

DISS. ETH NO. 23931

**Generalized and High-throughput
¹³C Metabolic Flux Ratio Analysis
by Machine Learning**

A thesis submitted to attain the degree of

DOCTOR OF SCIENCES OF ETH ZURICH

(Dr. sc. ETH Zurich)

presented by

MARIA ZIMMERMANN-KOGADEEVA

Specialist in Mathematics and System Programming,

Lomonosov Moscow State University

born on 03.05.1989

citizen of Russian Federation

accepted on the recommendation of

Prof. Dr. Uwe Sauer

Dr. Nicola Zamboni

Prof. Dr. Manfred Claassen

Prof. Dr. Vassily Hatzimanikatis

2016

„Psychical continuity is nothing but the preservation of form in the flux of metabolic changes taking place in a sentient organism“

Edward Douglas Fawcett (1866-1960)

Contents

Abstract	7
Zusammenfassung	11
Chapter 1 General introduction	15
Chapter 2 SUMOFLUX: A generalized method for targeted ^{13}C metabolic flux ratio analysis	49
Chapter 3 Targeted ^{13}C flux ratio analysis reveals different patterns of amino acid co-utilization in mycobacteria	75
Chapter 4 High-throughput targeted ^{13}C flux ratio analysis enables large-scale functional profiling of enzyme phosphorylation in <i>Escherichia coli</i>	107
Chapter 5 Concluding Remarks	139
Appendix I Supplementary information Chapter 2	147
Appendix II Supplementary information Chapter 3	159
Appendix III Supplementary information Chapter 4	183
Abbreviations	195
Acknowledgments	199
Curriculum vitae	201

Abstract

Metabolism is an essential process for all living creatures. It describes the chemical conversion of consumed nutrients into biomass precursors, redox equivalents and energy, and the release of waste products. The metabolic potential of each cell is represented by a network of metabolites connected via reactions catalyzed by enzymes encoded in its genome. Living cells adjust their metabolic reaction rates, known as fluxes, in response to the external conditions, internal status and cellular requirements. Hence, metabolic fluxes reflect actual cellular behavior, and their assessment is crucial for understanding and controlling metabolic processes of biotechnological and biomedical relevance.

Metabolic fluxes cannot be measured directly, and have to be indirectly inferred from measurable quantities, e.g. gene expression, protein abundance, or temporal profiles of metabolite concentration changes. The most explicit experimental technique for flux elucidation are stable isotope tracing experiments, highly informative when alternative reactions lead to distinct labeling patterns in metabolic intermediates. These labeling patterns are detectable by nuclear magnetic resonance spectroscopy or mass spectrometry, and have to be interpreted either manually, with analytic equations, or incorporated into mathematical models in order to retrieve information on the underlying metabolic fluxes.

Existing flux analysis methods are divided into global ^{13}C metabolic flux analysis, which is based on iterative fitting of flux distributions to the labeling data; and local analysis based on calculating flux ratios from labeling data with ad hoc analytic equations. The former is generally applicable, but it requires comprehensive measurements, provides multiple equally plausible flux solutions and is computationally costly. The latter, on the contrary, is rapid and easy to apply, provides unique relative flux estimates, but is limited to few nodes and experimental conditions. Therefore, there is a demand for a fast, robust and generally applicable method that is scalable to large datasets, conceivably obtained from parallel experiments.

In this work, we present SUMOFLUX, a conceptually novel generalized method for targeted ^{13}C metabolic flux ratio analysis. This method exploits machine learning to predict flux ratios of interest from measurable data, using an *in silico*

training dataset generated with **surrogate modelling**. SUMOFLUX is applicable to virtually any type of network, substrate and measured data that can be simulated; and the actual estimation is very rapid once the flux ratio predictor is built. In **Chapter 2**, we developed the SUMOFLUX workflow, performed a proof-of-principle experiment to resolve key flux ratios in central carbon metabolism of *Escherichia coli*, and demonstrated that SUMOFLUX estimates were in good agreement with results obtained with both local and global ^{13}C flux analysis methods. Additionally, we illustrated the scalability and ease of experimental design with SUMOFLUX on a cohort of 121 *Bacillus subtilis* transcription factor mutants.

A remarkable advantage of the targeted approach is its applicability in complex systems even in case of poorly determined networks and little amount of data. This benefit became especially apparent in **Chapter 3**, where we investigated amino acid metabolism in mycobacteria in defined media and in the infection setup with macrophage-like THP-1 cells. By formulating specific flux ratios characterizing amino acid utilization, we classified amino acids by their role for central metabolism in *Mycobacterium smegmatis* and *Mycobacterium tuberculosis* growing in composite media. Investigation of *M. tuberculosis* behavior in the complex infection setup, where inter-species and media exchange fluxes are unknown, was possible through extensive simulations of feasible flux distributions in the phagosome-bacterial network. It revealed that during infection, biosynthesis of several amino acids decreased compared to bacterial growth in rich media. This implies that in the scarce nutritional conditions inside the phagosome, the pathogen is forced to utilize any nutrient it encounters. These results underline the adaptability of mycobacterial metabolism and partially explain recurrent failures of multiple drug treatments.

We further exploited the speed and flexibility of SUMOFLUX in **Chapter 4**, coupling it with rapid labeling data acquisition by untargeted high-throughput metabolomics platform FIA-TOF (flow injection – time of flight), that enabled to perform several hundred flux analyses per day. Substantial gain in speed came with challenges of missing and overlapping isotopologue data, which we partially solved by adopting rigorous filtering and quality check procedures. We validated the high-throughput flux analysis protocol with a set of *E. coli* knockout mutants with known flux phenotypes, which SUMOFLUX succeeded to predict. The developed protocol allowed us to perform a fluxomics screen of 60 *E. coli* strains with mutations in enzyme phosphorylation sites to generate hypotheses on the functional role of this

post-translational modification, which yet has not been extensively studied in prokaryotes. Our flux screening revealed the deactivating function of isocitrate dehydrogenase phosphorylation reported earlier, and proposed several novel functionally relevant phosphorylation events. The developed high-throughput flux profiling protocol brings ^{13}C fluxomics to a new level comparable with the scale of other omics techniques.

In summary, the developed targeted ^{13}C flux ratio approach offers an unprecedented medley of advantages. First, due to the generalization power of machine learning, its application is not limited to a specific organism, experimental conditions, type of input data or flux ratios. Second, the embedded surrogate modeling allows to reduce assumptions on the metabolic network structure, and to perform analysis of poorly determined systems. Third, estimating local ratios and *in silico* testing ensures extraction of most flux information even from small amount of measurements. Finally, once the flux ratio predictors are built, their application is rapid and scalable for high-throughput analysis. Taken together, this approach is able to address local questions in complex setups, such as bacterial co-cultures, higher cells in complex media or host-microbe systems, and is best suited for targeted hypotheses validation and high-throughput flux screening. We believe that our approach, alone or in combination with global flux analysis methods, will open up new horizons in ^{13}C fluxomics, and advance future biological discoveries in cellular metabolism and its regulation.

Zusammenfassung

Der Stoffwechsel oder Metabolismus ist ein unverzichtbarer Prozess für jedes Lebewesen. Es beschreibt die chemische Umwandlung aufgenommener Nahrung in Biomassebausteine, Redox-Äquivalente, Energie und Abfallprodukte. Das metabolische Potential einer Zelle kann durch ein Netzwerk dargestellt werden, in dem die verschiedenen Metabolite durch chemische Reaktionen miteinander verbunden sind. Diese chemischen Reaktionen wiederum werden durch Enzyme katalysiert, die im Genom der Zelle codiert sind. Lebende Zellen passen ihre metabolischen Reaktionsraten, auch metabolische Flüsse genannt, an die vorgefundenen Umweltbedingungen, den intrazellulären Status und die zellulären Anforderungen an. So gesehen widerspiegeln metabolische Flüsse das eigentliche Zellverhalten und deren Bestimmung ist daher essentiell um metabolische Prozesse von biotechnologischer und biomedizinischer Wichtigkeit zu verstehen und zu kontrollieren.

Metabolische Flüsse können nicht direkt gemessen werden und müssen deshalb indirekt von messbaren Größen, wie beispielsweise Genexpression, Proteinlevels, oder Veränderungen der Metabolitkonzentrationen abgeleitet werden. Experimentelle Ansätze, die mit stabilen Massenisotopen markierte Substrate verwenden, werden speziell für Flussbestimmungen eingesetzt. Dabei werden die Massenisotope der verschiedenen Metabolite mittels Kernspinresonanzspektroskopie oder Massenspektrometrie quantifiziert und miteinander verglichen. Die Interpretation der resultierenden Massenisotopenverteilungen erfolgt entweder manuell, mittels analytischen Formeln oder eines mathematischen Modells und gibt Aufschluss über die metabolischen Flüsse, die den gemessenen Isotopenverteilungen zu Grunde liegen.

Metabolische Flussanalysen können in globale ^{13}C Flussanalysen, welche iterativ diejenige Flussverteilung suchen, die die gemessenen Isotopenverteilungen am besten reflektiert und lokale ^{13}C Flussanalysen, die das Verhältnis mehrerer Flüsse zueinander mittels spezifisch dafür hergeleiteten analytischen Formeln bestimmen. Erstere sind allgemein anwendbar, benötigen aber umfangreiche Messdaten, resultieren in mehreren möglichen Lösungen und sind rechnerisch aufwendig. Im Gegensatz dazu sind lokale Flussanalysen rechnerisch schnell, einfach anzuwenden,

und resultieren in einer einzigen Lösung. Allerdings ist deren Anwendung auf einige wenige metabolische Knoten und experimentelle Voraussetzungen beschränkt. Daher besteht die Nachfrage für eine schnelle, robuste und allgemein anwendbare Methode, mit deren Hilfe auch die heutzutage typisch grossen Datensätze paralleler Experimente analysiert werden können.

In dieser Arbeit stellen wir SUMOFLUX, eine konzeptionell neue und allgemein anwendbare Methode für ^{13}C Flussanalyse spezifischer Flussverhältnisse vor. Diese Methode basiert auf Maschinellem Lernen um bestimmte metabolische Flussverhältnisse von gemessenen Daten vorauszusagen. Dazu wird zuerst ein *in silico* Datensatz modelliert – „**Surrogate Modelling**“. SUMOFLUX kann auf jeden Netzwerktyp, jedes Substrat und alle möglichen Messdaten angewandt werden und ist nach dem Erstellen der sogenannten „Predictors“ mittels *in silico* Modellierung rechnerisch sehr schnell. Im **2. Kapitel** führen wir SUMOFLUX ein und zeigen dass die berechneten metabolischen Flüsse in *Escherichia coli* mit den Resultaten globaler und lokaler ^{13}C Flussanalysen im Einklang sind. Weiter demonstrieren wir die Skalierbarkeit unserer Methode durch ihre Anwendung auf 121 Transkriptionsfaktorenmutanten in *Bacillus subtilis* und wir illustrieren, wie die Methode auch für die Optimierung des Experimentaldesigns verwendet werden kann.

Ein bemerkenswerter Vorteil unseres Ansatzes ist seine mögliche Anwendung auf komplexe Systeme, auch wenn nur wenige Daten von Messungen und zum metabolischen Netzwerk verfügbar sind. Diesen Vorteil nutzten wir im **3. Kapitel** aus, in dem wir den Aminosäurestoffwechsel von Mykobakterien unter axenischen Wachstumsbedingungen und während der Infektion von Makrophagen untersuchten. Wir teilten Aminosäuren aufgrund ihrer *in vitro* Aufnahme durch *Mycobacterium tuberculosis* und *Mycobacterium smegmatis* und ihrer metabolischen Verwendung, charakterisiert durch die errechneten Flussverhältnisse, in verschiedene Klassen ein. Untersuchungen des metabolischen Verhaltens von *M. tuberculosis* während der Infektion von Makrophagen, wenn weder der Stoffaustausch mit dem Wachstumsmedium noch derjenige zwischen den beiden Organismen bekannt sind, waren mittels extensiver SUMOFLUX Simulationen aller möglicher Stoffflüsse im kombinierten metabolischen Netzwerk von Makrophagen und dem Pathogen möglich. Diese Simulationen zeigten, dass die zum Wachstum normalisierte Biosynthese bestimmter Aminosäuren während der Infektion im Vergleich zu *in vitro* Bedingungen reduziert wird. Aus diesen Daten schlossen wir, dass einerseits nur

bestimmte Aminosäuren für die im Phagosom replizierenden Tuberkelbakterien zugänglich sind und dass andererseits auch Aminosäuren zur Energiegewinnung verstoffwechselt werden. Unsere Resultate unterstreichen die Anpassungsfähigkeit des mykobakteriellen Metabolismus und erklären zum Teil die wiederholten Misserfolge vergangener Versuche chemotherapeutisch den Stoffwechsel dieses Pathogens zu inhibieren.

Im **4. Kapitel** nutzten wir die hohe Geschwindigkeit und Flexibilität von SUMOFLUX um mittels FIA-TOF (flow injection - time of flight) ^{13}C Daten mit hohem Durchsatz zu messen und zu analysieren. Dies resultierte in mehreren Hundert Flussanalysen pro Tag. Diese hohe Messgeschwindigkeit ist der Tatsache zu verdanken, dass vollends auf chromatographische Trennung verzichtet wird, da die Proben direkt ins Massenspektrometer injiziert werden, was zu fehlenden und überlappenden Isotopologen führt. Dieses Problem lösten wir durch die Entwicklung von spezifischen Datenfiltern und Qualitätskontrollen eines jeden Massensignals. Wir validierten die Kombination von SUMOFLUX mit FIA-TOF erfolgreich mittels Analyse von mehreren *E. coli* Mutanten mit bekannten Flussverteilungen. Anschliessend wandten wir das entwickelte Protokoll auf 60 *E. coli* Phosphorylierungsmutanten an um die Rolle dieser in Prokaryoten wenig untersuchten posttranslationalen Modifizierung zu untersuchen. Unsere Flussanalyse bestätigte die bereits beschriebene Inhibition der Isocitratehydrogenase durch deren Phosphorylierung und identifizierte verschiedene neue funktionelle Proteinphosphorylierungen. Der entwickelte Ansatz zur Flussanalyse mit hohem Durchsatz stellt ^{13}C Fluxomics messtechnisch endlich auf Augenhöhe mit anderen Omics-Analysen.

Die Vorteile der entwickelten Methode lassen sich wie folgt zusammenfassen: Erstens, dank der allgemeinen Gültigkeit Maschinellen Lernens ist die Anwendung unserer Methode weder auf spezifische Organismen, experimentelle Bedingungen, Inputdaten noch bestimmte Flussverhältnisse limitiert. Zweitens, die eingebettete *in silico* Modellierung erlaubt *a priori* Annahmen zur metabolischen Netzwerkstruktur zu reduzieren und Flussanalysen auch für schlecht annotierte Netzwerke durchzuführen. Drittens, Berechnungen von lokalen Flussverhältnissen und deren extensive *in silico* Prüfung maximiert die Menge an Flussinformationen bei einem Minimum an notwendigen Messdaten. Letztlich, sobald die Modellierung abgeschlossen ist und die „Predictors“ berechnet sind, ist Integration der Messdaten

sehr rasch und kann mit dem Durchsatz modernster Messmethoden Schritt halten. Zusammenfassend lässt sich sagen, dass unser Ansatz fähig ist, lokale Flussverhältnisse in einem komplexen Setup zu beantworten, wie beispielsweise bakterielle Co-Kulturen, höhere Zellen in heterogenen Wachstumsmedien, und Wirt-Gast-Wechselwirkungen. Er eignet sich am besten zur Beantwortung gezielter Fragestellungen und Screens mit hohem Messdurchsatz. Wir sind überzeugt, dass der hier entwickelte Ansatz alleine, oder in Kombination mit globalen Flussanalysen, eine neue Ära von ^{13}C Fluxomics Analysen einläutet und künftig einen Beitrag zur Erforschung des Stoffwechsels und dessen Regulation leisten wird.

Chapter 1

General Introduction

CELLULAR METABOLISM AND ITS APPLICATIONS

Metabolism is an essential process for all living organisms. It consists of chemical transformations of molecules that sustain main cellular processes, such as growth, reproduction, maintenance, elimination of toxic compounds and response to external stimuli. These transformations, called metabolic reactions, are catalyzed by enzymes – proteins encoded in the cellular genome. Although some reactions can occur spontaneously, enzymes increase reaction rates manifold by lowering the activation energy. Metabolic reactions are assigned to two major processes – catabolism and anabolism. Catabolic reactions break organic molecules down into simpler units (pyruvate, lactate, acetate) to release energy in the form of ATP and reduced electron carriers (NADPH and NADH). Anabolic reactions construct complex molecules necessary for growth and maintenance (amino acids, nucleic acids, lipids) from smaller units, and require energy and reducing power of NADH and NADPH.

Catabolic and anabolic processes constitute central carbon metabolism, consisting of about sixty reactions highly conserved among different organisms from microbes to human¹. The central metabolic network includes three major pathways: glycolysis, pentose phosphate (PP) pathway and tricarboxylic acid (TCA) cycle. Glycolysis converts glucose to pyruvate in several enzymatic steps with concurrent formation of two molecules of ATP. In pentose phosphate pathway, intermediates of glycolysis are converted into pentoses that serve as nucleotide precursors, and NADPH is released for biosynthetic needs. Tricarboxylic acid cycle is used by all aerobic organisms to generate energy and amino acid precursors by oxidizing acetyl-CoA into carbon dioxide. In total, 36 molecules of ATP can be produced from one molecule of glucose in central carbon metabolism. Nowadays, based on genome sequencing and sequence similarities, metabolic networks of thousands of organisms have been described, spanning all kingdoms of life^{2,3}.

Whereas metabolic potential of a cell is encoded in its genome, the rates of metabolic reactions, or fluxes, result from a complex interplay of diverse molecular processes including gene transcription, protein translation and post-translational regulation. Therefore, metabolic fluxes represent the functional readout of the cellular physiologic or pathologic state. Over the past decades, technological advances resulted in exponential increase of knowledge about DNA, RNA, proteins and metabolites, and gave rise to genomics⁴⁻⁶, transcriptomics^{7,8}, proteomics^{9,10} and metabolomics¹¹⁻¹⁴.

Among these omics disciplines, metabolomics initially received insufficient attention, since it was considered being the “dull workhorse process” of the cell, solely reflecting the processes governed by cellular genome, transcriptome and proteome¹⁵. This conception was based on the fact that the majority of metabolic reactions are catalyzed by enzymes¹⁶, most of which are proteins encoded in the genome, which thus determines the metabolic potential of the cell. Metabolism is strikingly conserved across all domains of life, with central metabolic pathways being present in all known organisms¹⁷. However, the operation and activity of these pathways vary drastically depending on the organism and conditions, which could not be explained by transcriptional regulation alone¹⁸, underlining the active control of metabolism that often occurs post-translationally, enabling rapid adaptations to the changing conditions and needs of the organism^{19,20}. Hence, investigation of metabolism became an important aspect of systems biology, which aims at understanding cellular processes as a complex interplay of many components of a living system^{15,21-23}.

Originally, metabolism was studied at the macroscopic level of the whole organism. For example, estimates were reported for the conservation of energy in the human body^{24,25} and changes in inorganic, carbohydrate and oxygen intermediates in urine and blood upon perturbations^{26,27}. Besides, since the first enzyme was purified²⁸, enzymatic reactions were studied *in vitro* by following the kinetics of reaction intermediates^{29,30}. Undoubtedly, kinetic information is essential to understand how a steady-state flux through a pathway can be maintained and how it can be changed from one rate to another. However, the discrepancies between *in vitro* enzyme kinetics and *in vivo* system behavior might be too large to deduce the information about metabolic fluxes from *in vitro* experiments³¹.

With the development of experimental, analytical and computational methods, it became possible to probe *in vivo* metabolic fluxes³²⁻³⁶, measure a vast variety of intracellular metabolites³⁷⁻⁴⁰, and build genome-scale metabolic models for investigation and prediction of cellular behavior⁴¹⁻⁴⁵. These techniques greatly advanced the fields of metabolic engineering, biomedical sciences and drug discovery.

Genetic engineering of bacteria has been used in modifying plant-microbe-pest interactions, biodegradation of xenobiotics and toxic waste transformations, production of chemicals and fuels, mineral processing, wastewater treatment applications^{46,47}. With the development of molecular biological techniques for DNA recombination, metabolic pathway modification became directed as opposed to random mutagenesis

and subsequent selection⁴⁸. Integrative metabolic pathway analysis and genome scale modeling became powerful methods for the systematic improvement of cellular properties in a broad range of contexts and applications⁴⁸⁻⁵³.

Another broad area of metabolic engineering applications is in the field of biomedicine and drug discovery^{48,54}. Metabolic dysregulation plays a role in various diseases, such as obesity⁵⁵, diabetes⁵⁶ and cancer^{57,58}. Genome scale modeling, omics measurements and metabolic flux analysis revealed the role of proliferative adaptation in causing the cancer-specific Warburg effect⁵⁹, predicted drug targets inhibiting cancer migration⁶⁰ and proliferation^{61,62}, explained metabolic changes caused by mutations in oncogenes⁶³⁻⁶⁵, and proposed combinatorial therapies based on synthetic lethality predictions⁶⁶. Despite the extensive knowledge about cancer metabolism, therapies often fail due to constant metabolic adaptations of tumor cells to the treatments⁶⁷⁻⁶⁹.

Similar to the vast applications in metabolic diseases, combination of omics techniques with metabolic flux analysis methods proved fruitful in investigating metabolism of pathogens and host-pathogen interactions. Metabolic enzymes are attractive drug targets, and genome-scale metabolic modeling is widely used to predict essential and synthetically lethal genes in various pathogens⁷⁰⁻⁷⁴. However, microbial resistance to antibiotics remains a serious threat for human health^{75,76}. A large-scale *in vivo* screen of *Salmonella enterica* mutants⁷⁷ and multi-omics studies in mycobacteria⁷⁸⁻⁸⁰ revealed robustness of bacterial metabolism, reflected in redundancy of the majority of the enzymes and rapid adaptation capability. This observation suggests that bacteria have evolved to consume multiple nutrients from their host during infection despite its elaborate defense mechanisms, which has been proposed in several studies^{77,81-83}. Although first attempts in modeling host-pathogen^{81,84} and host-microbiome^{85,86} interactions at a genome scale succeeded in predicting macroscopic parameters, such as biomass and ATP production or species co-occurrence, significant improvements in estimation and analysis of interspecies fluxes and metabolic adaptations of both the host and the bacteria have to be achieved in order to understand the behavior of this complex system and propose new therapeutic and prophylactic approaches to treat and prevent infectious diseases⁸⁶⁻⁹⁰.

EARLY METHODS FOR *IN VIVO* FLUX PROFILING

Already in the first half of the XX century isotope labeling techniques were used to probe cellular metabolism *in vivo*^{32,91}. The early studies of central carbon metabolism were performed by administering [1-¹⁴C] and [6-¹⁴C] glucose to the cells and measuring the ¹⁴C yield in CO₂ and triose phosphate derivatives such as lactate and fatty acids^{92,93}. Glucose recycling through the pentose phosphate pathway was manually calculated taking into account the possibility of multiple cycles⁹⁴. A general model for glycolysis, the PP pathway and the Entner-Doudoroff pathway was proposed along with equations estimating specific reaction activities³⁴. This approach was used to estimate exchange fluxes through transaldolase reactions⁹⁵, study the effect of diet and hormones on the PP pathway in rat adipose tissue^{96,97}, and the PP pathway activity in different tissues in rodents³³ and in human⁹⁸. The manual data analysis process included crude simplifications due to limited measurements (mostly CO₂, lactate, glycogen and glycerol were available), and insufficient network topology information (for example, TCA cycle activity was estimated solely based on ¹⁴C incorporation in excreted CO₂, ignoring other CO₂ producing reactions)⁹⁸. Moreover, these studies were impeded by the difficulty of conducting experiments, deficit and high price of the labeled substrates, and health concerns due to exposure to radioactivity⁹⁹.

In parallel to radioactive glucose ¹⁴C isotopes, stable ¹³C isotopes were also exploited to study malonate metabolism in mice¹⁰⁰ and carbon transport in plants¹⁰¹. With the advances of stable isotopes' chemical production and detection with nuclear magnetic resonance (NMR) spectroscopy and mass-spectrometry (MS) techniques, ¹³C labeling superseded the radioactive labeling experiments^{99,102}. Central carbon metabolism of *Escherichia coli* and higher cells were investigated by manual inspection of ¹³C labeling patterns detected by NMR and label "scrambling"¹⁰². Linear equations formulated for each labeled atom were solved manually in an iterative procedure in terms of the known specific activities of the substrates added to the cell suspension by the investigator¹⁰³. This process was prone to errors and an assay for complex models could take a year to perform. Computational power was only sufficient to solve stoichiometric equation systems of up to 50x50 in size¹⁰⁴. Few computers were capable of inverting a stoichiometric matrix of size 130x130, and even if the system was solved programmatically, the speed advantage gained by using the computer came with a loss of manual trial-and-error fitting¹⁰³.

EVOLUTION OF METABOLIC FLUX ANALYSIS: FINDING THE MISSING CONSTRAINT

Further developments of metabolic flux analysis methods went hand in hand with the rapid advances in experimental, analytical and computational technologies. Since molecular techniques for DNA manipulation became available, there was a boom in metabolic engineering, causing the demand for rational approaches to optimal strain design^{46,47,105}. Metabolic control theory¹⁰⁶ was proposed to identify enzymes having the most influence on the metabolic fluxes given kinetic parameters and metabolite concentrations¹⁰⁵. However, enzyme amplification in the anabolic pathway of the desired product not necessarily leads to the increase of its production, since primary metabolism is usually robust to large flux changes at the branch points¹⁰⁷. Therefore, it was necessary to develop rational methods to estimate metabolic fluxes in order to predict optimal yields and strategies to achieve them^{47,107}. Although the ¹³C labeling experiments could provide *in vivo* flux estimates, they were expensive, labor intensive and limited in the number of possible measurements and therefore resolvable fluxes. Instead, metabolic balancing approaches based on the steady state assumptions were used to calculate metabolic fluxes given the stoichiometric matrix, biomass composition, uptake, secretion and growth rates. For example, such approaches were applied to study citrate production in *Candida lipolytica*¹⁰⁸, environmental adaptations in *Clostridium acetobutylicum*¹⁰⁹, *Acinetobacter calcoaceticus*¹¹⁰, and lysine fermentation by *Corynebacterium glutamicum*¹¹¹. Additionally, procedures for singularity, sensitivity and consistency analysis in metabolic networks were proposed to identify principal nodes which have to be modified in order to drain primary metabolism towards the desired product formation¹¹¹.

Computational models and linear optimization algorithms^{112,113} enabled automatic analysis of the effect of stoichiometric constraints on the metabolic behavior of an organism in a given environment. For example, linear programming was used to examine the interactions between stoichiometric constraints and the requirement for efficiency in conversion of glucose to fat in adipocytes¹¹⁴. Most of the studied metabolic networks were underdetermined, since the number of fluxes exceeded the number of metabolites, and some reactions were bidirectional or formed futile cycles. Hence, there was a need for additional constraints or optimization principles in order to determine the flux distributions¹¹⁵.

Introducing additional measurements. The easiest way to reduce the number of degrees of freedom in the metabolic network is to include additional flux measurements. Usually, these constraints were added by measuring CO₂ or adding cofactor balancing equations (ATP, NAD(P)H, O₂, CO₂, NH₃)¹¹⁵⁻¹¹⁷. Unfortunately, some of the cofactor balances cannot be closed, due to inability to precisely measure cellular ATP maintenance, or due to existence of transhydrogenase converting NADPH to NADH. The latter issue can be solved by lumping these cofactors into one pool¹¹⁶. However, there are only a few degrees of freedom that can be reduced with cofactor balancing, and the system might still remain underdetermined. Generally, a method was proposed to determine which fluxes have to be additionally measured in order to minimize the errors in the estimated fluxes^{118,119}.

Objective functions. Searching for a flux distribution which solves a given optimization problem is called flux balance analysis (FBA). Optimization for growth as a representation of natural selection process was initially proposed for prokaryotic organisms⁴⁵. Growth is modeled with a biomass vector, for which the biomass coefficients have to be determined experimentally^{116,117,120,121}, as biomass composition may differ in different bacterial strains, cell lines and culture conditions¹¹⁶. Nowadays, however, generalized biomass vectors are usually used for flux estimation.

The maximization of biomass is not well suited to study prokaryotes which did not undergo evolutionary pressure, such as genetic knockouts. In such cases, it was proposed to predict flux distributions by minimizing the mutant's metabolic adjustment compared to the parent wild type flux distribution¹²². However, it was demonstrated that after several generations, the flux distributions of the knockouts approach those predicted by optimizing biomass^{123,124}. Although there is a unique optimal value for the optimization function, multiple flux distributions could lead to the solution, potentially representing alternative evolutionary strategies¹²⁵.

Alternative widely used objective functions, especially used to model higher cell metabolism, include maximization of ATP or NAD(P)H production or minimization of the Euclidean norm of fluxes¹²⁶. For optimal strain design, the desired compound production is optimized. In this case, FBA solutions represent the maximal theoretical yield, and can be used to analyze discrepancies between the optimal solution and experimental measurements^{48,127}. In this way, maximum yields were calculated for riboflavin producing *B. subtilis*¹²⁸ and ethanol producing yeast¹²⁹. With FBA, it is

possible to study the influence of different optimization functions on the flux distribution in the metabolic network¹³⁰, and the roles of different metabolites in sustaining different optimization requirements by computing their shadow prices^{130,131}.

FBA is sensitive to assumptions on biomass composition, cofactor balancing¹²⁰, constraints and objective function. Therefore, instead of predicting exact network behaviors, it is useful to analyze the states which the system theoretically can achieve and which it cannot under given conditions^{70,132}. By maximizing ATP production and analyzing the network constraints, FBA was used to explain methanol secretion in yeast and acetate secretion in *E. coli*¹³³, differences in lactate consuming and lactate producing higher cells¹³⁴, and investigate genetic knockout effect on the mitochondrial ATP production¹³⁵. Minimization of the Euclidean norm of fluxes was used as an optimization function to compare hybridoma cells metabolism in two culture media¹¹⁶.

Flux distributions obtained by solving different optimization problems can be validated with the flux estimates from ¹³C labeling data. In hybridoma cells, for example, flux solutions obtained by maximizing ATP or NADH maximization resulted in better concordance with ¹³C flux estimates compared to the flux solution obtained by minimizing the sum of fluxes¹³⁶. In a large study on the optimality of microbial metabolism, it was found that five objective functions are consistent with 44 *in vivo* flux distributions of different bacteria estimated with ¹³C labeling: ATP, biomass, acetate and carbon dioxide yields and minimum sum of absolute fluxes¹³⁷. However, instead of following a single optimization strategy, bacterial metabolic fluxes were close to Pareto surface describing three of the optimization functions – ATP and biomass maximization and optimal resource allocation. It was proposed and further supported by evolution experiments, that bacteria having suboptimal flux distributions achieve evolutionary benefit by their ability to rapidly switch between conditions¹³⁷.

Thermodynamic constraints. The second law of thermodynamics can be used to constrain the direction of all network fluxes by calculating the sign of the change in the Gibbs free energy for each reaction¹³⁸. The change in the Gibbs free energy depends on the temperature, concentrations of substrates and products, and change in the Gibbs free energies in a reference state, which can be estimated using group contribution^{139,140} or component contribution¹⁴¹ methods. In the absence of quantitative metabolomics measurements, thermodynamic assignment of reaction directionality can be made by assuming an accommodating range of reactant concentration (0.02 – 20 mM)¹⁴². The

flux distributions generated using thermodynamic constraints involve neither fluxes through any thermodynamically unfavorable loops, nor fluxes through thermodynamically unfavorable reactions given the concentration ranges found in the cell¹³⁹. Whereas the majority of *E. coli* reactions were found to be thermodynamically optimized, a few unfavorable reactions with estimated positive Gibbs free energy changes were identified as potential bottlenecks in the production of growth¹⁴⁰. Concurrently, the reactions with large negative Gibbs free energy changes were suggested as potential targets of active allosteric or genetic regulation^{139,143}. Perturbation analysis by varying energy values and metabolite concentrations proved to be useful for the determination of the thermodynamically feasible activity ranges for the metabolites¹⁴⁰ and identification of critical reactions for which fluxes are largely undetermined by the FBA approach¹⁴⁴. Thermodynamic approaches were applied to evaluate thermodynamic and kinetic quality of different pathways in central metabolism¹⁴⁵ and genome-scale network of *E. coli*¹⁴⁶. Although thermodynamic-based flux analysis improves flux predictions and provides crucial information on the kinetic properties and reaction capabilities in the model, its application is hampered by the lack of detailed information on enzyme kinetics, quantitative measurements and thermodynamic properties under different environmental conditions¹⁴⁷.

Integration of Omics data. With the development of large scale genomics, transcriptomics, proteomics and metabolomics techniques, flux analysis in metabolic networks was brought to a new level^{44,148}. Since the publication of the first *E. coli* genome sequence⁴, large-scale network topology reconstructions became possible with gene homology analysis by mapping genes to enzymes and reactions they can catalyze¹⁴⁹⁻¹⁵³. Nowadays, hundreds of genome annotations are available, and the process of genome-scale model reconstructions has been standardized^{42,154,155} and facilitated with a number of automatic tools^{43,156-158}. These models are gradually refined, as new genome annotations and functional characterizations of enzymes emerge^{159,160}.

While genomic information determines the topology of a metabolic network, transcript and protein expression data can be included as constraints to shrink the solution space and to build condition specific models^{161,162}. For instance, gene expression profile can impose binary activity constraints, i.e. if a gene is expressed below a certain threshold, the reaction is inactive^{163,164}. In a more flexible way,

expression values can be converted to continuous flux bounds¹⁶⁵ or used to discriminate flux solutions between conditions using differential analysis thresholds¹⁶⁶. Based on gene expression and proteomics data, specific human tissue models¹⁶⁷, cancer models¹⁶⁸ and host-pathogen interaction models⁸⁴ have been developed.

Apart from being translated into direct constraints on metabolic reactions, expression data can be exploited to incorporate mechanisms of transcriptional regulation or protein allocation into the models. First genome-scale transcriptional regulatory and metabolic network of *E. coli* was used not only to improve flux predictions, but at the same time to indicate knowledge gaps and identify novel components and regulatory interactions by analyzing discrepancies between model predictions and experimental outcomes¹⁶⁹. Steady state regulatory FBA incorporating Boolean regulatory constraints allowed to quantify the extent to which transcriptional regulation determined metabolic behavior in *E. coli* and identify redundantly expressed genes¹⁷⁰. Integrative omics-metabolic analysis, which incorporates measured proteomics and metabolomics data into flux rate constraints with kinetic equations, was shown to improve metabolic flux estimates in *E. coli* and human erythrocytes¹⁷¹.

Combined models of metabolism and molecular expression include proteomic constraints to predict gene expression along with growth rate and fluxes accounting for metabolic costs of proteins^{41,172,173}. Including cellular networks beyond metabolism, such as gene product expression coupled to metabolism, protein translocation in the cellular membrane, protein structures of metabolic enzymes, and transcriptional regulation, enable novel predictions on different cellular levels⁴⁹, and bring us closer to developing whole-cell models¹⁶¹. However, it is important to note that although different omics measurements provide additional layers of information for flux analysis, their incorporation into flux constraints relies on numerous assumptions. Therefore, the estimated flux distributions should be validated experimentally, for example, with stable isotope labeling techniques¹⁷⁴.

Global ¹³C metabolic flux analysis. Despite being the most direct experimental technique for flux elucidation, for a long time ¹³C labeling experiments were only occasionally used to validate metabolic fluxes predicted by FBA, mostly due to high cost of tracers and limited availability of metabolite measurements^{121,136}. Developments of NMR spectroscopy and mass spectrometry analytic techniques expanded the number of measurable metabolites and enabled the investigation of larger models^{40,175-177}.

Compared to FBA, ^{13}C metabolic flux analysis offers additional independent constraints, therefore allows to omit energy and cofactor balance equations and enables quantification of both flux directions of bidirectional reaction steps¹⁷⁸. Combined, metabolite balancing and ^{13}C tracer experiments could offer a much more powerful flux analysis approach. Thus, a general modeling framework for ^{13}C metabolic flux analysis (^{13}C -MFA) was proposed, which was based on principles of metabolic balancing, included bidirectional reaction steps and modeled carbon flow with atom transition equations^{178,179}. The system of carbon transition equations could no longer be handled with linear methods, because bimolecular reactions had to be described with quadratic equations. Also, the system could not be solved manually, since 25 fluxes between 20 pools were represented with 3200 isotopomer fluxes between 600 isotopomer pools³⁵. Further technical improvements included representation of atomic equations with the so-called cumomers, leading to networks with cascade structure that could be solved linearly¹⁸⁰. As an alternative to cumomer equations, genetic algorithms¹⁸¹ and the elementary metabolite unit framework based on a highly efficient decomposition method that identifies the minimum amount of information needed to simulate isotopic labeling within a reaction network were proposed, which considerably reduced the computation time¹⁸². Increase in available computational powers led to the development of efficient ^{13}C -MFA platforms facilitating metabolic flux analysis in different organisms¹⁸³⁻¹⁸⁶.

It was shown that ^{13}C -MFA produces more precise flux estimates in riboflavin producing *B. subtilis*¹⁸⁷, *E. coli*¹⁸⁸ and *S. cerevisiae*¹⁸⁹. Moreover, fast calculations and analytical expressions for metabolic flux analysis sensitivities enable experimental design by comparing different experimental setups (labeling techniques, measured metabolites) with respect to the achievable flux information¹⁹⁰⁻¹⁹⁵.

Although ^{13}C -MFA provides a global flux estimate for the given metabolic network, often there are multiple equally plausible solutions, and global optimum cannot be guaranteed³⁵. In practice, the parameter fitting is often performed several hundred times, and the best fit among all solutions is reported^{164,196,197}, what further increases computational time hindering large-scale applications of global ^{13}C -MFA. Moreover, with increased network complexity, the number of measurements required for constraints also increases, leading to infeasibility of accurate flux estimates in complex systems, such as multi-nutrient environments or co-cultures.

Local ^{13}C flux ratio analysis. As an alternative to global ^{13}C -MFA, targeted flux ratio analysis methods were developed based on analytic equations connecting a specific flux ratio, usually between two or more reactions forming a single metabolite, and measurable mass isotopologues^{198,199}. These targeted approaches require only few measurements and are easy to apply, interpret and compare estimates across samples. These advantages enabled large-scale ^{13}C flux ratio profiling of *B. subtilis* and *E. coli* transcription factor mutants^{200,201} and enzymatic knockouts of yeast²⁰². Flux ratio estimates can serve as additional constraints for FBA to provide global flux solutions^{203,204} also in genome-scale networks²⁰⁵. Flux ratio equations require manual derivation based on strong assumptions on the network structure¹⁹⁸. Therefore, to date only a handful of analytic formula exist, and they are applicable in restricted cases for specific organisms. Although an automated procedure was proposed to derive flux ratios using graph algorithms²⁰⁶, it was limited to reactions producing the same junction metabolite resolvable by linear means, and did not find broad application.

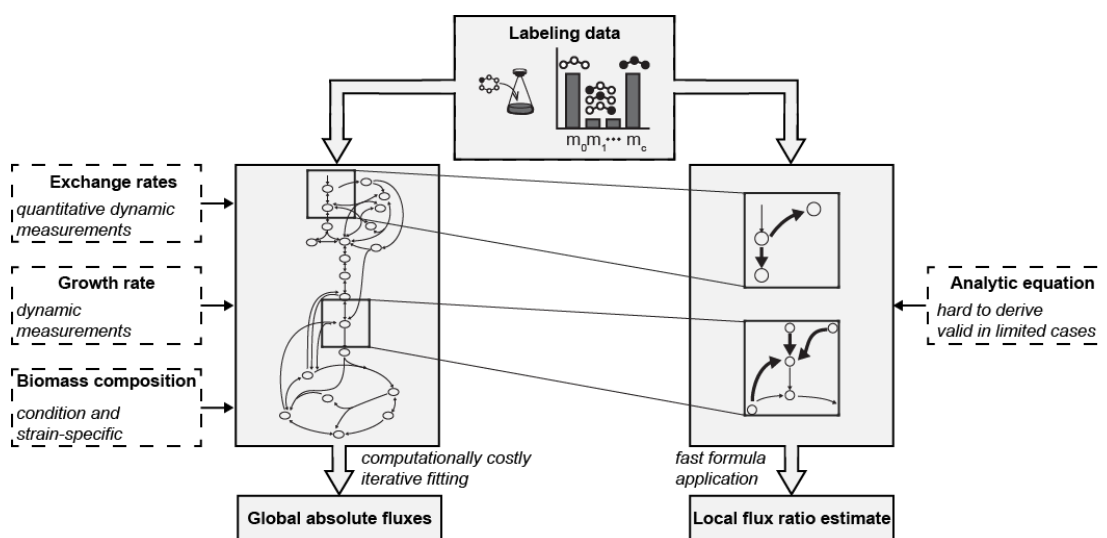


Figure 1. Schematic comparison of global and local stationary ^{13}C metabolic flux analysis approaches.

NON-STATIONARY FLUX ANALYSIS APPROACHES

The stationary flux analysis approaches are based on a strong assumption of metabolic and isotopic steady state. Although this assumption is valid in most experiments with prokaryotes growing in minimal media with one or two carbon sources, it might not hold true in case of sequential substrate utilization, or slow growing organisms and mammalian cells. A parallel branch of flux analysis methods aims at estimating metabolic fluxes in metabolic or isotopic non-stationary conditions. FBA and metabolic balancing have been successfully extended to dynamic systems by incorporating time-series metabolomics measurements and kinetic models²⁰⁷. The key assumption that must hold for dynamic flux analysis approaches is the maintenance of the pseudo-steady state, which means that the time-scale for process dynamics must be longer than the time-scale for intracellular equilibration with the extracellular environment²⁰⁷. These approaches proved useful to iteratively fine-tune metabolic flux estimations by incorporating dynamic measurements into kinetic models to simulate bioprocesses and generate hypotheses on optimal conditions and system behavior, which in turn has to be validated experimentally.

The ¹³C-MFA method was extended to INST-MFA (isotopically non-stationary MFA)²⁰⁸, which is performed in a transient labeling state, enabling short-time experiments. Compared to ¹³C-MFA, INST-MFA provides more information, especially in case of missing structural information, multiple carbon sources and growth on C₁ substrates²⁰⁹. Moreover, statements about the magnitude of some non-measurable pool sizes can be made²⁰⁸. Comparison of INST-MFA with stationary ¹³C-MFA on the data of *Corynebacterium glutamicum* grown in fed-batch cultivation revealed that only the isotopic non-stationary approach led to confidential flux estimations under studied conditions²¹⁰. However, INST-MFA is much more challenging to perform and interpret than other flux analysis techniques. It requires extensive knowledge on the reaction activity and kinetic parameters, large computational powers to solve the systems of ordinary differential equations, and accurate measurements of metabolite concentrations.

In principle, non-stationary flux analysis methods are complementary to the steady state approaches. Stationary analysis supplies cornerstones of dynamic modelling, as it describes system behavior around steady state. Resolving bidirectional fluxes may provide an additional input into the kinetic model. Kinetic models have

more parameters to depend on – reversible enzymatic mechanisms have to be taken into account^{35,209}. The constraints obtained with targeted non-stationary flux analysis methods, in turn, can be used to improve the global flux predictions using steady state approach²¹¹.

Targeted ¹³C flux ratio analysis performed in a dynamic system offers reduction of necessary measurements, simpler kinetic model, and independence from the network structure not involving the flux ratio of interest²¹¹. Since only the first seconds to minutes of label propagation are monitored, this approach allows to considerably reduce experiment duration and consumption of labeled media. Moreover, it enables to elucidate flux propagation through linear pathways, which are not accessible by stationary methods, therefore should be preferred in case such flux ratio formalism is applicable to the system of interest.

MACHINE LEARNING APPROACHES FOR FLUX ANALYSIS

Machine learning is a conceptually different way of data analysis based on inferring a predictive model from a set of known inputs and corresponding outputs, in order to generalize and provide output estimates for new input samples. Machine learning applications in systems biology are emerging^{212,213}. With the increasing number of omics datasets, the use of machine learning for predicting disease progression^{214,215}, regulatory events²¹³ or drug response²¹⁶⁻²¹⁸ becomes more and more widespread.

In the field of metabolomics, machine learning was used to predict metabolic pathways from genomic data²¹⁹, kinetic parameters of the models^{220,221}, or extract informative features distinguishing metabolic phenotypes²²². Since there is a lack of experimental data from ¹³C flux analysis, machine learning has not been extensively applied in fluxomics studies. The first attempt to predict fluxes from the information on organism, condition and genetic modification in bacteria was performed by training a predictive model on a dataset collected from 120 literature sources²²³. In another study, analogous input information was combined with basal gene expression levels to sequentially predict gene and protein expression, metabolite abundance and metabolic fluxes in *E. coli*²²⁴. Although in both studies machine learning predictors provided reasonable results for the test datasets, their application might be limited to closely related bacterial species, since the input information is restricted to a predefined set of strains and media composition.

The main drawback of machine learning is that it can only extract information from the provided training sample, therefore cannot predict fundamentally new outcomes. It is also important to consider that machine learning algorithms have to be adequately validated on a dataset not included in training or tuning their parameters. Due to the lack of data, it is not always possible to obtain large non-intersecting training and testing datasets, therefore some of the validation results might be overoptimistic²²⁵. Machine learning is a powerful tool for large-scale data analysis, however, its application is far from mechanical²¹². Successful application of machine learning requires careful formulation of the prediction task, choice of input parameters and prediction method, and appropriate procedures for tuning, testing and diagnosis of the predictors.

FRONTIERS AND PERSPECTIVES OF FLUX ANALYSIS IN COMPLEX SYSTEMS

Ideally, metabolism should be studied *in vivo* at a genome-scale and single cell levels^{54,174,226,227}. In reality, the complex problem of metabolic operation of a living system is reduced to a subset of questions which can be tested in controlled laboratory experiments with available measurement techniques. The development of new experimental, measurement and data analysis methods slowly but gradually enables to increase the complexity of laboratory studied systems, in order to move a step closer to understanding *in vivo* processes.

Whereas genome-wide transcriptomics and proteomics data are widely used to constrain genome-scale models for flux balance analysis^{148,171,173}, ¹³C flux analysis has not yet achieved this advanced level²²⁸. The first genome-scale ¹³C-MFA in *E. coli* proved that reducing the network by eliminating *a priori* blocked reactions and exploiting the elementary metabolic unit decomposition allows to solve the flux analysis problem computationally and obtain a better fit compared to the core metabolism model²²⁹. In this study, only 17 measured amino acids were used to fit the fluxes, hence some improvements in decreasing the uncertainties of the genome-scale model may be achieved by including more measurements or combining data from parallel labeling experiments^{229,230}. However, if alternative pathways produce identical intermediate carbon transitions, additional measurements and tracer labeling schemes would not be able to resolve such pathway ambiguity. In this case, combination of

targeted ^{13}C flux analysis with metabolite balancing or dynamic labeling experiments might be a better solution.

One of the traits of the *in vivo* systems that is often omitted *in vitro* is the inter-cellular interaction. Most studies of bacterial pathogens or cancer cells aimed at discovery of new therapeutic treatments investigate the system behavior in the defined (sometimes even minimal) media, what often leads to discrepancies between the expected and the observed *in vivo* behavior of the candidate knockout strain²³¹ or mutated cancer cells²³². Although *in vitro* host-pathogen interaction systems have been established^{77,233,234}, the inter-species flux analysis is hindered by the difficulty of the accurate material separation and quantification. Whereas it is possible to detect gene expression changes in multi-organism system with dual RNA-seq techniques^{235,236}, metabolomics measurements and thus more direct flux estimations are infeasible due to the universality of metabolic compounds. In a study of host-pathogen interactions during early infection, ^{13}C flux spectral analysis was performed using measurements of protein bound amino acids derived after separating bacteria from macrophages by centrifugation⁸¹. The combined metabolic network of host and pathogen central metabolism was used to predict potential nutrient fluxes by sampling thousands of flux distributions assuming different single, double and triple amino acid exchange fluxes and ranking the solutions by residual valine. Although this study pioneered inter-species ^{13}C flux analysis, there is a considerable imbalance between the computational effort and its outcome due to the limited amount of measurements and large system uncertainty, which could be potentially reduced by applying a less computationally intensive targeted flux ratio analysis approach. While measuring the protein-bound amino acids or recently proposed peptide²³⁷ or RNA²³⁸ labeling patterns is a promising approach to separate metabolic material in co-cultures, such data is only reliable assuming a steady state for fluxes and labeling patterns²³⁷. In case the label exposure time is short compared to the organism doubling time, these measurements represent an integration of previous labeling states, which has to be either included in the model or at least interpreted with caution. As an alternative, it was proposed to model the mixed metabolic measurements in the bacterial co-culture by adding a species fraction coefficient as an additional parameter of the optimization²³⁹, which proved applicable to a simple co-culture of two *E. coli* strains. Such approach might become the first step to analyze interactions between the gut bacteria, which recently receive more and more attention^{88,240,241}.

Further simplification of a complex inter-cellular behavior is studying one organism's metabolism in a complex media mimicking the interaction space. High-coverage untargeted metabolomics³⁸ can be used to measure all potential substrate and product exchange rates to be used as FBA constraints, or refine condition-specific metabolic models to predict intracellular states²⁴². Parallel labeling experiments with different substrates labeled at a time can provide additional information on substrate utilization and enable more precise flux estimates^{197,226,234,243}.

Finally, although flux analysis at the single cell level still sounds somewhat futuristic, it is inevitable to understand metabolic phenotypes of subcellular populations in order to fight cancer⁶⁸ or pathogen⁷⁶ drug resistance. Nowadays, when the design of FRET (Förster resonance energy transfer) biosensors providing quantitative dynamic measurements of a specific metabolite at a single-cell level is facilitated²⁴⁴, and advances in different mass spectrometry techniques enable metabolomics measurements in ultra-small volumes²⁴⁵, application of single-cell fluxomics becomes more and more plausible^{226,246}. Since analyzing single cell fluxes would require high-throughput fitting to potentially small amount of measurements lacking physiological parameters, targeted flux analysis methods would be among the first candidates to tackle this exciting challenge²²⁶.

THESIS OUTLINE

The main contribution of this thesis was the development of a fast and generalized method for targeted ¹³C flux ratio analysis, SUMOFLUX, which is applicable to virtually any metabolic system, steady state conditions and type of ¹³C measurements.

In **Chapter 2**, we develop the general workflow of SUMOFLUX, which consists of **surrogate modelling** of the labeling experiment in order to train and test machine learning predictors for targeted flux ratios given the metabolic network, substrate(s) label, available measurements and noise. As a proof of principle, we apply SUMOFLUX to resolve key flux ratios in a small set of *Escherichia coli* knockout strains, and demonstrate that our predictions are in agreement with estimates of classical analytic formulas and global ¹³C-MFA approach. We underline the generality and scalability of SUMOFLUX by resolving a novel flux ratio for glyoxylate shunt using the same data, and by performing flux ratio analysis in a large cohort of *Bacillus*

subtilis transcription factor mutants. Lastly, we illustrate the suitability of SUMOFLUX workflow for sensitivity analysis and optimization of experimental design.

In **Chapter 3**, we exploit SUMOFLUX to study amino acid metabolism of mycobacteria in complex environments. Amino acids play an important role during tuberculosis infection, but their co-metabolism and availability inside the host are scarcely understood. We first characterize amino acid utilization patterns of the model organism *Mycobacterium smegmatis* in glucose and amino acid two-substrate media by estimating specific catabolism and biosynthesis ratios with SUMOFLUX. Next, we perform co-utilization analysis in a more complex system of *M. tuberculosis* growing in rich medium and in the host-pathogen infection setup. We conclude that mycobacteria are rapidly adapting to co-feed, and are prepared to take up nutrients as soon as encounter them, which seems to occur even in the coarse phagosome environment.

In **Chapter 4**, we take advantage of the speed and generality of SUMOFLUX to develop a fast protocol for high-throughput flux profiling by applying it to the data measured with FIA-TOF (flow injection analysis – time of flight) mass spectrometer. With the proposed experimental and data analysis workflow, hundreds of samples can be analyzed within a day. Using a set of *E. coli* knockout strains with known flux phenotypes, we establish a data pre-processing procedure, which includes peak filtering and quality check, and demonstrate that SUMOFLUX can predict expected flux phenotypes given these data. We apply the proposed workflow to a large cohort of *E. coli* phosphorylation mutants to generate hypotheses on the functional role of enzyme phosphorylation. The developed high-throughput flux profiling workflow paves the way for large-scale comparative ¹³C fluxomics.

ADDITIONAL PROJECTS

These projects were also part of this work, but they are not included in this thesis.

Integration of metabolomics and transcriptomics reveals a complex diet of *Mycobacterium tuberculosis* during early macrophage infection.

Michael Zimmermann*, Maria Kogadeeva*, Martin Gengenbacher, Gayle McEwen, Hans Joachim Mollenkopf, Nicola Zamboni, Stefan Hugo Ernst Kaufmann, and Uwe Sauer.

Manuscript under revisions

* *Equally contributed to this work*

Contribution of MK: Analyzed and interpreted the data, wrote parts of the manuscript.

L-Arginine modulates T cell metabolism and enhances survival and anti-tumor activity.

Roger Geiger, Jan C. Rieckmann, Tobias Wolf, Camilla Basso, Yuehan Feng, Tobias Fuhrer, Maria Kogadeeva, Paola Picotti, Felix Meissner, Matthias Mann, Nicola Zamboni, Federica Sallusto, Antonio Lanzavecchia.

Cell 167 (2016): 829-842.

Contribution of MK: analyzed the data from metabolome and flux experiments

Next-Generation “-omics” approaches reveal a massive alteration of host RNA metabolism during bacteriophage infection of *Pseudomonas aeruginosa*.

Anne Chevallereau, Bob G. Blasdel, Jeroen De Smet, Marc Monot, Michael Zimmermann, Maria Kogadeeva, Uwe Sauer, Peter Jorth, Marvin Whiteley, Laurent Debarbieux, Rob Lavigne.

PLoS Genetics 12, no. 7 (2016): e1006134.

Contribution of MK: analyzed the data from metabolome experiments

High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection.

Jeroen De Smet, Michael Zimmermann, Maria Kogadeeva, Pieter-Jan Ceysens, Wesley Vermaelen, Bob Blasdel, Ho Bin Jang, Uwe Sauer, and Rob Lavigne.

The ISME journal (2016).

Contribution of MK: analyzed the data from metabolome experiments

Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*.

Olga T. Schubert, Christina Ludwig, Maria Kogadeeva, Michael Zimmermann, George Rosenberger, Martin Gengenbacher, Ludovic C. Gillet, Ben C. Collins, Hannes L. Röst, Stefan H.E. Kaufmann, Uwe Sauer, Ruedi Aebersold

Cell host & microbe 18, no. 1 (2015): 96-108.

Contribution of MK: Analyzed and interpreted the data, wrote parts of the manuscript.

Regulation of yeast central metabolism by enzyme phosphorylation.

Ana Paula Oliveira, Christina Ludwig, Paola Picotti, Maria Kogadeeva, Ruedi Aebersold, and Uwe Sauer.

Molecular Systems Biology 8, no. 1 (2012): 623.

Contribution of MK: Contributed to data analysis.

REFERENCES

- 1 Lehninger, A. L., Nelson, D. L. & Cox, M. M. Principles of biochemistry. *New York: Worth* **38** (1993).
- 2 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30 (2000).
- 3 Caspi, R. *et al.* The MetaCyc database of metabolic pathways and enzymes and the BioCyc collection of pathway/genome databases. *Nucleic acids research* **44**, D471-D480 (2016).
- 4 Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453-1462 (1997).
- 5 Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860-921 (2001).
- 6 Lander, E. S. The new genomics: global views of biology. *Science* **274**, 536 (1996).
- 7 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics* **10**, 57-63 (2009).
- 8 Nolan, T., Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. *Nature protocols* **1**, 1559-1582 (2006).
- 9 Tyers, M. & Mann, M. From genomics to proteomics. *Nature* **422**, 193-197 (2003).
- 10 Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198-207 (2003).
- 11 Zamboni, N. & Sauer, U. Novel biological insights through metabolomics and ¹³C-flux analysis. *Current opinion in microbiology* **12**, 553-558, doi:10.1016/j.mib.2009.08.003 (2009).
- 12 Goodacre, R., Vaidyanathan, S., Dunn, W. B., Harrigan, G. G. & Kell, D. B. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends in biotechnology* **22**, 245-252 (2004).
- 13 Nielsen, J. & Oliver, S. The next wave in metabolome analysis. *Trends in biotechnology* **23**, 544-546 (2005).
- 14 Reaves, M. L. & Rabinowitz, J. D. Metabolomics in systems microbiology. *Current opinion in biotechnology* **22**, 17-25 (2011).
- 15 Heinemann, M. & Sauer, U. Systems biology of microbial metabolism. *Current opinion in microbiology* **13**, 337-343, doi:10.1016/j.mib.2010.02.005 (2010).
- 16 Berg, J. M., Tymoczko, J. & Stryer, L. Biochemistry. *New York* (2006).
- 17 Pace, N. R. The universal nature of biochemistry. *Proceedings of the National Academy of Sciences* **98**, 805-808 (2001).
- 18 Kochanowski, K., Sauer, U. & Chubukov, V. Somewhat in control – the role of transcription in regulating microbial metabolic fluxes. *Current opinion in biotechnology* **24**, 1-7 (2013).
- 19 Wegner, A., Meiser, J. & Weindl, D. How metabolites modulate metabolic flux. *Current Opinion in Biotechnology* **34**, 16-22, doi:10.1016/j.copbio.2014.11.008 (2015).
- 20 Gerosa, L. & Sauer, U. Regulation and control of metabolic fluxes in microbes. *Current opinion in biotechnology* **22**, 566-575, doi:10.1016/j.copbio.2011.04.016 (2011).
- 21 Westerhoff, H. V. & Palsson, B. O. The evolution of molecular biology into systems biology. *Nature biotechnology* **22**, 1249-1252 (2004).
- 22 Eddy, J. A., Funk, C. C. & Price, N. D. Fostering synergy between cell biology and systems biology. *Trends in Cell Biology* **25**, 440-445, doi:10.1016/j.tcb.2015.04.005 (2015).
- 23 Sauer, U., Heinemann, M. & Zamboni, N. Getting closer to the whole picture. *Science(Washington)* **316**, 550-551 (2007).

- 24 Atwater, W. & Rosa, E. A new respiration calorimeter and experiments on the conservation of energy in the human body. I. *Physical Review (Series I)* **9**, 129 (1899).
- 25 Harris, J. A. & Benedict, F. G. A biometric study of human basal metabolism. *Proceedings of the National Academy of Sciences* **4**, 370-373 (1918).
- 26 Haldane, J., Wigglesworth, V. & Woodrow, C. The effect of reaction changes on human inorganic metabolism. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **96**, 1-14 (1924).
- 27 Haldane, J., Wigglesworth, V. & Woodrow, C. The effect of reaction changes on human carbohydrate and oxygen metabolism. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* **96**, 15-28 (1924).
- 28 Sumner, J. B. The isolation and crystallization of the enzyme urease. *Journal of Biological Chemistry* **69**, 435-441 (1926).
- 29 Blangy, D., Buc, H. & Monod, J. Kinetics of the allosteric interactions of phosphofructokinase from *Escherichia coli*. *Journal of molecular biology* **31**, 13-35, doi:10.1016/0022-2836(68)90051-X (1968).
- 30 Cleland, W. Enzyme kinetics. *Annual review of biochemistry* **36**, 77-112 (1967).
- 31 Newsholme, E. & Crabtree, B. Flux-generating and regulatory steps in metabolic control. *Trends in Biochemical Sciences* **6**, 53-56 (1981).
- 32 Schoenheimer, R., Rittenberg, D., Fox, M., Keston, A. S. & Ratner, S. The nitrogen isotope (N^{15}) as a tool in the study of the intermediary metabolism of nitrogenous compounds. *Journal of the American Chemical Society* **59**, 1768-1768 (1937).
- 33 Hostetler, K. Y. & Landau, B. R. Estimation of the pentose cycle contribution to glucose metabolism in tissue *in vivo*. *Biochemistry* **6**, 2961-2964 (1967).
- 34 Landau, B. R., Bartsch, G. E., Katz, J. & Wood, H. G. Estimation of pathway contributions to glucose metabolism and of the rate of isomerization of hexose 6-phosphate. *The Journal of biological chemistry* **239**, 686-696 (1964).
- 35 Wiechert, W. & de Graaf, A. A. *In vivo* stationary flux analysis by ^{13}C labeling experiments. *Advances in biochemical engineering/biotechnology* **54**, 109-154 (1996).
- 36 Zamboni, N. Toward metabolome-bases ^{13}C flux analysis: a universal tool for measuring *in vivo* metabolic activity. *Topics in current genetics* **18**, doi:10.1007/4735 (2007).
- 37 Buescher, J. M., Czernik, D., Ewald, J. C., Sauer, U. & Zamboni, N. Cross-platform comparison of methods for quantitative metabolomics of primary metabolism. *Analytical chemistry* **81**, 2135-2143, doi:10.1021/ac8022857 (2009).
- 38 Fuhrer, T., Heer, D., Begemann, B. & Zamboni, N. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Analytical chemistry* **83**, 7074-7080, doi:10.1021/ac201267k (2011).
- 39 Yuan, J., Bennett, B. D. & Rabinowitz, J. D. Kinetic flux profiling for quantitation of cellular metabolic fluxes. *Nature protocols* **3**, 1328-1340, doi:10.1038/nprot.2008.131 (2008).
- 40 Wittmann, C. & Heinzle, E. Mass spectrometry for metabolic flux analysis. *Biotechnology and bioengineering* **62**, 739-750 (1999).
- 41 O'Brien, E. J., Lerman, J. a., Chang, R. L., Hyduke, D. R. & Palsson, B. Ø. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology* **9**, doi:10.1038/msb.2013.52 (2013).
- 42 Schellenberger, J., Park, J. O., Conrad, T. M. & Palsson, B. Ø. BiGG : a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions Database. (2010).
- 43 Becker, S. A. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nature protocols* **2**, 727-738 (2007).

- 44 Palsson, B. *In silico* biology through “omics”. *Nature biotechnology* **20**, 649-650 (2002).
- 45 Varma, A. & Palsson, B. O. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature biotechnology* **12**, 994-998 (1994).
- 46 Lindow, S., Panopoulos, N. & McFarland, B. Genetic engineering of bacteria from managed and natural habitats. *Science(Washington)* **244**, 1300-1307 (1989).
- 47 Bailey, J., Birnbaum, S., Galazzo, J., Khosla, C. & Shanks, J. Strategies and challenges in metabolic engineering. *Annals of the New York Academy of Sciences* **589**, 1-15 (1990).
- 48 Stephanopoulos, G. Metabolic fluxes and metabolic engineering. *Metabolic engineering* **1**, 1-11 (1999).
- 49 King, Z. A., Lloyd, C. J., Feist, A. M. & Palsson, B. O. Next-generation genome-scale models for metabolic engineering. *Current Opinion in Biotechnology* **35**, 23-29, doi:10.1016/j.copbio.2014.12.016 (2015).
- 50 Long, M. R., Ong, W. K. & Reed, J. L. Computational methods in metabolic engineering for strain design. *Current Opinion in Biotechnology* **34**, 135-141, doi:10.1016/j.copbio.2014.12.019 (2015).
- 51 Gowen, C. M. & Fong, S. S. Systems metabolic engineering. **985**, doi:10.1007/978-1-62703-299-5 (2013).
- 52 Toya, Y. & Shimizu, H. Flux analysis and metabolomics for systematic metabolic engineering of microorganisms. *Biotechnology advances* **31**, 818-826 (2013).
- 53 Wiechert, W. Modeling and simulation: tools for metabolic engineering. *Journal of biotechnology* **94**, 37-63 (2002).
- 54 Cordes, T. & Metallo, C. M. Tracing insights into human metabolism using chemical engineering approaches. *Current Opinion in Chemical Engineering* **14**, 72-81 (2016).
- 55 Hotamisligil, G. S. Inflammation and metabolic disorders. *Nature* **444**, 860-867 (2006).
- 56 Lusis, A. J., Attie, A. D. & Reue, K. Metabolic syndrome: from epidemiology to systems biology. *Nature Reviews Genetics* **9**, 819-830 (2008).
- 57 Warburg, O. The metabolism of carcinoma cells. *The Journal of Cancer Research* **9**, 148-163 (1925).
- 58 Frezza, C., Pollard, P. J. & Gottlieb, E. Inborn and acquired metabolic defects in cancer. *Journal of molecular medicine* **89**, 213-220 (2011).
- 59 Shlomi, T., Benyamini, T., Gottlieb, E., Sharan, R. & Ruppin, E. Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. *PLoS computational biology* **7**, e1002018-e1002018, doi:10.1371/journal.pcbi.1002018 (2011).
- 60 Yizhak, K. *et al.* A computational study of the Warburg effect identifies metabolic targets inhibiting cancer migration. 1-12 (2014).
- 61 Yizhak, K. *et al.* Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. 1-23, doi:10.7554/eLife.03641 (2014).
- 62 Folger, O. *et al.* Predicting selective drug targets in cancer through metabolic networks. *Molecular systems biology* **7**, 501 (2011).
- 63 Gaglio, D. *et al.* Oncogenic K-Ras decouples glucose and glutamine metabolism to support cancer cell growth. *Molecular Systems Biology* **7**, 1-15, doi:10.1038/msb.2011.56 (2011).
- 64 Grassian, A. R. *et al.* IDH1 mutations alter citric acid cycle metabolism and increase dependence on oxidative mitochondrial metabolism. *Cancer Research* **74**, 3317-3331, doi:10.1158/0008-5472.CAN-14-0772-T (2014).
- 65 Wise, D. R. *et al.* Hypoxia promotes isocitrate dehydrogenase-dependent carboxylation of α -ketoglutarate to citrate to support cell growth and viability.

- Proceedings of the National Academy of Sciences of the United States of America* **108**, 19611-19616, doi:10.1073/pnas.1117773108 (2011).
- 66 Jerby-Arnon, L. *et al.* Predicting cancer-specific vulnerability via data-driven detection of synthetic lethality. *Cell* **158**, 1199-1209 (2014).
- 67 Yang, C. *et al.* Glutamine oxidation maintains the TCA cycle and cell survival during impaired mitochondrial pyruvate transport. *Molecular Cell* **56**, 414-424, doi:10.1016/j.molcel.2014.09.025 (2014).
- 68 Holohan, C., Van Schaeybroeck, S., Longley, D. B. & Johnston, P. G. Cancer drug resistance: an evolving paradigm. *Nature Reviews Cancer* **13**, 714-726 (2013).
- 69 Gottesman, M. M. Mechanisms of cancer drug resistance. *Annual review of medicine* **53**, 615-627 (2002).
- 70 Joyce, A. R. & Palsson, B. Ø. Predicting gene essentiality using genome-scale *in silico* models. *Microbial Gene Essentiality: Protocols and Bioinformatics*, 433-457 (2008).
- 71 Kim, H. U., Sohn, S. B. & Lee, S. Y. Metabolic network modeling and simulation for drug targeting and discovery. *Biotechnology journal* **7**, 330-342, doi:10.1002/biot.201100159 (2012).
- 72 Lee, D.-S. *et al.* Comparative genome-scale metabolic reconstruction and flux balance analysis of multiple *Staphylococcus aureus* genomes identify novel antimicrobial drug targets. *Journal of bacteriology* **191**, 4015-4024 (2009).
- 73 Jamshidi, N. & Palsson, B. Ø. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ 661 and proposing alternative drug targets. *BMC systems biology* **1**, 1 (2007).
- 74 Raman, K., Rajagopalan, P. & Chandra, N. Flux balance analysis of mycolic acid pathway: targets for anti-tubercular drugs. *PLoS Comput Biol* **1**, e46 (2005).
- 75 Norrby, S. R., Nord, C. E. & Finch, R. Lack of development of new antimicrobial drugs: a potential serious threat to public health. *The Lancet infectious diseases* **5**, 115-119 (2005).
- 76 Blair, J. M., Webber, M. A., Baylay, A. J., Ogbolu, D. O. & Piddock, L. J. Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology* **13**, 42-51 (2015).
- 77 Steeb, B. *et al.* Parallel exploitation of diverse host nutrients enhances *Salmonella* virulence. *PLoS Pathog* **9**, e1003301, doi:10.1371/journal.ppat.1003301 (2013).
- 78 Ehrt, S., Rhee, K. & Schnappinger, D. Mycobacterial genes essential for the pathogen's survival in the host. *Immunological reviews* **264**, 319-326 (2015).
- 79 de Carvalho, L. P. S. *et al.* Metabolomics of *Mycobacterium tuberculosis* reveals compartmentalized co-catabolism of carbon substrates. *Chemistry & biology* **17**, 1122-1131 (2010).
- 80 Chopra, T. *et al.* Quantitative mass spectrometry reveals plasticity of metabolic networks in *Mycobacterium smegmatis*. *Molecular & cellular proteomics : MCP*, M113.034082--M034113.034082-, doi:10.1074/mcp.M113.034082 (2014).
- 81 Beste, D. J. V. *et al.* ¹³C-Flux spectral analysis of host-pathogen metabolism reveals a mixed diet for intracellular *Mycobacterium tuberculosis*. *Chemistry & biology* **20**, 1012-1021, doi:10.1016/j.chembiol.2013.06.012 (2013).
- 82 Lee, W., Vanderven, B. C., Fahey, R. J. & Russell, D. G. Intracellular *Mycobacterium tuberculosis* exploits host-derived fatty acids to limit metabolic stress. *The Journal of biological chemistry* **288**, 6788-6800, doi:10.1074/jbc.M112.445056 (2013).
- 83 Niederweis, M. Nutrient acquisition by mycobacteria. *Microbiology (Reading, England)* **154**, 679-692, doi:10.1099/mic.0.2007/012872-0 (2008).
- 84 Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. Ø. & Jamshidi, N. Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Molecular systems biology* **6**, 422-422, doi:10.1038/msb.2010.68 (2010).

- 85 Levy, R. & Borenstein, E. Metabolic modeling of species interaction in the human microbiome elucidates community-level assembly rules. *Proceedings of the National Academy of Sciences* **110**, 12804-12809 (2013).
- 86 Thiele, I., Heinken, A. & Fleming, R. M. A systems biology approach to studying the role of microbes in human health. *Current opinion in biotechnology* **24**, 4-12 (2013).
- 87 Appelberg, R. Macrophage nutritive antimicrobial mechanisms. *Journal of leukocyte biology* **79**, 1117-1128 (2006).
- 88 Greenblum, S., Chiu, H.-C., Levy, R., Carr, R. & Borenstein, E. Towards a predictive systems-level model of the human microbiome: progress, challenges, and opportunities. *Current opinion in biotechnology* **24**, 810-820 (2013).
- 89 Wallace, B. D. & Redinbo, M. R. The human microbiome is a source of therapeutic drug targets. *Current opinion in chemical biology* **17**, 379-384 (2013).
- 90 Wilson, I. D. & Nicholson, J. K. Metabonomics and Gut Microbiota in Nutrition and Disease, 323-341, Springer (2015).
- 91 Calvin, M. *et al.* Isotopic Carbon. *The Journal of Physical Chemistry* **53**, 1139-1140 (1949).
- 92 Katz, J. & Wood, H. G. The use of $C^{14}O_2$ yields from glucose-1-and-6- C^{14} for the evaluation of the pathways of glucose metabolism. *Journal of Biological Chemistry* **238**, 517-523 (1963).
- 93 Bloom, B., Stetten, M. R. & Stetten, D. Evaluation of catabolic pathways of glucose in mammalian systems. *Journal of Biological Chemistry* **204**, 681-694 (1953).
- 94 Wood, H. G. & Katz, J. The distribution of C^{14} in the hexose phosphates and the effect of recycling in the pentose cycle. *Journal of Biological Chemistry* **233**, 1279-1282 (1958).
- 95 Ljungdahl, L., Wood, H. G., Racker, E. & Couri, D. Formation of unequally labeled fructose 6-phosphate by an exchange reaction catalyzed by transaldolase. *Journal of Biological Chemistry* **236**, 1622-1625 (1961).
- 96 Katz, J., Landau, B. R. & Bartsch, G. E. The pentose cycle, triose phosphate isomerization, and lipogenesis in rat adipose tissue. *Journal of Biological Chemistry* **241**, 727-740 (1966).
- 97 Landau, B. R. & Katz, J. A quantitative estimation of the pathways of glucose metabolism in rat adipose tissue *in vitro*. *Journal of Biological Chemistry* **239**, 697-704 (1964).
- 98 Segal, S., Berman, M. & Blair, A. The metabolism of variously C^{14} -labeled glucose in man and an estimation of the extent of glucose metabolism by the hexose monophosphate pathway. *Journal of Clinical Investigation* **40**, 1263 (1961).
- 99 Matwiyoff, N. & Ott, D. Stable isotope tracers in the life sciences and medicine. Los Alamos Scientific Lab., NM, (1973).
- 100 Lifson, N. & Stolen, J. A. Metabolism of C^{13} -Carboxyl-labeled malonate by the intact mouse. *Experimental Biology and Medicine* **74**, 451-453 (1950).
- 101 Rabideau, G. S. & Burr, G. O. The use of the C^{13} isotope as a tracer for transport studies in plants. *American Journal of Botany*, 349-356 (1945).
- 102 Shulman, R. *et al.* Cellular applications of ^{31}P and ^{13}C nuclear magnetic resonance. *Science* **205**, 160-166 (1979).
- 103 Blum, J. J. & Stein, R. B. On the analysis of metabolic networks. *Biological Regulation and Development*, 99-124 (1982).
- 104 Borowitz, M. J., Stein, R. B. & Blum, J. J. Quantitative analysis of the change of metabolite fluxes along the pentose phosphate and glycolytic pathways in *Tetrahymena* in response to carbohydrates. *Journal of Biological Chemistry* **252**, 1589-1605 (1977).

- 105 Kell, D. B. & Westerhoff, H. V. Towards a rational approach to the optimization of flux in microbial biotransformations. *Trends in Biotechnology* **4**, 137-142 (1986).
- 106 Kacser, H. a. & Burns, J. in *Symp. Soc. Exp. Biol.* 65-104.
- 107 Stephanopoulos, G. & Vallino, J. J. Network rigidity and metabolic engineering in metabolite overproduction. *Science* **252**, 1675-1681 (1991).
- 108 Aiba, S. & Matsuoka, M. Identification of metabolic model: citrate production from glucose by *Candida lipolytica*. *Biotechnology and Bioengineering* **21**, 1373-1386 (1979).
- 109 Reardon, K. F., Scheper, T. H. & Bailey, J. E. Metabolic pathway rates and culture fluorescence in batch fermentations of *Clostridium acetobutylicum*. *Biotechnology progress* **3**, 153-167 (1987).
- 110 Vallino, J. J. & Stephanopoulos, G. N. Intelligent Sensors in Biotechnology. *Annals of the New York Academy of Sciences* **506**, 415-430 (1987).
- 111 Vallino, J. J. & Stephanopoulos, G. Flux determination in cellular bioreaction networks: applications to lysine fermentations. *Frontiers in bioprocessing* **1**, 205-219 (1989).
- 112 Watson, M. Metabolic maps for the Apple II. *Biochemical Society Transactions* **12**, 1093-1094 (1984).
- 113 Watson, M. A discrete model of bacterial metabolism. *Computer applications in the biosciences: CABIOS* **2**, 23-27 (1986).
- 114 Fell, D. A. & Small, J. R. Fat synthesis in adipose tissue. An examination of stoichiometric constraints. *Biochemical Journal* **238**, 781-786 (1986).
- 115 Bonarius, H. P. J., Schmid, G. & Tramper, J. Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends in Biotechnology* **15**, 308-314, doi:10.1016/S0167-7799(97)01067-6 (1997).
- 116 Bonarius, H. P. *et al.* Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechnology and bioengineering* **50**, 299-318 (1996).
- 117 Jørgensen, H., Nielsen, J., Villadsen, J. & Møllgaard, H. Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations. *Biotechnology and Bioengineering* **46**, 117-131 (1995).
- 118 Savinell, J. M. & Palsson, B. O. Optimal selection of metabolic fluxes for in vivo measurement. I. Development of mathematical methods. *Journal of theoretical biology* **155**, 201-214 (1992).
- 119 Savinell, J. M. & Palsson, B. O. Optimal selection of metabolic fluxes for in vivo measurement. II. Application to *Escherichia coli* and hybridoma cell metabolism. *Journal of theoretical biology* **155**, 215-242 (1992).
- 120 Sauer, U. *et al.* Physiology and metabolic fluxes of wild-type and riboflavin-producing *Bacillus subtilis*. Physiology and Metabolic Fluxes of Wild-Type and Riboflavin-Producing *Bacillus subtilis*. **62** (1996).
- 121 Zupke, C. & Stephanopoulos, G. Intracellular flux analysis in hybridomas using mass balances and *in vitro* ¹³C NMR. *Biotechnology and bioengineering* **45**, 292-303 (1995).
- 122 Segre, D., Vitkup, D. & Church, G. M. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences* **99**, 15112-15117 (2002).
- 123 Ibarra, R. U., Edwards, J. S. & Palsson, B. O. *Escherichia coli* K-12 undergoes adaptive evolution to achieve *in silico* predicted optimal growth. *Nature* **420**, 186-189 (2002).
- 124 Fong, S. S. & Palsson, B. Ø. Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nature genetics* **36**, 1056-1058 (2004).

- 125 Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering* **5**, 264-276, doi:10.1016/j.ymben.2003.09.002 (2003).
- 126 Bonarius, H. P., Schmid, G. & Tramper, J. Flux analysis of underdetermined metabolic networks: the quest for the missing constraints. *Trends in Biotechnology* **15**, 308-314 (1997).
- 127 Edwards, J. S. & Palsson, B. O. How will bioinformatics influence metabolic engineering? *Biotechnology and Bioengineering* **58**, 162-169 (1998).
- 128 Sauer, U., Cameron, D. C. & Bailey, J. E. Metabolic capacity of *Bacillus subtilis* for the production of purine nucleosides, riboflavin, and folic acid. *Biotechnology and bioengineering* **59**, 227-238 (1998).
- 129 Van Gulik, W. & Heijnen, J. A metabolic network stoichiometry analysis of microbial growth and product formation. *Biotechnology and bioengineering* **48**, 681-698 (1995).
- 130 Savinell, J. M. & Palsson, B. O. Network analysis of intermediary metabolism using linear optimization. I. Development of mathematical formalism. *Journal of theoretical biology* **154**, 421-454 (1992).
- 131 Savinell, J. M. & Palsson, B. O. Network analysis of intermediary metabolism using linear optimization: II. Interpretation of hybridoma cell metabolism. *Journal of theoretical biology* **154**, 455-473 (1992).
- 132 Klamt, S., Schuster, S. & Gilles, E. D. Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria. *Biotechnology and Bioengineering* **77**, 734-751, doi:10.1002/bit.10153 (2002).
- 133 Majewski, R. & Domach, M. Simple constrained-optimization view of acetate overflow in *E. coli*. *Biotechnology and bioengineering* **35**, 732-738 (1990).
- 134 Martínez, V. S. *et al.* Flux balance analysis of CHO cells before and after a metabolic switch from lactate production to consumption. *Biotechnology and bioengineering* **110**, 660-666 (2013).
- 135 Ramakrishna, R., Edwards, J. S., McCulloch, A. & Palsson, B. O. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **280**, R695-R704 (2001).
- 136 Bonarius, H. P., Timmerarends, B., de Gooijer, C. D. & Tramper, J. Metabolite-balancing techniques vs. ¹³C tracer experiments to determine metabolic fluxes in hybridoma cells. *Biotechnology and bioengineering* **58**, 258-262 (1998).
- 137 Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. & Sauer, U. Multidimensional optimality of microbial metabolism. *Science* **336**, 601-604, doi:10.1126/science.1216882 (2012).
- 138 Soh, K. C. & Hatzimanikatis, V. Constraining the flux space using thermodynamics and integration of metabolomics data. *Metabolic Flux Analysis: Methods and Protocols*, 49-63 (2014).
- 139 Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. Thermodynamics-based metabolic flux analysis. *Biophysical journal* **92**, 1792-1805 (2007).
- 140 Henry, C. S., Jankowski, M. D., Broadbelt, L. J. & Hatzimanikatis, V. Genome-scale thermodynamic analysis of *Escherichia coli* metabolism. *Biophysical journal* **90**, 1453-1461, doi:10.1529/biophysj.105.071720 (2006).
- 141 Noor, E., Haraldsdóttir, H. S., Milo, R. & Fleming, R. M. T. Consistent estimation of Gibbs energy using component contributions. *PLoS Computational Biology* **9**, e1003098-e1003098, doi:10.1371/journal.pcbi.1003098 (2013).

- 142 Fleming, R. M., Thiele, I. & Nasheuer, H. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to *Escherichia coli*. *Biophysical chemistry* **145**, 47-56 (2009).
- 143 Kümmel, A., Panke, S. & Heinemann, M. Putative regulatory sites unraveled by network-embedded thermodynamic analysis of metabolome data. *Molecular systems biology* **2** (2006).
- 144 Hoppe, A., Hoffmann, S. & Holzhütter, H.-G. Including metabolite concentrations into flux balance analysis: thermodynamic realizability as a constraint on flux distributions in metabolic networks. *BMC systems biology* **1**, 1 (2007).
- 145 Noor, E. *et al.* Pathway thermodynamics highlights kinetic obstacles in central metabolism. *PLoS Comput Biol* **10**, e1003483 (2014).
- 146 Chakrabarti, A., Miskovic, L., Soh, K. C. & Hatzimanikatis, V. Towards kinetic modeling of genome-scale metabolic networks without sacrificing stoichiometric, thermodynamic and physiological constraints. *Biotechnology journal* **8**, 1043-1057 (2013).
- 147 Ataman, M. & Hatzimanikatis, V. Heading in the right direction: thermodynamics-based network analysis and pathway engineering. *Current Opinion in Biotechnology* **36**, 176-182 (2015).
- 148 Blazier, A. S. & Papin, J. Integration of expression data in genome-scale metabolic network reconstructions. *Frontiers in physiology* **3**, 299-299, doi:10.3389/fphys.2012.00299 (2012).
- 149 Schilling, C. H. *et al.* Genome-scale metabolic model of *Helicobacter pylori* 26695. *Journal of Bacteriology* **184**, 4582-4593, doi:10.1128/JB.184.16.4582 (2002).
- 150 Edwards, J. & Palsson, B. The *Escherichia coli* MG1655 *in silico* metabolic genotype: its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences* **97**, 5528-5533 (2000).
- 151 Oh, Y.-K., Palsson, B. O., Park, S. M., Schilling, C. H. & Mahadevan, R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *The Journal of biological chemistry* **282**, 28791-28799, doi:10.1074/jbc.M703759200 (2007).
- 152 Famili, I., Forster, J., Nielsen, J. & Palsson, B. O. *Saccharomyces cerevisiae* phenotypes can be predicted by using constraint-based analysis of a genome-scale reconstructed metabolic network. *Proceedings of the National Academy of Sciences of the United States of America* **100**, 13134-13139, doi:10.1073/pnas.2235812100 (2003).
- 153 Papin, J. A., Price, N. D., Wiback, S. J., Fell, D. A. & Palsson, B. O. Metabolic pathways in the post-genome era. *Trends in biochemical sciences* **28**, 250-258 (2003).
- 154 Thiele, I. & Palsson, B. Ø. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature protocols* **5**, 93-121, doi:10.1038/nprot.2009.203 (2010).
- 155 Heavner, B. D. & Price, N. D. Transparency in metabolic network reconstruction enables scalable biological discovery. *Current opinion in biotechnology* **34**, 105-109 (2015).
- 156 Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature protocols* **6**, 1290-1307, doi:10.1038/nprot.2011.308 (2011).
- 157 Henry, C. S. *et al.* High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nature biotechnology* **28**, 977-982, doi:10.1038/nbt.1672 (2010).
- 158 Hamilton, J. J. & Reed, J. L. Software platforms to facilitate reconstructing genome-scale metabolic networks. *Environmental microbiology* **16**, 49-59 (2014).

- 159 Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. O. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome biology* **4**, R54-R54, doi:10.1186/gb-2003-4-9-r54 (2003).
- 160 King, Z. A. *et al.* BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic acids research* **44**, D515-D522 (2016).
- 161 Terzer, M., Maynard, N. D. & Covert, M. W. Genome-scale metabolic networks. doi:10.1002/wsbm.037 (2009).
- 162 Reed, J. L. Shrinking the metabolic solution space using experimental datasets. *PLoS computational biology* **8**, e1002662-e1002662, doi:10.1371/journal.pcbi.1002662 (2012).
- 163 Becker, S. a. & Palsson, B. O. Context-specific metabolic networks are consistent with experiments. *PLoS computational biology* **4**, e1000082-e1000082, doi:10.1371/journal.pcbi.1000082 (2008).
- 164 Zur, H., Ruppin, E. & Shlomi, T. iMAT: an integrative metabolic analysis tool. *Bioinformatics (Oxford, England)* **26**, 3140-3142, doi:10.1093/bioinformatics/btq602 (2010).
- 165 Colijn, C. *et al.* Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS computational biology* **5**, e1000489-e1000489, doi:10.1371/journal.pcbi.1000489 (2009).
- 166 Jensen, P. A. & Papin, J. A. Functional integration of a metabolic network model and expression data without arbitrary thresholding. *Bioinformatics* **27**, 541-547 (2011).
- 167 Shlomi, T., Cabili, M. N., Herrgård, M. J., Palsson, B. Ø. & Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nature biotechnology* **26**, 1003-1010, doi:10.1038/nbt.1487 (2008).
- 168 Yizhak, K., Chaneton, B., Gottlieb, E. & Ruppin, E. Modeling cancer metabolism on a genome scale. *Molecular systems biology* **11**:817 (2015).
- 169 Covert, M. W., Knight, E. M., Reed, J. L., Herrgård, M. J. & Palsson, B. Ø. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* **429**, 2-6, doi:10.1038/nature02514.Published (2004).
- 170 Shlomi, T., Eisenberg, Y., Sharan, R. & Ruppin, E. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Molecular systems biology* **3**, 101-101, doi:10.1038/msb4100141 (2007).
- 171 Yizhak, K., Benyamini, T., Liebermeister, W., Ruppin, E. & Shlomi, T. Integrating quantitative proteomics and metabolomics with a genome-scale metabolic network model. *Bioinformatics (Oxford, England)* **26**, i255-260, doi:10.1093/bioinformatics/btq183 (2010).
- 172 Brien, E. J. O. & Palsson, B. O. Computing the functional proteome : recent progress and future prospects for genome-scale models. *Current Opinion in Biotechnology* **34**, 125-134, doi:10.1016/j.copbio.2014.12.017 (2015).
- 173 Lerman, J. a. *et al.* *In silico* method for modelling metabolism and gene product expression at genome scale. *Nature communications* **3**, 929-929, doi:10.1038/ncomms1928 (2012).
- 174 Wiechert, W. ¹³C metabolic flux analysis. *Metabolic engineering* **3**, 195-206, doi:10.1006/mben.2001.0187 (2001).
- 175 Szyperski, T. Biosynthetically directed fractional ¹³C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism. *European journal of biochemistry / FEBS* **232**, 433-448 (1995).
- 176 Dauner, M. & Sauer, U. GC-MS analysis of amino acids rapidly provides rich information for isotopomer balancing. *Biotechnology progress* **16**, 642-649, doi:10.1021/bp000058h (2000).

- 177 Christensen, B. & Nielsen, J. Isotopomer analysis using GC-MS. *Metabolic engineering* **1**, 282-290 (1999).
- 178 Wiechert, W. & Graaf, A. A. D. Bidirectional reaction steps in metabolic networks part I: Modelling and simulation of carbon isotope labelling experiments. *Biotechnology and bioengineering* (1996).
- 179 Wiechert, W., Siefke, C., Graaf, A. A. D. & Marx, A. Bidirectional reaction steps in metabolic networks part II: Flux estimation and statistical analysis. *Biotechnology and bioengineering* **55**, 118-135 (1996).
- 180 Wiechert, W., Möllney, M., Isermann, N., Wurzel, M. & de Graaf, a. a. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnology and bioengineering* **66**, 69-85 (1999).
- 181 Zhao, J. & Shimizu, K. Metabolic flux analysis of *Escherichia coli* K12 grown on ¹³C-labeled acetate and glucose using GC-MS and powerful flux calculation method. *Journal of biotechnology* **101**, 101-117 (2003).
- 182 Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metabolic engineering* **9**, 68-86, doi:10.1016/j.ymben.2006.09.001 (2007).
- 183 Weitzel, M. *et al.* 13CFLUX2--high-performance software suite for ¹³C-metabolic flux analysis. *Bioinformatics (Oxford, England)* **29**, 143-145, doi:10.1093/bioinformatics/bts646 (2013).
- 184 Young, J. D. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* **30**, 1333-1335, doi:10.1093/bioinformatics/btu015 (2014).
- 185 Quek, L.-E., Wittmann, C., Nielsen, L. K. & Krömer, J. O. OpenFLUX: efficient modelling software for ¹³C-based metabolic flux analysis. *Microbial cell factories* **8**, 25-25, doi:10.1186/1475-2859-8-25 (2009).
- 186 Yoo, H., Antoniewicz, M. R., Stephanopoulos, G. & Kelleher, J. K. Quantifying reductive carboxylation flux of glutamine to lipid in a brown adipocyte cell line. *Journal of Biological Chemistry* **283**, 20621-20627 (2008).
- 187 Sauer, U., Hatzimanikatis, V. & Bailey, J. E. Metabolic fluxes in riboflavin-producing *Bacillus subtilis*. *Nature biotechnology* **15**, 448-452 (1997).
- 188 Schmidt, K., Nielsen, J. & Villadsen, J. Quantitative analysis of metabolic fluxes in *Escherichia coli*, using two-dimensional NMR spectroscopy and complete isotopomer models. *Journal of biotechnology* **71**, 175-189 (1999).
- 189 Gombert, A. K., Moreira, M., Christensen, B., Nielsen, J. & Acteriol, J. B. Network identification and flux quantification in the central metabolism of *Saccharomyces cerevisiae* under different conditions of glucose repression. *Journal of bacteriology* **183**, 1441-1451, doi:10.1128/JB.183.4.1441 (2001).
- 190 Möllney, M., Wiechert, W., Kownatzki, D. & Graaf, A. A. D. Bidirectional reaction steps in metabolic networks part IV : Optimal design of isotopomer labeling experiments. *Biotechnology and bioengineering* **66**, 86-103(1999).
- 191 Millard, P., Sokol, S., Letisse, F. & Portais, J. C. IsoDesign: A software for optimizing the design of ¹³C-metabolic flux analysis experiments. *Biotechnology and Bioengineering* **111**, 202-208, doi:10.1002/bit.24997 (2014).
- 192 Crown, S. B., Ahn, W. S. & Antoniewicz, M. R. Rational design of ¹³C-labeling experiments for metabolic flux analysis in mammalian cells. *BMC systems biology* **6**, 43-43, doi:10.1186/1752-0509-6-43 (2012).
- 193 Walther, J. L., Metallo, C. M., Zhang, J. & Stephanopoulos, G. Optimization of ¹³C isotopic tracers for metabolic flux analysis in mammalian cells. *Metabolic engineering* **14**, 162-171, doi:10.1016/j.ymben.2011.12.004 (2011).

- 194 Metallo, C. M., Walther, J. L. & Stephanopoulos, G. Evaluation of ^{13}C isotopic tracers for metabolic flux analysis in mammalian cells. *Journal of biotechnology* **144**, 167-174, doi:10.1016/j.jbiotec.2009.07.010 (2009).
- 195 Araúzo-Bravo, M. An improved method for statistical analysis of metabolic flux analysis using isotopomer mapping matrices with analytical expressions. *Journal of Biotechnology* **105**, 117-133, doi:10.1016/S0168-1656(03)00169-X (2003).
- 196 Beste, D. J. V. *et al.* ^{13}C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in mycobacteria which requires isocitrate lyase and carbon dioxide fixation. *PLoS pathogens* **7**, e1002091-e1002091, doi:10.1371/journal.ppat.1002091 (2011).
- 197 Woo Suk, A. & Antoniewicz, M. R. Parallel labeling experiments with [1,2- ^{13}C]glucose and [U- ^{13}C]glutamine provide new insights into CHO cell metabolism. *Metabolic Engineering* **15**, 34-47, doi:10.1016/j.ymben.2012.10.001 (2013).
- 198 Fischer, E. & Sauer, U. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *European Journal of Biochemistry* **270**, 880-891, doi:10.1046/j.1432-1033.2003.03448.x (2003).
- 199 Christensen, B., Christiansen, T., Gombert, a. K., Thykaer, J. & Nielsen, J. Simple and robust method for estimation of the split between the oxidative pentose phosphate pathway and the Embden-Meyerhof-Parnas pathway in microorganisms. *Biotechnology and bioengineering* **74**, 517-523 (2001).
- 200 Fischer, E. & Sauer, U. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nature genetics* **37**, 636-640, doi:10.1038/ng1555 (2005).
- 201 Haverkorn van Rijsewijk, B. R. B., Nanchen, A., Nallet, S., Kleijn, R. J. & Sauer, U. Large-scale ^{13}C -flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Molecular systems biology* **7**, 477-477, doi:10.1038/msb.2011.9 (2011).
- 202 Blank, L. M., Kuepfer, L. & Sauer, U. Large-scale ^{13}C -flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome biology* **6**, R49-R49, doi:10.1186/gb-2005-6-6-r49 (2005).
- 203 Fischer, E., Zamboni, N. & Sauer, U. High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived ^{13}C constraints. *Analytical biochemistry* **325**, 308-316 (2004).
- 204 Zamboni, N., Fendt, S.-M., Rühl, M. & Sauer, U. ^{13}C -based metabolic flux analysis. *Nature protocols* **4**, 878-892, doi:10.1038/nprot.2009.58 (2009).
- 205 García Martín, H. *et al.* A method to constrain genome-scale models with ^{13}C labeling data. *PLOS Computational Biology* **11**, e1004363-e1004363, doi:10.1371/journal.pcbi.1004363 (2015).
- 206 Rantanen, A. *et al.* An analytic and systematic framework for estimating metabolic flux ratios from ^{13}C tracer experiments. *BMC bioinformatics* **9**, 266-266, doi:10.1186/1471-2105-9-266 (2008).
- 207 Antoniewicz, M. R. Dynamic metabolic flux analysis—tools for probing transient states of metabolic networks. *Current opinion in biotechnology* **24**, 973-978 (2013).
- 208 Nöh, K., Wahl, A. & Wiechert, W. Computational tools for isotopically instationary ^{13}C labeling experiments under metabolic steady state conditions. *Metabolic Engineering* **8**, 554-577, doi:10.1016/j.ymben.2006.05.006 (2006).
- 209 Wiechert, W. & Nöh, K. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Current Opinion in Biotechnology* **24**, 979-986, doi:10.1016/j.copbio.2013.03.024 (2013).

- 210 Noack, S., Nöh, K., Moch, M., Oldiges, M. & Wiechert, W. Stationary versus non-stationary ^{13}C -MFA: a comparison using a consistent dataset. *Journal of biotechnology* **154**, 179-190 (2011).
- 211 Hörl, M., Schnidder, J., Sauer, U. & Zamboni, N. Non-stationary ^{13}C -metabolic flux ratio analysis. *Biotechnology and bioengineering* **9999**, 1-13, doi:10.1002/bit.25004 (2013).
- 212 Tarca, A. L., Carey, V. J., Chen, X.-w., Romero, R. & Drăghici, S. Machine learning and its applications to biology. *PLoS Comput Biol* **3**, e116 (2007).
- 213 Angermueller, C., Pärnamaa, T., Parts, L. & Stegle, O. Deep learning for computational biology. *Molecular Systems Biology* **12**, 878 (2016).
- 214 Ye, Q.-H. *et al.* Predicting hepatitis B virus-positive metastatic hepatocellular carcinomas using gene expression profiling and supervised machine learning. *Nature medicine* **9**, 416-423 (2003).
- 215 Auslander, N., Wagner, A., Oberhardt, M. & Ruppin, E. Data-driven metabolic pathway compositions enhance cancer survival prediction. *PLoS Comput Biol* **12**, e1005125 (2016).
- 216 Shaked, I., Oberhardt, M. A., Atias, N., Sharan, R. & Ruppin, E. Metabolic network prediction of drug side effects. *Cell systems* **2**, 209-213 (2016).
- 217 Jerby, L. & Ruppin, E. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clinical Cancer Research* **18**, 5572-5584 (2012).
- 218 Menden, M. P. *et al.* Machine learning prediction of cancer cell sensitivity to drugs based on genomic and chemical properties. *PLoS one* **8**, e61318 (2013).
- 219 Dale, J. M., Popescu, L. & Karp, P. D. Machine learning methods for metabolic pathway prediction. *BMC bioinformatics* **11**, 1 (2010).
- 220 Andreozzi, S., Miskovic, L. & Hatzimanikatis, V. iSCHRUNK—*In silico* approach to characterization and reduction of uncertainty in the kinetic models of genome-scale metabolic networks. *Metabolic engineering* **33**, 158-168 (2016).
- 221 Kell, D. B. Metabolomics, modelling and machine learning in systems biology—towards an understanding of the languages of cells. *Febs Journal* **273**, 873-894 (2006).
- 222 Kell, D. B. Metabolomics and machine learning: explanatory analysis of complex metabolome data using genetic programming to produce simple, robust rules. *Molecular biology reports* **29**, 237-241 (2002).
- 223 Wu, S. G. *et al.* Rapid prediction of bacterial heterotrophic fluxomics using machine learning and constraint programming. *PLoS Comput Biol* **12**, e1004838 (2016).
- 224 Kim, M., Rai, N., Zorraquino, V. & Tagkopoulos, I. Multi-omics integration accurately predicts cellular state in unexplored conditions for *Escherichia coli*. *Nature Communications* **7**, 13090 (2016).
- 225 Michiels, S., Koscielny, S. & Hill, C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *The Lancet* **365**, 488-492 (2005).
- 226 Zamboni, N. ^{13}C metabolic flux analysis in complex systems. *Current opinion in biotechnology* **22**, 103-108, doi:10.1016/j.copbio.2010.08.009 (2011).
- 227 Dauner, M. From fluxes and isotope labeling patterns towards *in silico* cells. *Current opinion in biotechnology* **21**, 55-62, doi:10.1016/j.copbio.2010.01.014 (2010).
- 228 Gopalakrishnan, S. & Maranas, C. Achieving metabolic flux analysis for *S. cerevisiae* at a genome-scale: challenges, requirements, and considerations. *Metabolites* **5**, 521-535, doi:10.3390/metabo5030521 (2015).
- 229 Gopalakrishnan, S. & Maranas, C. ^{13}C Metabolic flux analysis at a genome-scale. *Metabolic Engineering*, doi:10.1016/j.ymben.2015.08.006 (2015).

- 230 Leighty, R. W. & Antoniewicz, M. R. COMPLETE-MFA: Complementary parallel labeling experiments technique for metabolic flux analysis. *Metabolic Engineering* **20**, 49-55, doi:10.1016/j.ymben.2013.08.006 (2013).
- 231 Muñoz-Elías, E. J. & McKinney, J. D. Carbon metabolism of intracellular bacteria. *Cellular microbiology* **8**, 10-22, doi:10.1111/j.1462-5822.2005.00648.x (2006).
- 232 Davidson, S. M. *et al.* Environment impacts the metabolic dependencies of Ras-driven non-small cell lung cancer. *Cell metabolism* **23**, 517-528 (2016).
- 233 Fontán, P., Aris, V., Ghanny, S., Soteropoulos, P. & Smith, I. Global transcriptional profile of *Mycobacterium tuberculosis* during THP-1 human macrophage infection. *Infection and immunity* **76**, 717-725, doi:10.1128/IAI.00974-07 (2008).
- 234 Eylert, E. *et al.* Isotopologue Profiling of *Legionella pneumophila* ROLE OF SERINE AND GLUCOSE AS CARBON SUBSTRATES. *Journal of biological chemistry* **285**, 22232-22243 (2010).
- 235 Westermann, A. J., Gorski, S. a. & Vogel, J. Dual RNA-seq of pathogen and host. *Nature reviews. Microbiology* **10**, 618-630, doi:10.1038/nrmicro2852 (2012).
- 236 Rienksma, R. A. *et al.* Comprehensive insights into transcriptional adaptation of intracellular mycobacteria by microbe-enriched dual RNA sequencing. *BMC genomics* **16**, 1 (2015).
- 237 Ghosh, A. *et al.* A peptide-based method for ¹³C metabolic flux analysis in microbial communities. *PLoS Comput Biol* **10**, e1003827 (2014).
- 238 Long, C. P., Au, J., Gonzalez, J. E. & Antoniewicz, M. R. ¹³C metabolic flux analysis of microbial and mammalian systems is enhanced with GC-MS measurements of glycogen and RNA labeling. *Metabolic Engineering* **38**, 65-72 (2016).
- 239 Gebreselassie, N. a. & Antoniewicz, M. R. ¹³C-Metabolic flux analysis of co-cultures: A novel approach. *Metabolic Engineering* **31**, 132-139, doi:10.1016/j.ymben.2015.07.005 (2015).
- 240 Mardinoglu, A. *et al.* The gut microbiota modulates host amino acid and glutathione metabolism in mice. *Molecular Systems Biology* **11**, 834-834, doi:10.15252/msb.20156487 (2015).
- 241 Karlsson, F., Tremaroli, V., Nielsen, J. & Bäckhed, F. Assessing the human gut microbiota in metabolic diseases. *Diabetes* **62**, 3341-3349 (2013).
- 242 Sigru, R. Prediction of intracellular metabolic states from extracellular metabolomic data. 603-619, doi:10.1007/s11306-014-0721-3 (2015).
- 243 Crown, S. B., Long, C. P. & Antoniewicz, M. R. Integrated ¹³C-metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. *Metabolic engineering* **28**, 151-158 (2015).
- 244 Peroza, E. A., Boumezbeur, A.-H. & Zamboni, N. Rapid, randomized development of genetically encoded FRET sensors for small molecules. *Analyst* **140**, 4540-4548 (2015).
- 245 Zenobi, R. Single-cell metabolomics: analytical and biological perspectives. *Science* **342**, 1243259 (2013).
- 246 Wiechert, W., Schweissgut, O., Takanaga, H. & Frommer, W. B. Fluxomics: mass spectrometry versus quantitative imaging. *Current opinion in plant biology* **10**, 323-330, doi:10.1016/j.pbi.2007.04.015 (2007).

Chapter 2

SUMOFLUX: A generalized method for targeted ^{13}C metabolic flux ratio analysis

Maria Kogadeeva and Nicola Zamboni

Published in: *PLoS Computational Biology* 12, no. 9 (2016): e1005109.

Contributions by MK:

Designed and performed experiments, analyzed the data, wrote the manuscript

Supplementary Information:

Available in Appendix I on page 147 et sqq.

ABSTRACT

Metabolic fluxes are a cornerstone of cellular physiology that emerge from a complex interplay of enzymes, carriers, and nutrients. The experimental assessment of *in vivo* intracellular fluxes using stable isotopic tracers is essential if we are to understand metabolic function and regulation. Flux estimation based on ^{13}C or ^2H labeling relies on complex simulation and iterative fitting; processes that necessitate a level of expertise that ordinarily preclude the non-expert user. To overcome this, we have developed SUMOFLUX, a methodology that is broadly applicable to the targeted analysis of ^{13}C -metabolic fluxes. By combining surrogate modeling and machine learning, we trained a predictor to specialize in estimating flux ratios from measurable ^{13}C -data. SUMOFLUX targets specific flux features individually, which makes it fast, user-friendly, applicable to experimental design and robust in terms of experimental noise and exchange flux magnitude. Collectively, we predict that SUMOFLUX's properties realistically pave the way to high-throughput flux analyses.

INTRODUCTION

Metabolic fluxes describe the *in vivo* flow of organic matter through the biochemical reaction network, as defined by enzymes and transporters. An improved knowledge of metabolic fluxes is crucial if we are to understand how cells utilize nutrients, and how they regulate metabolism in the face of dynamic environmental conditions, or in stressed pathologic states¹⁻⁴. Metabolic fluxes, as an emergent property of cellular systems, are prohibitively hard to predict using proteomics or metabolomics data, and are not, per-se, measurable. Hence, the task of assessing metabolic fluxes indirectly represents something of an analytic and mathematic tour-de-force.

The most informative approach to estimate metabolite fluxes involves stable isotope labeling. Cells grown in the presence of ^{13}C -enriched substrates incorporate heavy isotopes throughout their metabolic networks according to carbon fluxes and produce characteristic ^{13}C patterns in metabolites and products. Some of these can be measured by mass spectrometry or nuclear magnetic resonance and can ultimately be used to deduce fluxes using two basic approaches. The first is global isotopomer balancing, which seeks to estimate all metabolic fluxes by iterative fitting⁵⁻¹⁰. The power of this approach is that it integrates all available data simultaneously in order to

estimate metabolic fluxes across the entire system. The downside is that this approach is ill suited for high-throughput analyses as it necessitates quantification of all uptake and production rates, and analyzes each sample individually. In addition, the fitting procedure is mathematically cumbersome, and for complex or poorly calculable problems, can require extensive computation time. Finally, troubleshooting heavily relies on expert knowledge⁸.

The alternative approach is to use flux ratio analysis, which focuses on the resolution of local fluxes, centered on metabolic nodes of particular interest¹¹⁻¹⁶. For this purpose, flux ratio analysis adopts a targeted strategy in which relative (fractional) information on contributions from alternative pathways are calculated from a small subset of ¹³C-data using predefined analytic formulas. The advantage of this approach is that it is mathematically simple, rapid, well suited for large scale analyses¹⁶, and easily used by the non-expert user. However, this process suffers from the time-consuming procedure of deriving analytic formulas for each flux ratio of interest. These formulas, manually derived for each metabolic network, tracer, and environment, generally incorporate a mix of human intuition together with tacit assumptions regarding flux. Over the past 20 years, only a dozen have been derived to describe the central metabolism of microbes growing on single carbon sources. In practice, most experimental conditions cannot be addressed due to the lack of validated flux ratio predictors. In response to these limitations, automated tools have been developed to estimate flux ratios^{17,18}, although, thus far, these have been limited to linear cases and consequently have failed to find any broader application.

Here we present SUMOFLUX, a conceptually novel method to analyze, in a targeted fashion, flux ratios based on ¹³C-data. Our workflow circumvents concerns over the relevance and limitations of flux analyses by exploiting machine learning. A machine learning predictor is trained using *in silico* ¹³C-data, generated by surrogate modeling. The combination of surrogate modeling and machine learning permits the rapid estimation of flux ratios for virtually any metabolic network, label configuration, or available measurement. We now illustrate the proposed workflow for both canonical and novel flux ratios for central carbon metabolism. The speed and generality provided by machine learning makes SUMOFLUX particularly useful for optimizing experimental design, selecting metabolites to be measured, and merging data from several experiments. Moreover, we believe that the SUMOFLUX workflow provides a real prospect of high-throughput flux analyses.

RESULTS

Surrogate modeling and the workflow of ^{13}C -flux ratio analysis. In ^{13}C -metabolic flux ratio analysis, the goal is to estimate a flux ratio of interest. Typically, this is a number that indicates the relative fraction of a specific metabolite flowing through a chosen reaction or pathway. Flux ratios are estimated based on a stoichiometric model, knowledge of the ^{13}C -configuration of all of the relevant substrates, and the labeling patterns of metabolites as measured by mass spectrometry (or nuclear magnetic resonance). We formulated the derivation of flux ratio estimates from ^{13}C data as a nonlinear regression task to be solved using machine learning. By definition, the flux ratio of interest is the dependent variable that we aim to predict; measured ^{13}C isotope labeling patterns of intracellular metabolites are the independent variables, or the input features for the algorithm. A random forest predictor¹⁹ is then trained to build a functional relationship between the ^{13}C data and flux ratios using a training dataset. To build a generalized predictor, the training dataset should comprise hundreds, if not thousands of representative examples for which a flux ratio and ^{13}C data are available. Unfortunately, such a dataset is not accessible experimentally. First, because flux estimates are not amenable to direct measurement. Second, in the majority of cases it is impossible to select, or to construct, a cohort of cells with a phenotypic diversity that adequately represents the wide variety of fluxes and flux ratios that might exist. To overcome this fundamental problem, we have used surrogate modeling (hence the term SUMOFLUX). We built, *in silico*, a synthetic cohort of representative data points. Each data point is defined by a complete set of fluxes that fulfill the stoichiometric constraints of the metabolic network. This allows us to calculate a ratio (or any other derivative value) for the fluxes of interest, and to simulate the ^{13}C -labeling patterns of each metabolite, which is made possible because each flux distribution leads to a unique intracellular labeling pattern²⁰. It is therefore possible to construct an *in silico* dataset comprising thousands of data points, with flux ratios spanning the feasible range, and corresponding metabolite ^{13}C -labeling patterns. The synthetic *in silico* dataset is used to train, cross-validate, and then test the flux ratio predictor.

The full SUMOFLUX workflow for flux ratio estimation consists of five steps (Figure 1). First, a reference dataset of several thousand flux maps is sampled from that space of flux maps that fulfills certain stoichiometric constraints (mass balances) of the metabolic network. Extracellular flux constraints can be further refined by the

availability of substrates and a working knowledge of the major secreted products. Second, the ^{13}C -labeling patterns of the metabolites included in the network are simulated independently for each reference data point. Label propagation is then simulated using the existing algorithms⁹, given the ^{13}C -label of the substrate(s), and the map of the atom transition within the network. At this point, the simulated ^{13}C data, does not, as yet, reflect actual measurements. Third, to capture measurement data, we select only those ^{13}C features that are analytically accessible, and then superimpose noise values corresponding to those measured. Fourth, flux ratios of interest are calculated for all of the data points within the reference dataset, as the dependent variable in regression analyses. Fifth, we divide the reference dataset into independent training and test subsets, using the former to train a random forest with which to predict the calculated flux ratios from simulated ^{13}C data. We then assess the predictor's performance on the test dataset by calculating the mean absolute error of the predictions made. If the performance is insufficient (e.g. mean absolute error $\text{MAE} > 0.05$), we iteratively optimize our experimental strategy by changing the substrate label, or available measurements, then repeat the training. If the performance is judged to be satisfactory, we finally use the predictor to estimate flux ratios using real experimental data. To provide prediction intervals, we use quantile regression forests, which give a non-parametric and accurate estimates of conditional quantiles based on the full conditional distribution of the dependent variable²¹.

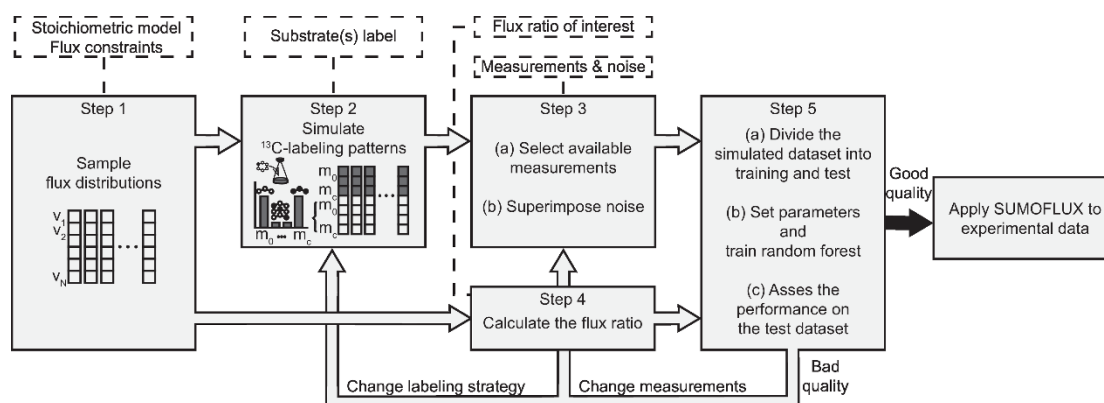


Figure 1. SUMOFLUX workflow for targeted ^{13}C flux ratio analysis. Input data are depicted in the dashed-line rectangles.

The most time consuming aspect of the workflow is the simulation of ^{13}C data in the reference dataset, which scales according to the number of samples and carbon atoms in the metabolites. For the model of central carbon metabolism with 39 reactions and 21 measured metabolites and fragments (Figure S1, Tables S1 and S2), 0.2 seconds are needed to simulate the labeling patterns for a single data point. Using a parallelization technique, this process can be accelerated to simulate the several thousand data points necessary for training and testing within a few minutes. Without parallelization, the simulation procedure for 20,000 data points takes ~ 1.5 hours, whereas the flux sampling and random forest training steps require less than a minute.

Overall, the SUMOFLUX workflow requires information on the stoichiometry of the metabolic network, and the carbon atom arrangement for all of the metabolites within the network. The choice of ^{13}C -tracer depends on the flux ratio of interest^{8,22}, but in practice is primarily constrained by commercial availability and costs. Hence, it is quite common to test flux calculability using multiple configurations of tracers²³⁻²⁵, which can be easily accomplished using the SUMOFLUX workflow due to its rapid computational time. In the following sections, we demonstrate the performance, generality, and scalability of SUMOFLUX, as well as its versatility in terms of feature selection and experimental design.

Analyses of flux ratios for central carbon metabolism. We chose to demonstrate SUMOFLUX by deriving estimates of the flux ratios for central carbon metabolism using the model organism, *Escherichia coli*. Its metabolic network includes the highly conserved pathways of glycolysis, the tricarboxylic acid (TCA) cycle, and the pentose phosphate (PP) pathway. Furthermore, it includes alternative pathways such as the Entner-Doudoroff pathway and the glyoxylate shunt that convey additional metabolic elements that might complicate flux estimates. As a reference, we considered the study of glucose metabolism in *E. coli* as described by flux ratio analyses using manually derived analytic equations¹⁵. We used our method to estimate five key flux ratios based on the labeling patterns measured by gas chromatography mass spectrometry (GC-MS) of proteinogenic amino acids upon silylation. We then sampled a reference dataset of 60,000 flux distributions using the *E. coli* central carbon metabolism network (Figure S1, Table S1), and simulated the labeling patterns of 21 intracellular metabolites and their fragments (Table S2), assuming growth on either 100% [1- ^{13}C] glucose, or a mix of 20% [U- ^{13}C] glucose and 80% naturally labeled glucose.

Several parameters had to be defined prior to predictor training. The performance of the random forest depends on the number of decision trees in the forest (*ntree*), and the number of input features used at each tree node (*mtry*). To choose these parameters we used five-fold cross-validation on the training dataset. We tested 16 combinations of *ntree* and *mtry* values for the five *E. coli* flux ratios. The combination of 100 prediction trees (*ntree*) with 20 *mtry* features delivered a good balance between predictor accuracy and computation time (Figure S2a). These two parameters were then applied throughout the study. The number of simulated points used for training also influences predictor accuracy and computation time. Our tests demonstrated that ~10,000 simulated points were generally adequate in terms of generating a sufficiently accurate estimate of the key flux ratios in the *E. coli* dataset; thereafter, any further increase in the number of data points provided no tangible improvement in accuracy (Figure S2b). We took these results into account when extracting the training sets for the predictors (see Materials and Methods for details).

We trained predictors for the five *E. coli* flux ratios on the simulated training dataset and then assessed their performance on an independent simulated test dataset. In all cases, the mean absolute error was < 0.1 (Figure 2, second column). For comparison, we also applied the analytic formulas manually derived for the *E. coli* study¹⁵ (Table S3) to the same simulated test dataset. For all tested flux ratios, SUMOFLUX outperformed the analytic formulas in terms of mean absolute error on the test dataset (Figure 2, fourth column). This possibly reflects the fact that the flux estimates for the test dataset were obtained through sampling of the entire solution space, and do not comply with some of the implicit simplifications and assumptions for the network, fluxes, and reaction reversibility, which are generally used to derive the analytic formulas¹⁵. For example, in calculating the fraction of oxaloacetate from phosphoenolpyruvate, the flux through the glyoxylate shunt was assumed to be zero, whereas in the test set it possessed a wide range of values (Figure 2e). Furthermore, the analytic formula for estimating the malic enzyme flux ratio provides only a lower bound value (Figure 2d). We also compared flux estimates generated by the two approaches using the real experimental ¹³C data from this study (Figure 2, third column, and Table S4); both produced concordant estimates (Pearson correlation coefficient, $PCC > 0.89$ for all ratios).

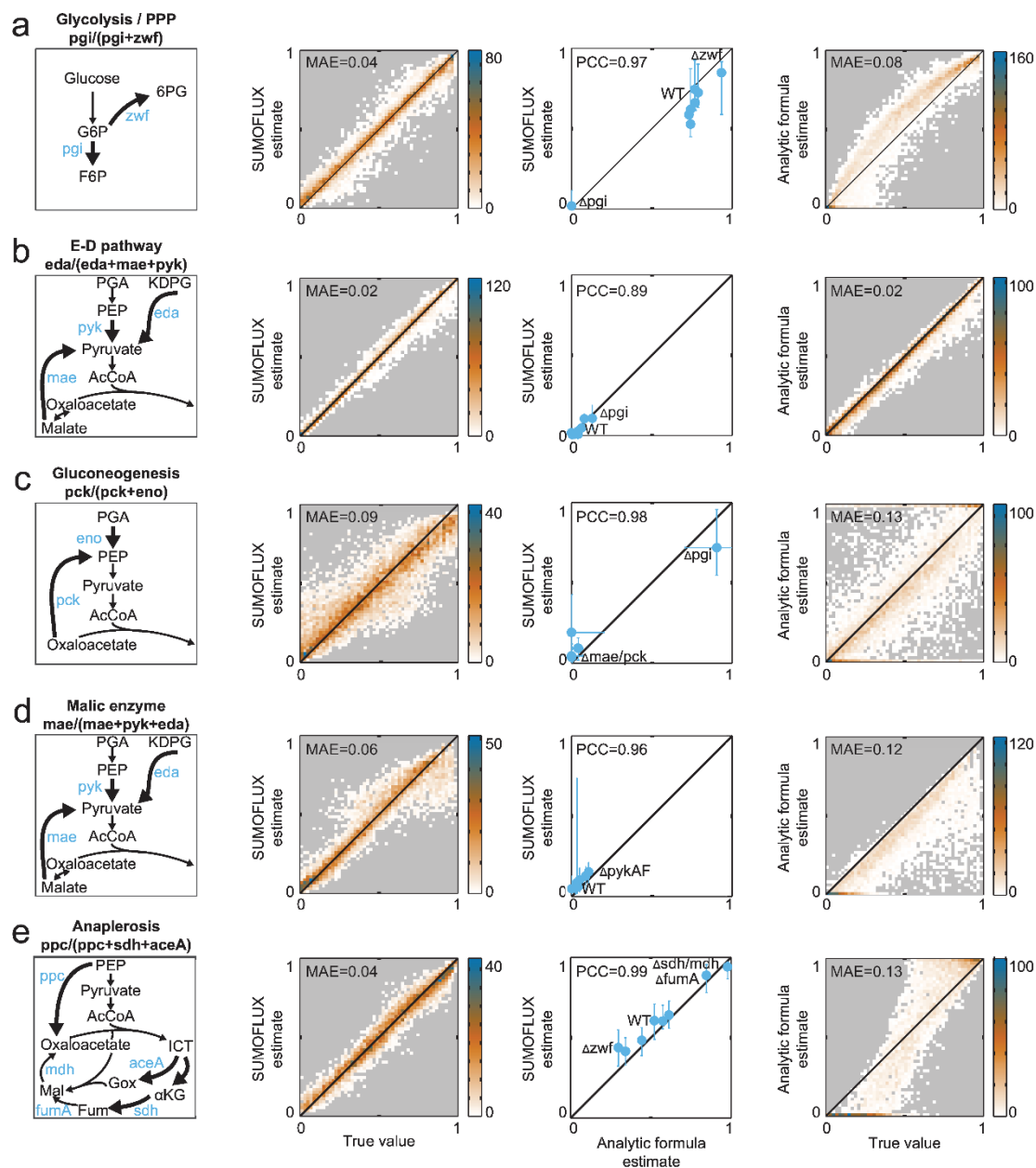


Figure 2. Comparison of SUMOFLUX and analytic formula estimates for flux ratios in *E. coli* central carbon metabolism. From left to right: a schematic representation of the flux ratio; density plot representing SUMOFLUX estimates versus the true flux ratios for *in silico* data; comparison of the SUMOFLUX and analytic formula estimates for the experimental data; density plot representing analytic formula estimates versus the true flux ratios for *in silico* data. Vertical error bars in the third panel represent [10%-90%] SUMOFLUX prediction quantiles, horizontal error bars represent standard deviation obtained with the analytic formula estimate. (a) Glycolysis versus PPP. (b) Pyruvate fraction from the E-D pathway. (c) PEP fraction from gluconeogenesis. (d) Pyruvate fraction from the malic enzyme flux. (e) Oxaloacetate fraction from anaplerosis from PEP. Ratios (a)-(c) were estimated from [1-¹³C] glucose experiment, ratios (d) and (e) were estimated from 20% [U-¹³C] and 80% naturally labeled glucose experiment. 6PG – 6-phospho-D-gluconate; α KG – α -ketoglutarate; AcCoA – acetyl-CoA; E-D – Entner-Doudoroff pathway; F6P – fructose-6-phosphate; Fum – fumarate; G6P – glucose-6-phosphate; Gox – glyoxylate; ICT – isocitrate; KDPG - 2-Keto-3-deoxy-6-phosphogluconate; MAE – mean absolute error; Mal – malate; PCC – Pearson correlation coefficient; PEP – phosphoenolpyruvate; PGA – phosphoglycerate; PPP – pentose phosphate pathway.

To further demonstrate the scalability and generality of SUMOFLUX, we applied the same approach and parameters to estimate four flux ratios using the GC-MS data for amino acids collected for 121 *Bacillus subtilis* transcription factor mutants grown on a mixture of 80% [1-¹³C] glucose and 20% [U-¹³C] glucose¹⁶. Again, the random forest predictor outperformed the analytic formulas for the *in silico* test dataset (Figure S3). For three flux ratios, the two approaches provided consistent estimates for the experimental data (PCC > 0.65). However, the malic enzyme ratio could not be resolved with sufficient precision using either method. Presumably, the mixture of tracers chosen was poorly suited to this task.

In order to highlight the scope of SUMOFLUX applicability in context of global ¹³C flux analysis methods, we compared it with the classical ¹³C-metabolic flux analysis by global isotopomer balancing (¹³C-MFA) approach, which seeks for a global flux solution that provides the best fit to the experimental data – measured metabolite labeling patterns and physiological parameters. We applied ¹³C-MFA to the data for the same eight *E. coli* strains and added glucose uptake rates¹⁵ as an additional input. With INCA software⁹, we calculated the best flux fit and flux confidence intervals using parameter continuation procedure (Table S5). SUMOFLUX and ¹³C-MFA differ in the demand for input information and produce different outcomes (flux ratios vs. net fluxes). To compare, we calculated flux ratios from the net fluxes estimated by ¹³C-MFA and directly compared those to SUMOFLUX results. Confidence intervals on flux ratios for ¹³C-MFA were obtained by repeating the optimization procedure 1000 times for each strain. Because it employs less input data, SUMOFLUX is expected to be worse than ¹³C-MFA. In general, however, the flux ratio estimates obtained with the two methods were in good agreement (PCC>0.83 for all ratios calculated for the best fit to either [1-¹³C] data, [U-¹³C] data, or combined dataset, Figure S4). Surprisingly, in several cases the confidence intervals of ¹³C-MFA flux ratio estimates were much larger than the prediction quantiles of SUMOFLUX and the accuracy of flux ratio estimates depended on the experimental dataset used for the fitting, perhaps pointing to the presence of inconsistent or overly noisy data that decrease the precision of ¹³C-MFA estimates. This example illustrates the complementarity of the two approaches. ¹³C-MFA provides global flux solutions, but in some cases the targeted approach performs better in resolving local fluxes.

Collectively, the *E. coli* and *B. subtilis* results demonstrate that SUMOFLUX is broadly applicable to real experimental data with an accuracy that is comparable, if not better, than that of manually derived formulas. Even though there is no guarantee that a specific flux ratio can be accurately estimated for a given metabolic network, tracer, or experimental data, SUMOFLUX does allow for rapid verification and ad-hoc experimental design. Beyond the speed and ease with which predictors can be generated for calculating metabolic flux ratios from ^{13}C data, SUMOFLUX offers additional benefits of robust prediction, the option to vary and optimize experimental design, and the estimation of novel ratios that we explore in the next sections.

SUMOFLUX is robust in terms of experimental noise and reversible reactions.

Excessive measurement noise and underestimation of exchange flux of bidirectional reactions are two frequent causes of inaccurate flux estimates. We set out to assess their influence on SUMOFLUX by performing an *in silico* experiment using *E. coli*, varying the values of the superimposed measurement noise by up to 0.10, i.e. 10-fold higher than that routinely obtained with careful peak integration. We also used exchange flux values of up to 100-fold that of the net flux value, i.e. a model approximation of full equilibration of the reactants. To exclude potential prediction accuracy differences arising from different training and test datasets, we used one set of flux vectors, divided into training and test subsets. Addition of four different noise levels and variation of four exchange flux magnitude values resulted in 16 datasets, which differ only in these two parameters. We again trained the predictors for each of the datasets using the training subset and calculated the mean absolute error (MAE) on an independent test subset. As a rule of thumb, we consider a ratio to be accurately predictable if the $\text{MAE} < 0.05$. This criterion was met by the Entner-Doudoroff, glycolysis/PP pathway, and anaplerosis ratio predictors, within the normal ranges for noise (~ 0.01) and exchange flux (~ 10 times the net flux value) (Figure 3a). The other two tested predictors were less precise, and are better suited for the analysis of substantial flux changes. Alternatively, different tracers and measurement techniques could be tested, as outlined below, to achieve more accurate analyses. We also performed robustness analyses of the analytic equations, and found that only the formula for the Entner-Doudoroff pathway was sufficiently robust in terms of noise and flux exchange that were within the normal ranges (Figure S5). The remaining four formulas were either extremely

sensitive to noise (gluconeogenesis ratio), or were poorly suited to the entire range of flux maps.

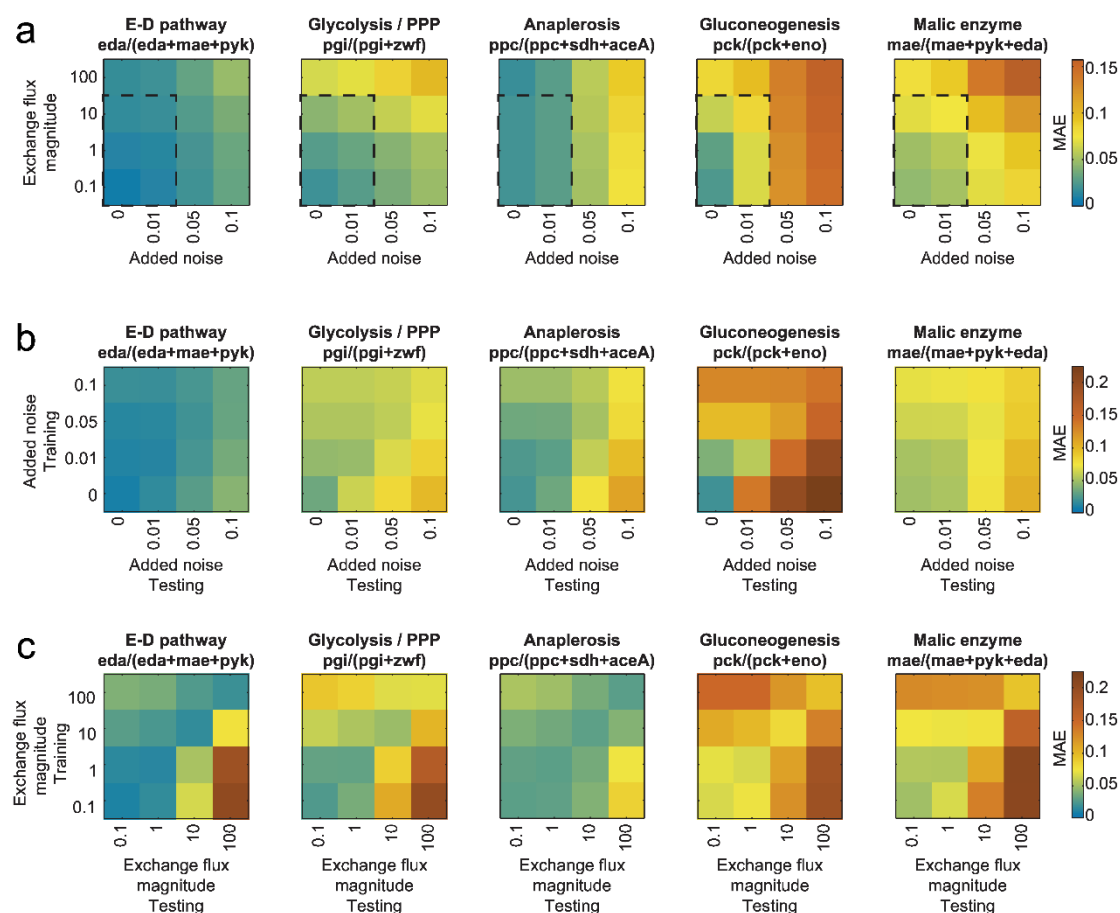


Figure 3. SUMOFLUX is robust in terms of experimental noise and exchange flux magnitude. (a) Mean absolute errors on the testing dataset of five flux ratio predictors applied to *in silico* data with different amount of measurement noise and exchange flux magnitude. The dashed rectangle indicates the normal range of noise (0.01) and exchange flux magnitude (10 times the net flux). (b) Mean absolute errors on the testing datasets with different noise levels of five flux ratio predictors trained on datasets with different amount of measurement noise. The exchange flux magnitude was set to 1 for all datasets. (c) Mean absolute errors on the testing dataset with different exchange flux magnitudes of five flux ratio predictors trained on datasets with different values of exchange flux magnitude. The noise level was set to 0.01 for all datasets. E-D – Entner-Doudoroff pathway, MAE – mean absolute error; PPP – pentose phosphate pathway.

Another important aspect that can be assessed with this type of analysis is to what extent erroneous assumptions in the training dataset affect the accuracy of the flux estimates in the test dataset. We used the simulated data described above to test the effects of noise and exchange flux magnitude values separately. First, we fixed the exchange flux magnitude to 1, and calculated the accuracy of the flux ratio predictors trained and tested on 16 combinations of train and test subsets with the four independently added noise levels. Notably, we observed that for all ratios

underestimating the level of noise in the data was detrimental for the flux ratio prediction accuracies. On the contrary, overestimating the noise in the training data resulted in better flux estimates in the less noisy test datasets compared to the test data with the same levels of noise as in training (Figure 3b). In practice, it is always desirable to only superimpose a realistic level of noise to *in silico* data, as the addition of noise inevitably decreases the prediction accuracy. However, it is advisable to adopt a conservative over-estimate noise to avoid overfitting.

Adequate magnitude of exchange fluxes appears to be even more important for predictor accuracy. We observed that both under- and over-estimated exchange flux magnitude values resulted in lower accuracy compared to the accuracy on the test dataset corresponding to the training (Figure 3c, diagonal values). Remarkably, in these simulated datasets we set one exchange value as upper bound for all reversible fluxes in the model, which presumably has a greater effect on prediction accuracy than an exchange flux magnitude of a single reaction. In biological systems, we expect large differences in the reversible flux magnitudes of different enzymes, and it is beneficial to include this information in the model, when available.

One way to control whether the simulated data used for training and testing adequately represents the experimental data, is to compare the distributions of inter-quantile ranges of the flux ratio estimates for the experimental data to the ones of the test data. In our examples, the distributions of the inter-quantile ranges of the flux estimates for under-estimated noise or inappropriately estimated exchange flux estimates are significantly different ($p < 0.01$, Wilcoxon-Mann-Whitney test, Figure S6). In practice, it is advisable to compare the inter-quantile range distributions of the flux ratio predictions for *in silico* and experimental data, although statistical tests should be used with caution due to very different sample sizes.

In summary, SUMOFLUX provides flux ratio predictors that are generally robust to noise and exchange fluxes, both of which are major confounding factors in labeling experiments. This robustness is dependent on flux ratio, labeling strategy, and the available measurements used for prediction, and can easily be assessed, if required, in each particular case.

Estimation of a novel flux ratio for the glyoxylate shunt. The glyoxylate shunt plays an essential role in bacterial adaptation to alternative carbon sources, such as acetate and fatty acids, as it replenishes the TCA cycle with C_2 carbon fragments. Hence, this

pathway has an important anaplerotic function besides phosphoenolpyruvate carboxylase. No analytic formulas were developed to resolve the relative contribution of the glyoxylate shunt due to the complexity of carbon rearrangement at this branch point, the additional complication introduced by multiple cycling in the TCA cycle, and the similarity of the labeling patterns of the relevant metabolites. Here, we opted to tentatively resolve this pathway using SUMOFLUX, and the ^{13}C -data available from GC-MS analyses of protein-bound amino acids in *E. coli*¹⁵. Using the same simulated dataset described above, we trained two more predictors to estimate flux contributions to the formation of oxaloacetate, one derived from the glyoxylate shunt, and the other from the TCA cycle (Figure 4a). The accuracy of the predictors achieved for the *in silico* test dataset was acceptable (MAE < 0.07) (Figure 4b, c). Collectively, these two novel ratio predictors, and the one previously trained to estimate the anaplerotic reaction from phosphoenolpyruvate to oxaloacetate (Figure 2e), allowed us to comprehensively assess the metabolic source of oxaloacetate. The prediction intervals of the estimates for the experimental data were in the range of 10% due to the difficulty of precisely resolving the glyoxylate shunt based on the available data. Nevertheless, the estimated values reflect those expected from the literature. Specifically, the differences between strains were consistent with their genotype (Figure 4d, S7, Table S6). The estimated glyoxylate shunt contribution for wild type bacteria was $16 \pm 10 \%$, with the highest glyoxylate shunt ratio ($32 \pm 15 \%$) estimated for the Δpgi mutant, which is consistent with other studies^{26,27}. In contrast, both the double $\Delta\text{mdh} \Delta\text{sdh}$ mutant and the ΔfumA mutant, in which the pathway from succinate to malate is disrupted, had an almost zero glyoxylate shunt and TCA cycle activity, with the major contribution to the oxaloacetate pool being the flux derived from phosphoenolpyruvate. The Δzwf mutant, with a compromised oxidative pentose phosphate pathway, exhibited the highest fraction for the TCA cycle flux ($49 \pm 18 \%$), which reflects a compensatory response to ensure NADPH equilibrium via isocitrate dehydrogenase²⁶. The glyoxylate shunt example shows how novel quantitative flux predictors can be rapidly generated using our approach.

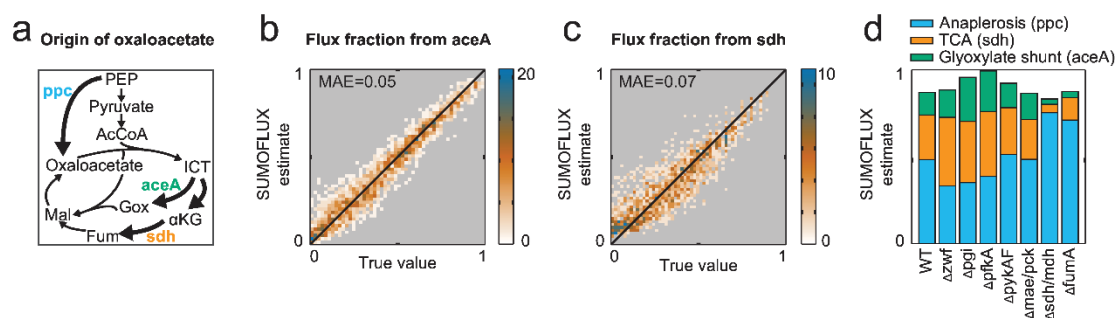


Figure 4. SUMOFLUX resolves a novel flux ratio in central carbon metabolism of *E. coli*. (a) A schematic representation of the glyoxylate shunt, TCA cycle and anaplerosis from PEP flux fractions. (b) Density plot representing SUMOFLUX estimates for the flux fraction from glyoxylate shunt versus the true flux ratios for *in silico* data. (c) Density plot representing SUMOFLUX estimates for the flux fraction from the TCA cycle versus the true flux ratios for *in silico* data. Both ratios were resolved for experiment with 20 % [U-¹³C] and 80 % naturally labeled glucose. (d) Predictions for the three flux fractions for the experimental data. αKG – α-ketoglutarate; AcCoA – acetyl-CoA; Fum – fumarate; Gox – glyoxylate; ICT – isocitrate; MAE – mean absolute error; Mal – malate; PEP – phosphoenolpyruvate; TCA – tricarboxylic acid cycle.

Experimental design. In the context of metabolic analyses, *a priori* experimental design aims at identifying the best settings from simulated data with which to accurately estimate the fluxes of interest. In global isotopomer balancing and fitting, numerical simulations have been frequently used to optimize tracer selection for one specific flux state, e.g. that of an unperturbed wild-type strain²³⁻²⁵. In targeted flux ratio analysis with manually derived analytic equations, simulation-assisted experimental design is not possible, as each equation is formulated for a specific experimental condition chosen by the researcher, and no simulation procedure is employed to assess its accuracy. In contrast, the speed and simplicity of SUMOFLUX facilitates the rapid testing of altered metabolic models, tracer choices, or data sets for the derivation of a flux ratio of interest. This enables us to systematically, yet rapidly, identify the optimal experimental strategy from those available.

We demonstrated this feature of SUMOFLUX by testing different settings for the *B. subtilis* labeling experiment. Using the same reference flux dataset as above, we simulated the ¹³C metabolite labeling patterns for eight different glucose labeling strategies. For each label, we simulated the measurements that could be obtained using four different measurement techniques: GC-MS analyses of amino acids, liquid chromatography LC-MS of intact intracellular metabolites, LC-MS/MS analyses of intact metabolites and their fragments²⁸ (Table S7), and all individual MS/MS traces used in multiple reaction monitoring of metabolites (Table S8). For each of the 32

experimental setups, we rapidly trained random forest predictors for the malic enzyme, gluconeogenesis, and glycolysis/PP pathway flux ratios, and assessed their performance *in silico* on the test dataset. To compensate for the different number of features and avoid over-fitting, we introduced a feature selection procedure using cross-validation on the training dataset prior to training (see Materials and Methods for details). As expected, flux calculability depends on the flux ratio of interest, the tracer, and the measurement platform (Figure 5). For the malic enzyme and glycolysis ratios, LC-based methods are preferable to GC-MS. Tracers such as [6-¹³C], [5,6-¹³C], or [4,5,6-¹³C] glucose offer the best overall accuracy (MAE < 0.05). Any of these tracers could be selected to quantify the three flux ratios in a single experiment. For specific flux ratios, the average error was reduced to about MAE 0.02-0.03 by selecting specific tracers. However, when taking into account the cost of tracers, a labeling experiment using 50% [U-¹³C] and 50% naturally labeled glucose might be seen as a compromise between prediction accuracy and cost. This analysis underlines the ease with which an experimental design targeted to address specific biological questions can be implemented using the SUMOFLUX workflow.

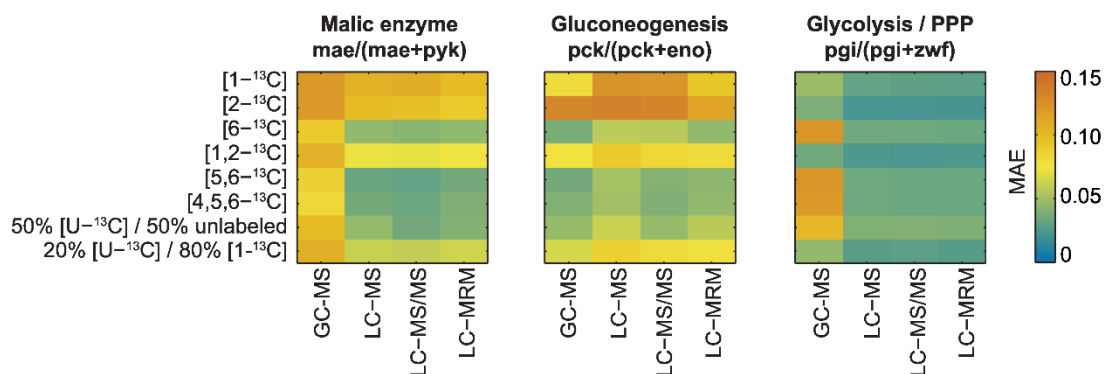


Figure 5. Optimizing experimental design to improve the estimation of three flux ratios in *Bacillus subtilis* central carbon metabolism. Mean absolute errors on the test dataset of three flux ratio predictors applied to *in silico* data simulated with different experimental setups. GC-MS – gas chromatography mass spectrometry, LC-MS – liquid chromatography mass spectrometry; LC-MS/MS - liquid chromatography-tandem mass spectrometry; LC-MRM – liquid chromatography with multiple reaction monitoring information; MAE – mean absolute error; PPP – pentose phosphate pathway.

DISCUSSION

We have developed a generalized method for targeted analysis of ^{13}C metabolic flux ratios, that builds on surrogate modeling (SUMOFLUX), i.e. uses a synthetic dataset to train a machine learning predictor to estimate a given flux ratio directly from ^{13}C -data. Synthetic datasets are constructed *in silico* on the sole basis of four easily accessible inputs: a stoichiometric model of metabolism, a list of possible metabolic substrates and their byproducts, a configuration of the ^{13}C -substrate, and a list of measurable metabolites with measurement error. These inputs are sufficient to generate a representative synthetic dataset covering a broad range of fluxes and flux ratios. A random forest predictor is then trained on this dataset to capture the relationship between simulated ^{13}C -data and the flux ratio of interest that holds true for all of the simulated data points. Therefore, the same predictor can be used to estimate flux ratios for normal cells, as well as for knockout mutants without the need for additional information on physiological parameters or their uptake/consumption rates. Due to the fact that the SUMOFLUX predictor targets only a single flux ratio at a time, it is very efficient in assessing calculability and eventually estimating flux values from real data. This feature is particularly relevant when tackling complex fluxes^{29,30}, as dozens of different experimental designs can be trialed within a few hours. If necessary, measurement data from parallel experiments using different ^{13}C tracers can be combined and passed as input features into the SUMOFLUX workflow. This approach has been proven to improve flux estimates in certain cases^{31,32}. The surrogate modeling of cells grown in rich media with multiple substrates is made possible because carbon labeling experiments can be simulated for large or even genome-wide networks³³, inclusive of all the key metabolic pathways. Overall, SUMOFLUX is generally applicable to virtually any combination of metabolic model (organism), medium composition, isotopic tracer, or measurement technique.

The crucial step in SUMOFLUX is the construction of the synthetic data used for predictor training. To obtain representative data, it is extremely important that the surrogate model is based on realistic assumptions of the metabolic network and experimental measurement accuracy. Prior knowledge can be integrated into the sampling procedure to limit the space of flux distributions and potentially improve the predictor's performance. Network simplification and constraining bear some risks. The metabolic model should encode all possible metabolic reactions such that ^{13}C -patterns

can be correctly assigned to the underlying flux states. If a reaction is omitted from the surrogate model, the predictor will provide biased estimates. Although the omission of reactions from a model leads to better accuracy *in silico*, that step would only be justified if the reaction was proven to be inactive under all conditions tested, e.g. by biochemical assay or enzyme quantification. Unless such information is available, it is recommended that all reactions be included in the model in order to achieve robust predictor training. For similar reasons, it is equally important to provide a real-life or conservative error model of the measurement data. According to *in silico* testing, overestimating noise in the simulated dataset does not lead to overestimating predictor's accuracy, on the contrary to underestimating noise. In our experience, a valid sanity check is to verify that the simulated data distribution covers the measured mass isotopomer fractions by comparing the distributions of simulated and experimental data. Another indicator of potential discrepancies between the simulated and experimental data is the difference between the distributions of interquantile ranges of the flux ratio predictions, which can be tested with a nonparametric test, such as Wilcoxon-Mann-Whitney. With these simple procedures, errors in the metabolic model, substrate composition, or experimental measurement can be detected.

Targeted flux ratio analysis using SUMOFLUX is best suited to the assessment of flux ratio with high accuracy, on selected metabolic nodes, or when mid to large throughput is necessary. High-throughput is made possible by the speed of the approach and by the fact that only ^{13}C -data are required. Once trained, the predictor can be applied to estimate flux ratios for all tested samples simultaneously. A further optimization of experimental measurement time can be explored by including feature selection during training to identify the most informative as well as negligible ^{13}C -features. SUMOFLUX complements the alternative global isotopomer balancing and iterative fitting method (^{13}C -MFA), which requires measurements of uptake/consumption rates, and more detailed analyses of each dataset, but provides net flux estimates for all reactions in the model. Our short comparison with the data of eight *E. coli* strains demonstrated that the two approach deliver consistent flux ratio estimates. In some instances, the confidence of SUMOFLUX estimates was better. Hence, it could be used before ^{13}C -MFA to increase its performance. In this case, multiple flux ratios could be estimated independently to obtain experimental information on different degrees of freedom prior to applying global ^{13}C flux analysis methods³⁴.

In principle, the concept of SUMOFLUX can be extended to isotopically non-stationary data. The simulation of dynamic ^{13}C -data can be completed with the inclusion of metabolite concentrations in the sampling procedure with simulation of ^{13}C dynamics at predefined time points to be matched in the experiment. The training of flux predictors from isotopically non-stationary data can use the same procedure outlined for stationary data, even though it is substantially more demanding because of the requirement to sample an increased number of degrees of freedom and measurable labeling features. However, it must be stressed that non-stationary labeling experiments are much more labor-intensive and data demanding, and can be performed only at low throughput³⁵. For practical reasons, the traditional approach of flux estimation by both global^{35,36} or local³⁷ iterative fitting is better suited to the analysis of small-scale non-stationary labeling experiments.

Overall, the concepts underlying the proposed SUMOFLUX workflow are easily transferrable and can be applied alone, or in combination with other methods, to address different flux analyses questions. We believe that SUMOFLUX has the potential to become a core tool in the analysis of metabolic fluxes, and opens new possibilities for high-throughput flux profiling of a wide variety of metabolic systems.

EXPERIMENTAL PROCEDURES

Network construction. Metabolic network with carbon atom transitions and the lists of input and output metabolites are defined by the user and are represented in the mat-file format required by the INCA software⁹. In order to reduce the dependency on the biomass vector coefficients, a separate output flux is defined for each of the biomass precursors, therefore biomass precursors are also added to the list of outputs. The substrates are defined as unbalanced compounds and do not participate in the stoichiometric equation system.

Flux sampling and ratio calculation. In the flux sampling procedure, the definitions of net, exchange, forward and backward fluxes are used³⁸. By default, the lower and upper bounds for reversible reactions are set to [-100 100], for irreversible reactions to [0 100], and the major uptake flux is set to 10. First, the initial net flux solution is found by minimizing the sum of squared fluxes with stoichiometric constraints, inequality constraints on the output fluxes, and flux bounds using the MATLAB solver `fmincon`. Second, a cohort of net flux vectors is generated with Monte Carlo sampling by adding linear combinations of null vectors of the stoichiometric matrix with random coefficients to the initial flux solution. Third, for each net flux, an exchange flux value is randomly generated in the order of magnitude relative to the net flux defined by the user (by default 1), and forward and backward flux values are calculated accordingly. Optionally, to achieve uniform coverage of values for a particular flux ratio or set of ratios, the ratio range is split into segments (for example, [0 0.1], [0.1 0.2] ... [0.9 1]), and the flux sampling procedure is repeated for each segment with the end points set as flux ratio constraints in the first step. The flux ratio of interest is calculated for each of the flux vectors with a formula defined by the user.

¹³C labeling patterns simulation and measured data simulation. Given the label of the substrate(s) and the list of metabolites and fragments, metabolite labeling patterns are simulated for each flux solution using the INCA software⁹. The INCA ‘simulate’ procedure is integrated into the SUMOFLUX workflow and is called internally for each of the sampled flux vectors. In case parallel computing is available, this procedure is parallelized.

The measurement data is simulated by extracting the measured compounds from the simulated data matrix and adding uniform noise to the measurements (0.01 by default). After adding noise, the mass distribution vectors for each metabolite are normalized.

Prediction procedures. Preparing the training and test samples and building the predictor. For each prediction task, a subset of data points is extracted from the entire simulated dataset to ensure that the dependent variable (flux ratio of interest) is uniformly distributed on the feasible range. The flux ratio values of the whole dataset are binned into segments (for example, [0 0.1], [0.1 0.2] ... [0.9 1]), and from each bin an equal amount of samples is drawn without replacement. This simulated subset is randomly divided in proportion 2:1 to form the training and test subsets. The MATLAB randomforest package (<https://code.google.com/archive/p/randomforest-matlab/>) modified to perform quantile regression²¹ is used to build the predictor. The `regRF_train` function is called with `ntree` trees and `mtry` variables for the node split (by default, `ntree=100` and `mtry=20`) to train the predictor on the training dataset. The predictor's estimates on the test dataset are obtained with the function `regRF_predict`, and the performance is assessed by calculating the mean absolute error between the estimates and the known simulated flux ratio values.

Cross-validation procedure. The five-fold cross-validation procedure is performed in the following way: the training dataset is randomly divided into five parts of equal size. One of the parts is the validation subset and is used to assess the predictor performance, whereas the other four are used to train the predictor with a certain set of parameters. The procedure is repeated until all subsets were used as validation subsets once, and the error is averaged across all the subsets.

Prediction quantile calculation. The prediction quantile calculation is based on the quantile regression forest algorithm²¹. To calculate prediction quantiles for a new data point, `regRF_predict` function is called with `extra_options` parameters (`extra_options.predict_all=true`, `extra_options.nodes=1`). For each tree in the forest, the terminal node where the new data point propagated is recorded. The same procedure is repeated for all data points in the training dataset. For each data point in the training dataset, a weight is calculated based on how many times this data point and the new

data point propagated to the same terminal node. The data points in the training dataset are sorted by the value of the dependent variable (flux ratio of interest), and the weights cumulative function is calculated. This function is used to estimate the prediction median (50%-quantile) and to report the [10% 90%] prediction interval.

Feature selection. If the number of features (measured labeling patterns) is large, the predictor's performance on the test dataset might decrease due to overfitting to the training dataset. To reduce the chance of overfitting, feature selection is performed using the cross-validation procedure. First, the predictor is trained based on all available features, and its performance is assessed with cross-validation. Second, the features are ranked according to the feature importance value calculated during training by `regRF_train` function. Third, the predictor is trained based on [50%, 25%, 10%, 5%] of the most important features. The mean absolute error is calculated with cross-validation. The percentage of features with the smallest MAE is selected for the further training.

Noise and exchange flux magnitude sensitivity analysis. To assess the sensitivity of the flux ratio predictors to noise and exchange flux magnitude, a set of 60'000 net fluxes was sampled, and the exchange flux values were randomly generated using each of the tested parameters as the upper bound: [0.1 1 10 100]. Isotope labeling was simulated for each of the four flux datasets as described previously. For each dataset, random noise was added using each of the tested parameters as the upper bound: [0 0.01 0.05 0.1]. The same separation into training and test subsets was used for all 16 simulated datasets. The training and testing was performed on the subsets with matching parameters. To assess the effect of the mismatched training and testing parameters, datasets with fixed noise level (0.01) and four different exchange flux magnitude values, or datasets with fixed exchange flux magnitude (1) and four different noise levels were used. The interquantile range distributions of the testing datasets were compared using right-tailed Wilcoxon-Mann-Whitney rank sum test (`ranksum` function in MATLAB).

¹³C Metabolic flux analysis (MFA). ¹³C-MFA was performed with INCA software⁹ in MATLAB 2013a (MathWorks Inc). The measured labeling data and glucose uptake rates¹⁵ were used as inputs to the model. Three sets of data were used to constrain the

model: the data from [1-¹³C] experiment only, the data from [U-¹³C] experiment only, and the combined dataset from both experiments. Each experimental strain was analyzed separately. Best fit flux solution and 95% flux confidence intervals were calculated with the parameter continuation procedure ('continue'). The [10% 90%] quantiles of the flux ratio distributions were estimated from the 1000 solutions found with the optimization procedure ('estimate') for each experimental strain.

Experimental data

Experimental data for *E. coli* and *B. subtilis* central carbon metabolism studies were downloaded from the supplementary materials available for the corresponding papers^{15,16}.

Code availability

MATLAB code for SUMOFLUX and example scripts are available at <http://www.imsb.ethz.ch/research/zamboni/resources.html>. All scripts are compatible with MATLAB 2013a (MathWorks Inc).

REFERENCES

- 1 Fendt, S.-M. *et al.* Reductive glutamine metabolism is a function of the α -ketoglutarate to citrate ratio in cells. *Nature communications* **4**, 2236-2236, doi:10.1038/ncomms3236 (2013).
- 2 Guzmán, S. *et al.* ¹³C metabolic flux analysis shows that resistin impairs the metabolic response to insulin in L6E9 myotubes. *BMC systems biology*, 1-12, doi:10.1186/s12918-014-0109-z (2014).
- 3 Yang, C. *et al.* Glutamine oxidation maintains the TCA cycle and cell survival during impaired mitochondrial pyruvate transport. *Molecular Cell* **56**, 414-424, doi:10.1016/j.molcel.2014.09.025 (2014).
- 4 Colijn, C. *et al.* Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS computational biology* **5**, e1000489-e1000489, doi:10.1371/journal.pcbi.1000489 (2009).
- 5 Wiechert, W., Möllney, M., Petersen, S. & de Graaf, A. A. A universal framework for ¹³C metabolic flux analysis. *Metabolic engineering* **3**, 265-283, doi:10.1006/mben.2001.0188 (2001).

- 6 Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metabolic engineering* **9**, 68-86, doi:10.1016/j.ymben.2006.09.001 (2007).
- 7 Quek, L.-E., Wittmann, C., Nielsen, L. K. & Krömer, J. O. OpenFLUX: efficient modelling software for ^{13}C -based metabolic flux analysis. *Microbial cell factories* **8**, 25-25, doi:10.1186/1475-2859-8-25 (2009).
- 8 Zamboni, N., Fendt, S. M., Rühl, M. & Sauer, U. ^{13}C -based metabolic flux analysis. *Nat Protoc* **4**, 878-892, doi:10.1038/nprot.2009.58 (2009).
- 9 Young, J. D. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* **30**, 1333-1335, doi:10.1093/bioinformatics/btu015 (2014).
- 10 Weitzel, M. *et al.* 13CFLUX2 - high-performance software suite for ^{13}C -metabolic flux analysis. *Bioinformatics (Oxford, England)* **29**, 143-145, doi:10.1093/bioinformatics/bts646 (2013).
- 11 Mancuso, a., Sharfstein, S. T., Tucker, S. N., Clark, D. S. & Blanch, H. W. Examination of primary metabolic pathways in a murine hybridoma with carbon-13 nuclear magnetic resonance spectroscopy. *Biotechnology and Bioengineering* **44**, 563-585, doi:10.1002/bit.260440504 (1994).
- 12 Sauer, U. *et al.* Metabolic flux ratio analysis of genetic and environmental modulations of *Escherichia coli* central carbon metabolism. *Journal of bacteriology* **181**, 6679-6688, doi:PMC94132 (1999).
- 13 Szyperski, T. *et al.* Bioreaction network topology and metabolic flux ratio analysis by biosynthetic fractional ^{13}C labeling and two-dimensional NMR spectroscopy. *Metabolic engineering* **1**, 189-197, doi:10.1006/mben.1999.0116 (1999).
- 14 Emmerling, M. *et al.* Metabolic flux responses to pyruvate kinase knockout in *Escherichia coli*. *Journal of bacteriology* **184**, 152-164, doi:10.1128/JB.184.1.152 (2002).
- 15 Fischer, E. & Sauer, U. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *European Journal of Biochemistry* **270**, 880-891, doi:10.1046/j.1432-1033.2003.03448.x (2003).
- 16 Fischer, E. & Sauer, U. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nature genetics* **37**, 636-640, doi:10.1038/ng1555 (2005).
- 17 Zamboni, N., Fischer, E. & Sauer, U. FiatFlux - a software for metabolic flux analysis from ^{13}C -glucose experiments. *BMC bioinformatics* **6**, 209-209, doi:10.1186/1471-2105-6-209 (2005).
- 18 Rantanen, A. *et al.* An analytic and systematic framework for estimating metabolic flux ratios from ^{13}C tracer experiments. *BMC bioinformatics* **9**, 266-266, doi:10.1186/1471-2105-9-266 (2008).
- 19 Breiman, L. Random forests. *Machine learning* **45**, 5-32, doi:10.1023/A:1010933404324 (2001).
- 20 Wiechert, W., Möllney, M., Isermann, N., Wurzel, M. & de Graaf, A. A. Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems. *Biotechnology and bioengineering* **66**, 69-85 (1999).
- 21 Meinshausen, N. Quantile Regression Forests. *Journal of Machine Learning Research* **7**, 983-999, doi:10.1111/j.1541-0420.2010.01521.x (2006).

- 22 Buescher, J. M. *et al.* A roadmap for interpreting ^{13}C metabolite labeling patterns from cells. *Curr Opin Biotechnol* **34**, 189-201, doi:10.1016/j.copbio.2015.02.003 (2015).
- 23 Möllney, M., Wiechert, W., Kownatzki, D. & de Graaf, A. A. Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments. *Biotechnology and bioengineering* **66**, 86-103 (1999).
- 24 Crown, S. B., Ahn, W. S. & Antoniewicz, M. R. Rational design of ^{13}C -labeling experiments for metabolic flux analysis in mammalian cells. *BMC systems biology* **6**, 43-43, doi:10.1186/1752-0509-6-43 (2012).
- 25 Millard, P., Sokol, S., Letisse, F. & Portais, J. C. IsoDesign: A software for optimizing the design of ^{13}C -metabolic flux analysis experiments. *Biotechnology and Bioengineering* **111**, 202-208, doi:10.1002/bit.24997 (2014).
- 26 Hua, Q., Yang, C., Baba, T., Mori, H. & Shimizu, K. Responses of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. *Journal of bacteriology* **185**, 7053-7067, doi:10.1128/JB.185.24.7053 (2003).
- 27 Toya, Y. *et al.* ^{13}C -metabolic flux analysis for batch culture of *Escherichia coli* and its Pyk and Pgi gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. *Biotechnology progress* **26**, 975-992, doi:10.1002/btpr.420 (2010).
- 28 Rühl, M. *et al.* Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of ^{13}C metabolic flux analysis. *Biotechnology and bioengineering* **109**, 763-771, doi:10.1002/bit.24344 (2012).
- 29 Zamboni, N. & Sauer, U. Novel biological insights through metabolomics and ^{13}C -flux analysis. *Current opinion in microbiology* **12**, 553-558, doi:10.1016/j.mib.2009.08.003 (2009).
- 30 Zamboni, N. ^{13}C metabolic flux analysis in complex systems. *Current opinion in biotechnology* **22**, 103-108, doi:10.1016/j.copbio.2010.08.009 (2011).
- 31 Leighty, R. W. & Antoniewicz, M. R. COMPLETE-MFA: Complementary parallel labeling experiments technique for metabolic flux analysis. *Metabolic Engineering* **20**, 49-55, doi:10.1016/j.ymben.2013.08.006 (2013).
- 32 Woo Suk, A. & Antoniewicz, M. R. Parallel labeling experiments with [1,2- ^{13}C]glucose and [U- ^{13}C]glutamine provide new insights into CHO cell metabolism. *Metabolic Engineering* **15**, 34-47, doi:10.1016/j.ymben.2012.10.001 (2013).
- 33 Gopalakrishnan, S. & Maranas, C. D. ^{13}C Metabolic flux analysis at a genome-scale. *Metabolic Engineering*, doi:10.1016/j.ymben.2015.08.006 (2015).
- 34 García Martín, H. *et al.* A method to constrain genome-scale models with ^{13}C labeling data. *PLOS Computational Biology* **11**, e1004363-e1004363, doi:10.1371/journal.pcbi.1004363 (2015).
- 35 Nöh, K., Wahl, A. & Wiechert, W. Computational tools for isotopically instationary ^{13}C labeling experiments under metabolic steady state conditions. *Metabolic Engineering* **8**, 554-577, doi:10.1016/j.ymben.2006.05.006 (2006).
- 36 Young, J. D., Walther, J. L., Antoniewicz, M. R. & Yoo, H. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnology* **99**, 686-699, doi:10.1002/bit (2008).
- 37 Hörl, M., Schnidder, J., Sauer, U. & Zamboni, N. Non-stationary ^{13}C -metabolic flux ratio analysis. *Biotechnology and bioengineering* **9999**, 1-13, doi:10.1002/bit.25004 (2013).

- 38 Wiechert, W. & de Graaf, A. A. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnology and bioengineering* **55**, 101-117, doi:10.1002/(SICI)1097-0290(19970705)55:1<101::AID-BIT12>3.0.CO;2-P (1997).

Chapter 3

Targeted ^{13}C flux ratio analysis reveals different patterns of amino acid co-utilization in mycobacteria

Maria Kogadeeva, Michael Zimmermann, Franziska Küng, Eva-Maria Link, Martin Osswald, Uwe Sauer, Nicola Zamboni

Contributions by MK:

Designed and performed experiments, analyzed the data, wrote the manuscript

Supplementary Information:

Available in Appendix II on page 159 et sqq.

ABSTRACT

The quest for finding the Achilles' heel of pathogenic mycobacteria to be exploited therapeutically continues after decades of efforts. Accumulated evidence suggests that during infection, mycobacteria have access to a variety of nutrients, among which amino acids play an important role. Understanding the metabolic adjustments necessary for amino acids utilization could provide targets for drug discovery. To characterize amino acid utilization in mycobacteria, we use ^{13}C tracing experiments and analyze metabolic fluxes in *Mycobacterium smegmatis* growing in glucose and amino acid containing media, and in *Mycobacterium tuberculosis* infecting macrophage-like THP-1 cells. Since global ^{13}C metabolic flux analysis in composite media is restricted by the complexity of the metabolic networks and the lack of measurement data to be used as constraints, we perform targeted ^{13}C flux ratio analysis with a recently developed SUMOFLUX workflow. Due to the combination of machine learning used to build flux ratio predictors, and surrogate modelling used to simulate a comprehensive training dataset, SUMOFLUX offers the possibility to estimate any calculable flux ratio from the given labeling data even in case of poorly determined networks. With SUMOFLUX, we calculated relative fractions of amino acid biosynthesis and catabolism, and resolved key central carbon metabolism ratios in *M. smegmatis* growing in defined co-feed media. All amino acids were consumed by the bacteria and affected metabolic fluxes in central metabolism, with glutamate having the largest impact and catabolic fraction. We investigated the speed of metabolic adaptation to the glucose and glutamate co-feed with dynamic medium shift experiments, and discovered that *M. smegmatis* rearranges the fluxes in central metabolism within seconds after glutamate supplementation. In order to quantify amino acid exchange fluxes between *M. tuberculosis* and its host during infection, we applied SUMOFLUX to resolve amino acid biosynthesis fractions from the labeling patterns of protein-bound amino acids derived from bacterial and THP-1 cells. This analysis revealed that during infection, biosynthesis fraction of most amino acids decreased compared to growth in rich medium, suggesting not only their availability for bacterial uptake, but sparsity of other carbon sources for their intracellular production. Overall, our results indicate that mycobacteria can rapidly adapt to nutrient availability in their environment and consume sparsely available amino acids as soon as they become available, which might

explain the failures of targeted therapeutic interventions, and underline the need for a complex therapies aiming at several targets simultaneously.

INTRODUCTION

Mycobacteria have been investigated for decades, since the species *Mycobacterium tuberculosis* and *Mycobacterium leprae* are among the most widespread deadly human pathogens threatening global health. Laboratory studies on these bacteria are hampered by the fact that bacteria reside in various states (acute or chronic infections), grow at slow rates, and are difficult to culture since humans are their only natural host. Despite advances of mycobacterial research, we are still far from understanding the *in vivo* host-pathogen interactions, necessary to control and defeat the disease¹⁻⁵.

Upon infection, *M. tuberculosis* resides in alveolar macrophages, where it is exposed to various host defense mechanisms, like an acidic environment and sparse nutrient supply^{3,6-8}. *M. tuberculosis* can autonomously produce most of the essential metabolites, which is believed to be one of its virulence strategies⁹. The role of metabolic pathways and substrates has been largely studied by infection experiments with auxotrophic mutants⁶. These studies highlighted the relevance of gluconeogenic carbon sources inside the host¹⁰⁻¹⁴, among those cholesterol being the major one^{15,16}. Later studies revealed the concomitant essentiality of glycolysis¹⁷. Furthermore, there is growing evidence that apart from carbohydrates, mycobacteria have access to amino acids derived from the host¹⁸⁻²⁰. Metabolomics experiments focusing on carbon co-utilization in *M. tuberculosis*²¹ and proteomics analysis of central carbon metabolism of its non-pathogenic relative *Mycobacterium smegmatis*²² revealed the ability of mycobacteria to adjust to simultaneous utilization of multiple carbon sources, which might be the key to its virulence success.

Despite accumulating knowledge on mycobacterial ability to co-metabolize carbohydrates and amino acids, there is a lack of quantitative understanding of the underlying metabolic adaptations. In this work, we set out to quantify metabolic fluxes in *M. smegmatis*, which shares many metabolic traits with its pathogenic relatives^{23,24}, during co-utilization of glucose and amino acids. The most direct experimental setup to quantify metabolic fluxes is ¹³C labeling, in which the label propagation from the substrate to the metabolic intermediates is detected by mass spectrometry or nuclear

magnetic resonance spectroscopy. With the classical ^{13}C metabolic flux analysis (^{13}C -MFA), the global flux distribution through the network is fitted to the labeling data, constrained by measured substrate uptake, secretion and growth rates. In a composite media, however, application of ^{13}C -MFA becomes challenging due to several reasons. First, with the number of substrates, the degrees of freedom of the network increase, thus more measurements are required to determine fluxes. Second, there is no guarantee that the labeling patterns contain sufficient information to estimate the underlying fluxes. For example, if one substrate is labeled and the other is not, and they fuel different parts of the metabolic network, most of the detected metabolites will be either fully labeled or unlabeled, and only the intermediates at the convolution point of the substrates' catabolic pathways will have distinct labeling patterns. Therefore, it would be necessary to perform several labeling experiments in parallel, with different tracers targeting elucidation of different parts of the metabolic network²⁵. Third, ^{13}C -MFA requires a closed carbon balance provided by additional measurements of physiological parameters, which are not always feasible.

Due to these challenges, the application of ^{13}C -MFA for flux estimation in composite media provides in most cases only qualitative information. For example, when multiple equally plausible flux solutions exist²⁶, or fluxes can be accurately fitted in a subnetwork of the studied metabolic network²⁷, the outcome of the flux analysis can be interpreted only in a conditional and comparative manner. In the to date most comprehensive ^{13}C labeling experiment with *M. tuberculosis* infecting macrophage-like THP-1 cells, ^{13}C flux spectral analysis based on ^{13}C -MFA was applied to investigate bacterial amino acid consumption during early infection²⁸. To predict which substrates *M. tuberculosis* consumed from the host, thousands of flux solutions generated for metabolic networks with different substrate configurations were fitted to the labeling patterns of host- and pathogen-derived amino acids. Despite extensive calculations, the outcome of the flux spectrum scanning was qualitative, stating that the model that included alanine, serine and acetate as substrates provided the best fit. Manual inspection of the labeling data was required to expand the potential substrate list to glutamate and aspartate²⁸. In several other co-metabolism studies, the labeling patterns of metabolic intermediates were solely manually compared to infer the contribution of each substrate to their formation^{21,29}.

Given the complexity and uncertainty of active metabolic networks in composite media, it might be beneficial to choose a local ^{13}C flux analysis method to

quantify specific nodes in the network rather than all fluxes at once²⁵. For example, the targeted ¹³C flux ratio analysis method SUMOFLUX (Chapter 2) has the potential to tackle the challenges posed by complex media conditions. First, SUMOFLUX exploits machine learning to build flux ratio predictors given the metabolic network, substrate label and available measurements. The predictors are trained on a large dataset simulated with surrogate modelling that consists of thousands of stoichiometrically feasible flux distributions and corresponding metabolite labeling patterns. Therefore, in case of an underdetermined network, a broad range of possible fluxes can be generated in the simulation procedure and further used to build a generalized flux ratio predictor. Second, the predictor testing procedure embedded in SUMOFLUX enables rapid evaluation of flux ratio resolvability given the available data and experimental setup. Therefore, even when only sparse measurements are provided, SUMOFLUX can be used to test which flux ratios can be estimated, in order to extract the most information from the available data. Third, due to the surrogate modelling, SUMOFLUX does not require additional measurements apart from the metabolite labeling patterns, since different configurations of physiological parameters can be simulated in the training dataset.

To investigate mycobacterial adaptations to co-metabolism of carbohydrates and amino acids, we analyzed metabolic fluxes of *M. smegmatis* grown in 20 defined two-substrate media comprised of [U-¹³C] glucose and a single unlabeled amino acid. We exploited SUMOFLUX to quantify specified biosynthesis and catabolism fractions along with several ratios describing central carbon metabolism for each amino acid condition. The ¹³C flux ratio analysis allowed us to compare metabolic rearrangements caused by different amino acids and rank them according to their impact on central metabolism. We discovered that glutamate is the most utilized amino acid, and followed up this condition by quantifying the fluxes with ¹³C-MFA from parallel labeling experiments, which confirmed global flux rearrangements in central metabolism. Additionally, we probed the speed of these rearrangements by following the label propagation in dynamic ¹³C experiments.

Since SUMOFLUX provided quantitative estimates of the flux ratios upon co-metabolism of glucose and amino acids in *M. smegmatis*, we decided to apply it to quantify flux ratios in publicly available dataset of *M. tuberculosis* growing in rich medium and infecting macrophage-like THP-1 cells²⁸. Comparison of our flux ratio estimates upon growth in defined media revealed that *M. smegmatis* and

M. tuberculosis have similar amino acid utilization patterns, with glutamate and histidine having a high intracellular turnover. Finally, we quantified amino acid biosynthesis and catabolism fractions of *M. tuberculosis* in the early infection setup. Inside the macrophage, both fractions of most amino acids decreased, indicating, on the one hand, sparse availability of nutrients hardly sufficient for catabolic activity, and, on the other hand, mycobacterial simultaneous uptake of several amino acids during infection. Our targeted ^{13}C metabolic flux ratio analysis with SUMOFLUX provides quantitative information on complex inter-species interactions, and elucidates the plasticity of mycobacterial amino acid utilization, which may partially explain the recurrent failures of the single target treatment strategies.

RESULTS

***M. smegmatis* has different patterns of amino acids co-utilization with glucose.** To investigate the ability of mycobacteria to co-utilize amino acids with glucose, we grew *M. smegmatis* in minimal media containing one out of the 20 amino acids (in concentration of 2 g L^{-1}) with $[\text{U-}^{13}\text{C}]$ glucose (2 g L^{-1}) (Figure 1a). First, to assess whether amino acid supplementation benefited bacterial growth, we measured bacterial culture density in microtiter plates over a period of 80 h and compared in to growth on glucose alone. In all 20 co-feed conditions, *M. smegmatis* grew faster and/or achieved a higher culture density compared to the growth in glucose media (Figures 1b and S1, Table S1), indicating that each amino acid was consumed and utilized by the bacteria.

Next, to assess the metabolic impact of the substrate amino acid in each condition, we measured steady state ^{13}C labeling patterns of intracellular protein-bound amino acids with gas chromatography – mass spectrometry (GC-MS) in the bacterial cultures at $\text{OD}_{600} \sim 1$. The fractional labeling of the intracellular amino acids reflects the carbon propagation from the carbon sources present in the media through the bacterial metabolic network. In all tested conditions, solely two carbon sources were provided: the fully labeled $[\text{U-}^{13}\text{C}]$ glucose and the unlabeled amino acid. Hence, the lower the intracellular fractional labeling of the measured amino acid, the more carbons were derived from the unlabeled amino acid consumed from the media, and vice versa.

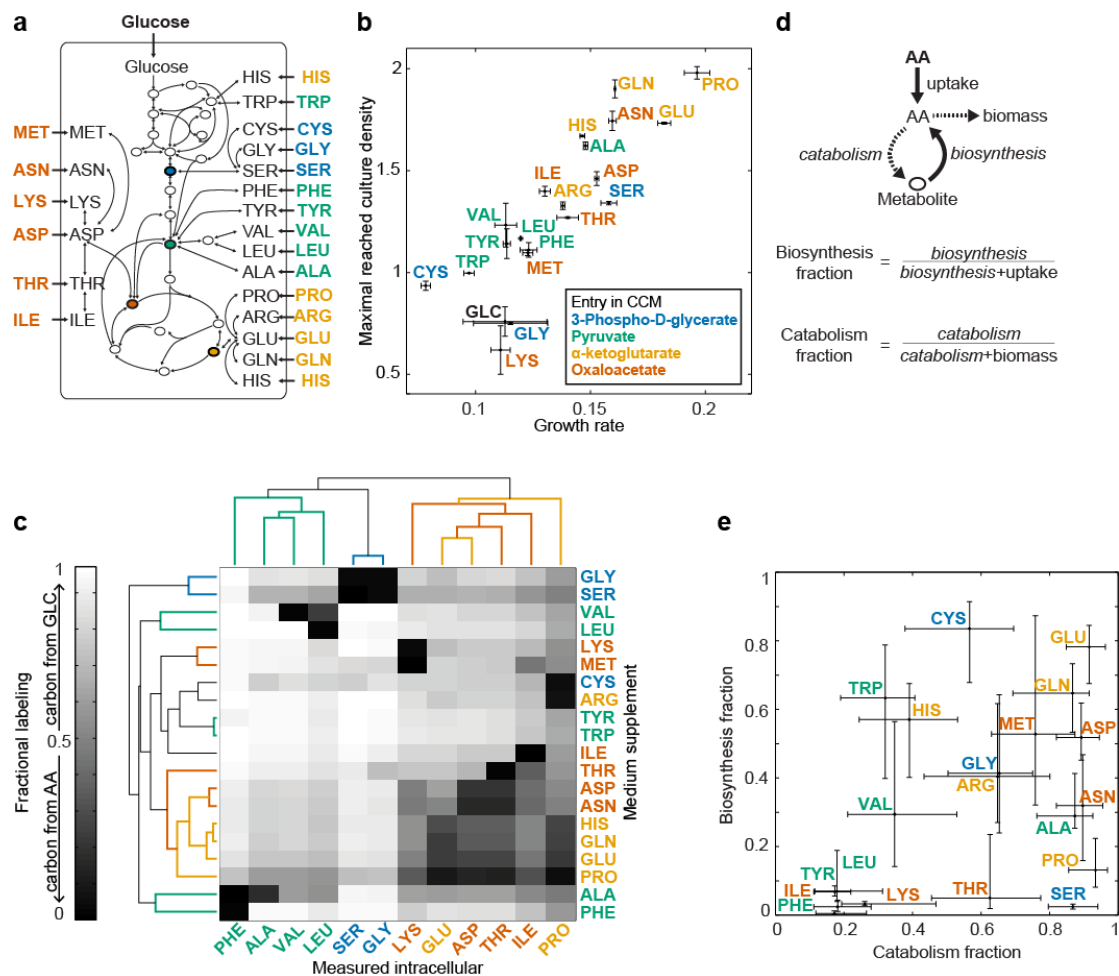


Figure 1. Different effects of amino acid co-metabolism in *M. smegmatis*. (a) Schematic representation of the central carbon metabolism network of *M. smegmatis* with entry points of the 20 amino acids. (b) Growth rate and maximal culture density of *M. smegmatis* grown on combinations of glucose and 20 single amino acids in 96 well plate. (c) Hierarchical clustering of the amino acid fractional labeling measured in *M. smegmatis* grown on combinations of [U-¹³C] glucose and 20 naturally labeled single amino acids. (d) Schematic definition of the biosynthesis and catabolism fraction of the amino acids. (e) SUMOFLUX estimates of biosynthesis and catabolism fractions for each amino acid in the corresponding medium containing [U-¹³C] glucose and a naturally labeled amino acid. Error bars indicate 50% prediction intervals. AA – amino acid; CCM – central carbon metabolism, GLC – glucose.

Hierarchical clustering of the fractional labeling measured for 12 amino acids displayed an association between metabolic proximity of amino acids in central metabolism (Figure 1a) and their mode of utilization (Figure 1c). For example, amino acids synthesized from the tricarboxylic acid (TCA) cycle intermediates α -ketoglutarate and oxaloacetate had the most influence on the fractional labeling of the other amino acids. On the other hand, amino acids synthesized from pyruvate were almost exclusively incorporated directly from the media, and only slightly influenced the fractional labeling of the other measured amino acids.

To obtain quantitative estimates of amino acid utilization, we exploited the targeted ^{13}C flux ratio analysis method SUMOFLUX (Chapter 2). Since SUMOFLUX can be used to predict any flux ratio that can be calculated from the network fluxes, we defined two ratios of interest characterizing amino acid utilization: the fraction of intracellular biosynthesis versus uptake from the medium; and the fraction of amino acid catabolism versus the direct use in protein synthesis (Figure 1d, Table S2). When an amino acid is added to the medium, its intracellular biosynthesis fraction will depend on the measured labeling in the corresponding intracellular amino acid, whereas its catabolism fraction will depend on the measured labeling in all the other intracellular amino acids. For each of the 20 amino acid co-feed media and for the glucose medium we built a metabolic model (Figure 1a, Table S3), simulated the training and testing datasets, trained the predictors and tested their performance *in silico* (see Materials and Methods or Chapter 2 for details). In general, all predictors provided accurate estimates of the corresponding biosynthesis or catabolism ratios (mean absolute error, $\text{MAE} < 0.1$) (Figures S2 and S3). Only four predictors had a lower accuracy, the biosynthesis and catabolism predictors for glutamine ($\text{MAE} = 0.11$) and for arginine ($\text{MAE} = 0.16$), which was expected for amino acids that were not measured with GC-MS.

SUMOFLUX estimates of biosynthesis and catabolism fractions for the experimental data complement the visual clustering of amino acid utilization patterns (Figure 1e). Indeed, relative catabolism fractions of amino acids produced in the TCA cycle were the highest, with more than 0.8 of the consumed amino acid being catabolized. With exception of alanine, which was highly catabolized, amino acids produced from pyruvate contributed less than 0.4 to catabolism. Serine was highly catabolized with very low intracellular turnover, since its intracellular biosynthesis fraction was less than 0.1 (Figure 1e).

Formulating the two flux ratios, intracellular biosynthesis and catabolism, for each amino acid in the network, allowed us to quantify their metabolic fate. Although all 20 amino acids were taken up and used to sustain bacterial growth, their utilization patterns significantly varied. With some exceptions, amino acids entering the central metabolism network at the same metabolic node had similar utilization patterns. Biosynthetically costly amino acids were diverted mostly to biomass and less to catabolism, reflecting a trend to spare energy equivalents³⁰ derived from glucose for other processes.

Central metabolic fluxes are rearranged upon availability of amino acids. Since all 20 amino acids were not only taken up by *M. smegmatis* to be used for protein synthesis, but were further catabolized, we decided to investigate their catabolic paths and corresponding adjustment of central metabolic fluxes. Using the labeling data obtained in the co-feed experiments, we applied SUMOFLUX to estimate the flux ratios at the “crossroads” of major catabolic fluxes in central metabolism (Table S2). We focused on the three important metabolic nodes – phosphoenolpyruvate (PEP), pyruvate and oxaloacetate, which are responsible for the distribution of the carbon flux among catabolism, anabolism and energy supply of the cells³¹.

First, we investigated the metabolic source of PEP, an important metabolite involved in glycolysis, gluconeogenesis and glucose uptake by the phosphotransferase system. Upon growth on glucose, PEP is derived from glycolysis in a reaction catalyzed by enolase (Eno), whereas upon growth on fatty acids or other gluconeogenic carbon sources, PEP is produced from oxaloacetate in a reaction catalyzed by PEP carboxykinase (Pck) (Figure 2a). PEP carboxykinase was reported to be essential in pathogenic mycobacteria during infection¹³, indicating a prevalence of gluconeogenic carbon sources inside the macrophage. With SUMOFLUX, we built two predictors for the flux fractions producing PEP from glycolysis (Eno) and gluconeogenesis (Pck). The accuracy of these predictors *in silico* was sufficiently good for a comparative analysis (MAE ~ 0.1, Figure S5). We found that upon growth on glucose, the relative contribution of glycolysis to the formation of PEP was 0.8 (Figure 2b), whereas upon addition of amino acids to the media the relative fraction of gluconeogenesis increased. The largest increase to more than 0.3 of the relative gluconeogenic flux fraction was caused by the addition of amino acids entering the central metabolism through the TCA cycle.

Second, we resolved the relative contributions of fluxes producing pyruvate, the main precursor for fatty acid biosynthesis, which is particularly important for mycobacteria³². We resolved the fractions of the glycolytic flux from PEP catalyzed by the pyruvate kinase (Pyk), which has an essential role during carbon co-catabolism, preventing accumulation of metabolic intermediates involved in metabolic regulation²⁰, and malic enzyme (Mae), which is usually dispensable on glucose. The *in silico* accuracy of the Pyk and Mae fraction predictors was sufficient for comparative analysis (MAE ~ 0.1, Figure S5). Similar to the PEP case, we observed that upon growth on glucose, pyruvate was produced mostly through glycolysis with Pyk contribution of

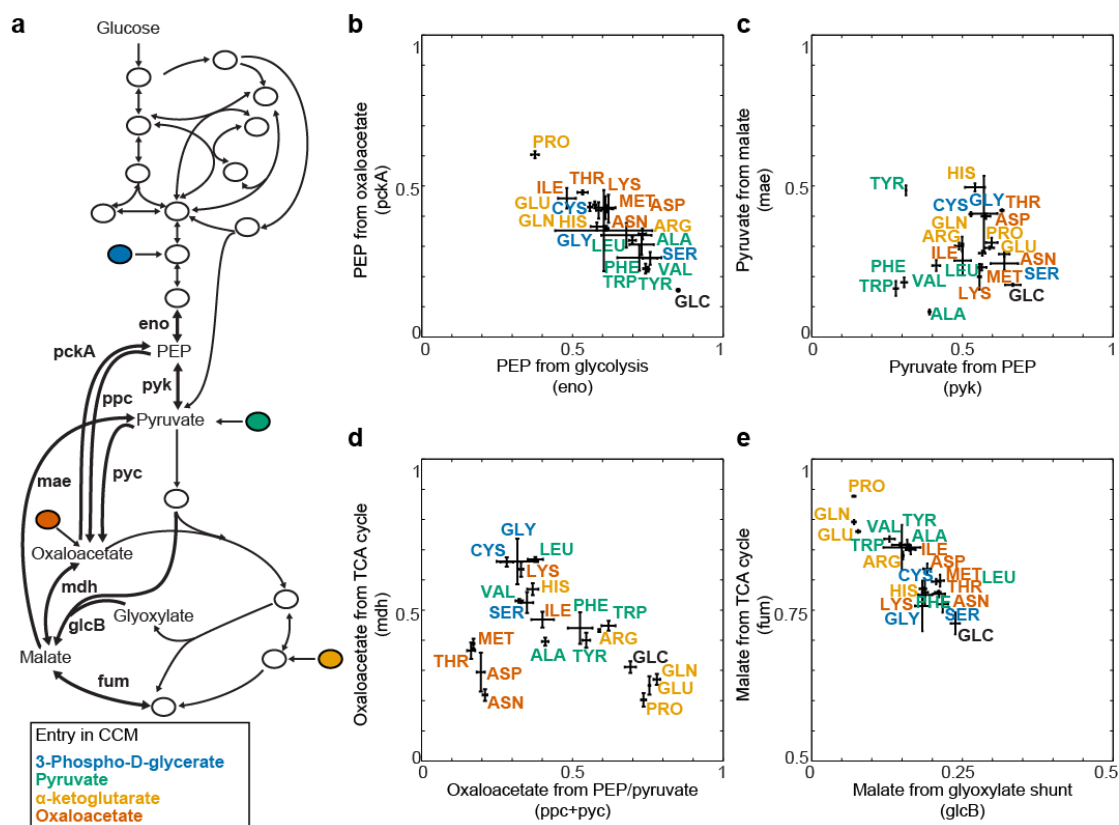


Figure 2. Amino acid co-metabolism in *M. smegmatis* leads to flux rearrangements in central carbon metabolism. (a) Schematic representation of the central carbon metabolic network of *M. smegmatis* with four amino acid entry points. (b) SUMOFLUX estimates for the fraction of Eno and PckA fluxes contributing to the PEP pool. (c) SUMOFLUX estimates for the fraction of Pyk and Mae fluxes contributing to the pyruvate pool. (d) SUMOFLUX estimates for the fraction of Ppc+Pyc and Mdh fluxes contributing to the oxaloacetate pool. (e) SUMOFLUX estimates for the fraction of GlcB and Fum fluxes contributing to the malate pool. All fractions were estimated from [U-¹³C] glucose and corresponding naturally labeled amino acid experiments. Error bars represent standard deviation of the estimates for four biological replicates. PEP – phosphoenolpyruvate.

more than 0.7, whereas the Mae fraction was less than 0.2. The Entner-Doudoroff pathway might have a residual contribution to the pyruvate pool (Figure 2a), however, it cannot be resolved with the applied labeling strategy. Upon addition of amino acids entering the central metabolism at pyruvate node, such as alanine, tryptophan and valine, both the relative contributions of Pyk and Mae decreased, implying that these amino acids partially replenished the pyruvate pool. The relative contribution of the gluconeogenic flux from malic enzyme increased upon addition of amino acids entering the TCA cycle (Figure 2c).

Third, we focused on the production of oxaloacetate at the crossroads of the TCA cycle, glyoxylate shunt, urea cycle, gluconeogenesis, fatty acid and amino acid biosynthesis. Oxaloacetate is mainly produced in the TCA cycle and consumed for the

biosynthesis of amino acids, therefore, cells have to replenish its pool through anaplerotic fluxes from PEP/pyruvate or the glyoxylate shunt (Figure 2a). With SUMOFLUX, we resolved two flux fractions contributing to the production of oxaloacetate: the anaplerotic flux fraction from PEP/pyruvate and the TCA cycle fraction coming from malate. These flux fractions were resolved with high accuracy *in silico* (MAE < 0.1 for all amino acid media, Figures S4 and S5). We found that anaplerotic fluxes undergo severe rearrangements upon addition of various amino acids (Figure 2d). Indeed, upon growth on glucose the relative contribution of anaplerosis was 0.3, which decreased upon addition of glutamate, glutamine and proline entering the TCA cycle at α -ketoglutarate and replenishing the metabolites used for biomass production (Figure 2d). On the contrary, addition of amino acids catabolized at the 3-phosphoglycerate and pyruvate nodes caused increase in the relative fraction of the flux from PEP/pyruvate. In analogy to the pyruvate case, amino acids catabolized through oxaloacetate directly replenished its pool (Figure 2d), causing the decrease of relative contributions of both anaplerosis and TCA cycle fluxes.

Compared to *Escherichia coli* and *Bacillus subtilis*, in which anaplerosis from PEP/pyruvate is the major source of oxaloacetate upon growth on glucose^{33,34}, we estimated surprisingly low relative contribution of 0.3 in *M. smegmatis*. It has been shown that in slow growing *E. coli* the glyoxylate shunt takes over the anaplerotic function³⁵. To test the hypothesis that the slow growing *M. smegmatis* undergoes similar flux adaptations with increased relative flux contribution of glyoxylate shunt, we estimated its fractional contribution to malate (Figure 2a). The accuracy of the glyoxylate shunt fraction predictors was good (MAE < 0.1, Figure S5). On glucose, the relative contribution of the glyoxylate shunt was ~ 0.25, and it decreased upon addition of any of the amino acids (Figure 2e). It can be explained either by the amino acids taking over the anaplerotic function by fueling metabolism either from the TCA cycle or from 3-phosphoglycerate and pyruvate, or by metabolic regulation taking place as the growth rate increased.

In summary, SUMOFLUX provided estimates for the major flux ratios in the TCA cycle and PEP/pyruvate nodes in *M. smegmatis* upon growth in 20 glucose and amino acid co-feed media. Apart from replenishment of the metabolite pool at their entry points in central carbon metabolism, most amino acids were propagated further resulting in more distant flux adjustments. Amino acids entering central metabolism through the TCA cycle had the largest influence on the resolved central metabolic flux

ratios. Upon growth on glucose, the relative contribution of glyoxylate shunt to the formation of malate was 0.25, and it decreased upon addition of amino acids. The glyoxylate shunt is required in bacteria upon growth on acetate or fatty acids, or under growth limiting conditions. In the case of *M. smegmatis*, the relatively high glyoxylate shunt activity might be explained by its slow growth, or by the need to quickly adapt to alternative carbon sources as soon as they become available.

Global metabolic flux rearrangements during co-utilization of glucose and glutamate. According to the SUMOFLUX estimates, glutamate is the most catabolized amino acid (Figure 1e), and belongs to the group of amino acids causing major flux rearrangements in the TCA cycle, PEP and pyruvate nodes (Figures 2b, d, e). With the data from the co-feed experiments, however, the fluxes in the upper glycolysis and pentose-phosphate (PP) pathway are unresolvable. Since glutamate is a standard major component of the commonly used 7H9 medium, we decided to follow up its co-metabolism with glucose and investigate whether its supplementation also affects fluxes in the upper glycolysis (Figure 3a).

In order to characterize fluxes in upper glycolysis, we performed two labeling experiments with a mixture of 60% [1-¹³C] and 40% [U-¹³C] glucose and a mixture of [1-¹³C] glucose and unlabeled glutamate. We used SUMOFLUX to build predictors for three flux ratios: the ratio between glycolysis (first step catalyzed by Pgi) and pentose-phosphate pathway (first step catalyzed by Zwf), the fraction of glyceraldehyde 3-phosphate produced from glycolysis, and the contribution of the Entner-Doudoroff pathway to the formation of pyruvate (Figure 3a). All three ratios could be resolved with high accuracy (MAE < 0.05, Figures S6a, b). During growth on glucose, the relative utilization of glucose through glycolysis was ~ 0.25, and it increased upon addition of glutamate to ~ 0.40 (Figure 3b). Such a low relative flux through glycolysis is surprising, since in *E. coli* or *B. subtilis*, for example, the relative glucose flux through glycolysis is ~ 0.70^{33,34}. On the other hand, the fractional contribution of glycolysis to glyceraldehyde 3-phosphate on glucose was ~ 0.60, suggesting high activity of transaldolases and transketolases in the PP pathway, routing the carbon flow back to glycolysis (Figure 3a). Upon addition of glutamate, the relative contribution of glycolysis to glyceraldehyde 3-phosphate increased to ~ 0.80. The fractional contribution of Entner-Doudoroff pathway to pyruvate upon growth on glucose was 0.1, which corresponds to the residual flux contribution calculated in the previous

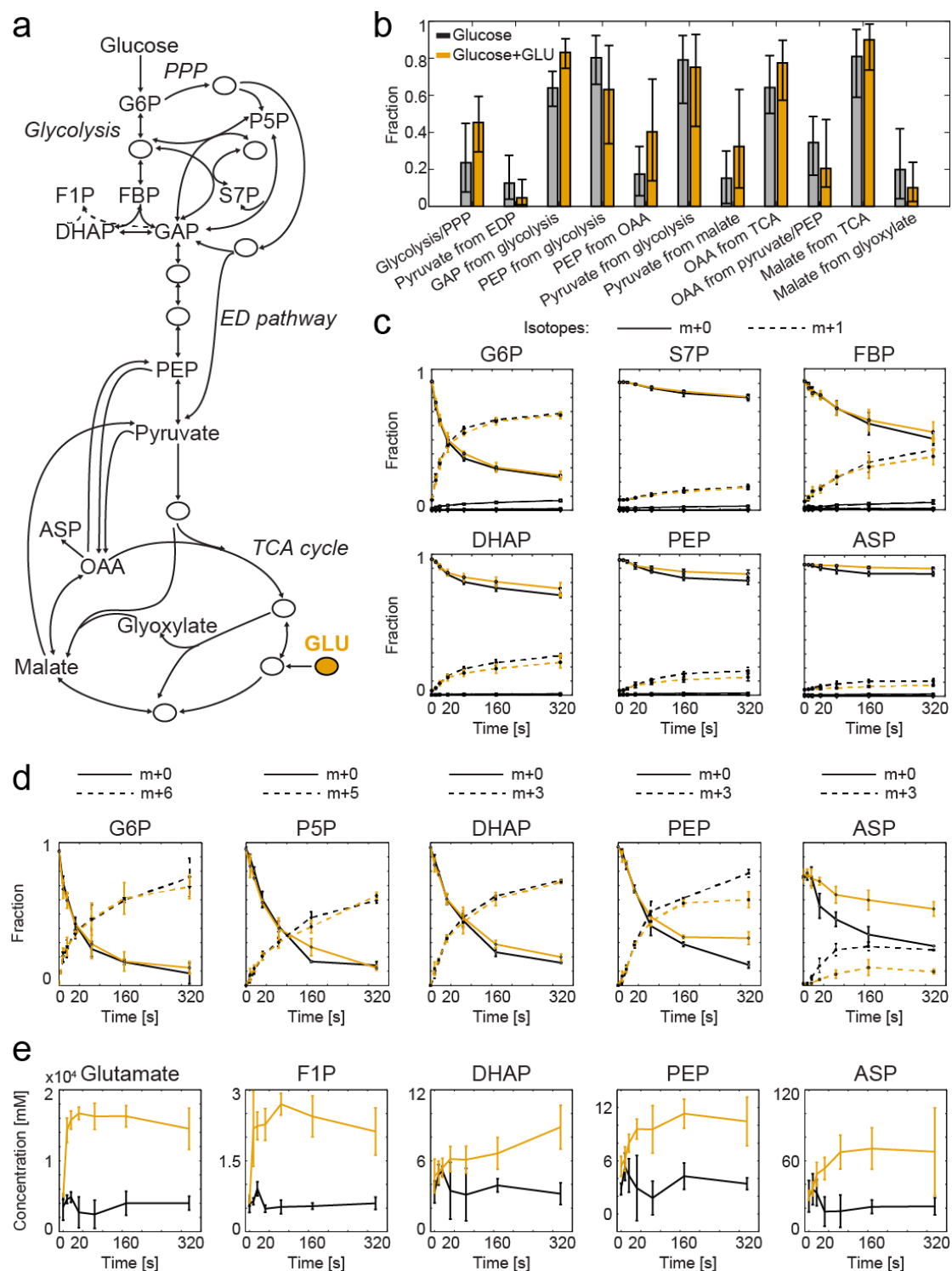


Figure 3. Flux ratio analysis and shift experiments demonstrate that glutamate added to the glucose medium is immediately taken up and affects central carbon fluxes in *M. smegmatis*. (a) Schematic representation of the central carbon metabolic network of *M. smegmatis* with major pathways. (b) SUMOFLUX estimates for the 11 key flux ratios and fractions in central metabolism. (c) Label propagation in CCM metabolites upon switching from naturally labeled glucose medium to $[1-^{13}\text{C}]$ glucose medium or the mixture of $[1-^{13}\text{C}]$ glucose and glutamate medium. (Continue on p. 88)

section (Figure 2c). In concordance with the decreased relative fraction of PP pathway upon addition of glutamate, the relative fraction of Entner-Doudoroff pathway to the formation of pyruvate also decreased (Figure 3b).

However, one should interpret the relative flux contribution changes with caution, as they may not always reflect the absolute flux changes. For example, while the flux fraction of the PP pathway decreased upon addition of glutamate, the absolute flux value depends on the glucose uptake, and could potentially stay the same. To examine whether glutamate substitution affects only the relative reaction rates or the absolute fluxes as well, we quantified extracellular fluxes and applied the global ^{13}C -MFA to fit the fluxes in central carbon metabolism. In the co-feed medium, glucose uptake decreased compared to the glucose medium, whereas the biomass yield on glucose was two times lower than upon growth on glucose and glutamate, implying a more efficient resource allocation during co-metabolism (Figure S7).

We estimated the absolute fluxes with ^{13}C -MFA³⁶ using the stoichiometric model of *M. smegmatis* (Table S3) and the measured glucose and glutamate uptake rates as constraints. The flux distributions were fitted to the data from the two parallel experiments with unlabeled glutamate and either [1- ^{13}C] or [U- ^{13}C] labeled glucose. The obtained flux distributions resulted in a good fit to the measured data (Figures S8-S10) and were in concordance with our flux ratio estimates (Figure 4, Tables S4 and S5). On glucose, the Zwf flux through the PP pathway was 3.5 times larger than the glycolytic Pgi flux. However, the fluxes further down the glycolysis pathway were high, confirming the estimated high fractional contribution to the glyceraldehyde 3-phosphate pool (Figures 4 and 3b). The anaplerotic flux from PEP/pyruvate was four times smaller than the TCA flux from malate, whereas the glyoxylate shunt had a high absolute flux in concordance with SUMOFLUX relative fraction estimates. In glucose plus glutamate medium, the ratio between glycolysis and PP pathway indeed increased, however, both Pgi and Zwf net fluxes decreased two and three times, respectively (Figure 4). Glutamate replenished the TCA cycle, as was predicted by the flux ratio analysis, and the flux through the glyoxylate shunt was low. Moreover, the exchange flux between intracellular glutamate and α -ketoglutarate was

(Continued from P. 87) (d) Label propagation in CCM metabolites upon switching of *M. smegmatis* from naturally labeled glucose medium to [U- ^{13}C] glucose medium or the mixture of [U- ^{13}C] glucose and glutamate medium. (e) Absolute metabolite concentration changes upon switching of *M. smegmatis* from naturally labeled glucose medium to the mixture of glucose and glutamate medium or glucose medium as a control. EDP – Entner-Doudoroff pathway; OAA – oxaloacetate; PEP – phosphoenolpyruvate; PPP – pentose phosphate pathway; TCA – tricarboxylic acid cycle.

high, with glutamate biosynthesis fraction ~ 0.86 , confirming SUMOFLUX estimates (Figure 1e).

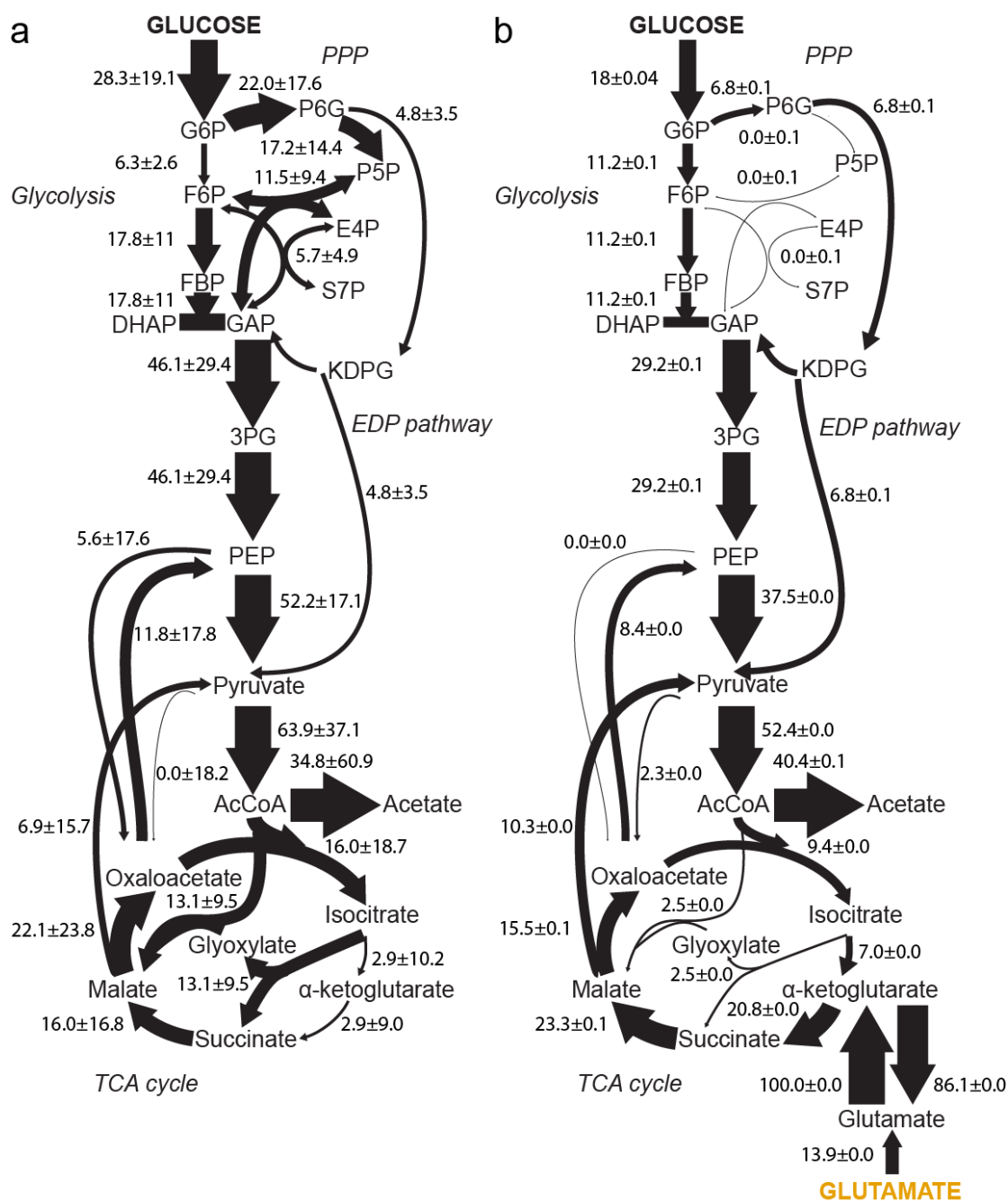


Figure 4. Flux distributions in central carbon metabolism of *M. smegmatis* grown in glucose or glucose plus glutamate medium. (a) Schematic representation of the global flux solution through the CCM network of *M. smegmatis* grown on combination of 60% [1- ^{13}C] and 40% [U- ^{13}C] glucose, obtained with ^{13}C -MFA (INCA software). (b) Schematic representation of the global flux solution through the CCM network of *M. smegmatis* grown on combination of labeled glucose and naturally labeled glutamate. The flux solution was obtained with ^{13}C -MFA (INCA software) using data from two labeling experiments, [1- ^{13}C] glucose and glutamate and [U- ^{13}C] glucose and glutamate. The thickness of flux lines is scaled to the glucose uptake vector in (a). CCM – central carbon metabolism, ^{13}C -MFA – ^{13}C metabolic flux analysis. EDP – Entner-Doudoroff pathway, PPP – pentose phosphate pathway, TCA – tricarboxylic acid cycle.

The absolute flux analysis confirmed substantial changes in central metabolism upon addition of glutamate. Since the glucose uptake decreased, both glycolytic and PP pathway fluxes decreased, despite the increase of the relative fraction of the PP pathway. Upon growth on glucose, the estimated flux distribution resembled the pyruvate dissimilation route characterized by flux through the glyoxylate shunt and anaplerotic reactions from pyruvate²⁶, presumably allowing bacteria to rapidly switch to growth on lipids, or even be able to incorporate substantial amounts of carbon dioxide. Reduced activity of the NADPH-producing TCA cycle enzymes might be the cause of the observed relatively high activity of the PP pathway, since its first reaction also produces NADPH. However, upon substitution of glutamate, the absolute flux through PP pathway and TCA cycle NADPH-producing reactions was low. Presumably, *M. smegmatis* exploits alternative cofactor balancing mechanisms. The high exchange flux between glutamate and α -ketoglutarate predicted by both SUMOFLUX and ¹³C-MFA might be potentially involved in cofactor balancing, since *M. smegmatis* possesses NADP⁺ dependent glutamate dehydrogenase Gdh1 (MSMEG_4699), and NAD⁺ dependent glutamate dehydrogenase Gdh2 (MSMEG_5442), which convert glutamate to α -ketoglutarate and vice versa.

***M. smegmatis* adapts to utilization of glutamate as soon as it becomes available.**

The steady state flux distributions in central carbon metabolism changed drastically upon addition of glutamate to the medium. The changes were observed not only locally, around the glutamate entry point in the TCA cycle, but also in the distant fluxes of upper glycolysis. Next, we set out to investigate how fast *M. smegmatis* adapts to the new environment. The adjustment speed provides further hypotheses on the regulatory mechanisms: slow adjustment within minutes or hours implies transcriptional regulation, since it involves *de novo* synthesis of enzymes; whereas fast adjustment on a second scale suggests post-translational regulation mechanisms such as protein modification or allosteric regulation.

To study the speed of metabolic adaptations to glutamate supplementation in the medium, we performed a fast nutrient shift experiment following a previously developed protocol³⁷. Labeled media was used to compare the label propagation through the metabolic network on glucose and on glucose glutamate co-feed. Briefly, *M. smegmatis* was grown in the medium with unlabeled glucose as a single carbon source, transferred to a filter, where it was washed with the new medium for 0, 10, 20,

40, 80, 160 and 320 seconds. We performed the shifts from unlabeled glucose to a mixture of [1-¹³C] glucose and unlabeled glutamate to focus on the fluxes in the upper part of central metabolism; and from unlabeled glucose to a mixture of [U-¹³C] glucose and unlabeled glutamate to focus on the fluxes around the TCA cycle. As a control, the shift was performed to the correspondingly labeled glucose media. We measured the labeled isotopes of intracellular metabolites with LC-MS/MS and compared the time profiles of the label propagation in glucose and glutamate medium to the ones in the glucose medium.

Already within the first 20 seconds, we observed differences in label propagation in glucose plus glutamate medium, indicating that *M. smegmatis* rearranges its fluxes as soon as glutamate becomes available. During the shift to [1-¹³C] glucose, the label propagation in glucose-6-phosphate was the same regardless of glutamate presence in the medium, whereas in dihydroxyacetone phosphate and PEP significant changes were observed already after 80 s (Figure 3c). Since the glucose label is cleaved off in the PP pathway, the slower label propagation in these metabolites indicates a higher activity of the PP pathway. However, according to the SUMOFLUX and ¹³C-MFA estimates, the PP pathway fluxes decrease in the co-feed medium. This discrepancy may be explained by the changes in metabolite concentrations, which we explore below.

During the shift to [U-¹³C] glucose, the label propagation in the upper glycolytic intermediates was similar between the two conditions, whereas the label propagation in PEP was considerably different. The label propagation in aspartate was changing already after the first 20 seconds (Figure 3d). The slow propagation of [¹³C₃] aspartate in the glucose plus glutamate medium reflects the immediate aspartate replenishment from the TCA cycle and decrease of the anaplerotic flux from PEP/pyruvate, which combines a fully labeled C₃ molecule with an unlabeled CO₂ molecule.

From the fast nutrient shift experiments, we concluded that *M. smegmatis* catabolizes glutamate as soon as it becomes available, rerouting it to fuel the TCA cycle and fulfil the anaplerotic demands. Differences in the label propagation in the upper glycolysis indicate that some slight changes might be occurring too, and probably the cellular response to glutamate utilization is a mixture of transcriptional and post-translational responses. The latter might include allosteric regulation by binding of effector molecules to the enzymes. If such regulation takes place, we expect that the effector's concentration would change fast upon glutamate supplementation. To

discover candidates for such putative regulators, we measured the absolute metabolite concentrations during the shift experiment.

Concentrations of several metabolites were changing within seconds after exposure to glutamate (Figure 3e). The intracellular levels of fructose-1-phosphate (F1P) followed the dynamics of glutamate concentration change and increased three-fold in the first 10 seconds of the shift. The concentration levels, however, remained low (~3mmol), hence we hypothesized that the F1P behavior is a side effect of concentration changes of other molecules in the upper glycolysis, causing the conversion of dihydroxyacetone phosphate and D-glyceraldehyde to F1P. Additionally, as F1P is one of the precursors of phenylalanine, tyrosine and tryptophan biosynthesis pathway, the observed changes might reflect global rearrangement in amino acid metabolism. The concentrations of aspartate and PEP were also rapidly increasing upon addition of glutamate. Aspartate and PEP have been reported to be allosteric regulators of fluxes in central metabolism of *E. coli*³⁸⁻⁴⁰, thus the observed rapid flux rerouting might be explained by the existence of similar regulation mechanisms in *M. smegmatis*.

***Mycobacterium tuberculosis* co-utilizes multiple amino acids upon infection.** In the previous sections, we demonstrated that *M. smegmatis* rapidly adapts to the amino acid availability in the medium, and has the capacity for their immediate consumption and catabolism. To investigate whether this metabolic capacity exists in the pathogen *M. tuberculosis* and is relevant for its virulence, we aimed at quantifying amino acid exchange fluxes between bacteria and its host during infection.

We used the available data from a ¹³C labeling experiment with *M. tuberculosis* H37Rv infecting macrophage-like THP-1 cells²⁸. In this experiment, THP-1 cells were pre-labeled in rich RPMI medium supplemented with [U-¹³C] glucose. The macrophages were infected with *M. tuberculosis* and incubated for 48 h in unlabeled RPMI medium. Subsequently, the protein material was separated, and the mass isotopologues of protein-bound amino acids of both organisms were measured with GC-MS. In the control experiments, THP-1 macrophages and *M. tuberculosis* were grown separately in RPMI medium supplemented with [U-¹³C] glucose.

To analyze metabolic fluxes in the infection setup, we built a dual host-pathogen metabolic network, which consisted of the central metabolic networks of THP-1 cell and *M. tuberculosis* connected via the phagosome exchange fluxes of glucose, glycerol, derivatives of pyruvate and acetyl-CoA representing fatty acids, and 20 amino acids.

The exchange fluxes between THP-1 and the environment included uptake of glucose and 20 amino acids according to the composition of RPMI medium. Since none of the exchange fluxes is measured, and the labeling data of only 10 amino acids are available, the global ^{13}C -MFA approach is not applicable to estimate metabolic fluxes²⁸. In the previous study, ^{13}C flux spectral analysis was proposed to identify the network topology providing the best fit to the measurement data. More than 600,000 flux optimizations describing networks with one, two or three active exchange fluxes between *M. tuberculosis* and THP-1 cell resulted in the best-fit model with acetate, alanine and serine being the substrates of *M. tuberculosis* inside the macrophage²⁸. Despite the extensive calculations, only qualitative information on no more than three substrates at a time could be extracted with this approach.

Here, we propose to quantify amino acid exchange between the bacteria and the host by using targeted ^{13}C flux ratio analysis analogous to the experiments with *M. smegmatis*. For each of the 20 amino acids, the biosynthesis and catabolism fractions in *M. tuberculosis* were formulated as described in the previous sections (Figure 1d). In order to train SUMOFLUX predictors, we simulated a comprehensive training dataset consisting of > 300,000 feasible flux distributions in the dual network, covering a broad range of biosynthesis and catabolism fraction values. Due to these rigorous simulations, SUMOFLUX overcomes the challenge of poorly determined network, since all exchange fluxes can be sampled without prior knowledge on their magnitude. Moreover, this *in silico* dataset can be used to train all the 20 biosynthesis and catabolism predictors due to the broad coverage of possible flux distributions. We divided the simulated dataset into a training and testing subset, in order to assess predictors' performance on the data not included in the training phase. Eight out of twenty predictors for alanine, glycine, histidine, leucine, phenylalanine, serine, tyrosine, and valine biosynthesis fraction had good accuracy *in silico* (MAE < 0.1, Figure S11), whereas the predictors for aspartate and glutamate had lower accuracy (MAE = 0.12 and MAE = 0.15, respectively). The other ten predictors performed poorly due to the absence of measurements of the corresponding amino acids, and were excluded from further analysis. The catabolism fraction predictors had low accuracy *in silico* (MAE > 0.11, Figure S12), but were retained for the purpose of comparison with *M. smegmatis* results.

According to the SUMOFLUX estimates, glutamate, alanine, aspartate, glycine, phenylalanine and leucine were to a larger extent consumed from the phagosome, as

their biosynthesis fraction was low (Figure 5a). On the contrary, tyrosine and histidine had a high biosynthesis fraction, whereas serine and valine had almost equal contributions to amino acid pools from biosynthesis and phagosomal uptake (Figure 5a). Generally, these estimates are in concordance with the results of the spectral ^{13}C -MFA, stating that alanine, glutamate, aspartate and serine are taken up from the host, whereas valine is mostly produced by the bacteria²⁸.

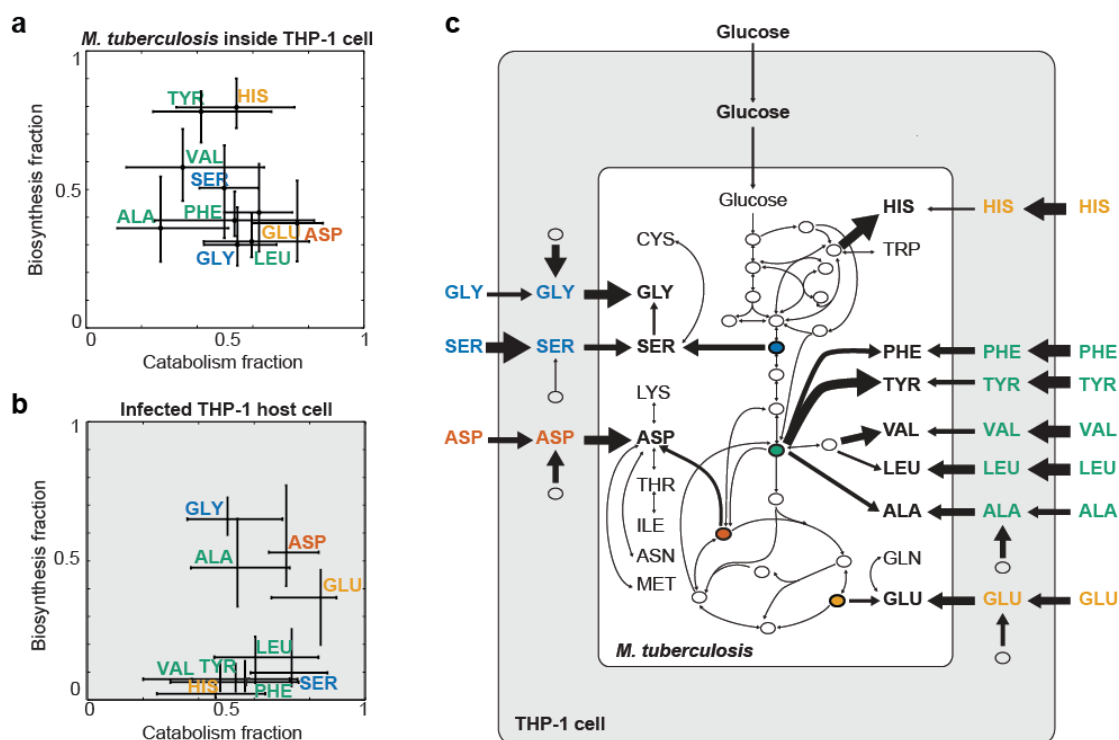


Figure 5. Amino acid exchange fluxes in *M. tuberculosis* and macrophage-like THP-1 cells during infection. (a) SUMOFLUX estimates of biosynthesis and catabolism fractions for 10 amino acids in *M. tuberculosis* infecting THP-1 macrophage. (b) SUMOFLUX estimates of biosynthesis and catabolism fractions for 10 amino acids in THP-1 macrophage during infection. (c) Schematic representation of the estimated amino acid uptake and biosynthesis fractions in the infection setup. Thickness of arrows represents relative contribution of uptake and biosynthesis fluxes to the amino acid pools. Central metabolism of the THP-1 cell is not depicted for illustration purposes. Error bars indicate 50% prediction intervals.

To investigate metabolic behavior of the host during infection, we build 20 amino acid biosynthesis and catabolism predictors for the THP-1 exchange fluxes with the RPMI medium. Again, only predictors for the measured amino acids had acceptable accuracy and were applied to the experimental data (Figure S13, S14). In THP-1 cells, biosynthesis fraction of the essential amino acids valine, tyrosine, leucine, phenylalanine and histidine was negligible, indicating that they were consumed from

the media (Figure 5b). Since the information on the essentiality was not included in the dual model, i.e. all amino acid utilization reactions were reversible, these results serve as a validation of the adequacy of SUMOFLUX estimates. Serine was the only non-essential amino acid with a very low biosynthesis fraction (Figure 5b).

The results of the amino acid uptake analysis in the infection setup are summarized in Figure 5c. We expanded the previously reported list of amino acids consumed by *M. tuberculosis* with glycine, leucine and phenylalanine. According to the SUMOFLUX estimates, histidine and tyrosine had to be synthesized *de novo* during infection. However, considering the analysis of amino acid and glucose co-utilization performed for *M. smegmatis*, we recognized that a high biosynthesis fraction does not necessarily reflect bacterial inability to consume the amino acid from the medium. For example, glutamate, which is incorporated into central metabolism, had a high biosynthesis fraction reflecting its high turnover inside the cell (Figure 1e). Therefore, in order to gain additional information on amino acid consumption during infection, it would be necessary to compare it to amino acid utilization in the medium without carbon limitation.

Hence, we set out to estimate amino acid exchange fluxes in *M. tuberculosis* growing in the rich RPMI medium in the control experiments performed in the infection study²⁸. For the analysis, we used a metabolic network consisting of central carbon metabolic fluxes and amino acid exchange reactions with the environment. As before, we built 20 SUMOFLUX predictors for biosynthesis and catabolism of amino acids, and only ten of them were applied for the experimental data due to the lack of measurements to accurately resolve the other ten fractions (Figure S15, S16). In the rich medium, amino acid utilization patterns of *M. tuberculosis* were similar to those of *M. smegmatis*, with a high biosynthesis fraction of glutamate, valine and histidine and a low biosynthesis fraction of serine (Figure S17). Comparison with the biosynthesis fraction estimates during infection suggests that in the phagosome, *M. tuberculosis* has access to most amino acids, but their amount is limited. Indeed, the biosynthesis fraction of serine is higher during infection than in rich medium, indicating that its uptake was not sufficient to fulfil cellular needs. On the contrary, the biosynthesis fractions of glutamate, valine and alanine were higher in rich medium compared to the infection setup, suggesting that in the phagosome, the uptake of alternative carbon sources required for anabolism of amino acids is impaired.

In summary, targeted ^{13}C metabolic flux ratio analysis with SUMOFLUX provided quantitative estimates of amino acid exchange fluxes between *M. tuberculosis* and its host during infection, which was not possible with global ^{13}C -MFA due to the complexity of the network and lack of data. Moreover, SUMOFLUX estimates enabled comparison of amino acid utilization modes in the infection setup and during growth in rich medium, contributing additional information on the availability of nutrients in the phagosome. Taken together, our results suggest that during infection, mycobacteria have access to most amino acids in limited quantities. The uptake of carbohydrates and fatty acids seems to be insufficient to fulfill the anabolic demand, forcing mycobacteria to make use of any nutrient they encounter.

DISCUSSION

In this study, we investigated metabolic adaptations of mycobacteria co-utilizing amino acids with other carbon sources. Our results underline the complexity of metabolic interactions between the host and the pathogen, which are difficult to quantify even with the most informative direct labeling of intracellular intermediates available. The comparison of amino acid utilization in *M. smegmatis* growing in glucose plus amino acid media, *M. tuberculosis* growing in RPMI medium and *M. tuberculosis* infecting THP-1 macrophages, indicates that during infection, *M. tuberculosis* has access to many sparsely available nutrients, rather than a single carbon source fueling its metabolism. This hypothesis is supported by growing evidence in the literature of mycobacterial metabolic adaptation plasticity²¹. Indeed, mycobacteria have a very robust metabolism with multiple copies of key enzymes, therefore double knockouts are often required to impair its growth^{12,14,41}. The gene and pathway essentiality might as well depend on the infection stage and metabolic state of the host^{11,15}. Altogether, these findings emphasize the need for a system view on mycobacterial metabolism per se, and its interaction with the host. Although certain genes and pathways appear essential for the pathogen in infection experiments and are defined as potential drug targets, a system analysis of the complex host-pathogen interactions could provide a better understanding of the metabolic plasticity of the pathogen and explain the recurrent failures of therapeutic interventions.

Investigating amino acid metabolism in terms of specific biosynthesis and catabolism fractions and focusing on key flux ratios in the central metabolism allowed

us to quantitatively characterize amino acid utilization patterns, group amino acids according to these patterns, and compare two species of mycobacteria in different experimental setups. Generally, amino acids metabolized in the TCA cycle (glutamate, histidine, aspartate) had a high turnover inside the cell, being both actively produced and catabolized, whereas amino acids metabolized through pyruvate (tyrosine, leucine) were mostly incorporated from the media and utilized for protein biosynthesis. Serine, as a special case, was mostly incorporated from the media and actively catabolized by both mycobacterial species. Such different utilization patterns might be partially explained by the different metabolic costs of amino acids in terms of biosynthetically required high-energy phosphate bonds³⁰. Along these lines, it has been previously reported that *E. coli* can regulate metabolic production and consumption during starvation, in order to spare more energetically costly amino acids⁴².

Our follow up experiment with glucose - glutamate co-feed in *M. smegmatis* revealed that as soon as this amino acid is supplemented, it is being taken up and utilized, causing distal flux rearrangements. This finding supports the hypothesis that mycobacterial survival strategy is to remain ready to consume any substrate that it encounters.

When we compared amino acid utilization in *M. tuberculosis* grown in RPMI medium to those in the infection setup, we observed a general decrease in the biosynthesis ratios. At first glance, this finding is contra-intuitive, as the environment inside the macrophage is nutrient-limited, and *M. tuberculosis* is expected to be forced to produce more biomass precursors *de novo*. On the other hand, in the RPMI medium the carbon source is not limiting, therefore bacteria utilize glucose to produce amino acids, as we observed for *M. smegmatis* growing on combinations of glucose and single amino acids. It is important to note that the fractional labeling and the similarity between the labeling patterns of the amino acid in the media and the intracellular amino acid do not unambiguously reflect its utilization patterns. The effect of decreasing biosynthesis fractions is likely explained by the fact that intracellular *M. tuberculosis* is able to consume many amino acids and nutrients from the host, but in very small quantities, so that the consumed metabolites are catabolized locally and do not propagate further through the network. To study *M. tuberculosis* amino acid metabolism more precisely, an *in vitro* experimental setup with a controlled environment, where the substrate of study is labeled, is required. It would be interesting to perform co-metabolism experiments with other carbon sources such as glycerol,

acetate or butyrate, and to vary the amount of carbon source to mimic the sparsity of carbon source during infection and to compare how the biosynthetic fraction of different amino acids changes upon carbon limitation.

For ^{13}C flux analysis during co-feed or co-culture, application of a targeted approach such as SUMOFLUX is particularly beneficial. Addressing specific questions in a complex infection setup, where the measurements are sparse, the nutrient uptake rates are inaccessible, and even the list of substrates may not be fully defined, enables to extract the most information from the available data. Moreover, in case of a dynamic infection setup, the measured protein bound amino acids contain the carry-over labeling from the biomass generated before the cellular exposure to the labeled medium. In the SUMOFLUX workflow, we could account for these uncertainties by simulating thousands of different flux distributions through the network, assuming uptake of multiple carbon sources simultaneously, and mimicking the unlabeled biomass contribution to the amino acid labeling by adding a random fraction of unlabeled isotopes. The combination of surrogate modelling with machine learning imbedded in SUMOFLUX enables to reduce the number of *a priori* assumptions about the system, and predict numerous local flux ratios extracting the most information from the available data. These local flux ratio estimates can be used together with global flux analysis approaches to model mycobacterial metabolism or infection process at a genome-scale. Over the past years, many steps have been made to move towards systemic modeling of mycobacterial metabolism⁴³⁻⁴⁸. Although these models were successful at predicting intracellular physiological parameters, such as biomass, ATP and cofactor production, or the outcome of a specific perturbation, their main limitation is sensitivity to numerous parameters and constrains, which in most cases are set based on prior knowledge and desired outcome. Incorporation of the data from labeling experiments and targeted flux ratio analysis will reduce the uncertainties in the model, help to decrease false positive predictions of drug targets, and improve our overall understanding of mycobacterial metabolism and its regulation.

EXPERIMENTAL PROCEDURES

Chemicals. Chemicals, where nothing else is mentioned, were purchased from Sigma Aldrich (Schnelldorf, Switzerland).

Bacterial culturing. First pre-cultures of *M. smegmatis* were grown on 7H9 Middlebrook medium supplemented with 0.5% glycerol and 0.1% Tween 80 adjusted to pH 6.6. Second pre-cultures and cultures of *M. smegmatis* were grown in 7H9 Middlebrook medium without glutamate and citrate supplemented with 0.05% Tyloxapol. Glucose and amino acids were added to a concentration of 2 gL⁻¹, if nothing else is mentioned. All cultures were grown at 37°C under constant shaking at 300 rpm.

Growth experiment in microtiter plates. Amino acid and glucose co-metabolism experiment was performed in Tecan plate reader (Tecan 200 infinity, Tecan, Männedorf, Switzerland). Second pre-cultures in mid-exponential growth phase (OD600 ~ 0.4 - 1.5) were inoculated in a 48-well plate at OD600 ~ 0.1 to a total culture volume of 500 µL.

Shake-flask experiments. Second pre-cultures in mid-exponential growth phase (OD600 ~ 0.4 - 1 for glucose cultures, OD600 ~ 0.4 - 1.5 for glucose - glutamate cultures) were inoculated in 500 mL shake flasks at 37°C (300 rpm). At multiple time points 1 mL of culture was centrifuged (RT, 16100 rpm, 3 min) and supernatant collected to determine uptake rates. Growth was monitored by OD600 measurement. For steady state experiment, samples were collected in mid-exponential growth phase according to metabolic sampling.

Calculation of physiological parameters. The growth parameters were calculated in Matlab 2013a (MathWorks) using a customized script. Linear curves were fitted to the log-transformed OD measurement data with the polyfit function, and the slope parameter was taken as the growth rate estimate.

Glucose and glutamate concentration measurements. Concentration of glucose in the supernatant was determined by HPLC (Agilent HP1100, Agilent Technologies, Santa Clara, CA, United States) with Aminex HPX-87 H ion exchange column (Bio-Rad, USA) and refractive index detector (RID-6A, Shimadzu, Japan). Mobile phase was 5 mM H₂SO₄, the temperature 45°C and the flow rate 0.6 mL/min.

Supernatants were centrifuged at RT, 14000 rpm, 10 min, and 20 μL was used for measurement. The concentration of glucose was calculated according to the calibration curve.

For glutamate uptake, the supernatant of the growth experiment was analyzed by Agilent 6550 iFunnel Q-TOF (Agilent Technologies, Santa Clara, CA, United States). Supernatants were centrifuged at RT, 14000 rpm, 10 min and the supernatant were diluted 1:200 with Milli-Q-water for measurement. The concentration of glutamate was calculated according to the calibration curve.

Labeling experiments. For labeling experiments [$\text{U-}^{13}\text{C}$] glucose and [$1\text{-}^{13}\text{C}$] glucose (Cambridge Isotope Laboratories, Andover, MA, United States) were added in concentration of 2 gL^{-1} . In case of 50% [$\text{U-}^{13}\text{C}$] glucose experiment, the total concentration of glucose was 2 gL^{-1} .

M. smegmatis cultures were grown until mid-exponential growth phase ($\text{OD}_{600} \sim 0.4 - 1$ for glucose cultures, $\text{OD}_{600} \sim 0.4 - 1.5$ for glucose plus glutamate cultures). Samples were vacuum filtered (HVLP $0.45\text{ }\mu\text{m}$). Sampling volume was calculated according to 1 mL corresponding to biomass of $\text{OD}_{600} = 2$. The filter paper with the bacterial residues was immediately transferred into 3 mL of cold extraction solution (acetonitrile:methanol:MilliQ water, 2:2:1). After extracting for at least one hour, the extracts were transferred to 15 mL Falcon tubes and dried under vacuum. For determining absolute concentrations, 100 μL internal standard was added to the extraction solution.

For analysis by LC-MS/MS, the dried samples were re-suspended in 100 μL MilliQ water. After centrifugation (4°C , 5000 rpm, 10 min) supernatants were transferred to 1.5 mL Eppendorf tube and centrifuged (4°C , 14000 rpm, 10 min).

For analysis by GC-MS, 1 mL of culture at mid-exponential growth phase ($\text{OD}_{600} \sim 0.4 - 1$ for glucose cultures, $\text{OD}_{600} \sim 0.4 - 1.5$ for glucose - glutamate cultures) was harvested. Cell pellets were collected after centrifugation (RT, 16100 rpm, 3 min) by removing the supernatant. After washing with 500 μL of 0.9% NaCl solution, pellets were re-suspended in 200 μL 6M HCl. The cell pellets were hydrolyzed for 12-24 h at 105°C in an oven. The cell hydrolysates were dried in a heating block at 95°C under constant stream of air for 12-24 h. The dried hydrolysates were again dried at 105°C

for 10 min. Derivatization of the amino acids was achieved by suspending the dried hydrolysates in 20 μ L DMF (Dimethylformamide) and adding it to 20 μ L of TBDMSTFA (N-tertbutyldimethylsilyl-N-methyltrifluoroacetamide with 1% (wt/wt) tertbutyldimethyl-chlorosilane). The GC-MS vials were then incubated for one hour at 85°C in a heating block⁴⁹.

Fast shift experiments. Glucose cultures were grown to mid-exponential growth phase $OD_{600} \sim 0.5 - 0.7$. Sampling volume was calculated corresponding to a biomass of an $OD_{600} = 2$. Volume of culture was vacuum filtrated. The filtrate was first washed for 10 s with glucose medium and immediately switched to a new medium and exposed for 0, 10, 20, 40, 80, 160 and 320 s. Samples were then transferred to 3 mL of cold extraction solution 40:40:20 acetonitrile:methanol:MilliQ water and processed as in described in metabolic sampling procedure.

¹³C-MFA analysis. For the ¹³C-MFA analysis, INCA software was used³⁶. The metabolic model for *M. smegmatis* was built on the basis of the provided metabolic model of *E. coli*, with additional output flux from pyruvate to simulate fatty acid production (Table S4). The mean and standard deviation of the measurement data for 60% [1-¹³C] and 40% [U-¹³C] glucose was used as the input data. The glucose uptake rate was constrained according to the 10x measured values (Table S4). For the glucose and glutamate condition, two reactions of glutamate uptake and conversion of glutamate to α -ketoglutarate were added. The glucose and glutamate uptake rates were constrained according to the 10x measured values (Table S4). The mean and standard deviations of the measurement data from [1-¹³C] glucose - glutamate and [U-¹³C] glucose - glutamate experiments were used simultaneously as input data. The “estimate” function of the INCA software was evoked 100 times, the distance of the simulated and measured data for each solution was calculated, and the solution with the smallest distance score was chosen as the final one.

SUMOFLUX analysis. Network construction. *M. smegmatis* metabolic networks for each amino acid utilization with carbon atom transitions and the lists of input and output metabolites were defined (Table S1). In order to reduce the dependency on the biomass vector coefficients, a separate output flux is defined for each of the biomass precursors, therefore biomass precursors are also added to the list of outputs. The substrates are defined as unbalanced compounds and do not participate in the stoichiometric equation

system. The network for *M. tuberculosis* growing in RPMI and dual host-pathogen infection model were built based on the original model²⁸ (Tables S2 and S3).

Flux sampling and ratio calculation. In the flux sampling procedure, the definitions of net, exchange, forwards and backward fluxes are used⁵⁰. By default, the lower and upper bounds for reversible reactions are set to [-100 100], for irreversible reactions to [0 100], and the major uptake flux is set to 10. The amino acid uptake upper bounds were set to 5. To achieve uniform coverage of values for a particular flux ratio or set of ratios, the ratio range was split into segments ([0 0.25], [0.25 0.5], [0.5 0.75], [0.75 1]), and the flux sampling procedure is repeated for each segment with the end points set as flux ratio constraints in the first step. The flux ratio of interest is calculated for each of the flux vectors with the corresponding formulas.

SUMOFLUX workflow. Surrogate modelling of the labeling data, training and testing the predictor, feature selection, cross validation procedure and quantile calculations were performed as described in Chapter 2.

REFERENCES

- 1 Boshoff, H. I. M. & Barry, C. Tuberculosis — metabolism and respiration in the absence of growth. *Nature Reviews Microbiology* **3.1**, 70-80 (2005).
- 2 Russell, D. G. *et al.* *Mycobacterium tuberculosis* wears what it eats. *Cell Host and Microbe* **8**, 68-76, doi:10.1016/j.chom.2010.06.002 (2010).
- 3 Eoh, H. Metabolomics: A window into the adaptive physiology of *Mycobacterium tuberculosis*. *Tuberculosis (Edinburgh, Scotland)* **94**, 538-543, doi:10.1016/j.tube.2014.08.002 (2014).
- 4 Niederweis, M. Nutrient acquisition by mycobacteria. *Microbiology (Reading, England)* **154**, 679-692, doi:10.1099/mic.0.2007/012872-0 (2008).
- 5 Rhee, K. Y. *et al.* Central carbon metabolism in *Mycobacterium tuberculosis*: an unexpected frontier. *Trends in microbiology* **19**, 307-314 (2011).
- 6 Appelberg, R. Macrophage nutritive antimicrobial mechanisms. *Journal of leukocyte biology* **79**, 1117-1128 (2006).
- 7 Eisenreich, W., Dandekar, T., Heesemann, J. & Goebel, W. Carbon metabolism of intracellular bacterial pathogens and possible links to virulence. *Nature reviews. Microbiology* **8**, 401-412, doi:10.1038/nrmicro2351 (2010).
- 8 Kwaik, Y. A. & Bumann, D. Host delivery of favorite meals for intracellular pathogens. *PLoS Pathog* **11**, e1004866 (2015).
- 9 Zhang, Y. J. & Rubin, E. J. Feast or famine: the host–pathogen battle over amino acids. *Cellular microbiology* **15**, 1079-1087 (2013).
- 10 Shi, S. & Ehrh, S. Dihydrolipoamide acyltransferase is critical for *Mycobacterium tuberculosis* pathogenesis. *Infection and immunity* **74**, 56-63 (2006).
- 11 McKinney, J. D. *et al.* Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* **406**, 735-738, doi:10.1038/35021074 (2000).

- 12 Muñoz-Elías, E. J. & McKinney, J. D. *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for in vivo growth and virulence. *Nature medicine* **11**, 638-644 (2005).
- 13 Marrero, J., Rhee, K. Y., Schnappinger, D., Pethe, K. & Ehrt, S. Gluconeogenic carbon flow of tricarboxylic acid cycle intermediates is critical for *Mycobacterium tuberculosis* to establish and maintain infection. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 9819-9824, doi:10.1073/pnas.1000715107 (2010).
- 14 Ganapathy, U. *et al.* Two enzymes with redundant fructose bisphosphatase activity sustain gluconeogenesis and virulence in *Mycobacterium tuberculosis*. *Nat Commun* **6**, 7912, doi:10.1038/ncomms8912 (2015).
- 15 Pandey, A. K. & Sasseti, C. M. Mycobacterial persistence requires the utilization of host cholesterol. *Proceedings of the National Academy of Sciences* **105**, 4376-4380 (2008).
- 16 Van der Geize, R. *et al.* A gene cluster encoding cholesterol catabolism in a soil actinomycete provides insight into *Mycobacterium tuberculosis* survival in macrophages. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1947-1952, doi:10.1073/pnas.0605728104 (2007).
- 17 Marrero, J., Trujillo, C., Rhee, K. Y. & Ehrt, S. Glucose phosphorylation is required for *Mycobacterium tuberculosis* persistence in mice. *PLoS Pathog* **9**, e1003116, doi:10.1371/journal.ppat.1003116 (2013).
- 18 Gouzy, A. *et al.* *Mycobacterium tuberculosis* exploits asparagine to assimilate nitrogen and resist acid stress during infection. *PLoS Pathogens* **10**, doi:10.1371/journal.ppat.1003928 (2014).
- 19 Gouzy, A., Poquet, Y. & Neyrolles, O. A central role for aspartate in *Mycobacterium tuberculosis* physiology and virulence. *Frontiers in cellular and infection microbiology* **3**, 68 (2013).
- 20 Noy, T. *et al.* Central role of pyruvate kinase in carbon co-catabolism of *Mycobacterium tuberculosis*. *Journal of Biological Chemistry* **291**, 7060-7069 (2016).
- 21 de Carvalho, L. P. S. *et al.* Metabolomics of *Mycobacterium tuberculosis* reveals compartmentalized co-catabolism of carbon substrates. *Chemistry & biology* **17**, 1122-1131 (2010).
- 22 Chopra, T. *et al.* Quantitative mass spectrometry reveals plasticity of metabolic networks in *Mycobacterium smegmatis*. *Molecular & cellular proteomics : MCP*, M113.034082--M034113.034082-, doi:10.1074/mcp.M113.034082 (2014).
- 23 Vissa, V. D., Sakamuri, R. M., Li, W. & Brennan, P. J. Defining mycobacteria: Shared and specific genome features for different lifestyles. *Indian journal of microbiology* **49**, 11-47 (2009).
- 24 Rastogi, N., Legrand, E. & Sola, C. The mycobacteria: an introduction to nomenclature and pathogenesis. *Revue Scientifique Et Technique-Office International Des Epizooties* **20**, 21-54 (2001).
- 25 Zamboni, N. ¹³C metabolic flux analysis in complex systems. *Current opinion in biotechnology* **22**, 103-108, doi:10.1016/j.copbio.2010.08.009 (2011).
- 26 Beste, D. J. V. *et al.* ¹³C metabolic flux analysis identifies an unusual route for pyruvate dissimilation in mycobacteria which requires isocitrate lyase and carbon dioxide fixation. *PLoS pathogens* **7**, e1002091-e1002091, doi:10.1371/journal.ppat.1002091 (2011).
- 27 Wendisch, V. F., Graaf, A. A. D., Sahm, H. & Eikmanns, B. J. Quantitative determination of metabolic fluxes during cointilization of two carbon sources: Comparative analyses with *Corynebacterium glutamicum* during growth on acetate

- and / or glucose. *Journal of bacteriology* **182**, 3088-3096, doi:10.1128/JB.182.11.3088-3096.2000. Updated (2000).
- 28 Beste, D. J. V. *et al.* ^{13}C -Flux spectral analysis of host-pathogen metabolism reveals a mixed diet for intracellular *Mycobacterium tuberculosis*. *Chemistry & biology* **20**, 1012-1021, doi:10.1016/j.chembiol.2013.06.012 (2013).
- 29 dos Santos, M. M., Gombert, A. K., Christensen, B., Olsson, L. & Nielsen, J. Identification of *in vivo* enzyme activities in the cometabolism of glucose and acetate by *Saccharomyces cerevisiae* by using ^{13}C -labeled substrates. *Eukaryotic Cell* **2**, 599-608 (2003).
- 30 Akashi, H. & Gojobori, T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of Sciences* **99**, 3695-3700 (2002).
- 31 Sauer, U. & Eikmanns, B. J. The PEP—pyruvate—oxaloacetate node as the switch point for carbon flux distribution in bacteria. *FEMS microbiology reviews* **29**, 765-794 (2005).
- 32 Neyrolles, O. & Guilhot, C. Recent advances in deciphering the contribution of *Mycobacterium tuberculosis* lipids to pathogenesis. *Tuberculosis (Edinburgh, Scotland)* **91**, 187-195, doi:10.1016/j.tube.2011.01.002 (2011).
- 33 Fischer, E. & Sauer, U. Large-scale *in vivo* flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nature genetics* **37**, 636-640, doi:10.1038/ng1555 (2005).
- 34 Fischer, E. & Sauer, U. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *European Journal of Biochemistry* **270**, 880-891, doi:10.1046/j.1432-1033.2003.03448.x (2003).
- 35 Fischer, E. & Sauer, U. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *Journal of Biological Chemistry* **278**, 46446-46451 (2003).
- 36 Young, J. D. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics* **30**, 1333-1335, doi:10.1093/bioinformatics/btu015 (2014).
- 37 Link, H., Kochanowski, K. & Sauer, U. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity *in vivo*. *Nature biotechnology* **31**, 357-361, doi:10.1038/nbt.2489 (2013).
- 38 Morikawa, M., Katsura, I., Taguchi, M. & Katsuki, H. Regulation of *Escherichia coli* phosphoenolpyruvate carboxylase by multiple effectors *in vivo*: I. Estimation of the activities in the cells grown on various compounds. *Journal of biochemistry* **87**, 441-449 (1980).
- 39 Xu, Y.-F., Amador-Noguez, D., Reaves, M. L., Feng, X.-J. & Rabinowitz, J. D. Ultrasensitive regulation of anaplerosis via allosteric activation of PEP carboxylase. *Nature Chemical Biology* **8**, 562-568, doi:10.1038/nchembio.941 (2012).
- 40 Blangy, D., Buc, H. & Monod, J. Kinetics of the allosteric interactions of phosphofructokinase from *Escherichia coli*. *Journal of molecular biology* **31**, 13-35, doi:10.1016/0022-2836(68)90051-X (1968).
- 41 Baughn, A. D., Garforth, S. J., Vilchèze, C. & Jacobs Jr, W. R. An anaerobic-type α -ketoglutarate ferredoxin oxidoreductase completes the oxidative tricarboxylic acid cycle of *Mycobacterium tuberculosis*. *PLoS Pathog* **5**, e1000662 (2009).
- 42 Link, H., Fuhrer, T., Gerosa, L., Zamboni, N. & Sauer, U. Real-time metabolome profiling of the metabolic switch between starvation and growth. *Nature Methods*, doi:10.1038/nmeth.3584 (2015).

- 43 Colijn, C. *et al.* Interpreting expression data with metabolic flux models: predicting *Mycobacterium tuberculosis* mycolic acid production. *PLoS computational biology* **5**, e1000489-e1000489, doi:10.1371/journal.pcbi.1000489 (2009).
- 44 Bordbar, A., Lewis, N. E., Schellenberger, J., Palsson, B. Ø. & Jamshidi, N. Insight into human alveolar macrophage and *M. tuberculosis* interactions via metabolic reconstructions. *Molecular systems biology* **6**, 422-422, doi:10.1038/msb.2010.68 (2010).
- 45 Ma, S. *et al.* Integrated modeling of gene regulatory and metabolic networks in *Mycobacterium tuberculosis*. *PLOS Computational Biology* **11**, e1004543-e1004543, doi:10.1371/journal.pcbi.1004543 (2015).
- 46 Fang, X., Wallqvist, A. & Reifman, J. Modeling phenotypic metabolic adaptations of *Mycobacterium tuberculosis* H37Rv under hypoxia. *PLoS computational biology* **8**, e1002688-e1002688, doi:10.1371/journal.pcbi.1002688 (2012).
- 47 Garay, C. D., Dreyfuss, J. M. & Galagan, J. E. Metabolic modeling predicts metabolite changes in *Mycobacterium tuberculosis*. *BMC systems biology* **9**, 1 (2015).
- 48 Jamshidi, N. & Palsson, B. Ø. Investigating the metabolic capabilities of *Mycobacterium tuberculosis* H37Rv using the *in silico* strain iNJ 661 and proposing alternative drug targets. *BMC systems biology* **1**, 1 (2007).
- 49 Zamboni, N., Fendt, S.-M., Rühl, M. & Sauer, U. ¹³C-based metabolic flux analysis. *Nature protocols* **4**, 878-892, doi:10.1038/nprot.2009.58 (2009).
- 50 Wiechert, W. & de Graaf, A. A. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnology and bioengineering* **55**, 101-117, doi:10.1002/(SICI)1097-0290(19970705)55:1<101::AID-BIT12>3.0.CO;2-P (1997).

Chapter 4

High-throughput targeted ^{13}C flux ratio analysis enables large-scale functional profiling of enzyme phosphorylation in *Escherichia coli*

Maria Kogadeeva, Zrinka Raguz Nakic, Uwe Sauer, Nicola Zamboni

Contributions by MK:

Designed and performed experiments, analyzed the data, wrote the manuscript

Supplementary Information:

Available in Appendix III on page 183 et sqq.

ABSTRACT

^{13}C metabolic flux analysis is the most informative tool to investigate the intracellular reaction rates that must support cellular needs under all environmental conditions. Unlike other omics techniques, such as transcriptomics, proteomics and metabolomics, ^{13}C fluxomics is still poorly suited for high-throughput, mainly due to the challenges of fast, extensive and sensitive labeling data acquisition. Here, we propose a general workflow for high-throughput ^{13}C -fluxomics. Our workflow combines parallel deep-well plate cultivation, untargeted metabolomics measurements with flow injection analysis – time of flight (FIA-TOF), and targeted flux analysis method SUMOFLUX. The complications of untargeted high-throughput mass spectrometry measurements, such as falsely annotated or missing ions, prevent application of standard ^{13}C flux analysis methods. However, SUMOFLUX is able to handle FIA-TOF labeling data due to its flexibility in terms of input information. To improve the data quality and reduce the dependence on faulty measurements, we developed a universal filtering procedure and an alternative input data representation. The proposed workflow was validated by resolving central flux ratios in *Escherichia coli* and its gene knockout mutants with known flux phenotypes. We further exploited ^{13}C flux profiling to estimate flux ratios in 60 *E. coli* strains with mutations in phosphosites and 26 corresponding gene knockout strains to investigate the functional role of enzyme phosphorylation. Our analysis confirmed the inactivating role of isocitrate dehydrogenase phosphorylation reported earlier, and hypothesized functions for several previously uncharacterized phosphorylation events. The established high-throughput flux ratio analysis workflow based on SUMOFLUX and untargeted metabolomics permits > 1000 flux analyses per day, opening up new horizons in the field of ^{13}C fluxomics and raising it to the level of other omics techniques.

INTRODUCTION

^{13}C metabolic flux analysis has become a primary approach to investigate intracellular reaction rates and their regulation, which is crucial for biological, biotechnological and biomedical applications^{1,2}. In the recent years, substantial improvements have been achieved in the speed and sensitivity of ^{13}C data acquisition³⁻⁶, efficacy of the flux estimation methods⁷⁻⁹, and miniaturized cell cultivation systems enabling parallel and

cost efficient experiments^{10,11}. Despite these advances, only a handful of large-scale ¹³C flux analysis studies investigating a few hundred strains have been reported to date¹²⁻¹⁵.

The main requirement for fast and high-throughput ¹³C-fluxomics is a standardized, transferable and robust experimental and computational workflow¹⁶. Such workflow should incorporate: i) parallel cultivation to increase throughput; ii) rapid metabolite sampling to reduce degradation effects; iii) fast and sensitive measurement platform; iv) automated ¹³C information extraction from raw analytical data and v) computationally efficient flux analysis tools. Recently, an automated robotic-based platform for ¹³C flux profiling of microorganisms combining these five steps has been developed¹⁷. However, the achieved throughput was only 20 flux analyses per day due to the long measurement time and complex raw data analysis required for nuclear magnetic resonance spectroscopy used in the workflow¹⁷. Therefore, the requirement for a fast and sensitive metabolite measurement platform remains to be the bottleneck of high-throughput ¹³C-flux analysis.

In order to tackle the bottleneck in ¹³C measurements, we propose to use flow injection analysis – time of flight (FIA-TOF) mass spectrometry, which has a throughput of > 1400 samples per day and is optimized to detect minute amounts of samples as obtained from cultivations in 96-well format¹⁸. With FIA-TOF platform, several thousands of metabolite ions per sample with the mass over charge (m/z) range spanning from 50 to 1000 Da/e can be detected in an untargeted manner within 1 min. Such decrease in measurement time comes with several challenges. First, the absence of a chromatographic separation inevitably leads to overlaps of ions with similar m/z. This problem is aggravated in the case of ¹³C data as each metabolite is represented with a set of isotopologues - identical compounds with different isotopic composition. Second, the untargeted nature of the measurements cannot guarantee detection of all the isotopologues of the same metabolite. Missing data prevent the calculation of normalized mass distribution vectors used as input data for ¹³C flux analysis. Third, even if all the isotopologues are detected, a single misannotated ion will distort the entire mass distribution vector of the metabolite.

The standard flux analysis methods are hardly applicable to analyze ¹³C data measured with FIA-TOF. The global ¹³C-metabolic flux analysis requires additional physiological measurements and poorly handles missing or faulty isotopologues, whereas the local flux ratio analysis depends on the measurements of specified

metabolites and fragments, which is not assured by untargeted metabolomics. One possible way to analyze ^{13}C FIA-TOF data is to perform multivariate analysis, which allows to unravel relative pathway activities, qualitative changes in pathway contributions, and nutrient contributions. However, it requires careful interpretation, and is often non-quantitative and less informative than other flux analysis methods².

At the same time, the previously developed targeted ^{13}C flux ratio analysis method SUMOFLUX (Chapter 2) is particularly suited to deal with the challenges of untargeted FIA-TOF measurements. SUMOFLUX relies on surrogate modelling and machine learning to build predictors for specified flux ratios given the list of available measurements, or input features. The training and testing data for machine learning are simulated *in silico* given the metabolic network, substrate label and the measurement list, and are subsequently used to train the predictors and assess their performance. Due to the flexibility in terms of input data and an embedded data quality check procedure, SUMOFLUX can be combined with filtering and automatic elimination of the faulty measurements.

On these grounds, we set out to establish an experimental and data analysis protocol for large-scale ^{13}C flux analysis based on data acquisition with FIA-TOF mass spectrometer and flux analysis with SUMOFLUX. For a proof-of-principle, we designed a pilot experiment with *Escherichia coli* wild type and a set of knockout mutants growing on labeled glucose. We adapted a cell cultivation and extraction protocol in deep-well plates, and developed rigorous filtering procedures to diminish spectral and annotation artifacts of FIA-TOF data acquisition. We demonstrated that SUMOFLUX could predict expected flux phenotypes of the knockout strains given even small subsets of filtered isotopologues. The proposed ^{13}C fluxomics workflow consists of five steps: i) 96 deep-well cultivation; ii) rapid metabolite sampling in 96 well plates; iii) metabolite measurement with FIA-TOF mass spectrometer; iv) ^{13}C data annotation and filtering v) flux ratio estimation with SUMOFLUX. Since SUMOFLUX estimation time is negligible (Chapter 2), the workflow time depends on the speed of data acquisition, which allows to perform > 1000 flux analyses per day.

We exhibited the applicability of the established ^{13}C fluxomics workflow to investigate functional roles of enzyme phosphorylation in *E. coli*. Phosphorylation is known to play a pivotal role in enzyme regulation in higher organisms, however, its importance in prokaryotes is poorly investigated¹⁹. Although advances in proteomics enabled to discover dozens of phosphorylation sites in *E. coli*²⁰, their functional roles

still remain to be discovered²¹. Since recent developments in genome editing allow large-scale generation of mutants with modified phosphosites^{22,23}, their high-throughput ¹³C flux profiling becomes a promising tool to elucidate the role of phosphorylation in metabolism. Here, we applied the developed ¹³C flux analysis workflow to estimate flux ratio changes in 60 mutants with modified phosphosites and 26 corresponding genetic knockouts grown on labeled glucose. Our analysis confirmed the deactivating role of isocitrate dehydrogenase phosphorylation reported before²⁴ and provided hypotheses on the functional role of several previously uncharacterized phosphorylation events.

RESULTS

High-throughput targeted ¹³C metabolic flux analysis workflow. High-throughput ¹³C metabolic flux analysis requires a combination of rapid experimental setup, data acquisition and a flux analysis method scalable for a large number of samples. The first two preconditions can be fulfilled, for example, with multiwell plate cultivation and untargeted metabolite measurements with flow injection analysis - time of flight (FIA-TOF) mass spectrometer¹⁸. This method allows detection of thousands of ions based on their mass over charge ratio in ~ 1 min per sample due to the omission of chromatographic separation step. Compared to targeted mass spectrometry, analysis of labeling data acquired with FIA-TOF is more demanding, since the risk of overlapping signal from ions with similar m/z and missing signal for some of the ion mass isotopologues is high. However, the targeted ¹³C flux ratio analysis method SUMOFLUX developed in Chapter 2 is well suited to handle missing and overlapping mass isotopologues. Indeed, since SUMOFLUX relies on machine learning to build flux ratio predictors, it is flexible in terms of input features, and practical to test what flux ratios can be estimated provided limited amount of data. Moreover, SUMOFLUX is easily scalable to large number of samples, therefore meets the third requirement for high-throughput flux analysis.

In order to assess the feasibility of high-throughput flux analysis by combining untargeted metabolomics with SUMOFLUX, we designed a pilot experiment using *Escherichia coli* and its genetic knockout mutants growing in labeled glucose. The aim of this experiment was to develop i) the experimental and data acquisition protocols,

ii) FIA-TOF data processing steps and iii) to apply SUMOFLUX to predict flux ratios in central carbon metabolism. Since the chosen knockouts were expected to show distinct flux phenotypes, they served as a validation for the SUMOFLUX flux ratio predictions. The metabolic model used for flux ratio estimation contained 34 metabolites corresponding to 202 mass isotopologues, which were used as a reference in the data processing steps (Tables S1 and S2). In the next sections, we describe each step of the high-throughput flux analysis in more detail.

Experimental workflow and data acquisition. To begin with, we aimed at developing an experimental and data acquisition workflow suitable for performing hundreds of flux analyses in parallel. The proposed workflow is based on plate cultivation and metabolite extraction protocols in 96-well format coupled with flow injection analysis – time of flight mass spectrometry (FIA-TOF)¹⁸. It consists of five steps (Figure 1a). First, cellular pre-cultures were grown in 96 deep-well plates in minimal media supplemented with either 100% [1-¹³C] glucose, or a mixture of 50% [U-¹³C] and 50% unlabeled glucose overnight (~ 18 h). Second, the culture plates prepared with the same media were inoculated at OD₆₀₀ ~ 0.05 and cultivated until cell density reached OD₆₀₀ ~ 1 (~ 5 h). Third, metabolism was quenched by centrifugation, and intracellular metabolites were extracted in the same plates with hot ethanol (~ 0.5 h per plate). Fourth, the extracts were transferred into a conical bottom 96-well plate, stored at -80° C, subsequently thawed, and directly measured with FIA-TOF (~ 1.5 h per plate). Fifth, the raw spectra of the samples were aligned and peaks were automatically picked with a procedure described before¹⁸. In total, 26,456 ions were detected with mass over charge ratio spanning the range [50 - 1000] Da/e, and intensities varying up to five orders of in-spectrum dynamic range.

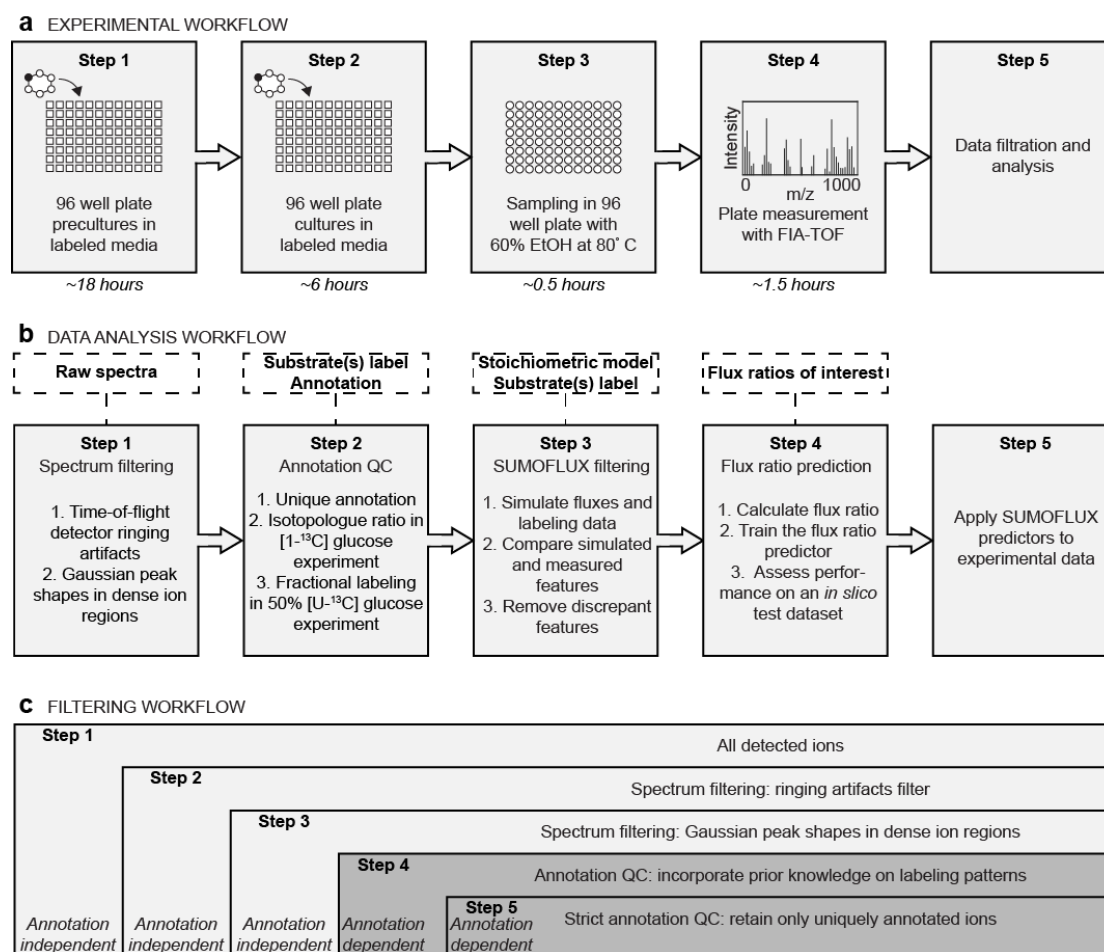


Figure 1. Experimental and data analysis workflows for high-throughput ^{13}C flux ratio analysis. (a) Experimental workflow for strain cultivation in 96 deep-well plates. Time estimates are provided per plate. (b) Data analysis workflow. Input information is depicted in dashed-line rectangles. (c) Visualization of the filtering layers. At each further step, the filters are cumulated.

Compared to the typical experimental workflow for ^{13}C flux analysis, which consists of shake flask cultivation and data acquisition with liquid chromatography or gas chromatography - mass spectrometry, the proposed workflow offers several advantages. First, the cultivation in deep-well plates enables convenient handling of at least 96 samples at once. Second, this type of cultivation requires much less media (2 mL per sample versus 100 mL per sample in a shake flask), thus drastically reducing the experimental costs. Third, FIA-TOF platform offers manifold decrease in measurement time (1 min per sample versus ~ 30 min per sample) and increase in ion coverage.

However, the time efficiency of the mass spectrometry measurements gained by omitting the chromatographic separation step and untargeted detection comes with additional challenges posed by detector artifacts, spectral crowding, and multiplicity of annotation. In the next sections, we propose a multi-step filtering procedure, which

deals both with technical aspects of peak picking (spectral filters) and multiple ion annotation (ratiometric filters) (Figure 1c).

Spectral filtering of FIA-TOF data. The outcome of a FIA-TOF measurement is a raw spectrum represented by a profile of ion counts versus m/z . In order to use the acquired data, ion peaks have to be picked and subsequently annotated. The peak picking procedure is complicated by two major problems. The first problem of the TOF detectors is the occurrence of detector ringing. The second problem is the convolution of peaks in crowded regions. Here, we propose a two-step procedure of spectral filtering (Figure 1b, c).

Detector ringing artifacts are caused by the time-of-flight detector and appear as small signals detected on the right (heavier) side of a high-intensity peak. These peaks do not exist in reality, but are artifacts caused by the detector electronics, therefore their intensity correlates strongly with the intensity of the leading peak. We used a filter that eliminates all peaks on the right side of a high-intensity peak (up to 0.3 Da/e) if their intensities have a Pearson correlation coefficient ≥ 0.9 . The threshold was chosen to be permissive to ensure that potentially important ions in the crowded regions are retained (for example, glutamine $m+1$ peak which is close to the glutamate $m+0$ peak) (Figure 2a). As a result, 12% of picked peaks were discarded (3,192 out of 26,456 picked peaks).

The second issue is the overlap of peaks with similar m/z , which leads to mismeasurement of their intensity. In order to correct for such overlaps or even discard peaks in overly crowded regions, we applied a Gaussian shape fitting procedure. Specifically, we attempted to reconstruct the detected spectrum with a sum of Gaussian functions centered on the known centroids (Figure 2b). If the goodness of fit of the reconstructed spectrum for a specific peak is low, or its intensity is largely influenced by its neighbors, the peak is discarded. Since for the ^{13}C flux analysis only the compounds contained in the metabolic model are relevant, we applied the Gaussian fit filtering to the peaks detected at m/z corresponding to those of model isotopologues (Table S2). Out of the 202 model isotopologues, 158 were detected in the spectra, and 111 passed the Gaussian fit filter with the goodness of fit threshold of 90% (Table S2, Materials and Methods).

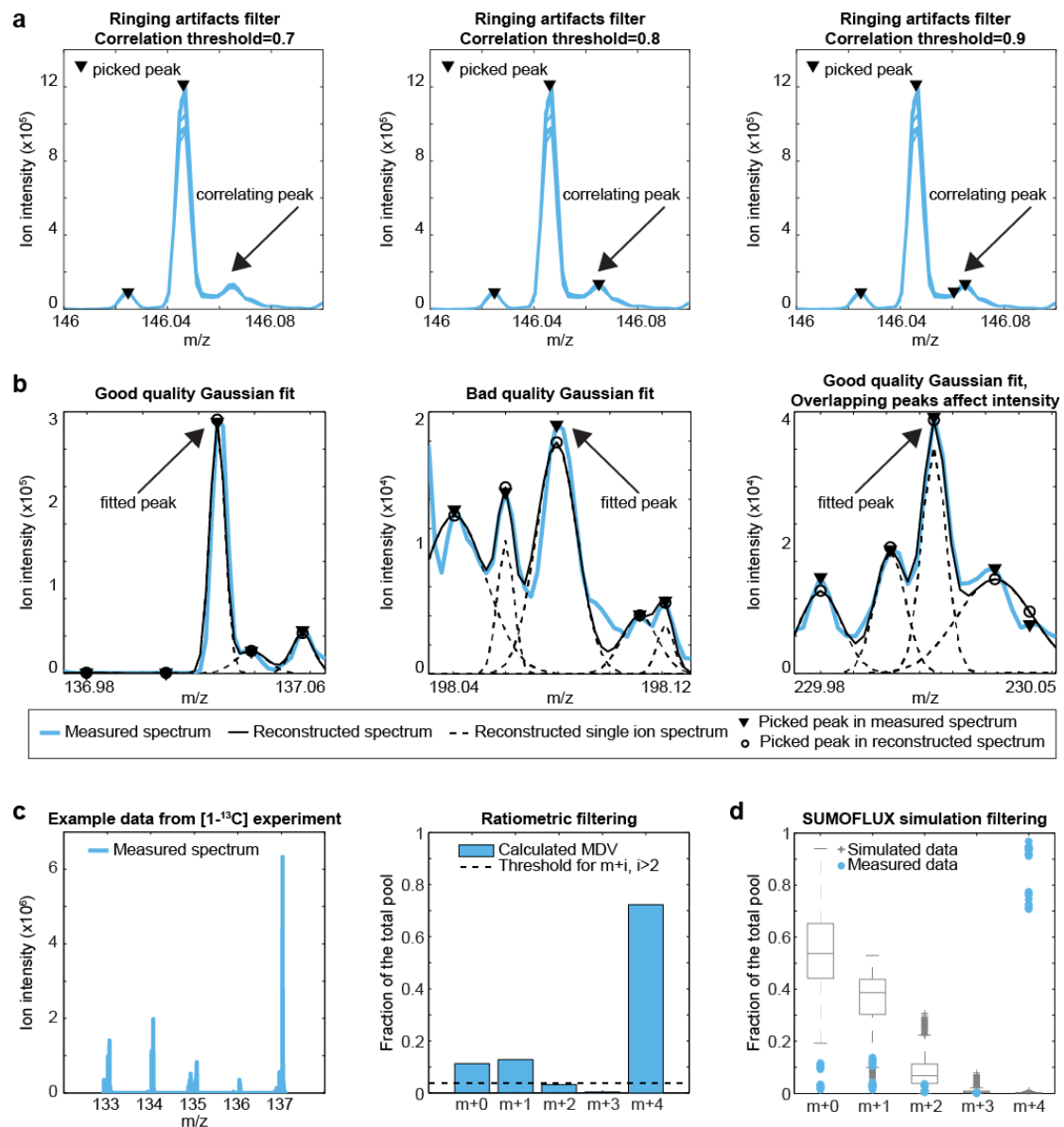


Figure 2. Examples of the FIA-TOF data filtering steps. (a) Spectral filtering: effect of the correlation threshold of the ringing artifacts filter on picking the glutamate $m+0$ (high intensity peak) and glutamine $m+1$ (the correlating peak) peaks. (b) Spectral filtering: examples of the Gaussian fit filtering: left, good quality fit of the restored profile; middle, low quality fit of the restored profile; right, good quality fit of the restored profile, however, the peak discarded due to high influence on its intensity by the neighboring peaks ($> 10\%$ of the measured peak intensity). (c) Ratiometric filtering: left, example spectrum of malate detected in a sample from [$1\text{-}^{13}\text{C}$] experiment with unexpectedly high intensity of $m+4$ isotopologue; right, the calculated mass distribution vector and ratiometric filtering threshold results in discarding the $m+4$ isotopologue (threshold of 33% of the $m+0$ value is depicted with the dashed line). (d) SUMOFLUX simulation filtering for the malate example. The $m+4$ and $m+0$ isotopologues are discarded since they are not covered by the simulated data distributions (threshold of $> 50\%$ of measured values to be outliers).

Annotation quality check and ratiometric filtering of FIA-TOF data. Out of the 111 isotopologue ions of interest that passed the spectral filtering, only 27 ions had a unique annotation (Table S2). This means that there was only one isotopologue with the matching m/z value in the whole library used to relate detected m/z to those of known compounds. Given the fact that the annotation library was compiled from the KEGG database for *E. coli*, which contains compounds detectable only in specific experimental setups (such as drugs and their derivatives), as well as compound isomers (such as D-alanine and L-alanine), we suspected that some of the non-uniquely annotated compounds can be informative and thus should be retained for further analysis. Based on the prior knowledge about the labeling strategies that were used in the pilot experiment, we proposed a ratiometric filtering procedure and annotation quality check suitable for evaluation of the labeled metabolites.

Two parallel labeling strategies were administered in the pilot *E. coli* experiment: 100% [$1-^{13}\text{C}$] glucose and a mixture of 50% [$\text{U}-^{13}\text{C}$] and 50% unlabeled glucose. The ratiometric filtering criterion is based on the fact that in the cells consuming [$1-^{13}\text{C}$] glucose the probability of a high abundance of a heavy isotopologue (with a mass shift by more than two ^{13}C) is low. Therefore, we calculated the ratio between the intensity of the heavy and natural mass isotopologues of each compound, and set a threshold of 0.33 to define an unreliable measurement. For example, the $m+4$ isotopologue of malate was four times more abundant than its $m+0$ isotopologue, indicating that the detected signal is biased by overlapping signals, despite its unique annotation (Figure 2c). After the ratiometric filtering, 104 out of 110 isotopologue ions of interest were retained (Table S2).

The experiment with [$\text{U}-^{13}\text{C}$] glucose provides additional filtering criterion. The fractional labeling of each compound is expected to correspond to the fraction of labeled glucose in the media (0.50 ± 0.05 in our case). This criterion incorporates information on all detected mass isotopologues of a compound. Therefore, it is metabolite and not isotopologue ion specific. The 104 mass isotopologues passing the ratiometric filter are related to 31 metabolites, for which we calculated the fractional labeling. Only five metabolites passed the quality check with fractional labeling values in the range 0.50 ± 0.05 (Figure S1). Since the fractional labeling calculation requires intensity measurements of all mass isotopologues of a compound, in the case when not all of them were detected or passed the filtering procedures, the calculated value is a lower bound for the actual fractional labeling. Therefore, we decided to keep the

isotopologue ions corresponding to metabolites with low fractional labeling and incomplete set of detected isotopologues for further analysis. The fractional labeling information can only be used for quality check and not for filtering, because it is not sufficient to detect which of the isotopologues might be overlapping with isobaric ions in case of contradictory values.

The ratiometric filtering and fractional labeling quality check are valuable ion selection criteria. However, they are only appropriate in specific experimental setups. In experiments with a mixed labeling strategy or multiple substrate co-metabolism, it is hardly possible to formulate selection rules. The strict filter of uniquely annotated compounds is applicable for any type of data, but might discard crucial information required to resolve fluxes of interest. In the next sections, we employ SUMOFLUX to test the effect of the filtering steps on flux ratio resolvability, and introduce an additional universal filtering procedure based on simulated labeling data generated in the SUMOFLUX workflow.

Feature selection for SUMOFLUX flux ratio analysis. After the spectral and ratiometric filtering of the FIA-TOF data, the list of isotopologue ions retained for ^{13}C metabolic flux analysis consisted of 105 ions (Table S2). For classical ^{13}C flux analysis, the labeling data is represented with mass distribution vectors, which indicate the fraction of each mass isotopologue in the total pool of the compound. Two potential problems affect the calculation of mass distribution vectors from the FIA-TOF data. First, even after filtering some of the isotopologues signals might originate from overlapping ions. Second, in case of missing isotopologues, the fractions of other isotopologues in the total compound pool will be overestimated.

The SUMOFLUX workflow consists of five steps: i) sampling of feasible flux solutions given the metabolic network; ii) simulating corresponding metabolite labeling patterns for each flux vector given the substrate label; iii) calculating the flux ratio of interest from the flux vectors; iv) dividing the simulated data in two parts to train the flux ratio predictor and test its performance; v) finally, applying good quality predictors to the experimental data (Figure 1b). Due to the extensive flux sampling, the labeling data simulated with SUMOFLUX is supposed to encompass all possible metabolic phenotypes for a given experiment. Therefore, if there are discrepancies between the *in silico* and measured data, they are likely caused by the erroneous FIA-TOF measurements. For example, comparing the malate data measured in the [$1\text{-}^{13}\text{C}$]

experiment (Figure 2c) with the corresponding *in silico* distribution reveals striking differences between the feasible and calculated fractions of the m+4 isotopologue (Figure 2d). In case more than half of the measured values were not covered by the *in silico* data, we decided to exclude mass isotopologue features from further analysis. However, in the example of malate, the natural m+0 isotopologue would be discarded as well, since the erroneous m+4 isotopologue causes the underestimation of other isotopologues' fractions.

Due to the machine learning procedure, SUMOFLUX is flexible in terms of input data, as virtually any vector of features can be used to train the flux ratio predictor. Therefore, we searched for a representation of the labeling data which would reduce the dependency on the erroneous features. As an alternative to the mass distribution vectors, we propose to calculate pairwise isotopologue ratios for each metabolite of interest. The advantage of this transformation is that the erroneous features will influence only the ratios they are involved in, while other features will remain unaffected (Figure 3a). Similarly, the pairwise isotopologue ratios do not depend on the missing isotopologues. The simulation filtering procedure described earlier is equally applicable to the isotopologue ratio features as to the mass distribution vectors. In the aforementioned example case of malate, only the features dependent on the erroneous m+4 isotopologue were discarded from further analysis (Figure S2). It is important to consider that the number of features increases, as the number of isotopologue ratios is larger than the length of the mass distribution vector, which may lead to predictor overfitting. Notably, the feature selection procedure embedded in SUMOFLUX helps to avoid this potential issue of overfitting (see Materials and Methods for details).

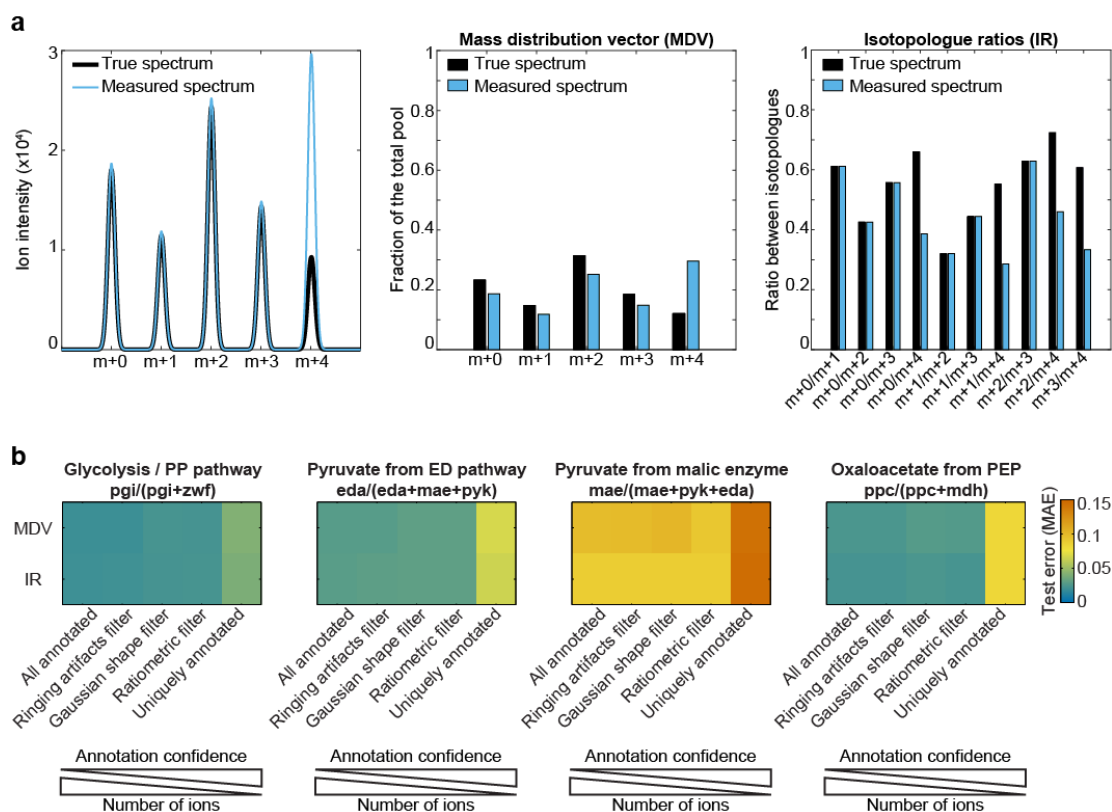


Figure 3. Robust feature selection and its effect on flux ratio predictors' performance *in silico*. (a) *In silico* simulation of an isotopologue overlapping with another compound, left; and its effect on mass distribution vector (MDV), middle; and isotopologue ratios (IR), right. (b) *In silico* testing mean absolute error (MAE) values of four flux ratio predictors build given either of the five sets of annotated ions represented with either MDV or IR.

***In silico* testing the effect of data filtering on flux ratio resolvability.** To verify whether the labeling data retained after the filtering steps is sufficient to perform flux ratio analysis, and whether the pairwise isotopologue ratios can be used instead of mass distribution vectors, we tested the resolvability of four key flux ratios in central metabolism of *E. coli* (Figure 4a).

We tested five sets of features resulting from the different filtering steps: all detectable isotopologue ions contained in the model, features passing ringing artifacts filter, Gaussian shape filter, ratiometric filter, and uniquely annotated ions. For each set of features, we calculated the mass distribution vectors and isotopologue ratios, and subsequently built ten predictors for each of the four flux ratios. The predictors' performance was assessed on *in silico* test datasets not involved in training by calculating the mean absolute error (MAE) of the estimates. Predictors with $MAE < 0.05$ were considered to be highly accurate, and with $MAE < 0.1$ of acceptable quality. For all tested ratios, there was no difference between the quality of the

predictors built using isotopologue ratios or mass distribution vectors as input (Figure 3b). The subset of annotated ions used for prediction, on the contrary, strongly affected the predictors' quality. As the confidence in the ion annotation increased (by eliminating dubious ions in the filtering steps), the number of ions decreased, and the predictors' quality decreased as well (Figure 3b). Only the ratio between glycolysis and pentose phosphate (PP) pathway could be resolved with good accuracy given the subset of uniquely annotated ions. For the other tested ratios, all subsets but the uniquely annotated ions were sufficient to achieve high or acceptable (in the case of pyruvate fraction from malate) prediction accuracy (Figure 3b).

To summarize, the *in silico* SUMOFLUX testing revealed that isotopologue ratios can substitute mass distribution vectors as input for targeted flux ratio analysis. The set of uniquely annotated ions was sufficient to resolve only one ratio, whereas other filtering steps retained sufficient information to resolve all tested flux ratios in central metabolism. Although with SUMOFLUX simulations we demonstrated the feasibility of flux ratio analysis with the FIA-TOF data, our results might be optimistic since the actual measurements might still contain artifacts not captured by *in silico* data. Therefore, in order to state the usability of the high-throughput flux analysis protocol, it has to be demonstrated on experimental data with *a priori* expectations on the flux estimates.

SUMOFLUX captures flux phenotypes of *E. coli* knockouts using FIA-TOF data.

Next we set out to verify whether SUMOFLUX can estimate flux ratios from real high-throughput mass spectrometry data. The five knockout strains included in the pilot experiment cannot catalyze specific reactions, therefore have expected flux phenotypes which serve as a reference for SUMOFLUX flux ratio predictions. In order to describe those metabolic phenotypes, we applied the four predictors described in the previous section, and trained four additional flux ratio predictors to complement the information about the branch points in central metabolism (Figure 4a).

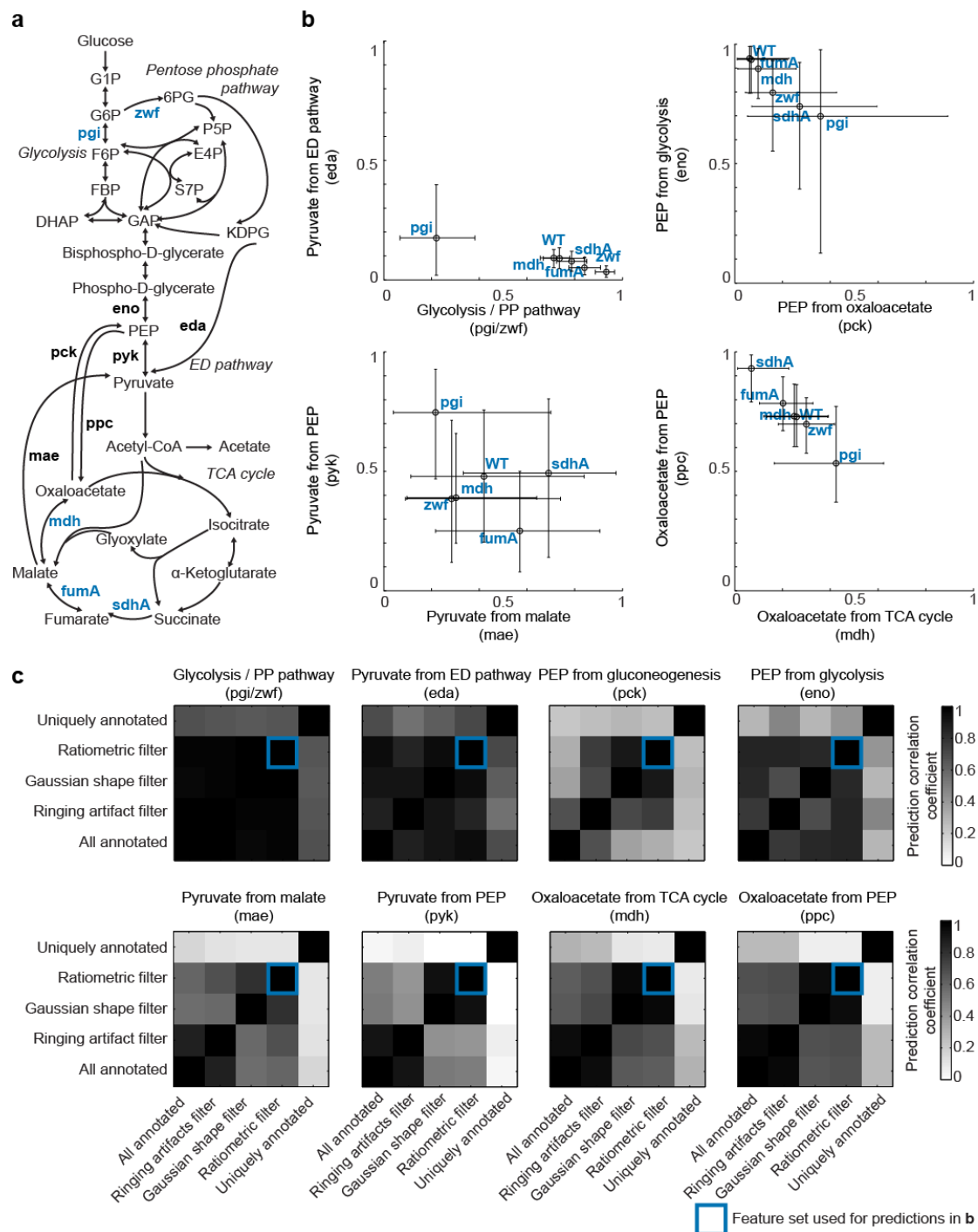


Figure 4. Metabolic flux phenotypes of *E. coli* gene knockouts are captured with high-throughput SUMOFLUX workflow. (a) Schematic representation of *E. coli* central carbon metabolism network and gene knockouts used for the proof of principle. (b) Estimated flux ratios for the wild type and knockout strains, error bars represent [10% 90%] prediction quantiles. (c) Correlation of flux ratio predictions made given the five different subsets of annotated ions. The ion subset used for predictions in (b) is marked with the blue rectangle.

As input data we used the isotopologue ratios calculated for the ions which passed all filtering steps but the unique annotation (Figure 1c). Additionally, the simulation filtering was performed before training the predictors. The eight flux ratio

predictors had good and acceptable accuracy *in silico*, although the accuracy of the two predictors for the phosphoenolpyruvate (PEP) fractions were rather low (MAE=0.14, Figure S3). Applied to the experimental data, the SUMOFLUX predictors were able to capture the expected flux phenotypes of the *E. coli* knockouts. Indeed, for the Δ pgi and Δ zwf mutants lacking enzymes in glycolysis and PP pathway, respectively, the ratio between glycolysis and PP pathway was significantly lower or higher than in wild type (FDR<10⁻¹², Figure 4b, upper left). Furthermore, the fraction of pyruvate from Entner-Doudoroff pathway was the largest in the Δ pgi and the smallest in Δ zwf, which has been shown before²⁵ (Chapter 2). For the Δ pgi and Δ sdhA mutants, the estimated fraction of PEP originating from gluconeogenesis was significantly higher than in wild type (FDR<0.05, Figure 4b, upper right), that has been observed previously (Chapter 2). The Δ sdhA and Δ fumA mutants lacking an enzyme in the tricarboxylic acid (TCA) cycle had higher fraction of oxaloacetate originating from anaplerosis from PEP. On the contrary, the Δ zwf mutant of the PP pathway and the slow growing Δ pgi mutant had a larger fraction of oxaloacetate originating from the TCA cycle compared to wild type, reflecting the compensation for cofactor production in case of Δ zwf, or the increased glyoxylate shunt activity in case of Δ pgi^{25,26} (Chapter 2), (Figure 4b, lower right). The estimates for the malic enzyme versus glycolysis fluxes contributing to the pyruvate pool had large prediction intervals (Figure 4b, lower left), and could not be used to distinguish specific flux phenotypes.

Flux ratio estimates for the knockout strains were in agreement with the expectations about their metabolic phenotypes, confirming the feasibility of ¹³C flux ratio analysis using FIA-TOF data. The input features used to train the SUMOFLUX predictors were rigorously filtered with the four-step filtering procedure (Figure 1c). We therefore decided to investigate the impact of these filtering steps on flux ratio estimation by calculating correlation coefficients between predictions made using the five filtered subsets of data. Similar to the *in silico* testing results, the predictions based on uniquely annotated ions were diverging from those made with other filtered subsets (Figure 4c). The estimates for glycolysis and PP pathway fractions and pyruvate from Entner-Doudoroff were in good agreement for any of the four filtering steps, whereas estimates for pyruvate, PEP and oxaloacetate origins were consistent between predictors trained on data filtered with either none and ringing artifacts filters, or Gaussian shape and ratiometric filters (Figure 4c).

Overall, in the pilot experiment with genetic knockouts of *E. coli* we demonstrated that the combination of untargeted FIA-TOF measurements with SUMOFLUX workflow is applicable for fast and high-throughput flux screening. The proposed filtering procedures helped to increase the data quality, whereas the representation of labeled data with isotopologue ratios reduced the potential influence of misannotated or missing isotopologues on the other isotopologues of the same compound. For all tested flux ratio predictors in central metabolism, the estimates obtained for the data filtered with Gaussian shape or ratiometric filters were in agreement, underlining i) the importance of filtering for successful flux ratio analysis ii) the redundancy of the experiment-specific ratiometric filter. Therefore, the proposed data processing procedure is not experiment specific, thus establishes the general applicability of the developed high-throughput ^{13}C flux ratio analysis workflow.

Large-scale flux profiling of *E. coli* phosphorylation mutants. Having established a fast high-throughput ^{13}C flux analysis workflow with SUMOFLUX, we set out to demonstrate its large-scale application for functional characterization of enzyme phosphorylation in *E. coli*. Since the advances in genome editing enable targeted mutations of specified phosphosites²³, their ^{13}C fluxomics profiling offers a unique opportunity to gain insights into their functional role in cellular metabolism.

We used a set of 60 *E. coli* strains with mutations in phosphosites (phosphomutants). The phosphoresidues serine or threonine were mutated to alanine, abolishing phosphorylation of the targeted amino acid (the phospho-OUT mutants). The phosphoresidues were additionally mutated to glutamic acid to mimic a constitutively phosphorylated state of the protein (the phospho-IN mutants). The wild type *E. coli* and 60 phosphomutants were grown on either [1- ^{13}C] glucose, or a mixture of 50% [U- ^{13}C] and 50% unlabeled glucose in 96 well plates, and intracellular metabolites were measured with FIA-TOF mass spectrometer, following the experimental and data analysis protocols developed in the previous sections.

The mutated enzymes catalyze reactions in central carbon (Figure 5a), amino acid and nucleotide metabolism and sugar transport (Table S3). Most of the mutants did not have any growth defect, or had a minor growth decrease compared to wild type. The most severe growth phenotype was observed for the phospho-IN mutation in isocitrate dehydrogenase (Icd), an enzyme catalyzing one of the reactions in the TCA

cycle; and galactitol permease (GatB), an enzyme involved in galactitol transport (Table S3).

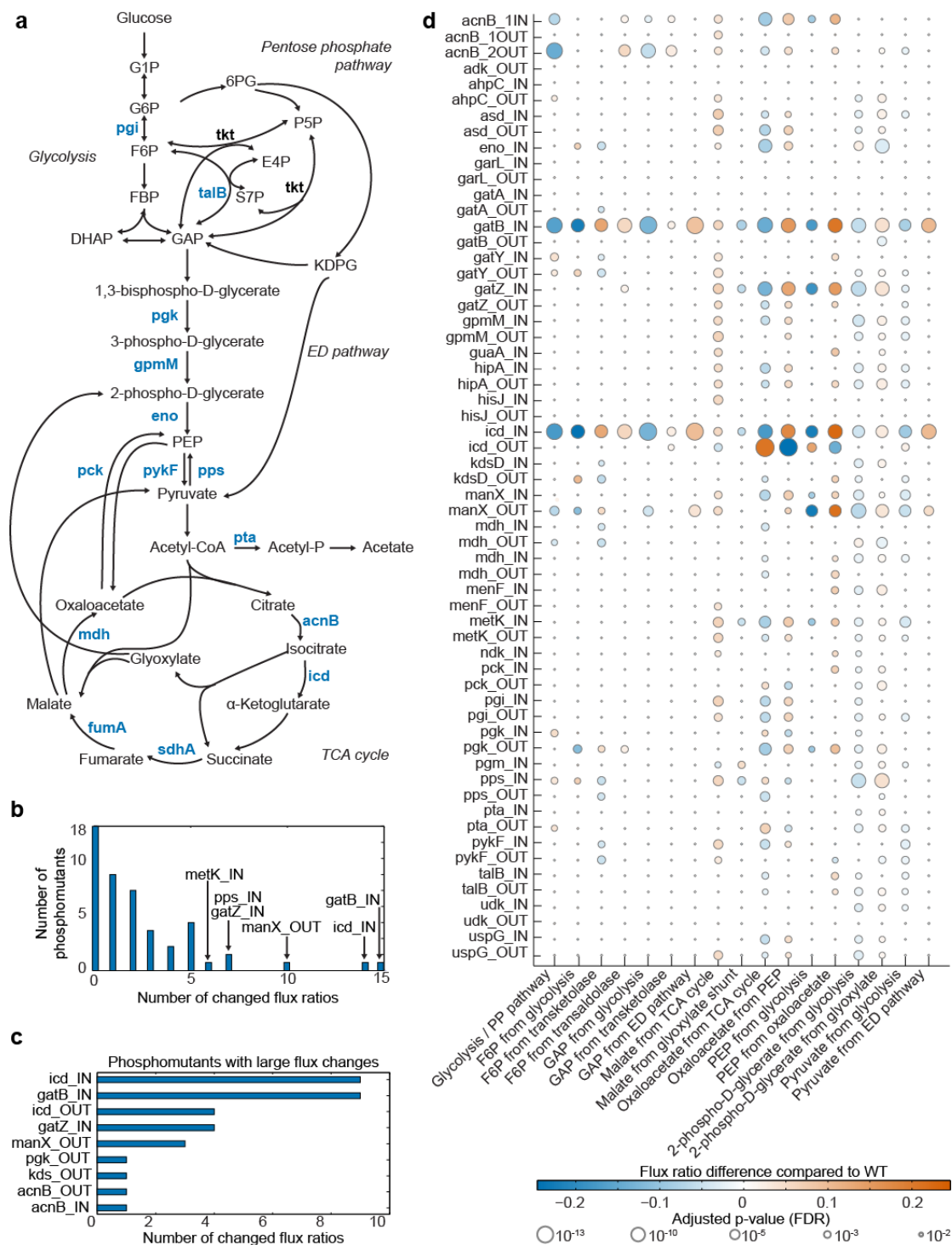


Figure 5. High-throughput metabolic flux profiling of *E. coli* phosphomutants. (a) Schematic representation of *E. coli* central carbon metabolism network and enzymes with mutated phosphosites (only subset of enzymes involved in central metabolism is shown). (b) Histogram representing the number of mutants exhibiting flux ratio changes of > 0.05 compared to the wild type values. (c) The phosphomutants exhibiting > 0.10 flux ratio change compared to the flux ratios of wild type. (d) An overview of the difference between mutant and wild type flux ratios for all profiled phosphomutants. The color corresponds to the magnitude of change, whereas the dot size corresponds to significance (false discovery rate, FDR).

In order to investigate metabolic flux change upon phosphorylation, we formulated 17 flux ratios that describe the source of metabolites produced via alternative reactions in the central metabolism network used for analysis (Figure 4d). All flux ratio predictors showed good or acceptable accuracy on the *in silico* testing dataset (Figure S4), and were subsequently applied to estimate the flux ratios for the phosphomutant samples. The effect of phosphorylation on the metabolic fluxes was quantified as the flux ratio difference between the mutant and the wild type. Since we did not expect major flux rearrangements for most of the mutants due to the absence of a growth defect, we set a permissive threshold of 0.05 to consider flux ratio to be changing. To distinguish between the global and local effects, we calculated how many flux ratios were changed for each phosphorylation event. Most of the mutants had few changes in metabolic ratios, one third of the mutations had no effect on central metabolism, and six mutants had global effects with more than five changing flux ratios (Figure 5b). Only nine phosphomutants had at least one large flux ratio change compared to wild type (> 0.1), among them the Icd and GatB phosphomutants were undergoing the most changes. This is little surprising given their large growth defect, however, the other mutants causing global or large changes did not have severe growth defects with the exception of aconitase (AcnB) mutants, with mutations in the enzyme catalyzing one of the TCA cycle steps (Figure 5c, Tables S3).

Intuitively, if phosphorylation activates or de-activates the corresponding enzyme under studied conditions, the phospho-IN and phospho-OUT mutants should exhibit opposite flux changes. To detect such cases, we calculated Pearson correlation between the flux ratios of the opposite phosphorylation mutants. Only in four cases the flux ratio changes had negative correlation: for Icd, GatB, phosphoglycerate kinase P_{gk}, a glycolytic enzyme, and isochorismate synthase MenF (Table S4). The icd_IN and icd_OUT mutants demonstrated opposite flux change patterns primarily affecting the flux through the TCA cycle. Consistent with the fact that enzyme phosphorylation deactivates the Icd enzyme in the TCA cycle²⁴, the icd_IN mutant had a decreased, and the icd_OUT mutant had an increased fraction of the TCA cycle contributing to the formation of oxaloacetate (Figures 5d, S5). Between the phospho-IN and phospho-OUT mutations in the P_{gk} and GatB, the largest flux ratio changes were observed in the transketolase forming fructose-6-phosphate and TCA cycle fluxes (Figures 5d, S5). In these cases, the role of phosphorylation is not straightforward to explain, since the flux changes are distal to the phosphorylation site. Phosphorylation might be triggering

some regulatory mechanisms affecting glucose transport (as GatB is involved in the phosphotransferase system²⁷) and pentose phosphate pathway activity. For the phosphomutants of isochorismate synthase, the correlation was too low to distinguish the flux change patterns (Pearson correlation -0.07, Table S4, Figure S5).

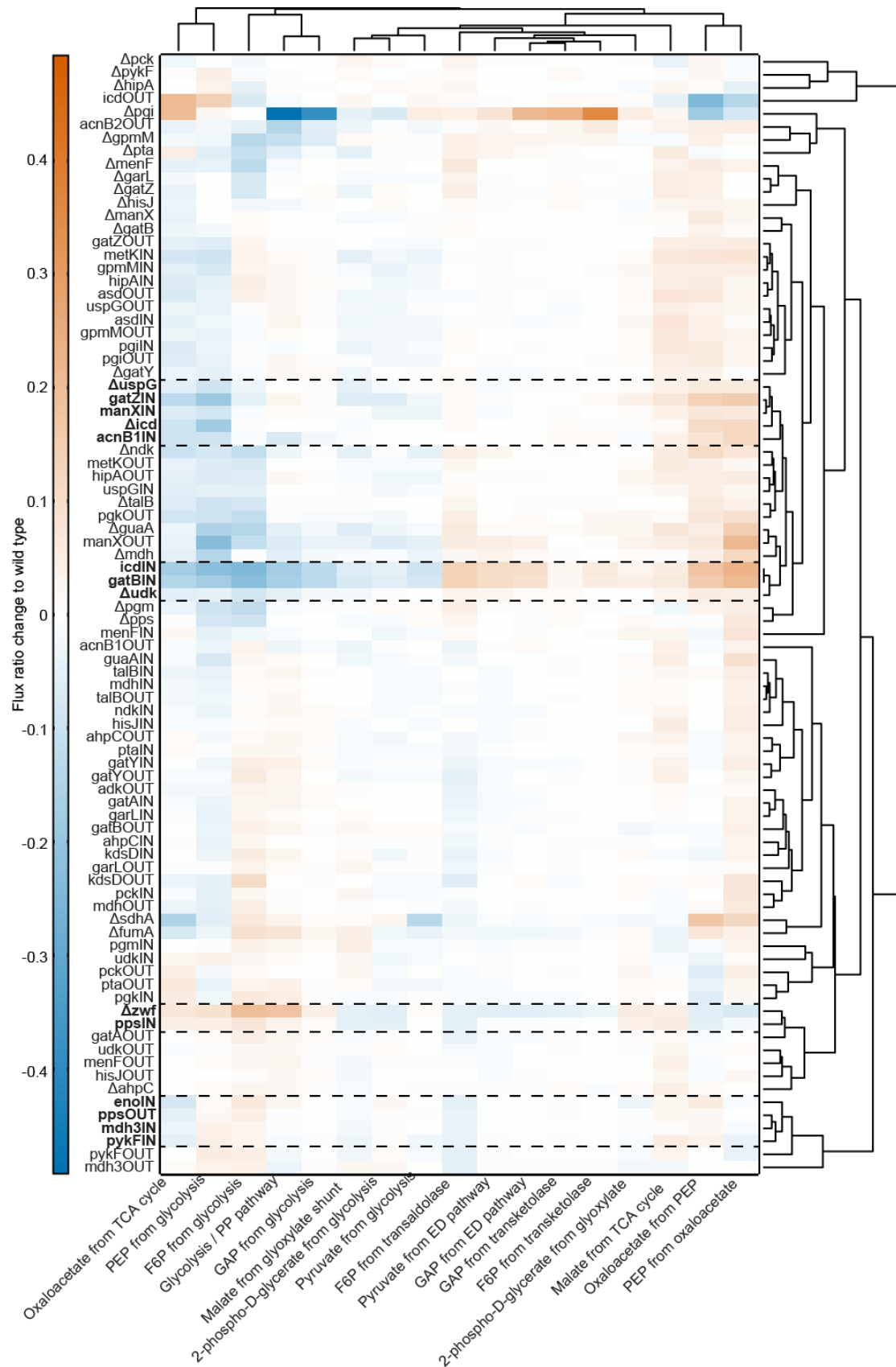
Even if a phosphomutant exhibits a flux phenotype distinct from the wild type, it is often difficult to infer the direct function of phosphorylation due to the caused chain of distal flux rearrangements. To get additional insights into the role of phosphorylation, we decided to compare metabolic flux ratios of the phosphomutants to the corresponding gene knockout mutants. We profiled 26 knockout strains following the same workflow, and performed hierarchical clustering of the flux changes in both knockouts and phosphomutants. The clustering revealed several remarkable flux patterns.

First, the *icd_IN* and *gatB_IN* mutants belong to the same cluster due to the strong and concordant flux ratio changes. Despite the inactivating role of Icd phosphorylation, the Δicd mutant did not occur in the same cluster, since the observed flux ratio changes were local and affected only the contribution of the TCA cycle and anaplerosis to the formation of oxaloacetate (Figure 6). This fact together with the absence of a growth defect in Δicd mutant (Table S3) indicate that Icd phosphorylation, apart from inactivating the enzyme, might be involved in global regulatory processes. The closest flux phenotype to the *icd_IN* and *gatB_IN* mutants was exhibited by the knockout of uridine kinase *Udk*, which had concordant, but minor flux ratio changes. At the same time, the flux phenotypes of *icd_IN* and *gatB_IN* cannot be explained solely by their severe growth impairment, since their flux ratio changes are not concordant with the other slow growing mutants, e.g. Δpgi (Figure 6, Table S3). Indeed, the slow growing *E. coli* is expected to have impaired glycolysis and anaplerosis²⁸, which was the case for Δpgi but not for *icd_IN* and *gatB_IN* mutants.

Second, the *acnB_IN* strain with mutation in the enzyme catalyzing a TCA cycle reaction converting citrate to isocitrate (Figure 5a) displayed a similar flux ratio phenotype as the Δicd knockout (Figure 6). Although the *acnB_OUT* mutant does not have a strong opposite flux ratio phenotype, the impairment of the TCA fluxes shown

Figure 6. Hierarchical clustering of the flux patterns of *E. coli* knockouts and phosphorylation mutants. Hierarchical clustering was performed for the flux ratio changes to the wild type values using Pearson correlation coefficient as distance metric. The color corresponds to the flux ratio changes between the mutant and the wild type strain. The dashed lines and bold text highlight the clusters discussed in the main text.

by *acnB*_IN mutant concordant with Δ *icd* knockout suggests an inactivating role of aconitase phosphorylation.



Peculiarly, similar flux ratio phenotypes were demonstrated by the phosphoenolpyruvate synthase *pps_IN* mutant and the glucose 6-phosphate dehydrogenase knockout Δzwf catalyzing one of the first step of the PP pathway. These flux phenotypes are characterized by an increased glycolytic and decreased anaplerotic activity (Figure 6). The phosphoenolpyruvate synthase catalyzes a gluconeogenic reaction forming phosphoenolpyruvate from pyruvate, and would not be expected to play a role upon growth on a glycolytic carbon source such as glucose. The similarity of flux patterns between *pps_IN* and Δzwf might indicate the inactivating role of Pps phosphorylation leading to even larger glycolytic flux ratio. On the other hand, the observed changes might be caused by changes in PEP concentrations, since this metabolite is involved in glucose transport. The *pps_OUT* mutant is clustered together with phospho-IN modifications of enolase (Eno) and pyruvate kinase (Pyk), two enzymes involved in lower glycolysis forming PEP and pyruvate (Figures 5a, 6). Their flux phenotypes are described by increase of both glycolytic and anaplerotic flux fractions. This observation suggests an activation function of Eno and Pyk phosphorylation, however, it might as well indicate a series of regulation events caused by changes levels of PEP and pyruvate, the known regulators of *E. coli* central metabolism²⁹.

In summary, we demonstrated the suitability of the high-throughput flux ratio profiling with SUMOFLUX on a set of *E. coli* phosphorylation mutants. By comparing the flux ratio profiles of phosphomutants to each other and knockout strains, we could confirm the inactivation function of isocitrate dehydrogenase phosphorylation and generate hypotheses for several other phosphorylation events. Overall, the established protocol for high-throughput large-scale flux ratio analysis raises the curtain for rapid flux profiling and provides an essential layer of information for integrative multi-omics analysis.

DISCUSSION

In this work, we established an experimental and data analysis workflow for high-throughput flux ratio screening by combining the fast untargeted FIA-TOF measurement platform and previously developed flux ratio analysis method SUMOFLUX. The experimental setup is cost and time efficient. First, for the 96 deep-

well cultivation, about 2 mL of culture medium per sample is required, thus the usage of expensive labeling substrate is drastically reduced compared to the shake flask cultivation, where 50-100 mL of culture medium per sample is used. Second, the metabolite extraction step in 96-well format takes less than 30 min per plate. Third, the measurement with FIA-TOF requires 1 min per sample, therefore one plate can be measured in ~ 1.5 h. The subsequent flux ratio prediction time is negligible. As soon as the SUMOFLUX predictor is built, it can be applied to all the samples simultaneously, which takes no more than a couple of minutes. Thus, with the proposed workflow, a thousand flux analyses can be performed in one day, which is an order of magnitude more than with most commonly used measurement techniques^{17,30}.

Such decrease in experimental time results in a decrease in data quality due to the untargeted nature of the measurements leading to missing or overlapping metabolite ions. Therefore, we proposed a two-step spectral filtering procedure, which eliminates faulty measurements based on the peak shapes and intensity correlation. This filtering is universally applicable to any type of data measured with FIA-TOF. Once the ions are filtered, they are annotated according to the detected mass, hence multiple annotations are unavoidable for the isobaric compounds. Since very few ions are annotated uniquely and, according to our analysis, do not contain necessary information for accurate flux estimates, a more permissive annotation filtering is required. In our experimental setup, we could apply a ratiometric filtering specific for the [1-¹³C] glucose labeling experiment, which was based on the prior assumptions on heavy isotopologue distributions. Although this specific annotation filtering is not applicable in other experimental conditions, we proved that the additional filtering step based on SUMOFLUX simulations, which is generally applicable to any type of data, is sufficient to discard unreliable measurements and lead to the concordant flux estimates.

SUMOFLUX offers several advantages for high-throughput large-scale flux analysis with FIA-TOF data. First, the possibility of using different input features enabled us to use isotopologue ratios instead of mass distribution vectors, which reduced the dependence of several features on one misannotated ion. Second, the flexible feature selection procedure helps to deal with missing data, as non-measured isotopologue ratios are easily eliminated from the predictor training. Third, the extensive data simulation procedure spans over a large cohort of possible flux phenotypes, therefore might serve as an additional filtering step. We proposed to compare the isotopologue ratios of the experimental data to the *in silico* ones, since the

discrepancies between the measured and simulated features likely indicate misannotated or overlapping ions. Of course, such discrepancies can also be caused by errors in the model or inappropriate assumptions on substrate labeling or measurement noise, hence a manual inspection of the corresponding distributions is advisable. Fourth, apart from the obvious advantage of the fast calculation time compared to the global ^{13}C flux analysis methods, SUMOFLUX provides targeted estimates of specified flux ratios, what simplifies the comparison between samples and statistical testing.

One of the experimental advantage of ^{13}C fluxomics screening is the independence between samples, since the flux analysis outcome for each sample depends solely on the ions detected in this sample. In contrast to comparative metabolomics profiling, for example, there is no need for careful culture density measurement and biomass normalization, since the ion intensities of different samples are not compared directly. In the experiment with phosphomutants, some of the strains were growing slowly, and their cell density at the time of sampling was only about OD600 ~ 0.15, what corresponds to < 0.05 mg of cell dry weight per well. Nevertheless, the metabolite measurements with FIA-TOF were sensitive enough to detect intracellular labeling patterns and perform flux ratio estimation for all strains involved in the study. This fact facilitates experimental design and administration and enables flux analysis in slow or non-growing cells.

Our analysis of *E. coli* phosphorylation mutants indicate that some phosphorylation events do not only change the activity of the corresponding enzyme, but also have global regulatory effects on cellular metabolism. Very few enzymes exhibited a strong negative correlation of the flux patterns, and yet two of them exhibited general flux changes. The metabolic and growth phenotypes of the phospho-IN mutations in Icd and GatB were much more severely impaired than in the corresponding knockout strains. Since the studied mutation in the phosphosites only mimics the protein phosphorylation state, it is possible that the global effects caused by these mutations are not linked to phosphorylation, but are caused by protein misfolding. Therefore, for the selected phosphoproteins, assessing their *in vivo* phosphorylation activity might be crucial. One possibility would be to perform ^{13}C flux profiling in several conditions coupled with large-scale phosphoproteomics³¹⁻³³. This approach proved to be useful to identify key phosphorylation events governing metabolic fluxes in yeast³⁴, even though fluxes were analyzed only in three conditions. With the developed high-throughput flux profiling approach, the analysis can be extended to

more media and stress conditions to detect concordant flux ratio changes and phosphorylation events.

Comparison between the metabolic phenotypes of phosphomutants and knockout strains revealed several hypotheses on the phosphorylation function. Whether these events are direct or indirect, still has to be tested. To obtain a more comprehensive picture, it would be desirable to include in the comparison the overexpression strains available for *E. coli*³⁵, to probe the opposite effects of the phosphorylation and investigate the flux changes caused by enzyme overexpression. Moreover, since it has been shown that some transcriptional regulatory events in *E. coli* have a larger impact on the absolute fluxes and not flux partitioning in the network¹⁴, measurement of absolute fluxes would provide additional information. In order to estimate absolute fluxes, flux balance analysis could be performed constrained by the flux ratio estimates obtained by ¹³C flux profiling, and uptake and secretion rates inferred from temporal FIA-TOF measurements of extracellular metabolites. Additionally, intracellular metabolite levels in different strains could be compared to complement the flux ratio comparison³⁶ and identify potential regulatory mechanisms following enzyme phosphorylation, such as interactions of increased or decreased metabolites with distal enzymes in central metabolism.

The proposed experimental and data analysis workflow is not limited to investigation of bacterial metabolism on single substrate media. Similar to the experimental setup described in Chapter 3, it would be possible to perform fluxomic profiling in multi-substrate conditions to study substrate efficiency usage and metabolic adaptations, screen large cohorts of strains relevant for biotechnological applications, or investigate drug effects and metabolic responses to thousands of compounds. The experimental protocol would have to be adapted for other organisms, however, it should be feasible since the high-throughput cultivation methods are already used for yeasts³⁷⁻³⁹ and higher cells^{10,40-43}. Although cellular behavior during culturing in different experimental setups might vary, it has been shown, at least for *E. coli* and *Corynebacterium glutamicum*, that flux distributions in shake flasks and 96 deep-well plates are identical^{44,45}.

While SUMOFLUX-predicted flux ratios were consistent with the expected knockout flux phenotypes in the proof-of-principle experiment, the precise flux ratio estimation using FIA-TOF data cannot be guaranteed. Even after filtering and quality checks, the annotated data might contain faulty ions or high level of noise if the signal

is low. The simulation filtering procedure embedded in the SUMOFLUX workflow may fail to identify discrepant features, or eliminate essentially informative ions in case there are not enough experimental samples to compare the *in silico* and measured isotopologue distributions. We hence recommend to consider SUMOFLUX estimates from FIA-TOF data as semi-quantitative and better suited for comparative studies. Still, the proposed high-throughput fluxomics profiling is a valuable hypotheses generating procedure, which presents a general overview of the flux patterns in all examined samples and identifies the most promising targets for the follow-up studies. Therefore, we believe that SUMOFLUX coupled to the untargeted metabolomics may drastically increase the speed and broaden the applicability of fluxomics, provide additional functional information layer and bring ^{13}C flux screening to the high-throughput level of the other omics techniques.

EXPERIMENTAL PROCEDURES

Strains and media. For the proof-of-principle, *E. coli* wild type K-12 strain BW25113 and the knockouts from the Keio collection were used⁴⁶. For the experiment with phosphorylation mutants, a progenitor wild type strain and the phosphosites mutants were constructed based on a multiplex automated genome engineering method²³ by Harris Wang, Columbia University, USA.

Chemicals. Chemicals, where nothing else is mentioned, were purchased from Sigma Aldrich (Schnelldorf, Switzerland).

Bacterial culturing. Frozen glycerol stocks of *E. coli* strains were used to inoculate agar plates of Luria-Bertani medium. Precultures were inoculated from agar plates in 96 deep-well plates with M9 minimal medium supplemented with either 5 g L^{-1} [$1\text{-}^{13}\text{C}$] glucose, or a mixture of 2.5 g L^{-1} [$\text{U-}^{13}\text{C}$] glucose and 2.5 g L^{-1} unlabeled glucose. Mid-exponentially M9 precultures at optical densities at 600 nm (OD_{600}) of 1-2 were then used to inoculate 96 deep-well plates with 1.2 mL M9 minimal medium to an OD_{600} of 0.05. In case of slow growing precultures, the cultures were inoculated to a maximal possible OD_{600} . All cultures were grown at 37°C under constant shaking at 300 rpm. Mutants were grown in triplicates on the 96-well plate while wild type was inoculated in triplicate in each plate.

The M9 medium contained: Na₂HPO₄ 2H₂O 211 mM, KH₂PO₄ 110 mM, NaCl 42.8 mM, (NH₄)₂SO₄ 56.7 mM, MgSO₄ 100mM, CaCl₂ 10mM, FeCl₃ 10mM, 500 x Thiamine-HCl 1.4 mM, ZnCl₂ 0.63 mM, MnSO₄ 0.71 mM, CoCl₂ 0.76 mM, CuCl₂ 0.7 mM. The prepared media were filter sterilized.

Metabolite sampling. Cell growth was monitored with the Tecan Sunrise Plate Reader by measuring OD₆₀₀. When cell density reached OD₆₀₀ ~ 1, the 96 deep-well plate was centrifuged (0° C, 4000 rpm, 2 min) and supernatant was discarded by tapping against a paper tissue. To each well, 180 uL of 60% EtOH at 80° C was added and the plate was boiled for 3 min in 80° C water bath with 2 vortexing steps. After extraction, the plate was centrifuged (0° C, 4000 rpm, 5 min) and supernatant was transferred to a 96-well storage plate AB 1058 (Thermo Scientific). The plates were kept at -80° C until measurement.

FIA-TOF measurement and ion annotation. Metabolites were measured by direct flow double injection of extracts on an Agilent 6550 series quadrupole TOF MS. Ions within a mass/charge ratio range of 50-1000 Da/e were measured in negative mode. The ions were annotated using the KEGG list of *E. coli* metabolites⁴⁷. Only ions corresponding to deprotonated metabolites were considered for further analysis.

All data analysis was conducted using Matlab (The Mathworks, Natick).

FIA-TOF data processing. Spectral filtering. For the ringing artifacts filtering, Pearson correlation coefficients across all measured samples were calculated for small intensity peaks in the proximity of 0.3 Da/e of a high intensity peak detected earlier. Correlation threshold of 0.9 was set to eliminate faulty peaks. For the Gaussian shape filtering, a maximum of 5 peaks detected at the maximum distance of 0.1 Da/e from the peak of interest were used to represent the spectrum as a sum of Gaussian functions. Matlab function fit was called with parameters of peak locations, peak height and peak width. Peak locations and heights were defined in the peak picking procedure, whereas peak width was estimated with the Matlab function findpeaks. The difference between the restored and measured peak intensity, as well as the difference between the single Gaussian peak fitting and measured intensity was calculated. A threshold of 0.1 of the measured intensity was set to define poor fit or large discrepancy between the measured and restored intensity to eliminate faulty peaks.

Ratiometric filtering. For the ratiometric filtering of data detected in [1-¹³C] glucose experiment, for each detected metabolite the ratio between the m+N isotopologue (N>2) and m+0 isotopologue was calculated. A threshold of 0.33 was set to eliminate potentially overlapping heavy isotopologues. For the quality check, fractional labeling for the data detected in [U-¹³C] experiment was calculated as the fraction of labeled atoms in the total pool of metabolites atoms.

SUMOFUX analysis. Network construction. *E. coli* metabolic network with carbon atom transitions and the lists of input and output metabolites were defined (Table S3). In order to reduce the dependency on the biomass vector coefficients, a separate output flux is defined for each of the biomass precursors, therefore biomass precursors are also added to the list of outputs. The substrates are defined as unbalanced compounds and do not participate in the stoichiometric equation system.

Flux sampling and ratio calculation. In the flux sampling procedure, the definitions of net, exchange, forwards and backward fluxes are used⁴⁸. By default, the lower and upper bounds for reversible reactions are set to [-100 100], for irreversible reactions to [0 100], and the major uptake flux is set to 10. To achieve uniform coverage of values for a particular flux ratio or set of ratios, the ratio range was split into segments ([0 0.25], [0.25 0.5], [0.5 0.75], [0.75 1]), and the flux sampling procedure is repeated for each segment with the end points set as flux ratio constraints in the first step.

The flux fraction for each studied metabolite from a specific reaction was defined as the ratio between the flux from this reaction and sum of all producing reactions of this metabolite. Surrogate modelling of the labeling data, training and testing the predictor, feature selection, cross validation procedure and quantile calculations were performed as described in Chapter 2.

Flux ratio comparison and clustering. Median predictions of the flux ratios were compared for each mutant and corresponding wild type strain estimates. Difference between the flux ratios were computed, and its significance was assessed with two-sample unpaired t-test with unequal variances. P-values were subsequently corrected for multiple hypotheses testing by calculating the false discovery rate with Benjamini-Hochberg procedure. A threshold of 0.05 was set to indicate small changes, a threshold of 0.1 was considered as a large change.

Pearson correlation coefficient of the flux ratio difference to the wild type was calculated for each pair of phospho-IN and phospho-OUT mutant with Matlab function `corr`.

Hierarchical clustering was performed with Matlab function `clustergram` on the matrix of mutants and flux ratio difference to wild type. Pearson correlation was used as distance metric for the mutants, and Euclidean distance as a metric for the flux ratios.

REFERENCES

- 1 Zamboni, N. ^{13}C metabolic flux analysis in complex systems. *Current opinion in biotechnology* **22**, 103-108, doi:10.1016/j.copbio.2010.08.009 (2011).
- 2 Buescher, J. M. *et al.* A roadmap for interpreting ^{13}C metabolite labeling patterns from cells. *Current Opinion in Biotechnology* **34**, 189-201, doi:10.1016/j.copbio.2015.02.003 (2015).
- 3 Wittmann, C. & Heinzle, E. Application of MALDI-TOF MS to lysine-producing *Corynebacterium glutamicum*: a novel approach for metabolic flux analysis. *European journal of biochemistry / FEBS* **268**, 2441-2455 (2001).
- 4 Rühl, M. *et al.* Collisional fragmentation of central carbon metabolites in LC-MS/MS increases precision of ^{13}C metabolic flux analysis. *Biotechnology and bioengineering* **109**, 763-771, doi:10.1002/bit.24344 (2012).
- 5 Choi, J. & Antoniewicz, M. R. Tandem mass spectrometry: a novel approach for metabolic flux analysis. *Metabolic engineering* **13**, 225-233 (2011).
- 6 Boisseau, R. *et al.* Fast spatially encoded 3D NMR strategies for ^{13}C -based metabolic flux analysis. *Analytical chemistry* **85**, 9751-9757 (2013).
- 7 Zamboni, N., Fendt, S.-M., Rühl, M. & Sauer, U. ^{13}C -based metabolic flux analysis. *Nature protocols* **4**, 878-892, doi:10.1038/nprot.2009.58 (2009).
- 8 Antoniewicz, M. R., Kelleher, J. K. & Stephanopoulos, G. Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions. *Metabolic engineering* **9**, 68-86, doi:10.1016/j.ymben.2006.09.001 (2007).
- 9 Weitzel, M. *et al.* 13CFLUX2 - high-performance software suite for ^{13}C -metabolic flux analysis. *Bioinformatics (Oxford, England)* **29**, 143-145, doi:10.1093/bioinformatics/bts646 (2013).
- 10 Balcarcel, R. R. & Clark, L. M. Metabolic screening of mammalian cell cultures using well-plates. *Biotechnology progress* **19**, 98-108 (2003).
- 11 Tang, Y. J. *et al.* Advances in analysis of microbial metabolic fluxes via ^{13}C isotopic labeling. *Mass spectrometry reviews* **28**, 362-375 (2009).
- 12 Fischer, E. & Sauer, U. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nature genetics* **37**, 636-640, doi:10.1038/ng1555 (2005).
- 13 Blank, L. M., Kuepfer, L. & Sauer, U. Large-scale ^{13}C -flux analysis reveals mechanistic principles of metabolic network robustness to null mutations in yeast. *Genome biology* **6**, R49-R49, doi:10.1186/gb-2005-6-6-r49 (2005).
- 14 Haverkorn van Rijsewijk, B. R. B., Nanchen, A., Nallet, S., Kleijn, R. J. & Sauer, U. Large-scale ^{13}C -flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Molecular systems biology* **7**, 477-477, doi:10.1038/msb.2011.9 (2011).

- 15 Fendt, S.-M. *et al.* Unraveling condition-dependent networks of transcription factors that control metabolic pathway activity in yeast. *Molecular systems biology* **6**, 432-432, doi:10.1038/msb.2010.91 (2010).
- 16 Niedenführ, S., Wiechert, W. & Nöh, K. How to measure metabolic fluxes: a taxonomic guide for ¹³C fluxomics. *Current opinion in biotechnology*, doi:10.1016/j.copbio.2014.12.003 (2015).
- 17 Heux, S., Poinot, J., Massou, S., Sokol, S. & Portais, J.-C. A novel platform for automated high-throughput fluxome profiling of metabolic variants. *Metabolic engineering* **25**, 8-19 (2014).
- 18 Fuhrer, T., Heer, D., Begemann, B. & Zamboni, N. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Analytical chemistry* **83**, 7074-7080, doi:10.1021/ac201267k (2011).
- 19 Kyriakis, J. M. In the beginning, there was protein phosphorylation. *Journal of Biological Chemistry* **289**, 9460-9462 (2014).
- 20 Macek, B. *et al.* Phosphoproteome analysis of *E. coli* reveals evolutionary conservation of bacterial Ser/Thr/Tyr phosphorylation. *Molecular & cellular proteomics* **7**, 299-307 (2008).
- 21 Chubukov, V., Gerosa, L., Kochanowski, K. & Sauer, U. Coordination of microbial metabolism. *Nature Reviews Microbiology* **12**, 327-340 (2014).
- 22 Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262-1278 (2014).
- 23 Wang, H. H. *et al.* Programming cells by multiplex genome engineering and accelerated evolution. *Nature* **460**, 894-898 (2009).
- 24 Dean, a. M., Lee, M. H. I. & Koshland, D. E. Phosphorylation inactivates *Escherichia coli* isocitrate dehydrogenase by preventing isocitrate binding. *Journal of Biological Chemistry* **264**, 20482-20486 (1989).
- 25 Fischer, E. & Sauer, U. Metabolic flux profiling of *Escherichia coli* mutants in central carbon metabolism using GC-MS. *European Journal of Biochemistry* **270**, 880-891, doi:10.1046/j.1432-1033.2003.03448.x (2003).
- 26 Hua, Q., Yang, C., Baba, T., Mori, H. & Shimizu, K. Responses of the central metabolism in *Escherichia coli* to phosphoglucose isomerase and glucose-6-phosphate dehydrogenase knockouts. *Journal of bacteriology* **185**, 7053-7067, doi:10.1128/JB.185.24.7053 (2003).
- 27 Nobelmann, B. & Lengeler, J. W. Molecular analysis of the gat genes from *Escherichia coli* and of their roles in galactitol transport and metabolism. *Journal of bacteriology* **178**, 6790-6795 (1996).
- 28 Fischer, E. & Sauer, U. A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry *Escherichia coli*. *Journal of Biological Chemistry* **278**, 46446-46451 (2003).
- 29 Link, H., Kochanowski, K. & Sauer, U. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity *in vivo*. *Nature biotechnology* **31**, 357-361, doi:10.1038/nbt.2489 (2013).
- 30 Fuhrer, T. & Zamboni, N. High-throughput discovery metabolomics. *Current opinion in biotechnology* **31**, 73-78 (2015).
- 31 Jünger, M. A. & Aebersold, R. Mass spectrometry-driven phosphoproteomics: patterning the systems biology mosaic. *Wiley Interdisciplinary Reviews: Developmental Biology* **3**, 83-112 (2014).
- 32 Engholm-Keller, K. & Larsen, M. R. Technologies and challenges in large-scale phosphoproteomics. *Proteomics* **13**, 910-931 (2013).
- 33 Zhou, H., Watts, J. D. & Aebersold, R. A systematic approach to the analysis of protein phosphorylation. *Nature biotechnology* **19**, 375-378 (2001).

- 34 Oliveira, A. P. *et al.* Regulation of yeast central metabolism by enzyme phosphorylation. *Molecular Systems Biology* **8**, 623 (2012).
- 35 Kitagawa, M. *et al.* 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 research* **12**, 291-299 (2006).
- 36 Schulz, J. C., Zampieri, M., Wanka, S., von Mering, C. & Sauer, U. Large-scale functional analysis of the roles of phosphorylation in yeast metabolic pathways. *Sci Signal* **7**, 1-13 (2014).
- 37 Back, A., Rossignol, T., Krier, F., Nicaud, J.-M. & Dhulster, P. High-throughput fermentation screening for the yeast *Yarrowia lipolytica* with real-time monitoring of biomass and lipid production. *Microbial Cell Factories* **15**, 147 (2016).
- 38 Allen, J. *et al.* High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nature biotechnology* **21**, 692-696 (2003).
- 39 Heux, S., Fuchs, T. J., Buhmann, J., Zamboni, N. & Sauer, U. A high-throughput metabolomics method to predict high concentration cytotoxicity of drugs from low concentration profiles. *Metabolomics* **8**, 433-443 (2012).
- 40 Ovcharenko, D., Jarvis, R., Hunicke-Smith, S., Kelnar, K. & Brown, D. High-throughput RNAi screening in vitro: from cell lines to primary cells. *Rna* **11**, 985-993 (2005).
- 41 Mousses, S. *et al.* RNAi microarray analysis in cultured mammalian cells. *Genome research* **13**, 2341-2347 (2003).
- 42 Huch, M. *et al.* Long-term culture of genome-stable bipotent stem cells from adult human liver. *Cell* **160**, 299-312 (2015).
- 43 Sato, T. & Clevers, H. Primary mouse small intestinal epithelial cell cultures. *Epithelial Cell Culture Protocols: Second Edition*, 319-328 (2013).
- 44 Fischer, E. High-throughput metabolic flux analysis based on gas chromatography–mass spectrometry derived ¹³C constraints. *Analytical Biochemistry* **325**, 308-316, doi:10.1016/j.ab.2003.10.036 (2004).
- 45 Wittmann, C., Kim, H. M. & Heinzle, E. Metabolic network analysis of lysine producing *Corynebacterium glutamicum* at a miniaturized scale. *Biotechnology and bioengineering* **87**, 1-6 (2004).
- 46 Baba, T. *et al.* Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* **2** (2006).
- 47 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids research* **28**, 27-30 (2000).
- 48 Wiechert, W. & de Graaf, a. a. Bidirectional reaction steps in metabolic networks: I. Modeling and simulation of carbon isotope labeling experiments. *Biotechnology and bioengineering* **55**, 101-117, doi:10.1002/(SICI)1097-0290(19970705)55:1<101::AID-BIT12>3.0.CO;2-P (1997).

Chapter 5

Concluding remarks

Approach. Stable isotope tracing is the most informative and direct way to estimate metabolic fluxes, which is essential for understanding and controlling metabolic behavior and adaptation to novel environments. Metabolic fluxes have to be inferred from the labeling data with computational modeling or analytic equations. Existing global ^{13}C metabolic flux analysis methods employ principles of constrained metabolic balancing to provide a flux distribution that best fits the labeling data. Therefore, extensive measurements of labeling data and physiological parameters, or even combination of parallel labeling experiments, are required to obtain the optimal flux distribution. Although generally applicable, these methods are poorly scalable due to the computational and experimental costs. Classical local ^{13}C flux ratio analysis with analytic formulas, on the contrary, is easily applicable to large cohorts of data and requires only a few metabolite measurements. However, these formulas are based on strong assumptions, valid under specified experimental conditions, and require expert knowledge to be derived. Due to the lack of a fast, data-flexible and generally applicable ^{13}C metabolic flux analysis approach, ^{13}C fluxomics has not yet gained the widespread use and throughput as omics techniques¹.

In this thesis, we developed SUMOFLUX, a generalized targeted ^{13}C metabolic flux ratio analysis method that combines the universality of global flux analysis and scalability of local flux ratio formulas. SUMOFLUX is a conceptually different approach that, for the first time to our knowledge, exploits machine learning to predict metabolic flux ratios from the ^{13}C labeling data. Since machine learning requires a comprehensive training dataset, which in this case is not accessible experimentally, we employ **surrogate modeling** to generate thousands of fluxes and corresponding labeling patterns, what is justified by the fact that each flux vector unambiguously defines the labeling patterns of intracellular metabolites². The prediction quality depends on the *in silico* training set, which should adequately describe the experimental setup. The training set is simulated given the information on the network structure, substrate label, and uptake and secretion fluxes. In case no physiological measurements are available, the sampling is performed to ensure a broad coverage of feasible range of exchange reactions and corresponding flux distributions. As a simple quality check of the *in silico* dataset, the simulated and measured isotopologue distributions can be compared in order to identify discrepancies, which indicate either errors in the model or inaccurate measurements. The latter can be corrected by omitting the discrepant features from the input list; the errors in the model, however, have to be curated by the user.

The combination of surrogate modeling with machine learning offers several advantages. First, SUMOFLUX allows to investigate any calculable ratio of fluxes given any measurable labeling data for any metabolic network and ^{13}C tracer configuration that can be simulated. Since alternative reactions might produce similar labeling patterns of the downstream metabolites, an accurate resolvability of a flux ratio of interest cannot be guaranteed. However, the *in silico* testing procedure allows to assess which flux ratios of interest can be resolved under given conditions, therefore helps to extract the most flux information from the available data. Second, in case of poor *in silico* resolvability, alternative experimental setups can be tested to optimize experimental design by varying the ^{13}C tracer or measurement data. The simulation time for the different setups can be reduced since already a small test dataset is sufficient to calculate the predictor error, which is a specific and easily interpretable metric. Third, combining data from parallel labeling experiments, which has been shown to improve flux estimation^{3,4}, is easily implemented by merging the data simulated for different substrate labels. A large number of input features might cause predictor overfitting, but this can be circumvented with the embedded feature selection procedure. Fourth, once the flux ratio predictor is trained, it is rapidly applied to any number of samples of interest, making SUMOFLUX particularly easy to scale.

High-throughput ^{13}C flux analysis. The scalability of SUMOFLUX together with its flexibility in terms of input data enabled us to develop a fast high-throughput experimental workflow for ^{13}C flux profiling using the untargeted metabolomics platform FIA-TOF (flow injection analysis – time of flight)⁵. With this workflow, about a thousand of flux ratio analyses can be routinely performed per day, assuming the labeled samples are available. Although the data generation with the untargeted FIA-TOF platform is fast, it is prone to potential ion overlaps and missing values. Therefore, we proposed rigorous filtering procedures and feature modification, which improved the data quality and proved to be sufficient to capture expected flux ratio phenotypes for a set of *E. coli* knockouts strains.

With the developed high-throughput workflow we performed functional profiling of a cohort of 86 *E. coli* strains with mutations in putative phosphorylation sites and corresponding knockout mutants. Comparison of flux ratio changes in central metabolism upon growth on glucose confirmed the reported deactivating function of isocitrate dehydrogenase phosphorylation, and suggested functional roles of several

previously uncharacterized phosphorylation events. Compared to the standard functional assays based on growth phenotypes or more detailed metabolomics analysis⁶, ¹³C flux profiling provides a more sensitive and interpretable readout of the functional changes. Since most phosphorylation events are functional only under specific circumstances, the performed screen should be expanded by including more experimental conditions and flux ratios in the analysis. Although the flux ratio estimates obtained from FIA-TOF labeling data may be semi-quantitative due to the high noise and overlaps in the data, the proposed high-throughput ¹³C flux profiling provides a valuable hypotheses generation tool.

Complex nutritional environments. In the natural environments, microorganisms often encounter a variety of nutrients, which they utilize either simultaneously or sequentially depending on the cellular needs and capacities. Flux analysis in composite media is complicated by the increased network complexity and often requires more measurements. The targeted approach offers an advantage of extracting the valuable information from limited amount of data or poorly determined networks.

We applied SUMOFLUX to characterize amino acid utilization in mycobacteria growing in composite media and infecting macrophage-like THP-1 cells. Although during infection the pathogens encounter nutrient-limited phagosomal environment⁷, there is emerging evidence that *Mycobacterium tuberculosis* utilizes multiple scarcely available nutrients, among which amino acids play an important role⁸. By formulating specific flux ratios, we quantified amino acid utilization patterns, classified them based on their influence on central metabolism, and compared the flux changes upon THP-1 infection to the growth in rich medium. In contrast to the global ¹³C flux spectral analysis, which could only provide binary uptake characteristic for a couple of amino acids⁸, the targeted flux ratio analysis provided quantitative estimates for the biosynthesis fractions of all measured amino acids. These results underline the incidence of multiple carbon and nitrogen sources inside the phagosome and mycobacterial ability to utilize them as soon as they become available, what partially explains recurrent failures of targeted therapeutic interventions.

Outlook. In this work, we exploited SUMOFLUX to analyze flux ratios in relatively small models of bacterial or dual bacterial-host central carbon metabolism. In principle, this approach can already be expanded to resolve flux ratios in networks of larger size and complexity. The scalability of SUMOFLUX to genome-scale models primarily depends on the possibility of ^{13}C data simulation for the training. Genome-scale ^{13}C metabolic flux analysis has already been performed in *E. coli* by reducing the network and using efficient data decomposition⁹. Although the computation time of genome-scale simulations to date remains a major challenge, with the advances of parallel computing it might be solved in the near future. However, it has been shown that the fluxes in the genome-scale model cannot be resolved with data from a single tracer experiment, thus multiple parallel labeling strategies are required to resolve specific branch points⁴. With the facilitated procedure for experimental design, SUMOFLUX might become the method of choice to select the optimal labeling strategies elucidating the key metabolic crossroads in genome-scale models.

Optimal tracer design alone is not sufficient to resolve fluxes in alternate pathways with the same atom transitions. In this case, only non-stationary flux analysis methods that follow dynamic label propagation might reveal the flux partitioning¹⁰. Non-stationary methods are highly informative, especially when the isotopic steady state cannot be guaranteed due to transient growth or sequentially consumed nutrients, but are highly computationally and experimentally demanding¹⁰. In principle, the concept of SUMOFLUX can be extended to isotopically non-stationary data. In case a comprehensive training dataset consisting of dynamic isotopologue profiles and corresponding flux ratios is available, the flux ratio prediction for the experimental data can be performed at high speed and throughput analogous to the stationary case. The simulation of non-stationary training data is substantially more demanding because of the requirement to sample an increased number of degrees of freedom and measurable labeling features. Moreover, non-stationary labeling experiments are much more labor-intensive and data demanding, and can be performed only at low throughput¹¹. For practical reasons, the traditional approach of flux estimation by both global^{11,12} or local¹³ iterative fitting is better suited to the analysis of small-scale non-stationary labeling experiments.

Due to its generalized nature, SUMOFLUX is also applicable to resolve fluxes in higher cells. Higher cells metabolism features consumption of multiple nutrients, therefore global flux estimation from a single labeling experiment becomes

impossible¹⁴. Local flux analysis methods are therefore more practical to test specific hypotheses, since they usually require less data and computations. Emerging approaches specialize in resolving flux nodes of relevance, such as detailed analysis of the tricarboxylic acid fluxes¹⁵, NADPH production¹⁶, or even direct assessment of anaplerotic fluxes in human¹⁷. The major challenge of higher cell flux analysis is that measured labeled compounds might represent an average of differently labeled molecules produced in different cellular compartments. With SUMOFLUX, this effect can be modeled in the training dataset by summarizing the simulated labeling patterns of the same metabolite produced by multiple reactions. Although in this case the accuracy of predictions might decrease, the *in silico* testing unravels the flux ratios which are theoretically resolvable under given conditions and those which are not.

The high throughput workflow combining FIA-TOF data acquisition with SUMOFLUX analysis provides a useful tool for ¹³C fluxomics screening. In a similar fashion as for functional profiling of *E. coli* phosphomutants, it is readily applicable to study metabolic changes in central metabolism of industrial strains, or flux rearrangements caused by stress or drug exposure. In contrast to the commonly used phenotypic screening, which does not provide mechanistic information¹⁸, ¹³C flux profiling might shed light on the modes of action of uncharacterized drugs, and even suggest combination therapies. However, in its current form, SUMOFLUX workflow estimates flux ratios defined by the user, which requires prior knowledge on their relevance. The untargeted data acquisition with FIA-TOF, in the meantime, offers a powerful tool for biological discovery¹⁹. Recently, untargeted stable isotope metabolomics advanced drug development by elucidating altered metabolic pathways in exposed cell lines or patient-derived primary cells²⁰⁻²². The underlying metabolic changes are analyzed in a qualitative way by visual inspection of changing isotopologues and assigning them to metabolic pathways. We believe that SUMOFLUX can benefit the untargeted ¹³C data analysis by flux ratio quantification in the proximity of the changing metabolites. Potentially, the high-throughput SUMOFLUX workflow can be integrated with data-driven metabolic subnetwork extraction²³ for subsequent targeted ¹³C flux ratio analysis. Such non-targeted approaches have great potential to reveal hitherto overlooked metabolic flux changes and provide means to get closer to the full picture²⁴.

REFERENCES

- 1 Zamboni, N. ^{13}C metabolic flux analysis in complex systems. *Current opinion in biotechnology* **22**, 103-108, doi:10.1016/j.copbio.2010.08.009 (2011).
- 2 Wiechert, W. & Graaf, A. A. D. Bidirectional reaction steps in metabolic networks part I: Modelling and simulation of carbon isotope labelling experiments. *Biotechnology and bioengineering* **55**: 101-117 (1997).
- 3 Crown, S. B., Long, C. P. & Antoniewicz, M. R. Integrated ^{13}C -metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. *Metabolic engineering* **28**, 151-158 (2015).
- 4 Leighty, R. W. & Antoniewicz, M. R. COMPLETE-MFA: Complementary parallel labeling experiments technique for metabolic flux analysis. *Metabolic Engineering* **20**, 49-55, doi:10.1016/j.ymben.2013.08.006 (2013).
- 5 Fuhrer, T., Heer, D., Begemann, B. & Zamboni, N. High-throughput, accurate mass metabolome profiling of cellular extracts by flow injection-time-of-flight mass spectrometry. *Analytical chemistry* **83**, 7074-7080, doi:10.1021/ac201267k (2011).
- 6 Schulz, J. C., Zampieri, M., Wanka, S., von Mering, C. & Sauer, U. Large-scale functional analysis of the roles of phosphorylation in yeast metabolic pathways. *Sci Signal* **7**, 1-13 (2014).
- 7 Kwaik, Y. A. & Bumann, D. Host delivery of favorite meals for intracellular pathogens. *PLoS Pathog* **11**, e1004866 (2015).
- 8 Beste, D. J. V. *et al.* ^{13}C -Flux spectral analysis of host-pathogen metabolism reveals a mixed diet for intracellular *Mycobacterium tuberculosis*. *Chemistry & biology* **20**, 1012-1021, doi:10.1016/j.chembiol.2013.06.012 (2013).
- 9 Gopalakrishnan, S. & Maranas, C. D. ^{13}C Metabolic flux analysis at a genome-scale. *Metabolic Engineering*, doi:10.1016/j.ymben.2015.08.006 (2015).
- 10 Wiechert, W. & Nöh, K. Isotopically non-stationary metabolic flux analysis: complex yet highly informative. *Current Opinion in Biotechnology* **24**, 979-986, doi:10.1016/j.copbio.2013.03.024 (2013).
- 11 Nöh, K., Wahl, A. & Wiechert, W. Computational tools for isotopically instationary ^{13}C labeling experiments under metabolic steady state conditions. *Metabolic Engineering* **8**, 554-577, doi:10.1016/j.ymben.2006.05.006 (2006).
- 12 Young, J. D., Walther, J. L., Antoniewicz, M. R. & Yoo, H. An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis. *Biotechnology* **99**, 686-699, doi:10.1002/bit (2008).
- 13 Hörl, M., Schnidder, J., Sauer, U. & Zamboni, N. Non-stationary ^{13}C -metabolic flux ratio analysis. *Biotechnology and bioengineering* **9999**, 1-13, doi:10.1002/bit.25004 (2013).
- 14 Walther, J. L., Metallo, C. M., Zhang, J. & Stephanopoulos, G. Optimization of ^{13}C isotopic tracers for metabolic flux analysis in mammalian cells. *Metabolic Engineering* **14**, 162-171, doi:10.1016/j.ymben.2011.12.004.Optimization (2012).
- 15 Alves, T. C. *et al.* Integrated, step-wise, mass-isotopomeric flux analysis of the TCA cycle. *Cell metabolism* **22**, 936-947 (2015).
- 16 Fan, J. *et al.* Quantitative flux analysis reveals folate-dependent NADPH production. *Nature* **510**, 298-302, doi:10.1038/nature13236 (2014).
- 17 Befroy, D. E. *et al.* Direct assessment of hepatic mitochondrial oxidative and anaplerotic fluxes in humans using dynamic ^{13}C magnetic resonance spectroscopy. *Nature medicine* **20**, 98-102, doi:10.1038/nm.3415 (2014).
- 18 Kell, D. B. Finding novel pharmaceuticals in the systems biology era using multiple effective drug targets, phenotypic screening and knowledge of transporters: where drug discovery went wrong and how to fix it. *FEBS Journal* **280**, 5957-5980 (2013).
- 19 Sévin, D. C., Kuehne, A., Zamboni, N. & Sauer, U. Biological insights through nontargeted metabolomics. *Current opinion in biotechnology* **34**, 1-8 (2015).

- 20 Weindl, D. *et al.* Bridging the gap between non-targeted stable isotope labeling and metabolic flux analysis. *Cancer & metabolism* **4**, 1 (2016).
- 21 Lane, A. N., Higashi, R. M. & Fan, T. W.-M. Preclinical models for interrogating drug action in human cancers using Stable Isotope Resolved Metabolomics (SIRM). *Metabolomics* **12**, 1-15 (2016).
- 22 Fan, T. W.-M. *et al.* Stable isotope-resolved metabolomics and applications for drug development. *Pharmacology & therapeutics* **133**, 366-391 (2012).
- 23 Jourdan, F. *et al.* Use of reconstituted metabolic networks to assist in metabolomic data visualization and mining. *Metabolomics* **6**, 312-321 (2010).
- 24 Weindl, D., Wegner, A. & Hiller, K. Metabolome-wide analysis of stable isotope labeling—is it worth the effort? *Frontiers in physiology* **6** (2015).

Appendix I

Supplementary information Chapter 2

Table S1. *E. coli* network of central carbon metabolism used throughout the study.

Reaction type	Reaction	Stoichiometry and carbon transitions
Uptake and secretion	glc_up	glucose (abcdef) -> G6P (abcdef)
	CO2up	CO2in (a) -> CO2 (a)
	accoa_ac	AcCoA (ab) -> Ac (ab)
	ac_out	Ac (ab) -> Acetate (ab)
	co2_out	CO2 (a) -> CO2out (a)
Biomass precursors	G6P_bm	G6P (abcdef) -> G6Pbm (abcdef)
	PGA_bm	PGA (abc) -> PGAbm (abc)
	P5P_bm	P5P (abcde) -> P5Pbm (abcde)
	PEP_bm	PEP (abc) -> PEPbm (abc)
	PYR_bm	PYR (abc) -> PYRbm (abc)
	OGA_bm	OGA (abcde) -> OGAbm (abcde)
	OAA_bm	OAA (abcd) -> OAAbm (abcd)
	E4P_bm	E4P (abcd) -> E4Pbm (abcd)
Glycolysis	pgi	G6P (abcdef) <-> F6P (abcdef)
	pfk	F6P (abcdef) -> FBP (abcdef)
	fba	FBP (abcdef) -> DHAP (cba) + GAP (def)
	tpi	DHAP (abc) <-> GAP (abc)
	gapdh	GAP (abc) -> BPG (abc)
	bpg	BPG (abc) -> PGA (abc)
	eno	PGA (abc) -> PEP (abc)
	pyk	PEP (abc) -> PYR (abc)
Pentose Phosphate Pathway	zwf	G6P (abcdef) -> PG6 (abcdef)
	gnd	PG6 (abcdef) -> P5P (bcdef) + CO2 (a)
	TK1	P5P (abcde) + P5P (fghij) <-> GAP (cde) + S7P (abfghij)
	TK2	P5P (abcde) + E4P (fghi) <-> GAP (cde) + F6P (abfghi)
	TA	S7P (abcdefg) + GAP (hij) <-> E4P (defg) + F6P (abchij)
Entner-Doudoroff	edp1	PG6 (abcdef) -> KDPG (abcdef)
	edp2	KDPG (abcdef) -> PYR (abc) + GAP (def)
TCA cycle	pdh	PYR (abc) -> AcCoA (bc) + CO2 (a)
	citl	OAA (cdef) + AcCoA (ab) -> Cit (fedcba)
	idh	Cit (abcdef) <-> OGA (abcef) + CO2 (d)
	sdh	OGA (abcde) -> Suc (bcde) + CO2 (a)
	fum	Suc (abcd) <-> Mal (abcd)
	mdh	Mal (abcd) <-> OAA (abcd)
Glyoxylate shunt	gs1	Cit (abcdef) -> Glx (ab) + Suc (dcef)
	gs2	Glx (ab) + AcCoA (cd) -> Mal (abcd)
Anaplerosis and gluconeogenesis	mae	Mal (abcd) -> PYR (abc) + CO2 (d)
	pyc	PEP (abc) + CO2 (d) -> OAA (abcd)
	pck	OAA (abcd) -> PEP (abc) + CO2 (d)

Table S2. Metabolites and metabolite fractions inferable from amino-acid measurements with GC-MS used in *E. coli* and *B. subtilis* studies. The metabolites and metabolite fragments were simulated in the in silico dataset for *E. coli* and *B. subtilis*.

Metabolite	Abbreviation	Fragment carbon positions
Oxoglutarate	OGA15	1 2 3 4 5
Oxoglutarate	OGA25	2 3 4 5
Oxoglutarate	OGA12	1 2
Phosphoenolpyruvate	PEP13	1 2 3
Phosphoenolpyruvate	PEP23	2 3
Phosphoenolpyruvate	PEP12	1 2
Erythrose 4-phosphate	E4P14	1 2 3 4
Biphosphoglycerate	BPG13	1 2 3
Biphosphoglycerate	BPG23	2 3
Biphosphoglycerate	BPG12	1 2
Pyruvate	PYR13	1 2 3
Pyruvate	PYR23	2 3
Pyruvate	PYR12	1 2
Oxaloacetate	OAA14	1 2 3 4
Oxaloacetate	OAA24	2 3 4
Oxaloacetate	OAA12	1 2
Acetyl-CoA	AcCoA12	1 2
Acetyl-CoA	AcCoA2	2
Ribose-5-phosphate	P5P15	1 2 3 4 5
Ribose-5-phosphate	P5P25	2 3 4 5
Ribose-5-phosphate	P5P12	1 2

Table S3. Analytic formulas used to calculate flux ratios in central carbon metabolism of *E. coli* and *B. subtilis*.

Flux ratio	Analytic formula for <i>E. coli</i>	Analytic formula for <i>B. subtilis</i>
Glycolysis / Pentose-phosphate pathway	$F = \frac{BPG_{13} - N_{13}}{0.5L_{13} \times 0.5N_{13} - N_{13}}$	$F = \frac{BPG_{13} - N_{13}}{0.5L_{13} \times 0.5M_{13} - M_{13}}$
Pyruvate from Entner-Doudoroff pathway	$F = \frac{PYR_{13} - BPG_{13}}{L_{13} - BPG_{13}}$	n/a
Oxaloacetate from anaplerosis	$\begin{bmatrix} F \\ F * CO_2 \end{bmatrix} = \frac{OAA_{14} - \alpha KG_{25}}{\begin{bmatrix} PEP_{13} & 0 \\ 0 & PEP_{13} \end{bmatrix} \begin{bmatrix} \alpha KG_{25} \\ PEP_{13} \end{bmatrix}}$	$\begin{bmatrix} F \\ F * CO_2 \end{bmatrix} = \frac{OAA_{14} - \alpha KG_{25}}{\begin{bmatrix} PYR_{13} & 0 \\ 0 & PYR_{13} \end{bmatrix} \begin{bmatrix} \alpha KG_{25} \\ PYR_{13} \end{bmatrix}}$
PEP from gluconeogenesis	$F = \frac{PEP_{12} - BPG_{12}}{OAA_{12} - BPG_{12}}$	$F = \frac{PEP_{12} - BPG_{12}}{OAA_{12} - BPG_{12}}$
Pyruvate from malate	$F = \frac{PYR_{23} - PEP_{23}}{M_1 \times M_1 - PEP_{23}}$	$F = \frac{PYR_{23} - PEP_{23}}{M_1 \times M_1 - PEP_{23}}$

M1-one carbon molecule labeled according to the substrate labeling

M13-last three carbon molecule labeled according to the substrate labeling

N13-three carbon molecule naturally labeled (C13 natural abundance 1%)

L13- first three carbon molecule fragment of the substrate propagated through glycolysis

α KG – α -ketoglutarate

BPG – biphosphoglycerate

OAA – oxaloacetate

PEP – phosphoenolpyruvate

PYR – pyruvate

Numbers in subscript indicate metabolic fragment

Table S4. SUMOFLUX predictions and analytic formulas' estimates of five flux ratios in central metabolism of *E. coli* for eight strains grown on 100% [1-¹³C] or 20% [U-¹³C] and 80% naturally labeled glucose; and for 121 *B. subtilis* strains grown on combination of 80% [1-¹³C] and 20% [U-¹³C] glucose. Flux ratio estimates (median) and 10% and 90% prediction quantiles are reported.

Table is available online at

journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005109#sec022

Table S5. ¹³C-MFA analysis of *E. coli* for eight strains grown on 100% [1-¹³C] or 20% [U-¹³C] and 80% naturally labeled glucose. The best fit flux values and 95% confidence intervals for the fluxes estimated using parameter continuation procedure in INCA software. The prediction quantiles for the flux ratios were calculated from the flux ratio distributions of 1000 flux solutions found with the optimization procedure.

Table is available online at

journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005109#sec022

Table S6. SUMOFLUX predictions for metabolic origin of oxaloacetate for eight *E. coli* strains grown on 20% [U-¹³C] and 80% naturally labeled glucose. Flux ratio estimates (median) and 10% and 90% prediction quantiles are reported.

Strain	Oxaloacetate from glyoxylate	Oxaloacetate from TCA cycle
MG1655	0.16 [0.06 0.26]	0.32 [0.13 0.48]
Δzwf	0.20 [0.05 0.41]	0.50 [0.31 0.67]
Δpgi	0.32 [0.13 0.47]	0.44 [0.19 0.60]
$\Delta pfkA$	0.30 [0.09 0.42]	0.47 [0.21 0.57]
$\Delta pykAF$	0.17 [0.02 0.30]	0.34 [0.17 0.49]
$\Delta mae/pck$	0.19 [0.05 0.34]	0.29 [0.08 0.46]
$\Delta sdh/mdh$	0.04 [0.01 0.12]	0.06 [0.01 0.24]
$\Delta fumA$	0.05 [0.01 0.12]	0.16 [0.03 0.30]

Table S7. Metabolites and metabolite fractions measurable with LC-MS or LC-MS/MS methods.
The metabolites and fragments that were used for *in silico* experimental design.

Metabolite	Abbreviation	Fragment carbon positions	LC-MS	LC-MS/MS
Oxoglutarate	OGA15	1 2 3 4 5	+	+
Oxoglutarate	OGA25	2 3 4 5		+
Oxoglutarate	OGA24	2 3 4		+
Phosphoenolpyruvate	PEP13	1 2 3	+	+
Phosphoenolpyruvate	PEP23	2 3		+
Biphosphoglycerate	BPG13	1 2 3	+	+
Pyruvate	PYR13	1 2 3	+	+
Pyruvate	PYR23	2 3		+
Oxaloacetate	OAA14	1 2 3 4	+	+
Oxaloacetate	OAA24	2 3 4		+
Ribose-5-phosphate	P5P15	1 2 3 4 5	+	+
Ribose-5-phosphate	P5P35	3 4 5		+
Ribose-5-phosphate	P5P45	4 5		+
Citrate	Cit16	1 2 3 4 5 6	+	+
Citrate	Cit12	1 2		+
Dihydroxyacetone phosphate	DHAP13	1 2 3	+	+
Fructose-6-phosphate	F6P16	1 2 3 4 5 6	+	+
Fructose-6-phosphate	F6P46	4 5 6		+
Fructose-6-phosphate	F6P56	5 6		+
Fructose-1,6-bisphosphate	FBP16	1 2 3 4 5 6	+	+
Glucose-6-phosphate	G6P16	1 2 3 4 5 6	+	+
Glucose-6-phosphate	G6P36	3 4 5 6		+
Glucose-6-phosphate	G6P46	4 5 6		+
Glucose-6-phosphate	G6P56	5 6		+
Malate	Mal14	1 2 3 4	+	+
Glycerate-3-phosphate	PGA13	1 2 3	+	+
Succinate	Suc14	1 2 3 4	+	+
Succinate	Suc13	1 2 3		+
Succinate	Suc24	2 3 4		+

Table S8. MRM transitions of the metabolites and metabolite fractions measurable with LC-MS/MS methods.

Table is available online at journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005109#sec022

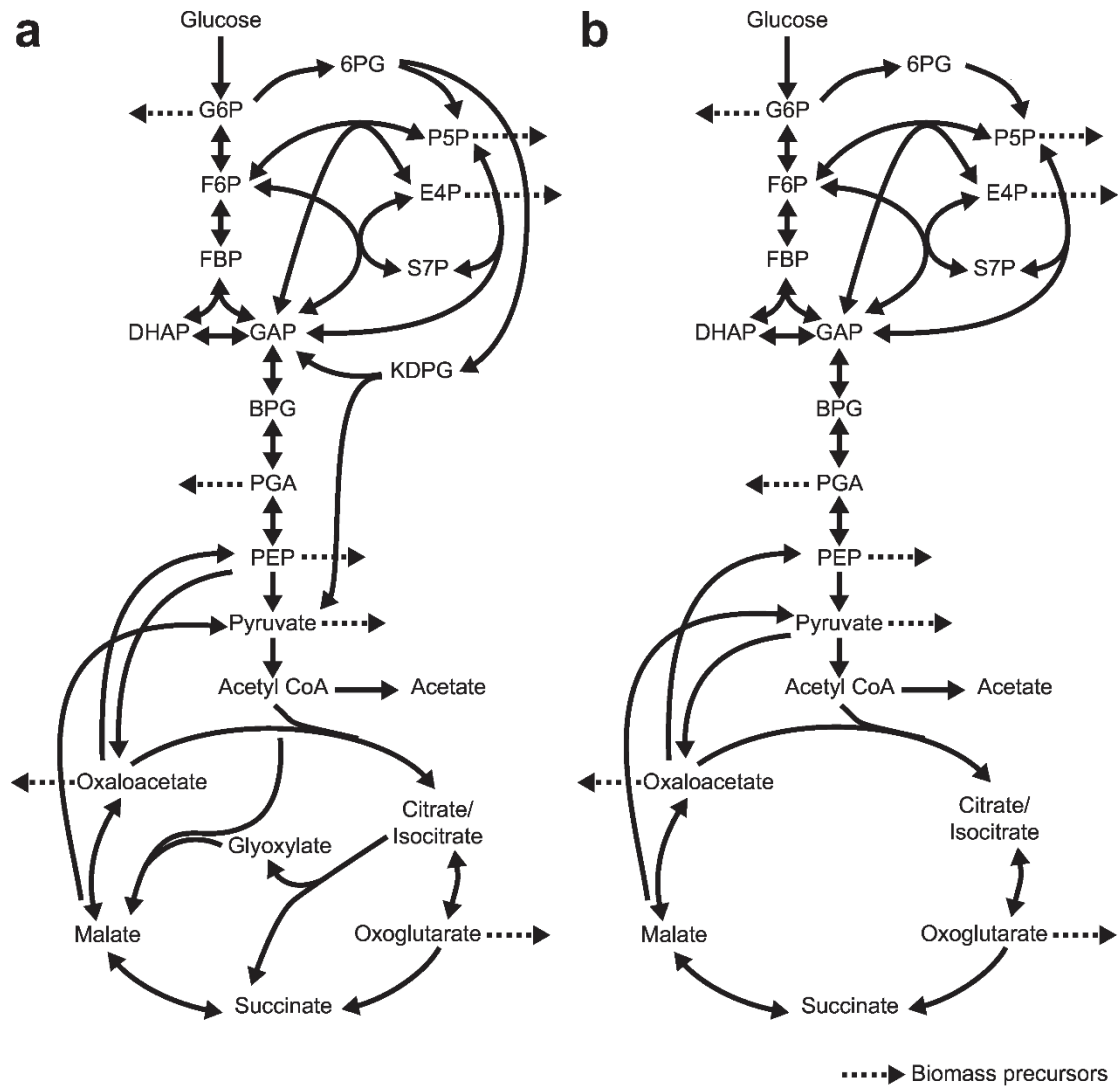


Figure S1. *Escherichia coli* and *Bacillus subtilis* metabolic networks used in the study. (a) *E. coli* metabolic network used for the simulations. (b) *B. subtilis* metabolic network used for the simulations. Biomass precursor fluxes are depicted with a dashed arrow.

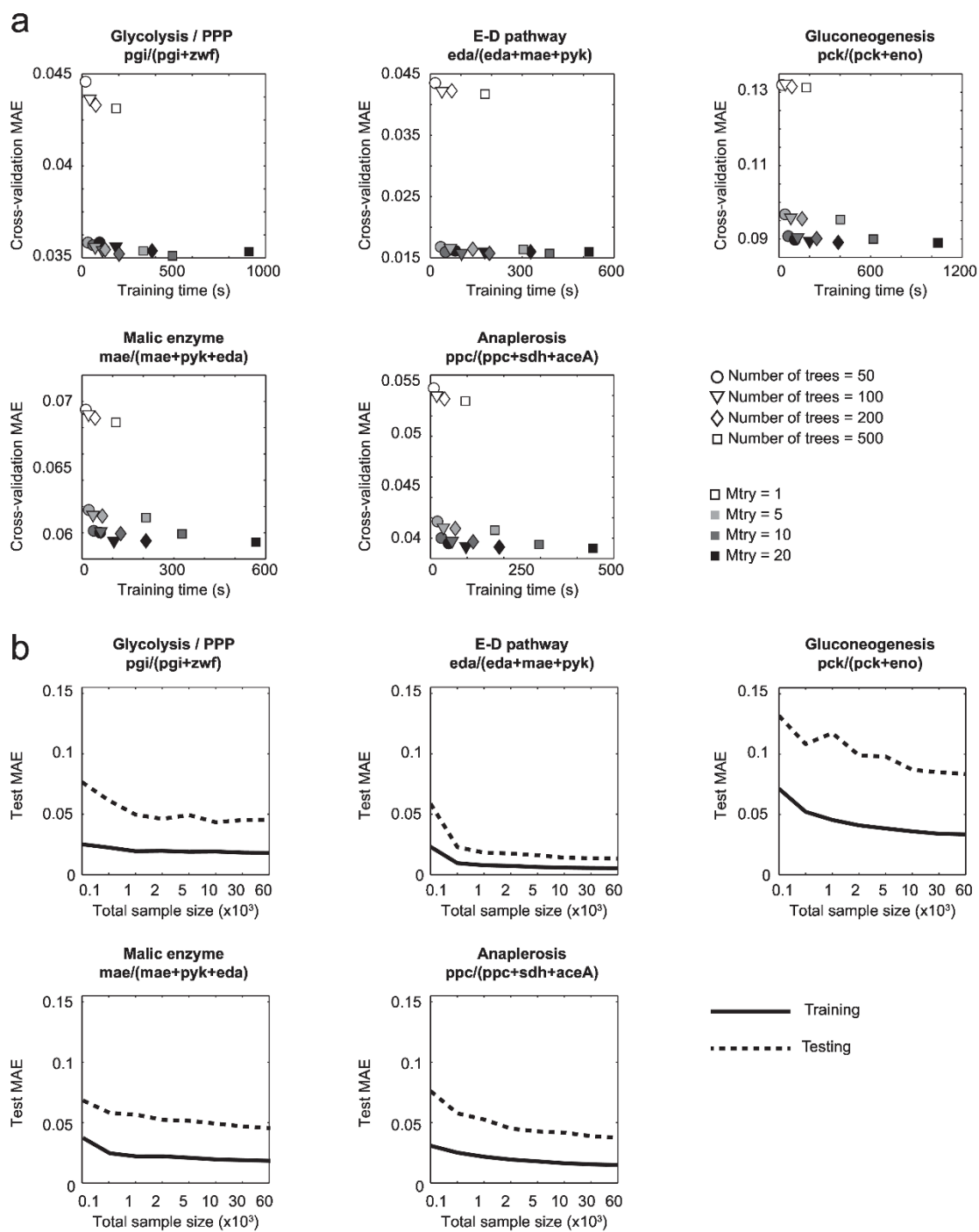


Figure S2. Choosing the SUMOFLUX predictor's parameters and sample size. (a) SUMOFLUX performance assessment with 5-fold cross-validation (CV) with different values of ntree and mtry parameters. (b) SUMOFLUX performance on the test dataset for different sample sizes.

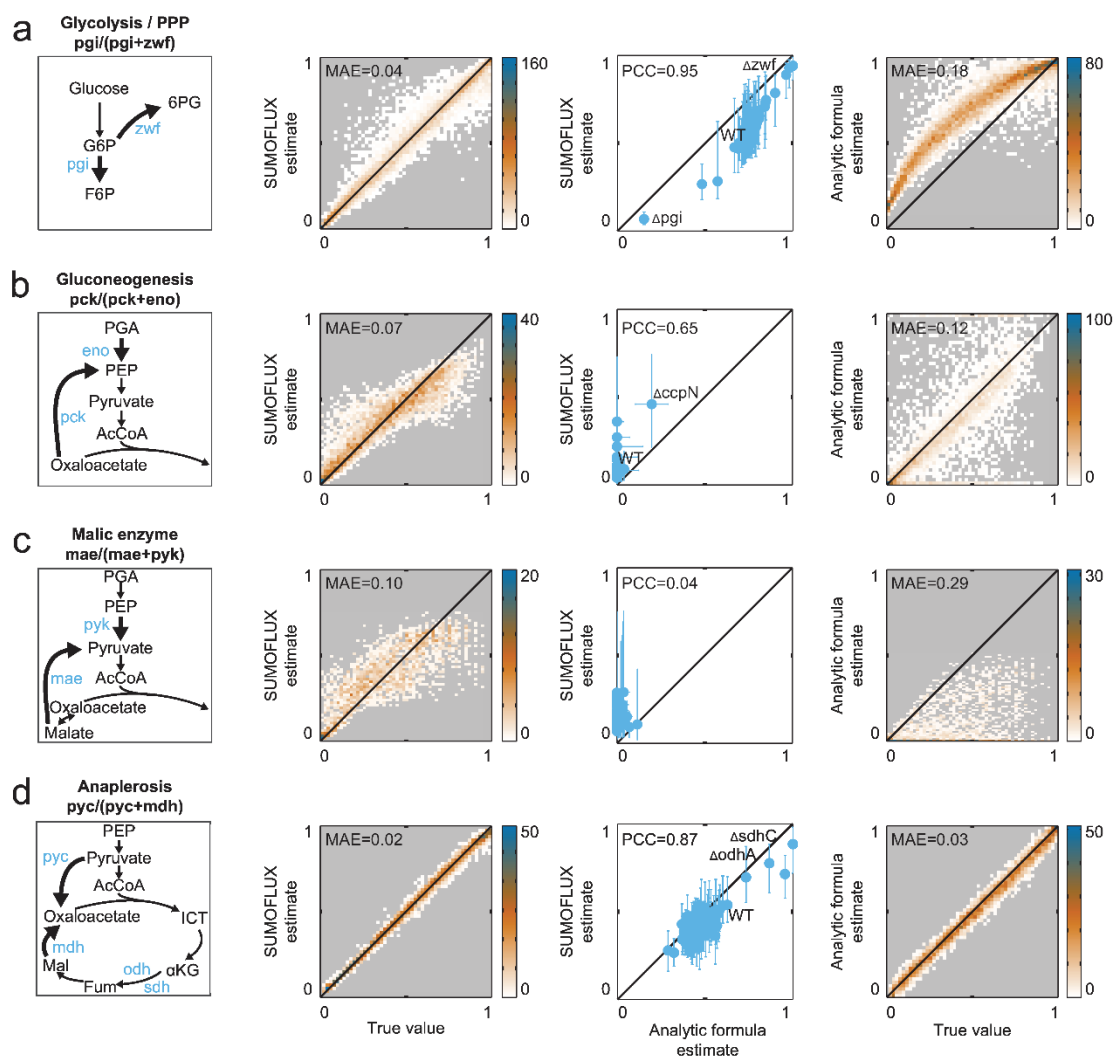


Figure S3. Comparison of SUMOFLUX and analytic formulas for key flux ratio estimation in *B. subtilis* central metabolism. From left to right: a schematic representation of the flux ratio; density plot representing SUMOFLUX estimates versus the true flux ratios for *in silico* data; comparison of the SUMOFLUX and analytic formula estimates for the experimental data; density plot representing analytic formula estimates versus the true flux ratios for *in silico* data. Vertical error bars in the third panel represent [10% 90%] SUMOFLUX prediction quantiles, horizontal error bars represent standard deviation of the analytic formula estimate. (a) Glycolysis versus PPP. (b) PEP fraction from gluconeogenesis. (c) Pyruvate fraction from the malic enzyme flux. (d) Oxaloacetate fraction from anaplerosis from pyruvate. Ratios were estimated for the experiment with 80% [^{13}C] and 20% [^{13}C] glucose. 6PG – 6-phospho-D-gluconate; αKG – α -ketoglutarate; AcCoA – acetyl-CoA; F6P – fructose-6-phosphate; Fum – fumarate; G6P – glucose-6-phosphate; ICT – isocitrate; MAE – mean absolute error; Mal – malate; PCC – Pearson correlation coefficient; PEP – phosphoenolpyruvate; PGA – phosphoglycerate; PPP – pentose phosphate pathway.

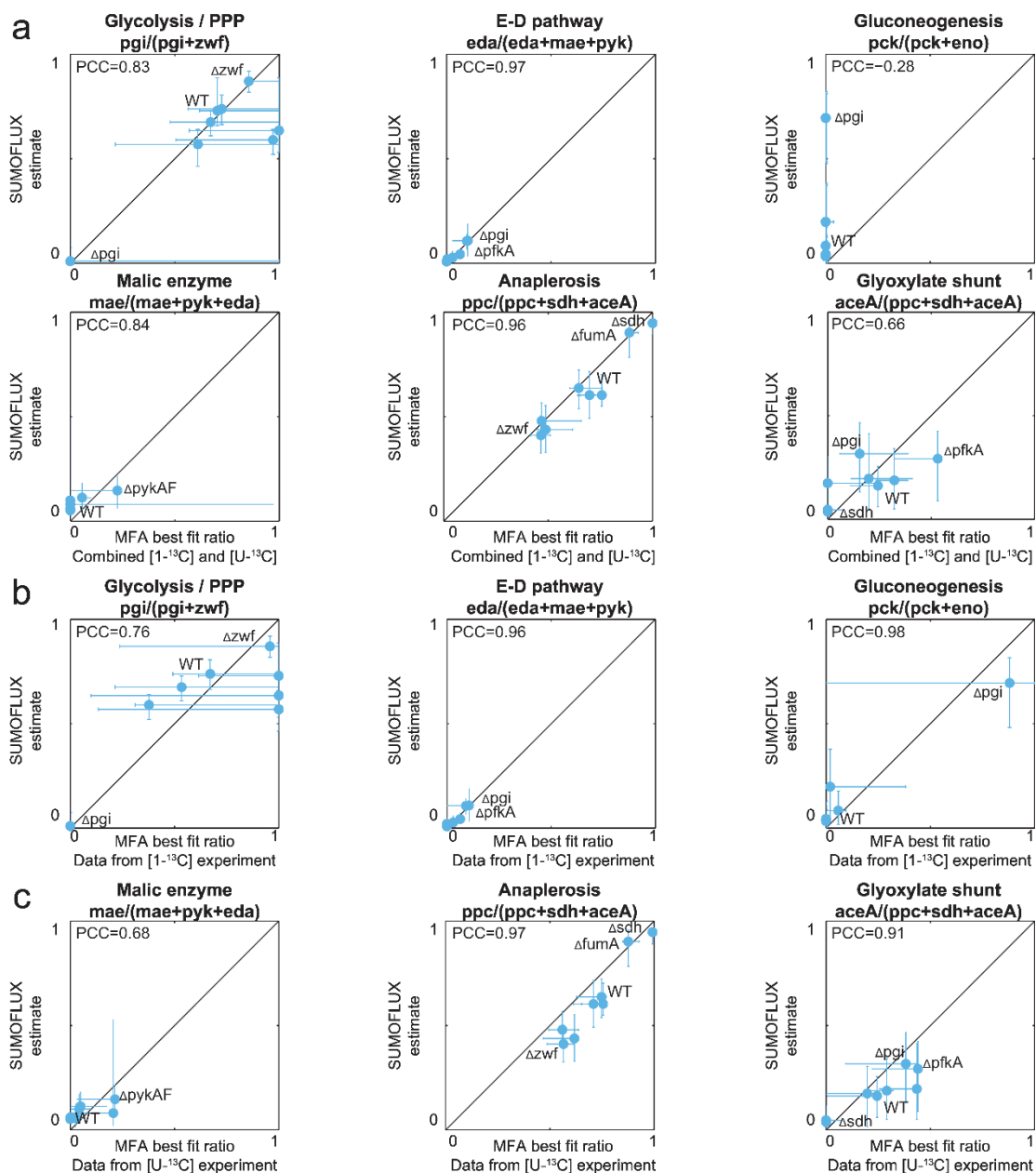


Figure S4. Comparison of SUMOFLUX and ^{13}C -MFA analysis of *E. coli* central metabolism. Comparison of the SUMOFLUX and ^{13}C -MFA flux ratio estimates for the experimental data. Error bars in represent [10% 90%] prediction quantiles. (a) ^{13}C -MFA flux ratios were calculated for the optimal solutions fitted to the combined data of [$1\text{-}^{13}\text{C}$] and [$\text{U-}^{13}\text{C}$] glucose labeling experiments. (b) ^{13}C -MFA flux ratios were calculated for the optimal solutions fitted to the data of [$1\text{-}^{13}\text{C}$] glucose labeling experiment only. (c) ^{13}C -MFA flux ratios were calculated for the optimal solutions fitted to the data of [$\text{U-}^{13}\text{C}$] glucose labeling experiment only. E-D – Entner-Doudoroff pathway; PCC – Pearson correlation coefficient; PPP – pentose phosphate pathway.

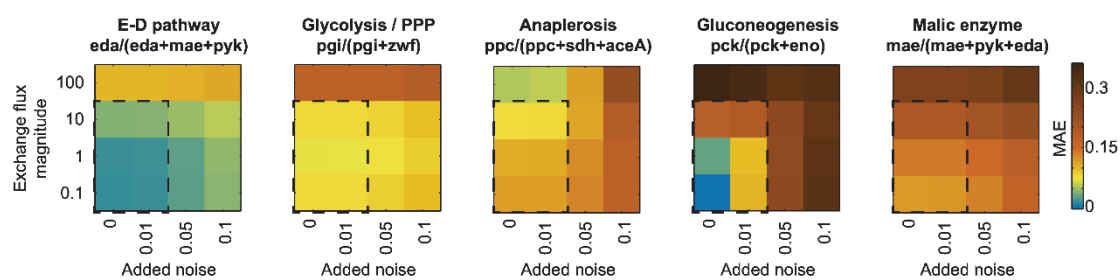


Figure S5. Robustness analysis of analytic formulas in terms of noise and exchange flux magnitude. Mean absolute errors on the test dataset of five analytic formulas applied to *in silico* data with different amount of measurement noise and exchange flux magnitude. The dashed rectangle indicates the normal range of noise (0.01) and exchange flux magnitude (10 times the net flux). E-D – Entner-Doudoroff pathway, MAE – mean absolute error; PPP – pentose phosphate pathway.

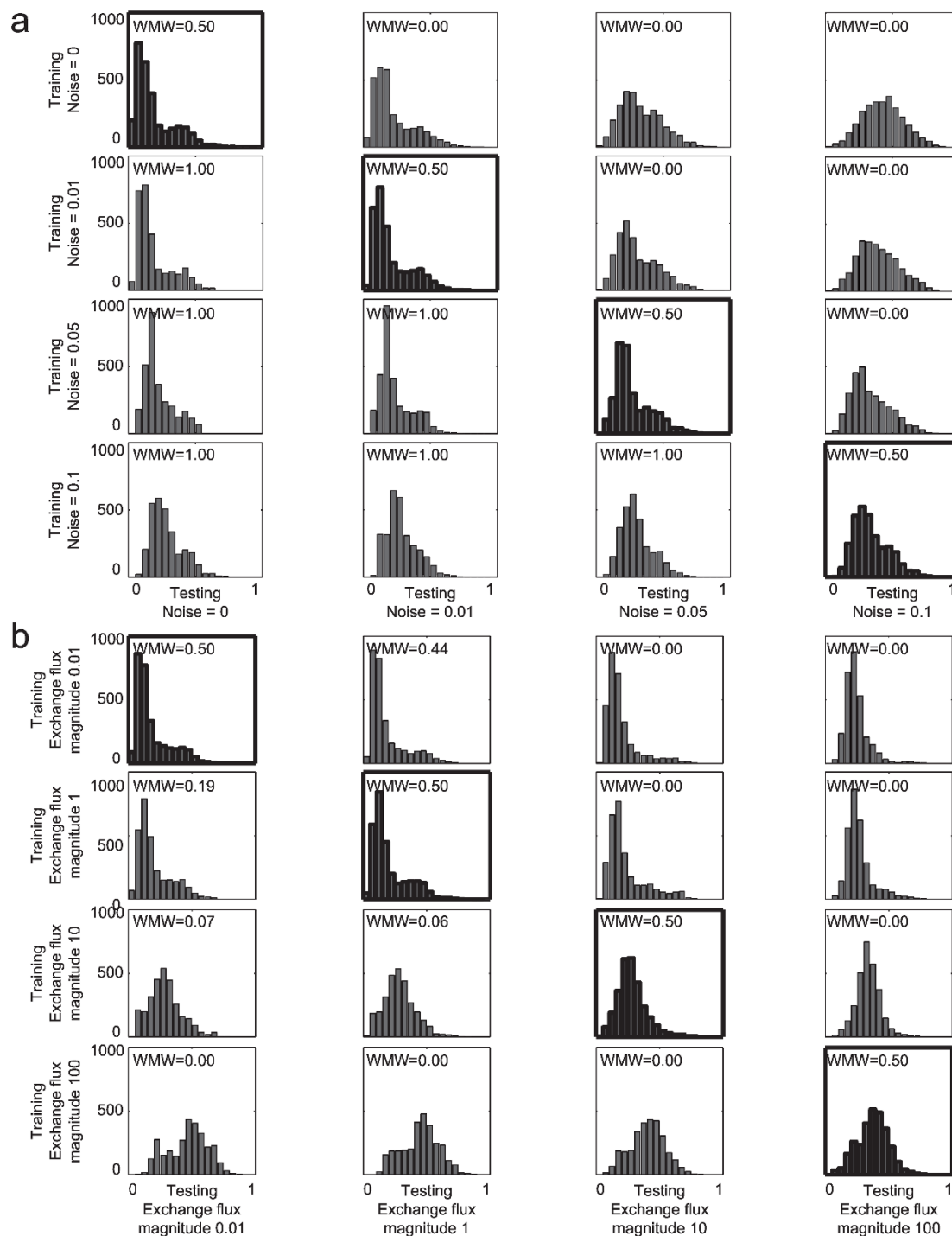


Figure S6. Distributions of interquartile ranges of the flux ratio predictions might indicate incompatibility of training and testing datasets. (a) Histograms representing the [10% 90%] interquartile ranges of the pyruvate fraction from the malic enzyme flux ratio predictions on the test dataset. The noise level in the training dataset is varied along the y-axis, the noise level in the testing dataset is varied along the x-axis. The exchange flux magnitude was set to 1. The histograms in black represent cases of compatible training and testing datasets with the same assumptions on the noise level and exchange flux magnitude. (b) Histograms representing the [10% 90%] interquartile ranges of the pyruvate fraction from the malic enzyme flux ratio predictions on the test dataset. The exchange flux magnitude in the training dataset is varied along the y-axis, the exchange flux magnitude in the testing dataset is varied along the x-axis. The noise level was set to 0.01. The histograms in black represent cases of compatible training and testing datasets with the same assumptions on the noise level and exchange flux magnitude. WMW – p-value of the Wilcoxon-Mann-Whitney right tail test comparing the distributions of interquartile range of each testing dataset to the distribution of interquartile range of the

testing dataset compatible with the noise and exchange level assumptions of the training dataset (diagonal plots in black). Low WMW values indicate that the median of the interquartile range distribution of the corresponding testing dataset is significantly larger than the median of the interquartile range distribution of the testing dataset compatible with the assumptions of the training dataset, which indicates that these assumptions are incompatible with the current testing set.

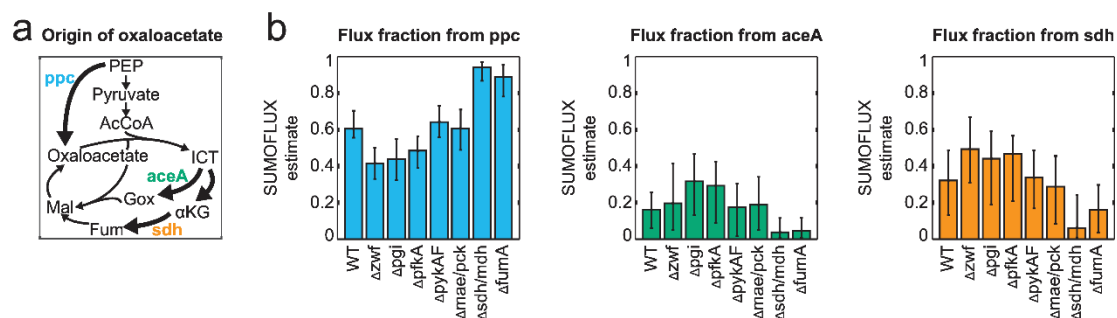


Figure S7. SUMOFLUX estimates for the relative contributions of fluxes to the oxaloacetate pool for the published *E. coli* data. (a) A schematic representation of the glyoxylate shunt, TCA cycle and anaplerosis from PEP fluxes contributing to the formation of oxaloacetate. (b) SUMOFLUX prediction of the relative contributions of the anaplerotic flux from phosphoenolpyruvate, glyoxylate shunt and TCA cycle flux to the oxaloacetate pool for the *in silico* test dataset. The error bars represent [10% 90%] prediction quantiles. Data from experiment with 20% [U - ^{13}C] and 80% naturally labeled glucose. α KG – α -ketoglutarate; AcCoA – acetyl-CoA; Fum – fumarate; Gox – glyoxylate; ICT – isocitrate; Mal – malate; PEP – phosphoenolpyruvate; TCA cycle – tricarboxylic acid cycle.

Appendix II

Supplementary information Chapter 3

Table S1. Growth rates of *M. smegmatis* in glucose and glucose and amino acid media. Optical density of bacterial cultures growing in amino acid and glucose media were monitored by Tecan plate reader (Tecan 200 infinity, Tecan, Männedorf, Switzerland).

Medium	Growth rate (mean)	Growth rate (std)	Max OD (mean)	Max OD (std)
Glucose	0.11	0.02	0.76	0.07
Glucose+alanine	0.15	0.00	1.62	0.02
Glucose+arginine	0.14	0.00	1.33	0.02
Glucose+asparagine	0.16	0.00	1.74	0.05
Glucose+aspartate	0.15	0.00	1.46	0.03
Glucose+cystein	0.08	0.00	0.93	0.02
Glucose+glutamine	0.16	0.00	1.90	0.04
Glucose+glutamate	0.18	0.00	1.73	0.01
Glucose+glycine	0.12	0.02	0.75	0.01
Glucose+histidine	0.15	0.00	1.67	0.00
Glucose+isoleucine	0.13	0.00	1.40	0.02
Glucose+leucine	0.12	0.00	1.16	0.00
Glucose+lysine	0.11	0.00	0.62	0.12
Glucose+methionine	0.12	0.00	1.09	0.01
Glucose+phenylalanine	0.12	0.00	1.11	0.04
Glucose+proline	0.20	0.01	1.98	0.03
Glucose+serine	0.16	0.00	1.34	0.01
Glucose+threonine	0.14	0.00	1.27	0.00
Glucose+tryptophan	0.10	0.00	0.99	0.00
Glucose+tyrosine	0.11	0.00	1.14	0.07
Glucose+valine	0.11	0.00	1.23	0.11

Table S2. Flux ratio definitions for reactions in the *M. smegmatis* metabolic model (Table S3).

Ratio name	Ratio formula
ALA biosynthesis	$\text{ala_out} / (\text{ala_up} + \text{ala_out})$
SER biosynthesis	$\text{ser_out} / (\text{ser_up} + \text{ser_out})$
GLU biosynthesis	$\text{glu_out} / (\text{glu_up} + \text{glu_out})$
ARG biosynthesis	$\text{glu_out} / (\text{arg_cat} + \text{glu_out})$
ASN biosynthesis	$\text{asp_out} / (\text{asn_cat} + \text{asp_out})$
ASP biosynthesis	$\text{asp_out} / (\text{asp_up} + \text{asp_out})$
GLN biosynthesis	$\text{glu_out} / (\text{gln_cat} + \text{glu_out})$
HIS biosynthesis	$(\text{his1_out} + \text{his2_out}) / (\text{his_up} + \text{his1_out} + \text{his2_out})$
ILE biosynthesis	$\text{ile_out} / (\text{ile_up} + \text{ile_out})$
LEU biosynthesis	$\text{leu_out} / (\text{leu_up} + \text{leu_out})$
MET biosynthesis	$\text{asp_out} / (\text{met_cat} + \text{asp_out})$
PHE biosynthesis	$\text{phe_out} / (\text{phe_up} + \text{phe_out})$
LYS biosynthesis	$(\text{lys1_out} + \text{lys2_out}) / (\text{lys_up} + \text{lys1_out} + \text{lys2_out})$
PRO biosynthesis	$\text{pro_out} / (\text{pro_up} + \text{pro_out})$
THR biosynthesis	$\text{thr_out} / (\text{thr_up} + \text{thr_out})$
TYR biosynthesis	$\text{tyr_out} / (\text{tyr_up} + \text{tyr_out})$
TRP biosynthesis	$\text{trp_out} / (\text{trp_up} + \text{trp_out})$
VAL biosynthesis	$\text{val_out} / (\text{val_up} + \text{val_out})$
CYS biosynthesis	$\text{ser_out} / (\text{cys_cat} + \text{ser_out})$
GLY biosynthesis	$\text{gly_out} / (\text{gly_up} + \text{gly_out})$
ALA catabolism	$\text{ala_cat} / (\text{ALA_bm} + \text{ala_cat})$
SER catabolism	$(\text{ser_cat} + \text{gly1_out}) / (\text{SER_bm} + \text{ser_cat} + \text{gly1_out})$
GLU catabolism	$(\text{glu_cat} + \text{his2_out} + \text{pro_out}) / (\text{GLU_bm} + \text{glu_cat} + \text{his2_out} + \text{pro_out})$
ARG catabolism	$\text{arg_cat} / (\text{ARG_bm} + \text{arg_cat})$
ASN catabolism	$\text{asn_cat} / (\text{ASN_bm} + \text{asn_cat})$
ASP catabolism	$(\text{asp_cat} + \text{thr_out} + \text{lys1_out} + \text{lys2_out}) / (\text{ASP_bm} + \text{asp_cat} + \text{thr_out} + \text{lys1_out} + \text{lys2_out})$
GLN catabolism	$\text{gln_cat} / (\text{GLN_bm} + \text{gln_cat})$
HIS catabolism	$\text{his1_cat} + \text{his2_cat} / (\text{HIS_bm} + \text{his1_cat} + \text{his2_cat})$
ILE catabolism	$\text{ile_cat} / (\text{ILE_bm} + \text{ile_cat})$
LEU catabolism	$\text{leu_cat} / (\text{LEU_bm} + \text{leu_cat})$
MET catabolism	$\text{met_cat} / (\text{MET_bm} + \text{met_cat})$
PHE catabolism	$\text{phe_cat} / (\text{PHE_bm} + \text{phe_cat})$
LYS catabolism	$(\text{lys1_cat} + \text{lys2_cat}) / (\text{LYS_bm} + \text{lys1_cat} + \text{lys2_cat})$
PRO catabolism	$\text{pro_cat} / (\text{PRO_bm} + \text{pro_cat})$
THR catabolism	$(\text{thr_cat} + \text{ile_out}) / (\text{THR_bm} + \text{thr_cat} + \text{ile_out})$
TYR catabolism	$\text{tyr_cat} / (\text{TYR_bm} + \text{tyr_cat})$
TRP catabolism	$\text{trp_cat} / (\text{TRP_bm} + \text{trp_cat})$
VAL catabolism	$\text{val_cat} / (\text{VAL_bm} + \text{val_cat})$
CYS catabolism	$\text{cys_cat} / (\text{CYS_bm} + \text{cys_cat})$
GLY catabolism	$\text{gly_cat} / (\text{GLY_bm} + \text{gly_cat})$
PEP from oxaloacetate	$\text{pck} / (\text{eno} + \text{pck})$
PEP from glycolysis	$\text{eno} / (\text{eno} + \text{pck})$
Pyruvate from malic enzyme	$\text{mae} / (\text{pyk} + \text{edp2} + \text{mae})$
Pyruvate from PEP	$\text{pyk} / (\text{pyk} + \text{edp2} + \text{mae})$
Oxaloacetate from TCA cycle	$\text{mdh} / (\text{mdh} + \text{pyc} + \text{ppc})$
Oxaloacetate from PEP/pyruvate	$\text{pyc} + \text{ppc} / (\text{mdh} + \text{pyc} + \text{ppc})$
Malate from glyoxylate shunt	$\text{gs2} / (\text{fum} + \text{gs2})$
Malate from TCA cycle	$\text{fum} / (\text{fum} + \text{gs2})$
Glycolysis/PPP	$\text{pgi} / (\text{pgi} + \text{zwf})$
Pyruvate from ED pathway	$\text{edp2} / (\text{edp2} + \text{pyk} + \text{mae})$
DHAP through glycolysis	$(\text{fba} + \text{tpi}) / (\text{fba} + \text{tpi} + \text{gapdh} + \text{TA})$

Table S3. *M. smegmatis* model of central carbon metabolism used throughout the study. Reactions in the sections "Glucose + amino acid" were included only for the specified models.

Reaction type	Reaction	Stoichiometry and carbon transitions
Uptake and secretion	glc_up	glucose (abcdef) -> G6P (abcdef)
	CO2up	CO2in (a) -> CO2 (a)
	accoa_ac	AcCoA (ab) -> Ac (ab)
	ac_out	Ac (ab) -> Acetate (ab)
	co2_out	CO2 (a) -> CO2out (a)
Biomass precursors	G6P_bm	G6P (abcdef) -> G6Pbm (abcdef)
	PGA_bm	PGA (abc) -> PGAbm (abc)
	P5P_bm	P5P (abcde) -> P5Pbm (abcde)
	PEP_bm	PEP (abc) -> PEPbm (abc)
	PYR_bm	PSuYR (abc) -> PYRbm (abc)
	OGA_bm	OGA (abcde) -> OGAbm (abcde)
	OAA_bm	OAA (abcd) -> OAAbm (abcd)
	E4P_bm	E4P (abcd) -> E4Pbm (abcd)
	PHE_bm	PHE (abcdefghi) -> PHEbm (abcdefghi)
	ALA_bm	ALA (abc) -> ALAbm (abc)
	ASP_bm	ASP (abcd) -> ASPbm (abcd)
	GLU_bm	GLU (abcde) -> GLUbm (abcde)
	GLY_bm	GLY (ab) -> GLYbm (ab)
	HIS_bm	HIS (abcdef) -> HISbm (abcdef)
	ILE_bm	ILE (abcdef) -> ILEbm (abcdef)
	LEU_bm	LEU (abcdef) -> LEUbm (abcdef)
	LYS_bm	LYS (abcdef) -> LYSbm (abcdef)
	SER_bm	SER (abc) -> SERbm (abc)
	THR_bm	THR (abcd) -> THRbm (abcd)
	TYR_bm	TYR (abcdefghi) -> TYRbm (abcdefghi)
VAL_bm	VAL (abcde) -> VALbm (abcde)	
PRO_bm	PRO (abc) -> PRObm (abc)	
CYS_bm	CYS (abc) -> CYSbm (abc)	
GLN_bm	GLN (abcde) -> GLNbm (abcde)	
ASN_bm	ASN (abcd) -> ASNbm (abcd)	
ARG_bm	ARG (abcdef) -> ARGbm (abcdef)	
MET_bm	MET (abcde) -> METbm (abcde)	
TRP_bm	TRP (hijklabcdef) -> TRPbm (hijklabcdef)	
Glycolysis	pgm	G1P (abcdef) <-> G6P (abcdef)
	pgi	G6P (abcdef) <-> F6P (abcdef)
	pfk	F6P (abcdef) -> FBP (abcdef)
	fba	FBP (abcdef) -> DHAP (cba) + GAP (def)
	tpi	DHAP (abc) <-> GAP (abc)
	gapdh	GAP (abc) -> BPG (abc)
	bpg	BPG (abc) -> PGA (abc)
	eno	PGA (abc) -> PEP (abc)
	pyk	PEP (abc) -> PYR (abc)
Pentose phosphate pathway	zwf	G6P (abcdef) -> PG6 (abcdef)
	gnd	PG6 (abcdef) -> P5P (bcdef) + CO2 (a)
	TK1	P5P (abcde) + P5P (fghij) <-> GAP (cde) + S7P (abfghij)
	TK2	P5P (abcde) + E4P (fghi) <-> GAP (cde) + F6P (abfghi)
	TA	S7P (abcdefg) + GAP (hij) <-> E4P (defg) + F6P (abchij)
Entner-Doudoroff	edp1	PG6 (abcdef) -> KDPPG (abcdef)
	edp2	KDPPG (abcdef) -> PYR (abc) + GAP (def)
TCA cycle	pdh	PYR (abc) -> AcCoA (bc) + CO2 (a)
	citl	OAA (cdef) + AcCoA (ab) -> Cit (fedcba)
	acnA	Cit (abcdef) <-> Act (abcdef)

	acnB	Act(abcdef) <-> Icit (abcdef)
	idh	Icit (abcdef) <-> OGA (abcef) + CO2 (d)
	sdh	OGA (abcde) -> Suc (bcde) + CO2 (a)
	frdA	Suc (abcd) <-> Fum (abcd)
	fum	Fum (abcd) <-> Mal (abcd)
	mdh	Mal (abcd) <-> OAA (abcd)
Glyoxylate shunt	gs1	Cit (abcdef) -> Glx (ab) + Suc (dcef)
	gs2	Glx (ab) + AcCoA (cd) -> Mal (abcd)
Glycerate pathway	gcl	Glx (ab) + Glx (cd) -> PGA (abd) + CO2 (c)
Anaplerosis and gluconeogenesis	mae	Mal (abcd) -> PYR (abc) + CO2 (d)
	pck	OAA (abcd) -> PEP (abc) + CO2 (d)
	ppc	PEP (abc) + CO2 (d) -> OAA (abcd)
Intermediates of aminoacid metabolism	kiv_out	PYR (abc) + PYR (def) -> KIV (abcef) + CO2 (d)
	cho_out	E4P (abcd) + PYR (efg) -> CHO (fgabcde)
Amino acid production	ala_out	PYR (abc) -> ALA (abc)
	asp_out	OAA (abcd) -> ASP (abcd)
	val_out	KIV (abcde) -> VAL (abcde)
	leu_out	KIV (abcde) + AcCoA (fg) -> LEU (bcdefg) + CO2 (a)
	ser_out	BPG (abc) -> SER (abc)
	gly_out	SER (abc) -> GLY (ab) + MTHF (c)
	his1_out	P5P (abcde) + MTHF (f) -> HIS (fedcba)
	his2_out	GLU (abcde) + MTHF (f) -> HIS (fedcba)
	thr_out	ASP (abcd) -> THR (abcd)
	lys1_out	PYR (abc) + ASP (defg) -> LYS (bcdefg) + CO2 (a)
	lys2_out	PYR (abc) + ASP (defg) -> LYS (abcefg) + CO2 (d)
	ile_out	THR (abcd) + PYR (efg) -> ILE (abcdfg) + CO2 (e)
	glu_out	OGA (abcde) -> GLU (abcde)
	pro_out	GLU (abcde) -> PRO (abcde)
	phe_out	CHO (abcdefg) + PYR (hij) -> PHE (hijabcdef) + CO2 (g)
	tyr_out	CHO (abcdefg) + PYR (hij) -> TYR (hijabcdef) + CO2 (g)
	arg_out	GLU (abcde) + urea (f) -> ARG (abcdef)
	asn_out	ASP (abcd) -> ASN (abcd)
	cys_out	SER (abc) -> CYS (abc)
	gln_out	GLU (abcde) -> GLN (abcde)
met_out	ASP (abcd) + MTHF (e) -> MET (abcde)	
trp_out	CHO (abcdefg) + P5P (hijkl) -> TRP (hijklabcdef) + CO2 (g)	
Glucose+alanine	ala_up	ALA_in (abc) -> ALA (abc)
	ala_cat	ALA (abc) -> PYR (abc)
Glucose+arginine	arg_up	ARG_in (abcdef) -> ARG (abcdef)
	ARG_bm	ARG (abcdef) -> ARGbm (abcdef)
	arg_cat	ARG (abcdef) -> GLU (abcde) + urea (f)
	glu_cat	GLU (abcde) -> OGA (abcde)
	arg_out	GLU (abcde) + urea (f) -> ARG (abcdef)
Glucose+asparagine	asn_up	ASN_in (abcd) -> ASN (abcd)
	ASN_bm	ASN (abcd) -> ASNbm (abcd)
	asn_cat	ASN (abcd) -> ASP (abcd)
	asp_cat	ASP (abcd) -> OAA (abcd)
	asn_out	ASP (abcd) -> ASN (abcd)
Glucose+aspartate	asp_up	ASP_in (abcd) -> ASP (abcd)
	asp_cat	ASP (abcd) -> OAA (abcd)
Glucose+cystein	cys_up	CYS_in (abc) -> CYS (abc)
	cys_cat	CYS (abc) -> SER (abc)
	CYS_bm	CYS (abc) -> CYSbm (abc)
	ser_cat	SER (abc) -> BPG (abc)

	cys_out	SER (abc) -> CYS (abc)
Glucose+glutamine	gln_up	GLN_in (abcde) -> GLN (abcde)
	GLN_bm	GLN (abcde) -> GLNbm (abcde)
	glu_cat	GLU (abcde) -> OGA (abcde)
	gln_cat	GLN (abcde) -> GLU (abcde)
	gln_out	GLU (abcde) -> GLN (abcde)
Glucose+glutamate	glu_up	GLU_in (abcde) -> GLU (abcde)
	glu_cat	GLU (abcde) -> OGA (abcde)
Glucose+glycine	gly_up	GLY_in (ab) -> GLY (ab)
	mthf_up	MTHF_in(a) -> MTHF(a)
	GLY_bm	GLY (ab) -> GLYbm (ab)
	gly_cat	GLY (ab) + MTHF(c) -> SER (abc)
	ser_cat	SER (abc) -> BPG (abc)
	gly_out	SER (abc) -> GLY (ab) + MTHF (c)
Glucose+histidine	his_up	HIS_in (abcdef) -> HIS (abcdef)
	his1_cat	HIS (fedcba) -> P5P (abcde) + MTHF (f)
	his2_cat	HIS (fedcba) -> GLU (abcde) + MTHF (f)
Glucose+isoleucine	ile_up	ILE_in (abcdef) -> ILE (abcdef)
	ile_cat	ILE (abcdefg) + CO2 (e) -> OAA (abcd) + PYR (efg)
Glucose+leucine	leu_up	LEU_in (abcdef) -> LEU (abcdef)
	leu_cat	LEU (bcdefg) + CO2 (a) -> KIV (abcde) + AcCoA (fg)
	kiv_cat	KIV (abcef) + CO2 (d) -> PYR (abc) + PYR (def)
Glucose+lysine	lys_up	LYS_in (abcdef) -> LYS (abcdef)
	lys1_cat	LYS (bcdefg) + CO2 (a) -> PYR (abc) + ASP (defg)
	lys2_cat	LYS (abcefg) + CO2 (d) -> PYR (abc) + ASP (defg)
Glucose+methionine	met_up	MET_in (abcd) -> MET (abcd)
	met_cat	MET (abcd) -> ASP (abcd)
	MET_bm	MET (abcd) -> METbm (abcd)
	asp_cat	ASP (abcd) -> OAA (abcd)
	met_out	ASP (abcd) -> MET (abcd)
Glucose+phenylalanine	phe_up	PHE_in (abcdefghi) -> PHE (abcdefghi)
	phe_cat	PHE (hijabcdef) + CO2 (g) -> CHO (abcdefg) + PYR (hij)
	cho_cat	CHO (fgabcde) -> E4P (abcd) + PYR (efg)
Glucose+proline	pro_up	PRO_in (abcde) -> PRO (abcde)
	glu_cat	GLU (abcde) -> OGA (abcde)
	pro_cat	PRO (abcde) -> GLU (abcde)
Glucose+serine	ser_up	SER_in (abc) -> SER (abc)
	ser_cat	SER (abc) -> BPG (abc)
Glucose+threonine	thr_up	THR_in (abcd) -> THR (abcd)
	thr_cat	THR (abcd) -> ASP (abcd)
	asp_cat	ASP (abcd) -> OAA (abcd)
Glucose+tryptophan	trp_up	TRP_in (abcdefghijk) -> TRP (abcdefghijk)
	TRP_bm	TRP (hijklabcdef) -> TRPbm (hijklabcdef)
	trp_cat	TRP (hijklabcdef) + CO2 (g) -> CHO (abcdefg) + P5P (hijkl)
	cho_cat	CHO (fgabcde) -> E4P (abcd) + PYR (efg)
	trp_out	CHO (abcdefg) + P5P (hijkl) -> TRP (hijklabcdef) + CO2 (g)
Glucose+tyrosine	tyr_up	TYR_in (abcdefghi) -> TYR (abcdefghi)
	tyr_cat	TYR (hijabcdef) + CO2 (g) -> CHO (abcdefg) + PYR (hij)
	cho_cat	CHO (fgabcde) -> E4P (abcd) + PYR (efg)
Glucose+valine	val_up	VAL_in (abcde) -> VAL (abcde)
	val_cat	VAL (abcde) -> KIV (abcde)
	kiv_cat	KIV (abcef) + CO2 (d) -> PYR (abc) + PYR (def)

Table S4. ¹³C-MFA predictions of fluxes in *M. smegmatis* in glucose and glucose and glutamate media. INCA software was used (Young 2014) to calculate fluxes. Glucose media contained 60% [1-¹³C] and 40% [U-¹³C] glucose. For estimates in glucose and glutamate medium, a combination of two experiments with [1-¹³C] glucose + glutamate and [U-¹³C] glucose + glutamate media was used.

Reaction	Type	Glucose+glutamate medium		Glucose medium	
		Flux value	Flux std	Flux value	Flux std
G6P <-> F6P	Net	11.2	0.1	6.3	2.6
G6P <-> F6P	Exch	0.0	0.6	0.0	5.9
F6P -> FBP	Net	11.2	0.1	17.8	11.0
FBP <-> DHAP + GAP	Net	11.2	0.1	17.8	11.0
FBP <-> DHAP + GAP	Exch	5.8	0.0	0.6	17.7
DHAP <-> GAP	Net	11.2	0.1	17.8	11.0
DHAP <-> GAP	Exch	3.6	0.0	0.6	17.7
GAP <-> PG3	Net	29.2	0.1	46.1	29.8
GAP <-> PG3	Exch	0.0	0.0	53.9	17.7
PG3 <-> PEP	Net	29.2	0.1	46.1	29.4
PG3 <-> PEP	Exch	70.8	1.0	0.0	15.8
PEP -> Pyr	Net	37.5	0.0	52.2	17.1
G6P -> PG6	Net	6.8	0.1	22.0	17.6
PG6 -> Ru5P + CO2	Net	0.0	0.1	17.2	14.4
Ru5P <-> X5P	Net	0.0	0.1	11.5	9.4
Ru5P <-> X5P	Exch	100.0	0.1	52.8	96.1
Ru5P <-> R5P	Net	0.0	0.0	5.7	5.1
Ru5P <-> R5P	Exch	100.0	0.0	1.2	18.4
X5P <-> GAP + EC2	Net	0.0	0.1	11.5	9.4
X5P <-> GAP + EC2	Exch	100.0	0.1	62.5	17.7
F6P <-> E4P + EC2	Net	0.0	0.0	-5.7	4.6
F6P <-> E4P + EC2	Exch	100.0	0.0	0.0	6.1
S7P <-> R5P + EC2	Net	0.0	0.0	-5.7	4.9
S7P <-> R5P + EC2	Exch	100.0	0.0	2.5	17.7
F6P <-> GAP + EC3	Net	0.0	0.0	-5.7	4.9
F6P <-> GAP + EC3	Exch	28.6	0.0	53.3	18.3
S7P <-> E4P + EC3	Net	0.0	0.0	5.7	4.9
S7P <-> E4P + EC3	Exch	100.0	0.0	0.0	21.0
PG6 -> KDPG	Net	6.8	0.1	4.8	3.5
KDPG -> Pyr + GAP	Net	6.8	0.1	4.8	3.5
Pyr -> AcCoA + CO2	Net	52.4	0.0	63.9	37.1
OAA + AcCoA -> Cit	Net	9.4	0.0	16.0	18.7
Cit <-> ICit	Net	9.4	0.0	16.0	18.7
Cit <-> ICit	Exch	0.0	0.0	0.1	17.7
ICit <-> AKG + CO2	Net	7.0	0.0	2.9	10.2
ICit <-> AKG + CO2	Exch	0.0	0.0	0.0	9.6
AKG -> SucCoA + CO2	Net	20.8	0.0	2.9	9.0
SucCoA <-> Suc	Net	20.8	0.0	2.9	9.0
SucCoA <-> Suc	Exch	0.0	0.0	2.2	17.7
Suc <-> Fum	Net	23.3	0.1	16.0	16.8
Suc <-> Fum	Exch	2.2	0.1	2.0	24.4
Fum <-> Mal	Net	23.3	0.1	16.0	16.8
Fum <-> Mal	Exch	76.7	0.1	2.0	24.4
Mal <-> OAA	Net	15.5	0.1	22.1	23.8
Mal <-> OAA	Exch	8.6	0.0	0.0	17.7
Mal -> Pyr + CO2	Net	10.3	0.0	6.9	15.7
PEP + CO2 -> OAA	Net	0.0	0.0	5.6	17.6
Pyr + CO2 -> OAA	Net	2.3	0.0	0.0	18.2
OAA -> PEP + CO2	Net	8.4	0.0	11.8	17.8
AcCoA -> Ac	Net	40.4	0.1	34.8	60.9
AKG -> Glu	Net	86.1	0.0	0.0	3.9
Glu -> Gln	Net	0.0	0.0	0.0	1.0

Glu -> Pro	Net	0.0	0.0	0.0	0.8
Glu + CO2 -> Arg	Net	0.0	0.0	0.0	1.1
OAA -> Asp	Net	0.0	0.0	0.0	5.7
Asp -> Asn	Net	0.0	0.0	0.0	0.9
Pyr -> Ala	Net	0.0	0.0	0.0	1.9
PG3 -> Ser	Net	0.0	0.0	0.0	3.7
Ser <=> Gly + MEETHF	Net	0.0	0.0	0.0	2.3
Ser <=> Gly + MEETHF	Exch	0.0	0.0	0.0	4.3
Thr -> Gly + AcCoA	Net	0.0	0.0	0.0	0.0
Ser -> Cys	Net	0.0	0.0	0.0	0.3
Pyr + Asp -> Lys + CO2	Net	0.0	0.0	0.0	1.3
Asp -> Thr	Net	0.0	0.0	0.0	2.1
Asp -> Met	Net	0.0	0.0	0.0	0.6
Pyr + Pyr -> Val + CO2	Net	0.0	0.0	0.0	1.6
AcCoA + Pyr + Pyr -> Leu + CO2 + CO2	Net	0.0	0.0	0.0	1.7
Thr + Pyr -> Ile + CO2	Net	0.0	0.0	0.0	1.1
PEP + PEP + E4P -> Phe + CO2	Net	0.0	0.0	0.0	0.7
PEP + PEP + E4P -> Tyr + CO2	Net	0.0	0.0	0.0	0.5
Ser + R5P + PEP + E4P -> Trp + CO2 + GAP	Net	0.0	0.0	0.0	0.2
R5P + FTHF -> His	Net	0.0	0.0	0.0	0.4
MEETHF -> METHF	Net	0.0	0.0	0.0	2.0
MEETHF -> FTHF	Net	0.0	0.0	0.0	0.4
Gluc.ext -> G6P	Net	18.0	0.0	28.3	19.1
Ac -> Ac.ext	Net	40.4	0.1	34.8	60.9
CO2 -> CO2.ext	Net	96.5	0.0	100.0	17.9
Biomass	Net	0.0	0.0	0.0	4.0
ICit -> Glx + Suc	Net	2.5	0.0	13.1	9.5
Glx + AcCoA -> Mal	Net	2.5	0.0	13.1	9.5
Pyr -> FattyAcid	Net	0.0	0.1	0.0	78.8
Gluc.ext -> Glu	Net	13.9	0.0	n/a	n/a
Glu -> AKG	Net	100.0	0.0	n/a	n/a

Table S5. Flux ratios calculated for the 13C-MFA solutions from Table S4.

Ratio name	Glucose medium	Glucose+glutamate medium
Glycolysis/PPP	0.22	0.62
Pyruvate from ED pathway	0.07	0.12
DHAP through glycolysis	0.77	0.77
PEP from glycolysis	0.80	0.78
PEP from oxaloacetate	0.20	0.22
Pyruvate from PEP	0.82	0.69
Pyruvate from malic enzyme	0.11	0.19
Oxaloacetate from TCA cycle	0.44	0.47
Oxaloacetate from PEP/pyruvate	0.20	0.13
Malate from TCA cycle	0.55	0.90
Malate from glyoxylate shunt	0.45	0.10

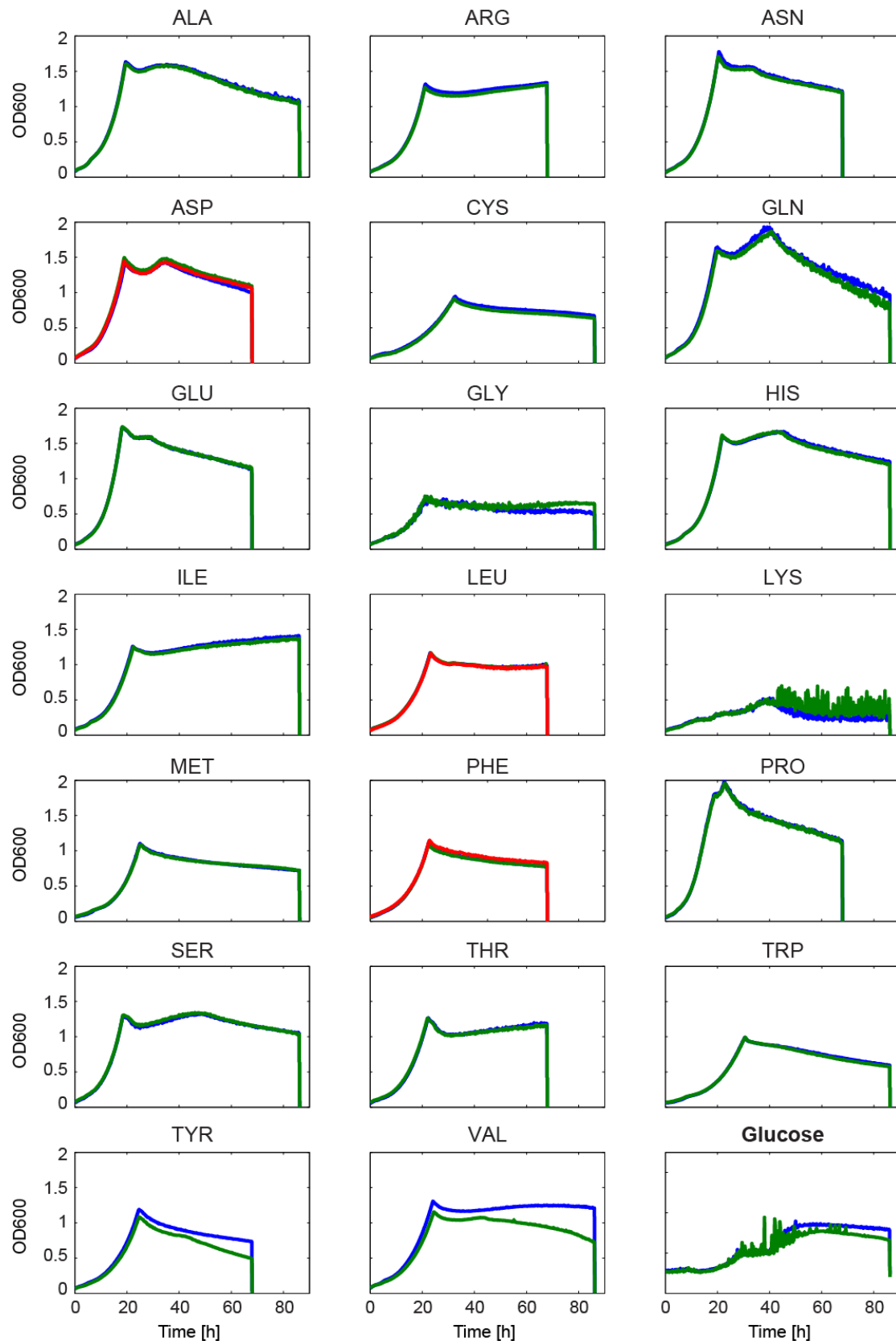


Figure S1. Single amino acids cause different growth effects in *M. smegmatis*. Growth curves of *M. smegmatis* grown on combinations of glucose and 20 single amino acids in 96 well plate. Culture density was measured and converted to OD600. Different colors in each plot depict biological replicates.

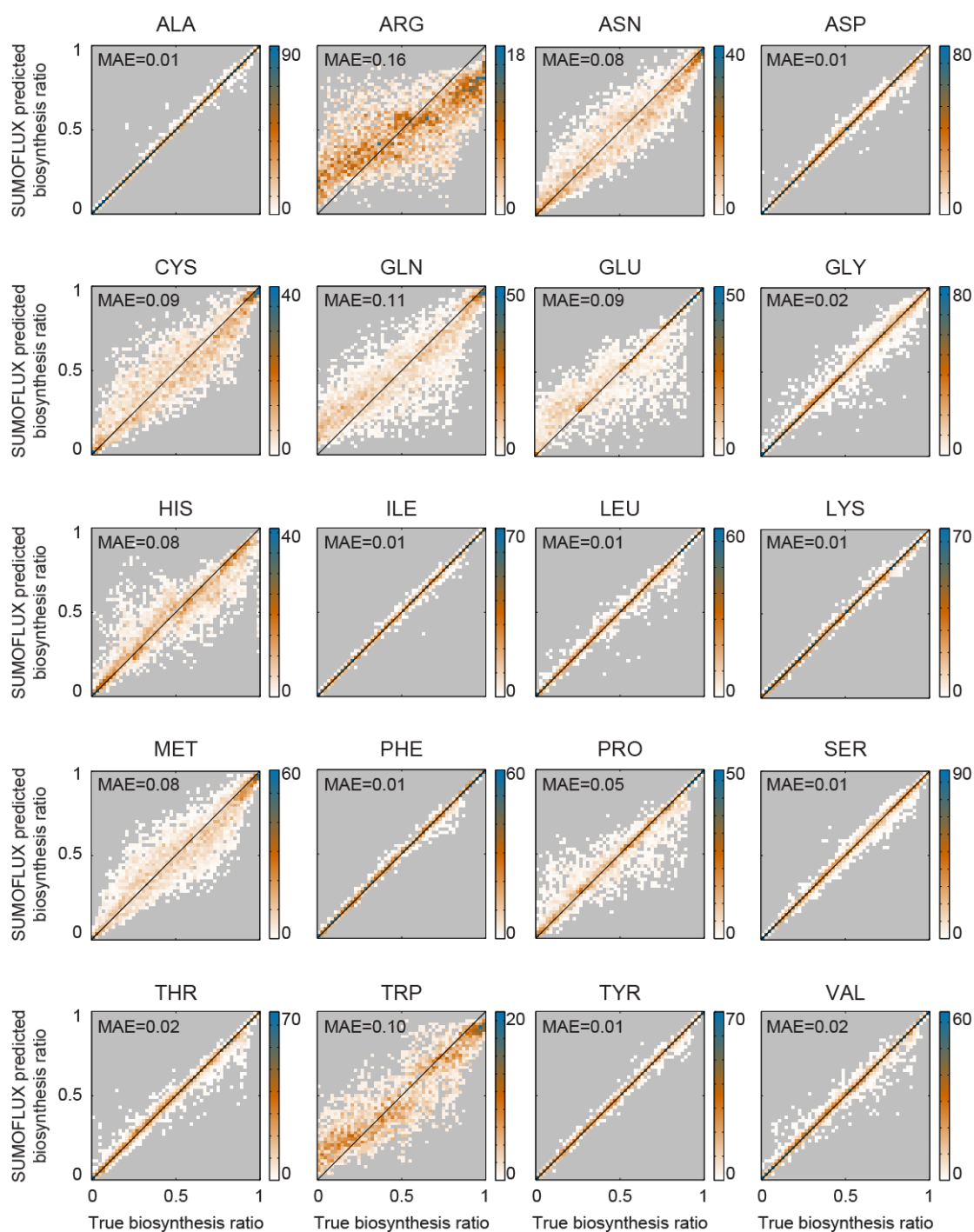


Figure S2. SUMOFLUX estimates single amino acid biosynthesis fractions for *M. smegmatis* with high accuracy. Density plot representing SUMOFLUX estimates versus the true flux ratios for *in silico* data; biosynthesis fraction for each amino acid was estimated for the experiment with [U-¹³C] glucose and single naturally labeled amino acid experiment.

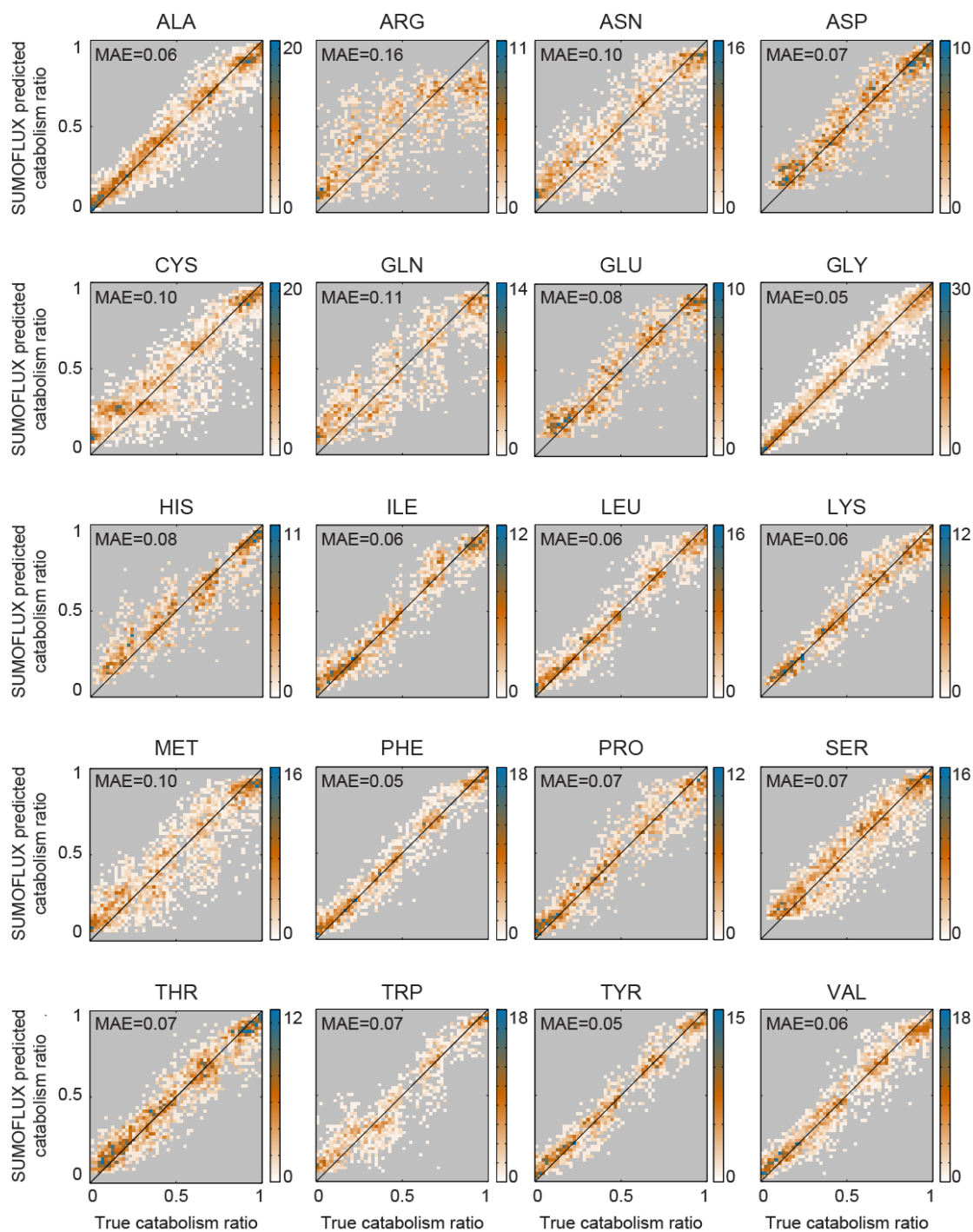


Figure S3. SUMOFLUX estimates single amino acid catabolism fractions for *M. smegmatis* with good accuracy. Density plot representing SUMOFLUX estimates versus the true flux ratios for *in silico* data; catabolism fraction for each amino acid was estimated for the experiment with [U-¹³C] glucose and single naturally labeled amino acid experiment.

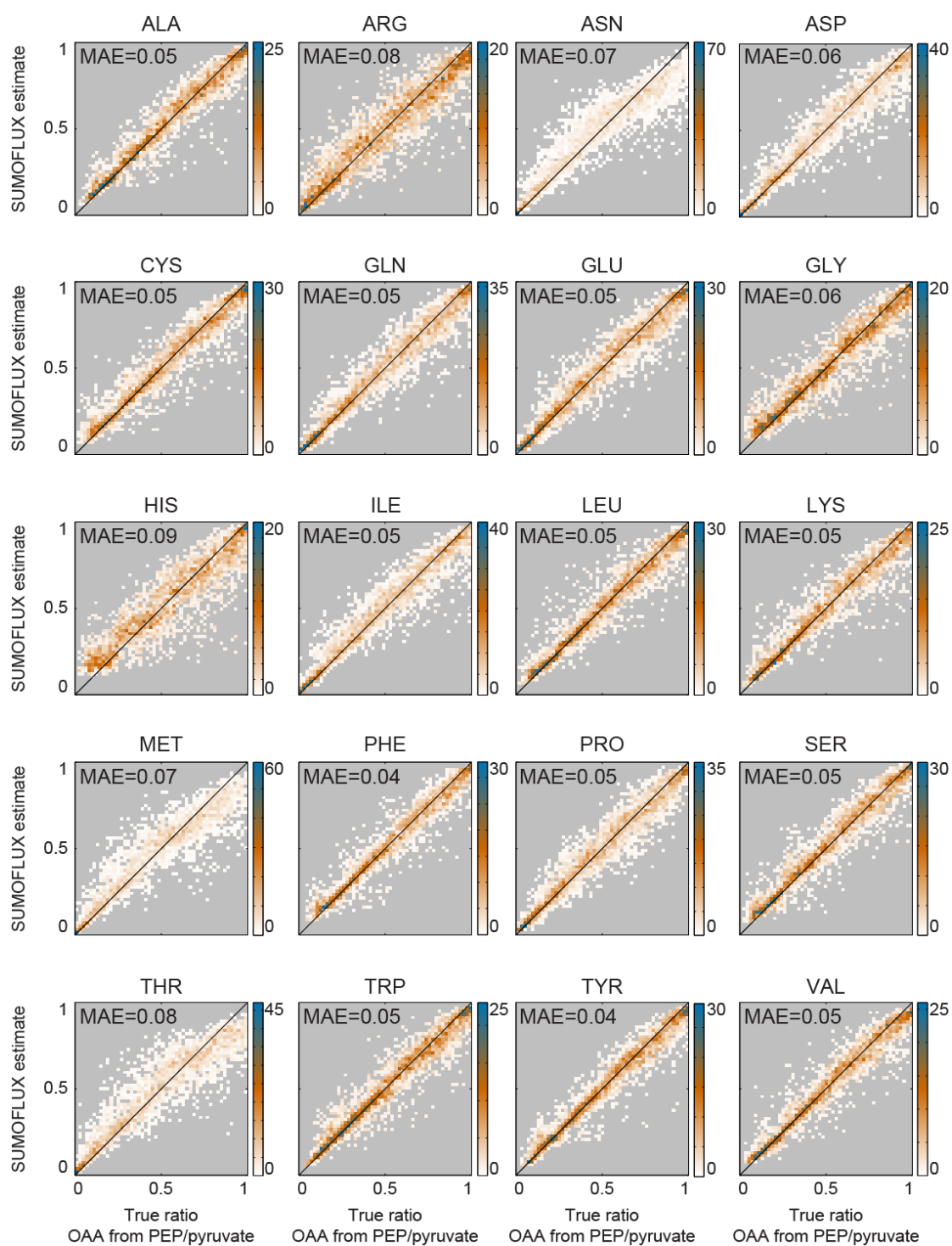


Figure S4. SUMOFLUX estimates ppc+pyc fractions to oxaloacetate pool for *M. smegmatis* with high accuracy. Density plot representing SUMOFLUX estimates versus the true flux ratios *in silico* data; ppc+pyc flux fraction to oxaloacetate pool for each amino acid media was estimated for the experiment with [U-¹³C] glucose and single naturally labeled amino acid experiment.

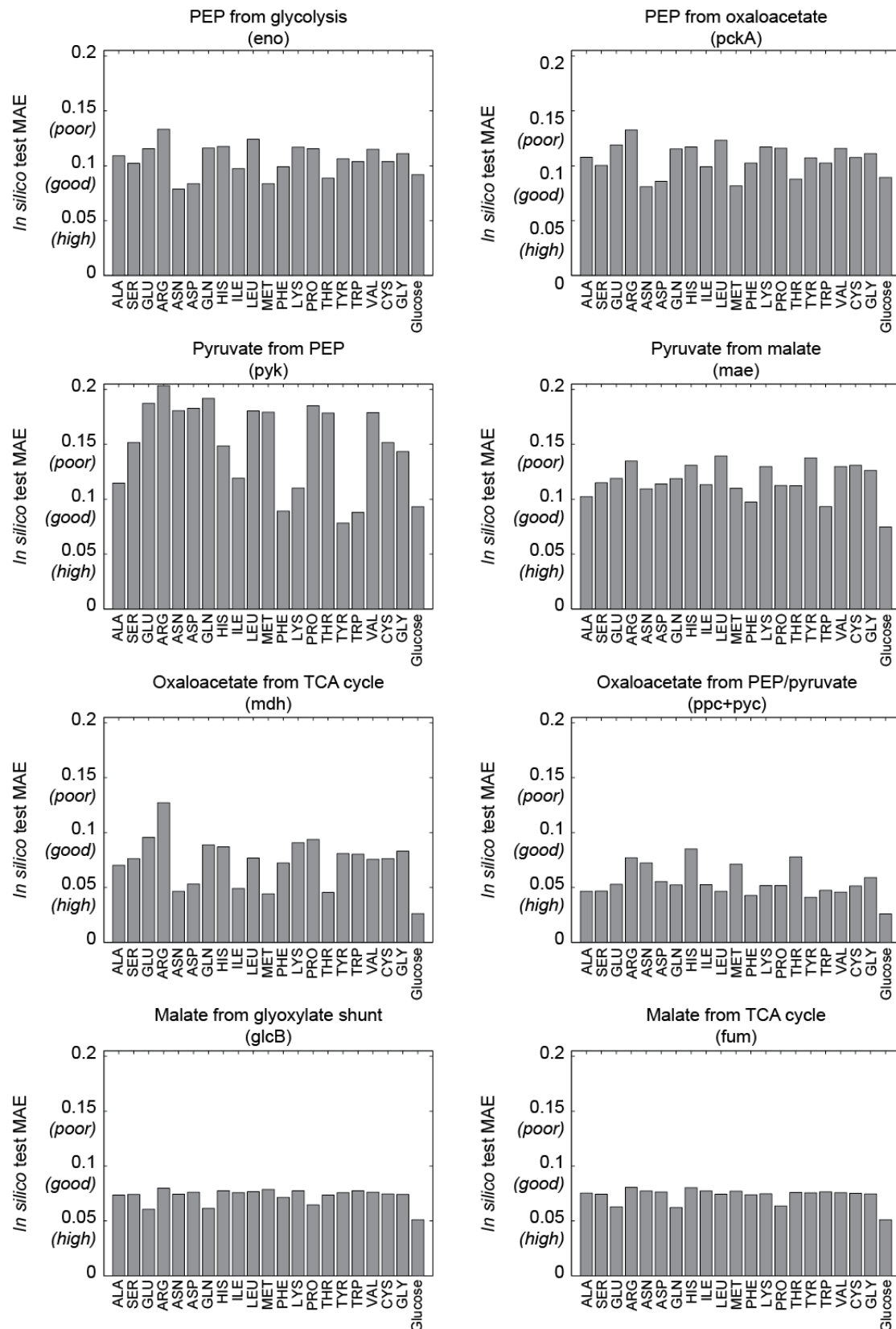


Figure S5. SUMOFLUX estimates key flux fractions in CCM with moderate to high accuracy. Bar plots representing the mean absolute error (MAE) on the *in silico* test datasets of the CCM flux fraction predictors in each of the tested 21 conditions (*M. smegmatis* growing on a combination of with $[U-^{13}C]$ glucose and single naturally labeled amino acid, or a mixture of 50% $[U-^{13}C]$ and 50% unlabeled glucose). The accuracy is considered high if MAE < 0.05, good is MAE < 0.1, and poor otherwise

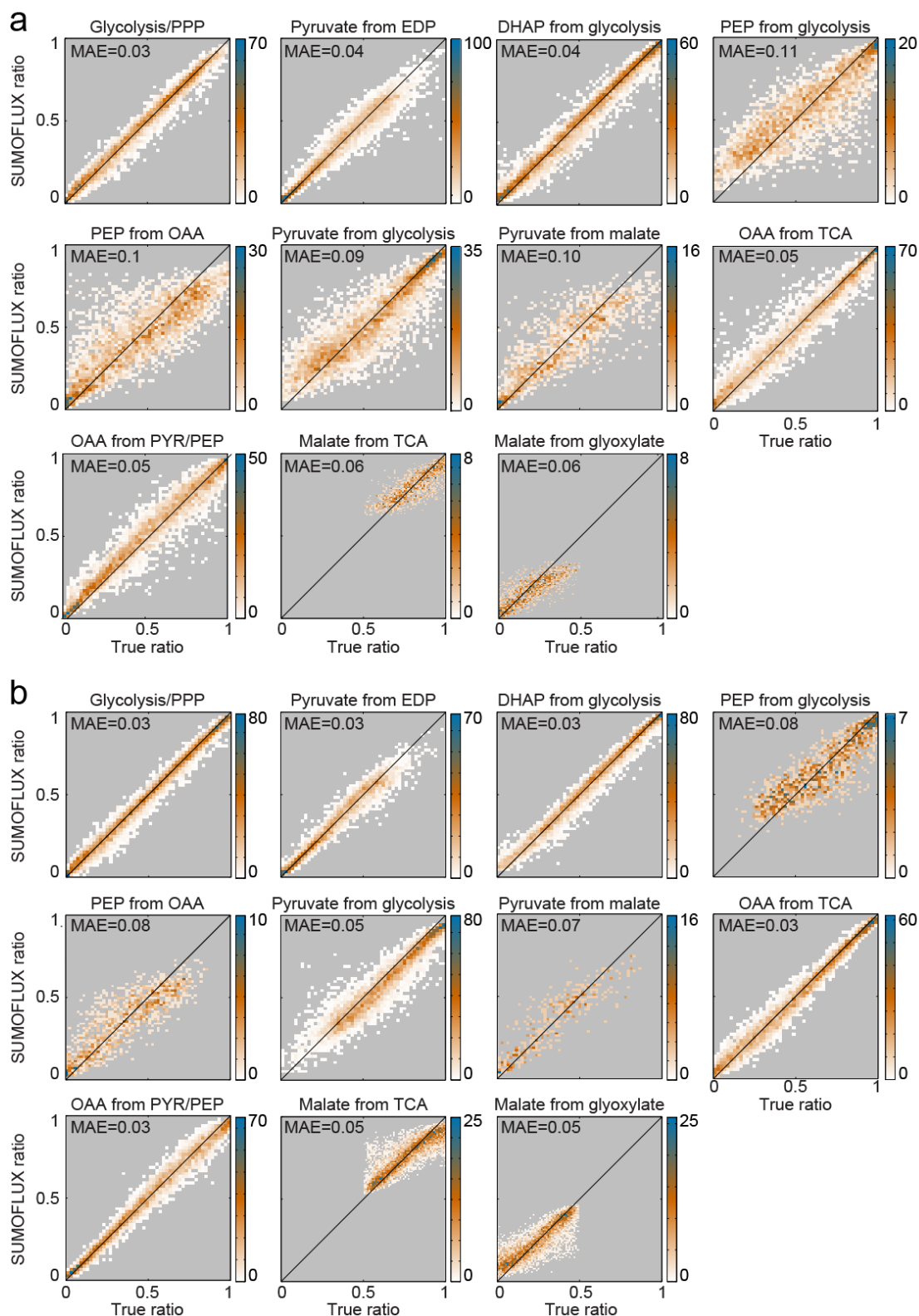


Figure S6. SUMOFLUX estimates key flux ratios in CCM for *M. smegmatis* grown in glucose or glucose and glutamate media with high accuracy. (a) Density plots representing SUMOFLUX estimates versus the true flux ratios for *in silico* data for 11 flux ratios and fractions in CCM. Ratios were estimated using combined data from two labeling experiments: *M. smegmatis* grown on mixture of [1-¹³C] glucose and glutamate or [U-¹³C] glucose and glutamate. (b) Density plots representing SUMOFLUX estimates versus the true flux ratios for *in silico* data for 11 flux ratios and fractions in CCM. Ratios were estimated using combined data from the experiment with *M. smegmatis* grown on mixture of 60% [1-¹³C] glucose and 40% [U-¹³C] glucose.

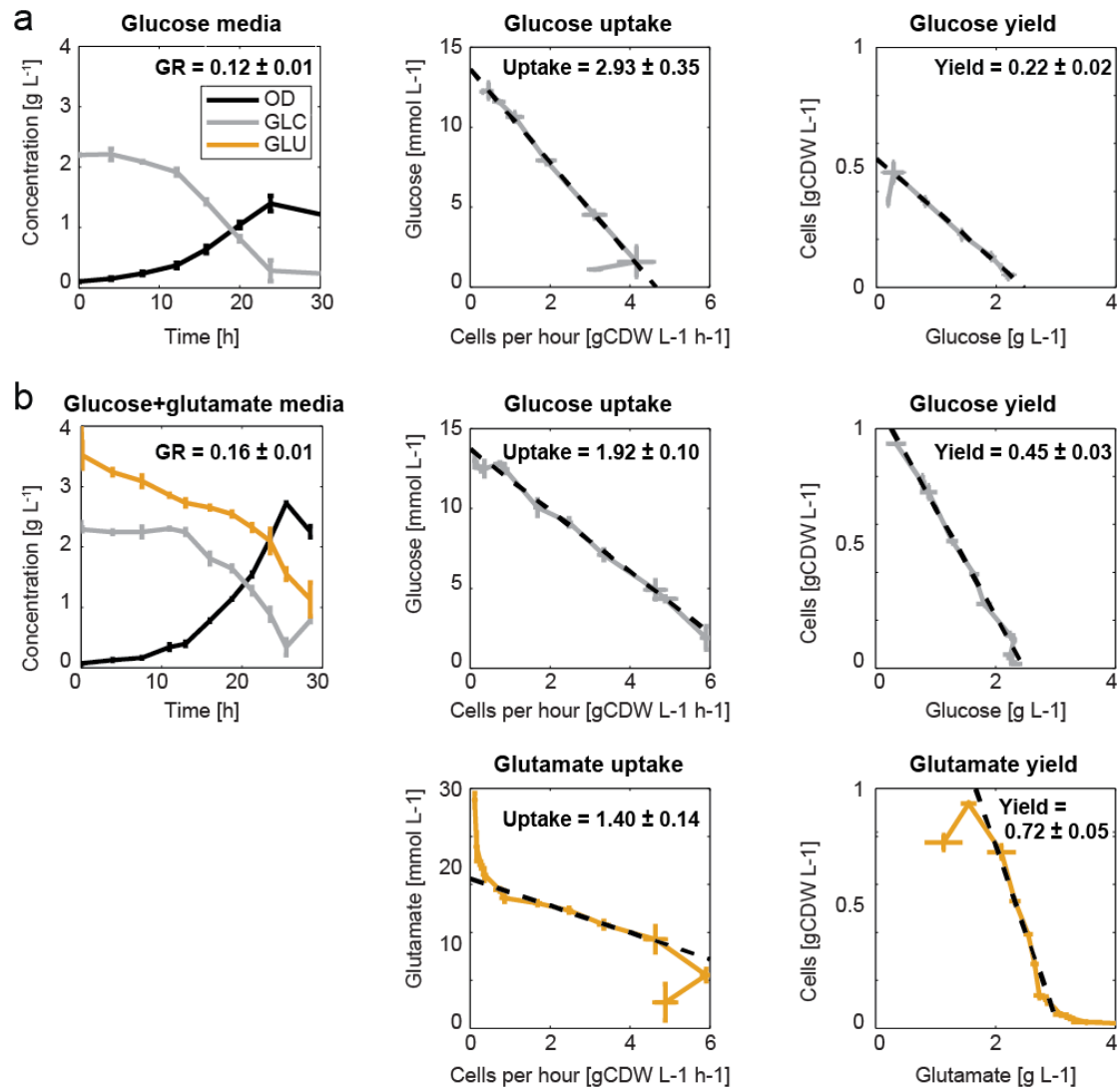


Figure S7. Effect of glutamate in the media on the growth and carbon uptake of *M. smegmatis*. (a) Left to right: *M. smegmatis* growth curve and glucose concentration curve over 30 h in glucose media; glucose concentration versus cell concentration over time plot used to calculate glucose uptake; cell concentration versus glucose concentration plot used to calculate glucose yield. (b) Upper panel, left to right: *M. smegmatis* growth curve and glucose concentration curve over 30 h in glucose and glutamate media; glucose concentration versus cell concentration over time plot used to calculate glucose uptake; cell concentration versus glucose concentration plot used to calculate glucose yield. Lower panel, left to right: glutamate concentration versus cell concentration over time plot used to calculate glutamate uptake; cell concentration versus glutamate concentration plot used to calculate glutamate yield.

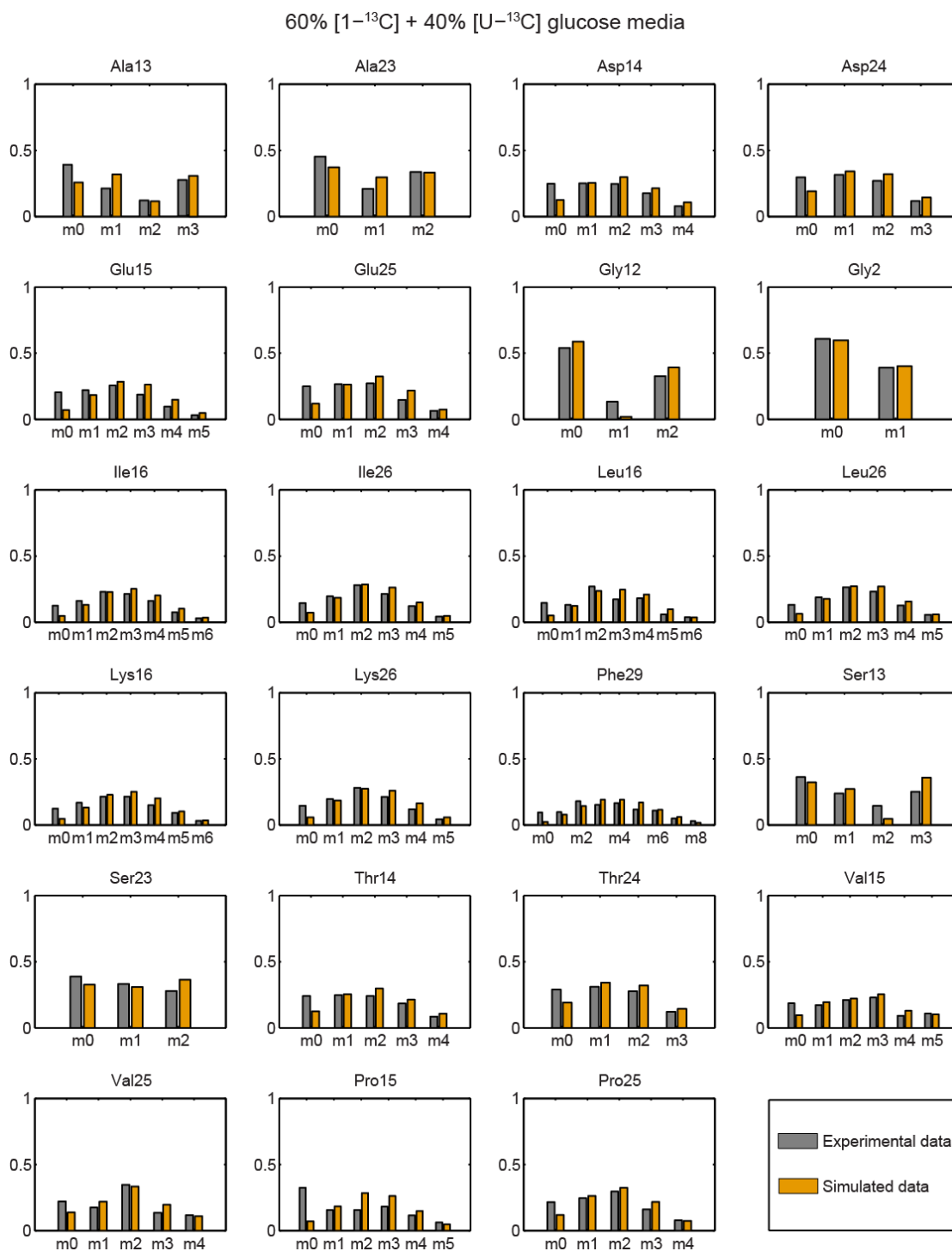


Figure S8. Comparison of experimental and simulated data corresponding to the flux solution for *M. smegmatis* growing on glucose. Measured isotope labeling patterns of intracellular amino acids, and corresponding *in silico* labeling patterns simulated for the flux solution for central carbon metabolism of *M. smegmatis* growing on 60% [1-¹³C] and 40% [U-¹³C] glucose.

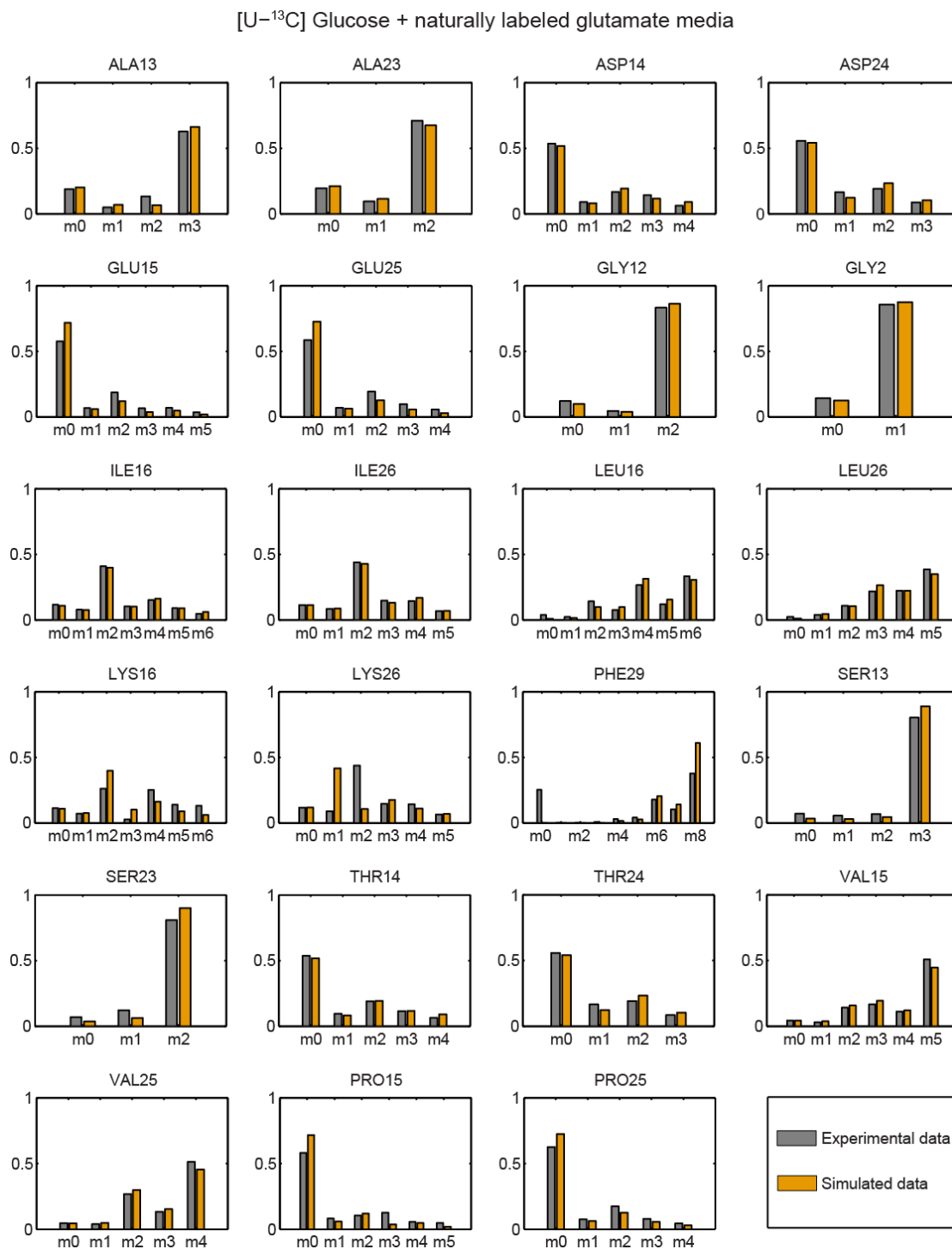


Figure S9. Comparison of experimental and simulated data corresponding to the flux solution for *M. smegmatis* growing on [U-¹³C] glucose and glutamate. Measured isotope labeling patterns of intracellular amino acids, and corresponding *in silico* labeling patterns simulated for the flux solution for central carbon metabolism of *M. smegmatis* growing on [U-¹³C] glucose and naturally labeled glutamate.

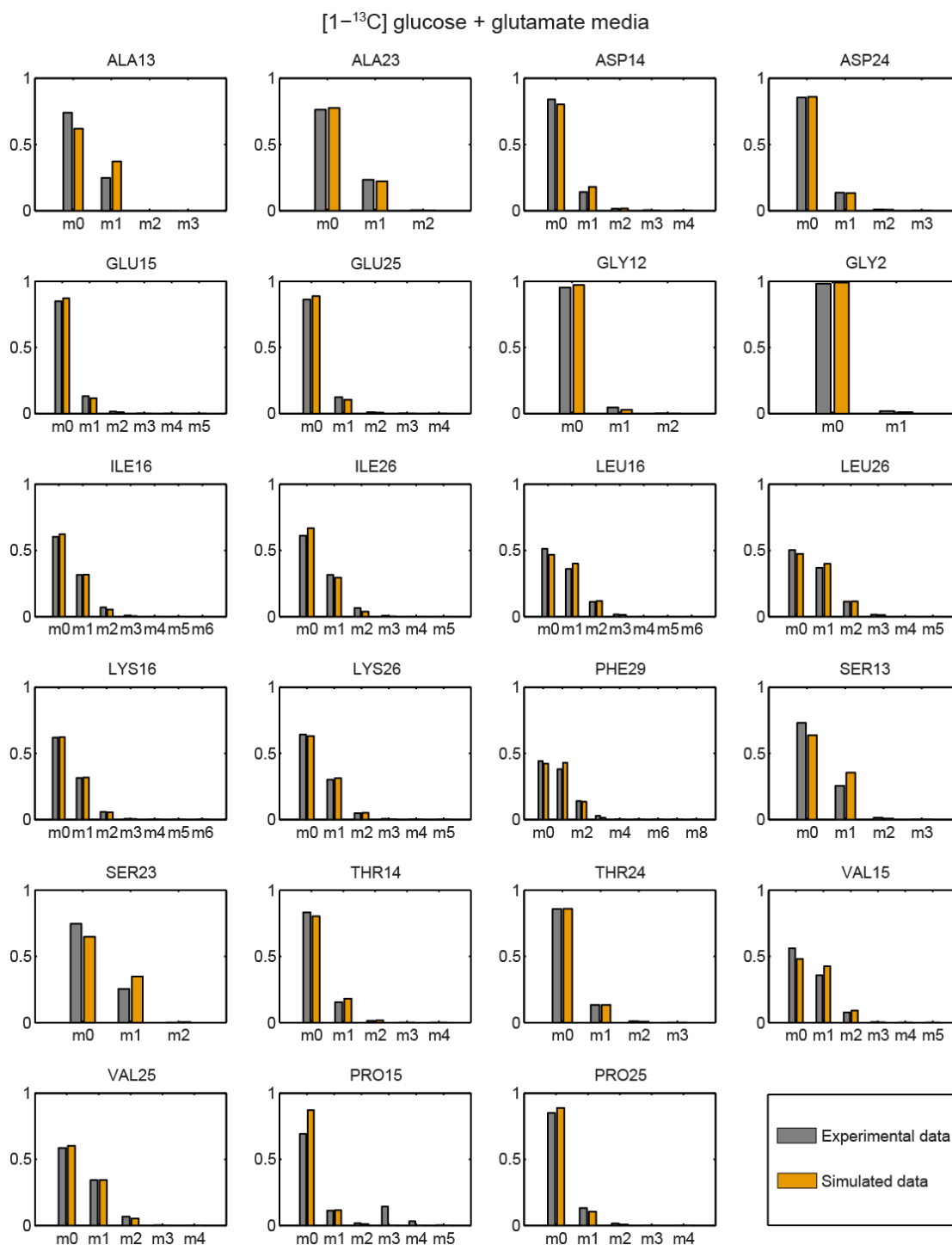


Figure S10. Comparison of experimental and simulated data corresponding to the flux solution for *M. smegmatis* growing on [1-¹³C] glucose and glutamate. Measured isotope labeling patterns of intracellular amino acids, and corresponding *in silico* labeling patterns simulated for the flux solution for central carbon metabolism of *M. smegmatis* growing on [1-¹³C] glucose and naturally labeled glutamate.

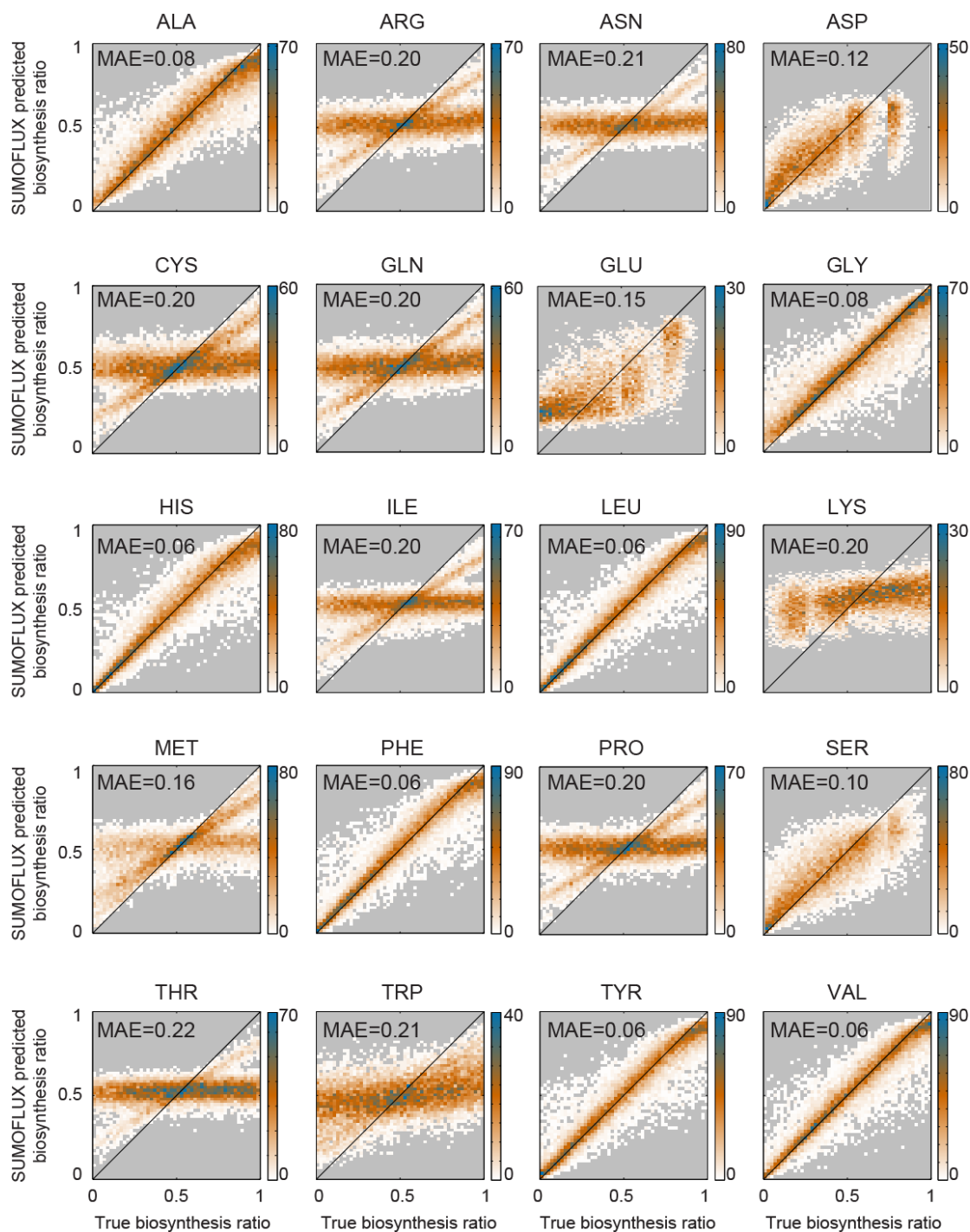


Figure S11. SUMOFLUX estimates amino acid biosynthesis fractions for *M. tuberculosis* infecting THP1 macrophage with good accuracy for 10 amino acids. Density plot representing SUMOFLUX estimates versus the true flux ratios for *in silico* data; biosynthesis fraction for each amino acid was estimated for the infection setup, where THP1 macrophage was pre-labeled in RPMI media containing [U - ^{13}C] glucose and 20 naturally labeled amino acids.

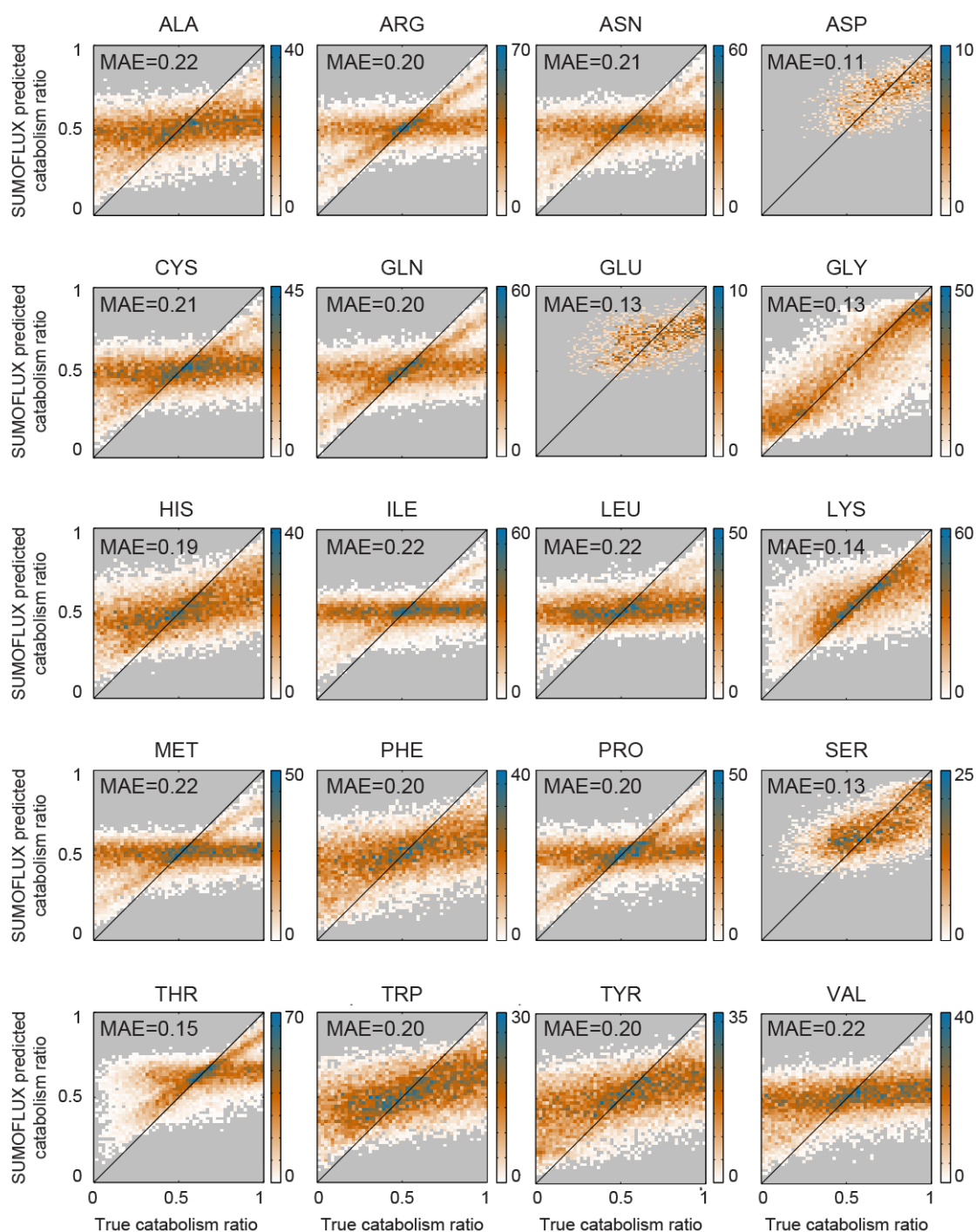


Figure S12. SUMOFLUX estimates amino acid catabolism fractions for *M. tuberculosis* infecting THP1 macrophage with poor accuracy. Density plot representing SUMOFLUX estimates versus the true flux ratios for *in silico* data; catabolism fraction for each amino acid was estimated for the infection setup, where THP1 macrophage was pre-labeled in RPMI media containing [U-¹³C] glucose and 20 naturally labeled amino acids.

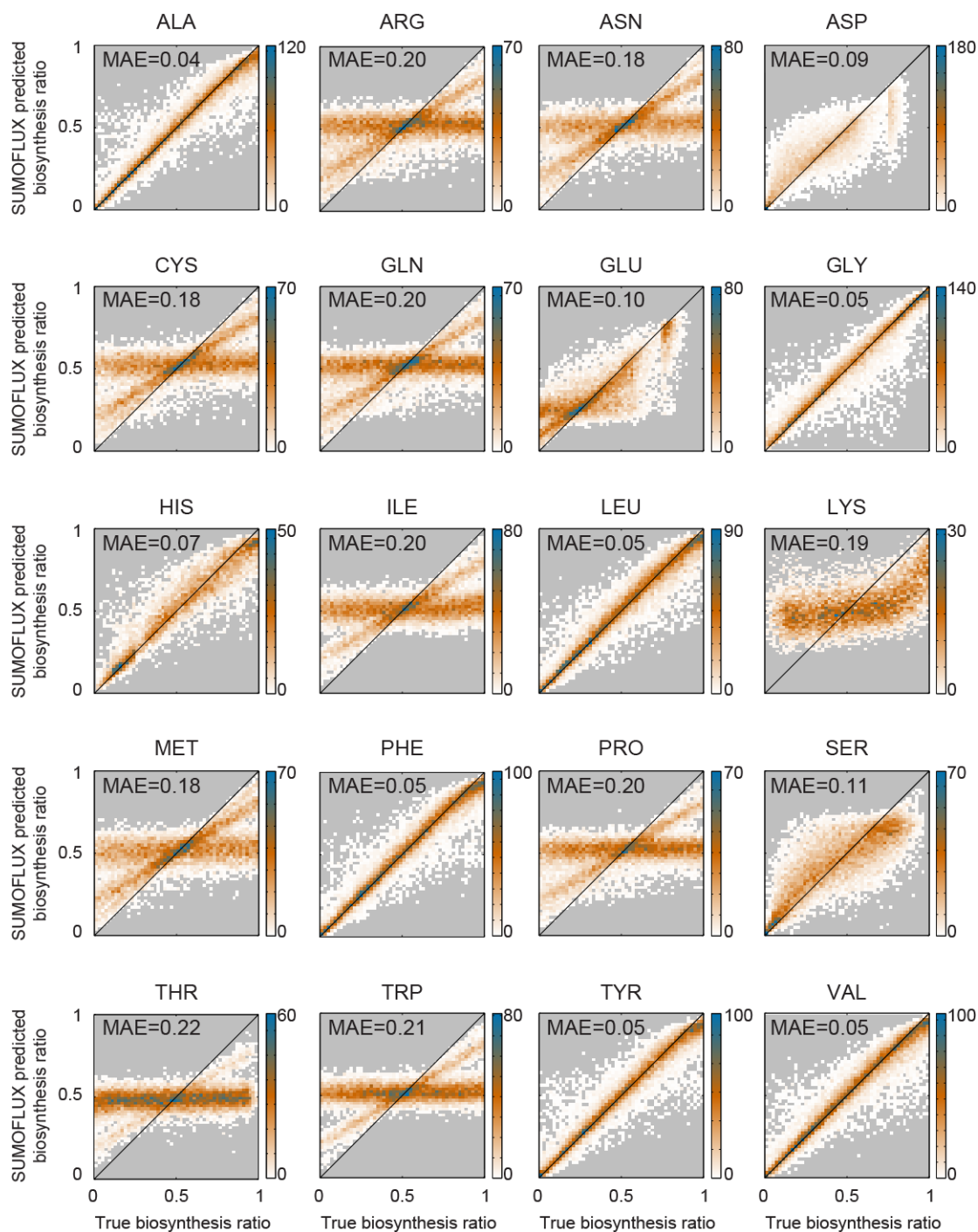


Figure S13. SUMOFLUX estimates amino acid biosynthesis fractions for *M. tuberculosis* grown in rich RPMI media with good accuracy for 10 amino acids. Density plot representing SUMOFLUX estimates versus the true flux ratios for *in silico* data; biosynthesis fraction for each amino acid was estimated for the experiment with RPMI media containing [U-¹³C] glucose and 20 naturally labeled amino acids.

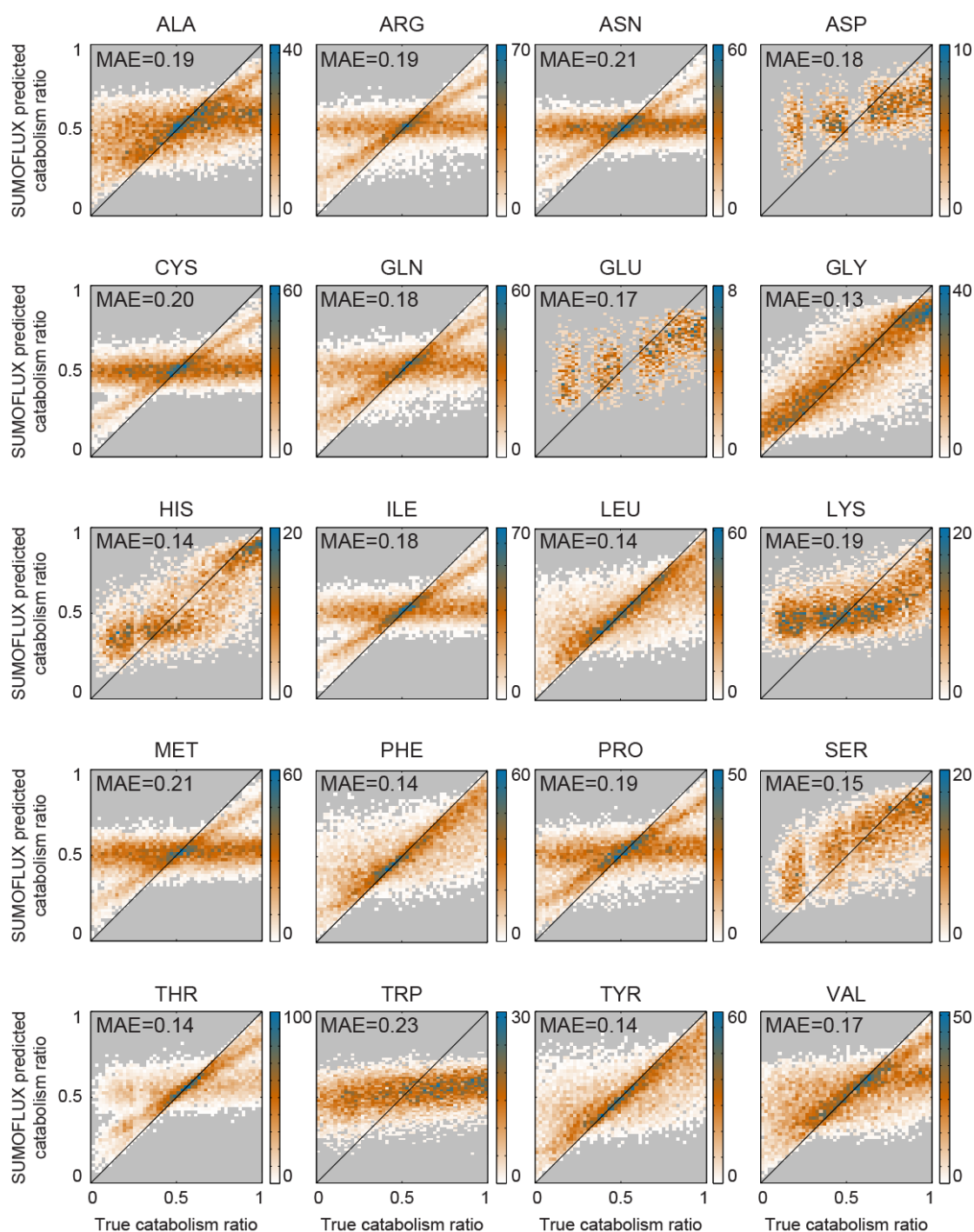


Figure S14. SUMOFLUX estimates amino acid catabolism fractions for *M. tuberculosis* grown in rich RPMI media with poor accuracy. Density plot representing SUMOFLUX estimates versus the true flux ratios for *in silico* data; catabolism fraction for each amino acid was estimated for the experiment with RPMI media containing [U-¹³C] glucose and 20 naturally labeled amino acids.

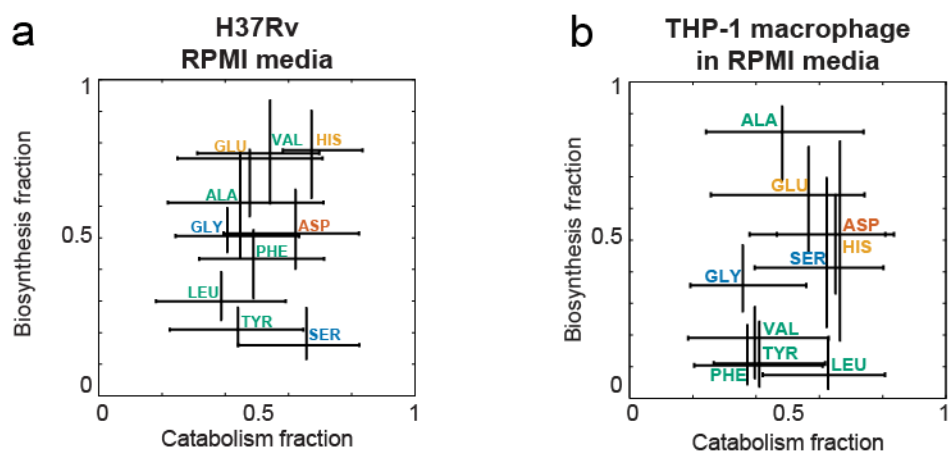


Figure S15. Amino acid exchange flux ratio quantification in RPMI media. (a) SUMOFLUX estimates of biosynthesis and catabolism fractions for 10 amino acids in *M. tuberculosis* H37Rv grown in RPMI media with [U-¹³C] glucose and 20 naturally labeled amino acids. (b) SUMOFLUX estimates of biosynthesis and catabolism fractions for 10 amino acids in THP-1 macrophage growing in RPMI media with [U-¹³C] glucose and naturally labeled amino acids. Error bars indicate 50% prediction interval.

Appendix III

Supplementary information Chapter 4

Table S1. *E. coli* network of central carbon metabolism used throughout the study.

Reaction type	Reaction	Stoichiometry and carbon transitions
Uptake and secretion	glc_up	glucose (abcdef) -> G6P (abcdef)
	CO2up	CO2in (a) -> CO2 (a)
	accoa_ac	AcCoA (ab) -> Ac (ab)
	ac_out	Ac (ab) -> Acetate (ab)
	co2_out	CO2 (a) -> CO2out (a)
Biomass precursors	G6P_bm	G6P (abcdef) -> G6Pbm (abcdef)
	PGA_bm	PGA (abc) -> PGAbm (abc)
	P5P_bm	P5P (abcde) -> P5Pbm (abcde)
	PEP_bm	PEP (abc) -> PEPbm (abc)
	PYR_bm	PSuYR (abc) -> PYRbm (abc)
	OGA_bm	OGA (abcde) -> OGAbm (abcde)
	OAA_bm	OAA (abcd) -> OAAbm (abcd)
	E4P_bm	E4P (abcd) -> E4Pbm (abcd)
	PHE_bm	PHE (abcdefghi) -> PHEbm (abcdefghi)
	ALA_bm	ALA (abc) -> ALAbm (abc)
	ASP_bm	ASP (abcd) -> ASPbm (abcd)
	GLU_bm	GLU (abcde) -> GLUbm (abcde)
	GLY_bm	GLY (ab) -> GLYbm (ab)
	HIS_bm	HIS (abcdef) -> HISbm (abcdef)
	ILE_bm	ILE (abcdef) -> ILEbm (abcdef)
	LEU_bm	LEU (abcdef) -> LEUbm (abcdef)
	LYS_bm	LYS (abcdef) -> LYSbm (abcdef)
	SER_bm	SER (abc) -> SERbm (abc)
	THR_bm	THR (abcd) -> THRbm (abcd)
	TYR_bm	TYR (abcdefghi) -> TYRbm (abcdefghi)
VAL_bm	VAL (abcde) -> VALbm (abcde)	
PRO_bm	PRO (abc) -> PRObm (abc)	
CYS_bm	CYS (abc) -> CYSbm (abc)	
GLN_bm	GLN (abcde) -> GLNbm (abcde)	
ASN_bm	ASN (abcd) -> ASNbm (abcd)	
ARG_bm	ARG (abcdef) -> ARGbm (abcdef)	
MET_bm	MET (abcde) -> METbm (abcde)	
TRP_bm	TRP (hijklabcdef) -> TRPbm (hijklabcdef)	
Glycolysis	pgm	G1P (abcdef) <-> G6P (abcdef)
	pgi	G6P (abcdef) <-> F6P (abcdef)
	pfk	F6P (abcdef) -> FBP (abcdef)
	fba	FBP (abcdef) -> DHAP (cba) + GAP (def)
	tpi	DHAP (abc) <-> GAP (abc)
	gapdh	GAP (abc) -> BPG (abc)
	bpg	BPG (abc) -> PGA (abc)
	eno	PGA (abc) -> PEP (abc)
	pyk	PEP (abc) -> PYR (abc)
Pentose phosphate pathway	zwf	G6P (abcdef) -> PG6 (abcdef)
	gnd	PG6 (abcdef) -> P5P (bcdef) + CO2 (a)
	TK1	P5P (abcde) + P5P (fghij) <-> GAP (cde) + S7P (abfghij)
	TK2	P5P (abcde) + E4P (fghi) <-> GAP (cde) + F6P (abfghi)
	TA	S7P (abcdefg) + GAP (hij) <-> E4P (defg) + F6P (abchij)
Entner-Doudoroff	edp1	PG6 (abcdef) -> KDPG (abcdef)
	edp2	KDPG (abcdef) -> PYR (abc) + GAP (def)
TCA cycle	pdh	PYR (abc) -> AcCoA (bc) + CO2 (a)
	cit1	OAA (cdef) + AcCoA (ab) -> Cit (fedcba)
	acnA	Cit (abcdef) <-> Act (abcdef)
	acnB	Act(abcdef) <-> Icit (abcdef)
	idh	Icit (abcdef) <-> OGA (abcef) + CO2 (d)

	sdh	OGA (abcde) -> Suc (bcde) + CO2 (a)
	frdA	Suc (abcd) <-> Fum (abcd)
	fum	Fum (abcd) <-> Mal (abcd)
	mdh	Mal (abcd) <-> OAA (abcd)
Glyoxylate shunt	gs1	Cit (abcdef) -> Glx (ab) + Suc (dcef)
	gs2	Glx (ab) + AcCoA (cd) -> Mal (abcd)
Glycerate pathway	gcl	Glx (ab) + Glx (cd) -> PGA (abd) + CO2 (c)
Anaplerosis and gluconeogenesis	mae	Mal (abcd) -> PYR (abc) + CO2 (d)
	pck	OAA (abcd) -> PEP (abc) + CO2 (d)
	ppc	PEP (abc) + CO2 (d) -> OAA (abcd)
Intermediates of aminoacid metabolism	kiv_out	PYR (abc) + PYR (def) -> KIV (abcef) + CO2 (d)
	cho_out	E4P (abcd) + PYR (efg) -> CHO (fgabcde)
Amino acid production	ala_out	PYR (abc) -> ALA (abc)
	asp_out	OAA (abcd) -> ASP (abcd)
	val_out	KIV (abcde) -> VAL (abcde)
	leu_out	KIV (abcde) + AcCoA (fg) -> LEU (bcdefg) + CO2 (a)
	ser_out	BPG (abc) -> SER (abc)
	gly_out	SER (abc) -> GLY (ab) + MTHF (c)
	his1_out	P5P (abcde) + MTHF (f) -> HIS (fedcba)
	his2_out	GLU (abcde) + MTHF (f) -> HIS (fedcba)
	thr_out	ASP (abcd) -> THR (abcd)
	lys1_out	PYR (abc) + ASP (defg) -> LYS (bcdefg) + CO2 (a)
	lys2_out	PYR (abc) + ASP (defg) -> LYS (abcefg) + CO2 (d)
	ile_out	THR (abcd) + PYR (efg) -> ILE (abcdfg) + CO2 (e)
	glu_out	OGA (abcde) -> GLU (abcde)
	pro_out	GLU (abcde) -> PRO (abcde)
	phe_out	CHO (abcdefg) + PYR (hij) -> PHE (hijabcdef) + CO2 (g)
	tyr_out	CHO (abcdefg) + PYR (hij) -> TYR (hijabcdef) + CO2 (g)
	arg_out	GLU (abcde) + urea (f) -> ARG (abcdef)
	asn_out	ASP (abcd) -> ASN (abcd)
	cys_out	SER (abc) -> CYS (abc)
	gln_out	GLU (abcde) -> GLN (abcde)
met_out	ASP (abcd) + MTHF (e) -> MET (abcde)	
trp_out	CHO (abcdefg) + P5P (hijkl) -> TRP (hijklabcdef) + CO2 (g)	

Table S2. Simulated isotopologues contained in the *E. coli* central metabolism model detected in the FIA-TOF spectra. Filtering steps: A – detected and annotated ions; S – spectral detector ringing filter; G – Gaussian fitting filter; R – ratiometric filter; U – uniquely annotated ions.

Model ID	Isotopologue name	Formula	m/z	A	S	G	R	U
ARG16_2	L-Arginine_2	C6H14N4O2	175.111	1	1	1	1	1
ARG16_3	L-Arginine_3	C6H14N4O2	176.114	1	1	1	1	1
E4P14_1	D-Erythrose 4-phosphate_1	C4H9O7P	200.004	1	1	1	1	1
GLY12_1	Glycine_1	C2H5NO2	75.028	1	1	1	1	1
GLY12_2	Glycine_2	C2H5NO2	76.031	1	1	1	1	1
HIS16_2	L-Histidine_2	C6H9N3O2	156.068	1	1	1	1	1
HIS16_3	L-Histidine_3	C6H9N3O2	157.072	1	1	1	1	1
OGA15_0	2-Oxoglutarate_0	C5H6O5	145.014	1	1	1	1	1
PEP13_0	Phosphoenolpyruvate_0	C3H5O6P	166.975	1	1	1	1	1
PRO15_2	L-Proline_2	C5H9NO2	116.062	1	1	1	1	1
PRO15_3	L-Proline_3	C5H9NO2	117.066	1	1	1	1	1
Suc14_0	Succinate_0	C4H6O4	117.019	1	1	1	1	1
Suc14_1	Succinate_1	C4H6O4	118.022	1	1	1	1	1
Suc14_2	Succinate_2	C4H6O4	119.025	1	1	1	1	1
Suc14_3	Succinate_3	C4H6O4	120.029	1	1	1	1	1
TRP111_2	L-Tryptophan_2	C11H12N2O2	205.089	1	1	1	1	1
TRP111_8	L-Tryptophan_8	C11H12N2O2	211.109	1	1	1	1	1
ARG16_4	L-Arginine_4	C6H14N4O2	177.117	1	1	0	0	1
ARG16_5	L-Arginine_5	C6H14N4O2	178.121	1	1	0	0	1
ASP14_4	L-Aspartate_4	C4H7NO4	136.043	1	1	0	0	1
E4P14_0	D-Erythrose 4-phosphate_0	C4H9O7P	199.001	1	1	0	0	1
E4P14_4	D-Erythrose 4-phosphate_4	C4H9O7P	203.014	1	1	0	0	1
LYS16_5	L-Lysine_5	C6H14N2O2	150.114	1	1	0	0	1
OGA15_2	2-Oxoglutarate_2	C5H6O5	147.020	1	1	0	0	1
OGA15_3	2-Oxoglutarate_3	C5H6O5	148.024	1	1	0	0	1
OGA15_5	2-Oxoglutarate_5	C5H6O5	150.030	1	1	0	0	1
PRO15_4	L-Proline_4	C5H9NO2	118.069	1	1	0	0	1
ALA13_0	L-Alanine_0	C3H7NO2	88.040	1	1	1	1	0
ALA13_1	L-Alanine_1	C3H7NO2	89.043	1	1	1	1	0
ALA13_2	L-Alanine_2	C3H7NO2	90.047	1	1	1	1	0
ALA13_3	L-Alanine_3	C3H7NO2	91.050	1	1	1	1	0
ARG16_0	L-Arginine_0	C6H14N4O2	173.104	1	1	1	1	0
ARG16_1	L-Arginine_1	C6H14N4O2	174.107	1	1	1	1	0
ASN14_0	L-Asparagine_0	C4H8N2O3	131.046	1	1	1	1	0
ASN14_1	L-Asparagine_1	C4H8N2O3	132.049	1	1	1	1	0
ASP14_0	L-Aspartate_0	C4H7NO4	132.030	1	1	1	1	0
ASP14_1	L-Aspartate_1	C4H7NO4	133.033	1	1	1	1	0
ASP14_2	L-Aspartate_2	C4H7NO4	134.036	1	1	1	1	0
BPG13_0	3-Phospho-D-glyceroyl phosphate_0	C3H8O10P2	264.951	1	1	1	1	0
BPG13_1	3-Phospho-D-glyceroyl phosphate_1	C3H8O10P2	265.955	1	1	1	1	0
BPG13_2	3-Phospho-D-glyceroyl phosphate_2	C3H8O10P2	266.958	1	1	1	1	0
BPG13_3	3-Phospho-D-glyceroyl phosphate_3	C3H8O10P2	267.962	1	1	1	1	0
Cit16_1	Citrate_1	C6H8O7	192.023	1	1	1	1	0
Cit16_2	Citrate_2	C6H8O7	193.026	1	1	1	1	0
Cit16_3	Citrate_3	C6H8O7	194.029	1	1	1	1	0
Cit16_4	Citrate_4	C6H8O7	195.033	1	1	1	1	0
Cit16_5	Citrate_5	C6H8O7	196.036	1	1	1	1	0
Cit16_6	Citrate_6	C6H8O7	197.039	1	1	1	1	0
DHAP13_0	Glycerone phosphate_0	C3H7O6P	168.990	1	1	1	1	0
F6P16_0	beta-D-Fructose 6-phosphate_0	C6H13O9P	259.022	1	1	1	1	0
F6P16_1	beta-D-Fructose 6-phosphate_1	C6H13O9P	260.025	1	1	1	1	0
FBP16_0	beta-D-Fructose 1,6-bisphosphate_0	C6H14O12P2	338.988	1	1	1	1	0
FBP16_1	beta-D-Fructose 1,6-bisphosphate_1	C6H14O12P2	339.992	1	1	1	1	0
FBP16_2	beta-D-Fructose 1,6-bisphosphate_2	C6H14O12P2	340.995	1	1	1	1	0
FBP16_6	beta-D-Fructose 1,6-bisphosphate_6	C6H14O12P2	345.008	1	1	1	1	0

G6P16_0	alpha-D-Glucose 6-phosphate_0	C6H13O9P	259.022	1	1	1	1	0
G6P16_1	alpha-D-Glucose 6-phosphate_1	C6H13O9P	260.025	1	1	1	1	0
GLN15_0	L-Glutamine_0	C5H10N2O3	145.061	1	1	1	1	0
GLN15_1	L-Glutamine_1	C5H10N2O3	146.065	1	1	1	1	0
GLU15_0	L-Glutamate_0	C5H9NO4	146.045	1	1	1	1	0
GLU15_1	L-Glutamate_1	C5H9NO4	147.049	1	1	1	1	0
GLU15_2	L-Glutamate_2	C5H9NO4	148.052	1	1	1	1	0
GLU15_3	L-Glutamate_3	C5H9NO4	149.055	1	1	1	1	0
GLU15_4	L-Glutamate_4	C5H9NO4	150.059	1	1	1	1	0
GLU15_5	L-Glutamate_5	C5H9NO4	151.062	1	1	1	1	0
GLY12_0	Glycine_0	C2H5NO2	74.024	1	1	1	1	0
ILE16_0	L-Isoleucine_0	C6H13NO2	130.087	1	1	1	1	0
ILE16_1	L-Isoleucine_1	C6H13NO2	131.090	1	1	1	1	0
ILE16_2	L-Isoleucine_2	C6H13NO2	132.094	1	1	1	1	0
ILE16_3	L-Isoleucine_3	C6H13NO2	133.097	1	1	1	1	0
ILE16_4	L-Isoleucine_4	C6H13NO2	134.100	1	1	1	1	0
ILE16_5	L-Isoleucine_5	C6H13NO2	135.104	1	1	1	1	0
LEU16_0	L-Leucine_0	C6H13NO2	130.087	1	1	1	1	0
LEU16_1	L-Leucine_1	C6H13NO2	131.090	1	1	1	1	0
LEU16_2	L-Leucine_2	C6H13NO2	132.094	1	1	1	1	0
LEU16_3	L-Leucine_3	C6H13NO2	133.097	1	1	1	1	0
LEU16_4	L-Leucine_4	C6H13NO2	134.100	1	1	1	1	0
LEU16_5	L-Leucine_5	C6H13NO2	135.104	1	1	1	1	0
Mal14_0	(S)-Malate_0	C4H6O5	133.014	1	1	1	1	0
Mal14_1	(S)-Malate_1	C4H6O5	134.017	1	1	1	1	0
Mal14_2	(S)-Malate_2	C4H6O5	135.020	1	1	1	1	0
Mal14_3	(S)-Malate_3	C4H6O5	136.024	1	1	1	1	0
P5P15_0	alpha-D-Ribose 1-phosphate_0	C5H11O8P	229.011	1	1	1	1	0
P5P15_1	alpha-D-Ribose 1-phosphate_1	C5H11O8P	230.015	1	1	1	1	0
P5P15_2	alpha-D-Ribose 1-phosphate_2	C5H11O8P	231.018	1	1	1	1	0
P5P15_3	alpha-D-Ribose 1-phosphate_3	C5H11O8P	232.021	1	1	1	1	0
P5P15_4	alpha-D-Ribose 1-phosphate_4	C5H11O8P	233.025	1	1	1	1	0
P5P15_5	alpha-D-Ribose 1-phosphate_5	C5H11O8P	234.028	1	1	1	1	0
PGA13_0	3-Phospho-D-glycerate_0	C3H7O7P	185.988	1	1	1	1	0
PGA13_1	3-Phospho-D-glycerate_1	C3H7O7P	185.988	1	1	1	1	0
PGA13_2	3-Phospho-D-glycerate_2	C3H7O7P	186.992	1	1	1	1	0
PGA13_3	3-Phospho-D-glycerate_3	C3H7O7P	187.995	1	1	1	1	0
PHE19_0	L-Phenylalanine_0	C9H11NO2	164.071	1	1	1	1	0
PHE19_1	L-Phenylalanine_1	C9H11NO2	165.075	1	1	1	1	0
PHE19_2	L-Phenylalanine_2	C9H11NO2	166.078	1	1	1	1	0
PHE19_4	L-Phenylalanine_4	C9H11NO2	168.085	1	1	1	1	0
PHE19_9	L-Phenylalanine_9	C9H11NO2	173.101	1	1	1	1	0
PRO15_1	L-Proline_1	C5H9NO2	115.059	1	1	1	1	0
PYR13_1	Pyruvate_1	C3H4O3	88.012	1	1	1	1	0
PYR13_3	Pyruvate_3	C3H4O3	90.018	1	1	1	1	0
SER13_0	L-Serine_0	C3H7NO3	104.035	1	1	1	1	0
SER13_1	L-Serine_1	C3H7NO3	105.038	1	1	1	1	0
SER13_2	L-Serine_2	C3H7NO3	106.041	1	1	1	1	0
THR14_0	L-Threonine_0	C4H9NO3	118.050	1	1	1	1	0
THR14_2	L-Threonine_2	C4H9NO3	120.057	1	1	1	1	0
THR14_4	L-Threonine_4	C4H9NO3	122.064	1	1	1	1	0
TYR19_2	L-Tyrosine_2	C9H11NO3	182.073	1	1	1	1	0
TYR19_3	L-Tyrosine_3	C9H11NO3	183.076	1	1	1	1	0
TYR19_6	L-Tyrosine_6	C9H11NO3	186.086	1	1	1	1	0
VAL15_0	L-Valine_0	C5H11NO2	116.071	1	1	1	1	0
VAL15_1	L-Valine_1	C5H11NO2	117.075	1	1	1	1	0
VAL15_2	L-Valine_2	C5H11NO2	118.078	1	1	1	1	0
VAL15_3	L-Valine_3	C5H11NO2	119.081	1	1	1	1	0
VAL15_4	L-Valine_4	C5H11NO2	120.085	1	1	1	1	0
VAL15_5	L-Valine_5	C5H11NO2	121.088	1	1	1	1	0

FBP16_3	beta-D-Fructose 1,6-bisphosphate_3	C6H14O12P2	341.998	1	1	1	0	0
HIS16_5	L-Histidine_5	C6H9N3O2	159.078	1	1	1	0	0
LYS16_3	L-Lysine_3	C6H14N2O2	148.108	1	1	1	0	0
LYS16_4	L-Lysine_4	C6H14N2O2	149.111	1	1	1	0	0
Mal14_4	(S)-Malate_4	C4H6O5	137.027	1	1	1	0	0
PHE19_6	L-Phenylalanine_6	C9H11NO2	170.091	1	1	1	0	0
ASN14_3	L-Asparagine_3	C4H8N2O3	134.056	1	1	0	0	0
ASN14_4	L-Asparagine_4	C4H8N2O3	135.059	1	1	0	0	0
ASP14_3	L-Aspartate_3	C4H7NO4	135.040	1	1	0	0	0
CYS13_0	L-Cysteine_0	C3H7NO2S	120.012	1	1	0	0	0
DHAP13_2	Glycerone phosphate_2	C3H7O6P	170.997	1	1	0	0	0
DHAP13_3	Glycerone phosphate_3	C3H7O6P	172.000	1	1	0	0	0
F6P16_2	beta-D-Fructose 6-phosphate_2	C6H13O9P	261.029	1	1	0	0	0
F6P16_3	beta-D-Fructose 6-phosphate_3	C6H13O9P	262.032	1	1	0	0	0
F6P16_5	beta-D-Fructose 6-phosphate_5	C6H13O9P	264.039	1	1	0	0	0
F6P16_6	beta-D-Fructose 6-phosphate_6	C6H13O9P	265.042	1	1	0	0	0
FBP16_5	beta-D-Fructose 1,6-bisphosphate_5	C6H14O12P2	344.005	1	1	0	0	0
G6P16_2	alpha-D-Glucose 6-phosphate_2	C6H13O9P	261.029	1	1	0	0	0
G6P16_3	alpha-D-Glucose 6-phosphate_3	C6H13O9P	262.032	1	1	0	0	0
G6P16_5	alpha-D-Glucose 6-phosphate_5	C6H13O9P	264.039	1	1	0	0	0
G6P16_6	alpha-D-Glucose 6-phosphate_6	C6H13O9P	265.042	1	1	0	0	0
HIS16_0	L-Histidine_0	C6H9N3O2	154.062	1	1	0	0	0
HIS16_1	L-Histidine_1	C6H9N3O2	155.065	1	1	0	0	0
MET14_2	L-Methionine_2	C5H11NO2S	150.050	1	1	0	0	0
PHE19_8	L-Phenylalanine_8	C9H11NO2	172.098	1	1	0	0	0
PRO15_0	L-Proline_0	C5H9NO2	114.056	1	1	0	0	0
PYR13_0	Pyruvate_0	C3H4O3	87.008	1	1	0	0	0
PYR13_2	Pyruvate_2	C3H4O3	89.015	1	1	0	0	0
SER13_3	L-Serine_3	C3H7NO3	107.045	1	1	0	0	0
THR14_3	L-Threonine_3	C4H9NO3	121.060	1	1	0	0	0
TRP111_6	L-Tryptophan_6	C11H12N2O2	209.102	1	1	0	0	0
TYR19_1	L-Tyrosine_1	C9H11NO3	181.069	1	1	0	0	0
TYR19_4	L-Tyrosine_4	C9H11NO3	184.079	1	1	0	0	0
TYR19_7	L-Tyrosine_7	C9H11NO3	187.090	1	1	0	0	0
TYR19_9	L-Tyrosine_9	C9H11NO3	189.096	1	1	0	0	0
Cit16_0	Citrate_0	C6H8O7	191.019	1	0	0	0	0
ILE16_6	L-Isoleucine_6	C6H13NO2	136.107	1	0	0	0	0
LEU16_6	L-Leucine_6	C6H13NO2	136.107	1	0	0	0	0
LYS16_0	L-Lysine_0	C6H14N2O2	145.098	1	0	0	0	0
LYS16_1	L-Lysine_1	C6H14N2O2	146.101	1	0	0	0	0
LYS16_2	L-Lysine_2	C6H14N2O2	147.104	1	0	0	0	0
LYS16_6	L-Lysine_6	C6H14N2O2	151.118	1	0	0	0	0
PRO15_5	L-Proline_5	C5H9NO2	119.072	1	0	0	0	0

Table S3. *E. coli* strains with mutations in phosphorylation cites used in the study. Growth rate relative to wild type strain, wild type strain growth rate = 0.62. IN – phospho-IN mutation, OUT – phospho-out mutation, KO – knock out.

Gene	Location	Protein	Growth rate (% WT)		
			IN	OUT	KO
acnB	622	bifunctional aconitate hydratase 2 and 2-methylisocitrate dehydratase	62.79	95.34	57.39
adk	30	adenylate kinase	-	90.58	-
ahpC	84	alkyl hydroperoxide reductase, AhpC component	90.24	82.19	95.89
asd	38	Aspartate-semialdehyde dehydrogenase	92.06	94.08	-
eno	372 & 375 & 379	enolase	98.38	-	-
garL	124	5-keto-4-deoxy-D-glucarate aldolase	97.91	93.50	108.80
gatA	70	PTS system galactitol-specific EIIA component	107.08	103.08	-
gatB	63	galactitol PTS permease - GatB subunit	58.10	109.80	106.91
gatY	115	D-tagatose-1,6-bisphosphate aldolase subunit GatY	86.50	93.05	85.17
gatZ	265	D-tagatose-1,6-bisphosphate aldolase subunit GatZ	101.38	105.66	104.97
gpmM	64	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	95.56	98.59	117.32
guaA	357	GMP synthase [glutamine-hydrolyzing]	108.58	-	114.90
hipA	150	Serine/threonine-protein kinase HipA	90.93	94.17	99.66
hisJ	217	Histidine-binding periplasmic protein	99.65	98.47	100.20
icd	113	Isocitrate dehydrogenase	0.00	91.80	123.29
kdsD	2	Arabinose 5-phosphate isomerase KdsD	95.31	89.34	90.63
manX	72	PTS system mannose-specific EIIAB component	98.40	91.84	108.10
mdh	280	malate dehydrogenase	100.58	104.37	113.32
menF	122	Isochorismate synthase MenF	79.04	90.82	95.31
metK	198	S-adenosylmethionine synthase	94.80	97.16	-
ndk	93	Nucleoside diphosphate kinase	104.46	-	93.68
pck	250 & 252	phosphoenolpyruvate carboxykinase (ATP)	96.97	104.51	99.41
pgi	105 & 107	Glucose-6-phosphate isomerase	97.05	94.62	38.34
pgk	192 & 196 & 199	phosphoglycerate kinase	68.30	91.68	-
pgm	144&146	Phosphoglucomutase	107.91	-	103.38
pps	419	phosphoenolpyruvate synthetase	98.16	92.59	97.59
pta	691	phosphate acetyltransferase / phosphate propionyltransferase	108.42	105.84	107.48
pykF	36	pyruvate kinase I	84.37	87.44	67.29
talB	226	transaldolase	107.23	109.95	102.61
udk	106 & 108	Uridine kinase	100.16	105.51	106.69
uspG	128	Universal stress protein	100.36	99.97	99.09

Table S4. Pearson correlation between relative flux ratio changes of phosho-IN and phospho-OUT mutants. Wild type strain was used as a reference to calculate the difference between the flux ratios. Pearson correlation and p-values were calculated with Matlab function corr.

Mutant	Pearson correlation	P-value
icd	-0.57	0.02
pgk	-0.42	0.10
gatB	-0.20	0.45
menF	-0.07	0.80
udk	0.04	0.89
pps	0.28	0.28
acnB	0.38	0.13
mdh3	0.53	0.03
pck	0.56	0.02
gatA	0.59	0.01
acnB1	0.60	0.01
metK	0.67	0.00
hipA	0.69	0.00
uspG	0.71	0.00
ahpC	0.73	0.00
pta	0.75	0.00
garL	0.77	0.00
hisJ	0.78	0.00
kdsD	0.79	0.00
manX	0.79	0.00
mdh	0.81	0.00
pykF	0.83	0.00
gatZ	0.87	0.00
gpmM	0.87	0.00
asd	0.89	0.00
pgi	0.92	0.00
gatY	0.93	0.00
talB	0.95	0.00

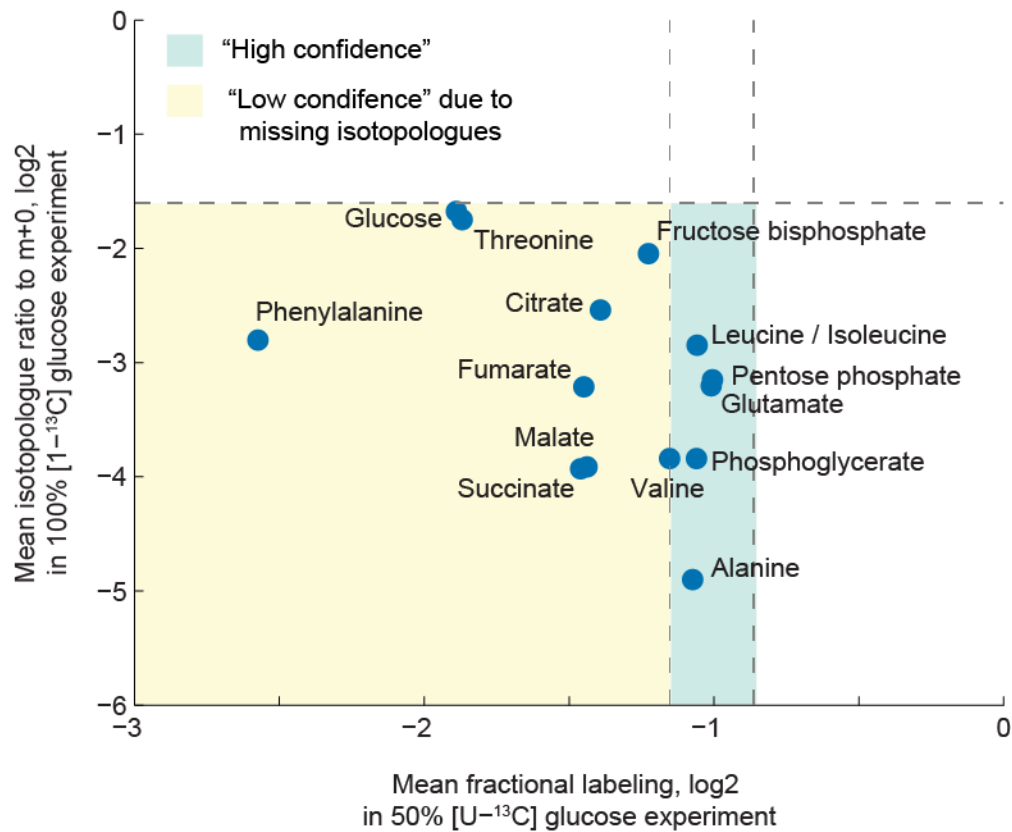


Figure S1. Annotated ions quality check based on ratiometric filtering. Metabolites of interest were ranked according to isotopologue ratio to the $m+0$ isotopologue in the $[1-^{13}\text{C}]$ glucose experiment and fractional labeling in the $[U-^{13}\text{C}]$ glucose experiments. Isotopologue ions with a value of heavy isotopologue to $m+0$ ratio > 0.33 have been discarded.

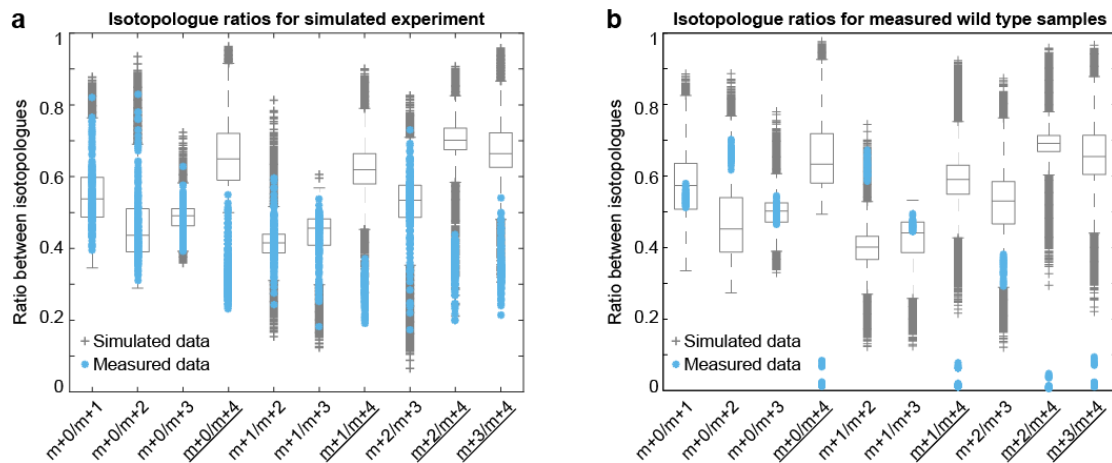


Figure S2. Robust feature selection and SUMOFLUX simulation filtering. Example of simulation filter comparing *in silico* and measured data represented as pairwise isotopologue ratios. (a) *In silico* example of overlapping ion. (b) Example of measured malate in *E. coli* wild type samples grown on $[1-^{13}\text{C}]$ glucose.

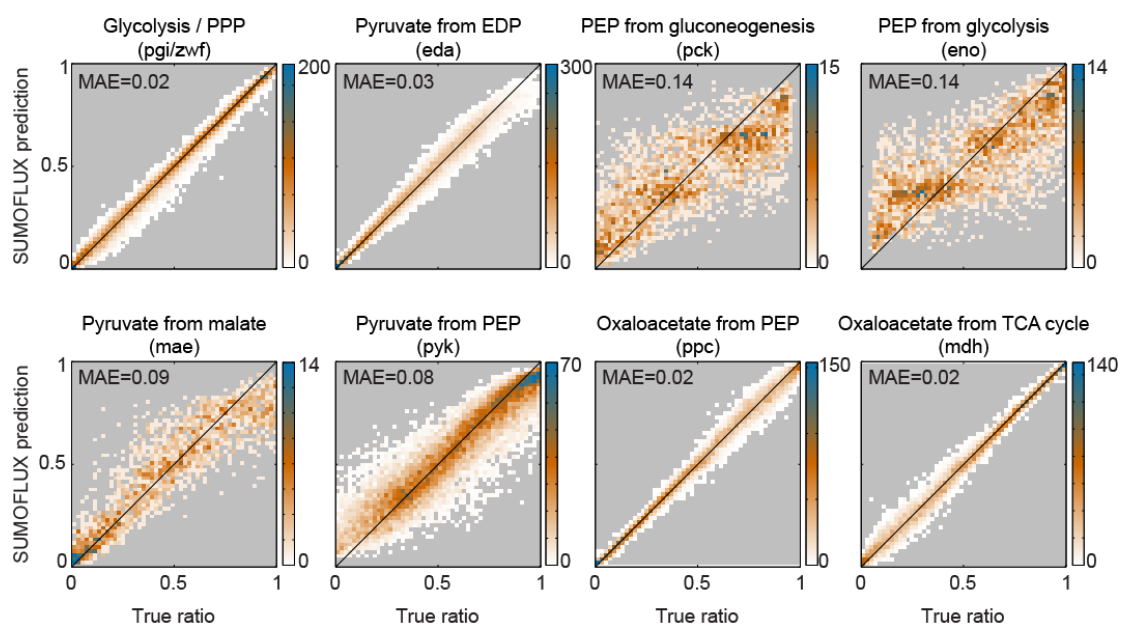


Figure S3. SUMOFLUX estimates eight flux ratios in central metabolism of *E. coli* with FIA-TOF data with high accuracy. Density plots representing SUMOFLUX estimates versus the true flux ratios for in silico data for eight flux ratios and fractions in central metabolism used for the proof-of-principle. Upper panel, ratios were estimated using the data from $[1-^{13}\text{C}]$ glucose experiment; lower panel, ratios were estimated using the data from $[U-^{13}\text{C}]$ glucose experiment. The features passing the ratiometric filtering were used for the predictions.

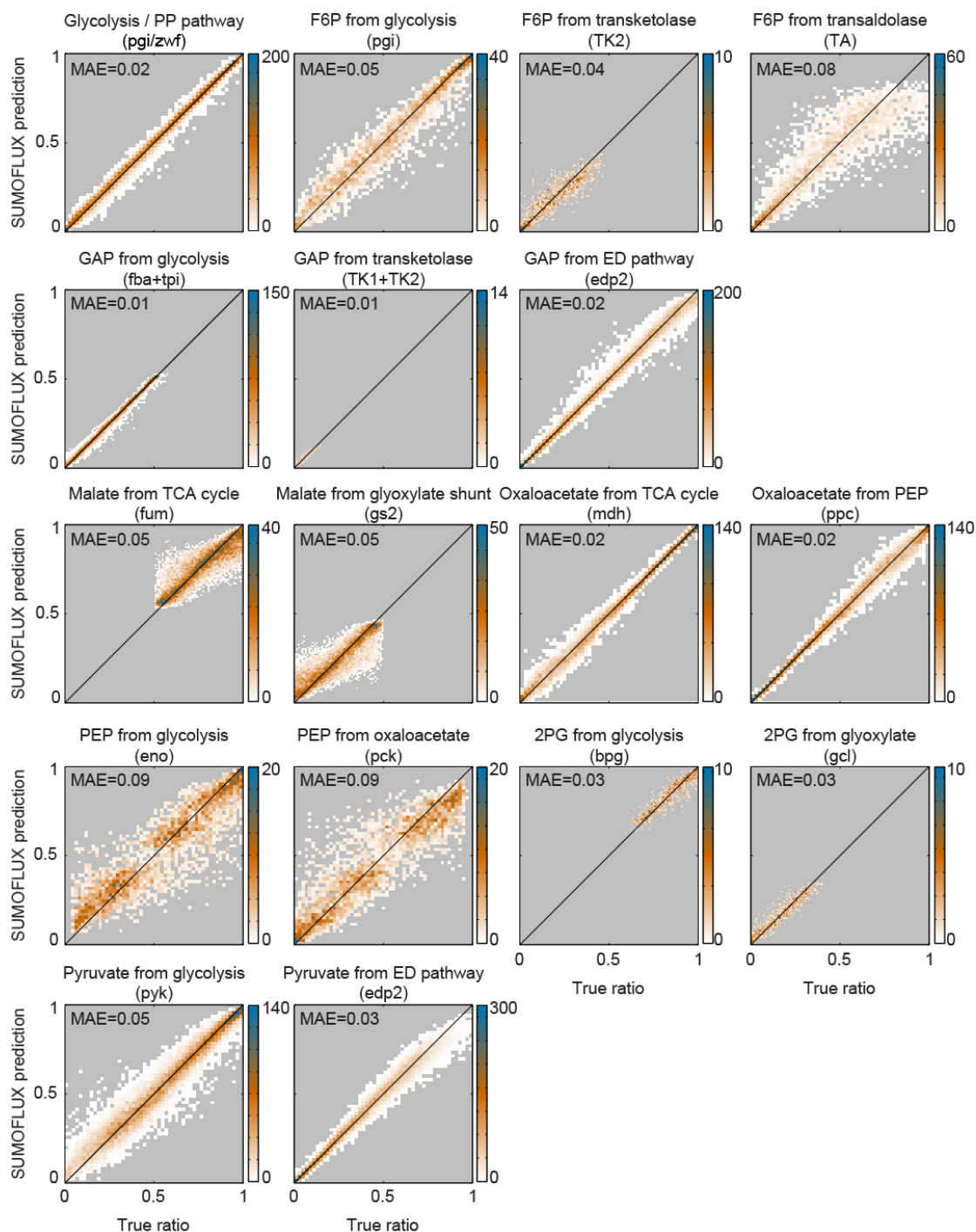


Figure S4. SUMOFLUX predictors' accuracies for the flux ratios calculated for *E. coli* phosphomutant profiling. Density plots representing SUMOFLUX estimates versus the true flux ratios for in silico data combined from 100% [1-¹³C] and a mixture of 50% [U-¹³C] and 50% unlabeled glucose experiments. The features passing the ratiometric filtering were used for the predictions.

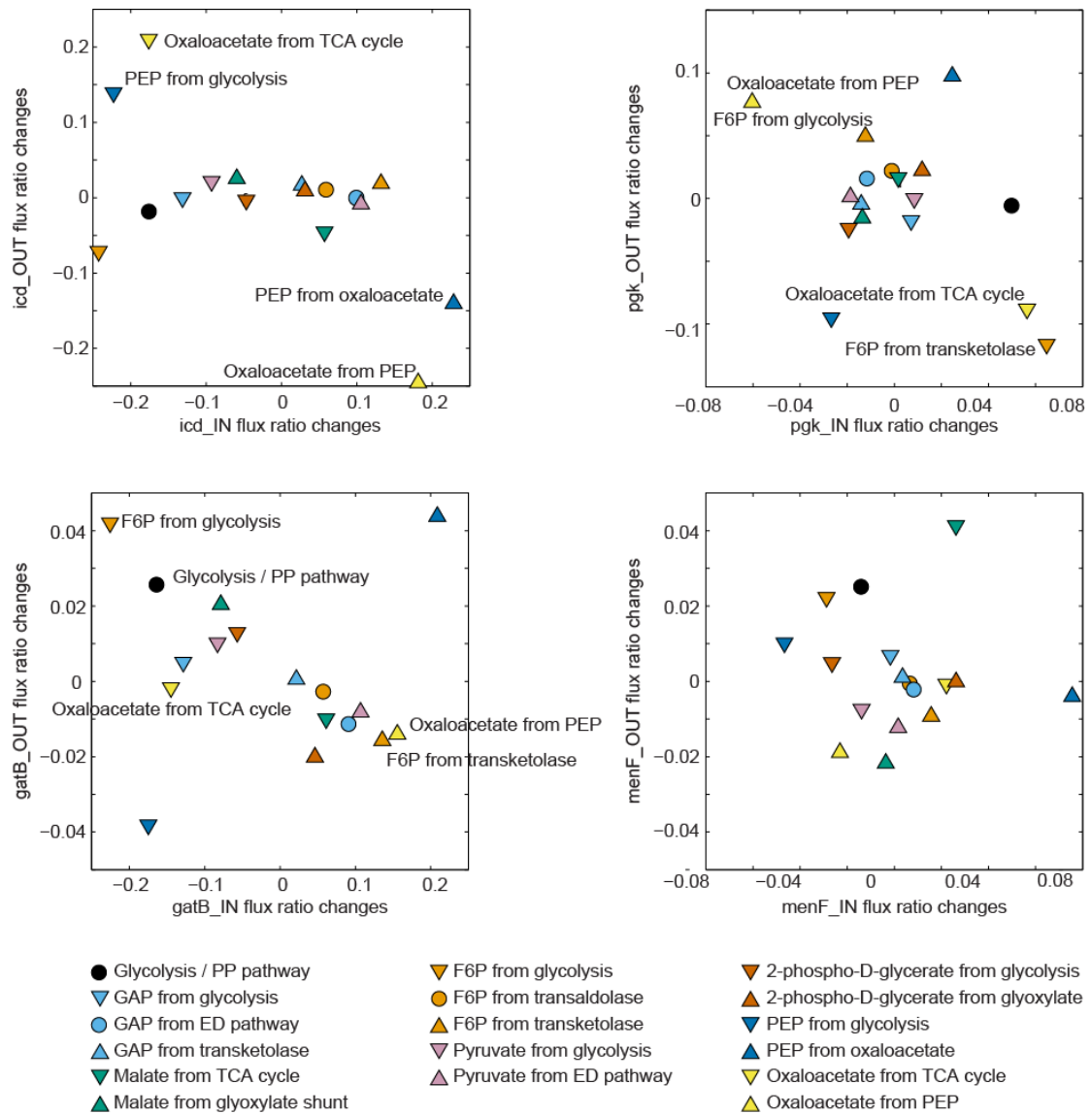


Figure S5. Flux ratio estimates of the phospho-IN and phospho-OUT mutants showing negative correlation.

Abbreviations

	¹³ C MFA	¹³ C metabolic flux analysis
	¹³ C MFRA	¹³ C metabolic flux ratio analysis
	6PG	6-phosphogluconate
A	AA	amino acid
	AcCoA	Acetyl-CoA
	αKG	alpha-ketoglutarate
	ALA	alanine
	ARG	arginine
	ASN	asparagine
	ASP	aspartate
	ATP	adenosine triphosphate
B	<i>B. subtilis</i>	<i>Bacillus subtilis</i>
	BPG	biphosphoglycerate
C	C	carbon
	c	concentration
	CCM	central carbon metabolism
	CDW	cell dry weight
	CoA	coenzyme A
	CYS	cysteine
D	Da	dalton
	DHAP	dihydroxyacetone phosphate
	DNA	deoxyribonucleic acid
E	e	charge
	<i>E. coli</i>	<i>Escherichia coli</i>
	E4P	erythrose-4-phosphate
	E-D pathway	Entner-Doudoroff pathway
	EtOH	ethanol
F	F1P	fructose-1-phosphate
	F6P	fructose-6-phosphate
	FBA	flux balance analysis
	FBP	fructose-1,6-bisphosphate

	FDR	false discovery rate
	FIA-TOF	flow injection analysis - time of flight
	FL	fractional labeling
	Fum	fumarate
G	g	grams
	G6P	glucose-6-phosphate
	GAP	glyceraldehyde-3-phosphate
	GC	gas chromatography
	GLC	glucose
	GLN	glutamine
	GLU	glutamate
	GLY	glycine
	Gox	glyoxylate
	GR	growth rate
H	h	hour
	HIS	histidine
	HPLC	high-performance liquid chromatography
I	ICT	isocitrate
	ILE	isoleucine
	IR	isotopologue ratio
K	KDPG	2-Keto-3-deoxy-6-phosphogluconate
	KEGG	Kyoto Encyclopedia of Genes and Genomes
L	L	liter
	LB	Luria-Bertani
	lb	lower bound
	LC	liquid chromatography
	LEU	leucine
	LYS	lysine
M	<i>M. smegmatis</i>	<i>Mycobacterium smegmatis</i>
	<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
	m/z	mass to charge ratio
	MAE	mean absolute error
	MAL	malate

	MDV	mass distribution vector
	MET	methionine
	mg	milligram
	min	minute
	mL	milliliter
	mM	millimolar
	MRM	multiple reaction monitoring
	MS	mass spectrometry
	MS/MS	tandem mass spectrometry
N	N	nitrogen
	NAD ⁺	nicotinamide adenine dinucleotide oxidized form
	NAD ⁺ /NADH	nicotinamide adenine dinucleotide
	NADH	nicotinamide adenine dinucleotide reduced form
	NADP ⁺	nicotinamide adenine dinucleotide phosphate oxidized form
	NADPH	nicotinamide adenine dinucleotide phosphate reduced form
	NMR	nuclear magnetic resonance
O	OAA	oxaloacetate
	OD	optical density
	ODE	ordinary differential equation
P	P5P	pentose-5-phosphate
	PCC	Pearson correlation coefficient
	PEP	phosphoenolpyruvate
	PGA	phospho-glycerate
	PHE	phenylalanine
	PPP	pentose phosphate pathway
	PRO	proline
	PYR	pyruvate
R	R5P	ribose-5-phosphate
	RNA	Ribonucleic acid
	RPMI medium	Roswell Park Memorial Institute medium
	RT	room temperature
	Ru5P	ribulose-5-phosphate
S	s	second

	S7P	sedoheptulose-7-phosphate
	SER	serine
T	TBDMS	tert-butyldimethylsilyl
	TCA cycle	tricarboxylic acid cycle
	THR	threonine
	TOF	time of flight
	TRP	tryptophan
	TYR	tyrosine
U	ub	upper bound
	uL	microliter
	VAL	valine
W	WMW test	Wilcoxon-Mann-Whitney test
	wt	wild-type

Acknowledgments

The presented work was only possible with the help of several people, whom I would like to thank at this point.

Dr. Nicola Zamboni, who gave me the opportunity to work in his group and who generously shared with me his ideas and allowed me to develop my own.

Prof. Dr. Uwe Sauer, for trusting in me in the first place, for providing a birds-eye perspective on my work, and sharing his wisdom and experience in the field of Systems Biology and in science in general.

Prof. Dr. Manfred Claassen and **Prof. Dr. Vassily Hatzimanikatis**, the members of my PhD Committee, for sharing their knowledge and giving the chance to look at my project from different perspectives.

All present and past members of Zamboni and Sauer labs for being great colleagues and friends, and contributing greatly to the fun and time spent at work. **Marieke** and **Dimitris**, for going through the whole Ph.D. journey with me from the very beginning, for the never-stop-learning and improving our teaching skills together. **Petrica**, for providing her shoulder (and food) when I needed it. **Manu**, for introducing me to the world of lab and still believing in me. **Andi** and **Sébastien** for their words of support and brother attitude. **Leila**, for the funny and serious discussions about science and life. **Zrinka**, for contributing to this work and patiently teaching me high-throughput techniques. **Amelia, Philipp** and **Yanfeng** for making D58 the nicest lab on the floor. Tobi, Brendan, Sarah, Karthik, Mattia, Elad, Markus, Ruben, Michiel, Birgitta, Maren, Pau, Karin, Karl, Hannes, Luca, Viktor, Laura, Juliane, Ana Paula, Este, Stifi, Boris, Daniel, Szymon, for making the lab the way it is, and the time spent in lab so great.

My shared students **Eva-Maria Link** and **Martin Osswald**, and **Franziska Küng**, who were of great assistance in various projects and gave me the opportunity to teach and learn.

Prof. Dr. Konstantin Vorontsov, Prof. Dr. Alexander Dyakonov and Dr. Evgeniy Riabenko, who have been key persons for my scientific path.

The whole Russian crew in Zürich, especially **Valera, Valya** and **Dima** without whom I would not be there where I am, and Zürich would not feel so much like home.

My friends from home, especially **Dashik** and **Alena**, for being there for me and sharing the everyday life despite the distance.

Elina and **Vera**, for cheering for me and understanding my failures to visit them.

Danka, my partner in crime, for the constant moral and professional support, week by week, no matter what.

Michi, for making the journey out of my comfort zone so enjoyable, and for the eternal support in all areas of life.

My parents, my family, and friends who supported and believed in me.

Curriculum vitae

Personal details:

name: Maria Zimmermann-Kogadeeva
 address: Lerchenberg 1
 CH-8046 Zurich
 date of birth: 03/05/1989
 nationality: Russian

Education:

06/2012 – present Ph.D. at ETH Zurich with Dr. Nicola Zamboni and Prof. Dr. Uwe Sauer
 09/2006 – 09/2011 Diploma in mathematics at Lomonosov Moscow State University, Moscow, Russia
 09/2002 – 09/2006 Lyceum #1533 of Informational Technologies
 Specialization: programming

Publications:

Geiger, Roger, Jan C. Rieckmann, Tobias Wolf, Camilla Basso, Yuehan Feng, Tobias Fuhrer, **Maria Kogadeeva**, Paola Picotti, Felix Meissner, Matthias Mann, Nicola Zamboni, Federica Sallusto, Antonio Lanzavecchia. "L-Arginine modulates T cell metabolism and enhances survival and anti-tumor activity." *Cell* 167 (2016): 829-842.

Kogadeeva, Maria, and Nicola Zamboni. "SUMOFLUX: A generalized method for targeted ¹³C metabolic flux ratio analysis." *PLoS Comput Biol* 12, no. 9 (2016): e1005109.

Chevallereau, Anne, Bob G. Blasdel, Jeroen De Smet, Marc Monot, Michael Zimmermann, **Maria Kogadeeva**, Uwe Sauer, Peter Jorth, Marvin Whiteley, Laurent Debarbieux, Rob Lavigne. "Next-Generation “-omics” approaches reveal a massive alteration of host RNA metabolism during bacteriophage infection of *Pseudomonas aeruginosa*." *PLoS Genet* 12, no. 7 (2016): e1006134.

De Smet, Jeroen, Michael Zimmermann, **Maria Kogadeeva**, Pieter-Jan Ceyskens, Wesley Vermaelen, Bob Blasdel, Ho Bin Jang, Uwe Sauer, and Rob Lavigne. "High coverage metabolomics analysis reveals phage-specific alterations to *Pseudomonas aeruginosa* physiology during infection." *The ISME journal* (2016).

Schubert, Olga T., Christina Ludwig, **Maria Kogadeeva**, Michael Zimmermann, George Rosenberger, Martin Gengenbacher, Ludovic C. Gillet et al. "Absolute proteome composition and dynamics during dormancy and resuscitation of *Mycobacterium tuberculosis*." *Cell host & microbe* 18, no. 1 (2015): 96-108.

Oliveira, Ana Paula, Christina Ludwig, Paola Picotti, **Maria Kogadeeva**, Ruedi Aebersold, and Uwe Sauer. "Regulation of yeast central metabolism by enzyme phosphorylation." *Molecular Systems Biology* 8, no. 1 (2012): 623.

„Even the one who is crawling is already moving forwards“

Alexander G. Dyakonov