotella</i>	Rnf	Nuo		NrfA		Sdh	Cyd	ATPase

Cyd, cytochrome bd oxidase; Mtr, methyltransferase; Ndh and Nuo, NADH dehydrogenases; NrfA, nitrite reductase; Rnf, FADH₂-NAD oxidoreductase; Sdh, fumarate reductase/succinate oxidase; ATPase, F-type ATPase.

Effects of Dietary NFC and Protein on the Levels of SLP-Coupled Energy Productivity Within the Ruminal Microbiome

Our metagenomics analysis showed that (1) the total abundance of SLP enzymes was significantly increased when the diet shifted from B to G, whereas it was no significant change when the diet shifted from B to P; (2) 21 of 68 archaeal genera and 258 of 549 bacterial genera were related to the SLP within the ruminal microbiome; and (3) the relative abundance of SLP-related species was no significant change when the diet shifted from B to G, whereas it showed the significant decrease when the diet shifted from B to P. By considering the low completeness of the

detected genomes, these results suggest that the NFC-rich diet provides the significant benefits with regard to the amount of SLP-related microorganisms/gene pool, whereas the protein-rich diet has no significant impact on the amount of SLP-related microorganisms/gene pool.

In the rumen, the conversion of glucose to two pyruvate molecules during glycolysis gives a net gain of two ATPs and two NADHs. Subsequently, the NADHs are oxidized to NAD⁺ by using pyruvate or one of its derivatives as an electron acceptor. This process, referred to as fermentation, may lead to the production of more ATPs, since the SLP is coupled with the oxidation of NADH in many kinds of fermentation. In the present study, four kinds of fermentation were detected within the rumen (Table 4). Among them, butanol fermentation,

heterolactate fermentation, and propionate fermentation give rise to the additional ATPs (Figure 1). According to our analysis, the total abundances of NADH oxidases located on above three fermentation pathways were significantly increased when the diet shifted from B to G (1437–3054) but were no significantly changes when the diet shifted from B to P (1437–1388). This result suggests that the level of SLP-coupled energy productivity is promoted by the NFC-rich diet but not affected by the protein-rich diet in a comparison with the basal diet. Besides, a recent study has shown that an increase of intracellular NAD^+ will promote the transfer of electrons along the ETC and thus promote the generation of ATPs in metal respiring microorganisms (Li et al., 2018). In present study, the total abundances of NADH oxidases in all fermentation pathways showed the same trends during the shift of diets, suggesting the amount of NAD^+ and its cycling rate is promoted by the dietary NFC but not affected by the dietary protein. Accordingly, we speculate that the energy productivity of ruminal microbiome might be promoted by dietary NFC via its effects on the levels of SLP coupled to the glucose fermentation and the amount of intracellular NAD^+ .

Effects of Dietary NFC and Protein on the Levels of ETP-Related Genes and Species Within the Ruminal Microbiome

According to our analysis, first, Ech, and cytochrome b0 were not present among the ruminal microorganisms. Second, the abundance of cytochrome c reductase (Pet and torC) and sulfate reductase Apr was less than 1 in all samples, and the species whose genomes encoded it were not detected in this study. We therefore speculated that the species whose genomes encoded cytochrome c reductase and Apr could not be present in the ruminal microbiome. Third, although the methyltransferase Met had low abundance (around 2 in all samples), it was only encoded on the genomes of archaeal methanogens that accounted for 1.2–1.4% of the ruminal microbiome. The abundances of NADH dehydrogenases Ndh and Hox were more than 10 in at least one group. However, the species whose genome encoded them were not detected. Since the completeness of bins was inadequate here, we therefore speculated that the species whose genomes encoded Met, Ndh and Hox existed in the rumen, but that their relative abundances were low. Fourth, the abundance of fumarate reductase Sdh, FADH_2 -NAD oxidoreductase Rnf, NADH dehydrogenase Nuo, nitrite reductase NrfA, and cytochrome bd complex cydA and cydB was high indicating their high abundance in the ruminal microbiome.

The microbial ETCs are diverse, and typically end up with a F-type ATPase. In present study, we evaluated the roles of ETP on the energy productivity of the ruminal microbiome by using the results from ATPase in present study, because the low completeness of received species made the detection of the integrity ETC in the specific species impossible. According to our analysis, the total mean abundance of ATPase subunits was significantly increased when the diet shifted from B to G, whereas no significant changes were seen when the diet shifted from B

to P. This result indicated the ETP-coupled energy productivity was promoted by the NFC-rich diet but not affected by the protein-rich diet in a comparison with the basal diet.

The comparison on the gene abundance of the SLP enzymes and the ATPase within the groups may imply that the ETP-based ATP productivity accounts for 13–18% of SLP-based ATP productivity in the rumen microbiome. At the species level, 15 kinds of bacteria were considered to have an ETP-type energy producing mechanism within the ruminal microbiome. The detected ETC components of 15 ETP-related species are listed in Table 5. According to previous studies, *Arthrobacter*, *Geobacter*, and *Paenibacillus* are metal-respiring bacteria (Bond and Lovley, 2003; Van Driessche et al., 2005). Our analysis showed that Cyd, NrfA, Nuo, Rnf, and Sdh were used as electron transporters in these bacteria. *Clostridium* and *Eggerthella* are CO_2 -respiring acetogenic bacteria (Harris et al., 2018). Our analysis showed that Nuo, Rnf, and Sdh were commonly used as the electron transporters in them. These results are supported by previous studies showing that cytochromes, NADH dehydrogenase, NrfA, and fumarate reductase play roles in the electron transport pathway of metal respiration. Rnf and NADH dehydrogenase have important functions in the electron transport pathway of reductive acetogenesis via the Wood–Ljungdahl pathway. Next, we showed Sdh to be the common electron transporter in the sulfate-respiring *Desulfovibrio* and *Desulfomicrobium* (Kushkevych, 2014) and in the fumarate-respiring *Syntrophobacter* (Plugge et al., 2012). Rnf was shown to be the common electron transporter in the nitrate-reducing bacteria *Bacillus* and *Selenomonas* (Zhao et al., 2015; Bruna et al., 2018), the nitrogen-fixing bacterium *Sinorhizobium* (Delamuta et al., 2015), the proteolytic bacterium *Bacteroides* (Macfarlane et al., 1995), lactate-fermented bacterium *Bifidobacterium* (Macfarlane et al., 1995), and the xenobiotic-degraded bacterium *Slackia* (Cho et al., 2016). This is similar to recently discovered ETCs, all of which use Rnf as their electron transporter. In the present study, most kinds of electron transporters were detected in *Prevotella*. The high abundance and the high species diversity of this genus in the rumen might be the reason for this result. Together, these results suggest that (1) various kinds of ETCs exist in the ruminal microorganism population, and (2) ETP-type energy-producing mechanism might also play important roles in the energy productivity of the ruminal microbiome.

CONCLUSION

In conclusion, our data showed that, on the gene level, the mean abundance of the subunits of ATPase accounted for less than 20% of the total abundances of SLP enzymes within the ruminal microbiome. On the genus level, 279 genera were showed to be able to generate ATPs via SLP mechanism, whereas only 15 genera were showed to be able to generate ATPs via ETP mechanism within 606 detected genera. Next, an increase in dietary energy level increased rumen MCP yields, whereas an increase in dietary protein level had no significant impact on it. On the gene level, an increase in dietary energy

level increased the total abundance of SLP enzymes that were coupled to the NADH oxidization in the butanol fermentation, heterolactate fermentation and propionate fermentation, and the mean abundance of F-type ATPase that catalyzed the generation of ATPs in the ETP within the ruminal microbiome, whereas an increase in dietary protein level had no significant impact on the abundances of these enzymes. On the genus level, an increase in dietary energy level increased the total abundance of 15 ETP-related genera, and the total abundance of 40 genera that own SLP-coupled fermentation pathway, whereas an increase in dietary protein level has no significant impact on the relative abundance of these genera. In summary, the comparison on the abundance of SLP enzymes and ATPase suggested that SLP occupied more important position than ETP in the energy generation within ruminal microbiome, accounting for more than 80% of total energy productivity. Furthermore, the increase in dietary energy levels promotes rumen energy productivity and MCP yield by improving levels of ETP and SLP coupled to glucose fermentation in the ruminal microbiome. But, an increase in dietary protein level has no such effects.

ETHICS STATEMENT

This study was approved by the Animal Care and Use Committee of Nanjing Agricultural University, in compliance with the

Regulations for the Administration of Affairs Concerning Experimental Animals (The State Science and Technology Commission of China, 2010).

AUTHOR CONTRIBUTIONS

HS wrote the manuscript. HS and ZX analyzed data. ZL and ZS designed the research. ZL and YT performed the experiments. All authors approved the final manuscript.

FUNDING

This work was supported by the National Natural Science Foundation of China (A0201800763), the Fundamental Research Funds for the Central Universities (Y0201801248), and the Science Foundation of Jiangsu Province (BK20180542).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00847/full#supplementary-material>

REFERENCES

- Alneberg, J., Bjarnason, B. S., De Bruijn, I., Schirmer, M., Quick, J., Ijaz, U. Z., et al. (2014). Binning metagenomic contigs by coverage and composition. *Nat. Methods* 11, 1144–1146. doi: 10.1038/nmeth.3103
- Bond, D. R., and Lovley, D. R. (2003). Electricity production by *Geobacter sulfurreducens* attached to electrodes. *Appl. Environ. Microbiol.* 69, 1548–1555. doi: 10.1128/AEM.69.3.1548-1555.2003
- Bongaerts, J., Zoske, S., Weidner, U., and Uden, G. (1995). Transcriptional regulation of the proton translocating NADH dehydrogenase genes (nuoA-N) of *Escherichia coli* by electron acceptors, electron donors and gene regulators. *Mol. Microbiol.* 16, 521–534. doi: 10.1111/j.1365-2958.1995.tb02416.x
- Bruna, F., Mousty, C., Besse-Hoggan, P., Batisson, I., and Prevot, V. (2018). Assembly of nitroreductase and layered double hydroxides toward functional biohybrid materials. *J. Colloid Interface Sci.* 533, 71–81. doi: 10.1016/j.jcis.2018.07.126
- Cameron, M. R., Klusmeyer, T. H., Lynch, G. L., Clark, J. H., and Nelson, D. R. (1991). Effects of urea and starch on rumen fermentation, nutrient passage to the duodenum, and performance of cows. *J. Dairy Sci.* 74, 1321–1336. doi: 10.3168/jds.S0022-0302(91)78288-X
- Casper, D. P., and Schingoethe, D. J. (1989). Lactational response of dairy cows to diets varying in ruminal solubilities of carbohydrate and crude protein. *J. Dairy Sci.* 72, 928–941. doi: 10.3168/jds.S0022-0302(89)79186-4
- Chen, X. B., and Gomes, M. J. (1995). *Estimation of Microbial Protein Supply to Sheep and Cattle based on Urinary Excretion of Purine Derivative – An Overview of the Technique Details*. Aberdeen: International Feed Resources Unit, Rowett Research Institute.
- Cho, G. S., Ritzmann, F., Eckstein, M., Huch, M., Briviba, K., Behnlian, D., et al. (2016). Quantification of *Slackia* and *Eggerthella* spp. in human feces and adhesion of representatives strains to Caco-2 cells. *Front. Microbiol.* 7:658. doi: 10.3389/fmicb.2016.00658
- Cunniff, P., and AOAC International (1997). *Official Methods of Analysis of AOAC International*. Gaithersburg, MD: AOAC International.
- De Rosa, E., Checchetto, V., Franchin, C., Bergantino, E., Berto, P., Szabo, I., et al. (2015). [NiFe]-hydrogenase is essential for cyanobacterium *Synechocystis* sp. PCC 6803 aerobic growth in the dark. *Sci. Rep.* 5:12424. doi: 10.1038/srep12424
- Delamuta, J. R., Ribeiro, R. A., Ormeno-Orrillo, E., Parma, M. M., Melo, I. S., Martinez-Romero, E., et al. (2015). *Bradyrhizobium tropiciagri* sp. nov. and *Bradyrhizobium embrapense* sp. nov., nitrogen-fixing symbionts of tropical forage legumes. *Int. J. Syst. Evol. Microbiol.* 65, 4424–4433. doi: 10.1099/ijsem.0.000592
- Demeyer, D., and Van Nevel, C. (1986). Influence of substrate and microbial interaction on efficiency of rumen microbial growth. *Reprod. Nutr. Dev.* 26, 161–179. doi: 10.1051/rnd:19860203
- Fu, H., Jin, M., Ju, L., Mao, Y., and Gao, H. (2014). Evidence for function overlapping of CymA and the cytochrome bc1 complex in the *Shewanella oneidensis* nitrate and nitrite respiration. *Environ. Microbiol.* 16, 3181–3195. doi: 10.1111/1462-2920.12457
- Harris, S. C., Devendran, S., Mendez-Garcia, C., Mythen, S. M., Wright, C. L., Fields, C. J., et al. (2018). Bile acid oxidation by *Eggerthella lenta* strains C592 and DSM 2243(T). *Gut Microbes* 9, 523–539. doi: 10.1080/19490976.2018.1458180
- Henning, P. H., Steyn, D. G., and Meissner, H. H. (1991). The effect of energy and nitrogen supply pattern on rumen bacterial growth in vitro. *Anim. Sci.* 53, 165–175. doi: 10.1017/S0003356100020080
- Herrmann, G., Jayamani, E., Mai, G., and Buckel, W. (2008). Energy conservation via electron-transferring flavoprotein in anaerobic bacteria. *J. Bacteriol.* 190, 784–791. doi: 10.1128/JB.01422-07
- Hoover, W. H., and Stokes, S. R. (1991). Balancing carbohydrates and proteins for optimum rumen microbial yield. *J. Dairy Sci.* 74, 3630–3644. doi: 10.3168/jds.S0022-0302(91)78553-6
- Kent, W. J. (2002). BLAT—the BLAST-like alignment tool. *Genome Res.* 12, 656–664. doi: 10.1101/gr.229202
- Kljak, K., Pino, F., and Heinrichs, A. J. (2017). Effect of forage to concentrate ratio with sorghum silage as a source of forage on rumen fermentation, N balance, and purine derivative excretion in limit-fed dairy heifers. *J. Dairy Sci.* 100, 213–223. doi: 10.3168/jds.2016-11383
- Kracke, F., Vassilev, I., and Kromer, J. O. (2015). Microbial electron transport and energy conservation - the foundation for optimizing

- bioelectrochemical systems. *Front. Microbiol.* 6:575. doi: 10.3389/fmicb.2015.00575
- Kushkevych, I. V. (2014). Acetate kinase activity and kinetic properties of the enzyme in *Desulfovibrio piger* Vib-7 and *Desulfomicrobium* sp. Rod-9 intestinal bacterial strains. *Open Microbiol. J.* 8, 138–143. doi: 10.2174/1874285801408010138
- Langmead, B., and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9, 357–359. doi: 10.1038/nmeth.1923
- Leahy, S. C., Kelly, W. J., Altermann, E., Ronimus, R. S., Yeoman, C. J., Pacheco, D. M., et al. (2010). The genome sequence of the rumen methanogen *Methanobrevibacter ruminantium* reveals new possibilities for controlling ruminant methane emissions. *PLoS One* 5:e8926. doi: 10.1371/journal.pone.0008926
- Li, F., Li, Y. X., Cao, Y. X., Wang, L., Liu, C. G., Shi, L., et al. (2018). Modular engineering to increase intracellular NAD[H/(+)] promotes rate of extracellular electron transfer of *Shewanella oneidensis*. *Nat. Commun.* 9:3637. doi: 10.1038/s41467-018-05995-8
- Macfarlane, G. T., Macfarlane, S., and Gibson, G. R. (1995). Co-culture of *Bifidobacterium adolescentis* and *Bacteroides thetaiotaomicron* in arabinogalactan-limited chemostats: effects of dilution rate and pH. *Anaerobe* 1, 275–281. doi: 10.1006/anae.1995.1027
- Maeng, W. J., Van Nevel, C. J., Baldwin, R. L., and Morris, J. G. (1976). Rumen microbial growth rates and yields: effect of amino acids and protein. *J. Dairy Sci.* 59, 68–79. doi: 10.3168/jds.S0022-0302(76)84157-4
- Meuer, J., Bartoschek, S., Koch, J., Kunkel, A., and Hedderich, R. (1999). Purification and catalytic properties of Ech hydrogenase from *Methanosarcina barkeri*. *Eur. J. Biochem.* 265, 325–335. doi: 10.1046/j.1432-1327.1999.00738.x
- Moriya, Y., Itoh, M., Okuda, S., Yoshizawa, A. C., and Kanehisa, M. (2007). KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* 35, W182–W185. doi: 10.1093/nar/gkm321
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Parks, D. H., Macdonald, N. J., and Beiko, R. G. (2011). Classifying short genomic fragments from novel lineages using composition and homology. *BMC Bioinformatics* 12:328. doi: 10.1186/1471-2105-12-328
- Peng, Y., Leung, H. C., Yiu, S. M., and Chin, F. Y. (2012). IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28, 1420–1428. doi: 10.1093/bioinformatics/bts174
- Plugge, C. M., Henstra, A. M., Worm, P., Swarts, D. C., Paulitsch-Fuchs, A. H., Scholten, J. C., et al. (2012). Complete genome sequence of *Syntrophobacter fumaroxidans* strain (MPOB(T)). *Stand. Genom. Sci.* 7, 91–106. doi: 10.4056/sigs.2996379
- Rho, M., Tang, H., and Ye, Y. (2010). FragGeneScan: predicting genes in short and error-prone reads. *Nucleic Acids Res.* 38:e191. doi: 10.1093/nar/gkq747
- Satter, L. D., and Roffler, R. E. (1975). Nitrogen requirement and utilization in dairy cattle. *J. Dairy Sci.* 58, 1219–1237. doi: 10.3168/jds.S0022-0302(75)84698-4
- Satter, L. D., and Slyter, L. L. (1974). Effect of ammonia concentration of rumen microbial protein production in vitro. *Br. J. Nutr.* 32, 199–208. doi: 10.1079/BJN19740073
- Schmieder, R., and Edwards, R. (2011). Fast identification and removal of sequence contamination from genomic and metagenomic datasets. *PLoS One* 6:e17288. doi: 10.1371/journal.pone.0017288
- Schuchmann, K., and Muller, V. (2014). Autotrophy at the thermodynamic limit of life: a model for energy conservation in acetogenic bacteria. *Nat. Rev. Microbiol.* 12, 809–821. doi: 10.1038/nrmicro3365
- Schwab, M. A., Kolker, S., Van Den Heuvel, L. P., Sauer, S., Wolf, N. I., Rating, D., et al. (2005). Optimized spectrophotometric assay for the completely activated pyruvate dehydrogenase complex in fibroblasts. *Clin. Chem.* 51, 151–160. doi: 10.1373/clinchem.2004.033852
- Storm, E., Brown, D. S., and Orskov, E. R. (1983). The nutritive value of rumen micro-organisms in ruminants. 3. The digestion of microbial amino and nucleic acids in, and losses of endogenous nitrogen from, the small intestine of sheep. *Br. J. Nutr.* 50, 479–485. doi: 10.1079/BJN19830116
- Tamminga, S. (1979). Protein degradation in the forestomachs of ruminants. *J. Anim. Sci.* 49, 1615–1630. doi: 10.2527/jas1979.4961615x
- Tas, M. V., Evans, R. A., and Axford, R. F. (1981). The digestibility of amino acids in the small intestine of the sheep. *Br. J. Nutr.* 45, 167–174. doi: 10.1079/BJN19810089
- Van Driessche, E., Houf, K., Vangroenweghe, F., De Zutter, L., and Van Hoof, J. (2005). Prevalence, enumeration and strain variation of *Arcobacter* species in the faeces of healthy cattle in Belgium. *Vet. Microbiol.* 105, 149–154. doi: 10.1016/j.vetmic.2004.11.002
- Van Soest, P. J., Robertson, J. B., and Lewis, B. A. (1991). Methods for dietary fiber, neutral detergent fiber, and nonstarch polysaccharides in relation to animal nutrition. *J. Dairy Sci.* 74, 3583–3597. doi: 10.3168/jds.S0022-0302(91)78551-2
- Virtanen, A. I. (1966). Milk production of cows on protein-free feed. *Science* 153, 1603–1614. doi: 10.1126/science.153.3744.1603
- Weatherburn, M. W. (1967). Phenol-hypochlorite reaction for determination of ammonia. *Anal. Chem.* 39, 971–974. doi: 10.1021/ac60252a045
- Yang, W., Shen, Z., and Martens, H. (2012). An energy-rich diet enhances expression of Na(+)/H(+) exchanger isoform 1 and 3 messenger RNA in rumen epithelium of goat. *J. Anim. Sci.* 90, 307–317. doi: 10.2527/jas.2011-3854
- Zhao, L., Meng, Q., Ren, L., Liu, W., Zhang, X., Huo, Y., et al. (2015). Effects of nitrate addition on rumen fermentation, bacterial biodiversity and abundance. *Asian Australas. J. Anim. Sci.* 28, 1433–1441. doi: 10.5713/ajas.15.0091

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

Copyright © 2019 Lu, Xu, Shen, Tian and Shen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Approaches to Computational Strain Design in the Multiomics Era

Peter C. St. John and Yannick J. Bomble*

National Renewable Energy Laboratory, Golden, CO, United States

OPEN ACCESS

Edited by:

Yinjie Tang,
Washington University in St. Louis,
United States

Reviewed by:

Hector Garcia Martin,
Lawrence Berkeley National
Laboratory, United States Department
of Energy (DOE), United States
Ilya R. Akberdin,
San Diego State University,
United States

*Correspondence:

Yannick J. Bomble
Yannick.bomble@nrel.gov

Specialty section:

This article was submitted to
Systems Microbiology,
a section of the journal
Frontiers in Microbiology

Received: 21 December 2018

Accepted: 08 March 2019

Published: 05 April 2019

Citation:

St. John PC and Bomble YJ
(2019) Approaches to Computational
Strain Design in the Multiomics Era.
Front. Microbiol. 10:597.
doi: 10.3389/fmicb.2019.00597

Modern omics analyses are able to effectively characterize the genetic, regulatory, and metabolic phenotypes of engineered microbes, yet designing genetic interventions to achieve a desired phenotype remains challenging. With recent developments in genetic engineering techniques, timelines associated with building and testing strain designs have been greatly reduced, allowing for the first time an efficient closed loop iteration between experiment and analysis. However, the scale and complexity associated with multi-omics datasets complicates manual biological reasoning about the mechanisms driving phenotypic changes. Computational techniques therefore form a critical part of the Design-Build-Test-Learn (DBTL) cycle in metabolic engineering. Traditional statistical approaches can reduce the dimensionality of these datasets and identify common motifs among high-performing strains. While successful in many studies, these methods do not take full advantage of known connections between genes, proteins, and metabolic networks. There is therefore a growing interest in model-aided design, in which modeling frameworks from systems biology are used to integrate experimental data and generate effective and non-intuitive design predictions. In this mini-review, we discuss recent progress and challenges in this field. In particular, we compare methods augmenting flux balance analysis with additional constraints from fluxomic, genomic, and metabolomic datasets and methods employing kinetic representations of individual metabolic reactions, and machine learning. We conclude with a discussion of potential future directions for improving strain design predictions in the omics era and remaining experimental and computational hurdles.

Keywords: constraint-based methods, kinetic metabolic models, machine learning, multiomics, strain engineering

INTRODUCTION

The biorefinery concept involves the development of sustainable and low-impact production routes for major commodity chemicals and fuels from biomass (Bozell and Petersen, 2010). Biomanufacturing using engineered microbes is a critical component of many production pathways, and offers the opportunity for high selectivity and yield (Nielsen and Keasling, 2016). However, optimizing microbial metabolism for a given process is time intensive and costly, limiting microbial bioconversions at present to only a few commercially successful compounds (Van Dien, 2013; Chubukov et al., 2016). This difficulty is primarily due to the complex relationship between genotype and phenotype, involving regulation at the metabolic, translational, and transcriptional levels. In recent years, the procedure of strain engineering has been formalized through the

Design-Build-Test-Learn (DBTL) cycle, which takes advantage of recent improvements in genetic engineering and high-throughput characterization in the Build and Test stages, respectively, to efficiently screen larger libraries of strain modifications (Liu et al., 2015). The Learn and Design stages use computational techniques to interpret experimental results and suggest further modification targets. The Learn step is perhaps the most weakly developed step of the DBTL cycle, and can take the form of a wide range of computational techniques from statistical analysis to detailed simulations (Nielsen and Keasling, 2016). In this minireview, we discuss recent research in methodology for integrating biological data – particularly in the form of multiomics analyses – into developing new and efficient strain designs. We first review relevant experimental considerations from the Test stage and summarize the types of data available for informing strain designs. We next cover constraint based methods, kinetic simulations, and machine learning approaches, as well as recent studies that have used these methods in strain design. Lastly, we finish by discussing available software implementations and future directions for tackling the Learn step.

EXPERIMENTAL INPUTS

A number of recent reviews have covered the growing usefulness of omics approaches in characterizing cell physiology (Petzold et al., 2015; Nielsen, 2017; Becker and Wittmann, 2018; Yurkovich and Palsson, 2018), and therefore we only briefly cover the relevant data generated in typical strain characterization experiments. Frequently used omics data include transcriptomics, proteomics, metabolomics, and fluxomics, which measure gene expression, protein expression, metabolite concentrations, and intracellular fluxes, respectively. Transcriptomics is typically performed using next-generation sequencing methods that quantify relative differences in RNA expression within a given biological sample (Petzold et al., 2015). Relative comparisons between samples are also possible using statistical techniques (Wagner et al., 2012). Due to the similar physical nature of RNA transcripts, transcriptomics approaches are among the easiest to perform at the genome-scale, but their distance from metabolic networks by several layers of regulation makes direct understanding of metabolic function using these data difficult. Proteomics is one step closer to the determination of metabolic fluxes and uses mass spectrometry to quantify protein expression through the amino acid sequences of digested peptides (Kolker et al., 2006). Similar to transcriptomics, proteomics experiments typically measure relative protein expression within a sample, although statistical and experimental methods for comparing relative protein expression between samples are possible (Petzold et al., 2015). Absolute quantification of protein expression is feasible but more difficult, with a range of accuracies depending on the method used (Ariake et al., 2012). While more involved than transcriptomics due to protein's 3D structure and lack of amplification techniques, proteomic analyses are still able to survey a similar fraction of the protein-coding genome

(Haider and Pal, 2013). Metabolomics poses an even greater challenge, as the high turnover of metabolites requires fast quenching and processing of samples (Petzold et al., 2015). As a result, the scope of metabolomic analyses are typically restricted to a smaller fraction of the organism's metabolism. Similar to transcriptomics and proteomics, metabolite concentrations are typically measured as relative quantities in high-throughput exploratory experiments (Lei et al., 2011). Absolute metabolite quantifications are possible in targeted metabolomic studies using external or isotope-labeled standards. Lastly, fluxomics is concerned with accurately measuring internal fluxes of key metabolic reactions directly using isotopic labeling. While an excellent indicator of metabolic state, fluxomics is performed with less frequency than the previously discussed methods due to its experimental difficulty (Blank, 2016). In addition to careful cell culture and sample processing, fluxomics requires an accurate mathematical model that tracks atom transitions during metabolic reactions (Wiechert, 2001). This mathematical model is used in conjunction with ^{13}C isotope labeling patterns to infer fluxes through each reaction, and as a result, inferred fluxes have typically been restricted to the main reactions in central carbon metabolism. However, extensions of MFA to include genome-scale flux analysis have been proposed (Gopalakrishnan and Maranas, 2015). Some genome-scale MFA methods leverage metabolism's bow-tie structure to constrain fluxes through peripheral pathways with a high degree of confidence (García Martín et al., 2015; Ando and Garcia Martin, 2018).

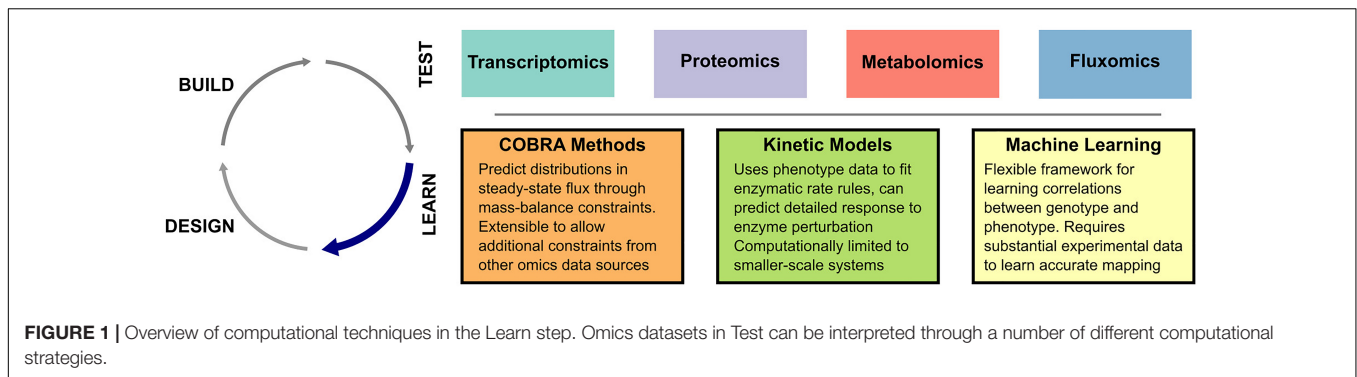
Even with access to direct measurements of activity for a wide range cellular machinery components, using these data to enhance metabolic flux for a desired pathway remains challenging. We next discuss Learn techniques that synthesize these vast data sources together with generalized knowledge of biological function.

LEARN METHODOLOGY

The goal of the Learn and Design steps is to use the characterization of previously engineered strains to develop improved strain designs. In its most basic form, this step can be accomplished by examining biological features (i.e., differentially expressed genes) correlated with improved strain performance, and overexpressing those likely involved in the pathway of interest (Yoshikawa et al., 2012). Designs based on rational consideration of omics data have proven successful (Guan et al., 2017), validating the human-in-the-loop approach. However, model driven designs will likely be critical to speeding up the DBTL cycle and revealing non-intuitive targets (Vickers, 2016). In the next sections, we review several lines of research into model-driven interpretation of omics data. A schematic of these approaches is shown in **Figure 1**.

Constraint-Based Methods

Constraint-Based Reconstruction and Analysis (COBRA) methods use biological knowledge and data to place constraints on intracellular fluxes, and in recent years have expanded to consider a wide range of recent omics techniques. Here we



focus on extensions of COBRA methods that pertain to guiding strain designs from omics data, while a number of recent reviews have covered COBRA methods in greater depth (O'Brien et al., 2015; Campbell et al., 2017; Stalidzans et al., 2018). A central technique to COBRA methods is flux balance analysis (FBA), which assumes that metabolite concentrations in the cell reach a pseudo steady-state when compared to the time scales associated with substrate uptake and cell division (Orth et al., 2010). This assumption allows fluxes to be constrained by mass balance equations developed from databases of biochemical reaction stoichiometry. Mass balance constraints alone (in the absence of ^{13}C isotope labeling or other product data) are often not sufficient to determine a unique vector of metabolic fluxes. By assuming a cellular objective such as maximizing biomass or ATP production, unique flux vectors can be predicted. The accuracy of these predicted flux values are dependent on the objective chosen, and some objective functions have shown good correlation with experimental omics data (Lewis et al., 2010). Since such models can be simulated quickly and rely primarily on well-curated databases of metabolic reactions, many genome-scale models (GEMs) of microbial metabolism have been created (Henry et al., 2010; King et al., 2015). While useful in understanding metabolic functionality and predicting the results of gene manipulation, these assumptions are not sufficient to fully incorporate the phenotypic observations resulting from omics analyses.

Extensions to the COBRA framework have therefore been proposed to impose additional constraints from experimental observations. One of the earliest such studies used transcriptomic data to block flux through reactions where gene expression for required enzymes was not observed (Åkesson et al., 2004). This method considered gene product expression through boolean logic, however, more recent studies have explicitly included gene product expression in the constraint-based framework (Becker and Palsson, 2008; Shlomi et al., 2008). Metabolism and gene-expression models (ME-models) explicitly model reactions involved in transcription and translations to build a quantitative model of enzyme production and usage (Lerman et al., 2012). These models therefore allow direct comparison of model predictions with transcriptomic and proteomic data (O'Brien et al., 2014, 2015). In a similar method, genome-scale models with protein structures (GEM-PROs) include structural information about each enzymatically catalyzed reaction (Chang et al., 2013). Such models allow

the explicit simulation of the proteome fraction devoted to different cellular activities (Basan et al., 2015), and therefore might also be used to add additional constraints from proteomic analyses. The GECKO method combines literature knowledge of enzyme kinetics with proteomics data to constrain metabolic fluxes (Sánchez et al., 2017). However, while many enzymes have been kinetically characterized for well-studied species, these data are typically not available for non-model microbes (Nilsson et al., 2017).

Metabolomics data are typically incorporated in constraint-based models through the explicit consideration of reaction thermodynamics. If absolute metabolite concentrations are available, thermodynamic metabolic flux analysis can provide more condition-specific information on irreversible reactions (Henry et al., 2007). These principles have been successfully applied to select the most promising pathways for the synthesis of a variety of products (Averesch et al., 2017). Further extensions to the COBRA framework will likely include even more cellular functionality. Toward this goal, whole-cell models that integrate gene expression, protein production, and cell cycle have been constructed (Karr et al., 2012).

Constraint-Based Reconstruction and Analysis methods therefore represent an extensible and computationally efficient framework for connecting omics data of different types and have been used to successfully interpret omics data and improve strain designs in a number of studies (Wisselink et al., 2010; Brunk et al., 2016). An advantage of COBRA methods is their limited number of parameters that must be fit from experimental data, and therefore they are often able to suggest strain designs without substantial experimental support. In particular, these methods are especially efficient in determining metabolic changes that couple product production to cell growth (Long et al., 2015). The accuracy of constraint-based models in predicting *de novo* experimental results has not been rigorously evaluated and would serve a useful study in measuring the progress in our understanding of cellular behavior. However, even modest success rates from predictive tools are useful in guiding experimental efforts where the search space is vast. A limitation of constrained-based methods is that they are often less suitable for suggesting improvements to fine-tune the enzyme expression of an existing pathway. Such a task typically requires a kinetic description of the reactions in question, which we discuss in the next section.

Kinetic Metabolic Models

The goal of kinetic metabolic models is to capture the dynamic behavior of individual enzymes and integrate these expressions into the behavior of the full metabolic network. These models allow the direct prediction steady-state flux distributions as a function of enzyme expression, which typically serve as the most reliable experimental data for validation. However, models that explicitly incorporate enzyme kinetics (if parameterized correctly) are capable of predicting finer details of pathway dynamics, including the effect of slight changes in enzyme activity on metabolic flux. In constraint-based models, metabolite pools are assumed to be in a pseudo steady-state, and thus the rate rules governing flux through each reaction can be ignored. While the steady-state assumption may be justified, the specific steady-state reached inside the cell is determined, among a multitude of factors, both by external metabolite conditions as well as the kinetics and expression levels of metabolic enzymes. Kinetic modeling frameworks therefore seek to estimate these reaction rate rules from observed metabolic phenotypes to predict how enzyme perturbation will affect steady-state concentrations and fluxes.

Small-scale kinetic models of core carbon metabolism can leverage enzyme kinetics *in vitro* and time-course metabolite concentration measurements in fitting parameter values (Chassagnole et al., 2002). However, transient cellular responses are difficult to measure at the genome-scale, and direct enzyme kinetic measurements are sparser for peripheral pathways. Large-scale dynamic metabolic simulations are therefore largely based on steady-state flux and concentration data (Vasilakou et al., 2016). Because of these limited data, quantifying parameter uncertainty is therefore a critical challenge in large-scale kinetic models (Tummler and Klipp, 2018). Metabolic ensemble modeling addresses this challenge directly by finding distributions in parameter values that all reproduce the observed experimental data (Tran et al., 2008). This approach has been used to suggest subsequent enzymes in a linear pathway for overexpression (Contador et al., 2009), and an ensemble-based kinetic model of *Escherichia coli* has demonstrated superior predictive ability of steady-state flux distributions (Khodayari et al., 2014).

Smaller-scale, hand-curated kinetic models can use rate rules for individual enzymes with experimentally validated functional forms. However, traditional rate rule expressions (such as Michaelis–Menten kinetics) become difficult to construct for reactions with many participating species. Accordingly, larger-scale kinetic models typically choose a generalizable framework for constructing rate rule expressions. These frameworks range in computational complexity and faithfulness to the underlying enzyme-substrate system, and we leave a detailed comparison of these approaches to a number of recently published reviews (Heijnen, 2005; Hadlich et al., 2009; Du et al., 2016; Saa and Nielsen, 2017). Software available for kinetic modeling has continued to improve, and typically allows the user to specify reaction stoichiometry and rate rules independently from the chosen simulation algorithm. Such software includes COPASI (Hoops et al., 2006), CellDesigner (Funahashi et al., 2008), and MATCONT (Dhooge et al., 2003).

Regardless of the framework chosen, a major hurdle in using kinetic models for interpretation of omics data is the computational effort required in parameter estimation. In metabolic ensemble modeling, parameters are sampled at random and retained in the final ensemble only if they match all the considered experimental data (Tran et al., 2008). As a result, as more data is added or the model expanded, the computational costs increase substantially. Methods for improving the computational speed of the approach have been developed (Greene et al., 2017), but calculating steady states of the dynamic model remains a computational bottleneck. Ensemble-based inference approaches are therefore typically applied to smaller, core-carbon metabolic networks (Khodayari et al., 2014). A recent genome-scale kinetic modeling study optimized only a single parameter set due to the added cost of ensemble-based parameter estimation (Khodayari and Maranas, 2016). However, this single parameter set demonstrated a superior ability to reproduce a wide range of experimental observations compared with constraint-based methods (Khodayari and Maranas, 2016). The ensemble modeling sampling approach has been recently formalized as a form of Bayesian inference (Saa and Nielsen, 2016), demonstrating that detailed posterior distributions in parameter estimates and model predictions could be found. Kinetic models therefore offer a promising future direction for incorporating vast quantities of omics data in metabolic reconstructions if computational bottlenecks can be circumvented (St. John et al., 2018). While difficult to fit, the added parameters from kinetic representations give these models more expressive power in fitting experimental data.

A factor complicating the analysis of experimental data with kinetic models is the stochasticity introduced by low cell volumes and small copy numbers of several key enzymes (Levine and Hwa, 2007; Kiviet et al., 2014). Cell to cell heterogeneity therefore imposes unique challenges in understanding microbial kinetics that might be resolved through the use of explicit stochastic simulation algorithms (Gillespie, 1977) as implemented in a variety of software packages (Hoops et al., 2006; Sanft et al., 2011; Abel et al., 2016). In the subsequent section we discuss machine learning approaches that add even more parameters to be fit, but may prove useful as high-throughput strain construction and characterization techniques improve.

Machine Learning

Machine learning methods for interpreting omics data have taken a wide range of forms, largely due to the many varied biological questions that can be asked. In this section, we focus on methods that predict future targets for strain engineering. Integrative omics analyses attempt to draw connects between disparate omics data sources, either with or without prior biological knowledge (Berger et al., 2013; Bersanelli et al., 2016). These methods have been used to predict key regulatory genes correlated with metabolic productivity (Larsen et al., 2018), and inferred regulatory networks have also been incorporated into FBA models (Chandrasekaran and Price, 2010). Other studies have used machine learning to understand and predict metabolic performance from hyperparameters associated with cell growth. Wu et al. (2016) explored methods for machine learning in

meta-analysis to predict likely pathway success as a function of the complexity of the engineered pathway and other factors. In Kim et al. (2016), machine learning methods are used both for data reconciliation between omics sources, as well as to directly map the genotype-phenotype relationship. Another interesting study used machine learning methods as a replacement for the traditional rate equation frameworks discussed in the previous section (Costello and Martin, 2018). In that study, rate equations were learned directly from time-series metabolomics and were successful in predicting medium-producing strains given high and low-producing varieties. Costello and Martin (2018) also quantified the amount of data required for accurate rate determination at approximately 10 strains. Given the rapid advancement of machine learning methods and biological data collection, these approaches may offer flexible and efficient ways of directly incorporate biological data in new strain designs.

DISCUSSION

Since Learn lags behind the rest of the DBTL methodology in the development of validated and standardized techniques, feasible computational techniques are still being explored and improved upon. As a result, software libraries for performing the analyses described in this minireview are relatively scarce. As the most mature method of the three, COBRA methods have relatively strong software support in both the MATLAB (Heirendt et al., 2017) and Python (Ebrahim et al., 2013) ecosystems. Dependent packages have also been created for a number of the COBRA extensions for integrating or predicting omics-level data. Kinetic models, alternatively, have relatively poor support in the software landscape. This is likely due to the multitude of kinetic frameworks available as well as their slow (but parallelizable) convergence, requiring hardware-dependent simulation strategies. For machine learning, several actively developed packages are available that implement common approaches. Scikit-Learn for Python implements a variety of machine learning strategies under a consistent API (Pedregosa et al., 2011).

REFERENCES

- Abel, J. H., Drawert, B., Hellander, A., and Petzold, L. R. (2016). GillesPy: a python package for stochastic model building and simulation. *IEEE Life Sci. Lett.* 2, 35–38. doi: 10.1109/lis.2017.2652448 doi: 10.1109/LLS.2017.2652448
- Åkesson, M., Förster, J., and Nielsen, J. (2004). Integration of gene expression data into genome-scale metabolic models. *Met. Eng.* 6, 285–293. doi: 10.1016/j.ymben.2003.12.002
- Ando, D., and Garcia Martin, H. (2018). Two-scale ¹³C metabolic flux analysis for metabolic engineering. *Synth. Metab. Pathways* 1671, 333–352. doi: 10.1007/978-1-4939-7295-1_21
- Arike, L., Valgepea, K., Peil, L., Nahku, R., Adamberg, K., and Vilu, R. (2012). Comparison and applications of label-free absolute proteome quantification methods on *Escherichia coli*. *J. Proteomics* 75, 5437–5448. doi: 10.1016/j.jprot.2012.06.020
- Averesch, N. J. H., Martínez, V. S., Nielsen, L. K., and Krömer, J. O. (2017). Toward synthetic biology strategies for adipic acid production: an in silico tool for combined thermodynamics and stoichiometric analysis of metabolic networks. *ACS Synth. Biol.* 7, 490–509. doi: 10.1021/acssynbio.7b00304
- Deep learning frameworks such as Tensorflow or PyTorch simplify the process of constructing deep neural networks and training them on specialized hardware. Compared to the availability of general-purpose machine learning, omics-specific machine learning analyses have substantially fewer libraries under active development. However, creating and distributing standardized Learn work flows will be critical to enabling the reproducible analyses required of the iterative DBTL cycle. Such standardized approaches will necessarily require the development and maintenance of software and best practices in the metabolic modeling community.
- ## AUTHOR CONTRIBUTIONS
- PSJ and YB contributed to the conception and writing of the manuscript. PSJ created the figure. YB supervised the research.
- ## FUNDING
- We thank the U.S. Department of Energy Bioenergy Technologies Office for funding under Contract DE-AC36-08GO28308 with the National Renewable Energy Laboratory. This work was authored by Alliance for Sustainable Energy, LLC, the Manager and Operator of the National Renewable Energy Laboratory for the United States Department of Energy (DOE) under Contract No. DE-AC36-08GO28308. The views expressed in the article do not necessarily represent the views of the DOE or the U.S. Government.
- ## ACKNOWLEDGMENTS
- The U.S. Government retains and the publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid-up, irrevocable, worldwide license to publish or reproduce the published form of this work, or allow others to do so, for U.S. Government purposes.
- Basan, M., Hui, S., Okano, H., Zhang, Z., Shen, Y., Williamson, J. R., et al. (2015). Overflow metabolism in *Escherichia coli* results from efficient proteome allocation. *Nature* 528, 99–104. doi: 10.1038/nature15765
- Becker, J., and Wittmann, C. (2018). From systems biology to metabolically engineered cells — an omics perspective on the development of industrial microbes. *Curr. Opin. Microbiol.* 45, 180–188. doi: 10.1016/j.mib.2018.06.001
- Becker, S. A., and Palsson, B. O. (2008). Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.* 4:e1000082. doi: 10.1371/journal.pcbi.1000082
- Berger, B., Peng, J., and Singh, M. (2013). Computational solutions for omics data. *Nat. Rev. Genet.* 14, 333–346. doi: 10.1038/nrg3433
- Bersanelli, M., Mosca, E., Remondini, D., Giampieri, E., Sala, C., Castellani, G., et al. (2016). Methods for the integration of multi-omics data: mathematical aspects. *BMC Bioinformatics* 17:S15. doi: 10.1186/s12859-015-0857-9
- Blank, L. M. (2016). Let's talk about flux or the importance of (intracellular) reaction rates. *Microbial Biotechnol.* 10, 28–30. doi: 10.1111/1751-7915.12455
- Bozell, J. J., and Petersen, G. R. (2010). Technology development for the production of biobased products from biorefinery carbohydrates—the us department of energy's "top 10" revisited. *Green Chem.* 12:539. doi: 10.1039/b922014c

- Brunk, E., George, K. W., Alonso-Gutierrez, J., Thompson, M., Baidoo, E., Wang, G., et al. (2016). Characterizing strain variation in engineered *E. coli* using a multi-omics-based workflow. *Cell Syst.* 2, 335–346. doi: 10.1016/j.cels.2016.04.004
- Campbell, K., Xia, J., and Nielsen, J. (2017). The impact of systems biology on bioprocessing. *Trends Biotechnol.* 35, 1156–1168. doi: 10.1016/j.tibtech.2017.08.011
- Chandrasekaran, S., and Price, N. D. (2010). Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17845–17850. doi: 10.1073/pnas.1005139107
- Chang, R. L., Andrews, K., Kim, D., Li, Z., Godzik, A., and Palsson, B. O. (2013). Structural systems biology evaluation of metabolic thermotolerance in *Escherichia coli*. *Science* 340, 1220–1223. doi: 10.1126/science.1234012
- Chassagnole, C., Noisommit-Rizzi, N., Schmid, J. W., Mauch, K., and Reuss, M. (2002). Dynamic modeling of the central carbon metabolism of *Escherichia coli*. *Biotechnol. Bioeng.* 79, 53–73. doi: 10.1002/bit.10288
- Chubukov, V., Mukhopadhyay, A., Petzold, C. J., Keasling, J. D., and Martin, H. G. (2016). Synthetic and systems biology for microbial production of commodity chemicals. *Syst. Biol. Appl.* 2:16009. doi: 10.1038/npsba.2016.9
- Contador, C. A., Rizk, M. L., Asenjo, J. A., and Liao, J. C. (2009). Ensemble modeling for strain development of l-lysine-producing *Escherichia coli*. *Metab. Eng.* 11, 221–233. doi: 10.1016/j.ymben.2009.04.002
- Costello, Z., and Martin, H. G. (2018). A machine learning approach to predict metabolic pathway dynamics from time-series multiomics data. *Syst. Biol. Appl.* 4:19. doi: 10.1038/s41540-018-0054-3
- Dhooze, A., Govaerts, W., and Kuznetsov, Y. A. (2003). MATCONT. *ACM Trans. Math. Softw.* 29, 141–164. doi: 10.1145/779359.779362
- Du, B., Zielinski, D. C., Kavvas, E. S., Dräger, A., Tan, J., Zhang, Z., et al. (2016). Evaluation of rate law approximations in bottom-up kinetic models of metabolism. *BMC Syst. Biol.* 10:40. doi: 10.1186/s12918-016-0283-2
- Ebrahim, A., Lerman, J. A., Palsson, B. O., and Hyduke, D. R. (2013). COBRApy: constraints-based reconstruction and analysis for python. *BMC Syst. Biol.* 7:74. doi: 10.1186/1752-0509-7-74
- Funahashi, A., Matsuoka, Y., Jouraku, A., Morohashi, M., Kikuchi, N., and Kitano, H. (2008). CellDesigner 3.5: a versatile modeling tool for biochemical networks. *Proc. IEEE* 96, 1254–1265. doi: 10.1109/jproc.2008.925458
- García Martín, H., Kumar, V. S., Weaver, D., Ghosh, A., Chubukov, V., Mukhopadhyay, A., et al. (2015). A method to constrain genome-scale models with ¹³C labeling data. *PLoS Comput. Biol.* 11:e1004363. doi: 10.1371/journal.pcbi.1004363
- Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81, 2340–2361. doi: 10.1021/j100540a008
- Gopalakrishnan, S., and Maranas, C. D. (2015). ¹³C metabolic flux analysis at a genome-scale. *Metab. Eng.* 32, 12–22. doi: 10.1016/j.ymben.2015.08.006
- Greene, J. L., Wächter, A., Tyo, K. E., and Broadbelt, L. J. (2017). Acceleration strategies to enhance metabolic ensemble modeling performance. *Biophys. J.* 113, 1150–1162. doi: 10.1016/j.bpj.2017.07.018
- Guan, N., Du, B., Li, J., Shin, H. D., Chen, R. R., Du, G., et al. (2017). Comparative genomics and transcriptomics analysis-guided metabolic engineering of propionibacterium acidipropionici for improved propionic acid production. *Biotechnol. Bioeng.* 115, 483–494. doi: 10.1002/bit.26478
- Hadlich, F., Noack, S., and Wiechert, W. (2009). Translating biochemical network models between different kinetic formats. *Metab. Eng.* 11, 87–100. doi: 10.1016/j.ymben.2008.10.002
- Haider, S., and Pal, R. (2013). Integrated analysis of transcriptomic and proteomic data. *Curr. Genomics* 14, 91–110. doi: 10.2174/1389202911314020003
- Heijnen, J. J. (2005). Approximate kinetic formats used in metabolic network modeling. *Biotechnol. Bioeng.* 91, 534–545. doi: 10.1002/bit.20558
- Heirendt, L., Arreckx, S., Pfau, T., Mendoza, S. N., Richelle, A., Heinken, A., et al. (2017). Creation and analysis of biochemical constraint-based models: the cobra toolbox v3.0. *arXiv* <https://arxiv.org/abs/1710.04038>
- Henry, C. S., Broadbelt, L. J., and Hatzimanikatis, V. (2007). Thermodynamics-based metabolic flux analysis. *Biophys. J.* 92, 1792–1805. doi: 10.1529/biophysj.106.093138
- Henry, C. S., DeJongh, M., Best, A. A., Frybarger, P. M., Lindsay, B., and Stevens, R. L. (2010). High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat. Biotechnol.* 28, 977–982. doi: 10.1038/nbt.1672
- Hoops, S., Sahle, S., Gauges, R., Lee, C., Pahle, J., Simus, N., et al. (2006). COPASI—a complex pathway simulator. *Bioinformatics* 22, 3067–3074. doi: 10.1093/bioinformatics/btl485
- Karr, J. R., Sanghvi, J. C., Macklin, D. N., Gutschow, M. V., Jacobs, J. M., Bolival, B., et al. (2012). A whole-cell computational model predicts phenotype from genotype. *Cell* 150, 389–401. doi: 10.1016/j.cell.2012.05.044
- Khodayari, A., and Maranas, C. D. (2016). A genome-scale *Escherichia coli* kinetic metabolic model k-ecoli457 satisfying flux data for multiple mutant strains. *Nat. Commun.* 7:13806. doi: 10.1038/ncomms13806
- Khodayari, A., Zomorodi, A. R., Liao, J. C., and Maranas, C. D. (2014). A kinetic model of *Escherichia coli* core metabolism satisfying multiple sets of mutant flux data. *Metab. Eng.* 25, 50–62. doi: 10.1016/j.ymben.2014.05.014
- Kim, M., Rai, N., Zorraquino, V., and Tagkopoulos, I. (2016). Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*. *Nat. Commun.* 7:13090. doi: 10.1038/ncomms13090
- King, Z. A., Lu, J., Dräger, A., Miller, P., Federowicz, S., Lerman, J. A., et al. (2015). BiGG models: a platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res.* 44, D515–D522. doi: 10.1093/nar/gkv1049
- Kiviet, D. J., Nghe, P., Walker, N., Boulineau, S., Sunderlikova, V., and Tans, S. J. (2014). Stochasticity of metabolism and growth at the single-cell level. *Nature* 514, 376–379. doi: 10.1038/nature13582
- Kolker, E., Higdon, R., and Hogan, J. M. (2006). Protein identification and expression analysis using mass spectrometry. *Trends Microbiol.* 14, 229–235. doi: 10.1016/j.tim.2006.03.005
- Larsen, P. E., Zerbis, S., Laible, P. D., Collart, F. R., Korajczyk, P., Dai, Y., et al. (2018). Modeling the *Pseudomonas* sulfur regulome by quantifying the storage and communication of information. *Msystems* 3:e00189-17. doi: 10.1128/msystems.00189-17
- Lei, Z., Huhman, D. V., and Sumner, L. W. (2011). Mass spectrometry strategies in metabolomics. *J. Biol. Chem.* 286, 25435–25442. doi: 10.1074/jbc.r111.238691
- Lerman, J. A., Hyduke, D. R., Latif, H., Portnoy, V. A., Lewis, N. E., Orth, J. D., et al. (2012). In silico method for modelling metabolism and gene product expression at genome scale. *Nat. Commun.* 3:929. doi: 10.1038/ncomms1928
- Levine, E., and Hwa, T. (2007). Stochastic fluctuations in metabolic pathways. *Proc. Natl. Acad. Sci. U.S.A.* 104, 9224–9229. doi: 10.1073/pnas.0610987104
- Lewis, N. E., Hixson, K. K., Conrad, T. M., Lerman, J. A., Charusanti, P., Polpitiya, A. D., et al. (2010). Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* 6:390. doi: 10.1038/msb.2010.47
- Liu, R., Bassalo, M. C., Zeitoun, R. I., and Gill, R. T. (2015). Genome scale engineering techniques for metabolic engineering. *Metab. Eng.* 32, 143–154. doi: 10.1016/j.ymben.2015.09.013
- Long, M. R., Ong, W. K., and Reed, J. L. (2015). Computational methods in metabolic engineering for strain design. *Curr. Opin. Biotechnol.* 34, 135–141. doi: 10.1016/j.copbio.2014.12.019
- Nielsen, J. (2017). Systems biology of metabolism. *Annu. Rev. Biochem.* 86, 245–275. doi: 10.1146/annurev-biochem-061516-044757
- Nielsen, J., and Keasling, J. D. (2016). Engineering cellular metabolism. *Cell* 164, 1185–1197. doi: 10.1016/j.cell.2016.02.004
- Nilsson, A., Nielsen, J., and Palsson, B. O. (2017). Metabolic models of protein allocation call for the kinetome. *Cell Syst.* 5, 538–541. doi: 10.1016/j.cels.2017.11.013
- O'Brien, E. J., Lerman, J. A., Chang, R. L., Hyduke, D. R., and Palsson, B. O. (2014). Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Mol. Syst. Biol.* 9, 693–693. doi: 10.1038/msb.2013.52
- O'Brien, E. J., Monk, J. M., and Palsson, B. O. (2015). Using genome-scale models to predict biological capabilities. *Cell* 161, 971–987. doi: 10.1016/j.cell.2015.05.019
- Orth, J. D., Thiele, I., and Palsson, B. O. (2010). What is flux balance analysis? *Nat. Biotechnol.* 28, 245–248. doi: 10.1038/nbt.1614
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.

- Petzold, C. J., Chan, L. J. G., Nhan, M., and Adams, P. D. (2015). Analytics for metabolic engineering. *Front. Bioeng. Biotechnol.* 3:135. doi: 10.3389/fbioe.2015.00135
- Saa, P. A., and Nielsen, L. K. (2016). Construction of feasible and accurate kinetic models of metabolism: a bayesian approach. *Sci. Rep.* 6:29635. doi: 10.1038/srep29635
- Saa, P. A., and Nielsen, L. K. (2017). Formulation, construction and analysis of kinetic models of metabolism: a review of modelling frameworks. *Biotechnol. Adv.* 35, 981–1003. doi: 10.1016/j.biotechadv.2017.09.005
- Sánchez, B. J., Zhang, C., Nilsson, A., Lahtvee, P., Kerkhoven, E. J., and Nielsen, J. (2017). Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Mol. Syst. Biol.* 13:935. doi: 10.15252/msb.20167411
- Sanft, K. R., Wu, S., Roh, M., Fu, J., Lim, R. K., and Petzold, L. R. (2011). StochKit2: software for discrete stochastic simulation of biochemical systems with events. *Bioinformatics* 27, 2457–2458. doi: 10.1093/bioinformatics/btr401
- Shlomi, T., Cabili, M. N., Herrgård, M. J., Pálsson, B. Ø., and Rupp, E. (2008). Network-based prediction of human tissue-specific metabolism. *Nat. Biotechnol.* 26, 1003–1010. doi: 10.1038/nbt.1487
- St. John, P., Strutz, J., Broadbelt, L. J., Tyo, K. E. J., and Bomble, Y. J. (2018). Bayesian inference of metabolic kinetics from genome-scale multiomics data. *bioRxiv* <https://doi.org/10.1101/450163>
- Stalidzans, E., Seiman, A., Peebo, K., Komasilovs, V., and Pentjuss, A. (2018). Model-based metabolism design: constraints for kinetic and stoichiometric models. *Biochem. Soc. Trans.* 46, 261–267. doi: 10.1042/bst20170263
- Tran, L. M., Rizk, M. L., and Liao, J. C. (2008). Ensemble modeling of metabolic networks. *Biophys. J.* 95, 5606–5617. doi: 10.1529/biophysj.108.135442
- Tummler, K., and Klipp, E. (2018). The discrepancy between data for and expectations on metabolic models: How to match experiments and computational efforts to arrive at quantitative predictions? *Curr. Opin. Syst. Biol.* 8, 1–6. doi: 10.1016/j.coisb.2017.11.003
- Van Dien, S. (2013). From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. *Curr. Opin. Biotechnol.* 24, 1061–1068. doi: 10.1016/j.copbio.2013.03.002
- Vasilakou, E., Machado, D., Theorell, A., Rocha, I., Nöh, K., Oldiges, M., et al. (2016). Current state and challenges for dynamic metabolic modeling. *Curr. Opin. Microbiol.* 33, 97–104. doi: 10.1016/j.mib.2016.07.008
- Vickers, C. (2016). Bespoke design of whole-cell microbial machines. *Microb. Biotechnol.* 10, 35–36. doi: 10.1111/1751-7915.12460
- Wagner, G. P., Kin, K., and Lynch, V. J. (2012). Measurement of mRNA abundance using rna-seq data: RPKM measure is inconsistent among samples. *Theory Biosci.* 131, 281–285. doi: 10.1007/s12064-012-0162-3
- Wiechert, W. (2001). 13C metabolic flux analysis. *Metab. Eng.* 3, 195–206. doi: 10.1006/mben.2001.0187
- Wisselink, H. W., Cipollina, C., Oud, B., Crimi, B., Heijnen, J. J., Pronk, J. T., et al. (2010). Metabolome, transcriptome and metabolic flux analysis of arabinose fermentation by engineered *saccharomyces cerevisiae*. *Metab. Eng.* 12, 537–551. doi: 10.1016/j.ymben.2010.08.003
- Wu, S. G., Shimizu, K., Tang, J. K.-H., and Tang, Y. J. (2016). Facilitate collaborations among synthetic biology, metabolic engineering and machine learning. *ChemBioEng Rev.* 3, 45–54. doi: 10.1002/cben.201500024
- Yoshikawa, K., Furusawa, C., Hirasawa, T., and Shimizu, H. (2012). “Design of superior cell factories based on systems wide omics analysis,” in *Systems Metabolic Engineering*, eds C. Wittmann and S. Lee (Dordrecht: Springer), 57–81. doi: 10.1007/978-94-007-4534-6_3
- Yurkovich, J. T., and Pálsson, B. O. (2018). Quantitative -omic data empowers bottom-up systems biology. *Curr. Opin. Biotechnol.* 51, 130–136. doi: 10.1016/j.copbio.2018.01.009

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

Copyright © 2019 St. John and Bomble. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



¹³C-Metabolic Flux Analysis Reveals the Metabolic Flux Redistribution for Enhanced Production of Poly-γ-Glutamic Acid in *dlt* Over-Expressed *Bacillus licheniformis*

Penghui He¹, Ni Wan², Dongbo Cai¹, Shiyong Hu¹, Yaozhong Chen¹, Shunyi Li¹ and Shouwen Chen^{1,3*}

OPEN ACCESS

Edited by:

Wei Xiong,
National Renewable Energy
Laboratory (DOE), United States

Reviewed by:

Deng Liu,
Washington University in St. Louis,
United States
Joshua Chan,
Colorado State University,
United States

*Correspondence:

Shouwen Chen
mel212@126.com

Specialty section:

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

Received: 30 October 2018

Accepted: 17 January 2019

Published: 01 February 2019

Citation:

He P, Wan N, Cai D, Hu S,
Chen Y, Li S and Chen S (2019)
¹³C-Metabolic Flux Analysis Reveals
the Metabolic Flux Redistribution
for Enhanced Production
of Poly-γ-Glutamic Acid in *dlt*
Over-Expressed *Bacillus licheniformis*.
Front. Microbiol. 10:105.
doi: 10.3389/fmicb.2019.00105

¹ State Key Laboratory of Biocatalysis and Enzyme Engineering, Environmental Microbial Technology Center of Hubei Province, College of Life Sciences, Hubei University, Wuhan, China, ² Mechanical Engineering and Materials Science, Washington University, St. Louis, MO, United States, ³ State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan, China

Poly-γ-glutamic acid (γ-PGA) is an anionic polymer with various applications. Teichoic acid (TA) is a special component of cell wall in gram-positive bacteria, and its D-alanylation modification can change the net negative charge of cell surface, autolysin activity and cationic binding efficiency, and might further affect metabolic production. In this research, four genes (*dltA*, *dltB*, *dltC*, and *dltD*) of *dlt* operon were, respectively, deleted and overexpressed in the γ-PGA producing strain *Bacillus licheniformis* WX-02. Our results implied that overexpression of these genes could all significantly increase γ-PGA synthetic capabilities, among these strains, the *dltB* overexpression strain WX-02/pHY-*dltB* owned the highest γ-PGA yield (2.54 g/L), which was 93.42% higher than that of the control strain WX-02/pHY300 (1.31 g/L). While, the gene deletion strains produced lower γ-PGA titers. Furthermore, ¹³C-Metabolic flux analysis was conducted to investigate the influence of *dltB* overexpression on metabolic flux redistribution during γ-PGA synthesis. The simulation data demonstrated that fluxes of pentose phosphate pathway and tricarboxylic acid cycle in WX-02/pHY-*dltB* were 36.41 and 19.18 mmol/g DCW/h, increased by 7.82 and 38.38% compared to WX-02/pHY300 (33.77 and 13.86 mmol/g DCW/h), respectively. The synthetic capabilities of ATP and NADPH were also increased slightly. Meanwhile, the fluxes of glycolytic and by-product synthetic pathways were all reduced in WX-02/pHY-*dltB*. All these above phenomena were beneficial for γ-PGA synthesis. Collectively, this study clarified that overexpression of *dltB* strengthened the fluxes of PPP pathway, TCA cycle and energy metabolism for γ-PGA synthesis, and provided an effective strategy for enhanced production of γ-PGA.

Keywords: *Bacillus licheniformis*, poly-γ-glutamic acid, ¹³C-metabolic flux analysis, *dltB*, cell surface negative charge

INTRODUCTION

Poly- γ -glutamic acid (γ -PGA) is an important anionic polypeptide consisting of D-glutamic acid and/or L-glutamic acid residues, which linked together via amide bonds between α -amino and γ -carboxyl (Feng et al., 2015; Sirisansaneeyakul et al., 2017). Since γ -PGA owns the excellent features of biocompatibility, biodegradability, water solubility, edibility, non-toxicity, environmentally friendliness, etc. (Cao et al., 2018), it has the wide-range applications in the areas of food, medicine, cosmetics industries, etc. (Cai et al., 2017).

Bacillus species have been proven as the efficient γ -PGA producers, and a number of metabolic engineering strategies have been developed to improve γ -PGA yield. For example, knocking out glutamate dehydrogenase genes *rocG* and *gudB* improved glutamic acid accumulation, which led to a 38% increase of γ -PGA yield in *Bacillus amyloliquefaciens* (Zhang et al., 2015). The synthesis of extracellular polysaccharide and lipopolysaccharide were blocked to decrease by-product yields for γ -PGA production in *B. amyloliquefaciens* LL3 (Feng et al., 2015). In addition, strengthening of NADPH and ATP supplies all benefited γ -PGA production (Cai et al., 2017, 2018). Recently, cell surface engineering was proven to be an effective strategy for enhancement production of metabolics. Overexpression of phosphatidylserine synthase gene *pssA* could enhance the cell membrane integrity and hydrophilicity, and further improved the cell tolerance and biorenewable yields (short-chain fatty acids, organic alcohols, organic acids and other aromatic compounds, etc) in *E. coli* (Tan et al., 2017). Elevation of membrane cardiolipin levels via overexpressing cardiolipin synthase gene *clsA* significantly increased hyaluronic acid titer by 204% in *Bacillus subtilis* (Westbrook et al., 2018). However, no research has been focused on the relationship between cell surface engineering and γ -PGA synthesis.

Teichoic acids (TAs) is an anionic polymer composed of ribitol and glycerol residues, and it was linked by phosphodiester bonds in the cell wall. The *dlt* operon, which consisted of four genes *dltA*, *dltB*, *dltC*, and *dltD*, owns the function of incorporating D-alanine to TAs in *Bacillus* (Koprivnjak et al., 2006). Previously, removal of D-alanyl esters from TAs of *Staphylococcus aureus* increased the amounts of Mg^{2+} bound to cell wall, which might affect the activities of several membrane proteins or enzymes (Lambert et al., 1975). Previous researches implied that the expression of *dlt* operon could regulate the cell surface net negative charge, and the changes on surface charge could mediate the recruitment of signaling molecules (lignin)

(Hyrylainen et al., 2000; Heit et al., 2011). γ -PGA synthesis is regulated by two component systems ComP~ComA and DegS~DegU, the regulators DegQ and SwrA (Tran et al., 2000). The recruitment of ComP might be affected by the reduction of cell surface negative charge, and further activated these regulators for γ -PGA synthesis. Furthermore, overexpression of *dltA* triggered the electrostatic repulsion between *S. aureus* and daptomycin, which represented the keystone of DAP resistance (Cafiso et al., 2014). γ -PGA and TAs are all anionic polymer, and the electrostatic repulsion generated between γ -PGA and TAs was not conducive to γ -PGA secretion. Thus, reducing cell surface negative charge might be an effective strategy for enhancement production of γ -PGA.

The biochemical reactions in cells were difficult to be profoundly described with the existing techniques. While, acting as a new method, ^{13}C metabolic flux analysis (^{13}C -MFA) is a promising tool to quantify cellular metabolism reaction rates in a network via isotopomer tracer experiments (Long et al., 2016). Moreover, ^{13}C -MFA is a powerful tool to quantify the energy generation and consumption rates in the metabolic pathways (He et al., 2014; Yao et al., 2016), and it also applied to comprehend and analyze the changes in central carbon metabolism under aerobic and anaerobic conditions. For instance, the optimal tracers [1,2- ^{13}C] glucose, [1,6- ^{13}C] glucose, [1,2- ^{13}C] xylose, and [5- ^{13}C] xylose were applied in *Escherichia coli* under aerobic and anaerobic conditions, and their results demonstrated that the fluxes of EMP pathway and TCA cycle were all increased under anaerobic conditions, which generated more energy, formate, alcohol and succinate to adapt the adverse environments (Gonzalez et al., 2017).

In this study, the genes *dltA*, *dltB*, *dltC*, and *dltD* of *dlt* operon were, respectively, deleted and overexpressed in *B. licheniformis* WX-02, and ^{13}C -MFA was performed to expound the effect of *dltB* overexpression on metabolic flux redistribution. In addition, the transcriptional level, by-products contents were also measured during γ -PGA synthesis. The aim of this study is to illustrate the relationship between *dlt* operon overexpression and γ -PGA synthesis by ^{13}C -MFA, and provides an efficient strategy of strain improvement for γ -PGA production.

MATERIALS AND METHODS

Strains and Culture Conditions

Strains and plasmids used in this study were provided in Table 1. *B. licheniformis* WX-02 was acted as the original strain for constructing recombinants, and *E. coli* DH5 α was served as the host strain for plasmid construction. *B. licheniformis* and *E. coli* were grown in LB medium with responsible antibiotic (20 μ g/mL kanamycin, 100 μ g/mL ampicillin, or 20 μ g/mL tetracycline), when required.

The γ -PGA fermentation medium contains 20 g/L glucose, 5 g/L NH_4NO_3 , 11.35 g/L Na_2HPO_4 , 8.15 g/L KH_2PO_4 , 0.20 g/L $MgSO_4 \cdot 7H_2O$, 0.01 g/L $MnSO_4 \cdot H_2O$, 0.0008 g/L $CaCl_2$, and 0.0045 g/L Na_2 -EDTA. *B. licheniformis* were grown in twenty-four well plates with 2 mL working volume, and cultivated at 220 rpm and 37°C for 14 h. For ^{13}C -MFA, glucose was replaced

Abbreviations: 2,3-BDO, 2,3-butanediol; 3PG, 3-phosphoglyceric acid; 6PG, 6-phosphogluconolactone; AceCoA, acetyl-CoA; AKG, α -ketoglutarate; CIT, citrate; DHAP, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; EMP pathway, glycolysis pathway; F6P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; FUM, fumarate; G6P, glucose-6-phosphate; GAP, glyceraldehyde 3-phosphate; Glu, glutamic acid; GLX, glyoxylate; ICIT, isocitrate; MAL, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PPP pathway, pentose phosphate pathway; PYR, pyruvate; R5P, ribose 5-phosphate; Ru5P, ribulose 5-phosphate; S7P, sedoheptulose 7-phosphate; SUC, succinate; SucCoA, succinyl-CoA; TCA cycle, tricarboxylic acid cycle; X5P, xylulose 5-phosphate.

TABLE 1 | The strains and plasmids used in this study.

Strains or plasmids	Description	Source of reference
<i>Bacillus licheniformis</i>		
WX-02	Poly- γ -glutamate producing strain (CCTCC M208065)	CCTCC
WX-02 $\Delta dltA$	WX-02 ($\Delta dltA$)	This study
WX-02 $\Delta dltB$	WX-02 ($\Delta dltB$)	This study
WX-02 $\Delta dltC$	WX-02 ($\Delta dltC$)	This study
WX-02 $\Delta dltD$	WX-02 ($\Delta dltD$)	This study
WX-02/pHY300	WX-02 harboring pHY300PLK	This study
WX-02/pHY- <i>dltA</i>	WX-02 harboring pHY- <i>dltA</i>	This study
WX-02/pHY- <i>dltB</i>	WX-02 harboring pHY- <i>dltB</i>	This study
WX-02/pHY- <i>dltC</i>	WX-02 harboring pHY- <i>dltC</i>	This study
WX-02/pHY- <i>dltD</i>	WX-02 harboring pHY- <i>dltD</i>	This study
Plasmids		
T ₂ (2)-Ori	<i>Bacillus</i> knockout vector; Kan ^r	This study
T ₂ - $\Delta dltA$	T ₂ (2)-Ori - <i>dltA</i> (A+B); to knock out <i>dltA</i>	This study
T ₂ - $\Delta dltB$	T ₂ (2)-Ori - <i>dltB</i> (A+B); to knock out <i>dltB</i>	This study
T ₂ - $\Delta dltC$	T ₂ (2)-Ori - <i>dltC</i> (A+B); to knock out <i>dltC</i>	This study
T ₂ - $\Delta dltD$	T ₂ (2)-Ori - <i>dltD</i> (A+B); to knock out <i>dltD</i>	This study
pHY300PLK	<i>E. coli</i> - <i>B. licheniformis</i> shuttle vector, Ap ^r (<i>E. coli</i>), Tc ^r (<i>E. coli</i> and <i>B. licheniformis</i>)	This study
pHY- <i>dltA</i>	pHY300PLK containing P43 promoter, the gene <i>dltA</i> and <i>amyL</i> terminator	This study
pHY- <i>dltB</i>	pHY300PLK containing P43 promoter, the gene <i>dltB</i> and <i>amyL</i> terminator	This study
pHY- <i>dltC</i>	pHY300PLK containing P43 promoter, the gene <i>dltC</i> and <i>amyL</i> terminator	This study
pHY- <i>dltD</i>	pHY300PLK containing P43 promoter, the gene <i>dltD</i> and <i>amyL</i> terminator	This study

by [1, 2-¹³C] glucose (Sigma-Aldrich, CAS#138079-87-5) in γ -PGA production medium.

Strain Construction

The construction procedures of *dltB* deletion and overexpression strains were served as the examples. For *dltB* deletion strain, the upstream and downstream arms of gene *dltB* were, respectively, amplified by the primers $\Delta dltB$ -F1/R1 and $\Delta dltB$ -F2/R2 (Supplementary Table S1), based on the genomic DNA of *B. licheniformis* WX-02, and fused by Splicing Overlapping Extension PCR (SOE-PCR) with primers $\Delta dltB$ -F1/R2. The fused fragment was inserted into the plasmid T₂(2)-Ori at *SacI/XbaI*, colony PCR and DNA sequence confirmed that the *dltB* deletion vector was constructed successfully, named as T₂- $\Delta dltB$. Then, T₂- $\Delta dltB$ was electro-transferred into *B. licheniformis* WX-02, and the positive transformant was cultivated in LB medium with 20 μ g/mL kanamycin at 220 rpm and 45°C, and sub-cultured for three times to obtain the single-crossover recombinants. The recombinants were grown in LB medium at 37°C with six subcultures, and the kanamycin sensitive colonies were further confirmed by colony PCR and DNA sequence, and the *dltB* deletion strain was named WX-02 $\Delta dltB$.

For gene overexpression strain, P43 promoter from *B. subtilis* 168, gene *dltB* and *amyL* terminator from *B. licheniformis* WX-02 were, respectively, amplified (Supplementary Table S1), and fused by SOE-PCR. The fused fragment was inserted into pHY300PLK at the restriction sites *EcoRI/XbaI*, resulting in the plasmid pHY-*dltB*. Then, pHY-*dltB* was introduced into WX-02 by electro-transformation, resulting in the *dltB* overexpression strain, named WX-02/pHY-*dltB*.

Analytic Methods

The cell biomass was monitored by measuring OD₆₀₀ using a UV-spectrophotometer-752 N (Shanghai Instrument Analysis Instrument Co., Ltd., Shanghai, China), glucose concentrations were determined via using the Enzyme Electrode Analyzer SBA-40E (Shandong Academy of Sciences, Shandong, China). The γ -PGA yield was measured according to our previously reported method (Cai et al., 2018). Briefly, the volume of 2 mL fermentation broth was mixed with 4 mL distilled water, and cells were separated by centrifugation at 12000 rpm for 6 min after adjusting the pH to 2.5~3.0. While, three volumes of absolute ethanol were added into the supernatant after adjusting pH to 7.0, and the precipitate was dried at 80°C to a constant weight for measuring γ -PGA yield. The net negative charge was measured by determining the cation binding rate of cell surface in LB medium, according to the previously described method (Cao et al., 2017; Chen et al., 2018).

The concentrations of by-products, acetoin, 2,3-butanediol and acetic acid, were measured by GC/MS (Gas chromatograph Trace, Thermo; Triple Quadrupole Mass Spectrometer, Thermo; column: DB-WAXMS, 30 m \times 0.25 mm \times 0.25 μ m, Thermo; United States). Briefly, the volume of 0.5 mL fermentation broth was mixed with 1.5 mL absolute ethanol, and the supernatant was separated by centrifugation at 12000 rpm for 10 min. Equal volume of ethyl acetate was added into the supernatant for extracting acetoin, 2,3-butanediol and acetate. The parameters of GC-MS were set as follow: Gas chromatographic inlet temperature was 220°C, the injection volume was 1 μ L, carrier gas flow rate was 1 mL/min and the split ratio was 20:1. The temperature program: hold at 40°C for 1 min, and increased to

160°C at 6°C/min, hold for 1.5 min, and then raised to 220°C at 30°C/min, hold for 2 min. Solvent delay was set as 4 min. The range of mass to charge ratio (m/z) in MS was set between 35 and 500.

The gene transcriptional level were measured with TRIzol® Reagent (Invitrogen, United States) and PrimeScript™ II 1st Strand cDNA Synthesis Kit (TaKaRa, Japan) according to our previously reported method (Cai et al., 2017). The primers used for RT-qPCR were listed in **Supplementary Table S2**, and the housekeeping gene 16S rRNA was used to normalize the gene expression data. The data were averaged and presented as the mean \pm SD.

Mass Isotopomer Distribution Analysis of Amino Acids

To analyze the mass isotopomer distributions (MIDs) of amino acids, [1,2- ^{13}C] glucose was served as the sole carbon source

for cell growth and γ -PGA synthesis. The cell pellets at the mid-logarithmic phase (7 h) were collected and hydrolyzed by 6 M HCl at 100°C for 24 h. The supernatant of hydrolyzate was air-dried, and the precipitate was subsequently derivatized with *N*-tert-butyldimethylsilyl-*N*-methyltrifluoroacetamide (TBDMS). Then, the derivatization products were quantified by GC/MS (Gas chromatograph Trace, Thermo; Triple Quadrupole Mass Spectrometer, Thermo; column: TG-5MS, 30 m \times 0.25 mm \times 0.25 μm , Thermo; United States). The injection volume and injection split ratio were 1 μL and 1:10, respectively, and carrier gas helium was 1.2 mL/min. GC temperature program was set as follows: hold at 150°C for 2 min, and increased to 280°C at 3°C/min, and increased to 300°C at 20°C/min, and then hold for 5 min. Solvent delay was set as 5 min, and the range of mass to charge ratio (m/z) in MS was set between 60 and 500 (You et al., 2012). The software WuFlux-Ms Tool was used to analyze and correct amino acid MS data (fragments of $[\text{M}-57]^+$, $[\text{M}-159]^+$, or $[\text{M}-85]^+$,

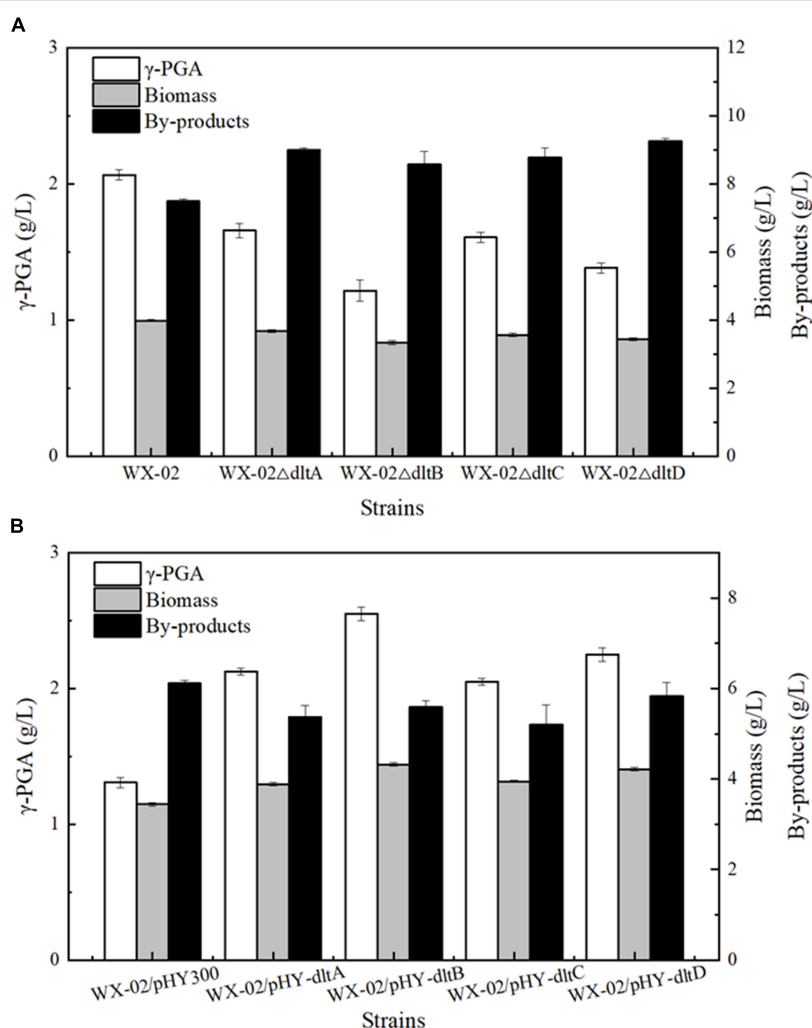


FIGURE 1 | Effects of deletion and overexpression of *dlt* operon on γ -PGA and by-products production. **(A)** The γ -PGA yields, biomass formation and by-products yields of WX-02, WX-02 Δ dltA, WX-02 Δ dltB, WX-02 Δ dltC, and WX-02 Δ dltD. **(B)** The γ -PGA yields, biomass formation and by-products yields of WX-02/pHY300, WX-02/pHY-dltA, WX-02/pHY-dltB, WX-02/pHY-dltC, and WX-02/pHY-dltD.

and [f302]) (Wahl et al., 2004). Isotopomer labeling fractions (M0, M1, M2, etc.) represent fragments containing unlabeled, singly labeled, doubly labeled amino acids, etc (Wan et al., 2017).

¹³C-Metabolic Flux Analysis With INCA

¹³C-MFA was performed by using the INCA software (Young, 2014), which is based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007). The central metabolic network was modified according to the previous research (He et al., 2017) and KEGG pathway database. As the target product and main by-products, several steps for γ -PGA, acetoin and 2,3-butanediol syntheses, such as 2 Pyruvate \rightarrow α -acetolactate + CO₂, α -acetolactate \rightarrow Acetoin + CO₂, Acetoin + NADH \leftrightarrow 2,3-Butanediol + NAD⁺, Acetoin + NAD⁺ \rightarrow Acetyl-CoA + NADH, Acetoin + ATP \rightarrow Acetoin_{ex}, 2,3-Butanediol \rightarrow 2,3-Butanediol_{ex}, Glutamate + ATP \rightarrow γ -PGA, and γ -PGA \rightarrow γ -PGA_{ex}, were added into the metabolic network (seeing in the **Supplementary Table S4**). The information of labeling amino acids (alanine, glycine, valine, leucine, serine, threonine, phenylalanine, aspartic acid, glutamic acid, and histidine) (**Supplementary Table S3**) were substituted into the model for calculating metabolic flux of central metabolism (Wan et al., 2017). The standard deviation of ¹³C-enrichment was set as 0.0107 for statistical analysis (Nanchen et al., 2007), and the flux estimations were performed by using MATLAB R2014a (The Mathworks Inc.).

Statistical Analyses

All samples were analyzed in triplicate, and the data were presented as the mean \pm SD for each sample point. Significant differences were determined by one-way analysis of variance (ANOVA). Statistical significance was defined as $P < 0.05$.

RESULTS

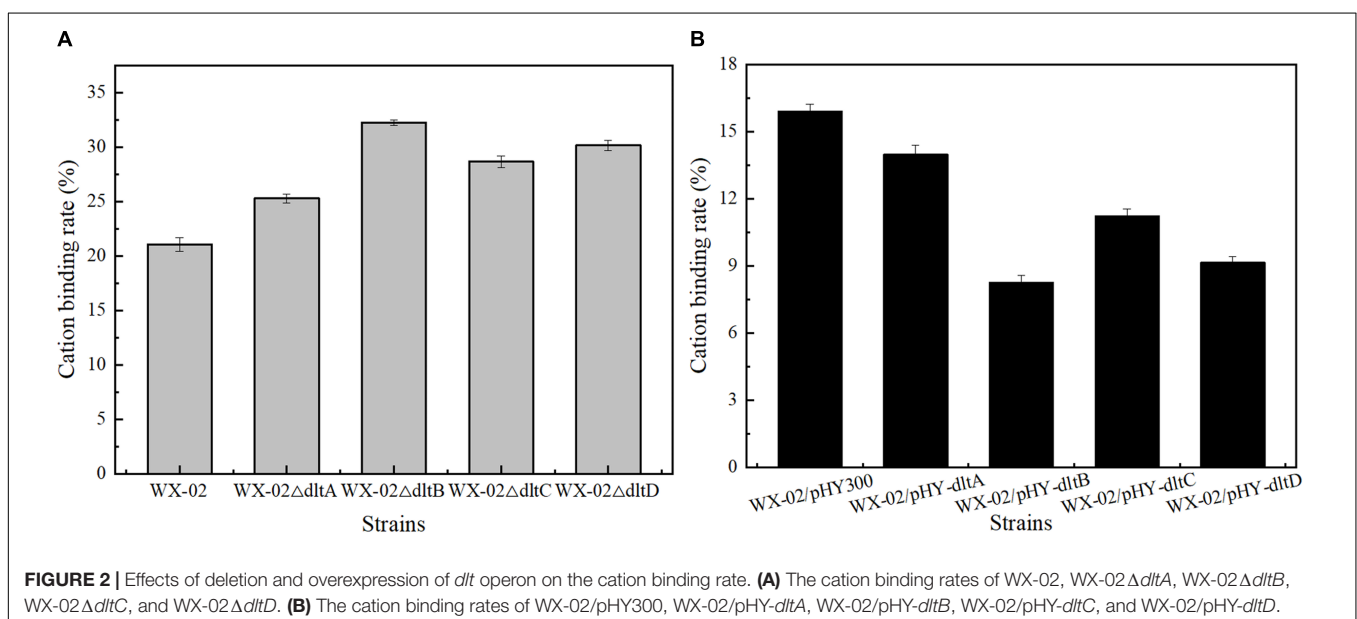
Effects of Deletion and Overexpression of *dlt* Operon on γ -PGA Production

To evaluate the relationship between *dlt* operon expression and γ -PGA synthesis, the genes *dltA*, *dltB*, *dltC* and *dltD* of *dlt* operon were deleted and overexpressed in the γ -PGA production strain *B. licheniformis* WX-02, resulting in the recombinant strains WX-02 Δ *dltA*, WX-02 Δ *dltB*, WX-02 Δ *dltC*, WX-02 Δ *dltD* and WX-02/pHY-*dltA*, WX-02/pHY-*dltB*, WX-02/pHY-*dltC*, WX-02/pHY-*dltD*, respectively. Based on our results of **Figure 1**, deletions of *dltA*, *dltB*, *dltC*, and *dltD* all reduced γ -PGA yield, which decreased by 19.81, 41.55, 22.22, and 33.33% compared with that of control strain WX-02 (2.07 g/L), respectively (**Figure 1A**). And deletion of one gene has no effects on the transcriptional levels of other genes involved in *dlt* operon (**Supplementary Figure S2**). Furthermore, γ -PGA yields produced by the *dlt* overexpression strains were 2.13, 2.55, 2.05, and 2.25 g/L, increased by 62.21, 94.66, 56.49, and 71.76% compared to WX-02/pHY300 (1.31 g/L), respectively (**Figure 1B**). Collectively, these above results indicated that overexpression of *dlt* operon benefited γ -PGA synthesis.

Previously, the *dlt* operon was proved to play the important role on the net negative charge of cell surface. As shown in **Figure 2**, the negative charges of cell surface were all increased in the *dlt* deletion strains, and decreased in the *dlt* overexpression strains. The negative charge was reduced by 48.93% in the *dltB* overexpression strain WX-02/pHY-*dltB*.

Fermentation Process Curves of WX-02/pHY300 and WX-02/pHY-*dltB*

Furthermore, the process curves of WX-02/pHY-300 and WX-02/pHY-*dltB* were determined, and the cell growth, glucose uptake, γ -PGA yield, by-product (acetoin, 2,3-butanediol and



acetic acid) concentrations were measured during γ -PGA production. As shown in **Figure 3A**, the cell biomass of WX-02/pHY-*dltB* were slightly lower than those of control strain WX-02/pHY-300 before 9 h, and then increased faster subsequently. Besides, the logarithmic growth period is delayed by 2 or 3 h, and there is no obvious decline phase in the *dltB* overexpression strain WX-02/pHY-*dltB*. The specific cell growth rate of WX-02/pHY-*dltB* was 0.32 h^{-1} , increased by 7.58% compared with that of WX-02/pHY-300 (**Table 2**). The glucose uptake rate of WX-02/pHY-*dltB* was $7.32 \text{ mmol/g DCW/h}$, similar with that of WX-02/pHY-300 ($7.23 \text{ mmol/g DCW/h}$), and the γ -PGA synthetic rate of WX-02/pHY-*dltB* ($0.92 \text{ mmol/g DCW/h}$) was increased by 41.54% compared to WX-02/pHY-300 ($0.65 \text{ mmol/g DCW/h}$). The concentrations of by-products (acetoin, 2,3-butanediol and acetic acid) was significantly lower in WX-02/pHY-*dltB* than that of WX-02/pHY-300 (**Figure 3B**). Moreover, the formation rates of acetic acid and 2,3-butanediol showed no significant difference between these strains, while the acetoin formation rate was decreased by 50.00% in WX-02/pHY-*dltB* (0.56 mmol/g

DCW/h) (**Table 2**). Assuming that there was no other significant by-products and CO_2 produced, and the carbon recoveries of WX-02/pHY300 and WX-02/pHY-*dltB* were roughly consistent for each strain (**Table 2**).

¹³C-Metabolic Flux Analysis

The metabolic flux distributions were determined based on the measured mass isotopomer distributions of proteinogenic amino acids and synthetic rates of γ -PGA, acetic acid, acetoin and 2,3-butanediol. The measured mass isotopomer distributions (MIDs) of proteinogenic amino acids coincided well with the simulated MIDs (**Supplementary Table S3**), indicating the good fit and high flux precision between the measured and simulated data (Yao et al., 2016). The measured biomass formation rates were not employed as the constraints of ¹³C-MFA model (**Table 2**). The flux values and exchange coefficients were estimated with the ¹³C MFA model listed in **Supplementary Table S4**.

As shown in **Figure 4**, the fluxes in the central metabolism of *B. licheniformis* were redistributed in WX-02/pHY-*dltB*. Firstly, the flux from pyruvate to acetyl-CoA was increased by 11.12%, and the flux distributed into tricarboxylic acid cycle was enhanced by 38.38% in WX-02/pHY-*dltB*. The flux for acetic acid production was increased slightly, whereas, the flux of overflow metabolism (from pyruvate to acetoin and 2,3-butanediol) was decreased by 26.93%. Secondly, the flux of PP pathway was increased by 7.82% in the *dltB* overexpression strain, which might generate more NADPH for γ -PGA synthesis. Thirdly, the flux through pyruvate carboxylation ($\text{PYR} + \text{CO}_2 \rightarrow \text{OAA}$), the major anaplerotic flux into TCA cycle, was increased by 25.40% in WX-02/pHY-*dltB*. The increase of pyruvate carboxylation could enable more carbon flow from glucose to TCA cycle, and further increased the demand for oxaloacetate based biomass synthesis (**Table 2**; He et al., 2014). Fourthly, the fluxes from α -ketoglutaric acid to glutamic acid and γ -PGA were increased by 19.47 and 57.92%, respectively. Fifthly, the flux of EMP pathway was decreased by 4.13%, which further reduced the NADH generation and by-products (acetoin, 2,3-butanediol and acetic acid) syntheses. Lastly, the estimated biomass biosynthesis flux was slightly higher (about 6.40%) in WX-02/pHY-*dltB* (**Figure 4**), which positively correlated with the results of cell biomass (**Table 1**).

The cofactors (such as NADH, NADPH, etc.) and ATP play the important roles in metabolic production (Kind et al., 2013), as well as for γ -PGA synthesis (Cai et al., 2017, 2018). Based on our metabolic flux model in *B. licheniformis*, the NADPH and ATP formation rates in WX-02/pHY-*dltB* were increased by 12.50 and 3.54%, which were beneficial for γ -PGA synthesis (**Table 3**).

Transcriptional Level Analysis

The transcriptional levels of genes related to glucose metabolism and γ -PGA synthesis were measured in WX-02/pHY300 and WX-02/pHY-*dltB* (**Figure 5**). Compared to WX-02/pHY300, the transcription levels of glucose transporter genes *ptsG*, *glcU*, and *glcP* were reduced obviously in *dltB* overexpression strain. The glucose-6-phosphate dehydrogenase gene *zwf* in PP pathway showed higher expression levels, while the transcription level of glucose-6-phosphate isomerase gene *pgi* in EMP pathway

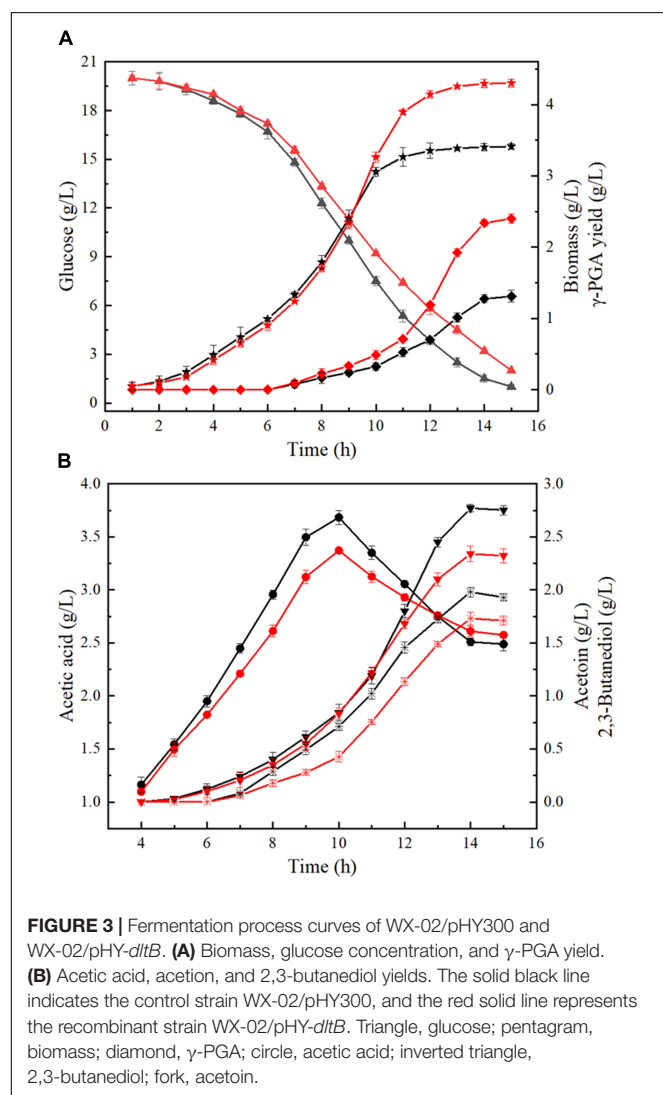


TABLE 2 | The fermentation characteristics of WX-02/pHY300 and WX-02/pHY-*dltB*.

Strains	Specific growth rate μ (h ⁻¹)	Glucose uptake rate (mmol/g/h)	γ -PGA formation rate (mmol/g/h)	Acetate formation rate (mmol/g/h)	Acetoin formation rate (mmol/g/h)	2,3-Butanediol formation rate (mmol/g/h)	carbon recovery (mmol/g/h)
WX-02/pHY300	0.29	7.23	0.65	3.48	1.12	1.14	31.04
WX-02/pHY- <i>dltB</i>	0.32	7.32	0.92	3.57	0.56	1.10	31.39

The carbon recoveries of WX-02/pHY300 and WX-02/pHY-*dltB* were calculated as follows:
In control strain WX-02/pHY300:
Carbon consumption of glucose: 7.23*6 = 43.38 mM/g/h.
Carbon consumption for biomass and product synthesis:
0.65*5+3.48*2+1.12*4+1.14*4+0.29/24.6*1000 = 31.04 mM/g/h.
In engineered strain WX-02/pHY-*dltB*:
Carbon consumption of glucose: 7.32*6 = 43.92 mM/g/h.
Carbon consumption for biomass and product synthesis:
0.92*5+3.57*2+0.56*4+1.10*4+0.32/24.6*1000 = 31.39 mM/g/h.
(Biomass was calculated by C₇H_{1.8}O_{0.5}N_{0.2}. The calculation of carbon recovery is products formation rate * the number of carbon).

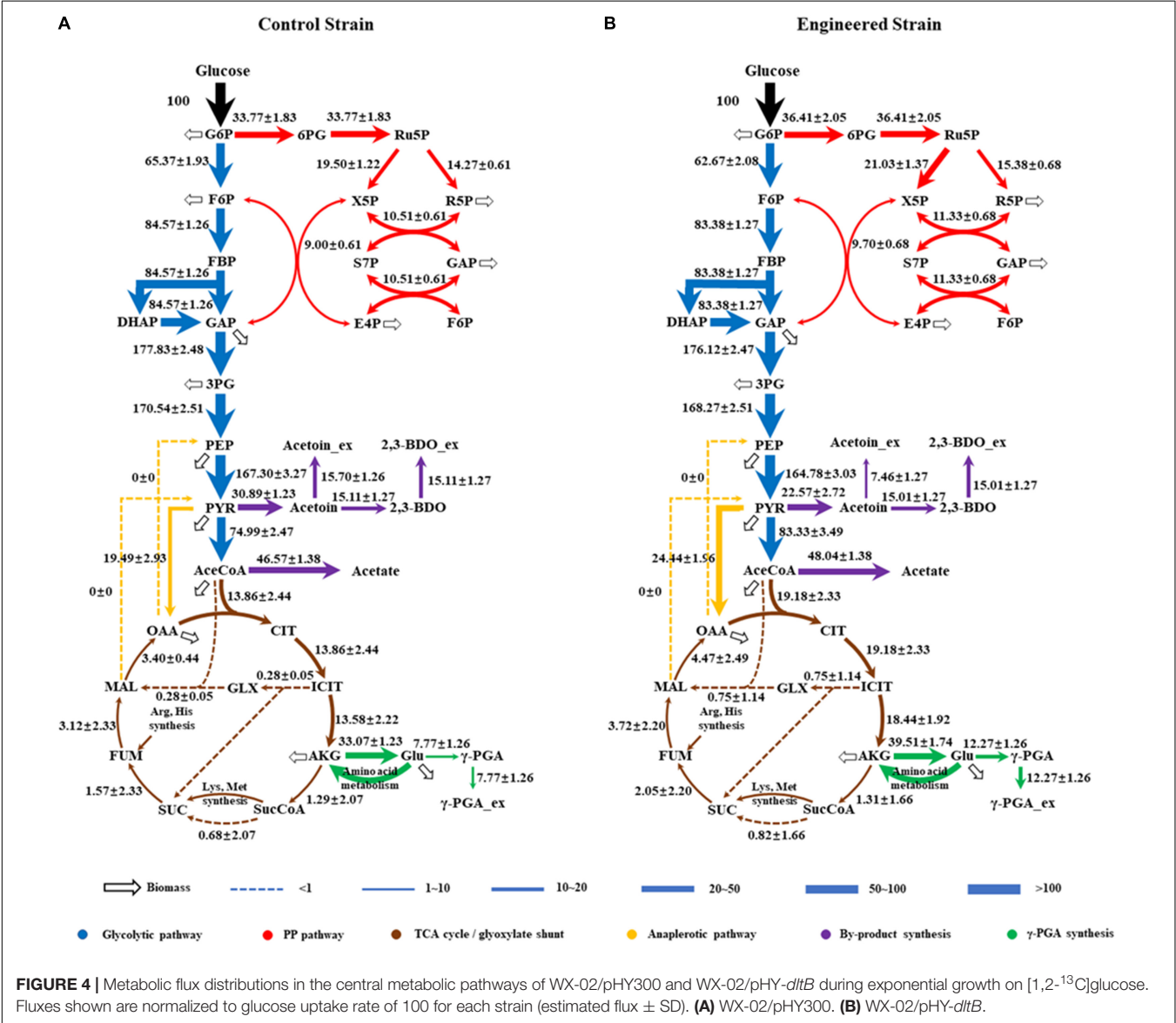


TABLE 3 | The synthesis and consumption contents of NADPH, NADH, ATP, and FADH2 of WX-02/pHY300 and WX-02/pHY-*dltB* (mmol/g CDW/h).

Pathway	NADPH		NADH		FADH2		ATP	
	Control	Engineered	Control	Engineered	Control	Engineered	Control	Engineered
Glycolysis	0	0	177.83	176.12	0	0	160.56	157.52
PP pathway	67.54	72.82	0	0	0	0	0	0
TCA cycle	13.58	18.44	79.68	89.11	1.57	2.05	0.68	0.82
Amino acid synthesis	-58.67	-67.08	3.79	4.08	0	0	-23.36	-25.11
γ -PGA formation	0	0	0	0	0	0	-7.77	-12.27
Acetic acid formation	0	0	0	0	0	0	46.57	48.04
Acetoin and 2,3-butanediol formation	0	0	-15.02	-14.91	0	0	-15.70	-7.46
Biomass formation	-22.47	-24.19	6.10	6.56	0	0	-139.32	-149.96
Oxidative phosphorylation	0	0	-253.93	-262.64	-1.57	-2.05	764.93	792.02

All the flux values are normalized to the glucose uptake rate of 100 mmol h⁻¹ for each strain.

was decreased by 25.01%, indicating the elevated PP pathway and depressed EMP pathway fluxes in WX-02/pHY-*dltB*. In addition, the transcription levels of genes *citB* (encoding citrate synthase) and *icd* (encoding isocitrate dehydrogenase) in TCA cycle were enhanced by 90.03 and 105.12%, which strengthened the metabolic flux in TCA cycle.

Furthermore, the transcription levels of glutamate dehydrogenase gene *rocG*, which mainly catalyzes the formation of glutamic acid from α -ketoglutaric acid in *B. licheniformis* WX-02 (Tian et al., 2017), was improved by 83.25% in WX-02/pHY-*dltB*, and γ -PGA synthetase genes *pgsB*, *pgsC* were, respectively, increased by 2.62- and 2.12-fold, which indicated that more glutamic acid was synthesized and converted to γ -PGA in WX-02/pHY-*dltB*. Besides, the gene *pgdS* (encoding γ -DL-glutamyl hydrolase) was also increased in WX-02/pHY-*dltB*, and the improvement of *pgdS* could decrease the molecular weight of γ -PGA (Tian et al., 2017), positively correlated with the actual measurement results (Date not shown). Moreover, transcription factor genes *comA*, *comP*, *degS*, and *degU*, which associated with the synthesis of γ -PGA, were all increased in the *dltB* overexpression strain.

DISCUSSION

Cell wall and cell membrane are the selective permeation barrier of bacteria, and engineering of cell surface negative charge, membrane phospholipid head and fatty acid hydrophobic tail composition could affect bacterial growth and product production (Hyyrylainen et al., 2007; Ghorbal et al., 2013; Tan et al., 2017). Previous researches demonstrated that the increases of negative charge on the cell surface could improve the secretion efficiency of target protein, and the increase rate of target protein with lower isoelectric points (PI) was higher than that of protein with higher PI (Cao et al., 2017; Chen et al., 2018). While, based on our results, the increases of cell surface negative charge led to the decrease of γ -PGA yield in the *dlt* operon mutants.

On the one hand, overexpression of *dltB* slowed down the absorption and utilization rate of glucose (Figure 3A), and further decreased the flux of EMP pathway. Previous research implied that reduction of glycolysis flux in *E. coli* THRD

could reduce the accumulation of acetate and increase NADPH and L-threonine generations (Xie et al., 2014), which was consistent with our results (Figure 3A and Table 2). Moreover, the alanylation of teichoic acids could modulate the negative charge of cell wall to protect secretory or cell wall-associated proteins against degradation during the post-translocational folding (Hyyrylainen et al., 2000), and enhancement of cell surface negative charge could increase the binding capacity of autolysin, and accelerated autolysis of bacteria (Steen et al., 2005). Figure 3A showed that the logarithmic period of WX-02/pHY-*dltB* was extended by 2~3 h, and the growing status of WX-02/ Δ *dltB* was obviously badness in post-fermentation compared with WX-02 (Supplementary Figure S1). This results indicated that overexpression of *dltB* could promote the cell growth by reducing the autolysis of cells. Therefore, overexpression of *dltB* reduced the absorption and utilization rate of glucose, which decreased the flux of EMP pathway and synthesis capabilities of by-products (acetoin, 2,3-butanediol and acetic acid), and extended logarithmic growth phase, and all these phenomenons were beneficial for γ -PGA synthesis.

Then, ¹³C-labeled isotope tracer and ¹³C-MFA were applied to further evaluate the metabolic flux redistributions in the *dltB* overexpression strain. Based on our results, the flux of EMP pathway is weakened, which reduced the NADH generation and by-products synthesis, and our results were positively consisted with the previous research (Lee and Oh, 2016). Secondly, the flux of PPP pathway and NADPH supply were all enhanced in the *dltB* overexpression strain. Since the conversion of α -ketoglutarate to glutamic acid requires NADPH in *B. licheniformis* WX-02, enhancement of NADPH supply contributes to γ -PGA synthesis (Cai et al., 2017; Tian et al., 2017). In addition, the flux of TCA cycle was enhanced by 38.38% in the *dltB* overexpression strain, and the increase of α -ketoglutaric acid synthesis flux was beneficial for cell growth (Wan et al., 2017) and γ -PGA production. Besides, the flux of complement-deficient pathway (PYR \rightarrow OAA) was enhanced by 25.40%, which is beneficial for oxaloacetate accumulation and cell growth (Sauer et al., 1999). Finally, the flux from α -ketoglutarate to glutamate and further generate γ -PGA were, respectively, increased by 19.47 and 57.92%, which were consistent with fermentation characters (Figure 1).

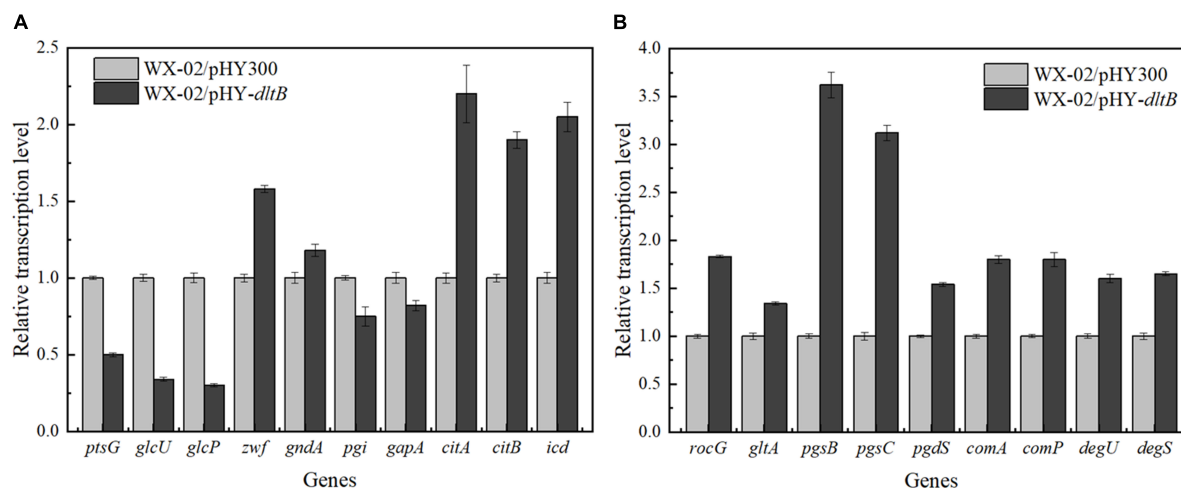


FIGURE 5 | Transcription levels of genes in glucose metabolism and γ -PGA synthesis in WX-02/pHY300 and WX-02/pHY-*dltB*. **(A)** Genes in glucose metabolism. **(B)** Genes in γ -PGA synthesis.

Furthermore, overexpression of *dlt* operon led to the decrease of cell surface negative charge (Figure 2), and further changed the extracellular microenvironment (Hyyrylainen et al., 2000), which might promote the recruitment of signal molecular ComP to activate two component system ComP-ComA for γ -PGA synthesis. Based on our results, the transcriptional levels of *comP/comA* and *degU/degS* were all increased in the *dltB* overexpression strain, which further affected the transcription levels of genes involved in γ -PGA synthesis and carbon metabolism (Figure 5). Moreover, the decrease of cell surface negative charge could reduce the binding of cationic (e.g., Mg^{2+} and Ca^{2+} , etc.) and cationic antimicrobial peptides on cell surface (Neuhaus and Baddiley, 2003), and the low levels of cationic or cationic antimicrobial peptides could activate PhoQ(R)~PhoP system (Garcia Vescovi et al., 1996; Ren et al., 2017). In this study, the concentration of phosphate in the culture medium was too high to activate PhoQ(R)~PhoP system (Devine, 2018) in WX-02/pHY300, however, due to the low level of Mg^{2+} and Ca^{2+} on cell surface of *dltB* overexpression strain, the PhoQ(R)~PhoP might be activated in WX-02/pHY-*dltB*, which was consistent with our unpublished results that strengthening PhoQ(R)~PhoP system benefits γ -PGA synthesis. Finally, γ -PGA is an anionic polypeptide, and the decrease of cell surface negative charge might reduce the strength of electrostatic repulsion (Cafiso et al., 2014) between TAs and γ -PGA, which further benefited γ -PGA secretion. To test this hypothesis, the negatively charged products (such as lichenysin, bacitracin, pulcherrimin, et al.) were analyzed, and our results showed that the yields of those products with negative charge were all increased in *dlt* overexpression strains (Supplementary Figure S3).

CONCLUSION

Cell surface engineering is a promising tactic for enhanced production of metabolites. This study demonstrated that

overexpression of *dlt* operon could increase γ -PGA production for the first time. Furthermore, ^{13}C -MFA was applied to illustrate the affect of *dltB* overexpression on γ -PGA synthesis. Our results showed that overexpression of *dltB* could reduce the negative charge of cell surface, which further reduced the absorption and utilization rate of glucose and the flux of EMP pathway. Meanwhile, the increases of TCA cycle flux, NADPH and ATP supplies, glutamic acid formation benefited γ -PGA synthesis. This work demonstrated that ^{13}C -MFA provided the distinct metabolic insights in engineered microbes for the changes in central carbon metabolism.

AUTHOR CONTRIBUTIONS

PH and SC designed the study. PH, SH, and YC carried out the molecular biology studies and construction of recombinant strains. PH, DC, SH, and YC carried out the fermentation studies. PH and NW carried out the metabolic flux analysis. PH, DC, SL, and SC analyzed the data and wrote the manuscript. All authors read and approved the final manuscript.

FUNDING

This work was supported by the National Program on Key Basic Research Project (973 Program, No. 2015CB150505), the Technical Innovation Special Fund of Hubei Province (2018ACA149), and the China Postdoctoral Science Foundation (2018M642802).

ACKNOWLEDGMENTS

We thank to Lian He from University of Washington for his help on ^{13}C -metabolic flow analysis.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2019.00105/full#supplementary-material>

REFERENCES

- Antonieiwicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2007). Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* 9, 68–86. doi: 10.1016/j.ymben.2006.09.001
- Cafiso, V., Bertuccio, T., Purrello, S., Campanile, F., Mammina, C., Sartor, A., et al. (2014). *dltA* overexpression: a strain-independent keystone of daptomycin resistance in methicillin-resistant *Staphylococcus aureus*. *Int. J. Antimicrob. Agents* 43, 26–31. doi: 10.1016/j.ijantimicag.2013.10.001
- Cai, D., Chen, Y., He, P., Wang, S., Mo, F., Li, X., et al. (2018). Enhanced production of poly- γ -glutamic acid by improving ATP supply in metabolically engineered *Bacillus licheniformis*. *Biotechnol. Bioeng.* 115, 2541–2553. doi: 10.1002/bit.26774
- Cai, D., He, P., Lu, X., Zhu, C., Zhu, J., Zhan, Y., et al. (2017). A novel approach to improve poly- γ -glutamic acid production by NADPH regeneration in *Bacillus licheniformis* WX-02. *Sci. Rep.* 7:43404. doi: 10.1038/srep43404
- Cao, H., Van Heel, A. J., Ahmed, H., Mols, M., and Kuipers, O. P. (2017). Cell surface engineering of *Bacillus subtilis* improves production yields of heterologously expressed α -amylases. *Microb. Cell Fact.* 16:56. doi: 10.1186/s12934-017-0674-0
- Cao, M., Feng, J., Sirisansaneeyakul, S., Song, C., and Chisti, Y. (2018). Genetic and metabolic engineering for microbial production of poly- γ -glutamic acid. *Biotechnol. Adv.* 36, 1424–1433. doi: 10.1016/j.biotechadv.2018.05.006
- Chen, Y., Cai, D., He, P., Mo, F., Zhang, Q., Ma, X., et al. (2018). Enhanced production of heterologous proteins by *Bacillus licheniformis* with defective D-alanylation of lipoteichoic acid. *World J. Microb. Biotechnol.* 34:135. doi: 10.1007/s11274-018-2520-x
- Devine, K. M. (2018). Activation of the PhoPR-mediated response to phosphate limitation is regulated by wall teichoic acid metabolism in *Bacillus subtilis*. *Front. Microbiol.* 9:2678. doi: 10.3389/fmicb.2018.02678
- Feng, J., Gu, Y., Quan, Y., Cao, M., Gao, W., Zhang, W., et al. (2015). Improved poly- γ -glutamic acid production in *Bacillus amyloliquefaciens* by modular pathway engineering. *Metab. Eng.* 32, 106–115. doi: 10.1016/j.ymben.2015.09.011
- Garcia Vescovi, E., Soncini, F. C., and Groisman, E. A. (1996). Mg^{2+} as an extracellular signal: environmental regulation of *Salmonella* virulence. *Cell* 84, 165–174. doi: 10.1016/S0092-8674(00)81003-X
- Ghorbal, S. K., Chatti, A., Sethom, M. M., Maalej, L., Mihoub, M., Kefacha, S., et al. (2013). Changes in membrane fatty acid composition of *Pseudomonas aeruginosa* in response to UV-C radiations. *Curr. Microbiol.* 67, 112–117. doi: 10.1007/s00284-013-0342-5
- Gonzalez, J. E., Long, C. P., and Antoniewicz, M. R. (2017). Comprehensive analysis of glucose and xylose metabolism in *Escherichia coli* under aerobic and anaerobic conditions by (^{13}C) metabolic flux analysis. *Metab. Eng.* 39, 9–18. doi: 10.1016/j.ymben.2016.11.003
- He, L., Xiao, Y., Gebreselassie, N., Zhang, F., Antoniewicz, M. R., Tang, Y. J., et al. (2014). Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on (^{13}C) -metabolic flux analysis. *Biotechnol. Bioeng.* 111, 575–585. doi: 10.1002/bit.25124
- He, L., Xiu, Y., Jones, J. A., Baidoo, E. E. K., Keasling, J. D., Tang, Y. J., et al. (2017). Deciphering flux adjustments of engineered *E. coli* cells during fermentation with changing growth conditions. *Metab. Eng.* 39, 247–256. doi: 10.1016/j.ymben.2016.12.008
- Heit, B., Yeung, T., and Grinstein, S. (2011). Changes in mitochondrial surface charge mediate recruitment of signaling molecules during apoptosis. *Am. J. Physiol. Cell Physiol.* 300, C33–C41. doi: 10.1152/ajpcell.00139.2010
- Hyrylainen, H. L., Pietiainen, M., Lunden, T., Ekman, A., Gardemeister, M., Murtomaki-Repo, S., et al. (2007). The density of negative charge in the cell wall influences two-component signal transduction in *Bacillus subtilis*. *Microbiology* 153, 2126–2136. doi: 10.1099/mic.0.2007/008680-0
- Hyrylainen, H. L., Vitikainen, M., Thwaite, J., Wu, H., Sarvas, M., Harwood, C. R., et al. (2000). D-Alanine substitution of teichoic acids as a modulator of protein folding and stability at the cytoplasmic membrane/cell wall interface of *Bacillus subtilis*. *J. Biol. Chem.* 275, 26696–26703. doi: 10.1074/jbc.M003804200
- Kind, S., Becker, J., and Wittmann, C. (2013). Increased lysine production by flux coupling of the tricarboxylic acid cycle and the lysine biosynthetic pathway—metabolic engineering of the availability of succinyl-CoA in *Corynebacterium glutamicum*. *Metab. Eng.* 15, 184–195. doi: 10.1016/j.ymben.2012.07.005
- Koprivnjak, T., Mlakar, V., Swanson, L., Fournier, B., Peschel, A., and Weiss, J. P. (2006). Cation-induced transcriptional regulation of the *dlt* operon of *Staphylococcus aureus*. *J. Bacteriol.* 188, 3622–3630. doi: 10.1128/JB.188.10.3622-3630.2006
- Lambert, P. A., Hancock, I. C., and Baddiley, J. (1975). Influence of alanyl ester residues on the binding of magnesium ions to teichoic acids. *Biochem. J.* 151, 671–676. doi: 10.1042/bj1510671
- Lee, S. W., and Oh, M. K. (2016). Improved production of N-acetylglucosamine in *Saccharomyces cerevisiae* by reducing glycolytic flux. *Biotechnol. Bioeng.* 113, 2524–2528. doi: 10.1002/bit.26014
- Long, C. P., Au, J., Gonzalez, J. E., and Antoniewicz, M. R. (2016). (^{13}C) metabolic flux analysis of microbial and mammalian systems is enhanced with GC-MS measurements of glycogen and RNA labeling. *Metab. Eng.* 38, 65–72. doi: 10.1016/j.ymben.2016.06.007
- Nanchen, A., Fuhrer, T., and Sauer, U. (2007). Determination of metabolic flux ratios from (^{13}C) -experiments and gas chromatography-mass spectrometry data: protocol and principles. *Methods Mol. Biol.* 358, 177–197. doi: 10.1007/978-1-59745-244-1_11
- Neuhaus, F. C., and Baddiley, J. (2003). A continuum of anionic charge: structures and functions of D-alanyl-teichoic acids in gram-positive bacteria. *Microbiol. Mol. Biol. Rev.* 67, 686–723. doi: 10.1128/MMBR.67.4.686-723.2003
- Ren, J., Sang, Y., Lu, J., and Yao, Y. F. (2017). Protein acetylation and its role in bacterial virulence. *Trends Microbiol.* 25, 768–779. doi: 10.1016/j.tim.2017.04.001
- Sauer, U., Lasko, D. R., Fiaux, J., Hochuli, M., Glaser, R., Szyperski, T., et al. (1999). Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *J. Bacteriol.* 181, 6679–6688.
- Sirisansaneeyakul, S., Cao, M., Kongklom, N., Chuensangjun, C., Shi, Z., and Chisti, Y. (2017). Microbial production of poly- γ -glutamic acid. *World J. Microbiol. Biotechnol.* 33:173. doi: 10.1007/s11274-017-2338-y
- Steen, A., Palumbo, E., Deghorain, M., Cocconcelli, P. S., Delcour, J., Kuipers, O. P., et al. (2005). Autolysis of *Lactococcus lactis* is increased upon D-alanine depletion of peptidoglycan and lipoteichoic acids. *J. Bacteriol.* 187, 114–124. doi: 10.1128/JB.187.1.114-124.2005
- Tan, Z., Khakbaz, P., Chen, Y., Lombardo, J., Yoon, J. M., Shanks, J. V., et al. (2017). Engineering *Escherichia coli* membrane phospholipid head distribution improves tolerance and production of biorenewables. *Metab. Eng.* 44, 1–12. doi: 10.1016/j.ymben.2017.08.006
- Tian, G., Wang, Q., Wei, X., Ma, X., and Chen, S. (2017). Glutamate dehydrogenase (RocG) in *Bacillus licheniformis* WX-02: enzymatic properties and specific functions in glutamic acid synthesis for poly- γ -glutamic acid production. *Enzyme Microb. Technol.* 99, 9–15. doi: 10.1016/j.enzmictec.2017.01.002
- Tran, L. S., Nagai, T., and Itoh, Y. (2000). Divergent structure of the ComQXPA quorum-sensing components: molecular basis of strain-specific communication mechanism in *Bacillus subtilis*. *Mol. Microbiol.* 37, 1159–1171. doi: 10.1046/j.1365-2958.2000.02069.x
- Wan, N., Delorenzo, D. M., He, L., You, L., Immethun, C. M., Wang, G., et al. (2017). Cyanobacterial carbon metabolism: fluxome plasticity and oxygen dependence. *Biotechnol. Bioeng.* 114, 1593–1602. doi: 10.1002/bit.26287

- Wahl, S. A., Dauner, M., and 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. doi: 10.1002/bit.10909
- Westbrook, A. W., Ren, X., Moo-Young, M., and Chou, C. P. (2018). Engineering of cell membrane to enhance heterologous production of hyaluronic acid in *Bacillus subtilis*. *Biotechnol. Bioeng.* 115, 216–231. doi: 10.1002/bit.26459
- Xie, X., Liang, Y., Liu, H., Liu, Y., Xu, Q., Zhang, C., et al. (2014). Modification of glycolysis and its effect on the production of L-threonine in *Escherichia coli*. *J. Ind. Microbiol. Biotechnol.* 41, 1007–1015. doi: 10.1007/s10295-014-1436-1
- Yao, R., Xiong, D., Hu, H., Wakayama, M., Yu, W., Zhang, X., et al. (2016). Elucidation of the co-metabolism of glycerol and glucose in *Escherichia coli* by genetic engineering, transcription profiling, and (^{13}C) metabolic flux analysis. *Biotechnol. Biofuels* 9:175. doi: 10.1186/s13068-016-0591-1
- You, L., Page, L., Feng, X., Berla, B., Pakrasi, H. B., and Tang, Y. J. (2012). Metabolic pathway confirmation and discovery through (^{13}C) -labeling of proteinogenic amino acids. *J. Vis. Exp.* 59:e3583. doi: 10.3791/3583
- Young, J. D. (2014). INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* 30, 1333–1335. doi: 10.1093/bioinformatics/btu015
- Zhang, W., He, Y., Gao, W., Feng, J., Cao, M., Yang, C., et al. (2015). Deletion of genes involved in glutamate metabolism to improve poly-gamma-glutamic acid production in *B. amyloliquefaciens* LL3. *J. Ind. Microbiol. Biotechnol.* 42, 297–305. doi: 10.1007/s10295-014-1563-8

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

Copyright © 2019 He, Wan, Cai, Hu, Chen, Li and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Tandem Mass Spectrometry for ^{13}C Metabolic Flux Analysis: Methods and Algorithms Based on EMU Framework

Jungik Choi and Maciek R. Antoniewicz*

Department of Chemical and Biomolecular Engineering, Metabolic Engineering and Systems Biology Laboratory, University of Delaware, Newark, DE, United States

OPEN ACCESS

Edited by:

Yinjie Tang,
Washington University in St. Louis,
United States

Reviewed by:

Lifeng Peng,
Victoria University of Wellington,
New Zealand

Chao Wu,
National Renewable Energy
Laboratory (DOE), United States

*Correspondence:

Maciek R. Antoniewicz
mraanton@udel.edu

Specialty section:

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

Received: 30 October 2018

Accepted: 09 January 2019

Published: 24 January 2019

Citation:

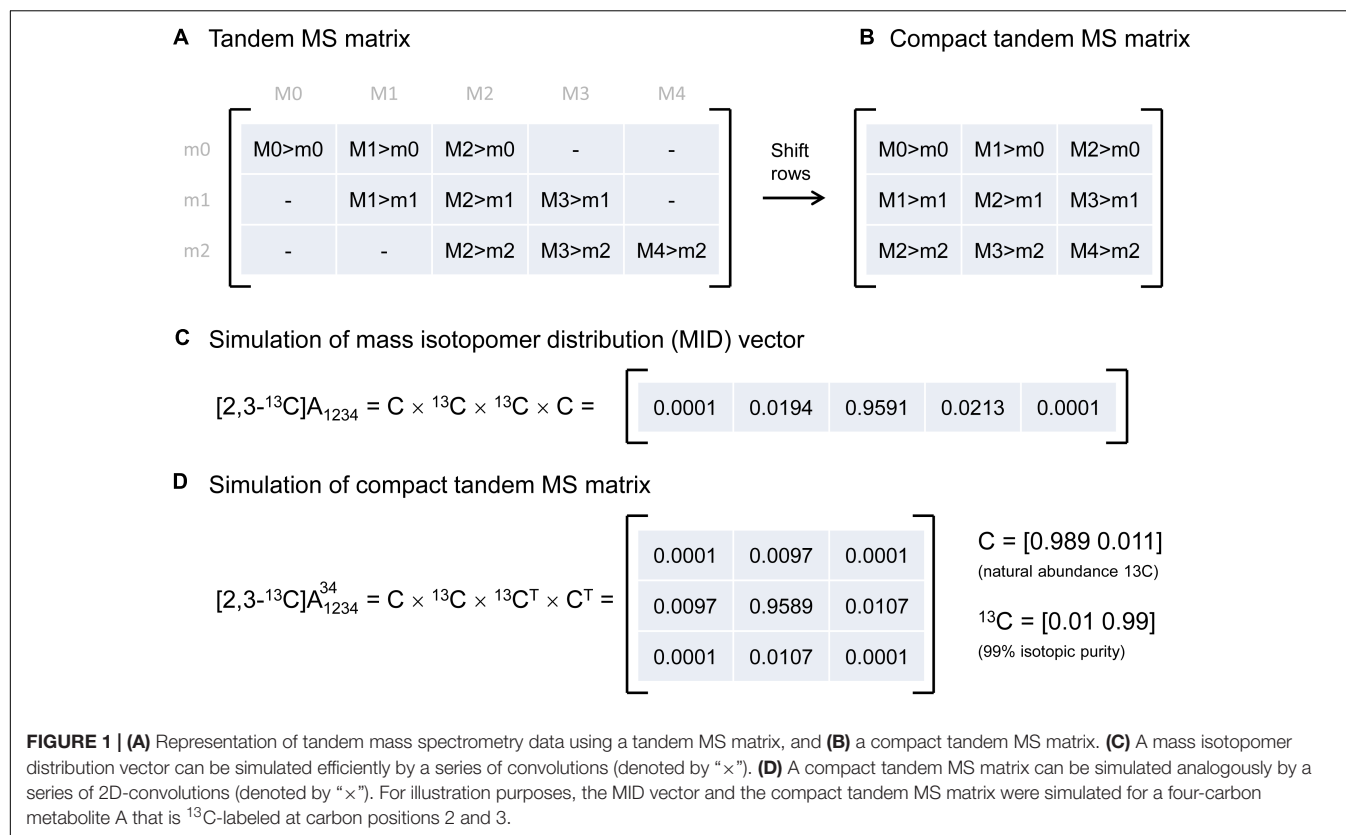
Choi J and Antoniewicz MR (2019)
Tandem Mass Spectrometry for ^{13}C
Metabolic Flux Analysis: Methods
and Algorithms Based on EMU
Framework. *Front. Microbiol.* 10:31.
doi: 10.3389/fmicb.2019.00031

In the past two decades, ^{13}C metabolic flux analysis (^{13}C -MFA) has matured into a powerful and widely used scientific tool in metabolic engineering and systems biology. Traditionally, metabolic fluxes have been determined from measurements of isotopic labeling by means of mass spectrometry (MS) or nuclear magnetic resonance (NMR). In recent years, tandem MS has emerged as a new analytical technique that can provide additional information for high-resolution quantification of metabolic fluxes in complex biological systems. In this paper, we present recent advances in methods and algorithms for incorporating tandem MS measurements into existing ^{13}C -MFA approaches that are based on the elementary metabolite units (EMU) framework. Specifically, efficient EMU-based algorithms are presented for simulating tandem MS data, tracing isotopic labeling in biochemical network models and for correcting tandem MS data for natural isotope abundances.

Keywords: elementary metabolite units, metabolic flux analysis, tandem mass spectrometry, stable isotope tracers, metabolism

INTRODUCTION

^{13}C -Metabolic flux analysis (^{13}C -MFA) is a widely used technique in metabolic engineering and biomedical sciences for quantifying rates of metabolite interconversions inside living cells (i.e., *in vivo* metabolic fluxes) (Antoniewicz, 2015a,b; Ahn et al., 2016). In ^{13}C -MFA, labeling experiment are performed by introducing a ^{13}C labeled substrate (the tracer), followed by measurement of ^{13}C labeling incorporation, and calculation of metabolic fluxes using one of several available software packages for ^{13}C -MFA (Yoo et al., 2008; Quek et al., 2009; Weitzel et al., 2013; Young, 2014).



Currently, the main techniques used to measure ^{13}C labeling are mass spectrometry (MS) (Antoniewicz et al., 2007a, 2011; McConnell and Antoniewicz, 2016), nuclear magnetic resonance (NMR) spectroscopy (Masakapalli et al., 2014), and tandem MS (Antoniewicz, 2013; Okahashi et al., 2016). Previous studies have demonstrated that tandem MS measurements can provide more labeling information than MS resulting in improved performance of ^{13}C -MFA in complex biological systems (Jeffrey et al., 2002; Choi and Antoniewicz, 2011; Choi et al., 2012).

Various modeling approaches have been applied to simulate tandem MS data for applications in ^{13}C -MFA. These include manually derived algebraic equations for specific network models (Jeffrey et al., 2002), isotopomers (Choi and Antoniewicz, 2011; Chandra and Peng, 2012), and tandemers (Tepper and Shlomi, 2015). Despite the demonstrated advantages of tandem MS, the technique is still not fully embraced by the ^{13}C -MFA community, likely because the modeling approaches have not been based on the elementary metabolite units (EMU) framework (Antoniewicz et al., 2007b; Crown and Antoniewicz, 2012), which is the most widely used approach for modeling isotopic labeling in ^{13}C -MFA (Young et al., 2008). The EMU framework is at the core of all major software packages for ^{13}C -MFA (Yoo et al., 2008; Quek et al., 2009; Weitzel et al., 2013; Young, 2014). In this methods paper, we describe efficient algorithms for modeling tandem MS data that are firmly based on the EMU framework. By building upon the EMU framework, we illustrate that the

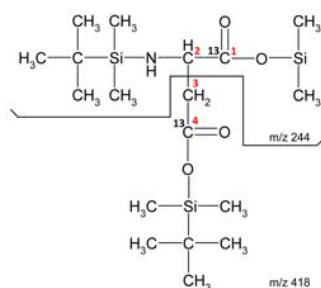
presented algorithms can be easily incorporated into existing software packages to take full advantage of the additional information provided by tandem MS for high-resolution flux measurements.

SIMULATION OF TANDEM MS DATA

Compact Tandem MS Matrix

First, we introduce here the new concept of the compact tandem MS matrix that will be used throughout this paper to describe tandem MS data. In later sections, it will be demonstrated that the compact tandem MS matrix is compatible with the EMU framework to allow tandem MS data to be incorporated into existing EMU algorithms for ^{13}C -MFA.

Tandem MS data was previously represented by the so-called tandem MS matrix (Figure 1A; Choi and Antoniewicz, 2011), where the columns correspond to m/z of the parent fragment and rows correspond to m/z of the daughter fragment. As an example, consider a metabolite A that has four carbon atoms, and assume that carbon atoms C1–C4 are present in the parent fragment, and that carbon atoms C3–C4 are present in the daughter fragment. In this case, the tandem MS matrix is a 3×5 matrix (rows m0–m2, columns M0–M4; Figure 1A). A disadvantage of this representation is that several matrix fields will have zero values by definition. For example, in the first column of the tandem MS matrix,

A Daughter fragment and complement fragment

P = core C-atoms in daughter fragment (C_2)

Q = other atoms in daughter fragment ($\text{C}_8\text{H}_{22}\text{O}_2\text{NSi}_2$)

R = core C-atoms in complement fragment (C_2)

S = other atoms in complement fragment ($\text{C}_6\text{H}_{18}\text{O}_2\text{Si}$)

B Simulation of compact tandem MS matrix

$$\text{Asp}_{1234}^{12} \text{ (core C-atoms)} = \mathbf{R} \times \mathbf{P}^T$$

$$\text{Asp}_{m/z244, m/z418}^{m/z244} = \text{Asp}_{1234}^{12} \text{ (core C-atoms)} \times \mathbf{S} \times \mathbf{Q}^T$$

C Correction for natural isotope abundances

$$\text{Asp}_{1234}^{12} \text{ (core C-atoms)} = \text{Measured (Asp}_{m/z244, m/z418}^{m/z244}) \div (\mathbf{S} \times \mathbf{Q}^T)$$

FIGURE 2 | (A) Definitions of daughter fragment and complement fragment; here, illustrated for TBDMS-derivatized aspartate, where the daughter fragment contains carbon atoms C1 and C2 of aspartate, and the complement fragment contains carbon atoms C3 and C4 of aspartate. **(B)** The compact tandem MS matrix of TBDMS-derivatized aspartate is simulated using 2D-convolutions (denoted by “ \times ”), where Q and S are the natural abundance MID vectors for the other atoms (i.e., non-core C-atoms) of the daughter and complement fragments, respectively. **(C)** Correction for natural isotope abundances of non-core C-atoms is accomplished by a 2D-deconvolution (denoted here by “ \div ”).

which corresponds to the unlabeled parent fragment (M0), the only possible daughter fragment is unlabeled (m0). A more convenient way of representing tandem MS data is by defining the compact tandem MS matrix, which is constructed by shifting rows of the tandem MS matrix to eliminate fields that are infeasible (i.e., which are zero by definition). In this example, the compact tandem MS matrix is 3×3 matrix (Figure 1B).

Simulating Tandem MS Data Using 2D-Convolutions

Next, we describe an algorithm to simulate the compact tandem MS matrix for metabolites that are naturally labeled and for isotopic tracers that are labeled at specific carbon positions. Previously, we demonstrated that mass isotopomer distributions (MID) can be simulated efficiently using a series of convolutions (function *conv* in Matlab) (Antoniewicz et al., 2007b), essentially by reconstructing a metabolite atom-by-atom (Figure 1C). Here, we demonstrate that the compact tandem MS matrix can be simulated analogously using a series of 2D-convolutions (function *conv2* in Matlab), again

by reconstructing a metabolite atom-by-atom. This approach can be used to simulate both naturally labeled compounds and tracers that are labeled at one or more specific carbon positions.

As an example, assume that we want to simulate the compact tandem MS matrix for metabolite A that is labeled at the second and third carbon positions, i.e., $[2,3-^{13}\text{C}]\text{A}$, and assume that the parent fragment contains all four carbon atoms, and the daughter fragment contains carbon atoms 3 and 4. Simulation of the compact tandem MS matrix is achieved by a sequence of 2D-convolutions as shown in Figure 1D. The 2D-convolution is performed with the transpose of the atom's MID vector if the atom is present in the daughter fragment, and using the atom's MID vector if the atom is not present in the daughter fragment. The simulated compact tandem MS matrix for $[2,3-^{13}\text{C}]\text{A}$ is shown in Figure 1D. The same approach can be used to simulate compact tandem MS matrices for metabolites of any size. This approach is not limited to simulating only carbon atoms, but can also be used to include any and all atoms. This is important since in LC-MS/MS natural abundances of e.g., sulfur (4.2% M+2) and oxygen (0.2% M+2) contribute to shifts in tandem MS distributions,

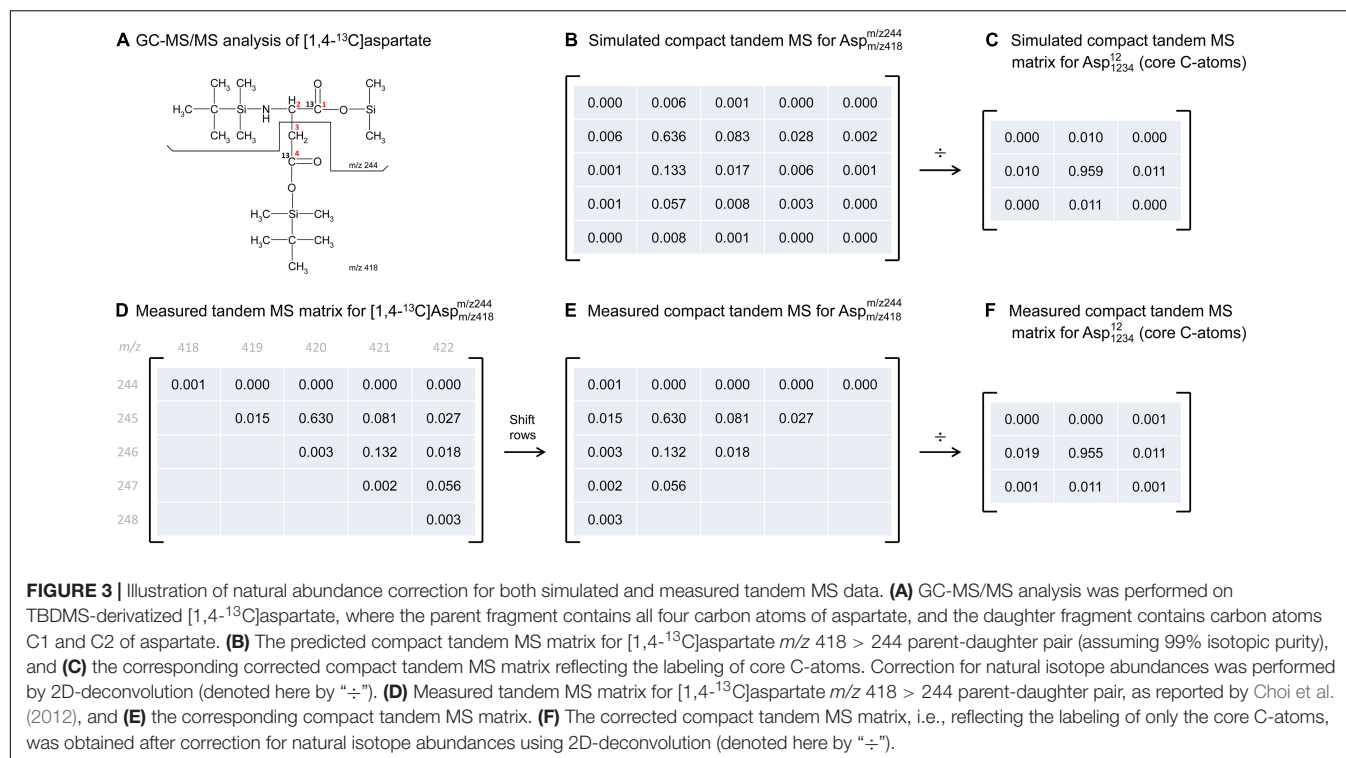


FIGURE 3 | Illustration of natural abundance correction for both simulated and measured tandem MS data. **(A)** GC-MS/MS analysis was performed on TBDMS-derivatized [1,4- ^{13}C]aspartate, where the parent fragment contains all four carbon atoms of aspartate, and the daughter fragment contains carbon atoms C1 and C2 of aspartate. **(B)** The predicted compact tandem MS matrix for [1,4- ^{13}C]aspartate m/z 418 > 244 parent-daughter pair (assuming 99% isotopic purity), and **(C)** the corresponding corrected compact tandem MS matrix reflecting the labeling of core C-atoms. Correction for natural isotope abundances was performed by 2D-deconvolution (denoted here by “÷”). **(D)** Measured tandem MS matrix for [1,4- ^{13}C]aspartate m/z 418 > 244 parent-daughter pair, as reported by Choi et al. (2012), and **(E)** the corresponding compact tandem MS matrix. **(F)** The corrected compact tandem MS matrix, i.e., reflecting the labeling of only the core C-atoms, was obtained after correction for natural isotope abundances using 2D-deconvolution (denoted here by “÷”).

and even more importantly, in GC-MS/MS analysis compounds are often derivatized (e.g., with TBDMS) which adds other atoms to the fragment that cause even more dramatic shifts in tandem MS distributions (Choi et al., 2012; Okahashi et al., 2016).

CORRECTION OF TANDEM MS DATA

Parent, Daughter, and Complement Fragments

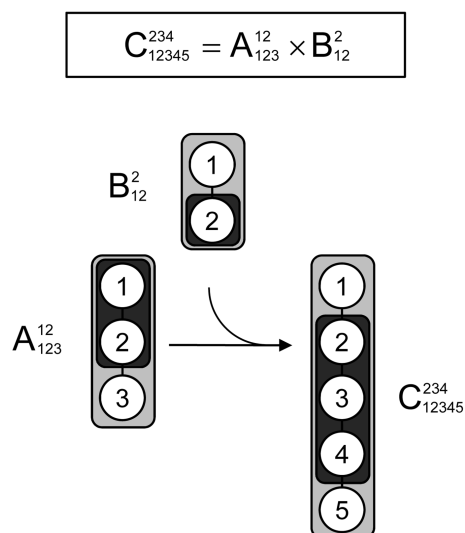
In order to use tandem MS data for ^{13}C -flux calculations, the data must be corrected for natural isotope abundances. Efficient algorithms have been developed for correcting MS data for use in ^{13}C -MFA (Fernandez et al., 1996); however, existing algorithms for correcting tandem MS data tend to be more tedious (Rantanen et al., 2002; Niedenfuhr et al., 2016). Here, we describe a more convenient algorithm based on the compact tandem MS matrix formulation described above. First, we must define a new term, the “complement fragment,” shown in **Figure 2A**. The complement fragment is defined as the part of the parent fragment that is not present in the daughter fragment. The daughter fragment and the complement fragment can then be further imagined to consist of core carbon atoms that originate from the metabolite that is measured, and other atoms (which may include both carbon and non-carbon atoms) from the derivatizing agent. For ^{13}C -MFA calculations, we are only interested in the labeling of the core C-atoms; thus, tandem MS data must be corrected for the skewing effects resulting from the presence of carbon atoms that are not core C-atoms.

Correcting Tandem MS Data for Natural Abundances

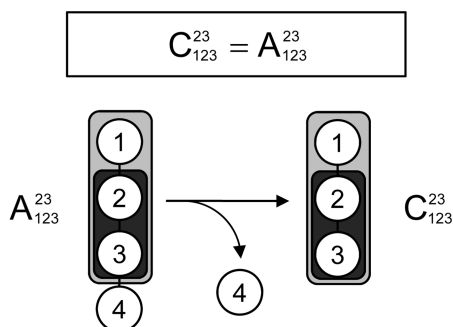
The correction of tandem MS data for natural isotope abundances is accomplished using 2D-deconvolution. To illustrate this, consider the case of TBDMS-derivatized aspartate. In a previous study, we validated several parent-daughter GC-MS/MS fragments for analysis of aspartate labeling (Choi et al., 2012). One of the validated parent-daughter pairs was m/z 418 > 244 (**Figure 2A**). In this case, the parent fragment (m/z 418, $\text{C}_{18}\text{H}_{40}\text{O}_4\text{NSi}_3$) contains all four C-atoms of aspartate and the daughter fragment (m/z 244, $\text{C}_{10}\text{H}_{22}\text{O}_2\text{NSi}_2$) contains the first two C-atoms of aspartate. Following the definitions in the previous section, the daughter fragment is thus imagined to consist of two core C-atoms of aspartate (i.e., C_2 , first two carbon atoms), and various other atoms ($\text{C}_8\text{H}_{22}\text{O}_2\text{NSi}_2$); the complement fragment consists of two core C-atoms of aspartate (i.e., C_2 , last two carbon atoms), and various other atoms ($\text{C}_6\text{H}_{18}\text{O}_2\text{Si}$). If the labeling of the core C-atoms is known, then we can predict the theoretical measured compact tandem MS matrix by 2D-convolutions shown in **Figure 2B**, where S is the natural abundance MID vector of $\text{C}_6\text{H}_{18}\text{O}_2\text{Si}$, and Q is the natural abundance MID vector of $\text{C}_8\text{H}_{22}\text{O}_2\text{NSi}_2$. The inverse operation, or 2D-deconvolution, can therefore be used to correct the measured tandem MS data for natural isotope abundances (**Figure 2C**). It should be noted that 2D-deconvolutions of this type are widely used in many fields of science such as image processing and filtering (e.g., functions *fft2* and *ifft2* in Matlab).

To illustrate the application of this correction algorithm, we have applied it to correct the tandem MS data that was previously measured for [1,4- ^{13}C]aspartate. **Figure 3D** shows the measured

A Condensation reaction



B Cleavage reaction



C Unimolecular reaction

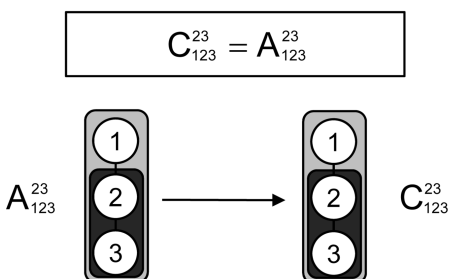


FIGURE 4 | Three types of biochemical reactions (A) condensation reaction, (B) cleavage reactions, and (C) unimolecular reaction, that must be considered for construction of EMU models. The labeling of an EMU product is expressed solely as a function of the labeling of the EMU educts. For the condensation reaction, the compact tandem MS matrix of the EMU product is calculated by 2D-convolution (denoted by “ \times ”) of the compact tandem MS matrices of the EMU educts.

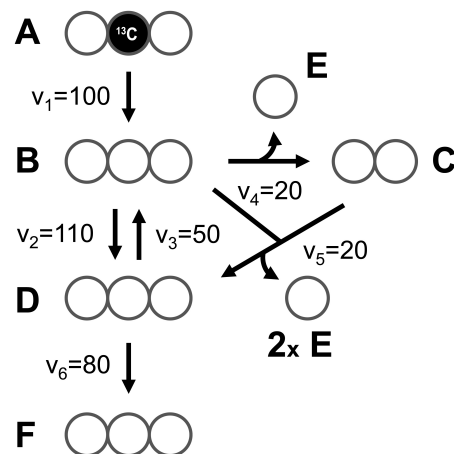


FIGURE 5 | A simple example network model (taken from the original EMU paper), used here to illustrate simulation of tandem MS data using the EMU approach.

TABLE 1 | Stoichiometry and atom transitions for the reactions in the example metabolic network.

Reaction No.	Reaction stoichiometry	Atom transitions*
1	$A \rightarrow B$	$abc \rightarrow abc$
2	$B \leftrightarrow D$	$abc \leftrightarrow abc$
3	$B \rightarrow C + E$	$abc \rightarrow bc + a$
4	$B + C \rightarrow D + E + E$	$abc + de \rightarrow bcd + a + e$
5	$D \rightarrow F$	$abc \rightarrow abc$

*For each metabolite atoms are identified using lower case letters to represent successive atoms in the metabolite.

tandem MS matrix for the parent-daughter fragment pair m/z 418 > 244, as reported by Choi et al. (2012) (Figure 3A). The measured tandem MS matrix was first transformed into the compact tandem MS matrix by row shifting (Figure 3E), and then corrected for natural isotope abundances using 2D-deconvolution, which produced a 3×3 matrix that now reflects the labeling of only the core C-atoms of aspartate (Figure 3F). For comparison, we also simulated the theoretical compact tandem MS matrix for [1,4- ^{13}C]aspartate, assuming 99% isotopic purity of the tracer (Figure 3B), as well as the corresponding corrected compact tandem MS matrix (Figure 3C). Overall, there was very good agreement between the measured and simulated matrices. This example demonstrates that correction for natural isotope abundances can be accomplished easily in a single step by 2D-deconvolution using the compact tandem MS matrices.

^{13}C -METABOLIC FLUX ANALYSIS WITH TANDEM MS DATA

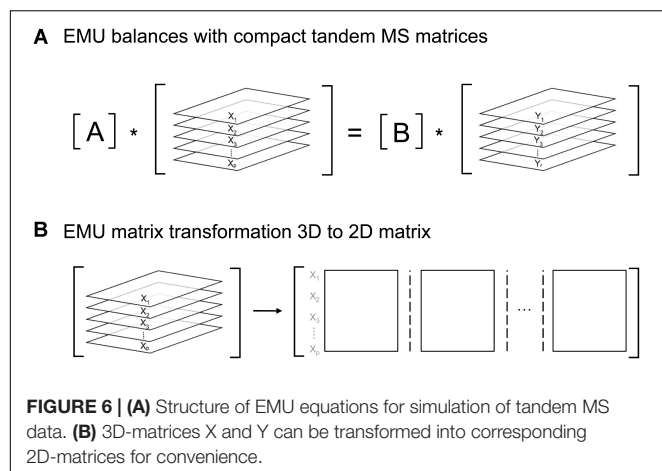
EMU Framework for Simulating Tandem MS

Lastly, we demonstrate that tandem MS data can be used for ^{13}C -MFA calculations using the widely used EMU framework. As

TABLE 2 | EMU decomposition for F_{123}^{23} in the example metabolic network.

Reaction No.	EMU reaction
6	$D_{123}^{23} \rightarrow F_{123}^{23}$
2	$B_{123}^{23} \rightarrow D_{123}^{23}$
5	$B_{23}^3 + C_1^1 \rightarrow D_{123}^{23}$
1	$A_{123}^{23} \rightarrow B_{123}^{23}$
3	$D_{123}^{23} \rightarrow B_{123}^{23}$
1	$A_{123}^{23} \rightarrow B_{123}^{23}$
3	$D_{23}^3 \rightarrow B_{23}^3$
2	$B_{23}^3 \rightarrow D_{23}^3$
5	$B_3 + C_1^1 \rightarrow D_{23}^3$
4	$B_2^2 \rightarrow C_1^1$
1	$A_2^2 \rightarrow B_2^2$
3	$D_2^2 \rightarrow B_2^2$
2	$B_2^2 \rightarrow D_2^2$
5	$B_3^3 \rightarrow D_2^2$
1	$A_3^3 \rightarrow B_3^3$
3	$D_3^3 \rightarrow B_3^3$
2	$B_3^3 \rightarrow D_3^3$
5	$C_1^1 \rightarrow D_3^3$
4	$B_2 \rightarrow C_1$
1	$A_2 \rightarrow B_2$
3	$D_2 \rightarrow B_2$
2	$B_2 \rightarrow D_2$
5	$B_3 \rightarrow D_2$
1	$A_3 \rightarrow B_3$
3	$D_3 \rightarrow B_3$
2	$B_3 \rightarrow D_3$
5	$C_1 \rightarrow D_3$

described in the original EMU paper (Antoniewicz et al., 2007b), there are three types of reactions that must be considered when constructing EMU models: a condensation reaction, a cleavage reaction, and a unimolecular reaction. **Figure 4** shows that for

**EMU Size 1**

$$\begin{bmatrix} -v_4 & v_4 & 0 & 0 & 0 \\ 0 & -v_1 - v_3 & v_3 & 0 & 0 \\ 0 & v_2 & -v_2 - v_5 & v_5 & 0 \\ 0 & 0 & 0 & -v_1 - v_3 & v_3 \\ v_5 & 0 & 0 & v_2 & -v_2 - v_5 \end{bmatrix} \begin{bmatrix} C_1 \\ B_2 \\ D_2 \\ B_3 \\ D_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -v_1 & 0 \\ 0 & 0 \\ 0 & -v_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_2 \\ A_3 \end{bmatrix}$$

$$\begin{bmatrix} -v_4 & v_4 & 0 & 0 & 0 \\ 0 & -v_1 - v_3 & v_3 & 0 & 0 \\ 0 & v_2 & -v_2 - v_5 & v_5 & 0 \\ 0 & 0 & 0 & -v_1 - v_3 & v_3 \\ v_5 & 0 & 0 & v_2 & -v_2 - v_5 \end{bmatrix} \begin{bmatrix} C_1^1 \\ B_2^2 \\ D_2^2 \\ B_3^3 \\ D_3^3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -v_1 & 0 \\ 0 & 0 \\ 0 & -v_1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} A_2^2 \\ A_3^3 \end{bmatrix}$$

EMU Size 2

$$\begin{bmatrix} -v_5 - v_2 & v_2 \\ v_3 & -v_1 - v_3 \end{bmatrix} \begin{bmatrix} D_{23}^3 \\ B_{23}^3 \end{bmatrix} = \begin{bmatrix} -v_5 & 0 \\ 0 & -v_1 \end{bmatrix} \begin{bmatrix} B_3 \times C_1^1 \\ A_{23}^3 \end{bmatrix}$$

EMU Size 3

$$\begin{bmatrix} -v_6 & v_6 & 0 \\ 0 & -v_5 - v_2 & v_2 \\ 0 & v_3 & -v_1 - v_3 \end{bmatrix} \begin{bmatrix} F_{123}^{23} \\ D_{123}^{23} \\ B_{123}^{23} \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ -v_5 & 0 \\ 0 & -v_1 \end{bmatrix} \begin{bmatrix} B_{23}^3 \times C_1^1 \\ A_{123}^{23} \end{bmatrix}$$

FIGURE 7 | EMU balances for simulation of tandem MS data in the simple example network model (shown in **Figure 5**). The EMU balances were constructed based on the EMU model decomposition shown in **Table 2**.**EMU Size 1**

$$\begin{bmatrix} C_1 \\ B_2 \\ D_2 \\ B_3 \\ D_3 \end{bmatrix} = \begin{bmatrix} -20 & 20 & 0 & 0 & 0 \\ 0 & -150 & 50 & 0 & 0 \\ 0 & 110 & -130 & 20 & 0 \\ 0 & 0 & 0 & -150 & 50 \\ 20 & 0 & 0 & 110 & -130 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ -100 & 0 \\ 0 & 0 \\ 0 & -100 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 0.067 & 0.933 \\ 0.067 & 0.933 \\ 0.200 & 0.800 \\ 0.933 & 0.067 \\ 0.800 & 0.200 \end{bmatrix}$$

EMU Size 2

$$\begin{bmatrix} D_{23}^3 \\ B_{23}^3 \end{bmatrix} = \begin{bmatrix} -130 & 110 \\ 50 & -150 \end{bmatrix}^{-1} \begin{bmatrix} -20 & 0 \\ 0 & -100 \end{bmatrix} \begin{bmatrix} 0.062 & 0.004 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0.871 & 0.062 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.013 & 0.787 \\ 0.004 & 0.929 \end{bmatrix} \begin{bmatrix} 0.187 & 0.013 \\ 0.062 & 0.004 \end{bmatrix}$$

EMU Size 3

$$\begin{bmatrix} F_{123}^{23} \\ D_{123}^{23} \\ B_{123}^{23} \end{bmatrix} = \begin{bmatrix} -80 & 80 & 0 \\ 0 & -130 & 110 \\ 0 & 50 & -150 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 0 \\ -20 & 0 \\ 0 & -100 \end{bmatrix} \begin{bmatrix} 0.000 & 0.062 & 0.008 & 0.867 & 0.058 & 0.004 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0.000 & 0.013 & 0.788 & 0.186 & 0.012 & 0.001 \\ 0.000 & 0.013 & 0.788 & 0.186 & 0.012 & 0.001 \\ 0.000 & 0.004 & 0.929 & 0.062 & 0.004 & 0.000 \end{bmatrix}$$

Compact tandem MS matrix

$$F_{123}^{23} = \begin{bmatrix} 0.000 & 0.013 \\ 0.788 & 0.186 \\ 0.012 & 0.001 \end{bmatrix}$$

FIGURE 8 | Numerical solution of the EMU balances for the simple example network to simulate tandem MS data for metabolite F (parent fragment C1–C3, daughter fragment C2–C3). For this simulation, the fluxes shown in **Figure 5** were used and metabolite A was assumed to be 100% ^{13}C -labeled at the second carbon position.

all three types of reactions the EMU product can be computed from the corresponding EMU educts. To simulate tandem MS data, compact tandem MS matrices can be used as state variables (note that MID vectors were used as state variables for simulating MS data in the original EMU paper). For an EMU condensation reaction, the compact tandem MS matrix of the EMU product is computed by a 2D-convolution as described above (**Figure 4A**). For the cleavage reaction and the unimolecular reaction, the compact tandem MS matrix of the EMU product is equal to the compact tandem MS matrix of the EMU educt.

EMU Decomposition of an Example Metabolic Network Model

At the core of the EMU methodology is the decomposition of the biochemical reaction network into EMU networks, which are then solved subsequently. The EMU decomposition is performed by tracing the origin of carbon atoms of a particular metabolite to carbon atoms of substrates. For more details, the reader is referred to the original EMU paper (Antoniewicz et al., 2007b). In the original EMU framework, EMU decomposition was accomplished by keeping track of C-atoms for the parent fragment only. To generate EMU networks for simulation of tandem MS data, we must also keep track of C-atoms of the daughter fragment. To illustrate this, we will use a simple example network model shown in **Figure 5** (that was also used in the original EMU paper), with the corresponding atoms transitions shown in **Table 1**. The network model was decomposed here in order to simulate tandem MS data for the parent-daughter fragment pair $F_{123} > F_{23}$. The complete EMU decomposition is shown in **Table 2**. It is noted that the approach described above is conceptually the same as the tandemers approach described by Tepper and Shlomi (2015).

^{13}C -MFA With Tandem MS Data and the EMU Framework

To determine fluxes with ^{13}C -MFA, isotopic labeling must be simulated by solving the EMU network models, which are represented mathematically in the form (see the original EMU paper for more details) (Antoniewicz et al., 2007b):

$$A * X = B * Y$$

Here, A and B are matrices that are functions of fluxes, and X and Y are matrices that contain the unknown and known EMU variables, respectively. In the original EMU framework, each row in the X and Y matrices contained MID vectors of the respective EMU variables. For simulations of tandem MS data, X and Y are now 3-dimensional matrices that contain the compact tandem MS matrices of the respective EMU variables (**Figure 6A**). For convenience, these 3D matrices can be transformed into 2D

matrices as shown in **Figure 6B**. By performing this 3D-to-2D transformation, the same EMU algorithms can be used to simulate tandem MS data that are used currently to simulate MS data. Thus, this allows current software packages to be upgraded easily to accommodate tandem MS measurements for flux calculations.

To illustrate the simulation of tandem MS data using the EMU framework and compact tandem MS matrices, **Figure 7** shows the EMU balances for the simple example model, and **Figure 8** shows the numerical simulation results. Here, we used the fluxes shown in **Figure 5** with metabolite A being 100% ^{13}C -labeled at the second carbon position. This simple example clearly illustrates that tandem MS data can be efficiently simulated using the existing EMU framework without any major modifications. Recent work has demonstrated that EMU decompositions of large-scale models are computationally tractable (Gopalakrishnan and Maranas, 2015). Thus, the methods and algorithms presented in this paper can be applied to realistically sized network models.

CONCLUDING REMARKS

Tandem mass spectrometry is a promising new analytical approach that provides additional labeling information for ^{13}C -flux studies. Previously, we demonstrated that this additional labeling information can significantly improve flux precision and resolution in complex biological systems (Choi and Antoniewicz, 2011). In this paper, we have presented a set of tools and algorithms for efficient simulation of tandem MS data using the EMU framework and for correction of tandem MS data for natural isotope abundances. By building upon the EMU framework, which is used by all current software packages for ^{13}C -MFA, we hope to accelerate the acceptance of tandem MS technique by ^{13}C -MFA community and encourage software developers to include capabilities for tandem MS ^{13}C -flux analysis in future software updates.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct and intellectual contribution to the work, and approved it for publication.

FUNDING

This work was supported by the NSF CAREER Award (CBET-1054120).

REFERENCES

- Ahn, W. S., Crown, S. B., and Antoniewicz, M. R. (2016). Evidence for transketolase-like TKTL1 flux in CHO cells based on parallel labeling experiments and (^{13}C) -metabolic flux analysis. *Metab. Eng.* 37, 72–78. doi: 10.1016/j.ymben.2016.05.005
- Antoniewicz, M. R. (2013). Tandem mass spectrometry for measuring stable-isotope labeling. *Curr. Opin. Biotechnol.* 24, 48–53. doi: 10.1016/j.copbio.2012.10.011
- Antoniewicz, M. R. (2015a). Methods and advances in metabolic flux analysis: a mini-review. *J. Ind. Microbiol. Biotechnol.* 42, 317–325. doi: 10.1007/s10295-015-1585-x

- Antoniewicz, M. R. (2015b). Parallel labeling experiments for pathway elucidation and ¹³C metabolic flux analysis. *Curr. Opin. Biotechnol.* 36, 91–97. doi: 10.1016/j.copbio.2015.08.014
- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2007a). Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. *Anal. Chem.* 79, 7554–7559.
- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2007b). Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metab. Eng.* 9, 68–86.
- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2011). Measuring deuterium enrichment of glucose hydrogen atoms by gas chromatography/mass spectrometry. *Anal. Chem.* 83, 3211–3216. doi: 10.1021/ac200012p
- Chandra, R., and Peng, L. (2012). “Application of Cooperative Convolution Optimization for ¹³C Metabolic Flux Analysis: Simulation of Isotopic Labeling Patterns Based on Tandem Mass Spectrometry Measurements,” in *Simulated Evolution and Learning. SEAL 2012. Lecture Notes in Computer Science*, Vol. 7673, eds L. T. Bui, Y. S. Ong, N. X. Hoai, H. Ishibuchi, and P. N. Suganthan (Heidelberg: Springer).
- Choi, J., and Antoniewicz, M. R. (2011). Tandem mass spectrometry: a novel approach for metabolic flux analysis. *Metab. Eng.* 13, 225–233. doi: 10.1016/j.ymben.2010.11.006
- Choi, J., Grossbach, M. T., and Antoniewicz, M. R. (2012). Measuring complete isotopomer distribution of aspartate using gas chromatography/tandem mass spectrometry. *Anal. Chem.* 84, 4628–4632. doi: 10.1021/ac300611n
- Crown, S. B., and Antoniewicz, M. R. (2012). Selection of tracers for ¹³C-metabolic flux analysis using elementary metabolite units (EMU) basis vector methodology. *Metab. Eng.* 14, 150–161. doi: 10.1016/j.ymben.2011.12.005
- Fernandez, C. A., Des Rosiers, C., Previs, S. F., David, F., and Brunengraber, H. (1996). Correction of ¹³C mass isotopomer distributions for natural stable isotope abundance. *J. Mass Spectrom.* 31, 255–262. doi: 10.1002/(SICI)1096-9888(199603)31:3<255::AID-JMS290>3.0.CO;2-3
- Gopalakrishnan, S., and Maranas, C. D. (2015). ¹³C metabolic flux analysis at a genome-scale. *Metab. Eng.* 32, 12–22. doi: 10.1016/j.ymben.2015.08.006
- Jeffrey, F. M., Roach, J. S., Storey, C. J., Sherry, A. D., and Malloy, C. R. (2002). ¹³C isotopomer analysis of glutamate by tandem mass spectrometry. *Anal. Biochem.* 300, 192–205. doi: 10.1006/abio.2001.5457
- Masakapalli, S. K., Ratcliffe, R. G., and Williams, T. C. (2014). Quantification of (1)³C enrichments and isotopomer abundances for metabolic flux analysis using 1D NMR spectroscopy. *Methods Mol. Biol.* 1090, 73–86. doi: 10.1007/978-1-62703-688-7_5
- McConnell, B. O., and Antoniewicz, M. R. (2016). Measuring the composition and stable-isotope labeling of algal biomass carbohydrates via gas chromatography/mass spectrometry. *Anal. Chem.* 88, 4624–4628. doi: 10.1021/acs.analchem.6b00779
- Niedenfuhr, S., Ten Pierick, A., Van Dam, P. T., Suarez-Mendez, C. A., Noh, K., and Wahl, S. A. (2016). Natural isotope correction of MS/MS measurements for metabolomics and (13)C fluxomics. *Biotechnol. Bioeng.* 113, 1137–1147. doi: 10.1002/bit.25859
- Okahashi, N., Kawana, S., Iida, J., Shimizu, H., and Matsuda, F. (2016). GC-MS/MS survey of collision-induced dissociation of tert-butylidimethylsilyl-derivatized amino acids and its application to (13)C-metabolic flux analysis of *Escherichia coli* central metabolism. *Anal. Bioanal. Chem.* 408, 6133–6140. doi: 10.1007/s00216-016-9724-4
- Quek, L. E., Wittmann, C., Nielsen, L. K., and Kromer, J. O. (2009). OpenFLUX: efficient modelling software for ¹³C-based metabolic flux analysis. *Microb. Cell Fact.* 8:25. doi: 10.1186/1475-2859-8-25
- Rantanen, A., Rousu, J., Kokkonen, J. T., Tarkkainen, V., and Ketola, R. A. (2002). Computing positional isotopomer distributions from tandem mass spectrometric data. *Metab. Eng.* 4, 285–294. doi: 10.1006/mben.2002.0232
- Tepper, N., and Shlomi, T. (2015). Efficient modeling of MS/MS data for metabolic flux analysis. *PLoS One* 10:e0130213. doi: 10.1371/journal.pone.0130213
- Weitzel, M., Noh, K., Dalman, T., Niedenfuhr, S., Stute, B., and Wiechert, W. (2013). 13CFLUX2-high-performance software suite for ¹³C-metabolic flux analysis. *Bioinformatics* 29, 143–145. doi: 10.1093/bioinformatics/bts646
- Yoo, H., Antoniewicz, M. R., Stephanopoulos, G., and Kelleher, J. K. (2008). Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line. *J. Biol. Chem.* 283, 20621–20627. doi: 10.1074/jbc.M706494200
- Young, J. D. (2014). INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* 30, 1333–1335. doi: 10.1093/bioinformatics/btu015
- Young, J. D., Walther, J. L., Antoniewicz, M. R., Yoo, H., and Stephanopoulos, G. (2008). An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnol. Bioeng.* 99, 686–699. doi: 10.1002/bit.21632

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

Copyright © 2019 Choi and Antoniewicz. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Flux Connections Between Gluconate Pathway, Glycolysis, and Pentose–Phosphate Pathway During Carbohydrate Metabolism in *Bacillus megaterium* QM B1551

Julie A. Wushensky¹, Tracy Youngster², Carol M. Mendonca¹ and Ludmilla Aristilde^{1,2*}

¹ Department of Biological and Environmental Engineering, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, United States, ² Soil and Crop Sciences Section, School of Integrative Plant Science, College of Agriculture and Life Sciences, Cornell University, Ithaca, NY, United States

OPEN ACCESS

Edited by:

Yinjie Tang,
Washington University in St. Louis,
United States

Reviewed by:

Wei Xiong,
National Renewable Energy
Laboratory (DOE), United States
Junyoung O. Park,
University of California, Los Angeles,
United States
Shan Yi,
University of California, Berkeley,
United States

*Correspondence:

Ludmilla Aristilde
ludmilla@cornell.edu;
la31@cornell.edu

Specialty section:

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

Received: 15 September 2018

Accepted: 30 October 2018

Published: 21 November 2018

Citation:

Wushensky JA, Youngster T,
Mendonca CM and Aristilde L (2018)
Flux Connections Between Gluconate
Pathway, Glycolysis,
and Pentose–Phosphate Pathway
During Carbohydrate Metabolism
in *Bacillus megaterium* QM B1551.
Front. Microbiol. 9:2789.
doi: 10.3389/fmicb.2018.02789

Bacillus megaterium is a bacterium of great importance as a plant-beneficial bacterium in agricultural applications and in industrial bioproduction of proteins. Understanding intracellular processing of carbohydrates in this species is crucial to predicting natural carbon utilization as well as informing strategies in metabolic engineering. Here, we applied stable isotope-assisted metabolomics profiling and metabolic flux analysis to elucidate, at high resolution, the connections of the different catabolic routes for carbohydrate metabolism immediately following substrate uptake in *B. megaterium* QM B1551. We performed multiple ¹³C tracer experiments to obtain both kinetic and long-term ¹³C profiling of intracellular metabolites. In addition to the direct phosphorylation of glucose to glucose-6-phosphate (G6P) prior to oxidation to 6-phosphogluconate (6P-gluconate), the labeling data also captured glucose catabolism through the gluconate pathway involving glucose oxidation to gluconate followed by phosphorylation to 6P-gluconate. Our data further confirmed the absence of the Entner–Doudoroff pathway in *B. megaterium* and showed that subsequent catabolism of 6P-gluconate was instead through the oxidative pentose–phosphate (PP) pathway. Quantitative flux analysis of glucose-grown cells showed equal partition of consumed glucose from G6P to the Embden–Meyerhof–Parnas (EMP) pathway and from G6P to the PP pathway through 6P-gluconate. Growth on fructose alone or xylose alone was consistent with the ability of *B. megaterium* to use each substrate as a sole source of carbon. However, a detailed ¹³C mapping during simultaneous feeding of *B. megaterium* on glucose, fructose, and xylose indicated non-uniform intracellular investment of the different carbohydrate substrates. Flux of glucose-derived carbons dominated the gluconate pathway and the PP pathway, whereas carbon flux from both glucose and fructose populated the EMP pathway; there was no assimilatory flux of xylose-derived carbons. Collectively, our findings provide new quantitative insights on the contribution of the different catabolic routes involved in initiating carbohydrate catabolism in *B. megaterium* and related *Bacillus* species.

Keywords: metabolomics of carbohydrate catabolism, glucose and fructose assimilation, gluconate uptake and metabolism, fructose metabolism, oxidative pentose–phosphate pathway, *Bacillus megaterium*

INTRODUCTION

Bacillus megaterium, an aerobic bacterium ubiquitous in a diverse range of environments, has been of special investigative interest for its applications in promoting plant health (Eppinger et al., 2011; Santos et al., 2014), bioremediation of contaminants (Quinn et al., 1989), and industrial bioproduction (Vary, 1994; Vary et al., 2007; Biedendieck et al., 2010). Despite extensive genetic characterization of *B. megaterium*, detailed elucidation of its metabolic network is still lacking (Kanehisa and Goto, 2000; Eppinger et al., 2011; Kanehisa et al., 2017). Much of what is known about carbon metabolism in *Bacillus* species is derived from *Bacillus subtilis*, the primary model organism for Gram-positive bacteria (Sauer et al., 1997; Dauner and Sauer, 2001; Dauner et al., 2001; Fuhrer et al., 2005). Previous metabolic studies highlighted incongruent metabolic fluxes between *B. subtilis* and *B. megaterium* species (Sauer et al., 1997; Dauner and Sauer, 2001; Dauner et al., 2001; Fuhrer et al., 2005; Furch et al., 2007a,b; Tannler et al., 2008; Youngster et al., 2017). Metabolic flux modeling has been performed on mutant strains of *B. megaterium* that lacked specific metabolic functions (Furch et al., 2007b). Therefore, the operational network of wild-type *B. megaterium* remains to be evaluated experimentally. Of particular interest is an investigation of the different catabolic routes that initiate carbohydrate metabolism in *B. megaterium* due to the ubiquitous presence of carbohydrate-containing feedstocks.

Here, we investigated the metabolic fluxes underlying the catabolism of glucose, fructose, and xylose in *B. megaterium* QM B1551. Glucose, a monomer of the biopolymers cellulose and starch, is a common carbohydrate in environmental matrices (Koegel-Knabner, 2002). The disaccharide sucrose, a dimer with glucose and fructose, is common in plant materials and can serve as a carbon source to *B. megaterium* (Youngster et al., 2017). Xylose, a pentose monosaccharide, is a major monomer in hemicellulose, an abundant component of plant cell walls (Koegel-Knabner, 2002). The enzymes involved in the uptake pathway of all three carbohydrates (i.e., glucose, fructose, and xylose) have been annotated in the genome of *B. megaterium* (Figure 1). The metabolic pathways that can be involved in carbohydrate catabolism are the Embden–Meyerhof–Parnas (EMP) pathway, the gluconate pathway, the oxidative pentose-phosphate (PP) pathway, the non-oxidative PP pathway, and the Entner–Doudoroff (ED) pathway (Figure 1).

It is well established that *Bacillus* species undergo glycolysis via the EMP pathway wherein glucose is first phosphorylated to glucose-6-phosphate (G6P), then isomerized to fructose-6-phosphate (F6P) before phosphorylation to fructose-1,6-bisphosphate (FBP) (Figure 1). An important step that provides the driving force for the yields of ATP downstream of the EMP pathway is the lysis of FBP to dihydroxyacetone phosphate (DHAP) and glyceraldehyde-3-phosphate (GAP) (Figure 1; Rabinowitz et al., 2015). The metabolites in the EMP pathway can also contribute to the PP pathway (Figure 1). The oxidative phase of the PP pathway involves the oxidation of G6P to 6-phosphogluconate (6P-gluconate) followed by decarboxylation to PPs; the non-oxidative phase involves the transketolase and

transaldolase reactions that combine F6P and GAP to yield intermediates in the PP pathway [xylulose-5-phosphate (Xu5P), erythrose-4-phosphate (E4P), and sedoheptulose-7-phosphate (S7P)] (Figure 1). Carbon fluxes from the EMP and PP pathways, which are eventually channeled to the tricarboxylic acid (TCA) cycle, generate reducing equivalents [NAD(P)H], energy (ATP), and metabolite precursors for biomass growth (i.e., biosynthesis of ribonucleotides for RNA and DNA, biosynthesis of amino acids for proteins) (Figure 1).

There is a lack of consensus regarding the distribution of the flux from G6P to the EMP pathway relative to the flux from G6P to the PP pathway via 6P-gluconate (Sauer et al., 1997; Dauner et al., 2001; Furch et al., 2007a,b; Tannler et al., 2008). Certain strains of *B. subtilis* (PRF93) were reported to utilize the PP pathway more than the direct EMP pathway (Sauer et al., 1997), whereas the opposite was reported for other *B. subtilis* strains (wild-type strain 168 and mutant strains RB50::[pRF69]_n) (Dauner et al., 2001; Tannler et al., 2008). In one mutant strain of *B. megaterium* (WH323), carbon starvation was shown to promote usage of the EMP pathway over the PP pathway, whereas recombinant gene induction increased activity in the PP pathway relative to the EMP pathway (Furch et al., 2007a,b). Furthermore, only a forward flux of the EMP pathway has been portrayed in previous models of *Bacillus* species, including *B. megaterium* (Sauer et al., 1997; Dauner et al., 2001; Furch et al., 2007b; Tannler et al., 2008). However, the genome of *B. megaterium* QM B1551 annotates the enzymes that can operate both forward and backward fluxes through the EMP pathway (Kanehisa and Goto, 2000; Kanehisa et al., 2017; Figure 1). The contribution of both of these fluxes, which has implications for the net thermodynamic driver of the EMP pathway, has been unexplored in previous *Bacillus* studies (Sauer et al., 1997; Dauner et al., 2001; Furch et al., 2007a,b; Tannler et al., 2008).

Regarding the gluconate pathway, which involves the oxidation of glucose to gluconate followed by phosphorylation to 6P-gluconate (Figure 1), annotation of the *B. megaterium* QM B1551 genome implied an incomplete gluconate pathway (Kanehisa and Goto, 2000; Kanehisa et al., 2017). Specifically annotated in the genome were glucose-1-dehydrogenase and gluconate kinase, which convert glucose to glucono-1,5-lactone and gluconate to 6P-gluconate, respectively (Figure 1; Kanehisa and Goto, 2000; Kanehisa et al., 2017). However, the enzyme gluconolactonase for the hydrolysis of glucono-1,5-lactone to gluconate was not present, albeit this reaction can happen spontaneously in highly alkaline conditions (Figure 1; Kanehisa and Goto, 2000; Kanehisa et al., 2017). Interestingly, despite the lack of genome annotation of gluconolactonase, evidence of the gluconate pathway was found in germinating spores of *B. megaterium* QM B1551 wherein significant gluconate evolution from glucose was detected (Otani et al., 1986; Sano et al., 1988). The connection between the gluconate pathway and the remaining network for carbohydrate catabolism in *B. megaterium* remains unclear. *Pseudomonas* species, which are well known to exhibit the gluconate pathway, rely significantly on the ED pathway to connect the gluconate pathway to downstream metabolism (del Castillo et al., 2007; Nikel et al., 2015; Sasnow et al., 2016; Wilkes et al., 2018). Due to the

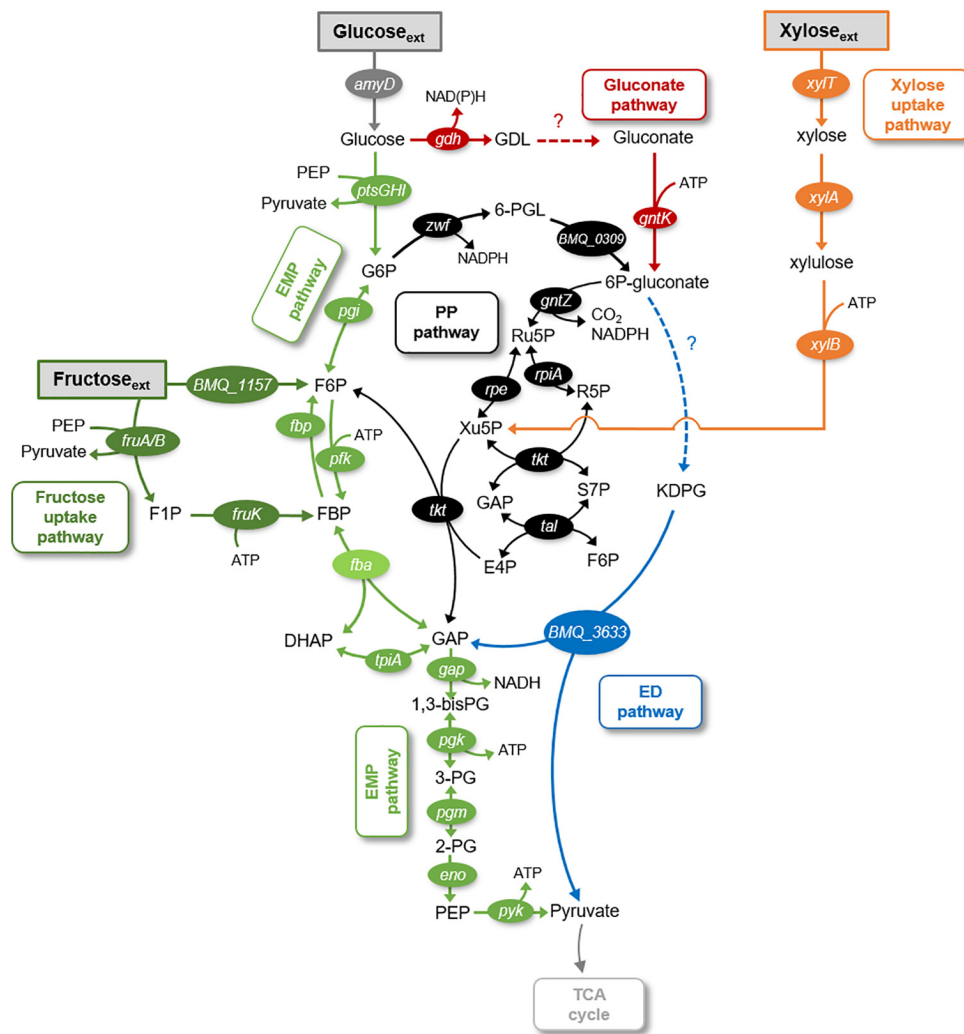


FIGURE 1 | Schematic of metabolic pathways involved in glucose, fructose, and xylose catabolism in *B. megaterium* QM B1551. The following pathways are emphasized: gluconate pathway (in red), the Entner–Doudoroff (ED) pathway (in blue), the Embden–Meyerhof–Parnas (EMP) pathway (in light green), the pentose–phosphate (PP) pathway (in black), the fructose uptake pathway (in dark green), and the xylose uptake pathway (in orange). The gene annotated corresponding to each metabolic enzyme is placed next to the respective reaction arrow. The information for the network was compiled from the KEGG database (Kanehisa and Goto, 2000; Kanehisa et al., 2017) and MetaCyc database (Caspi et al., 2012). The metabolite abbreviations are as follows: GDL, glucono-1,5-lactone; G6P, glucose-6-phosphate; 6-PGL, 6-phosphogluconolactone; 6P-gluconate, 6-phosphogluconate; KDPG, 2-keto-3-deoxy-6-phosphogluconate; Ru5P, ribulose 5-phosphate; Xu5P, xylulose 5-phosphate; S7P, sedoheptulose 7-phosphate; E4P, erythrose 4-phosphate; F6P, fructose 6-phosphate; FBP, fructose 1,6-bisphosphate; DHAP, dihydroxyacetone-3-phosphate; GAP, glyceraldehyde 3-phosphate; 1,3-bisPG, 1,3-bisphosphoglycerate; 3PG, 3-phosphoglycerate; 2PG, 2-phosphoglycerate; PEP, phosphoenolpyruvate. The enzymes corresponding to the gene annotations are as follows: *amyD*, carbohydrate ABC transporter permease AmyD; *ptsG*, PTS system glucose-specific transporter subunit IIBC; *ptsH*, phosphocarrier protein HPr; *ptsI*, PTS system transporter I; *gdh*, glucose 1-dehydrogenase; *pgi*, glucose-6-phosphate isomerase; *gntK*, gluconate kinase; *BMQ_1157*, fructokinase; *xylT*, xylose permease; *xylA*, D-xylose isomerase; *xylB*, xylokinase; *BMQ_0309*, 6-phosphogluconolactonase; *BMQ_3633*, 2-dehydro-3-deoxyphosphogluconate aldolase; *fruA*, PTS system fructose-specific II subunit IIA; *fruB*, PTS system fructose-specific II subunit IIB; *zwf*, glucose-6-phosphate 1-dehydrogenase; *gntZ*, phosphogluconate dehydrogenase; *rpe*, ribulose-phosphate 3-epimerase; *rpiA*, ribose 5-phosphate isomerase A; *tkt*, transketolase; *tal*, transaldolase; *pfk*, 6-phosphofructokinase; *fbp*, fructose-1,6-bisphosphatase; *fba*, fructose-1,6-bisphosphate aldolase; *tpiA*, triosephosphate isomerase; *gap*, glyceraldehyde-3-phosphate dehydrogenase; *pgk*, phosphoglycerate kinase; *pgm*, phosphoglucomutase; *eno*, enolase; *pyk*, pyruvate kinase.

absence of all the relevant enzymes, a functional ED pathway in *Bacillus* species has not been considered in previous metabolic studies (Furch et al., 2007a,b; Tannler et al., 2008). Specifically, in *B. megaterium* QM B1551, genome-level characterization has annotated one of the two essential enzymes required for the ED pathway (Kanehisa and Goto, 2000; Kanehisa

et al., 2017) (Figure 1). This characterization has noted the presence of 2-keto-3-deoxyphosphogluconate aldolase but the upstream enzyme, 6-phosphogluconate dehydratase, is absent from the genome of *B. megaterium* QM B1551 (Figure 1). Both the gluconate pathway and the ED pathway have not been considered in previously reported metabolic network models

of *Bacillus* species (Sauer et al., 1997; Dauner et al., 2001; Furch et al., 2007a,b; Tannler et al., 2008; Korneli et al., 2012).

Previous metabolic flux analysis (MFA) of *Bacillus* species has relied on ^{13}C -labeled amino acid labeling to estimate labeling at specific metabolic nodes (Sauer et al., 1997; Dauner et al., 2001; Furch et al., 2007a,b; Tannler et al., 2008; Korneli et al., 2012). However, because there are only two amino acid precursors from intermediates in the PP pathway [ribose-5-phosphate (R5P) and E4P], this method is not suitable for high-resolution determination of the carbon fluxes through the PP pathway as well as the other vicinal metabolic pathways such as the gluconate pathway and the ED pathway. Furthermore, labeling of free metabolites is needed for more accurate quantitative determination of flux distribution between the EMP pathway and the adjacent pathways. Here, we employed a ^{13}C -assisted metabolomic approach using high-resolution liquid chromatography-mass spectrometry (LC-MS) to detect free metabolites involved in all four pathways. We obtained both kinetic and long-term ^{13}C labeling of intracellular metabolites. We combined the cellular ^{13}C mapping with growth phenotypes to perform quantitative flux analysis of carbohydrate catabolism. Our results provide new insights on the metabolic network structure and flux distribution in *B. megaterium* (strain QM B1551). These findings have broader implications regarding the optimization of *Bacillus* and related bacterial species in biotechnological applications.

EXPERIMENTAL METHODS

Culturing Conditions

The *B. megaterium* QM B1551 cells were obtained from the Bacillus Genetic Stock Center (Columbus, OH, United States). Cell cultures were grown at 30°C in a G24 environmental incubator shaker (New Brunswick Scientific, Edison, NJ, United States) at 220 rpm. The growth medium, which was pH-adjusted (pH 7.0) and filter-sterilized (0.22 μm nylon; Waters Corporation, MA, United States), contained the following major salts: 18.7 mM NH_4Cl , 0.81 mM MgSO_4 , 0.034 mM CaCl_2 , 2H₂O, 89.4 mM K_2HPO_4 , 56.4 mM NaH_2PO_4 ·H₂O, 8.6 mM NaCl. The trace metal concentrations were as follows: 30 μM FeSO_4 ·7H₂O, 1.9 μM H_3BO_3 , 0.86 μM CuSO_4 ·5H₂O, 7.7 μM ZnSO_4 ·7H₂O, 0.75 μM MnSO_4 ·H₂O, 0.26 μM NiCl_2 ·6H₂O, and 0.31 μM Na_2MoO_4 ·2H₂O. The carbohydrate composition was 330 mM C total as glucose alone (equivalent to 55 mM or 9.91 g L⁻¹ glucose), fructose, alone, xylose alone, 1:1 glucose:gluconate mixture, or equimolar glucose:fructose:xylose mixture. All chemicals listed above were analytical grade, purchased from Sigma-Aldrich (St. Louis, MO, United States) and Fisher Scientific (Pittsburg, PA, United States). For all growth conditions, the cells were transferred twice into fresh growth media to ensure that the cells were well acclimated to their nutrient growth conditions prior to experimental sampling. Cell growth (three biological replicates per condition) was monitored *via* both optical density at 600 nm (OD₆₀₀) using an Agilent Cary UV-visible spectrophotometer (Santa Clara, CA, United States)

and measuring cell dry weight (in grams, gCDW). To obtain accurate reading at OD₆₀₀ above 0.5, cell suspensions were diluted. To determine the cell dry weight (three biological replicates), we centrifuged (9,391 g for 5 min at 4°C) harvested 1.5-mL culture aliquots and the retentate was frozen (−20°C) overnight prior to lyophilization using a Labconco freeze-dryer (Kansas City, MO, United States). The conversion factors of gCDW L⁻¹ per OD₆₀₀ were found to be: 0.61 ± 0.08 for glucose; 0.43 ± 0.10 for glucose:fructose:xylose; and 0.53 ± 0.07 for glucose:gluconate.

Carbohydrate Consumption

The depletion of substrate in the extracellular medium was taken to account for sugar consumption; this was confirmed by intracellular ^{13}C labeling (Aristilde et al., 2015). To quantify substrate depletion by exponentially growing cells under each growth condition, 0.7 mL aliquots were harvested at different times (two independent biological replicates at each timepoint) throughout cell growth and analyzed *via* ^1H nuclear magnetic resonance (NMR). The extracted aliquots were centrifuged for 30 min at 15,871 g and 4°C in filter-containing (0.22- μm pore size, nylon) microcentrifuge tubes. The filtered supernatants were frozen at −20°C until further processing. In preparation for the NMR analysis, 200 μL of the filtered samples were mixed with 60 μL of 100% D₂O, 50 μL of 6 mM 2,2-dimethyl-2-silapentane-5-sulfonate (DSS) as an internal standard, 240 μL of 100 mM sodium bicarbonate as a pH control, and 50 μL of 10 mM sodium-azide as an antimicrobial agent. Samples were stored at 4°C until analysis (Aristilde et al., 2015). The ^1H NMR measurements were performed using a Varian Unity INOVA 600-MHz NMR spectrometer at 25°C, with a relaxation delay of 5 s, recording of 16 scans per sample, and receiver gain of 32 dB (Aristilde et al., 2015). Substrate depletion rates (in mmol gCDW⁻¹ h⁻¹) were determined subsequently *via* regression analysis of carbohydrate depletion over time combined with biomass growth rate.

Metabolite Excretion

To determine excretion rates of metabolites, 50 μL of filtered cell suspensions (three biological replicates per condition) were obtained during growth and subsequently diluted with LC-MS grade water (Fisher Scientific, Pittsburgh, PA, United States) at 1:20 v/v during early exponential phase, or 1:200 v/v during late exponential and stationary phases. Different dilution ratios were used to account for elevated concentrations of extracellular metabolites in the media as a function of cell growth. The samples were stored at 4°C (for 5 h or less) prior to processing *via* LC-MS. Excretion rates (μmol gCDW⁻¹ h⁻¹) were calculated *via* regression analysis.

Kinetic Metabolite ^{13}C Labeling

Kinetic flux experiments were performed to monitor *in vivo* cellular incorporation of glucose (Sasnow et al., 2016). Batch cultures (two independent biological replicates) were grown as described above for the glucose-growth condition until early exponential phase, corresponding to OD₆₀₀ 0.4–0.6. Aliquots (3 mL) of the cultures were filtered (0.22- μm nylon filter). Each

cell-containing filter disk was then placed cell-side up on an agar plate containing unlabeled carbon in minimal media to allow the cells to acclimate and reach an exponentially growing phase, corresponding to OD₆₀₀ 0.5. To determine when this OD₆₀₀ value was reached, cells on filters from parallel plates were rinsed off into a 3-mL suspension for OD₆₀₀ reading. After the cells were adequately acclimated on the agar plate, the cell-containing filter disks were transferred to agar plates containing [U-¹³C₆]-glucose and, after a set period of time following the isotopic switch (0 and 20 s; 1, 4, 12, and 30 min), the filter disks were removed from the labeled plate and metabolism was immediately quenched by flipping the disks cell-side down into a 2-mL cold (4°C) methanol:acetonitrile:water quenching solution (2:2:1). The quenched solution containing the lysed cells scraped from the filter disks was then centrifuged (5 min at 9,391 g and 4°C). We dried 100-μL aliquots of the supernatants under N₂ gas prior to re-suspension in 100-μL LC-MS water for use in metabolomics analysis as described below.

Long-Term Intracellular Metabolite ¹³C Labeling

For long-term isotopic enrichment of the intracellular metabolites, liquid cultures were prepared with the growth medium containing the major and minor salts as listed above and the following ¹³C-labeled substrates: [1,2-¹³C₆]-glucose alone, equimolar [U-¹³C₆]-glucose and unlabeled gluconate, or equimolar [1,2,3-¹³C₃]-glucose, [1,6-¹³C₂]-fructose, and unlabeled xylose. All labeled carbohydrates were purchased from Cambridge Isotopes (Tewksbury, MA, United States) or Omicron Biochemicals (South Bend, IN, United States). For these labeling experiments, batch cultures (two independent biological replicates) were grown twice in labeled minimal media and sampling was done at two timepoints (when the cells reached OD₆₀₀ of 1.0 and 2.0) during exponential growth (Supplementary Figure S1). At each sampling time, 3 mL aliquots of cell suspensions were filtered through 0.22-μm nylon filters in a sterile environment. The cell-containing filter disks were immediately quenched in 2 mL of cold (4°C) methanol:acetonitrile:water solution (2:2:1). The lysed cells in the quenching solution were processed as described in the previous section in preparation for metabolomics analysis as described below.

Metabolomics Analysis via LC-MS

The samples prepared as described above were analyzed by reversed-phase ion-pairing ultra-high performance LC (Thermo Scientific Dionex UltiMate 3000) coupled to high-resolution/accurate MS (Thermo Scientific Q Exactive quadrupole-Orbitrap hybrid MS) with electrospray ionization run in full-scan negative mode (Aristilde et al., 2017). Details of the LC-MS protocol used here were previously reported (Aristilde et al., 2017). The following metabolites were isolated by LC-MS: gluconate, 6P-gluconate, G6P, F6P, FBP, pyruvate, Xu5P, R5P, DHAP, and S7P. As previously detailed and illustrated in Aristilde, 2017, the analytical isolation of the different compound isomers (i.e., G6P/F6P, Xu5P/R5P, and DHAP/GAP) was made

possible due to their chromatographic separations, despite their similar mass-over-charge ratios. Metabolite standards at various concentrations (10–1,000 nM) were also run in parallel to verify LC-MS identification and quantitation. The ¹³C labeling patterns were analyzed with the MAVEN (Metabolomic Analysis and Visualization Engine) software suite (Melamud et al., 2010; Clasquin et al., 2012). The labeling data were corrected subsequently for natural abundance of ¹³C.

Quantitative Metabolic Flux Analysis

The MFA of *B. megaterium* cells grown on glucose alone or the glucose:fructose:xylose mixture was constrained by the following experimental data: substrate consumption rates, metabolite excretion rates, long-term ¹³C labeling data of intracellular metabolites, and biomass growth. We employed previously reported stoichiometric biomass composition for *B. subtilis* (Dauner and Sauer, 2001) to estimate metabolite effluxes (from G6P, R5P, E4P, 3-PG, pyruvate, and DHAP) to sustain the biomass requirements for cell growth under each condition. The modeling software suite 13CFLUX2 was used to conduct the MFA analysis of the pathway network involving metabolic nodes for substrate uptake, the gluconate pathway, the EMP pathway, the PP pathway, and the biomass effluxes (Weitzel et al., 2012). Metabolite secretion rates were also included for gluconate. The model was initialized using a number of free fluxes, informed by previous values of fluxes in *B. megaterium* (Furch et al., 2007a). From these free fluxes, which were unconstrained values assigned by the user to provide an initial set of flux values, the modeling algorithm optimized all flux values in accordance with the experimental constraints through an iterative process (Supplementary Tables S1, S2). At each iteration, the quality of estimated fluxes was evaluated by calculating residual errors between the model-estimated metabolite labeling data and the corresponding experimental data (Supplementary Figures S2, S3).

RESULTS

Glucose Catabolism Involves the Gluconate Pathway Without the ED Pathway

Following uptake of extracellular glucose, glucose catabolism can be initiated by either phosphorylation to G6P or oxidation to gluconate (Figure 1). Kinetic ¹³C-labeling data following feeding of cells on [U-¹³C₆]-glucose showed, by 30 min, nearly complete isotopic enrichment of G6P and 6P-gluconate and ~75% isotopic enrichment of gluconate (Figure 2A). The ¹³C labeling of gluconate was indicative of glucose oxidation to gluconate, which was also reported in spores of *B. megaterium* QM B1551 (Otani et al., 1986; Sano et al., 1988). The plateau of gluconate ¹³C labeling below 100% may be due to carry over of residual non-labeled gluconate in the extracellular milieu because gluconate secretion was evident during growth of *B. megaterium* on glucose alone (Supplementary Table S1). Evidence of the gluconate pathway, which would include phosphorylation of

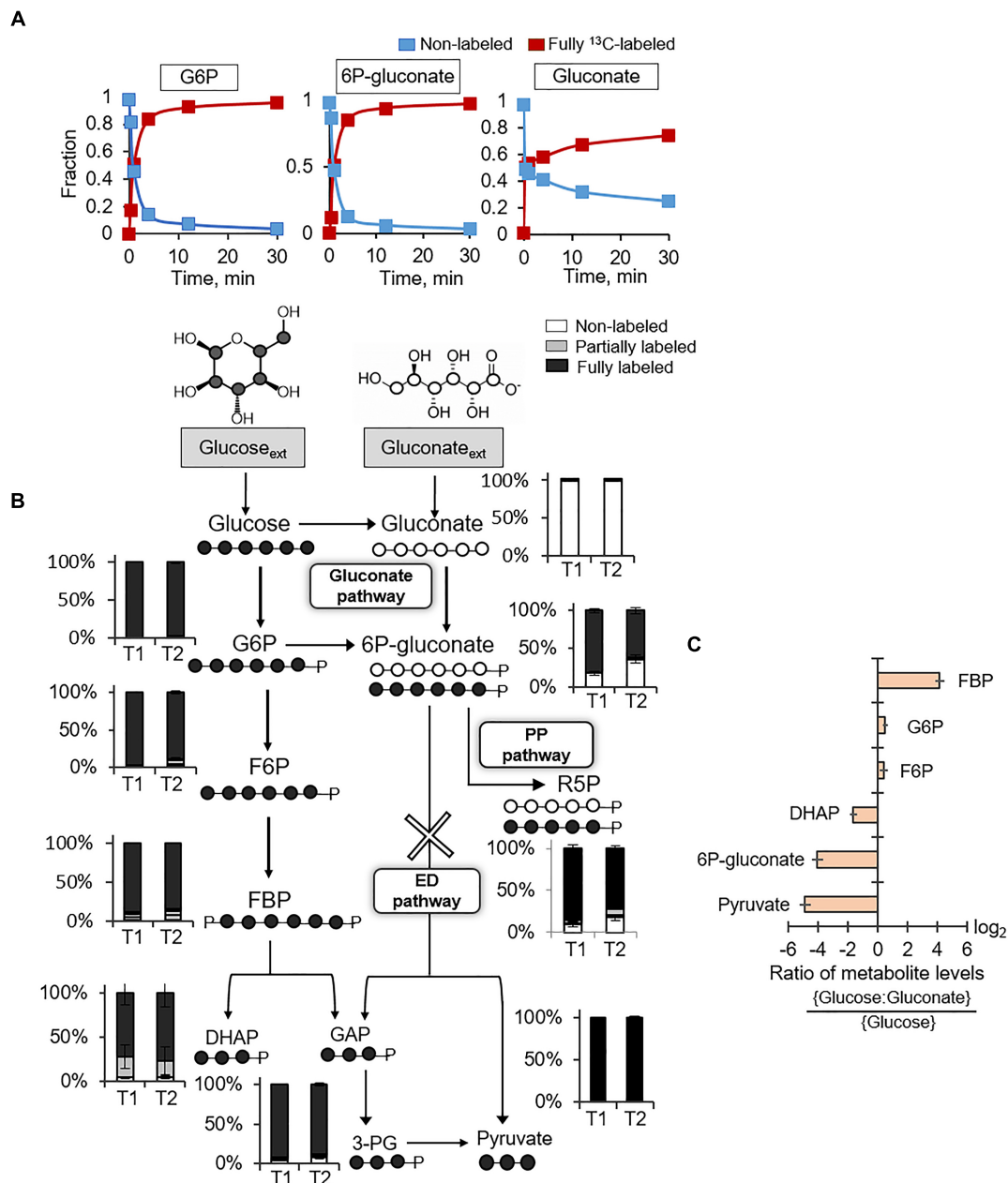


FIGURE 2 | Presence of gluconate pathway and absence of Entner–Doudoroff pathway. **(A)** Kinetic ^{13}C profiling of G6P, 6P-gluconate, and gluconate in cells grown on fully ^{13}C -labeled glucose. **(B)** Carbon mapping and long-term intracellular ^{13}C labeling at two timepoints during growth on fully ^{13}C -labeled glucose and unlabeled gluconate ($\text{OD}_{600} = 1$, T1; $\text{OD}_{600} = 2$, T2). The ED pathway is marked with a cross. **(C)** Ratio of metabolite levels in cells grown on glucose alone versus cells grown on both glucose and gluconate. Measured data (average \pm data range) were from two technical replicates of two independent biological replicates for each timepoint. Growth curves in **Supplementary Figure S1** indicate when data were acquired during biomass growth. Definitions for metabolite abbreviations are as shown in **Figure 1** caption.

gluconate to 6P-gluconate could not be resolved from the kinetic intracellular labeling from ^{13}C -glucose alone. The similar labeling kinetics for 6P-gluconate and G6P implied that 6P-gluconate was primarily produced from G6P but these data could not confirm either the presence or lack of gluconate flux to 6P-gluconate (**Figure 2A**). To capture specifically the incorporation of gluconate-derived carbons into 6P-gluconate, we fed the

cells equimolar concentrations of fully ^{13}C -labeled glucose and unlabeled gluconate (**Figure 2B**). Across the two timepoints during exponential growth on this glucose:gluconate mixture, the fraction of non-labeled 6P-gluconate increased from 18% to 36%, confirming the presence of an active gluconate pathway (**Figure 2B**). However, given the high-fraction of ^{13}C -labeled 6P-gluconate (above 60%), these kinetic data also indicated

that the assimilation route of glucose-derived ^{13}C carbons from G6P to 6P-gluconate was favored over the gluconate pathway (Figure 2B).

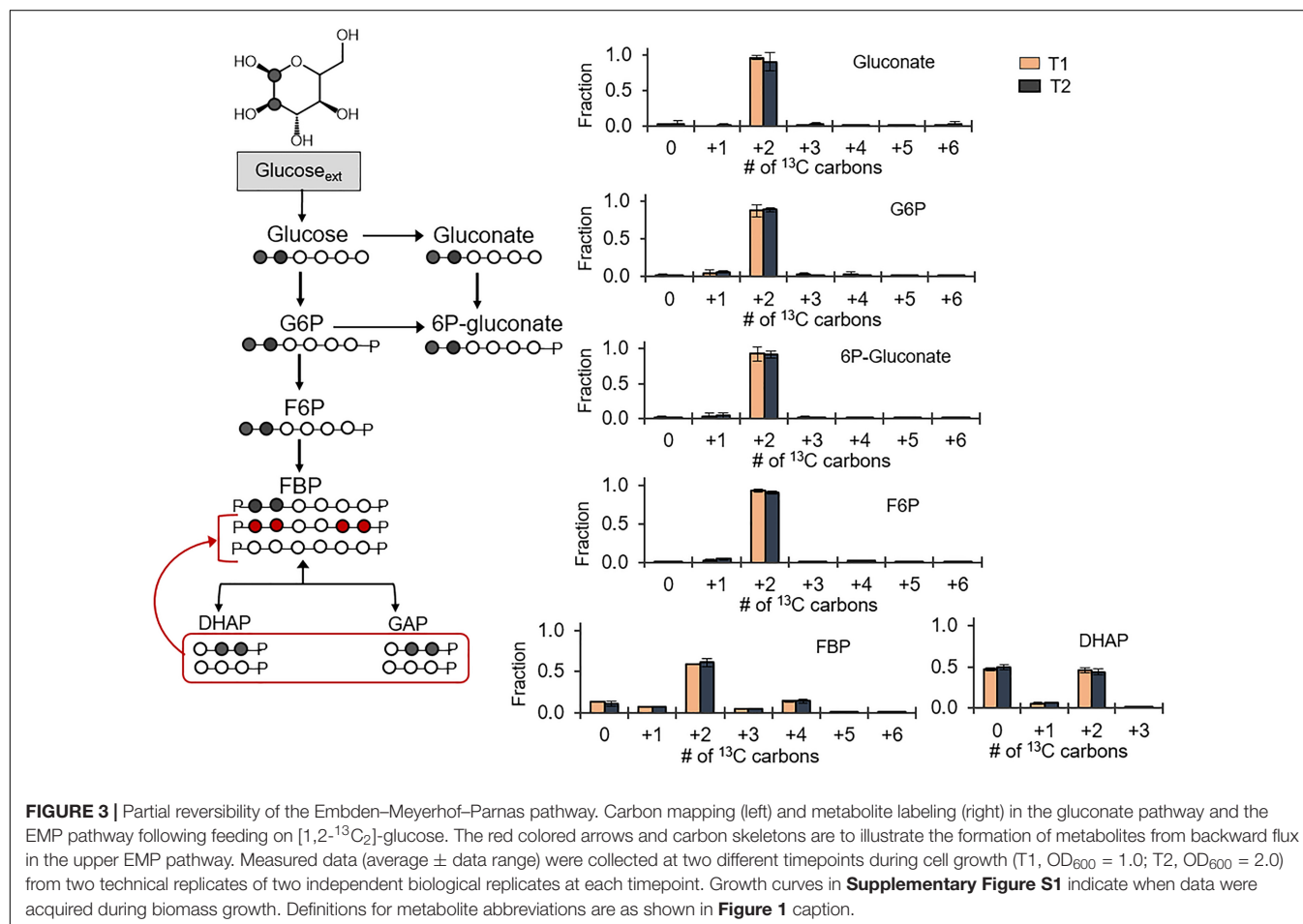
There are two possible fates of 6P-gluconate: the oxidative PP pathway and the ED pathway (Figure 2B). The ED pathway involves the splitting of 6P-gluconate into GAP and pyruvate (Figure 2B). At the timepoint when 6P-gluconate had the highest non-labeled fraction ($\sim 36\%$), the non-labeled fractions of both pyruvate and DHAP (an isomer of GAP) were less than 5% (i.e., within the error range of the LC-MS detection) (Figure 2B). The oxidative PP pathway was expected to generate non-labeled PPs (such as R5P) as evidence of catabolism of gluconate-derived carbons *via* this pathway (Figure 2B). We found that nearly 20% of R5P was non-labeled, implying that about 72% of the R5P pool was synthesized from 6P-gluconate *via* the oxidative PP pathway (Figure 2B). Therefore, these data collectively demonstrated that the fate of 6P-gluconate was through the oxidative PP pathway and that the ED pathway was inactive in *B. megaterium* QM B1551.

We compared the metabolite pools in cells grown on the equimolar glucose:gluconate mixture to those in cells grown on glucose alone – the total substrate carbon concentration (330 mM C) was the same in both conditions (Figure 2C). There was no appreciable change in the levels of G6P and F6P

(Figure 2C). However, during growth on the glucose:gluconate mixture, the level of 6P-gluconate was severely depleted (16-fold reduction), in agreement with limited flux from gluconate to 6P-gluconate, as indicated by the labeling data (Figures 2B,C). In accordance with the lack of the ED pathway, we obtained a decrease in both DHAP (~ 3 -fold reduction) and pyruvate (~ 30 -fold reduction) during feeding on the glucose:gluconate mixture (Figure 2C). Interestingly, despite the depletion in DHAP, which is downstream of FBP, there was an accumulation in FBP (~ 18 -fold increase) during feeding on the glucose:gluconate mixture (Figure 2C). Given the constant level of F6P upstream of FBP, these data implied that, when gluconate is present with glucose, the flux of FBP production from F6P exceeded the flux of FBP lysis to generate DHAP and GAP (Figure 2C).

Glucose Catabolism to Pyruvate Is Through a Partially Reversible EMP Pathway

Following the ^{13}C probing approach described above, $[1,2-^{13}\text{C}_2]$ -glucose was chosen as a substrate to evaluate possible backward fluxes in the EMP pathway (Figure 3). These backward fluxes, which are implied from the annotated genome of *B. megaterium* (Kanehisa and Goto, 2000; Kanehisa et al., 2017), have not been



probed previously. The forward EMP pathway involves carbon flux from glucose through G6P, F6P, FBP, and onward to DHAP and GAP following FBP lysis (**Figure 3**). In agreement with the assimilation of the doubly ^{13}C -labeled glucose, we obtained ~90% doubly ^{13}C -labeled fractions for gluconate, 6P-gluconate, G6P, and F6P (**Figure 3**). However, there was a discrepancy between the labeling patterns of F6P and those of FBP (**Figure 3**). Compared to F6P, there was a 30% depletion in the doubly ^{13}C -labeled fraction of FBP (**Figure 3**). Moreover, non-labeled FBP was 10%, compared to 1% in F6P; the quadruply ^{13}C -labeled fraction was 14% in FBP, compared to 2% in F6P (**Figure 3**). Non-labeled intermediates in the EMP pathway were generated during lysis of the doubly $[1,2-^{13}\text{C}_2]$ -FBP into $[2,3-^{13}\text{C}_2]$ DHAP and non-labeled GAP (**Figure 3**). Equilibrium isotopic labeling between the isomers DHAP and GAP led to the measurement of near equal fraction (~50%) of non-labeled and doubly ^{13}C -labeled fractions in DHAP (**Figure 3**). Operating the EMP pathway in the reverse direction by combining the labeling scheme for DHAP and GAP to produce FBP explained the non-labeled and quadruply ^{13}C -labeled fraction in FBP, in addition to the doubly ^{13}C -labeled fraction (**Figure 3**). However, the fact that F6P was primarily doubly ^{13}C -labeled indicated that the EMP pathway was only partially reversible (**Figure 3**).

Glucose Catabolism and Gluconate Pathway Are Both Connected to a Primarily Oxidative PP Pathway

The PP pathway possesses both an oxidative route and a non-oxidative route (**Figures 4A,B**). The oxidative PP pathway involves the decarboxylation of 6P-gluconate to Ru5P, which subsequently interconverts to R5P and Xu5P; the non-oxidative PP pathway combines metabolites from the EMP pathway, F6P and GAP, to generate eventually Xu5P and R5P (**Figures 4A,B**). According to our labeling scheme, decarboxylation of doubly ^{13}C -labeled 6P-gluconate through the oxidative PP pathway would produce singly ^{13}C -labeled R5P and Xu5P but the non-oxidative PP pathway would use EMP-pathway metabolites to generate triply ^{13}C -labeled R5P and Xu5P (**Figures 4A,B**). We found that the metabolite labeling patterns of R5P and Xu5P were primarily singly ^{13}C -labeled (77 and 72%, respectively) and only up to 18% of these metabolites were triply ^{13}C -labeled (**Figure 4**); in accordance with the labeling of Xu5P and R5P, we also obtained primarily doubly ^{13}C -labeled S7P (**Figure 4**). Thus, the labeling patterns of Xu5P and R5P were consistent with a significant involvement of the oxidative PP pathway with relatively minor contribution from the non-oxidative pathway (**Figure 4**). Therefore, through the oxidative PP pathway, PP metabolites were primarily produced from 6P-gluconate, which was generated from both the gluconate pathway and the direct glucose catabolism through G6P.

Simultaneous Processing of Different Carbohydrates Is Partitioned Into Different Catabolic Routes

Previous reports of carbon metabolism and metabolic flux modeling have focused on glucose-grown *Bacillus* species

(Sauer et al., 1997; Dauner et al., 2001; Tannler et al., 2008). However, glucose is often present with other carbohydrates in carbon feedstocks (Wendisch et al., 2016). In addition to glucose, we were able to obtain growth of the *B. megaterium* cells when the sole carbon source was given as fructose (another common hexose) or xylose (a common pentose) (**Supplementary Figure S1**). These data thus indicated that, in accordance with the genome annotation (**Figure 1**), the cells have the transporters and catabolic pathways to assimilate these carbohydrates. We also investigated how *B. megaterium* QM B1551 processes simultaneously glucose, fructose, and xylose by monitoring intracellular incorporation of equimolar carbon-equivalent concentrations of $[1,2,3-^{13}\text{C}_3]$ -glucose, $[1,6-^{13}\text{C}_2]$ -fructose, and unlabeled xylose (**Figure 5**). The labeling data revealed that a non-uniform intracellular routing of each substrate in the metabolic network (**Figure 5**).

Consistent with the oxidation of the triply ^{13}C -labeled glucose, gluconate was predominantly triply ^{13}C -labeled (at ~85%) (**Figure 5**). From the phosphorylation of triply ^{13}C -labeled glucose to G6P, the G6P pool had a high triply ^{13}C -labeled fraction (60–73%) with a lesser amount of doubly ^{13}C -labeled (19%) fraction (**Figure 5**). Downstream of G6P and gluconate, 6P-gluconate was ~85% triply ^{13}C -labeled and 11% doubly ^{13}C -labeled (**Figure 5**). The enrichment in triply ^{13}C -labeled fraction (by up to 20%) from G6P to 6P-gluconate emphasized the contribution of gluconate to generate 6P-gluconate, thus underscoring an active gluconate pathway in initiating the catabolism of glucose-derived carbons during growth on multiple carbohydrates (**Figure 5**). The small fraction of doubly ^{13}C -labeled fractions in both G6P and 6P-gluconate reflected the minor incorporation of fructose-derived carbons in these metabolites (**Figure 5**).

Notably, for the F6P pool, we obtained an equal amount (~45% each) of triply ^{13}C -labeled fraction and doubly ^{13}C -labeled fractions, respectively from glucose-derived and fructose-derived carbons (**Figure 5**). Forward flux through the EMP pathway was expected to generate ^{13}C labeling for FBP similar to F6P and, subsequently, the splitting of the triply ^{13}C -labeled and doubly ^{13}C -labeled fractions of FBP would generate equal proportions of non-labeled, singly ^{13}C -labeled, and triply ^{13}C -labeled fractions of triose-phosphates (**Figure 5**). In accordance with this labeling scheme in the EMP pathway, we did obtain an equal proportion of non-labeled, singly ^{13}C -labeled and triply ^{13}C -labeled fractions of DHAP, 3-PG, and pyruvate (**Figure 5**). However, in addition to doubly and triply ^{13}C -labeled FBP (at 20–32%), there were other ^{13}C -labeled fractions in FBP (at 5–15%) (**Figure 5**). Specifically, the latter fractions were non-labeled, singly ^{13}C -labeled, quadruply ^{13}C -labeled, and sextuply ^{13}C -labeled fractions of FBP (**Figure 5**). Thus, the labeling pattern of FBP reflected backward flux from the combination of DHAP and GAP to produce FBP (**Figure 5**). However, this backward flux was not extended to F6P (**Figure 5**). Therefore, the partial reversibility of carbon flux in the upper EMP pathway previously discussed with the glucose-grown cells was also evident in cells grown on the glucose:fructose:xylose mixture (**Figures 3, 5**). It is important to note that, in addition to indicating that the upper EMP pathway was only partially reversible, the distinct difference

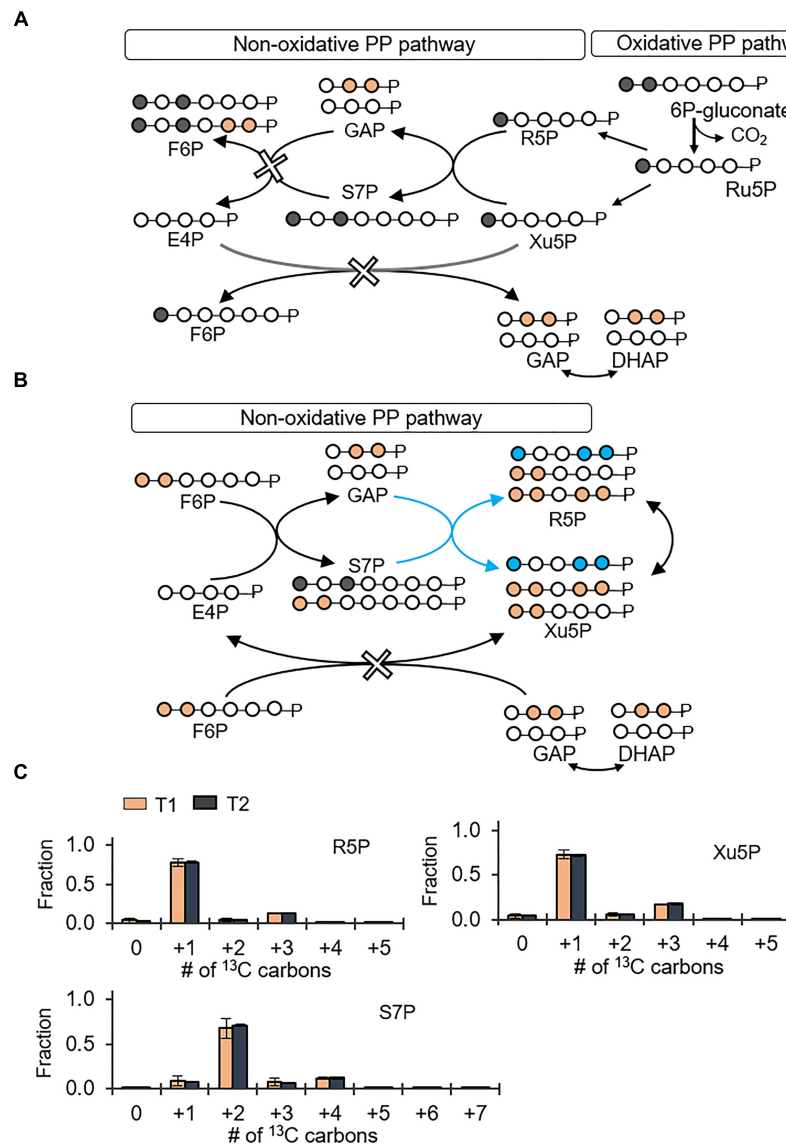
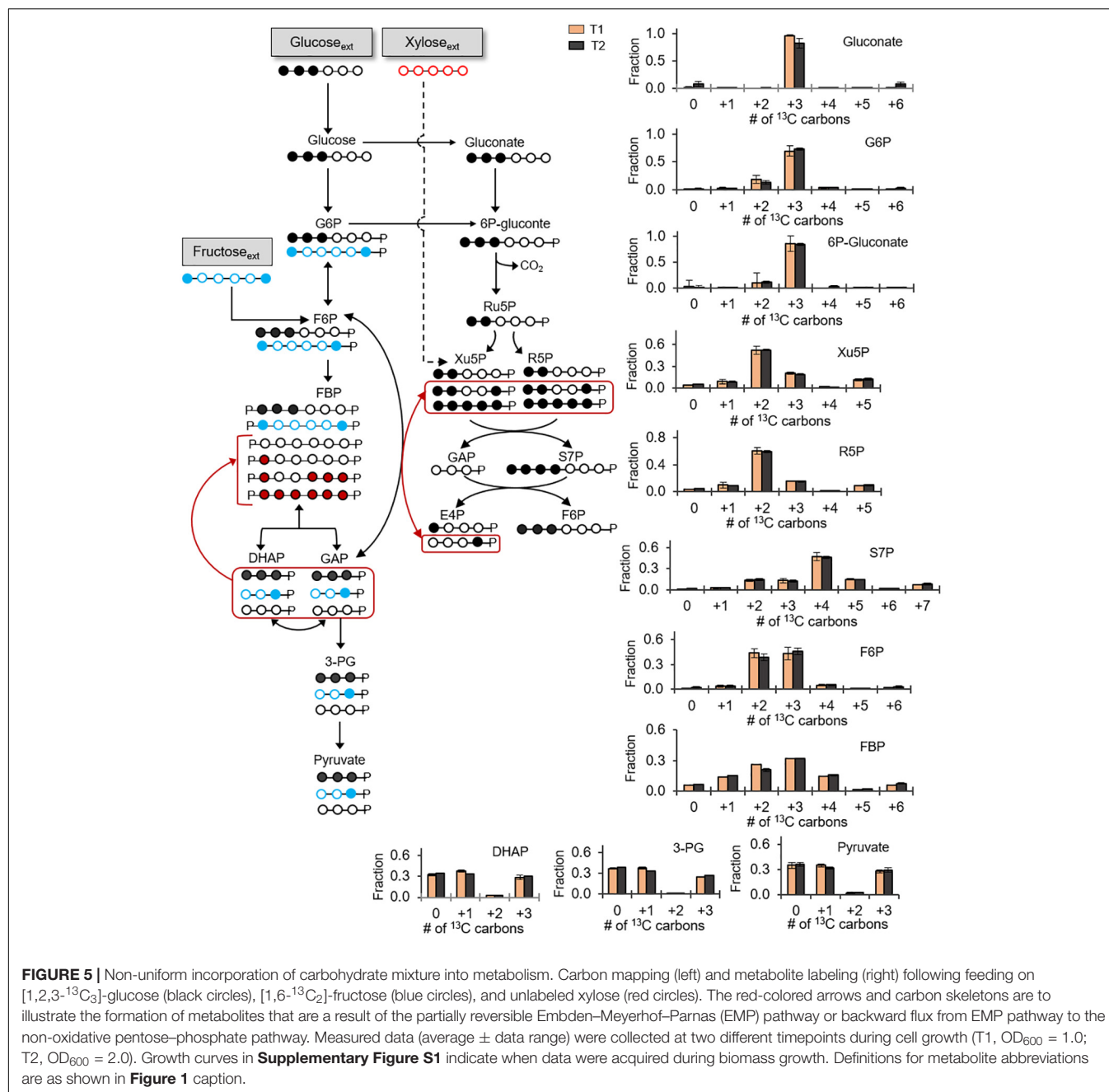


FIGURE 4 | Predominance of the oxidative pentose phosphate (PP) pathway over the non-oxidative PP pathway. Carbon mapping (**A,B**) and metabolite labeling (**C**) in the PP pathway following feeding on [1,2-¹³C₂]-glucose. Carbon mapping is shown for carbon flow from the oxidative PP pathway (**A**) or from the metabolites in the EMP pathway incorporated into the non-oxidative PP pathway (**B**). Blue-colored circles and arrows indicate metabolite labeling patterns derived from the EMP pathway operating in the direction of the non-oxidative PP pathway. Measured data (average ± data range) were collected at two different timepoints (T1, OD₆₀₀ = 1.0; T2, OD₆₀₀ = 2.0) from two technical replicates of two independent biological replicates. Growth curves in **Supplementary Figure S1** indicate when data were acquired during biomass growth. Definitions for metabolite abbreviations are as shown in **Figure 1** caption.

between the labeling patterns of F6P and FBP revealed that F6P was the entry point for the catabolism of fructose-derived carbons (**Figure 5**).

With respect to xylose assimilation, the incorporation of xylose-derived carbons was expected to introduce non-labeled carbons in the PP pathway. However, non-labeled fractions in the PP pathway metabolites (R5P, Xu5P, and S7P) were only 5% or less (**Figure 5**). By contrast, in accordance with the oxidative PP pathway, triply ¹³C-labeled 6P-gluconate from assimilated glucose-derived carbons led to 52–60% doubly ¹³C-labeled fractions in both R5P and Xu5P (**Figure 5**). Consistent

with the combination of these latter two metabolites to produce S7P, we obtained quadruply ¹³C-labeled S7P (~46%) (**Figure 5**). Despite the predominance of the oxidative PP pathway, minor fractions of other ¹³C isotopologs of the PP pathway metabolites were consistent with minor flux of the non-oxidative PP pathway (**Figure 5**). In sum, during growth on the carbohydrate mixture, glucose-derived carbons primarily populated the gluconate pathway and the oxidative PP pathway, both fructose-derived and glucose-derived carbons equally contributed to the EMP pathway, and the assimilation of xylose was insignificant (**Figure 5**).



Quantitative Flux Modeling During Growth on Glucose Alone Versus a Mixture With Glucose, Fructose, and Xylose

Using the ¹³C labeling data and the growth phenotypes, we performed MFA to quantitate explicitly up to 18 fluxes through the different catabolic routes for carbohydrate utilization in *B. megaterium* during growth on glucose alone or with fructose and xylose (**Figure 6**). The fluxes were normalized to the glucose uptake rate in both growth conditions (**Figure 6**) – the absolute fluxes are presented in **Supplementary Table S2**. The

cellular fluxes demonstrated that different catabolic routes were accentuated in *B. megaterium* depending on whether the cells were processing a single or a mixture of carbohydrate substrates (**Figure 6**).

Compared to the metabolism of glucose alone, there was a two-fold increase of the flux through the gluconate pathway in cells grown on the mixture (**Figures 6A,B**). This increased flux through the gluconate pathway during processing of the mixture was accompanied by a decrease (by up to 12%) in the carbon flux from G6P to F6P (**Figures 6A,B**). The increased carbon flux through the gluconate pathway led to a 17% increase in the activity of the oxidative PP pathway, from 6P-gluconate to

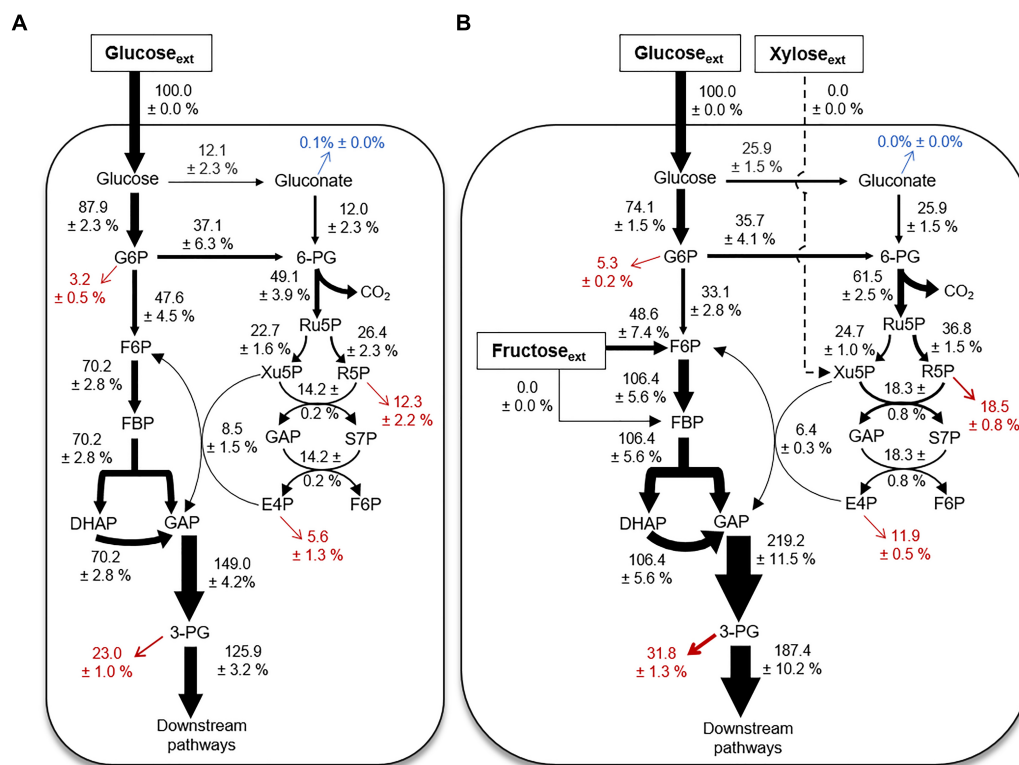


FIGURE 6 | Cellular metabolic flux distributions. Metabolic flux analysis of *B. megaterium* QM B1551 during feeding on (A) [1,2-¹³C₂]-glucose or (B) a mixture of [1,2,3-¹³C₃]-glucose, [1,6-¹³C₂]-fructose, and unlabeled xylose. All fluxes were normalized to 100% glucose uptake and the thickness of each arrow was scaled to this flux percentage. Fluxes toward biomass production are shown in red. Gluconate excretion is shown in blue. Flux values (average ± data range) were generated using experimental data from two independent biological replicates, at OD₆₀₀ = 2. Growth curves in **Supplementary Figure S1** indicate when data were acquired during biomass growth. Definitions for metabolite abbreviations are as shown in **Figure 1** caption.

Ru5P, in cells grown on the mixture relative to glucose alone (**Figures 6A,B**). Subsequent fluxes through the PP pathway were also increased (**Figures 6A,B**). The exception was the increase (by 25%) in the reaction flux that combines Xu5P and E4P to produce F6P and GAP (**Figures 6A,B**), which may be influenced by the influx of fructose-derived carbons into the EMP pathway (**Figures 5, 6A,B**). During growth on the mixture, the uptake flux from fructose to F6P was about half of the glucose uptake flux (**Figure 6B**). The second annotated pathway for fructose uptake (i.e., from fructose to FBP), was found to be inactive (**Figures 1, 6B**). Due to the incorporation of fructose-derived carbons at the metabolic node of F6P, the flux through the remaining EMP pathway downstream of F6P, was increased by up to 70% (**Figures 6A,B**). Thus, the overall flux profile in the gluconate pathway, EMP pathway, and PP pathway was reprogrammed to accommodate the assimilation of fructose (**Figures 6A,B**).

DISCUSSION

Here, we employed ¹³C-carbon mapping of intracellular metabolism to address knowledge gaps regarding the routes for carbohydrate catabolism in *B. megaterium* QM B1551. We also resolved inconsistencies from previous studies on *Bacillus* species

regarding the relative flux from glucose catabolism through either the EMP pathway or the PP pathway (Sauer et al., 1997; Dauner et al., 2001; Furch et al., 2007a,b; Tannler et al., 2008; Korneli et al., 2012). Our ¹³C labeling data presented evidence of both glucose oxidation to gluconate and flux from gluconate to 6P-gluconate, thus confirming the presence of the complete gluconate pathway in *B. megaterium* (**Figures 2A,B**). Our data also revealed that flux through the gluconate pathway was compromised by congestion at the gluconate node (**Figures 2A,B**). Further studies are needed to elucidate the regulatory mechanism for the gluconate pathway in *B. megaterium* QM B1551, which was beyond the scope of our data. Interestingly, in *Pseudomonas* species, gluconate pathway is strongly linked to a very active ED pathway (Nikel et al., 2015; Sasnow et al., 2016; Wilkes et al., 2018). Although the ED pathway was widely neglected in prior MFAs conducted on *Bacillus* species (Furch et al., 2007a,b; Tannler et al., 2008), genomic characterization of *B. megaterium* QM B1551 annotated one of the two essential enzymes in the ED pathway (**Figure 1**; Kanehisa and Goto, 2000; Kanehisa et al., 2017). However, our data demonstrated that the ED pathway was not active in *B. megaterium* QM B1551 (**Figure 2B**). Regarding the traditional glycolytic pathway (i.e., the EMP pathway), our flux modeling determined a net forward flux through the EMP pathway (**Figure 6**), in agreement with previous MFA studies

on *Bacillus* species including *B. megaterium* (Sauer et al., 1997; Dauner et al., 2001; Furch et al., 2007a,b; Tannler et al., 2008). However, our discrete metabolite labeling data revealed that the FBP lysis was reversible (Figures 3, 5). The consequence of this reversibility regarding the optimization of bioproduct generation in *B. megaterium* species remains to be determined.

The resolution of metabolite labeling in the PP pathway of *Bacillus* species has previously been deduced from ^{13}C labeling of their amino acid derivatives (Sauer et al., 1997; Dauner et al., 2001; Furch et al., 2007b; Korneli et al., 2012). Here, using a LC-MS approach to obtain the labeling patterns of three free metabolites specific to the PP pathway, we determined that the flux through the oxidative PP pathway (i.e., from 6P-gluconate to Ru5P) was ~ 6 -fold to ~ 10 -fold greater than the flux through the non-oxidative PP pathway (from F6P and GAP) (Figures 4C, 6B). This relative contribution of the two routes in the PP pathway was larger than previous published estimates, which reported a 3.5- to 5-fold greater activity of the oxidative PP pathway activity than non-oxidative PP pathway in *Bacillus* species (Sauer et al., 1997; Dauner et al., 2001; Furch et al., 2007b). The increased flux through the oxidative PP pathway determined from our modeling fluxes implied that the *B. megaterium* QM B1551 cells can afford a greater production of reducing power through the generation of NADPH than previously estimated.

While glucose is the substrate of choice to elucidate the metabolism of most bacterial model systems, a mixture of carbohydrates is typically present in environmental systems or in engineered bioreactors for industrial applications (Wendisch et al., 2016). Biomass growth on only glucose, fructose, or mannose indicated that *B. megaterium* can rely on each substrate as a sole source of carbon in its metabolism (Supplementary Figure S1). However, our metabolic study of cells fed on equimolar carbon-equivalent concentrations of glucose, fructose, and xylose revealed that *B. megaterium* QM B1551 utilizes different catabolic routes to process this carbohydrate mixture (Figures 5, 6). No appreciable incorporation of xylose-derived carbons was detected (Figure 5). Carbons derived from fructose and glucose contributed equally to the flux in the EMP pathway from F6P to the triose phosphates (Figure 5). The gluconate pathway was only populated by glucose-derived

carbons (Figure 5). Flux of glucose-derived carbons was also significant from the gluconate pathway through the oxidative PP pathway (Figures 5, 6). Our findings thus demonstrated that *B. megaterium* QM B1551 grown on a mixture can exhibit both complete repression of a given carbohydrate assimilation and preferential channeling of the assimilated carbohydrates into different catabolic routes. These findings have important implications for the processing of different carbon feedstocks. Follow-up metabolic flux studies and proteomics profiling are needed to gain further insights on how *B. megaterium* regulates this hierarchy in carbohydrate catabolism in response to different composition and concentration of carbohydrate mixtures.

Given the emergent usage of different feedstocks in engineered bioproduction, it is critical to understand the metabolic network in biotechnologically important bacterial species in response to various carbon sources (Wendisch et al., 2016). Due to the abundance of carbohydrate-containing feedstocks, we have focused here on determining the flux through the different catabolic pathways available for carbohydrate processing in *B. megaterium* QM B1551. The ^{13}C -metabolomics approaches employed here present a powerful tool to elucidate new pathways and gain insights on existing pathways involved in the intracellular network of cellular metabolism. These approaches could be instrumental to the broader field of metabolic characterization of environmentally-relevant bacteria and optimization of bacterial cells in biotechnological applications.

AUTHOR CONTRIBUTIONS

LA supervised the research. JW and LA designed the research, conducted data analysis, and wrote the manuscript. JW and TY performed the experiments. JW and CM performed the quantitative flux modeling. All authors contributed to the final draft of the manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02789/full#supplementary-material>

REFERENCES

- Aristilde, L. (2017). Metabolite labelling reveals hierarchies in *Clostridium acetobutylicum* that selectively channel carbons from sugar mixtures towards biofuel precursors. *Microb. Biotechnol.* 10, 162–174. doi: 10.1111/1751-7915.12459
- Aristilde, L., Lewis, I. A., Park, J. O., and Rabinowitz, J. D. (2015). Hierarchy in pentose sugar metabolism in *Clostridium acetobutylicum*. *Appl. Environ. Microbiol.* 81, 1452–1462. doi: 10.1128/AEM.03199-14
- Aristilde, L., Reed, M. L., Wilkes, R. A., Youngster, T., Kukurugya, M. A., Katz, V., et al. (2017). Glyphosate-induced specific and widespread perturbations in the metabolome of soil *Pseudomonas* species. *Front. Environ. Sci.* 5:34. doi: 10.3389/fenvs.2017.00034
- Biedendieck, R., Bunk, B., Furch, T., Franco-Lara, E., Jahn, M., and Jahn, D. (2010). "Systems biology of recombinant protein production in *Bacillus megaterium*," in *Biosystems Engineering I*, eds C. Wittmann and R. Krull (Heidelberg: Springer-Verlag Berlin Heidelberg), 133–162.
- Caspi, R., Altman, T., Dreher, K., Fulcher, C. A., Subhraveti, P., Keseler, I. M., et al. (2012). The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic Acids Res.* 40, D742–D753. doi: 10.1093/nar/gkr1014
- Clasquin, M. F., Melamud, E., and Rabinowitz, J. D. (2012). LC-MS data processing with MAVEN: a metabolomic analysis and visualization engine. *Curr. Protoc. Bioinformatics* 37:Unit14.11. doi: 10.1093/nar/gkr1014
- Dauner, M., Bailey, J. E., and Sauer, U. (2001). Metabolic flux analysis with a comprehensive isotopomer model in *Bacillus subtilis*. *Biotechnol. Bioeng.* 76, 144–156. doi: 10.1002/0471250953.bi1411s37
- Dauner, M., and Sauer, U. (2001). Stoichiometric growth model for riboflavin-producing *Bacillus subtilis*. *Biotechnol. Bioeng.* 76, 132–143. doi: 10.1002/bit.1154

- del Castillo, T., Ramos, J. L., Rodriguex-Herva, J. J., Fuhrer, T., Sauer, U., Duque, E., et al. (2007). Convergent peripheral pathways catalyze initial glucose catabolism in *Pseudomonas putida*: genomic and flux analysis. *J. Bacteriol.* 189, 5142–5152. doi: 10.1002/bit.1153
- Eppinger, M., Bunk, B., Johns, M. A., Edirisinghe, J. N., Kutumbaka, K. K., Koenig, S. S. K., et al. (2011). Genome sequences of the biotechnologically important *Bacillus megaterium* strains QM B1551 and DSM319. *J. Bacteriol.* 193, 4199–4213. doi: 10.1128/JB.00449-11
- Fuhrer, T., Fischer, E., and Sauer, U. (2005). Experimental identification and quantification of glucose metabolism in seven bacterial species. *J. Bacteriol.* 187, 1581–1590. doi: 10.1128/JB.187.5.1581-1590.2005
- Furch, T., Hollmann, R., Wittmann, C., Wang, W., and Decker, W. D. (2007a). Comparative study on central carbon metabolic fluxes of *Bacillus megaterium* strains in continuous culture using ¹³C labeled substrates. *Bioprocess Biosyst. Eng.* 30, 47–59. doi: 10.1128/JB.00449-11
- Furch, T., Wittmann, C., Wang, W., Granco-Lara, E., Jahn, D., and Deckwer, W. D. (2007b). Effect of different carbon sources on central metabolic fluxes and the recombinant production of a hydrolase from *Thermobifida fusca* in *Bacillus megaterium*. *J. Biotechnol.* 132, 385–394.
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* 45, D353–D361. doi: 10.1093/nar/gkw1092
- Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Koegel-Knabner, I. (2002). The macromolecular organic composition of plant and microbial residues as inputs to soil organic matter. *Soil Biol. Biochem.* 34, 139–162. doi: 10.1016/S0038-0717(01)00158-4
- Korneli, C., Bolten, C. J., Godard, T., Franco-Lara, E., and Wittmann, C. (2012). Debottlenecking recombinant protein production in *Bacillus megaterium* under large-scale conditions – targeted precursor feeding designed from metabolomics. *Biotechnol. Bioeng.* 109, 1538–1550. doi: 10.1002/bit.24434
- Melamud, E., Vastag, L., and Rabinowitz, J. D. (2010). Metabolomic analysis and visualization engine for LC-MS data. *Anal. Chem.* 82, 9818–9826. doi: 10.1021/ac1021166
- Nikel, P. I., Chavarria, M., Fuhrer, T., Sauer, U., and de Lorenzo, V. (2015). *Pseudomonas putida* KT2440 strain metabolizes glucose through a cycle formed by enzymes of the entner-doudoroff, embden-meyerhof-parnas, and pentose phosphate pathways. *Biol. Chem.* 290, 25920–25932. doi: 10.1074/jbc.M115.687749
- Otani, M., Ihara, N., Umezawa, C., and Sano, K. (1986). Predominance of gluconate formation from glucose during germination of *Bacillus megaterium* QM B1551 spores. *J. Bacteriol.* 167, 148–152. doi: 10.1128/jb.167.1.148-152.1986
- Quinn, P. J., Peden, J. M. M., and Dick, R. E. (1989). Carbon-phosphorus bond cleavage by gram-positive and gram-negative soil bacteria. *Appl. Microbiol. Biotechnol.* 31, 283–287. doi: 10.1007/BF00258410
- Rabinowitz, J. D., Aristilde, L., and Amador-Noguez, D. (2015). *Metabolomics of Clostridial Biofuel Production*. Technical Report No. SC0006839. Washington, DC: United States Department of Energy. doi: 10.2172/1213974
- Sano, K., Otani, M., and Umezawa, C. (1988). Glucose metabolism via the embden-meyerhof pathway is not involved in ATP production during spore germination of *Bacillus megaterium* QM B1551: a study with a mutant lacking hexokinase. *Biochem. Biophys. Res. Commun.* 151, 48–52. doi: 10.1016/0006-291X(88)90557-8
- Santos, S., Neto, I. F., Machado, M. D., Soares, H. M., and Soares, E. V. (2014). Siderophore production by *Bacillus megaterium*: effect of growth phase and cultural conditions. *Appl. Microbiol. Biotechnol.* 172, 549–560. doi: 10.1007/s12010-013-0562-y
- Sasnow, S. S., Wei, H., and Aristilde, L. (2016). Bypasses in intracellular glucose metabolism in iron-limited *Pseudomonas putida*. *Microbiologyopen* 5, 3–20. doi: 10.1002/mbo3.287
- Sauer, U., Hatzimanikatis, V., Bailey, J. E., Hochull, M., Szyperksy, T., and Wüthrich, K. (1997). Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. *Nat. Biotechnol.* 15, 448–452. doi: 10.1038/nbt0597-448
- Tannler, S., Decasper, S., and Sauer, U. (2008). Maintenance metabolism and carbon fluxes in *Bacillus* species. *Microb. Cell Fact.* 7:19. doi: 10.1186/1475-2859-7-19
- Vary, P. S. (1994). Prime time for *Bacillus megaterium*. *Microbiology* 140, 1001–1013. doi: 10.1099/13500872-140-5-1001
- Vary, P. S., Biedendieck, R., Furch, T., Meinhardt, F., Rodhe, M., Deckwer, W. D., et al. (2007). *Bacillus megaterium* – from simple soil bacterium to industrial protein production host. *Appl. Microbiol. Biotechnol.* 76, 957–967. doi: 10.1007/s00253-007-1089-3
- Weitzel, M., Nöh, K., Dalman, T., Niedenführ, S., Stute, B., and Wiechert, W. (2012). 13CFLUX2 - high-performance software suite for ¹³C-metabolic flux analysis. *Bioinformatics* 29, 143–145. doi: 10.1093/bioinformatics/bts646
- Wendisch, V. F., Brito, L. F., Lopez, M. G., Hennig, G., Pfeifenschneider, J., Sgobba, E., et al. (2016). The flexible feedstock concept in industrial biotechnology: metabolic engineering of *Escherichia coli*, *Cornibacterium glutamicum*, *Pseudomonas*, *Bacillus* and yeast strains for access to alternative carbon sources. *J. Biotechnol.* 234, 139–157. doi: 10.1016/j.jbiotec.2016.07.022
- Wilkes, R. A., Mendonca, C. M., and Aristilde, L. (2018). Metabolic flux hierarchy prioritizes the entner-doudoroff pathway for carbohydrate co-utilization in *Pseudomonas protegens* PF-5. *Appl. Environ. Microbiol.* 85:e02084-18. doi: 10.1128/AEM.02084-18
- Youngster, T., Wushensky, J. A., and Aristilde, L. (2017). Profiling glucose-induced selective inhibition of disaccharide catabolism in *Bacillus megaterium* QM B1551 by stable isotope labelling. *Microbiology* 163, 1509–1514. doi: 10.1099/mic.0.000540

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

Copyright © 2018 Wushensky, Youngster, Mendonca and Aristilde. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Bacterial Metabolism During Biofilm Growth Investigated by ^{13}C Tracing

Ni Wan¹, Hao Wang², Chun Kiat Ng^{3,4†}, Manisha Mukherjee^{3,4}, Dacheng Ren^{2,5}, Bin Cao^{3,4*} and Yinjie J. Tang^{6*}

¹ Mechanical Engineering and Materials Science, Washington University, St. Louis, MO, United States, ² Department of Biomedical and Chemical Engineering, Syracuse Biomaterials Institute, Syracuse University, Syracuse, NY, United States, ³ School of Civil and Environmental Engineering, Nanyang Technological University, Singapore, Singapore, ⁴ Singapore Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore, Singapore, ⁵ Department of Civil and Environmental Engineering, and Biology, Syracuse Biomaterials Institute, Syracuse University, Syracuse, NY, United States, ⁶ Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO, United States

OPEN ACCESS

Edited by:

Weiwen Zhang,
Tianjin University, China

Reviewed by:

Saori Kosono,
The University of Tokyo, Japan
Chao Wu,
National Renewable Energy
Laboratory (DOE), United States

*Correspondence:

Bin Cao
BinCao@ntu.edu.sg
Yinjie J. Tang
yinjie.tang@seas.wustl.edu

† Present address:

Chun Kiat Ng,
Department of Engineering Science,
University of Oxford, Oxford,
United Kingdom

Specialty section:

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

Received: 01 June 2018

Accepted: 17 October 2018

Published: 20 November 2018

Citation:

Wan N, Wang H, Ng CK,
Mukherjee M, Ren D, Cao B and
Tang YJ (2018) Bacterial Metabolism
During Biofilm Growth Investigated by
 ^{13}C Tracing. *Front. Microbiol.* 9:2657.
doi: 10.3389/fmicb.2018.02657

This study investigated the metabolism of *Pseudomonas aeruginosa* PAO1 during its biofilm development via microscopy imaging, gene expression analysis, and ^{13}C -labeling. First, dynamic labeling was employed to investigate glucose utilization rate in fresh biofilms (thickness 40~60 micrometer). The labeling turnover time of glucose-6-P indicated biofilm metabolism was substantially slower than planktonic cells. Second, PAO1 was cultured in continuous tubular biofilm reactors or shake flasks. Then ^{13}C -metabolic flux analysis of PAO1 was performed based on the isotopomer patterns of proteinogenic amino acids. The results showed that PAO1 biofilm cells during growth conserved the flux features as their planktonic mode. (1) Glucose could be degraded by two cyclic routes (the TCA cycle and the Entner-Doudoroff-Embden-Meyerhof-Parnas loop) that facilitated NAD(P)H supplies. (2) Anaplerotic pathways (including pyruvate shunt) increased flux plasticity. (3) Biofilm growth phenotype did not require significant intracellular flux rewiring (variations between biofilm and planktonic flux network, normalized by glucose uptake rate as 100%, were less than 20%). (4) Transcription analysis indicated that key catabolic genes in fresh biofilm cells had expression levels comparable to planktonic cells. Finally, PAO1, *Shewanella oneidensis* (as the comparing group), and their c-di-GMP transconjugants (with different biofilm formation capabilities) were ^{13}C -labeled under biofilm reactors or planktonic conditions. Analysis of amino acid labeling variances from different cultures indicated *Shewanella* flux network was more flexibly changed than PAO1 during its biofilm formation.

Keywords: c-di-GMP, dynamic labeling, Entner-Doudoroff pathway, pyruvate shunt, tubular biofilm reactors

INTRODUCTION

Biofilm is a heterogeneous and dynamic system. Its development consists of steps of adhesion of planktonic microbes, colony formation and growth, and detachment/migration of dispersed cells to new surfaces. Moreover, cells at different locations inside a biofilm may have distinct metabolisms (e.g., different transcriptomic and proteomic profiles) due to intrinsic chemical gradients (Williamson et al., 2012). The physiological differences between biofilm and planktonic cells have attracted extensive studies (O'Toole et al., 2000; Bester et al., 2005). To quantify biofilm

physiologies, diverse technologies including crystal violet assay, transcription/protein/metabolite analyses, and imaging (e.g., SEM, TEM, confocal microscopy) have been applied (Pantarella et al., 2013). Moreover, genetic mutations are used to reveal regulatory mechanisms of cell survival in various biofilm environments (Ding et al., 2014; Zhang et al., 2014). However, there is still little knowledge of metabolic fluxomes that describe *in vivo* enzyme activities inside biofilm cells for carbon/energy metabolism.

To decipher flux distributions in biofilm cells, the present study investigated the opportunistic pathogen *Pseudomonas aeruginosa* PAO1 for its metabolic functions under both planktonic and biofilm modes. Particularly, ^{13}C -fingerprinting of proteinogenic amino acids was used to trace carbon fluxes for substrate utilization and biomass synthesis. In parallel, dynamic labeling via ^{13}C -glucose pulses was used to reveal the speed of ^{13}C percolating through central pathways in fresh biofilms as well as planktonic cells. This study also examined the c-di-GMP transconjugant of PAO1 via ^{13}C -fingerprinting. The transconjugant overexpressed c-di-GMP and produced excess extracellular polymer substances (EPS) to enhance the biofilm formation (Chua et al., 2015). To broaden our perception of the degree of flux profile conservations between planktonic and biofilm cells, the same isotopic approaches were also used to investigate *Shewanella oneidensis* MR-1 (a metal reducing bacterium capable of proliferating in both aerobic and anaerobic conditions) (Tang et al., 2007). The outcomes improved our understanding of the mechanisms about how bacterial species reorganized their flux network during biofilm development.

MATERIALS AND METHODS

Strains and Cultivations

Pseudomonas aeruginosa PAO1 and its c-di-GMP transconjugants (i.e., a high c-di-GMP transconjugant with twice more biofilm formation and a low c-di-GMP transconjugant with reduced biofilm formation by ~30%) were grown in an M9 medium using 20 mM glucose. *S. oneidensis* MR-1 and its high c-di-GMP transconjugant for enhanced biofilm formation, as additional tests, were grown in a modified MR-1 medium (Cao et al., 2011) using 20 mM sodium lactate. For planktonic cultures, bacteria (20 mL) were grown in flasks (150 mL) with inoculation volume ratio of 0.5% (at the room temperature, shaking at 200 rpm). To produce sufficient biofilm biomass for metabolic flux analysis (Figure 1), PAO1 or MR-1 was grown in tubular biofilm reactors (sets of 20 cm long O_2 -permeable silicon tubing with an inner diameter of 3 mm) at the room temperature (Ding et al., 2014), where the respective media were continuously pumped through the tubular reactor by Low-Speed Digital Peristaltic Pump system (Cole-Parmer, Singapore) (Sternberg and Tolker-Nielsen, 2006). Each tubular reactor was inoculated by injecting diluted planktonic culture using a syringe and resulted in initial OD_{600} of ~0.01. After inoculation, the media flow was stopped for 1 h to allow initial attachment followed by continuous media flow with a flow rate of 6 mL/h. The biofilm

reactor had a pseudo-steady state operation for 3 days and the average wet biomass generation rate was ~0.03 g/day/reactor.

^{13}C -Fingerprinting Amino Acids to Trace Flux Distributions

For labeled experiments, 20 mM [$1,2\text{-}^{13}\text{C}$] glucose was used for cultivating both PAO1 and its c-di-GMP transconjugants, while [$3\text{-}^{13}\text{C}$] sodium lactate was used for cultivating MR-1 and its c-di-GMP transconjugant. In planktonic mode, pseudo-steady-state shake flask cultures were harvested by centrifugation during mid-exponential phases. Cell pellets and supernatant were stored at -20°C before further analysis. For biofilm mode, ^{13}C -labeled biomass in the tubular reactor was squeezed out for amino acid analysis. For labeled experiments, substrate concentrations (including glucose, lactate, and acetate) were measured using HPLC (Sivakumar et al., 2014). EPS formation was also determined (Jiao et al., 2010). To analyze proteinogenic amino acids, biomass pellets were hydrolyzed by 6 M HCl at 100°C , then air-dried and derivatized with *N*-tert-butyldimethylsilyl-*N*-methyltrifluoroacetamide (TBDMS) prior to GC-MS measurement (You et al., 2012). A published software was used to correct MS peaks (i.e., [M-57] and [M-159]) (Wahl et al., 2004). Mass isotopomer distributions (MID) (M0, M1, M2...) represent fragments with (0, 1, 2...) labeled carbons in amino acids. Due to overlapping peaks or product degradation, proline, arginine, cysteine, and tryptophan were not analyzed (Antoniewicz et al., 2007).

Biofilm Imaging and Viability Analysis

Fresh PAO1 cells were grown on glass slides (1 cm^2) for biofilm imaging and viability analyses. Briefly, PAO1 overnight cultures were used to inoculate Petri dishes containing M9 medium supplemented with 1 g/L unlabeled glucose. Biofilm cultures were incubated for 96 h (replacing spent medium with fresh M9 medium containing 1 g/L glucose every 48 h). After washing with PBS buffer, glass slides with attached biofilms (thickness 40~60 μm) were transferred into new Petri dishes containing fresh M9 medium. To observe the attachment/settlement of planktonic cells on biofilms, biofilm cells were stained using SYTO 9 green fluorescent nucleic acid stain (Thermo Fisher Scientific, Waltham, MA, United States), then PAO1 planktonic cells (OD_{600} 0.7~0.8) stained by orange dye Alexa Fluor 555 (Thermo Fisher Scientific, Waltham, MA, United States) were added into Petri dishes and incubated with biofilm slides for 1 h. The resulting biofilm was imaged using an Axio Imager M1 fluorescence microscope (Carl Zeiss, Inc., Germany) (note: the green color represents biofilm cells and orange color represents planktonic cells settled on slide surface). For parallel samples, live/dead staining images of PAO1 biofilm were also collected, where biofilm slides were stained with SYTO 9 (green) and propidium iodide (red) for 15 min at the room temperature before imaging (note: green stains all cells; while red indicates DNA in dead cells or extracellular DNA).

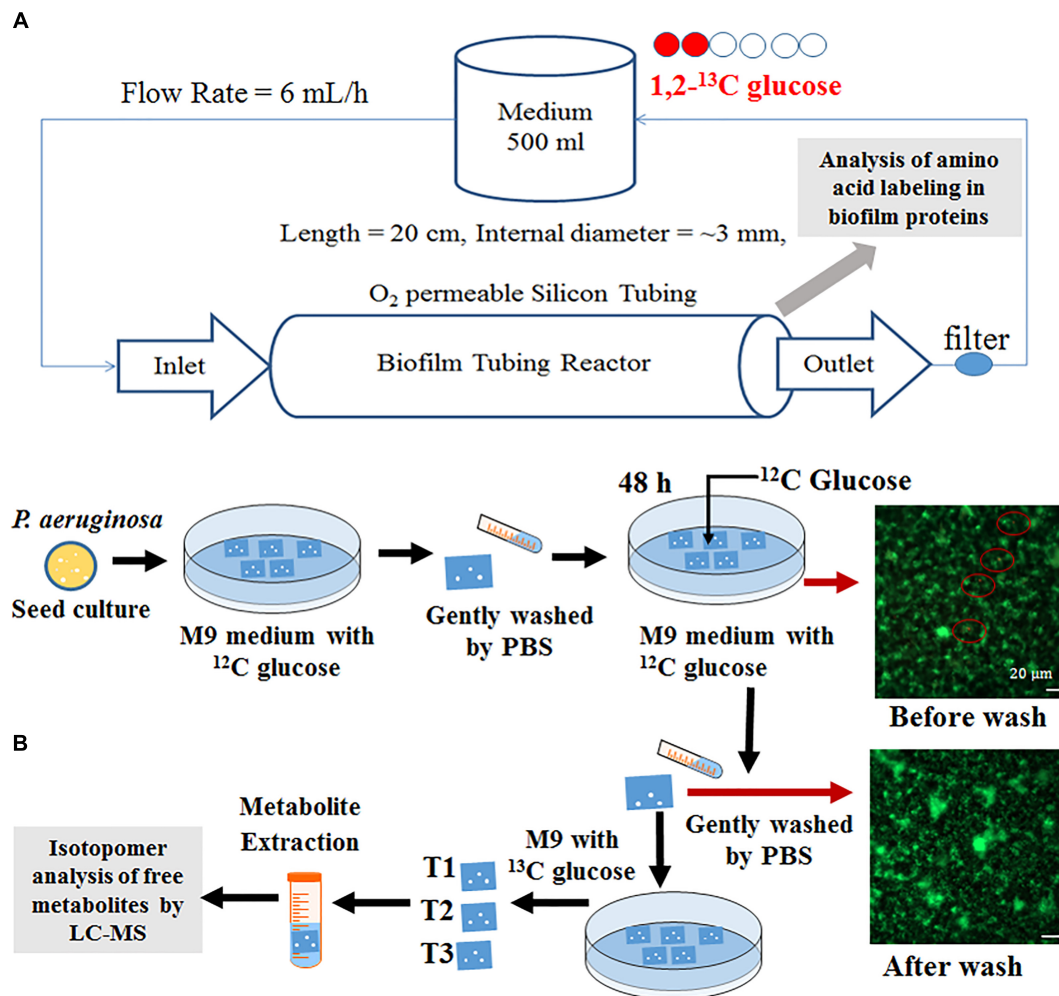
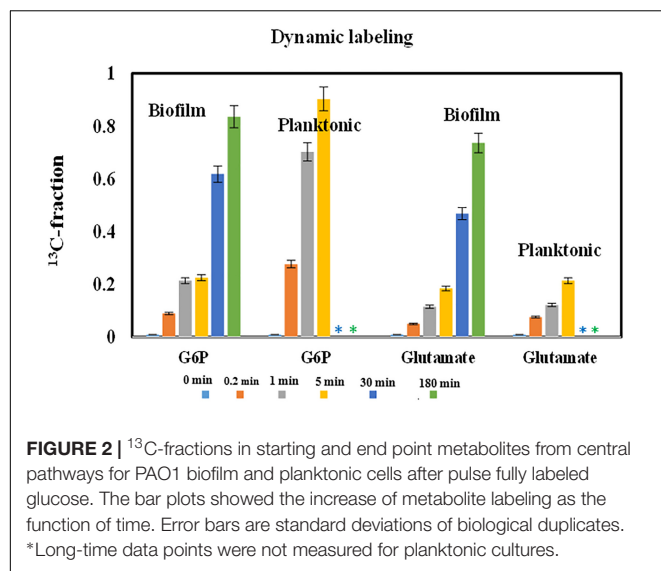


FIGURE 1 | Experimental platform schematic illustration of (A) tubular biofilm reactor and (B) dynamic labeling in Petri dish system. Microscopic analysis showed majority cells on glass slides before labeling experiments were biofilm (in green) rather than planktonic cells (in orange, highlighted by red circles).

Comparison of Glucose Catabolic Rates in PAO1 Cultures Using Dynamic ¹³C Labeling

Glucose uptake in planktonic and biofilm PAO1 were measured by tracking ¹³C incorporation rates of two key metabolites (glucose-6-P and glutamate) after pulsing fully labeled glucose into unlabeled cultures at the room temperature. For planktonic ¹³C-experiments, PAO1 was grown in shake flasks with 1 g/L unlabeled glucose. Once cells reached late exponential phase (OD₆₀₀ 0.7~0.8) and ~90% non-labeled glucose was consumed, fully labeled ¹³C glucose was added into the culture with final concentration of 2 g/L. After ¹³C-glucose additions, 15 mL of cell cultures were harvested by mixing cultures with 5 mL pre-cold M9-ice solutions at four sampling points (0, 0.2, 1, and 5 min). The samples were further quenched with ethanol-dry ice bath (−70°C) to reduce culture temperature to ~0°C. Samples (with ice particles) were centrifuged at 8,000 rpm for 1 min and the pellets were kept at −20°C before LC-MS

measurement. For dynamic ¹³C-experiments on biofilm, fresh PAO1 biofilm cells were prepared using glass slides (same as that for cell imaging). Before labeling experiments, glass slides with fresh unlabeled biofilm cells were washed by phosphate-buffered saline (PBS, 1X) buffer then soaked in 25 mL M9 medium containing 1 g/L fully labeled ¹³C-glucose for 0.2, 1, 5, 30, and 180 min. To harvest time-course samples, glass slides were placed in PBS-ice solution to quench cell metabolisms. Free metabolites were measured by LC-MS (Hollinshead et al., 2016). Briefly, quenched planktonic or biofilm cells were placed in cold methanol/chloroform solution (7:3 v/v) and shaken at 150 rpm at 4°C overnight. Deionized water was added to the solvent mix to extract cell metabolites. The aqueous phase was filtered through an Amicon Ultra centrifuge filter (3000 Da; EMD Millipore, Billerica, MA, United States) then lyophilized. The dried samples were dissolved in acetonitrile and water (6:4, v/v) solution for LC-MS analysis (Agilent Technologies 1200 Series equipped with a SeQuant Zic-pHILIC



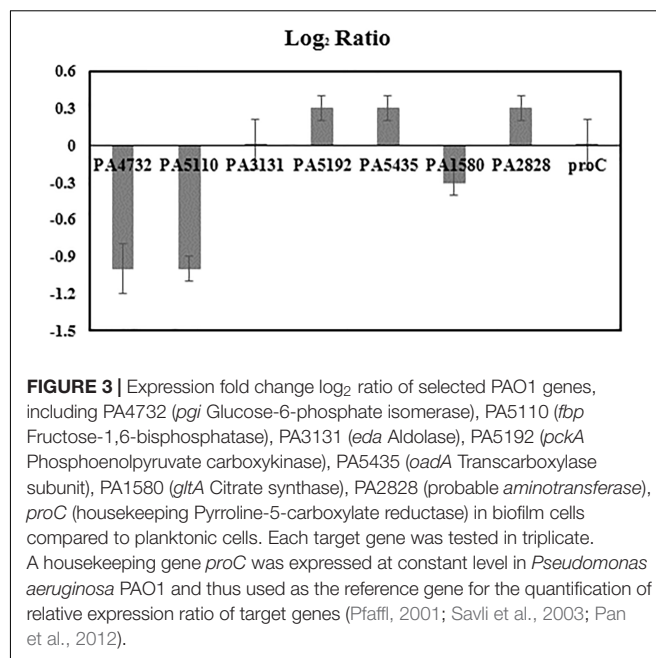
column) to determine MS distributions of targeted metabolites (Figure 2).

Gene Expressions in Fresh Biofilm Cells

The qPCR was used to compare the expressions of glycolytic pathway genes between fresh biofilm cells from glass slides and planktonic cells. The protocol has been reported in our previous research (Choudhary et al., 2015). Generally, the cDNA was synthesized from the isolated RNA samples of PAO1 planktonic cells and glass slide biofilms using iScript cDNA Synthesis Kit (Biorad, Hercules, CA, United States). The primers were designed within primer blast (NCBI). The qPCR samples were prepared by mixing cDNA, primers, and iTaqTM universal SYBR Green Supermix (Biorad, Hercules, CA, United States). The qPCR reactions were accomplished with an Eppendorf Mastercycler Realplex thermal cycler (Eppendorf, Hauppauge, NY, United States). The condition of qPCR reactions was: heat activation at 95°C for 1 min, 40 cycles of denaturation at 95°C for 10 s, and annealing/extension at 60°C for 1 min. The melting curve was set at 95°C for 30 s, 45°C for 30 s, 20 min hold with temperature gradient, and 95°C for 1 min. The relative expression ratios of the selected genes were analyzed using the LinRegPCR program (Heart Failure Research Center, Netherlands) and equation below (Pfaffl, 2001):

$$\text{Log}_2\text{Ratio} = \text{Log}_2 \left[\frac{\Delta C_q^{\text{Target}} (\text{Planktonic} - \text{Biofilm})}{\Delta C_q^{\text{Reference}} (\text{Planktonic} - \text{Biofilm})} \right]$$

ΔC_q represented the difference in value of quantitation cycle between planktonic and biofilm samples. E described the qPCR efficiency. Both ΔC_q and E were calculated by the LinReg PCR program based on the raw data of qPCR experiments. The target samples were seven selected genes (PA4732, PA5110, PA3131, PA5192, PA5435, PA1580, and PA2828) related to glucose metabolism of *P. aeruginosa*



(Figure 3). The reference sample was housekeeping gene *proC* (Savli et al., 2003).

Metabolic Flux Analysis of Planktonic Culture and Tubular Reactor Biofilm Cells

^{13}C -MFA was performed based on isotopomer data from proteinogenic amino acids from biofilm reactor and shake flask cultures (Supplementary Material). The software WUflux (He et al., 2016) was used for flux calculations. Biomass composition was modified based on previous study (Bartell et al., 2016). The MFA model included the EMP (Embden-Meyerhof-Parnas) pathway, the OPP (oxidative pentose phosphate) pathway, the ED (Entner-Doudoroff) pathway, the TCA cycle, the glyoxylate shunt, and biomass synthesis (Stover et al., 2000). Based on KEGG database, PAO1 contains fructose-1,6-bisphosphatase but lacks phosphofructokinase and thus the reaction (F6P → FBP) was deleted from the model. Since the precise measurement of actual glucose utilization for biofilm production was very difficult due to the presence of both planktonic cells and biofilm cells in tubular reactors, ^{13}C -MFA profiled relative fluxes by assuming glucose uptake rate as 100 units. The relative fluxes were solved by minimizing a quadratic error function that calculated the differences between predicted and measured isotopomer patterns ($n = 2$). The confidence intervals of fluxes were estimated as following. The model randomly perturbed both biomass equation for EPS formations by $\pm 10\%$ and amino acid MID data within measurement standard deviations for 500 times to simulate experimental uncertainty. Based on each new dataset, the model re-calculated fluxes. Then confident intervals were estimated based on the variations of resulting fluxes (He et al., 2016).

RESULTS

Dynamic Labeling of Free Metabolites in Planktonic and Biofilm Cells

Some bacterial species favor the growth on solid surfaces, while others favor planktonic mode. Comparisons between biofilm and planktonic cell growths have been extensively

reported (Heffernan et al., 2009). To understand overall PAO1 biofilm physiology, we prepared fresh PAO1 biofilms on glass slides. Before pulsing ^{13}C -glucose for the dynamic labeling of biofilm cells, we washed glass slides to remove planktonic cells attached on the biofilm surface. Fluorescence microscope imaging confirmed that few planktonic cells (pre-labeled with orange dye) remained on the biofilm surface (in green color) (Figure 1). These biofilm cells on glass slides could be easily

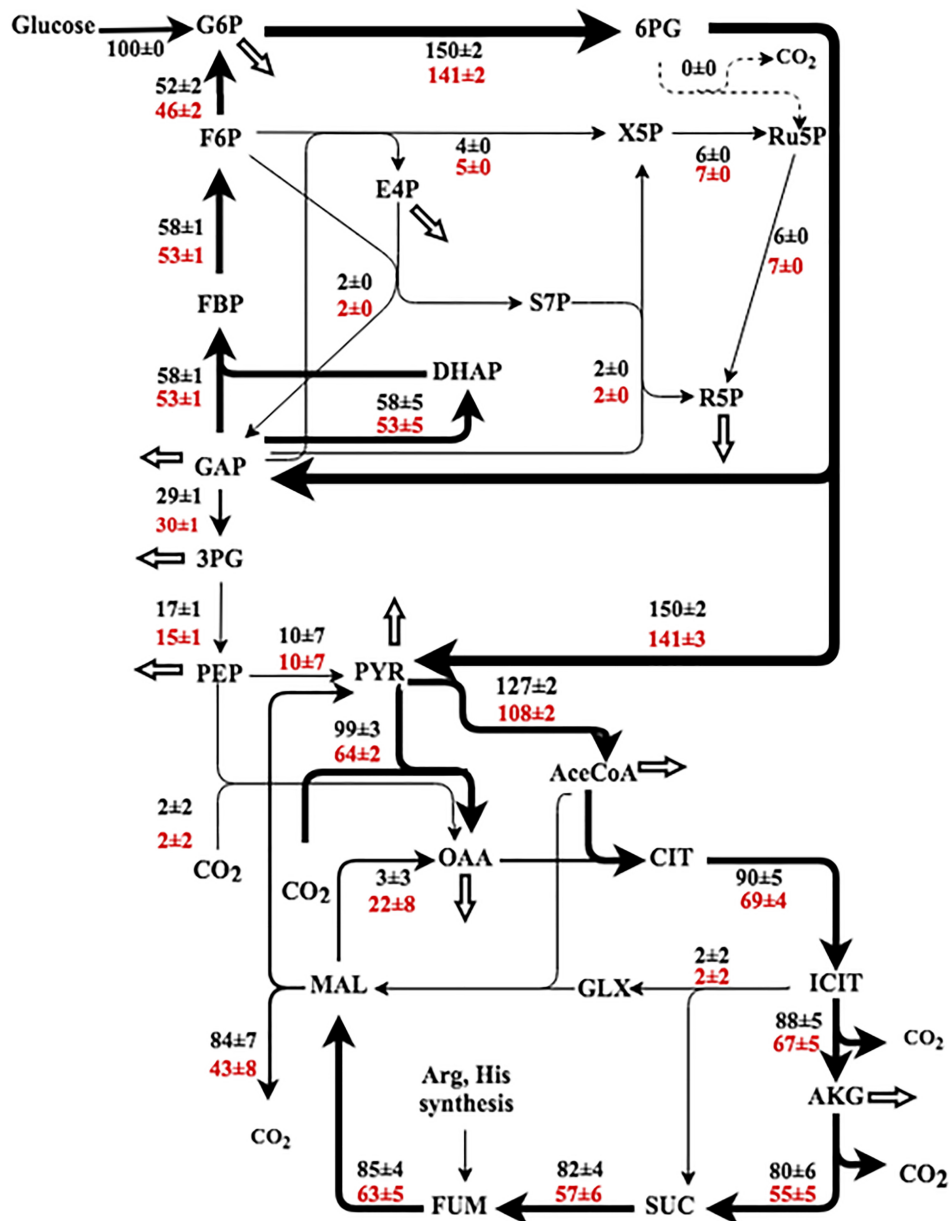


FIGURE 4 | Flux ratio of *Pseudomonas aeruginosa* as planktonic (black) and biofilm (red). The fluxes were normalized to the glucose uptake rate (represented as 100), and the fluxes are represented as 'best fit \pm confidence intervals' based on the measured isotopomer distributions (biological duplicates). The arrow thickness relates to the magnitude of flux. The white arrows represent the fluxes toward biomass synthesis. 3PG, 3-phosphoglycerate; 6PG, 6-phosphogluconate; AceCoA, acetyl-CoA; DHAP, dihydroxyacetone phosphate; E4P, erythrose 4-phosphate; FBP, fructose 1,6-bisphosphate; F6P, fructose 6-phosphate; G6P, glucose 6-phosphate; GAP, glyceraldehyde 3-phosphate; GLX, glyoxylate; ICT, isocitrate; MAL, malate; OAA, oxaloacetate; PEP, phosphoenolpyruvate; PYR, pyruvate; R5P, ribose 5-phosphate; Ru5P, ribulose-5-phosphate; RuBP, ribulose-1,5-diphosphate; S7P, sedoheptulose-7-phosphate; SUC, succinate; X5P, xylulose-5-phosphate.

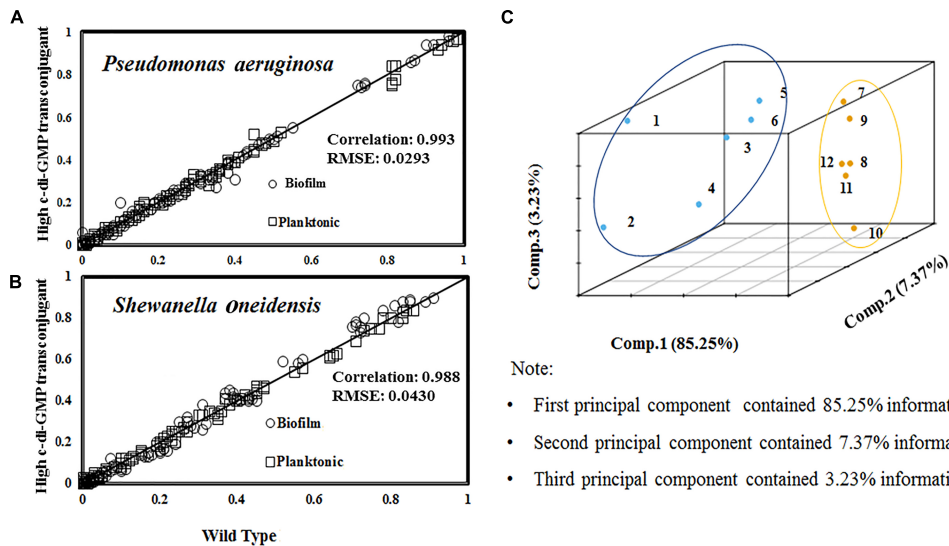


FIGURE 5 | Correlation of amino acids labeling between wild type and c-di-GMP transconjugants (represented by R^2). The RMSE was calculated based on the variations of amino acid labeling data between planktonic and biofilm samples from the same strain. **(A)** Comparison of MID from PAO1 and high c-di-GMP transconjugant samples under planktonic and biofilm modes (grown with [1,2- 13 C] glucose). **(B)** Comparison of MID from MR-1/high c-di-GMP transconjugant samples under planktonic and biofilm modes (grown with [3- 13 C] lactate). **(C)** Principal component analysis (PCA) of amino acid labeling for 12 conditions. **Condition 1~6 for MR-1:** (1) planktonic cells at the mid-log phase; (2) high c-di-GMP transconjugant planktonic cells at the mid-log phase; (3) MR-1 in tubular biofilm reactor at day 3; (4) MR-1 in tubular biofilm reactor at day 4; (5) high c-di-GMP transconjugant in tubular biofilm reactor at day 3; (6) high c-di-GMP transconjugant in tubular biofilm reactor at day 4. **Condition 7~12 for PAO1:** (7) planktonic cells at the mid-log phase; (8) high c-di-GMP transconjugant planktonic cells at the mid-log phase; (9) low c-di-GMP planktonic cells at the mid-log phase; (10) PAO1 in the tubular biofilm reactor; (11) high c-di-GMP transconjugant in the tubular biofilm reactor; (12) low c-di-GMP transconjugant in the tubular biofilm reactor. Principal Component Analysis was generated by R (version 3.2.2) for 12 biofilm/planktonic conditions. Principal components PC1, PC2 and PC3 were included in this study.

sampled and quenched for fast turnover metabolite analysis or cell imaging. Here, dynamic labeling technique was used to measure metabolite turnover rates in biofilm cells from glass slides, which were then compared with shake flask cultures. **Figure 2** showed labeling rates for two key metabolites after 13 C-glucose was pulsed into biofilm or planktonic cells. As expected, labeling rates of G6P (first metabolic node after glucose uptake) for planktonic cells were much faster than biofilm cells, and the 13 C enrichment reached saturation within 5 min. However, it took 180 min for G6P labeling to reach saturation in biofilm cells. Interestingly, final labeling percentages of G6P reached > 85% in biofilm cells, indicating that the majority of biofilm cells were metabolically active for glucose utilizations despite the slow rate. Spatial stratification of oxygen and glucose within the biofilm was a possible explanation. Moreover, free glutamate (the key downstream product from the TCA cycle for biomass synthesis) from both planktonic cells and biofilm cells were labeled much slower than that of G6P (20~25% after 5 min). This observation could be explained by the fact that metabolite turnover rates in amino acid synthesis pathways were much slower than the glucose uptake rates under both biofilm and planktonic modes.

Fluxomes of Planktonic and Biofilm *Pseudomonas* Cells

Planktonic fluxes in *P. aeruginosa* have been reported (Berger et al., 2014; Lassek et al., 2016; Opperman and Shachar-Hill,

2016). These studies highlighted the glucokinase (phosphorylate glucose to G6P then to 6PG) and ED pathways that are mainly responsible for glucose catabolism. The magnitude of fluxes through the oxidative pentose phosphate pathway, glyoxylate shunt, and the TCA cycle varied among different reports. This study examined the *P. aeruginosa* metabolism in both planktonic and biofilm modes at the room temperature. By cultivation with [1, 2- 13 C] labeled glucose in tubular reactors, the resulting proteinogenic amino acids stored labeling information (i.e., 13 C-fingerprinting) that could be used for 13 C-MFA. In contrast to dynamic labeling experiments (i.e., G6P turnover rates) that showed overall metabolic rates in the biofilm were much slower than their free-floating counterparts, metabolic flux distributions had smaller variations between biofilm and planktonic modes after normalizing glucose uptake as 100%. Most flux values in biofilm cells differed within 20% compared to planktonic cells. For both cultivation modes, flux network showed a complete carbohydrate degradation loop: Entner-Doudoroff-Embden-Meyerhof-Parnas (EDEMP) cycle (G6P→6PG→GAP→F6P→G6P) (**Figure 4**), possibly due to metabolic congestion at the lower segment of glycolysis. Compared to biofilm cells, planktonic cells had moderately higher fluxes through the TCA cycle. Similar EDEMP cycle has been observed in *Pseudomonas putida* (Nikel et al., 2015). *Pseudomonas* is well-known for using the ED pathway rather than the EMP for the glucose catabolism due to the absence of phosphofructokinase (Berger et al., 2014). The ED pathway

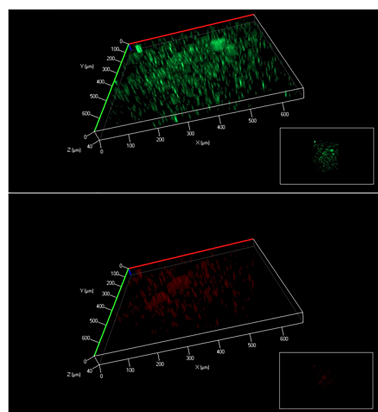


FIGURE 6 | Biofilm cell viability analysis (above: all cells; below: dead cells). The biofilm thickness was about 40–60 μm .

is not beneficial to ATP generation, but it reduces metabolic cost for enzyme synthesis (Stettner and Segrè, 2013). More importantly, the formation of EDMP cycle could improve NADPH generation to diminish oxidative stress and to promote the biosynthesis of C6 sugar phosphates (the precursor of EPS). The TCA cycle in *Pseudomonas* species was reported to operate with the pyruvate shunt, which was catalyzed by malic enzyme and pyruvate carboxylase (malate \rightarrow pyruvate \rightarrow OAA) (Fuhrer et al., 2005; del Castillo et al., 2007). Same pyruvate shunt was observed in both PAO1 planktonic and biofilm cells (Figure 4). For example, very little malate dehydrogenase flux was observed in planktonic cells, and a significant amount of OAA was synthesized from pyruvate. The pyruvate shunt coupled with other anaplerotic pathways (including glyoxylate shunt) could regulate fluxes between glycolysis nodes (PEP and Pyruvate) and the TCA nodes (Malate and OAA) to increase flux network plasticity.

The variations of flux network between planktonic and biofilm cells were further investigated via qPCR analysis. We compared the expression levels of seven key genes related to glucose metabolism (including *pgi*, *fbp*, *edaB*, *pckA*, *oada*, *gltA*, *aminotransferase*) along with the housekeeping gene *proC*. According to the results shown in Figure 3, *proC* was not differentially expressed under the two culture modes, which was consistent with a previous research (Savli et al., 2003). The qPCR results also indicated that expression levels of all selected genes between biofilm and planktonic samples have relatively small differences (note: less than twofold). This result, though incomplete to reflect global genetic regulations, suggested PAO1 could maintain normal functions of many central genes for glucose catabolism during the active biofilm growth phase.

¹³C Fingerprinting of the PAO1 Transconjugant and *Shewanella* Under Planktonic and Biofilm Conditions

We examined the transconjugants (i.e., high or low c-di-GMP expressions) of PAO1 via ¹³C-labeling of proteinogenic amino

acids from tubular reactors or shake flask cultures. The ¹³C-fingerprints (MID of amino acid labeling) of PAO1 and high c-di-GMP transconjugant were collected from planktonic cultures and biofilm reactors and plotted in Figure 5A. Compared to PAO1 wild type, its high c-di-GMP transconjugant produced 1.9-fold more EPS and twice more biofilm in tubular reactors. Moreover, using labeling data of PAO1 planktonic culture as the baseline, MID data were found to have high correlations ($R^2 = 0.99$) correlations between PAO1 and its transconjugant samples from planktonic and biofilm cultures (Figure 5A). This observation inferred that the mutant and the wild type shared similar flux distributions (i.e., change of planktonic or biofilm growth rate does not require significant intracellular flux rewiring). To obtain a broader understanding of flux regulations, similar ¹³C-fingerprinting experiments on *S. oneidensis* MR-1 and its high c-di-GMP transconjugant were performed. The MID of proteinogenic amino acids also demonstrated strong correlations ($R^2 = 0.99$) among MR-1 and its c-di-GMP transconjugant (Figure 5B). However, the root-mean-square error (RMSE) of labeling data variations between planktonic and biofilm cells in MR-1 was 1.5-fold higher than the RMSE obtained from PAO1 cultures (Figure 5A). Further principal component analysis (PCA) examined MID (as the features) of amino acids from different ¹³C-cultures (planktonic or biofilm cultures of PAO1, MR-1, and their transconjugants) (Figure 5C). Both the RMSE and PCA results indicated that the MR-1 metabolism could be more affected by its biofilm growth mode than the PAO1. This observation (i.e., MR-1 flux network was more flexible) was consistent to the reported versatility of MR-1 metabolisms (Guo et al., 2015). For example, O₂ conditions could influence acetate overflows and intracellular fluxome in MR-1 (Tang et al., 2007). Nevertheless, different bacteria may have different capabilities for minimizing the change of flux network when cells switch from planktonic to biofilm growth.

DISCUSSION

There is a consensus that cell attachment onto surfaces strongly influences microbial metabolism. For example, *P. aeruginosa* displays phenotypic changes during biofilm development (Sauer et al., 2002). Because of temporal and structural variations, conflicting observations have been reported on biofilm growth kinetics and metabolic activities compared to free-floating cells (van Loosdrecht et al., 1990; Heffernan et al., 2009). In this study, glucose uptake by fresh biofilm cells (based on G6P labeling) was found to be much slower than planktonic cells, while both planktonic and biofilm cells had sluggish glutamate synthesis (Figure 2). Moreover, biofilm cells employed a relatively similar flux network as planktonic cultures: PAO1 glucose catabolism was mainly dependent on the EDMP/TCA loops, pyruvate shunt, and several anaplerotic pathways. Meanwhile, expression levels of essential genes in PAO1 central pathways were analyzed and no target gene in glucose catabolism was highly up-regulated or down-regulated (Log₂ ratio of 2 as the cutoff, Figure 3) between planktonic and fresh biofilm cells (note: only two genes in glycolysis, *pgi* and *fbp* appeared to be moderately repressed in

biofilms compared to planktonic cells). The gene expressions in fresh biofilm cells indicated that their metabolism could maintain stable catabolic functions. These biofilm metabolic features could be explained by three reasons. First, our cultivations offered optimal biofilm growth and minimized biofilm heterogeneity. For example, majority of cells were alive in the freshly prepared biofilm (thickness of only 40~60 μm) on glass slides (dominating green signals compared to red signals in **Figure 6**), while the use of silicone tubing bioreactor improved oxygen and nutrient transports for biofilm biomass generation. Second, cells located in the peripheral layers of biofilms might contribute significantly to biofilm growth since these cells received nutrients at a level similar to that of planktonic cells. Third, bacterial metabolism inherently demonstrated robust ratios for resource allocations.

In a broader perspective, bacterial flux networks are not straightforwardly correlated with gene expressions (Chubukov et al., 2013) or proteomic profiles (Lassek et al., 2016). Although bacterial physiologies are sensitive to nutrient and growth conditions, flux ratios/network may demonstrate small perturbations or certain rigidity against genetic and environmental changes (Fischer and Sauer, 2005). For example, bacterial flux distribution under salt stresses could remain the same as normal growth conditions, which was in stark contrast to slower growth rate and high changes of transcript profiles (Tang et al., 2009). The conservation of microbial fluxomics (i.e., metabolic robustness) is regarded as the principle of how cell metabolism distributes resources for biomass growth, while microbial species may demonstrate different degrees of flux conservations during their biofilm growth.

The methods and observations in this study still have limitations. First, the variation in growth conditions and surface materials from different lab cultures may influence cell metabolisms. Second, a biofilm culture includes at least three sub-populations (planktonic cells, fast growing biofilm cells, and dormant/dead biofilm cells in deep layers, as shown in **Figure 6**). ^{13}C -fingerpringing of proteinogenic amino acids could only track these actively growing cells (i.e., on the top of the biofilms or deposited from planktonic phase) that consumed major nutrient resources for biomass synthesis. This approach failed to provide unique insights into the metabolic topology or flux network plasticity for these dormant/slow-growth biofilm cells under environmental stresses. To further reveal metabolic activities in heterogeneous biofilm, new tools (such as population snapshot measurements by cell sorting) are required to integrate with ^{13}C -labeling techniques. Some cell patterning technologies may also be adapted to obtain biofilms with well-defined structures (thus reduced heterogeneity) to allow better understanding of biofilm metabolism (Ren et al., 2012, 2013a,b; Gu et al., 2013). This is part of our ongoing work.

REFERENCES

- Antoniewicz, M. R., Kelleher, J. K., and Stephanopoulos, G. (2007). Accurate assessment of amino acid mass isotopomer distributions for metabolic flux analysis. *Anal. Chem.* 79, 7554–7559.
- Bartell, J. A., Blazier, A. S., Yen, P., Thøgersen, J. C., Jelsbak, L., Goldberg, J. B., et al. (2016). Reconstruction of the metabolic network of *Pseudomonas aeruginosa*

CONCLUSION

The flux network in biofilm cell is not yet well understood. This study elucidated metabolic features of PAO1 biofilm cells via comparative ^{13}C labeling. Bacterial cells within biofilms differ in physiologies because of nutrient and oxygen limitations, but biofilm flux distributions could still show certain degree of invariability. Specifically, PAO1 cells could fairly maintain its flux distributions and gene expressions as its planktonic culture during active biofilm development. To further decipher biofilm metabolism and regulations in different bacterial species, our future work aims to expand metabolite coverage as well as spatial and temporal analysis of biofilm subpopulations.

AVAILABILITY OF DATA AND MATERIALS

All isotopomer data and metabolic reactions are included as supporting information.

AUTHOR CONTRIBUTIONS

YT, DR, and BC initiated the project and designed experiments. NW, HW, CN, and MM performed experiments and modeling analysis. All authors wrote and approved the final version of the manuscript.

FUNDING

This research was supported by US NSF grants (CBET 1700881 and 1706061). This research is also supported by the Ministry of Education (MOE) in Singapore with Academic Research Fund (AcRF) Tier 1 Grant RG50/16 (M4011622.030 to BC), AcRF Tier 2 Grant (MOE2017-T2-2-042 to BC), and the National Research Foundation under its Research Centre of Excellence Program (Centre for Environmental Life Sciences Engineering, Nanyang Technological University, Singapore).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.02657/full#supplementary-material>

to interrogate virulence factor synthesis. *Nat. Commun.* 8:14631. doi: 10.1038/ncomms14631

- Berger, A., Dohnt, K., Tielen, P., Jahn, D., Becker, J., and Wittmann, C. (2014). Robustness and plasticity of metabolic pathway flux among uropathogenic isolates of *Pseudomonas aeruginosa*. *PLoS One* 9:e88368. doi: 10.1371/journal.pone.0088368

- Bester, E., Wolfaardt, G., Joubert, L., Garny, K., and Saftic, S. (2005). Planktonic-cell yield of a pseudomonad biofilm. *Appl. Environ. Microbiol.* 71, 7792–7798.
- Cao, B., Shi, L., Brown, R. N., Xiong, Y., Fredrickson, J. K., Romine, M. F., et al. (2011). Extracellular polymeric substances from *Shewanella* sp. HRCR-1 biofilms: characterization by infrared spectroscopy and proteomics. *Environ. Microbiol.* 13, 1018–1031. doi: 10.1111/j.1462-2920.2010.02407.x
- Choudhary, G. S., Yao, X., Wang, J., Peng, B., Bader, R. A., and Ren, D. (2015). Human granulocyte macrophage colony-stimulating factor enhances antibiotic susceptibility of *Pseudomonas aeruginosa* persister cells. *Sci. Rep.* 5:17315. doi: 10.1038/srep17315
- Chua, S. L., Sivakumar, K., Rytbke, M., Yuan, M., Andersen, J. B., Nielsen, T. E., et al. (2015). C-di-GMP regulates *Pseudomonas aeruginosa* stress response to tellurite during both planktonic and biofilm modes of growth. *Sci. Rep.* 5:10052. doi: 10.1038/srep10052
- Chubukov, V., Uhr, M., Le Chat, L., Kleijn, R. J., Jules, M., Link, H., et al. (2013). Transcriptional regulation is insufficient to explain substrate-induced flux changes in *Bacillus subtilis*. *Mol. Syst. Biol.* 9:709. doi: 10.1038/msb.2013.66
- del Castillo, T., Ramos, J. L., Rodríguez-Herva, J. J., Fuhrer, T., Sauer, U., and Duque, E. (2007). Convergent peripheral pathways catalyze initial glucose catabolism in *Pseudomonas putida*: genomic and flux analysis. *J. Bacteriol.* 189, 5142–5152.
- Ding, Y., Peng, N., Du, Y., Ji, L., and Cao, B. (2014). Disruption of putrescine biosynthesis in *Shewanella oneidensis* enhances biofilm cohesiveness and performance in Cr(VI) immobilization. *Appl. Environ. Microbiol.* 80, 1498–1506. doi: 10.1128/AEM.03461-13
- Fischer, E., and Sauer, U. (2005). Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nat. Genet.* 37, 636–640.
- Fuhrer, T., Fischer, E., and Sauer, U. (2005). Experimental identification and quantification of glucose metabolism in seven bacterial species. *J. Bacteriol.* 187, 1581–1590.
- Gu, H., Hou, S., Yongyat, C., De Tore, S., and Ren, D. (2013). Patterned biofilm formation reveals a mechanism for structural heterogeneity in bacterial biofilms. *Langmuir* 29, 11145–11153. doi: 10.1021/la402608z
- Guo, W., Luo, S., He, Z., and Feng, X. (2015). ¹³C pathway analysis of biofilm metabolism of *Shewanella oneidensis* MR-1. *RSC Adv.* 5, 39840–39843.
- He, L., Wu, S. G., Zhang, M. H., Chen, Y. X., and Tang, Y. J. (2016). WUFlux: an open-source platform for ¹³C metabolic flux analysis of bacterial metabolism. *BMC Bioinformatics* 17:444. doi: 10.1186/s12859-016-1314-0
- Heffernan, B., Murphy, C., and Casey, E. (2009). Comparison of planktonic and biofilm cultures of *Pseudomonas fluorescens* DSM 8341 cells grown on fluoroacetate. *Appl. Environ. Microbiol.* 75, 2899–2907. doi: 10.1128/AEM.01530-08
- Hollinshead, W. D., Rodriguez, S., Martin, H. G., Wang, G., Baidoo, E. E. K., Sale, K. L., et al. (2016). Examining *Escherichia coli* glycolytic pathways, catabolite repression, and metabolite channeling using Δ pfk transconjugants. *Biotechnol. Biofuels* 9:212.
- Jiao, Y., Cody, G. D., Harding, A. K., Wilmes, P., Schrenk, M., Wheeler, K. E., et al. (2010). Characterization of extracellular polymeric substances from acidophilic microbial biofilms. *Appl. Environ. Microbiol.* 76, 2916–2922. doi: 10.1128/AEM.02289-09
- Lassek, C., Berger, A., Zühlke, D., Wittmann, C., and Riedel, K. (2016). Proteome and carbon flux analysis of *Pseudomonas aeruginosa* clinical isolates from different infection sites. *Proteomics* 16, 1381–1385. doi: 10.1002/pmic.201500228
- Nikel, P. I., Chavarria, M., Fuhrer, T., Sauer, U., and de Lorenzo, V. (2015). *Pseudomonas putida* KT2440 metabolizes glucose through a cycle formed by enzymes of the entner-doudoroff, embden-meyerhof-parnas, and pentose phosphate pathways. *J. Biol. Chem.* 290, 25920–25932. doi: 10.1074/jbc.M115.687749
- Opperman, M. J., and Shachar-Hill, Y. (2016). Metabolic flux analyses of *Pseudomonas aeruginosa* cystic fibrosis isolates. *Metab. Eng.* 38, 251–263. doi: 10.1016/j.ymben.2016.09.002
- O'Toole, G., Kaplan, H. B., and Kolter, R. (2000). Biofilm formation as microbial development. *Annu. Rev. Microbiol.* 54, 49–79.
- Pan, J., Bahar, A. A., Syed, H., and Ren, D. (2012). Reverting antibiotic tolerance of *Pseudomonas aeruginosa* PAOI persister cells by (Z)-4-bromo-5-(bromomethylene)-3-methylfuran-2(5H)-one. *PLoS One* 7:e45778. doi: 10.1371/journal.pone.0045778
- Pantarella, F., Valenti, P., Natalizi, T., Passeri, D., and Berlutti, F. (2013). Analytical techniques to study microbial biofilm on abiotic surfaces: pros and cons of the main techniques currently in use. *Ann. Ig.* 25, 31–42. doi: 10.7416/ai.2013.1904
- Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* 29:e45.
- Ren, D., Cui, M., Xia, Y., and You, Z. (2012). Micropatterning and its applications in biomedical research. *Prog. Biochem. Biophys.* 39, 931–944. doi: 10.3724/SP.J.1206.2011.00228
- Ren, D., Xia, Y., Wang, J., and You, Z. (2013a). Micropatterning of single cell arrays using PEG-Silane and Biotin-(Strept)Avidin System by photolithography and chemical vapor deposition. *Sens. Actuators B Chem.* 188, 340–346. doi: 10.1016/j.snb.2013.07.037
- Ren, D., Xia, Y., and You, Z. (2013b). Multiplexed living cells stained with quantum dot bio probes for multiplexed detection of single-cell array. *J. Biomed. Opt.* 18, 1–9. doi: 10.1117/1.JBO.18.9.096005
- Sauer, K., Camper, A. K., Ehrlich, G. D., Costerton, J. W., and Davies, D. G. (2002). *Pseudomonas aeruginosa* displays multiple phenotypes during development as a biofilm. *J. Bacteriol.* 184, 1140–1154.
- Savli, H., Karadenizli, A., Kolayli, F., Gundes, S., Ozbek, U., and Vahaboglu, H. (2003). Expression stability of six housekeeping genes: a proposal for resistance gene quantification studies of *Pseudomonas aeruginosa* by real-time quantitative RT-PCR. *J. Med. Microbiol.* 52, 403–408.
- Sivakumar, K., Wang, V. B., Chen, X., Bazan, G. C., Kjelleberg, S., Loo, S. C. J., et al. (2014). Membrane permeabilization underlies the enhancement of extracellular bioactivity in *Shewanella oneidensis* by a membrane-spanning conjugated oligoelectrolyte. *Appl. Microbiol. Biotechnol.* 98, 9021–9031. doi: 10.1007/s00253-014-5973-3
- Sternberg, C., and Tolker-Nielsen, T. (2006). Growing and analyzing biofilms in flow cells. *Curr. Protoc. Microbiol.* 1B.2.1–1B.2.15.
- Stettner, A. I., and Segrè, D. (2013). The cost of efficiency in energy metabolism. *Proc. Natl. Acad. Sci. U.S.A.* 110, 9629–9630.
- Stover, C. K., Pham, X. Q., Erwin, A. L., Mizoguchi, S. D., Warrenner, P., Hickey, M. J., et al. (2000). Complete genome sequence of *Pseudomonas aeruginosa* PAOI, an opportunistic pathogen. *Nature* 406, 959–964.
- Tang, Y. J., Hwang, J. S., Wemmer, D. E., and Keasling, J. D. (2007). *Shewanella oneidensis* MR-1 fluxome under various oxygen conditions. *Appl. Environ. Microbiol.* 73, 718–729.
- Tang, Y. J., Martin, H. G., Deutschbauer, A., Feng, X., Huang, R., Llorca, X., et al. (2009). Invariability of central metabolic flux distribution in *Shewanella oneidensis* MR-1 under environmental or genetic perturbations. *Biotechnol. Prog.* 25, 1254–1259. doi: 10.1002/btpr.227
- van Loosdrecht, M. C., Lyklema, J., Norde, W., and Zehnder, A. J. (1990). Influence of interfaces on microbial activity. *Microbiol. Mol. Biol. Rev.* 51, 75–87.
- Wahl, S. A., Dauner, M., and Wiechert, W. (2004). New tools for mass isotopomer data evaluation in ¹³C flux analysis: mass isotope correction, data consistency checking, and precursor relationships. *Biotechnol. Bioeng.* 85, 259–268.
- Williamson, K. S., Richards, L. A., Perez-Osorio, A. C., Pitts, B., McInerney, K., Stewart, P. S., et al. (2012). Heterogeneity in *Pseudomonas aeruginosa* biofilms includes expression of ribosome hibernation factors in the antibiotic-tolerant subpopulation and hypoxia-induced stress response in the metabolically active population. *J. Bacteriol.* 194, 2062–2073. doi: 10.1128/JB.00022-12
- You, L., Page, L., Feng, X., Berla, B., Pakrasi, H. B., and Tang, Y. J. (2012). Metabolic pathway confirmation and discovery through ¹³C-labeling of proteinogenic amino acids. *J. Vis. Exp.* 59:e3583.
- Zhang, Y., Ng, C. K., Cohen, Y., and Cao, B. (2014). Cell growth and protein expression of *Shewanella oneidensis* in biofilms and hydrogel-entrapped cultures. *Mol. Biosyst.* 10, 1035–1042. doi: 10.1039/c3mb70520j

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

Copyright © 2018 Wan, Wang, Ng, Mukherjee, Ren, Cao and Tang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Isotope-Assisted Metabolite Analysis Sheds Light on Central Carbon Metabolism of a Model Cellulolytic Bacterium *Clostridium thermocellum*

Wei Xiong*, Jonathan Lo, Katherine J. Chou, Chao Wu, Lauren Magnusson, Tao Dong and PinChing Maness*

National Renewable Energy Laboratory, Golden, CO, United States

OPEN ACCESS

Edited by:

Ivan A. Berg,
Universität Münster, Germany

Reviewed by:

Wolfgang Buckel,
Philipps-Universität Marburg,
Germany
Wolfgang Eisenreich,
Technische Universität München,
Germany

*Correspondence:

Wei Xiong
Wei.Xiong@nrel.gov
PinChing Maness
PinChing.Maness@nrel.gov

Specialty section:

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

Received: 23 May 2018

Accepted: 31 July 2018

Published: 23 August 2018

Citation:

Xiong W, Lo J, Chou KJ, Wu C,
Magnusson L, Dong T and Maness P
(2018) Isotope-Assisted Metabolite
Analysis Sheds Light on Central
Carbon Metabolism of a Model
Cellulolytic Bacterium *Clostridium*
thermocellum.
Front. Microbiol. 9:1947.
doi: 10.3389/fmicb.2018.01947

Cellulolytic bacteria have the potential to perform lignocellulose hydrolysis and fermentation simultaneously. The metabolic pathways of these bacteria, therefore, require more comprehensive and quantitative understanding. Using isotope tracer, gas chromatography-mass spectrometry, and metabolic flux modeling, we decipher the metabolic network of *Clostridium thermocellum*, a model cellulolytic bacterium which represents as an attractive platform for conversion of lignocellulose to dedicated products. We uncover that the Embden–Meyerhof–Parnas (EMP) pathway is the predominant glycolytic route whereas the Entner–Doudoroff (ED) pathway and oxidative pentose phosphate pathway are inactive. We also observe that *C. thermocellum*'s TCA cycle is initiated by both *Si*- and *Re*-citrate synthase, and it is disconnected between 2-oxoglutarate and oxaloacetate in the oxidative direction; *C. thermocellum* uses a citramalate shunt to synthesize isoleucine; and both the one-carbon pathway and the malate shunt are highly active in this bacterium. To gain a quantitative understanding, we further formulate a fluxome map to quantify the metabolic fluxes through central metabolic pathways. This work represents the first global *in vivo* investigation of the principal carbon metabolism of *C. thermocellum*. Our results elucidate the unique structure of metabolic network in this cellulolytic bacterium and demonstrate the capability of isotope-assisted metabolite studies in understanding microbial metabolism of industrial interests.

Keywords: ^{13}C -isotope tracer, cellulolytic bacteria, citrate synthase, glycolytic pathways, isoleucine biosynthesis, metabolic flux analysis, tricarboxylic acid cycle

INTRODUCTION

Lignocellulosic biomass is the most abundant bio-renewable resource on earth. It has the potential to substitute the current sugar feedstocks for sustainable biofuels and biochemicals production. The use of lignocellulosic biomass has significant merits including improving net carbon and energy balances, reducing production cost, and avoiding the competition between food vs. fuel (Lynd et al., 2005, 2008). However, effective application of lignocellulose in bio-refineries is currently impeded by biomass recalcitrance: its inherent resistance to depolymerization. A potential solution is consolidated bioprocessing (CBP), in which microbes of outstanding cellulolytic ability are recruited for combined cellulose hydrolysis and fermentation without extra addition of cellulases (Lynd et al., 2005). *Clostridium thermocellum* is among the most attractive CBP host microbes.

C. thermocellum is a Gram-positive, thermophilic, and strict anaerobic bacterium. By taking advantage of an extracellular cellulase system called the cellulosome (Gold and Martin, 2007), *C. thermocellum* can degrade cellulose into soluble oligosaccharides. The latter can be further utilized by the cells and fermented to products of industrial interest (e.g., H₂, formate, lactate, acetate, ethanol) (Ellis et al., 2012; Holwerda et al., 2014). Recent advances in genetic modification tools (Tripathi et al., 2010; Argyros et al., 2011) further enable heterologous expression of new functional pathways (e.g., pentose sugar utilization Xiong et al., 2018, isobutanol production Lin et al., 2015) into the bacterium, making this CBP bacterium an attractive and engineerable platform for biofuels or biochemicals production.

Applications of *C. thermocellum* in CBP require not only knowledge in molecular and genetic details of its cellulolytic system but also global understanding of its carbon metabolism. Detailing the carbon fluxes from metabolic pathways in *C. thermocellum* will guide design principles for efficient conversion of lignocellulosic sugars into dedicated products. To date, there have been a few published works describing the fermentation pathways in *C. thermocellum* (Zhang and Lynd, 2005; Lu et al., 2006; Nataf et al., 2010; Olson et al., 2010; Deng et al., 2013; Zhou et al., 2013), whereas metabolic characteristics of this species have not been comprehensively and quantitatively understood. Although the availability of genome sequence of *C. thermocellum* strains (DSM 1313 and ATCC 27405) has provided essential information and blue prints in terms of metabolic pathways, elaborate experimental validation is still required to fully realize its biotechnological applications.

Recently, ¹³C-tracer and metabolic flux analysis has emerged as a powerful tool for validating genome annotation and this technique has enabled us to identify a unique CO₂-fixing one-carbon metabolic pathway in *C. thermocellum* (Xiong et al., 2016). Complementing classic biochemical approaches in understanding metabolism such as enzyme activity assay, gene mutagenesis and the use of specific inhibitors to key metabolic steps, it provides a uniquely non-invasive approach to delineate the cellular metabolism *in vivo*. Here we use ¹³C-labeled substrates as isotopic tracers to follow the operation of *C. thermocellum*'s central carbon metabolism and other primary metabolic pathways directly in living cells. As a verification of genomic information, we find that the Embden–Meyerhof–Parnas (EMP) pathway is the predominant glycolytic route whereas the Entner–Doudoroff (ED) pathway and oxidative pentose phosphate pathway are inactive. We also observe that *C. thermocellum*'s TCA cycle is initiated with both *Si*- and *Re*-citrate synthase and it is broken between 2-oxoglutarate and oxaloacetate in the oxidative direction. *C. thermocellum* uses a citramalate shunt to synthesize isoleucine; and the one-carbon metabolism and the malate shunt are active in this bacterium.

To gain a quantitative insight into the newly identified metabolic activities, we constructed a fluxomic model that enables us to quantify the metabolic fluxes through the EMP pathway, the nonoxidative pentose phosphate pathway, the TCA cycle, the malate shunt, and amino acid biosynthesis pathways. This study represents the first global characterization of the central carbon metabolic pathways in a model cellulolytic species.

It also showcases the capability of isotope-assisted metabolite analysis in complementing genome annotation with a more in-depth understanding of microbial metabolism.

MATERIALS AND METHODS

Strains, Culture Conditions, and Medium

C. thermocellum DSM 1313 derived strain Δhpt (Xiong et al., 2016) was grown anaerobically at 55°C on 5 g/L glucose. We used the DSM122-defined medium referred to as CTFUD minimum medium (Olson and Lynd, 2012). The rich medium was added with 1 g/L yeast extract. The growth medium was deoxygenated by gassing with argon for 20 min and autoclaved before use.

Quantitative Analysis of Fermentation Products

Cell growth was measured as a function of optical density by spectrophotometry (DU800; Beckman Coulter) at OD_{600nm}. An OD₆₀₀ of 1 correlated to 1.04 g/L cell dry weight ($R^2 = 0.9918$). Hydrogen gas was measured using a gas chromatograph (7890A GC; Agilent Technologies) equipped with a thermal conductivity detector and a stainless-steel Supelco 60/80 Mol Sieve column (6 ft × 1/8 in) with argon as the carrier gas. Peak areas were compared with a standard curve. Glucose, lactate, formate, acetate, and ethanol were measured by HPLC (1200 series; Agilent Technologies) with a mobile phase of 4 mM H₂SO₄ at 0.6 mL/min flow rate using an Aminex HPX-87H column with a Micro Guard Cation H Cartridge. The column temperature was set to 55°C.

¹³C-Tracer Experiment

We adopted a steady-state labeling strategy. The ¹³C-labeling experiment was performed with different nutrient combinations: (1) 5.56 mM [U-¹³C₆] glucose (20%) + 22.22 mM unlabeled glucose (80%); (2) 20 mM ¹³C-bicarbonate + 27.78 mM (5 g/L) glucose; (3) 20 mM ¹³C-formate + 27.78 mM glucose. (4) 20 mM [U-¹³C₅] glutamate + 27.78 mM glucose. These nutrient combinations are supplemented into the CTFUD defined medium, respectively. For cell growth, *C. thermocellum* strains were inoculated into these media with a starting OD₆₀₀ of 0.05. When late log phase was reached (OD₆₀₀ above 0.6), 3 mL of cultures was sampled.

GC-MS and Isotopomer Analysis for Proteinogenic Amino Acids

The sample treatment and GC-MS analysis were done as previously reported (5), with a few modifications. Briefly, 3 mL of sampled cultures was centrifuged at 10,000 × g for 1 min, and the cell pellets were digested with 500 μL of 6 M HCl at 105°C for 12 h. The hydrolysate was dried under nitrogen gas flow at 65°C and dissolved in 50 μL of water-free dimethylformamide. For the GC/MS measurement, the proteinogenic amino acids were derivatized before analysis. The dried hydrolysate, dissolved in N, N-dimethylformamide, was derivatized by 1% tert-butyl-dimethylchlorosilane (TBDMS) at 85°C for 60 min. One microliter of the sample in the organic

phase was loaded onto the GC/MS instrument [Agilent GC-6890 gas chromatograph equipped with an Agilent 19091J-413 column (30 m × 0.32 mm × 0.25 μm) directly connected to a MS-5975C mass spectrometer]. Helium was the carrier gas. The oven temperature was initially held at 50°C for 2 min; it was then raised to 150°C at 5°C/min and held at that value for 2 min. Finally, it was raised to 320°C at 7°C/min and held at that final value for 2 min. Other settings included splitless and electron impact ionization at 70 eV. The amino acids, including alanine, aspartate, glutamate, glycine, isoleucine, leucine, methionine, phenylalanine, proline, serine, threonine, tyrosine, and valine, were separated and analyzed. Histidine was not detected due to its low abundance in cell biomass.

To analyze the isotope labeling pattern of amino acids, a mass isotopomer distribution vector, MDV_{α} , was assigned according to Nanchen et al. (2007).

$$MDV_{\alpha} = \begin{bmatrix} (m_0) \\ (m_1) \\ \vdots \\ (m_n) \end{bmatrix} \sum_{i=0}^n m_i = 1 \quad (1)$$

where m_0 is the fractional abundance of molecules with mono-isotopic mass and $m_i > 0$ is the abundance of fragments with heavier masses. The GC-MS data were corrected for the naturally occurring isotopes of oxygen (O), hydrogen (H), and carbon (C) atoms using a correction matrix (Equation 2) as described by Nanchen et al. (2007).

$$MDV_{\alpha}^* = C_{\text{corr,COH}}^{-1} \cdot MDV_{\alpha} \quad (2)$$

where MDV_{α}^* is the corrected mass isotopomer distribution vector and $C_{\text{corr,COH}}^{-1}$ is the correction matrix. According to Equation 3, the resulting MDV_{α}^* values were then used to assess the fractional labeling enrichment of amino acids, respectively.

$$FL = \frac{\sum_{i=0}^n i \cdot m_i}{n \cdot \sum_{i=0}^n m_i} \quad (3)$$

where n represents the number of carbon atoms in the amino acid and i is the mass isotopomer.

Quantification of Metabolic Fluxes

The central carbon network of *C. thermocellum* DSM 1313 was constructed based on genome knowledge which has been validated by isotope tracer experiments in this work. Specifically, the network includes the EMP pathway, the nonoxidative pentose phosphate pathway, the malate shunt, the incomplete TCA Cycle led by both *Si*- and *Re*-citrate synthase, and amino acids biosynthesis pathways (Complete reactions list please see **Supplementary File 1**). The biomass composition (**Supplementary File 1**, R53) was defined according to a previous report describing genome-scale reconstruction of *C. thermocellum* metabolic network (Roberts et al., 2010). Minor modifications include: (1) We formulate the biomass equation by using the metabolites appeared in the reactions list (**Supplementary File 1**). (2) Experimentally measured fatty

acids profiles are employed (see **Supplementary File 1**). The ^{13}C -metabolic fluxes are quantified by minimizing the sum-of-squared residuals (SSR) between computationally simulated and experimentally determined measurements. INCA, a ^{13}C -flux software based on Matlab platform (Young, 2014) was utilized for flux estimation.

RESULTS

Validation of the Experimental System for ^{13}C Tracer Analysis of *C. thermocellum*

The first task for ^{13}C -tracer analysis is to validate the experimental system by examining the effect of medium nutrients on ^{13}C -labeling. This is because complex carbon nutrients (i.e., amino acids in yeast extract) can be incorporated into cell biomass, and thus interfere with a quantitative flux analysis. To test medium effect, we used 20% fully ^{13}C -labeled and 80% fully unlabeled glucose as the carbon source and cultivated the cells in CTFUD rich medium containing yeast extract (1 g/L) and CTFUD defined medium without yeast extract, respectively. We used cultures when cells were at late exponential phase of growth (OD_{600} above 0.6). The labeling steady state is assumed to be reached at this growth stage, we then quantify the fractional labeling of each proteinogenic amino acid (AA) fragments by GC-MS. As shown in **Figure S1**, the labeling ratio of most proteinogenic AA fragments in rich medium is far less than the theoretical value (20%), indicating that unlabeled amino acids in yeast extract were assimilated by *C. thermocellum* directly and thus diluting the labeling ratio significantly. In contrast, the labeling ratio of AA fragments in CTFUD define medium is much close to theoretical value (20%), confirming that ^{13}C -labeling is not interfered by medium nutrients in the defined medium and the pseudo steady state for labeling has been reached. Overall, our results have shown that tracer experiment using minimal CTFUD medium is a valid system for ^{13}C metabolic flux analysis.

Key Amino Acids Pathways

To examine amino acids biosynthetic pathways in *C. thermocellum*, the labeling profiles of 13 proteinogenic amino acids were analyzed. We used ^{13}C -bicarbonate labeling data as the blueprint, since ^{13}C -bicarbonate labeling leads to the $[1-^{13}\text{C}]$ pyruvate through reversed pyruvate ferredoxin oxidoreductase (PFOR) (Xiong et al., 2016), and pyruvate may serve as the precursor of several key amino acids. For example, valine and leucine exhibited similar carbon molecule-labeling pattern of pyruvate (see **Figure S2**) and the labeled carbons were shown on the first position of these two amino acids (carboxylic group). The finding supports that pyruvate is the common precursor of both valine and leucine in *C. thermocellum* (see **Figure S2**).

With respect to aspartate biosynthesis, we detected that 15% of aspartate has two labeled carbons (**Figure S3**). It is consistent with a strong activity of anaplerotic reactions (the malate shunt), which enable the combination of $[1-^{13}\text{C}]$ pyruvate and a ^{13}C -bicarbonate to form a two-carbon labeled malate and then oxaloacetate, the precursor of aspartate. We further observed that threonine has identical labeling pattern as that of aspartate,

confirming that its carbon skeleton is derived from oxaloacetate directly (**Figure S3**).

Next, we examine the biosynthetic pathway for isoleucine, which is a typical branch chain amino acid. Genomic information (<https://biocyc.org>) suggests that two possible pathways may contribute to the biosynthesis of isoleucine: the threonine pathway and the citramalate pathway (Please see **Figure S4** for detailed pathways). These two routes may lead to differential features in isotopomer labeling, thus enabling us to evaluate their relative contribution accordingly. The pathway from threonine may generate two-carbon labeled isoleucine (see **Figure S4**), while the citramalate pathway leads to only one ^{13}C -carbon on the first position of isoleucine. The measured Mass Distribution Vector (MDV) of isoleucine showed that main labeling of isoleucine is exactly located at the carboxylic group of isoleucine (see **Figure S4** Ile M-15, and Ile M-85) and the labeling probability on another carbon atom is comparatively low (m1: 7% in Ile M-85, see **Figure S4**). Apparently, this data suggests that the citramalate pathway serve as the dominant pathway for isoleucine biosynthesis.

Glycolytic Fluxes Are Dominantly Contributed by the EMP Pathway

The pathway activities in *C. thermocellum* including the glycolytic pathway and the one-carbon pathway were then quantitatively analyzed by ^{13}C -tracer analysis. Specifically, we utilized $[1-^{13}\text{C}]$ glucose as the carbon tracer to distinguish the glycolytic flux from the EMP pathway, the oxidative pentose phosphate pathway, and the ED pathway. Proteinogenic serine was used as the readout as it can be diversely synthesized from multiple routes of the glycolytic pathways. The deduced serine labeling pattern from each pathway is shown in **Figure 1**. Since the EMP pathway will split the hexose in the middle and results in 50% of serine labeled with one ^{13}C atom, whereas other pathways only generate unlabeled serine (labeling ratio equals to 0%). The flux from the EMP pathway can be quantified accordingly and the results are shown in **Figure 2**. Our data show 62% of serine are derived from the EMP pathway, supporting that the EMP pathway is the dominant glycolytic route in *C. thermocellum*. However, origin of the rest 38% of serine remains as an intriguing question.

We further tracked which other pathway(s) may contribute to the biosynthesis of the remaining labeled serine. With respect to the ED pathway, genes for ED pathway are not annotated in *C. thermocellum* genome but are identified in other *Clostridia* species (Bender and Gottschalk, 1973). The potential contribution from ED pathway therefore should be taken into account. We analyzed the ED pathway activity by checking the labeling pattern of alanine. If fluxes go through the ED pathway, $[1-^{13}\text{C}]$ glucose tracer will lead to labeling at the carboxylic group (C1) of pyruvate which is the precursor of alanine (Carbon destination of $[1-^{13}\text{C}]$ glucose by the ED pathway is shown in **Figure 3**). However, main labeling of alanine is found on C3 (**Figure 3**) identical to the EMP pathway activity, and negligible labeling is at C1 position. It clearly indicates very low activity of the ED pathway, which is consistent with the fact that key genes

for ED pathway (i.e., 2-keto-3-deoxyphosphogluconate aldolase) is absent or not annotated based on the genomic information.

We next considered the activity of oxidative pentose phosphate pathway. Genes for the glucose 6-phosphate dehydrogenase and 6-phosphogluconate dehydrogenase are not found in *C. thermocellum* genome either. Additionally, their biochemical activity was absent from cell-free extracts (Patni and Alexander, 1971), making it unlikely that serine can be derived from this pathway. With respect to other pathways which may contribute to the biosynthesis of serine, we evaluated the importance of one-carbon metabolism. Serine can be reversibly synthesized from glycine by combining a methyl group which is derived from the formate via the one-carbon metabolism. To validate this pathway, we labeled the cells with ^{13}C -formate and detected ~25% of serine labeled at C3 position (**Figure 4**). This result suggests the flux through one-carbon metabolism is quite high and dramatically contributes to serine synthesis. Overall, this analysis quantitatively informs us of the relative activities of glycolytic pathways and one-carbon metabolism.

C. thermocellum Carries Both Re- and Si-Citrate Synthase

As a ubiquitous energetic and biosynthetic pathway, the tricarboxylic acid (TCA) cycle however remains elusive for anaerobic bacteria. As the first step of the TCA cycle, biosynthesis of citrate from acetyl-coenzyme A and oxaloacetate is catalyzed in most organisms by a Si-citrate synthase, which is Si-face stereospecific with respect to C-2 of oxaloacetate. Whereas in *C. kluyveri* and some other clostridia, the reaction can be catalyzed by a Re-citrate synthase whose homolog was also found in *C. thermocellum* (Li et al., 2007). ^{13}C tracer and isotopomer analysis offers an opportunity to visualize the stereospecificity of citrate synthase, *in vivo* (Wu et al., 2010). With respect to *C. thermocellum*, the stereospecificity of citrate synthase can be deduced by a ^{13}C -bicarbonate labeling experiment (**Figure 5**) in which ^{13}C -carbons can be propagated to both carboxylic group of oxaloacetate via the reactions of reversed PFOR (rPFOR) and malate shunt, and then to citrate by the stereospecific citrate synthases. The stereotyped citrate synthase can be mirrored by the labeling patterns of downstream glutamate (GC-MS fragments C2-5 and C1-2, respectively). As shown in **Figure 5B**, over 25% C2-5 fragment has one ^{13}C -carbon, presumably on the δ -carboxylic group of glutamate. This is consistent with Re-type citrate synthase. We then checked the labeling pattern of glutamate C1-2. Interestingly 28% glutamate C1-2 fragment also has one ^{13}C -carbon, suggesting the first carboxyl group of glutamate was labeled as well. This labeling pattern is consistent with the notion that *C. thermocellum* also carries an active Si-citrate synthase. To confirm this observation, we investigated the genome of *C. thermocellum* DSM 1313 and indeed the gene for Si-citrate synthase (CLO1313_RS02945) is present. It has 67% sequence identity with a known Clostridial Si-citrate synthase (Li et al., 2007). For Re-citrate synthase, we did the BLAST analysis using the sequence of a biochemically verified gene from *C. kluyveri* DSM 555 (Li et al., 2007). A candidate gene (CLO1313_RS03665)

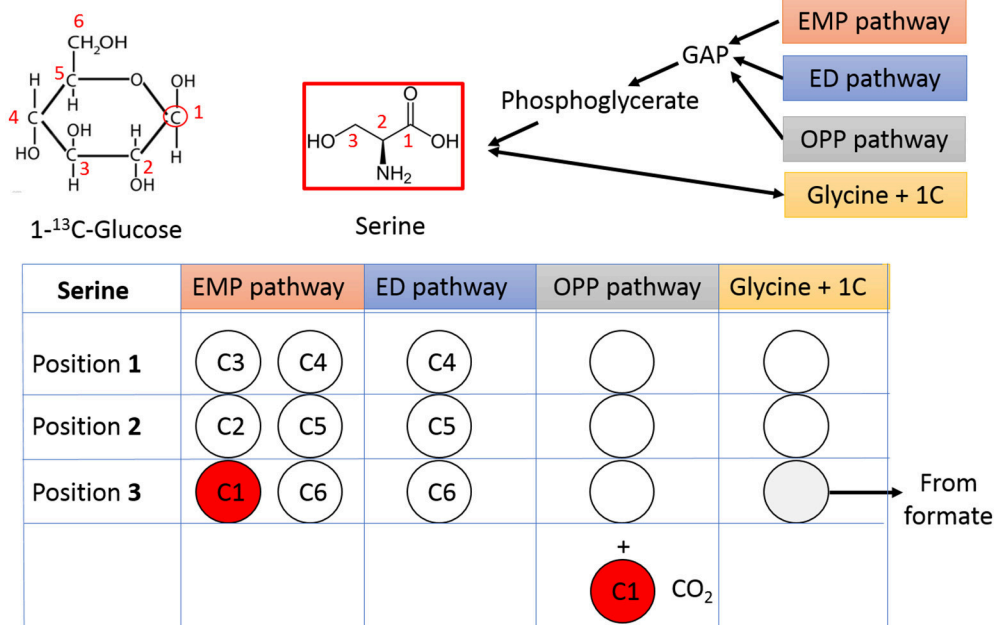


FIGURE 1 | Tracing the glycolytic flux using $1\text{-}^{13}\text{C}$ -glucose. Relative flux from different glycolytic pathways can be mirrored by the labeling pattern of serine. The unlabeled carbons are represented by open circles and labeled carbons are shown in red. The carbon number shown in each circle is from corresponding carbon in $1\text{-}^{13}\text{C}$ -glucose. GAP, Glyceraldehyde 3-phosphate; OPP pathway, Oxidative Pentose Phosphate pathway.

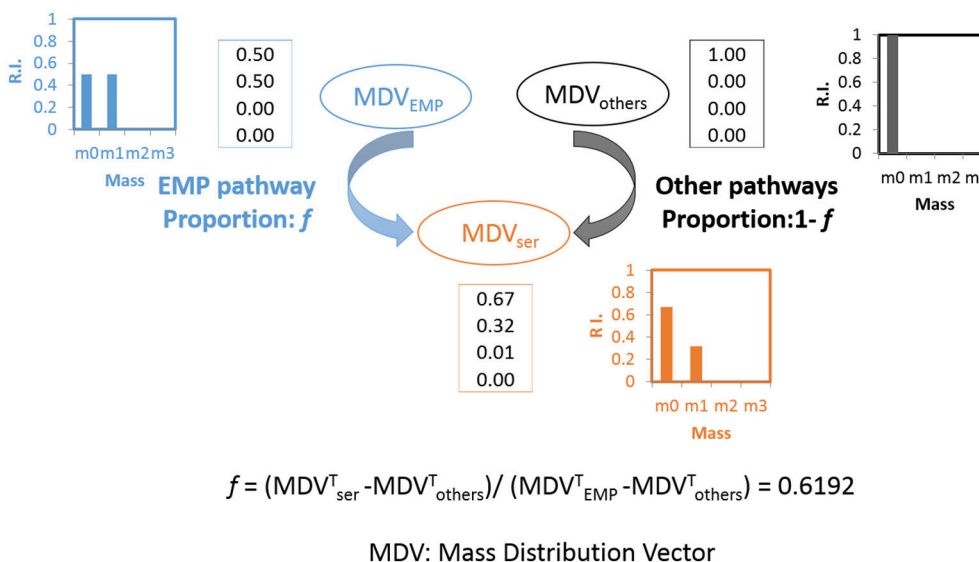


FIGURE 2 | Calculation of flux ratio: Serine from the EMP pathway (f). Upon $[1-^{13}\text{C}]$ glucose labeling, the serine molecules originate through the EMP pathway will lead to half of the serine molecules labeled at position 3 of serine ($m1: 0.5$) while the other half will be unlabeled ($m0: 0.5$) (shown in blue). If the serine molecules originate from other pathways, none of the molecules will be labeled ($m0: 1$, shown in black). The final MDV of serine (shown in orange) is the cooperative result of these two possibilities and the splitting ratio (f) can be analyzed quantitatively. The algorithm is based on (Nanchen et al., 2007).

shares 61% identity with *C. kluyveri* Re- citrate synthase and a neighbor gene (CLO1313_RS03670) shares 73% homology with aconitase, strongly indicating their joint functionality in the TCA cycle by converting acetyl CoA and oxaloacetate to isocitrate.

Integrating all the acquired knowledge, our data provides labeling evidence for the first time to support that *C. thermocellum* carries both *Re*- and *Si*-Citrate synthases to initiate the TCA cycle.

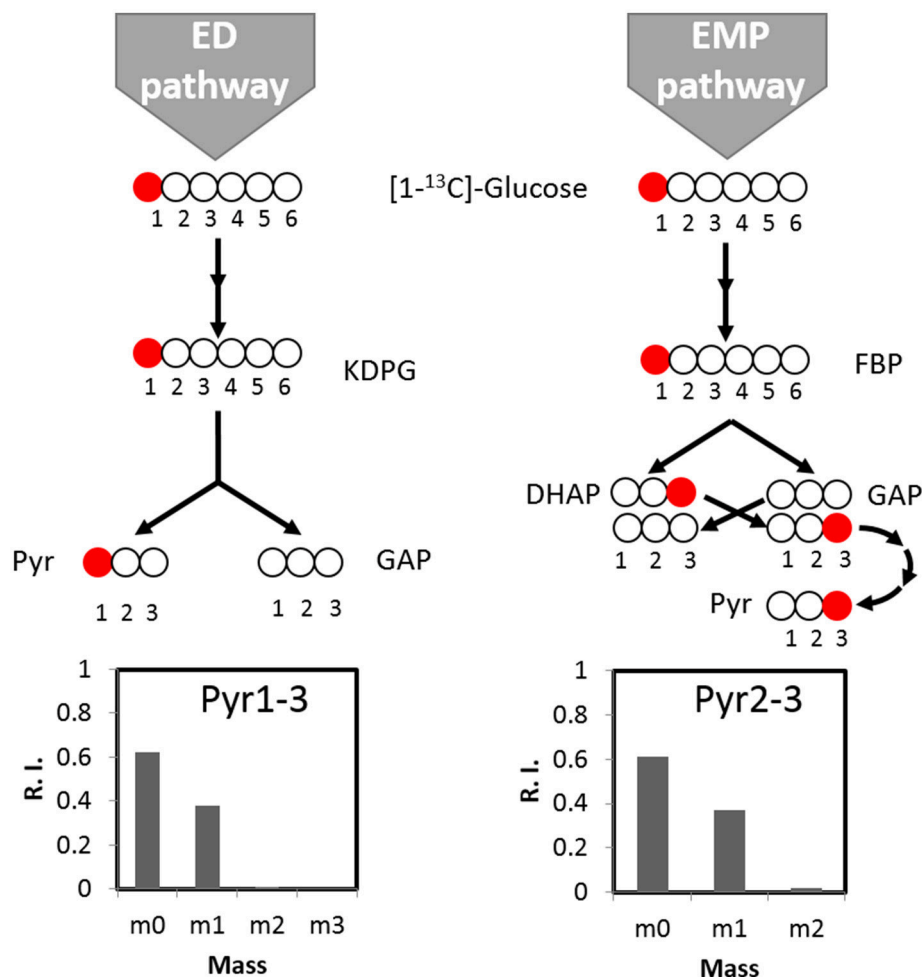


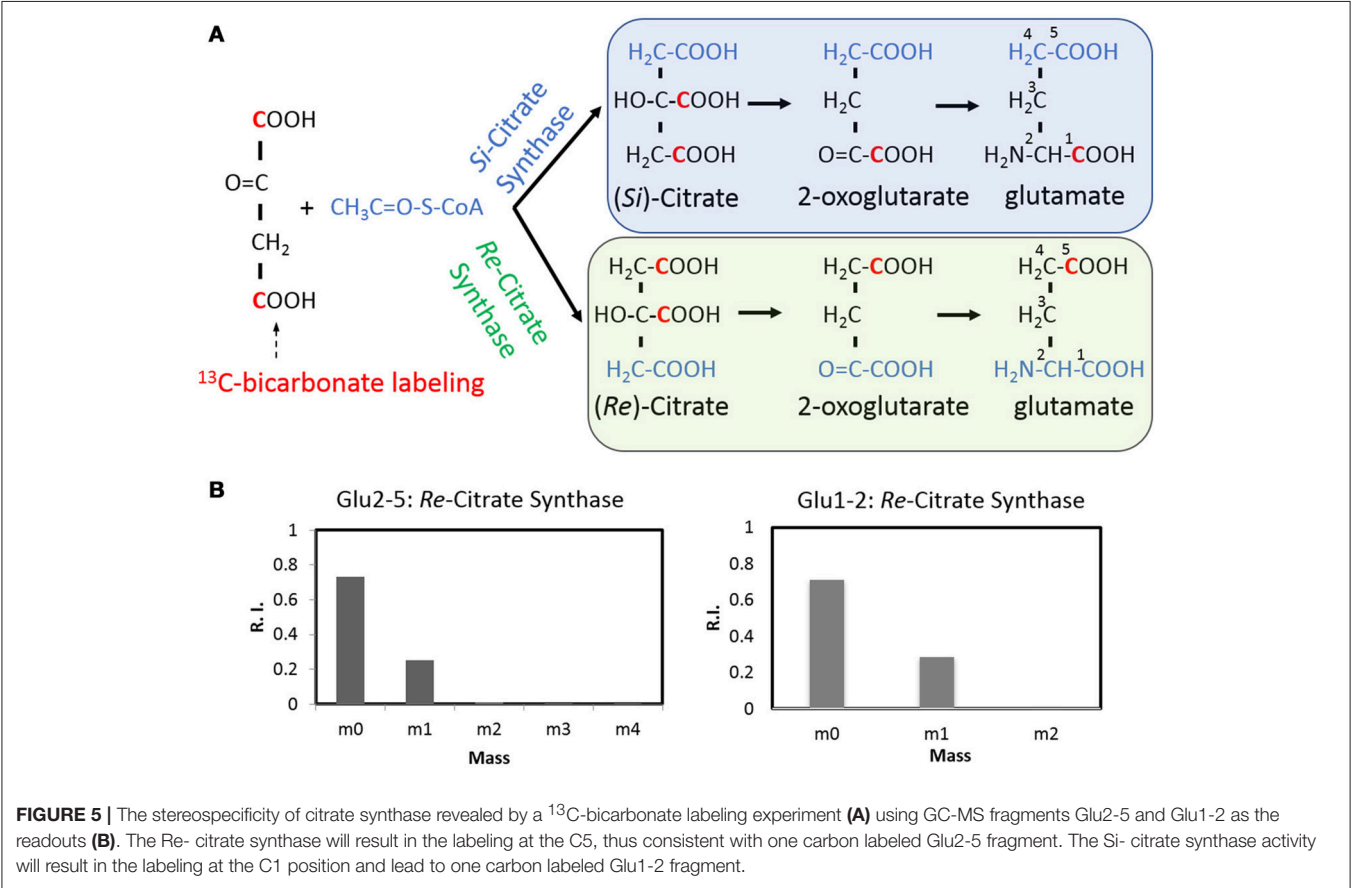
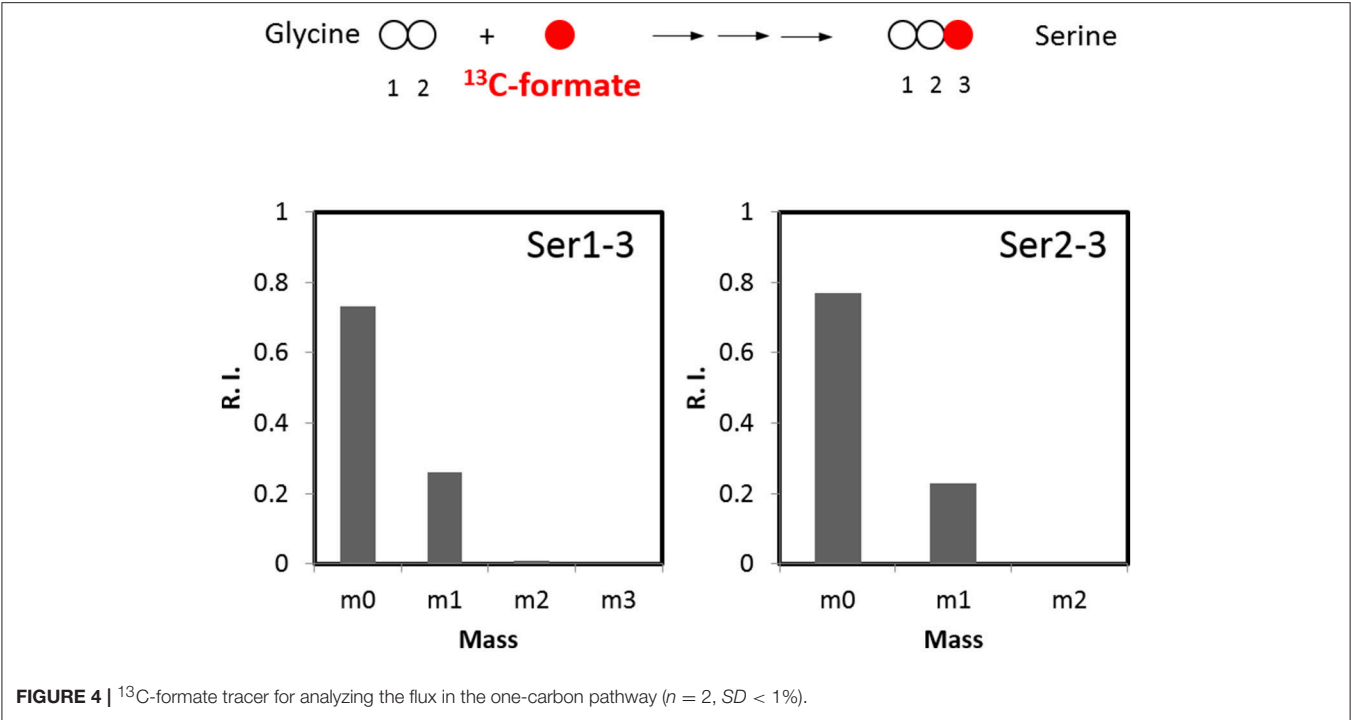
FIGURE 3 | The activity of the ED pathway can be mirrored by pyruvate labeling pattern. If fluxes go through the ED pathway, $[1-^{13}\text{C}]$ glucose tracer will lead to pyruvate labeling at carboxylic group (C1). If fluxes go through the EMP pathway, pyruvate can be labeled at the C3 position. According to labeling pattern of alanine, the direct product of pyruvate, there is no significant labeling at C1 of pyruvate, indicating negligible activity of the ED pathway. KDPG, 2-keto-3-deoxygluconate-6-phosphate; FBP, Fructose 1,6-bisphosphate; DHAP, Dihydroxyacetone phosphate; GAP, glyceraldehyde 3-phosphate; Pyr, pyruvate.

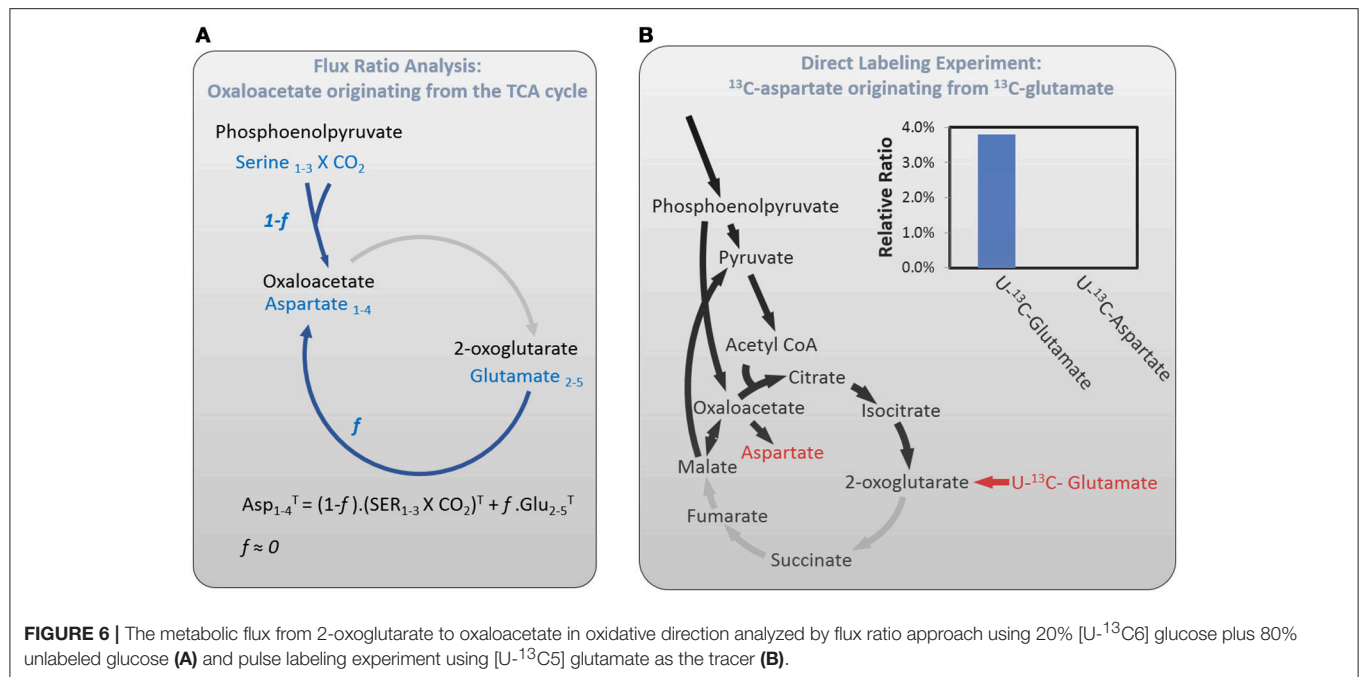
The TCA Cycle of *C. thermocellum* Is Incomplete

It is generally accepted that most obligatory anaerobes do not harbor a complete oxidative TCA cycle, as energy can be acquired mostly from anaerobic fermentation. Recent studies however showed that a few model anaerobic bacteria such as *Proteus mirabilis* (Alteri et al., 2012) and *C. acetobutylicum* (Amador-Noguez et al., 2010) harbor a complete TCA cycle, even if genome annotation suggests the absence of certain TCA cycle enzymes. It is therefore worthwhile to decipher the structure of the TCA cycle in *C. thermocellum* specifically. First, we adopted the flux ratio analysis (Nanchen et al., 2007) to probe the completeness of the TCA cycle. Using $[U-^{13}\text{C}_6]$ glucose experiments, we quantified the fraction of oxaloacetate originated from 2-oxoglutarate via the TCA cycle (f) vs. the ratio of oxaloacetate derived through anaplerotic reaction by e.g., phosphoenolpyruvate (PEP) carboxylase ($1-f$) (The algorithm

please, see Figure 6A and Nanchen et al., 2007). Our calculation returned a $f \approx 0$, representing the least squares solution of the relative flux from oxidative TCA cycle. This result indicates that the oxaloacetate is mainly synthesized by the anaplerotic reactions, whereas the flux barely reaches the oxaloacetate from the TCA cycle in the oxidative direction.

To further validate this calculation, we performed an additional pulse labeling experiment using $[U-^{13}\text{C}_5]$ glutamate as the isotope tracer. After 12h of labeling, we analyzed the proteinogenic glutamate and aspartate, respectively, which are produced from the TCA cycle (Figure 6B). Our result shows that nearly 4% fully labeled glutamate can be detected from the final glutamate pool, suggesting that the ^{13}C -tracer was incorporated into the metabolic pathway. In comparison, no fully labeled aspartate was detectable during the labeling period. It indicates that the ^{13}C carbons entering the 2-oxoglutarate branch cannot be propagated to oxaloacetate, the precursor of aspartate. Indeed,





these fluxomic evidences are consistent with the genomic data that the absence of a few key enzymes including 2-oxoglutarate decarboxylase, fumarate reductase, and fumarase prevents *C. thermocellum* from forming a complete TCA cycle.

Quantitative Flux Mapping

In vivo biochemical reaction rates are among the most important metabolic phenotypes. To obtain a quantitative insight into the fluxes in *C. thermocellum*, we developed a metabolic model that describes the isotope labeling of metabolites upon the addition of 20% universally labeled ^{13}C -glucose and 80% unlabeled glucose (see section Materials and Methods, **Figure S5** and **Supplementary File 1** for modeling details). The biochemical reactions adopted in the model are based on genomic information and involve newly validated pathways that have been identified by isotope tracer experiments as described above. Specifically, the principal metabolic network includes the EMP pathway, anaplerotic reactions (malate shunt), citramalate pathway for isoleucine biosynthesis, *Re*- and *Si*-citrate synthases, incomplete TCA cycle, all of which have been verified by isotope tracer experiments in this work. In addition, we excluded the oxidative pentose phosphate pathway and the ED pathway from the central carbon metabolism which have been shown absent in *C. thermocellum* as described above. The balancing of energetic co-factors such as NAD(P)H is not input into the model either, because whether a particular redox reaction involves NADH vs. NADPH needs to be validated experimentally, and any assumptions for NAD(P)H-specific reactions will bring uncertainty to the modeling. Experimental inputs to the model include steady-state labeling data from proteinogenic amino acids, specific growth rates, sugar uptake rates, excretion rates for fermentation products (see **Table S1**),

and flux ratio data at certain branch points: e.g., the relative contribution of the EMP pathway vs. one-carbon metabolism to serine biosynthesis as described above and the relative malate shunt and pyruvate phosphate dikinase fluxes to pyruvate, which has been quantitatively addressed recently (Olson et al., 2017).

Figure 7 shows representative results for the fitting data and a map of the identified flux values in central metabolism. The complete results are presented in **Supplementary File 1** and the goodness for fitting in the **Figure S5**. The metabolic model fits the observed data. Most of the identified fluxes were tightly constrained, indicating that they reliably mirror the available experimental data. The results show that glycolytic flux predominates and is directed primarily toward the fermentation of lactate, formate, acetate and ethanol. Other apparent fluxes include aspartate production via malate shunt, and pentose-phosphate production from glycolytic intermediates. Within the incomplete TCA cycle, the flux through the oxidative branch is limited. The production of glutamate from citrate can occur via *Re*-citrate synthase but is also expected to occur via the *Si*-citrate synthase. Compared with the flux map of other *Clostridia* e.g., *C. acetobutylicum* (Amador-Noguez et al., 2010), the computational results presented herein demonstrate a featured fluxomic structure unique in *C. thermocellum*.

DISCUSSION

Using ^{13}C -tracer studies, we have developed a quantitative flux model outlining the central carbon metabolism in *C. thermocellum* and elucidated the key pathways responsible for carbon flux distribution. This model could serve as a platform to probe cellular phenotypes upon subjecting to different growth

Our isotope experiment also demonstrates an incomplete TCA cycle in the oxidative direction from 2-oxoglutarate to oxaloacetate. This observation further prompts the question whether the TCA cycle in *C. thermocellum* can be operated reversibly in the reductive direction. In fact, the isotope data has excluded this possibility. In ^{13}C -bicarbonate labeling experiment, we detected two-carbon labeled oxaloacetate. If it follows reductive operation of the TCA cycle, then 3-carbon labeled glutamate should be visualized due to one more ^{13}C incorporation into the 2-oxoglutarate from the 2-oxoglutarate synthase. However, this is not observed in the glutamate labeling pattern (Figure 5), indicating no activity for the rTCA cycle. *C. thermocellum*'s TCA cycle is different from that of *C. acetobutylicum*, the latter showed a complete, albeit bifurcated structure (Amador-Noguez et al., 2010). Our findings strengthen the notion that TCA cycle in anaerobes plays a major role in biosynthesis rather than bioenergetics, although the architecture of the cycle is species-specific.

The approach used in this study represents a standard and applicable methodology for metabolic flux analysis using steady-state isotope tracer experiments. The isotope labeling approach (steady-state isotopic approach) we used is different from the kinetic flux profiling which is another promising fluxomics method (Yuan et al., 2008). One major merit of steady-state isotopic approach is its capability in providing accurate ratios of fluxes at branch points. Additionally, only a few data points are needed for the analysis, satisfying the requirements for high-throughput and large-scale fluxomic analysis (Fischer et al., 2004; Fischer and Sauer, 2005). Other advantages include easy experimental procedures, simple modeling algorithm. Nevertheless, steady-state isotopic approach still has its limitation: e.g., it cannot discriminate the fluxes from multiple parallel pathways if no distinguishable labeling patterns can be generated. In the case of *C. thermocellum*, for example, such a scenario occurs in deciphering fluxes from pyruvate phosphate dikinase (*ppdk*) vs. malate shunt, both

of which synthesize pyruvate from PEP but without forming distinguishable carbon-carbon bond. In this case, kinetic flux profiling which records labeling trajectories of intermediate as a function of time can provide a complementary solution (Olson et al., 2017). Joint and appropriate utilization of steady-state and kinetic fluxomics technique will enable a comprehensive understanding of the metabolic network by rationally designed isotope tracer experiments.

In conclusion, this work represents the first global investigation of the central carbon metabolism in *C. thermocellum* and exemplifies the ability of isotope tracer experiments and metabolic modeling in understanding microbial metabolism of industrial interests.

AUTHOR CONTRIBUTIONS

WX and PM led the research. WX designed the experiments. WX, KC, CW, LM, and TD performed the experiments. WX, JL, KC, and PM wrote the article.

ACKNOWLEDGMENTS

This work was supported by the US Department of Energy (DOE) Energy Efficiency and Renewable Energy (EERE) Fuel Cell Technologies Office under Contract DE-AC36-08-GO28308, and the NREL Laboratory Directed Research and Development (LDRD) project. We thank Chris Urban for his technical assistance. The authors declare that they have no conflicts of interest with the contents of this article.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2018.01947/full#supplementary-material>

REFERENCES

- Alteri, C. J., Himpf, S. D., Engstrom, M. D., and Mobley, H. L. (2012). Anaerobic respiration using a complete oxidative TCA cycle drives multicellular swarming in proteus mirabilis. *MBio* 3:e00365-12. doi: 10.1128/mBio.00365-12
- Amador-Noguez, D., Feng, X. J., Fan, J., Roquet, N., Rabitz, H., and Rabinowitz, J. D. (2010). Systems-level metabolic flux profiling elucidates a complete, bifurcated tricarboxylic acid cycle in clostridium acetobutylicum. *J. Bacteriol.* 192, 4452–4461. doi: 10.1128/JB.00490-10
- Argyros, D. A., Tripathi, S. A., Barrett, T. F., Rogers, S. R., Feinberg, L. F., Olson, D. G., et al. (2011). High ethanol titers from cellulose by using metabolically engineered thermophilic, anaerobic microbes. *Appl. Environ. Microbiol.* 77, 8288–8294. doi: 10.1128/AEM.00646-11
- Bender, R., and Gottschalk, G. (1973). Purification and properties of d-gluconate dehydratase from clostridium pasteurianum. *Eur. J. Biochem.* 40, 309–321.
- Deng, Y., Olson, D. G., Zhou, J., Herring, C. D., Joe Shaw, A., and Lynd, L. R. (2013). Redirecting carbon flux through exogenous pyruvate kinase to achieve high ethanol yields in *Clostridium thermocellum*. *Metab. Eng.* 15, 151–158. doi: 10.1016/j.ymben.2012.11.006
- Ellis, L. D., Holwerda, E. K., Hogsett, D., Rogers, S., Shao, X., Tschaplinski, T., et al. (2012). Closing the carbon balance for fermentation by *Clostridium thermocellum* (ATCC 27405). *Bioresour. Technol.* 103, 293–299. doi: 10.1016/j.biortech.2011.09.128
- Fischer, E., and Sauer, U. (2005). Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism. *Nat. Genet.* 37, 636–640. doi: 10.1038/ng1555
- Fischer, E., Zamboni, N., and Sauer, U. (2004). High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived ^{13}C constraints. *Anal. Biochem.* 325, 308–316. doi: 10.1016/j.ab.2003.10.036
- Gold, N. D., and Martin, V. J. (2007). Global view of the *Clostridium thermocellum* cellulosome revealed by quantitative proteomic analysis. *J. Bacteriol.* 189, 6787–6795. doi: 10.1128/JB.00882-07
- Holwerda, E. K., Thorne, P. G., Olson, D. G., Amador-Noguez, D., Engle, N. L., Tschaplinski, T. J., et al. (2014). The exometabolome of *Clostridium thermocellum* reveals overflow metabolism at high cellulose loading. *Biotechnol. Biofuels* 7:155. doi: 10.1186/s13068-014-0155-1
- Li, F., Hagemeyer, C. H., Seedorf, H., Gottschalk, G., and Thauer, R. K. (2007). Re-citrate synthase from clostridium kluyveri is phylogenetically related to homocitrate synthase and isopropylmalate synthase rather than to si-citrate synthase. *J. Bacteriol.* 189, 4299–4304. doi: 10.1128/JB.00198-07

- Lin, P. P., Mi, L., Morioka, A. H., Yoshino, K. M., Konishi, S., Xu, S. C., et al. (2015). Consolidated bioprocessing of cellulose to isobutanol using *Clostridium thermocellum*. *Metab. Eng.* 31, 44–52. doi: 10.1016/j.ymben.2015.07.001
- Liu, L., Zhang, L., Tang, W., Gu, Y., Hua, Q., Yang, S., et al. (2012). Phosphoketolase pathway for Xylose catabolism in *clostridium acetobutylicum* revealed by 13C metabolic flux analysis. *J. Bacteriol.* 194, 5413–5422. doi: 10.1128/JB.00713-12
- Lu, Y., Zhang, Y. H., and Lynd, L. R. (2006). Enzyme-microbe synergy during cellulose hydrolysis by *Clostridium thermocellum*. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16165–16169. doi: 10.1073/pnas.0605381103
- Lynd, L. R., Laser, M. S., Bransby, D., Dale, B. E., Davison, B., Hamilton, R., et al. (2008). How biotech can transform biofuels. *Nat. Biotechnol.* 26:169. doi: 10.1038/nbt0208-169
- Lynd, L. R., Van Zyl, W. H., McBride, J. E., and Laser, M. (2005). Consolidated bioprocessing of cellulosic biomass: an update. *Curr. Opin. Biotechnol.* 16, 577–583. doi: 10.1016/j.copbio.2005.08.009
- Müller, V., Imkamp, F., Biegel, E., Schmidt, S., Dilling, S. (2008). Discovery of a ferredoxin:NAD⁺-Oxidoreductase (Rnf) in *Acetobacterium woodii*. *Ann. N.Y.Acad. Sci.* 1125, 137–146. doi: 10.1196/annals.1419.011
- Nanchen, A., Fuhrer, T., and Sauer, U. (2007). Determination of metabolic flux ratios from 13C-experiments and gas chromatography-mass spectrometry data: protocol and principles. *Methods Mol. Biol.* 358, 177–197. doi: 10.1007/978-1-59745-244-1_11
- Nataf, Y., Bahari, L., Kahel-Raifer, H., Borovok, I., Lamed, R., Bayer, E. A., et al. (2010). *Clostridium thermocellum* cellulosomal genes are regulated by extracytoplasmic polysaccharides via alternative sigma factors. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18646–18651. doi: 10.1073/pnas.1012175107
- Olson, D. G., Hörl, M., Fuhrer, T., Cui, J., Zhou, J., Maloney, M. I., et al. (2017). Glycolysis without pyruvate kinase in *Clostridium thermocellum*. *Metab. Eng.* 39, 169–180. doi: 10.1016/j.ymben.2016.11.011
- Olson, D. G., and Lynd, L. R. (2012). Transformation of *Clostridium thermocellum* by electroporation. *Meth. Enzymol.* 510, 317–330. doi: 10.1016/B978-0-12-415931-0.00017-3
- Olson, D. G., Tripathi, S. A., Giannone, R. J., Lo, J., Caiazza, N. C., Hogsett, D. A., et al. (2010). Deletion of the Cel48S cellulase from *Clostridium thermocellum*. *Proc. Natl. Acad. Sci. U.S.A.* 107, 17727–17732. doi: 10.1073/pnas.1003584107
- Patni, N. J., and Alexander, J. K. (1971). Utilization of glucose by *Clostridium thermocellum*: presence of glucokinase and other glycolytic enzymes in cell extracts. *J. Bacteriol.* 105, 220–225.
- Roberts, S. B., Gowen, C. M., Brooks, J. P., and Fong, S. S. (2010). Genome-scale metabolic analysis of *Clostridium thermocellum* for bioethanol production. *BMC Syst. Biol.* 4:31. doi: 10.1186/1752-0509-4-31
- Taillefer, M., Rydzak, T., Levin, D. B., Oresnik, I. J., and Sparling, R. (2015). Reassessment of the transhydrogenase/malate shunt pathway in *Clostridium thermocellum* ATCC 27405 through Kinetic characterization of Malic enzyme and Malate dehydrogenase. *Appl. Environ. Microbiol.* 81, 2423–2432. doi: 10.1128/AEM.03360-14
- Tripathi, S. A., Olson, D. G., Argyros, D. A., Miller, B. B., Barrett, T. F., Murphy, D. M., et al. (2010). Development of pyrF-Based genetic system for targeted gene deletion in *Clostridium thermocellum* and creation of a pta mutant. *Appl. Environ. Microbiol.* 76, 6591–6599. doi: 10.1128/AEM.01484-10
- Wang, S., Huang, H., Moll, J., and Thauer, R. K. (2010). NADP⁺ reduction with reduced ferredoxin and NADP⁺ Reduction with NADH are coupled via an electron-bifurcating enzyme complex in *Clostridium kluyveri*. *J. Bacteriol.* 192, 5115–5123. doi: 10.1128/JB.00612-10
- Wu, B., Zhang, B., Feng, X., Rubens, J. R., Huang, R., Hicks, L. M., et al. (2010). Alternative isoleucine synthesis pathway in cyanobacterial species. *Microbiology* 156, 596–602. doi: 10.1099/mic.0.031799-0
- Xiong, W., Lin, P. P., Magnusson, L., Warner, L., Liao, J. C., Maness, P. C., et al. (2016). CO₂-fixing one-carbon metabolism in a cellulose-degrading bacterium *Clostridium thermocellum*. *Proc. Natl. Acad. Sci. U.S.A.* 113, 13180–13185. doi: 10.1073/pnas.1605482113
- Xiong, W., Reyes, L. H., Michener, W. E., Maness, P. C., and Chou, K. J. (2018). Engineering cellulolytic bacterium *Clostridium thermocellum* to co-ferment cellulose- and hemicellulose-derived sugars simultaneously. *Biotech. Bioeng.* 115, 1755–1763. doi:10.1002/bit.26590
- Young, J. D. (2014). INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* 30, 1333–1335. doi: 10.1093/bioinformatics/btu015
- Yuan, J., Bennett, B. D., and Rabinowitz, J. D. (2008). Kinetic flux profiling for quantitation of cellular metabolic fluxes. *Nat. Protoc.* 3, 1328–1340. doi: 10.1038/nprot.2008.131
- Zhang, Y. H., and Lynd, L. R. (2005). Cellulose utilization by *Clostridium thermocellum*: bioenergetics and hydrolysis product assimilation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 7321–7325. doi: 10.1073/pnas.0408734102
- Zhou, J., Olson, D. G., Argyros, D. A., Deng, Y., Van Gulik, W. M., Van Dijken, J. P., et al. (2013). Atypical glycolysis in *Clostridium thermocellum*. *Appl. Environ. Microbiol.* 79, 3000–3008. doi: 10.1128/AEM.04037-12

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

Copyright © 2018 Xiong, Lo, Chou, Wu, Magnusson, Dong and Maness. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read
for greatest visibility
and readership



FAST PUBLICATION

Around 90 days
from submission
to decision



HIGH QUALITY PEER-REVIEW

Rigorous, collaborative,
and constructive
peer-review



TRANSPARENT PEER-REVIEW

Editors and reviewers
acknowledged by name
on published articles

Frontiers

Avenue du Tribunal-Fédéral 34
1005 Lausanne | Switzerland

Visit us: www.frontiersin.org

Contact us: info@frontiersin.org | +41 21 510 17 00



REPRODUCIBILITY OF RESEARCH

Support open data
and methods to enhance
research reproducibility



DIGITAL PUBLISHING

Articles designed
for optimal readership
across devices



FOLLOW US

@frontiersin



IMPACT METRICS

Advanced article metrics
track visibility across
digital media



EXTENSIVE PROMOTION

Marketing
and promotion
of impactful research



LOOP RESEARCH NETWORK

Our network
increases your
article's readership