

RESEARCH ARTICLE

A Method to Constrain Genome-Scale Models with ^{13}C Labeling Data

Héctor García Martín^{1,2*}, Vinay Satish Kumar^{1,2}, Daniel Weaver^{1,2}, Amit Ghosh^{1,2}, Victor Chubukov^{1,2}, Aindrila Mukhopadhyay^{1,2}, Adam Arkin^{1,3}, Jay D. Keasling^{1,2,3,4}

1 Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, United States of America, **2** Joint BioEnergy Institute, Emeryville, United States of America, **3** Department of Bioengineering, University of California, Berkeley, Berkeley, United States of America, **4** Department of Chemical Engineering, University of California, Berkeley, Berkeley, United States of America

* hgmartin@lbl.gov



OPEN ACCESS

Citation: García Martín H, Kumar VS, Weaver D, Ghosh A, Chubukov V, Mukhopadhyay A, et al. (2015) A Method to Constrain Genome-Scale Models with ^{13}C Labeling Data. *PLoS Comput Biol* 11(9): e1004363. doi:10.1371/journal.pcbi.1004363

Editor: Costas D. Maranas, The Pennsylvania State University, UNITED STATES

Received: February 17, 2015

Accepted: May 29, 2015

Published: September 17, 2015

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All experimental data were obtained from Toya et al.: "13C-metabolic flux analysis for batch culture of Escherichia coli and its pykand pgigene knockout mutants based on mass isotopomer distribution of intracellular metabolites". *Biotechnology progress* 26: 975(992).

Funding: This work conducted by the Joint BioEnergy Institute was supported by the Office of Science, Office of Biological and Environmental Research, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. JDK has financial interests in Amyris and Lygos. The funders had no role in study design, data collection and

Abstract

Current limitations in quantitatively predicting biological behavior hinder our efforts to engineer biological systems to produce biofuels and other desired chemicals. Here, we present a new method for calculating metabolic fluxes, key targets in metabolic engineering, that incorporates data from ^{13}C labeling experiments and genome-scale models. The data from ^{13}C labeling experiments provide strong flux constraints that eliminate the need to assume an evolutionary optimization principle such as the growth rate optimization assumption used in Flux Balance Analysis (FBA). This effective constraining is achieved by making the simple but biologically relevant assumption that flux flows from core to peripheral metabolism and does not flow back. The new method is significantly more robust than FBA with respect to errors in genome-scale model reconstruction. Furthermore, it can provide a comprehensive picture of metabolite balancing and predictions for unmeasured extracellular fluxes as constrained by ^{13}C labeling data. A comparison shows that the results of this new method are similar to those found through ^{13}C Metabolic Flux Analysis (^{13}C MFA) for central carbon metabolism but, additionally, it provides flux estimates for peripheral metabolism. The extra validation gained by matching 48 relative labeling measurements is used to identify where and why several existing COntstraint Based Reconstruction and Analysis (COBRA) flux prediction algorithms fail. We demonstrate how to use this knowledge to refine these methods and improve their predictive capabilities. This method provides a reliable base upon which to improve the design of biological systems.

Author Summary

While metabolic fluxes constitute the most direct window into a cell's metabolism, their accurate measurement is non trivial. The gold standard for flux measurement involves providing a labeled feed where some of the carbon atoms have been substituted by isotopes with higher atomic mass (^{13}C instead of ^{12}C). The ensuing labeling found in intracellular metabolites is then used to computationally infer the metabolic fluxes that produced the

analysis, decision to publish, or preparation of the manuscript.

Competing Interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: JDK has financial interests in Amyris and Lygos. All other authors have no competing interests to declare.

observed pattern. However, this procedure is typically performed with small metabolic models encompassing only central carbon metabolism. The genomic revolution has afforded us easily available genomes and, with them, comprehensive genome-scale models of cellular metabolism. It would be desirable to use the ^{13}C labeling experimental data to constrain genome-scale models: these data constrain fluxes very effectively and provide in the labeling data fit an obvious proof that the underlying model correctly explains measured quantities. Here, we introduce a rigorous, self-consistent method that uses the full amount of information contained in ^{13}C labeling data to constrain fluxes for a genome-scale model where underlying assumptions are explicitly stated.

Introduction

Systems biology aims to understand and predict how a cell's behavior emerges from the interaction of its molecular parts [1–3]. Determination of metabolic fluxes (i.e., the number of metabolites traversing each biochemical reaction per unit time [4, 5]) is crucial to this effort because they map how carbon and electrons flow through metabolism to enable cell function [2, 5].

Metabolic fluxes typically cannot be measured directly, but must be inferred from experimental data through computational algorithms [4]. Among the most popular methods for studying metabolic fluxes are Metabolic Flux Analysis (MFA, [6, 7]), Flux Balance Analysis (FBA, [8]) and ^{13}C Metabolic Flux Analysis (^{13}C MFA, [4, 5, 9]). MFA calculates fluxes by using a stoichiometric model for the major intracellular reactions and assuming no metabolite accumulation [6]. The inputs are extracellular fluxes obtained through measurements of external concentrations of metabolites such as glucose or lactate as a function of time. If more flux measurements are available than degrees of freedom, the system is said to be overdetermined and a unique solution can be obtained. In the opposite case, the system is underdetermined and several flux profiles are compatible with the experimental data. MFA has been used to study fluxes in (e.g.) chinese hamster ovary [10], *S. cerevisiae* [11] and hybridoma cells [12].

Present-day FBA enhances MFA by expanding the network to include all reactions in metabolism, or at least as many as can be inferred from the genome through a metabolic reconstruction that yields a genome-scale stoichiometric model [13]. Since the degrees of freedom for such a model are usually over a hundred and measured fluxes are usually an order of magnitude less, the system is grossly underdetermined. Hence, fluxes are determined through linear programming (LP) by assuming that metabolism is tuned, due to evolutionary pressure, to maximize growth rate (typically; but see Schuetz *et al.* [14] for other suggested alternatives). The use of this objective function to interpret stoichiometric models is a key feature of FBA, even when stoichiometric models were not genome-scale. FBA forms the basis of a family of flux analysis methods named COntstraint-Based Reconstruction and Analysis (COBRA) methods [15], some of which can be used to produce flux predictions which are often used in bioengineering. These predictions can be full (when fluxes are determined without data from the actual experiment [16]) or partial (when some data from the experiment, like glucose consumption, is used for the prediction [17]), qualitative (e.g., prediction of whether an organism will grow or not under given conditions [18]) or quantitative (e.g., the value of the flux is predicted [19]). Full predictions are particularly useful for bioengineering purposes since they enable quick testing of the consequences of engineering approaches [20]. These methods have been used to facilitate the large-scale industrial production of 1,4-butanediol, a commodity chemical used to manufacture over 2.5 million tons annually of high-value polymers [21]. Recently, this rationally developed strain was used for a 5 million pound commercial production [22], and BASF has licensed this

strain for future production of renewable 1,4-butanediol [23]. FBA has applications beyond bio-engineering, including the prediction of ratios of microbial species in a simple microbial community [24] and providing valuable insights into tumor cell metabolism [25, 26].

^{13}C MFA improves on MFA by using data obtained from ^{13}C substrate labeling experiments together with a limited reaction stoichiometry and measured extracellular fluxes to measure intracellular fluxes. In these experiments, the organism under study is grown on ^{13}C labeled substrate and the labeling pattern (i.e., the fraction of molecules with 0,1,2, . . . ^{13}C atoms incorporated or Mass Distribution Vector, MDV [27]) is measured for a set of metabolites. Since the labeling pattern is highly dependent on the flux profile, it is possible to back-calculate the fluxes that best explain the measured labeling pattern if we know the fate of each carbon atom (carbon transitions, see [9]) for all reactions in the model. This involves a nonlinear fitting problem where the fluxes are the parameters. The usual approach is to consider only a small subset of all metabolism assumed to be comprehensive enough to explain the labeling of the measured metabolites, typically central carbon metabolism [28, 29]. Hence the model exhibits relatively few degrees of freedom that can be fully constrained by the labeling data. ^{13}C MFA has found applications in metabolic engineering [30], biotechnology [31] and biomedicine [32].

Each of these methods involves its own advantages and disadvantages. Unlike ^{13}C MFA, FBA uses the comprehensive description of metabolism contained in a genome-scale model and explicitly takes into account the system-wide balances of metabolites that can be crucial for host engineering [33]. Because of the exhaustive description of metabolism incorporated in genome-scale models, they often point towards completely unexpected regions of metabolism involved in the studied processes. For example, these models have been used to show that biosynthesis and degradation of heme compensates for the lack of a functional TCA cycles in cancer cells [26]. Furthermore, FBA can be used in combination with COBRA methods to make full predictions. ^{13}C MFA, on the other hand, is a descriptive method for determining the metabolic fluxes compatible with the accrued experimental data but does not postulate general principles that can be used to make predictions for experiments that have not been performed. However, ^{13}C MFA does not rely on maximum growth assumptions, the general applicability of which has been questioned [14, 34, 35] and shown to be inaccurate for engineered strains that are not under long-term evolutionary pressure [17]. Moreover, the comparison of measured and fit labeling patterns provides a degree of validation and falsifiability that FBA does not possess: an inadequate fit to the experimental data indicates that the underlying model assumptions are wrong. In contrast, FBA produces a solution for almost any input.

Several attempts to combine the complementary virtues of ^{13}C MFA and FBA have been reported. For example, they have often been combined to test new FBA-based methods, since ^{13}C MFA is considered to be the most authoritative determination of fluxes. Both Segrè *et al* [16] and Yizhak *et al* [19] used ^{13}C MFA to validate MOMA and IOMA, respectively. More recently, Schuetz *et al* [14], used ^{13}C MFA-derived fluxes to compare predictions from an MFA model containing ~ 100 reactions and using different objectives, and also to demonstrate the applicability of pareto optimality to predict fluxes [35]. Choi *et al* [36] combined both methods by using flux ratios obtained from ^{13}C MFA to constrain FBA for genome-scale models through the use of artificial metabolites. Although genome-scale models were not used, Suthers *et al* [27] presented ^{13}C MFA for a large scale model containing 350 reactions. The OPENflux open source software by Quek *et al* [37], allows for certain reactions in ^{13}C MFA to be used only for stoichiometric modeling purposes. More recently, Chen *et al* [38] for the first time modeled the same *E. coli* strain under the same conditions using FBA and ^{13}C MFA, using some of the information of the latter to constrain the former. In a similar vein, Kuepfer *et al* [39] constrained a *S. cerevisiae* genome-scale model with fluxes obtained from ^{13}C MFA and determined the flux distribution by using the minimization of the overall intracellular flux as the objective function.

However, to date there has been no attempt to use the data from ^{13}C labeling experiments to constrain fluxes for a genome-scale model without assuming that metabolism is evolutionarily tuned to optimize an objective function (such as the growth rate optimization typically used in FBA). The traditional underlying assumption in the field of ^{13}C MFA is that the degrees of freedom of the model must carefully match the amount of information obtained from the labeling patterns. Indeed, if the mathematical formulation of ^{13}C MFA were that of a linear programming problem (as for FBA) it would be pointless to try to constrain over a hundred degrees of freedom with the approximately 50 measurements of labeling used in this study, for example. However, ^{13}C MFA is a *nonlinear* fitting problem (with fluxes being the parameters) and these problems behave very differently for the underdetermined case, a special case of what is known as “sloppy” models in statistical mechanics [40–43]. These underdetermined nonlinear fits exhibit some degrees of freedom which are highly constrained (the parameters cannot be changed without noticeable measurable effects) and other degrees of freedom barely constrained at all (parameters can be changed freely, see section 3.3 in Brown *et al* [40] for a concrete example). Even if all degrees of freedom are not fully determined, the model can still be used to test hypotheses effectively [44]. These characteristics have been shown to be of general nature and apply to a variety of nonlinear problems appearing in systems biology [44], insect flight and variational quantum wave functions [45], interatomic potentials [46] and a model of the next-generation international linear collider [42]. In this paper, we present a systematic and rigorous framework to take full advantage of the nonlinear nature of the flux fitting problem and find all fluxes compatible with the ^{13}C labeling data for a genome-scale model. In order to constrain fluxes effectively, we make the biologically relevant assumption that metabolic flux flows from core metabolism (defined below) to peripheral metabolism and does not flow back.

The simultaneous use of ^{13}C labeling data and genome-scale models produces a new computational approach combining the advantages of both FBA and ^{13}C MFA: two-scale ^{13}C Metabolic Flux Analysis (2S- ^{13}C MFA, see Fig 1). 2S- ^{13}C MFA determines fluxes for a full genome-scale model, taking into account the system-wide balances of metabolites. However, 2S- ^{13}C MFA does not rely on maximum growth assumptions: instead it uses the data obtained from ^{13}C labeling experiments to constrain feasible fluxes. The use of this data is shown to constrain glycolytic and pentose phosphate pathway fluxes 8–50 fold more effectively than using only measured extracellular fluxes. We use this new method to compare different predictive methods and show where and why they fail. Based on that information, we develop a new predictive method that is able to produce a full quantitative prediction of 48 labeling measurements, going beyond the usual qualitative (e.g. grow/no grow) predictions.

Results and Discussion

Constraining genome-scale models with ^{13}C labeling data

In ^{13}C MFA, most reactions in the cell’s enormous metabolic network have a very limited contribution to the ^{13}C labeling of the observed metabolites (regardless of whether these are free metabolites [47] or proteogenic amino acids [4, 9]). Thus, since the complexity of the problem of flux measurement from labeling information scales nonlinearly with the number of reactions, the metabolic network in ^{13}C MFA consists of a minimal set of reactions (core set) that most influence labeling patterns (typically central carbon metabolism but may include other extra reactions, such as those describing protein turnover [48]), stemming from the literature or the researcher’s experience. This approach is able to convincingly explain labeling patterns for amino acids and intracellular metabolites for model organisms (e.g. *E. coli* [49, 50] and *S. cerevisiae* [51, 52]) under well-studied conditions (e.g. glucose feed). The good fits to experimental data support, to a good approximation, the underlying assumptions that carbon

Method overview

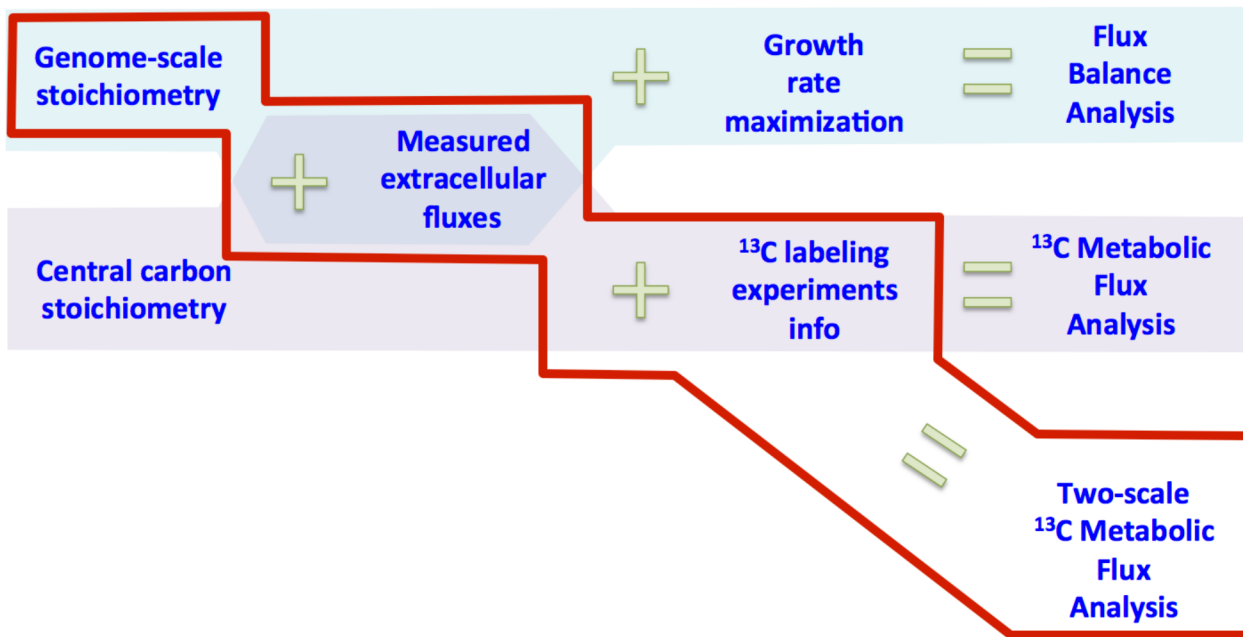


Fig 1. Method overview. FBA uses genome-scale stoichiometry and measured extracellular fluxes to constrain fluxes, which are finally determined by assuming growth rate maximization. ^{13}C MFA measures fluxes by using the highly informative ^{13}C labeling experimental data along with central carbon stoichiometry and measured extracellular fluxes. $2\text{S-}^{13}\text{C}$ MFA calculates fluxes by using the ^{13}C labeling experimental data and the measured extracellular fluxes to constrain fluxes for a genome-scale model.

doi:10.1371/journal.pcbi.1004363.g001

precursors flow from a central core metabolism to peripheral metabolism and do not flow back. However, it would be desirable to generate the minimal core network systematically through a computational method, so the technique can be generally applied to non-standard cases: e.g. non-model organisms, bioengineered strains, alternative non-standard carbon sources, human cells or microbial communities. Most crucially, such a method would explicitly test the assumption that reactions not included in the minimal network do not significantly influence labeling. As an added benefit, it would take advantage of the significant community efforts that have been put into developing and improving genome-scale metabolic models [53].

Another important benefit of integrating ^{13}C MFA with genome-scale metabolic models is to check for consistency of the inferred core fluxes with peripheral metabolism. The central metabolic network needs to produce not only carbon precursors for peripheral metabolism, but also ATP and reducing equivalents. Genome scale networks offer the possibility to track every reaction consuming and producing energy or reducing equivalents, and therefore can be used effectively to study the interplay between peripheral and core metabolism. On one hand, since core metabolism involves the reactions with largest fluxes in metabolism, once these are set by the ^{13}C labeling data, peripheral metabolism is expected to be highly constrained. On the other hand, changes in the usage of core intermediates in peripheral metabolism will also have significant effects on core fluxes since they need to provide these carbon intermediates, on top of ATP and reducing equivalents. While this has been already predicted in terms of small changes in biomass composition having significant effects on central carbon fluxes [54], the effect is an order of magnitude more important in bioengineered strains [55]. In such cells, peripheral metabolism can be vastly altered, and it is crucial to have a method that can take the diverse changes into account in a systematic manner.

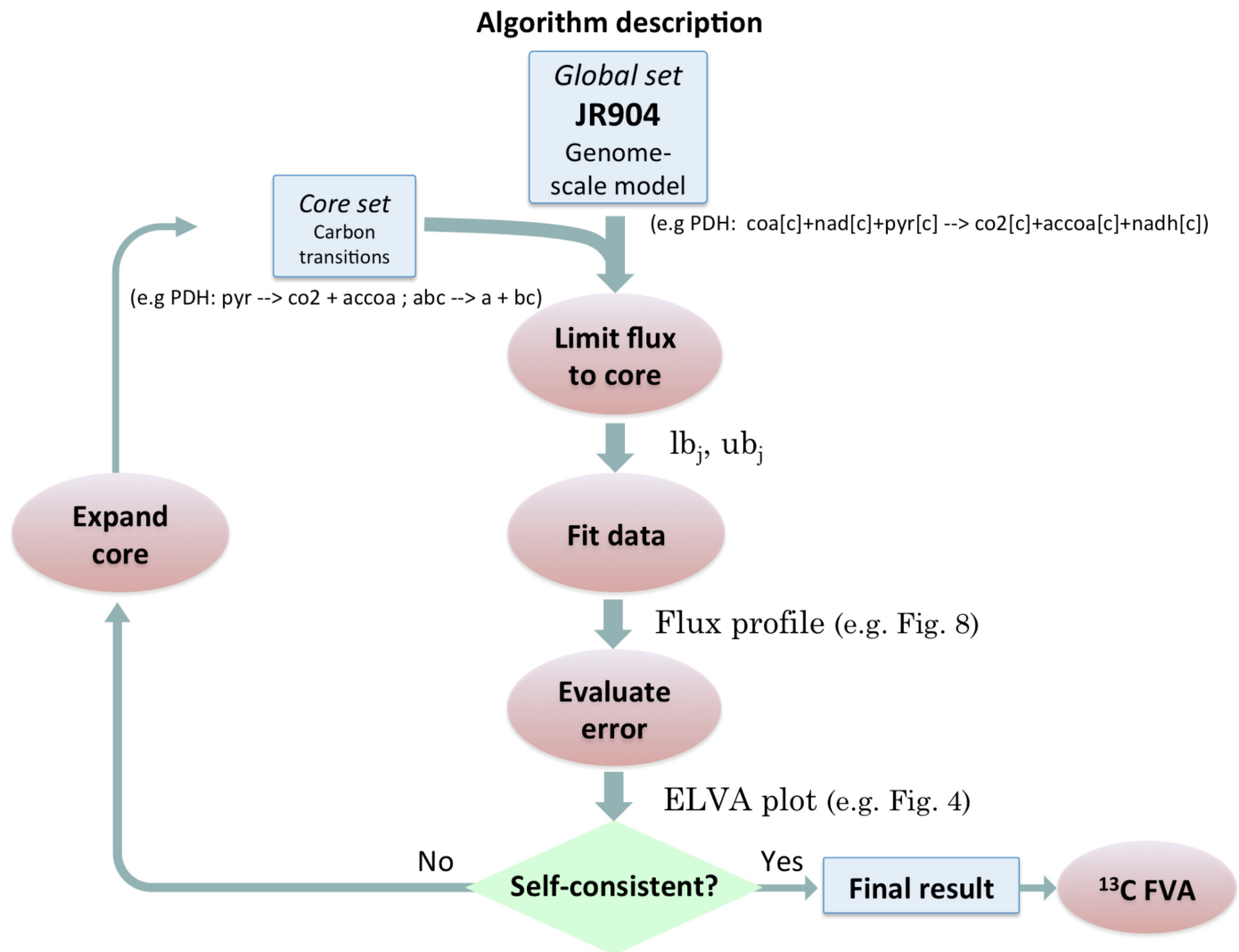


Fig 2. Algorithm description. Algorithm flow diagram for 2S- ^{13}C MFA showing a recursive procedure to achieve self-consistent results. The full model consists of a genome-scale model (iJR904 in this case) to which information on carbon transitions for the core sets of reactions is added (blue box on the left). The genome-scale model carries the measured extracellular fluxes information as upper and lower bounds (ub_j and lb_j). Carbon transitions (example line below the blue box) indicate the fate of each carbon atom in the reaction. The first step in the algorithm involves limiting the amount of flux that flows into the core set of metabolites and reactions, so as to enforce the two-scale approximation (i.e. that non-core contributions to labeling are negligible). The second step involves finding the set of fluxes that best fit the experimentally observed data, ignoring the non-core contributions. The final step tests that the error incurred by ignoring non-core reactions is negligible through External Labeling Variability Analysis (ELVA). If the ELVA does not indicate that the non-core contributions are negligible, the core set and the EMU model are expanded and the procedure repeated. When a self-consistent result is found, flux ranges compatible with the experimental data are obtained through ^{13}C Flux Variability Analysis.

doi:10.1371/journal.pcbi.1004363.g002

2S- ^{13}C MFA addresses both of these issues and provides a complete rigorous framework for estimating fluxes using ^{13}C labeling data in the context of a genome scale metabolic network. The algorithm is comprised of a set of optimization problems to be applied sequentially, as shown in the outline in Fig 2. Our approach divides the metabolites and reactions in a genome-scale model into two groups to be modeled at different scales of resolution (Fig 3A), in the spirit of the multi-scale approach in engineering and physics [56]. For “core” metabolites and reactions both stoichiometry and carbon labeling are tracked. For the remaining “non-core” metabolites and reactions, only stoichiometry is tracked and their contribution to the core set labeling is ignored. Crucially, the algorithm recursively adjusts the division into core and non-core reactions

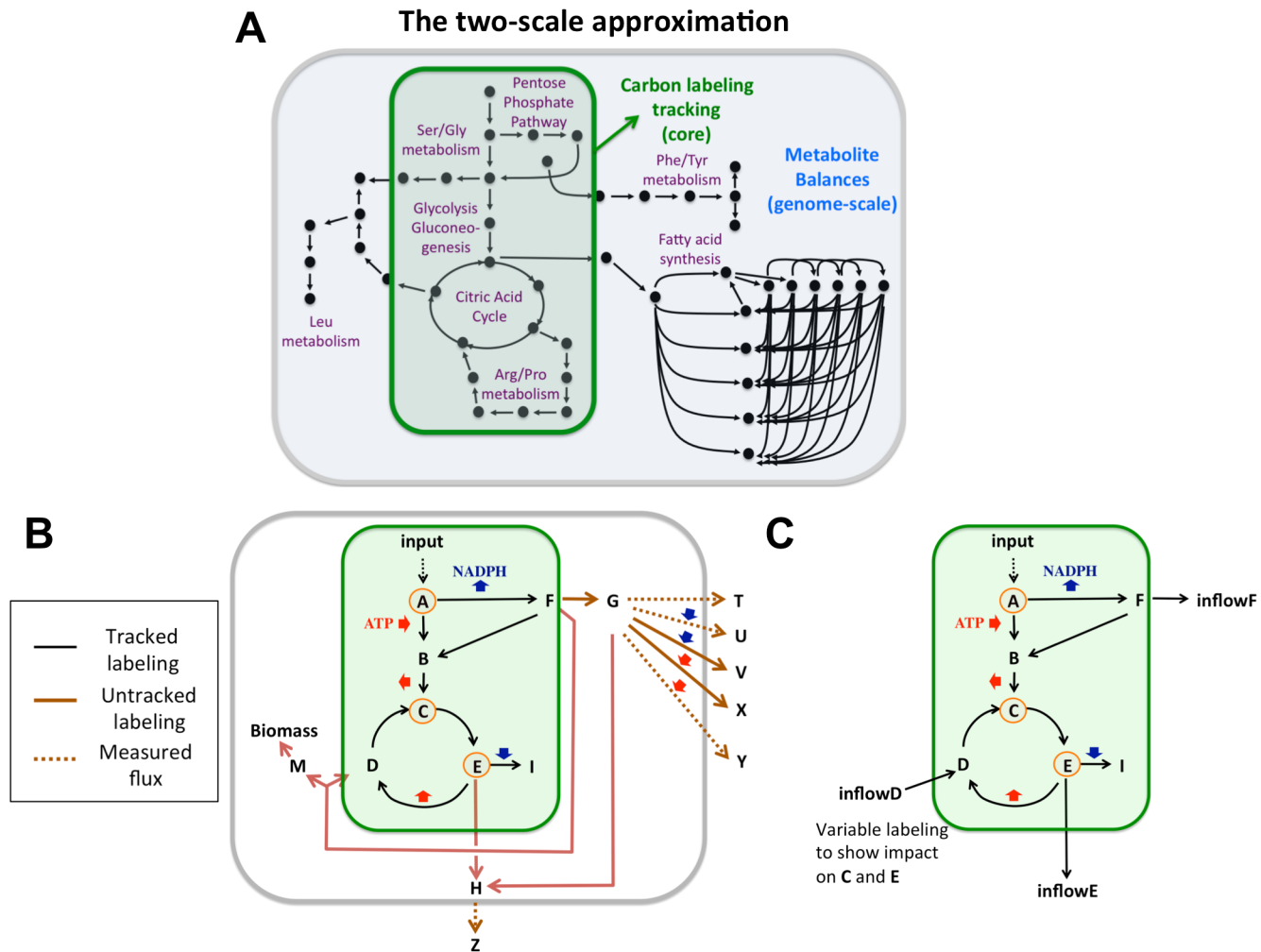


Fig 3. The two-scale approximation. **A)** $2\text{S-}^{13}\text{C}$ MFA models microbial metabolism at two different scales of resolution, hence minimizing the computational effort to explain the experimental data. While stoichiometric balances are taken into account for the full genome-scale model (iJR904 in this case), metabolite labeling originating from the ^{13}C feed in the labeling experiments is only tracked for the core set of reactions responsible for the main fraction of metabolite labeling (green box). The two-scale approximation assumes that non-core metabolites do not directly affect core metabolite labeling. The core set is expandable through the recursive procedure shown in Fig 2. **B)** Exemplary network of 20 reactions that illustrates the two-scale approximation and the approach. Measured data involves the MDV for metabolites A, C and E and extracellular fluxes for reactions producing metabolites T, U, Y and Z. The initial core set involves reactions and metabolites in the green box. The fit involves finding fluxes which best match the measured labeling and the values of the measured extracellular fluxes, where only the contribution of reactions inside the green box is taken into account to fit the labeling of metabolites A, C and E. However, the metabolite balance is global. In this way the fluxes are not overconstrained by e.g. NADPH balance: any excess NADPH can be balanced by the non-core fluxes that consume NADPH. **C)** Right lower panel illustrates External Labeling Variability Analysis (ELVA) for the exemplary network. ELVA gauges the effect of non-core reactions by considering only the core network and simulating the impact of non-core metabolite labeling through inflow metabolites (inflowD, inflowE, inflowF). The ELVA optimization problem (Eqs 9–15) finds the maximum impact that the unknown inflow metabolite labeling can have on the measured labeling pattern.

doi:10.1371/journal.pcbi.1004363.g003

according to the error in the labeling fitting. This procedure is illustrated with a small illustrative network (see Fig 3B and 3C) throughout the following subsections. Notice that, because of this core readjustment and the inclusion of other reactions needed to fully explain labeling [48], the core set needs not be the same as is usually referred to as central carbon metabolism.

Limiting flux to core reactions. $2\text{S-}^{13}\text{C}$ MFA starts with an initial guess for the core set and then limits the flux into core metabolism, so as to be consistent with the initial assumption that non-core metabolism does not directly contribute to core ^{13}C labeling. We will refer to

this assumption as the two-scale approximation or assumption. In the context of a genome-scale model (iJR904 [57] in this particular case, see [Materials and Methods](#)), this assumption amounts to limiting the upper bound of all reactions with products in core metabolism to zero, or the lowest value consistent with the observed growth rate (see [Materials and Methods](#) for the technical details).

Constraints introduced in this step replicate the same conditions as for ^{13}C MFA: by limiting the flux into core reactions, the degrees of freedom for the system of core reactions are similar to the case of standard ^{13}C MFA. This fact is reflected in very similar solutions for the core, as explained in the “Comparison with ^{13}C MFA” section. Hence, this is not a radically new assumption, but rather the same underlying assumption used in ^{13}C MFA. 2S- ^{13}C MFA simply explores its consequences in the context of genome-scale models. The degrees of freedom for the core system only differ from ^{13}C MFA because of the availability of more reactions to flow out of the core (since these are not set to zero). Rather than having the outgoing fluxes being prescribed by the modeler, 2S- ^{13}C MFA allows the system to channel this outgoing flux as it best fits the available data. This can lead to unforeseen results, as in the glycolate case shown in the “Comparison with ^{13}C MFA” section. Furthermore, these constraints automatically eliminate some of the biologically unrealistic solutions that have been shown to appear in fully unconstrained large-scale ^{13}C MFA studies [27].

In the illustrative network ([Fig 3B](#)), the core set of reactions is initially chosen as the reactions in the green box, which are expected to appropriately explain the labeling of the measured metabolites (A, C and E, in orange circles). The first step (“Limit flux to core”) involves setting to zero the flux of the reaction flowing from metabolite F to D. However, this reaction produces a metabolite M that is needed for cell growth, so it cannot be fully set to zero. A minimal level of flux is allowed (e.g. 5% of the glucose uptake rate), as shown in the “Limiting flux to core” section in materials and methods.

Fitting experimental labeling data. The second step fits the experimentally determined labeling data through the Elementary Metabolite Unit (EMU) method [28, 58]. While in this study we use labeling of central metabolic intermediates, the framework could be equally well applied to free amino acids or amino acids from hydrolyzed protein. The constraints include the experimentally determined growth, uptake, and secretion rates. The data fit is set as a Non-Linear Programming (NLP) optimization problem that minimizes the difference between computational and experimental labeling values and uses a combination of constraints from FBA and ^{13}C MFA, as shown in Eqs 1–7 (see [S1 Text](#)). Results can be seen in [S1–S3](#) and [S4–S12](#) Figs. Unlike FBA, there is no growth (or other flux) maximization involved: the growth rate is constrained to its measured value. Unlike ^{13}C MFA, the full genome-scale model reaction fluxes are used in the fit although, as explained below and observed in other nonlinear fits [27, 41, 42], not all fluxes are fully determined.

For the illustrative model, this step involves finding the fluxes that best fit the measured labeling of metabolites A, C and E where the labeling is calculated from only the contributions of the reactions in the green box (core) but metabolite balances are imposed for all reactions. In this way, instead of NADPH overconstraining the core fluxes by forcing the production inside the core (reaction A to F) to equal the consumption (reaction E to I), these fluxes are free to fit the experimental data while excess NADPH can be compensated by the reactions outside the core.

Self-consistency test through External Labeling Variability Analysis (ELVA). The key next step of 2S- ^{13}C MFA then determines if the fluxes obtained from the fit are consistent with the two-scale approximation. This extra procurement sets the method apart from ^{13}C MFA, which does not check the effect of ignored reactions, even for large-scale models [27] or by marking some reactions to be excluded from the isotopomer balance [37]. The impact of non-core reactions is calculated through External Labeling Variability Analysis (ELVA, see Eqs 9–

15). ELVA considers only core metabolism with the impact of non-core metabolism being represented through inflow metabolites, dummy metabolites with unfixed labeling since their labeling is, by definition, unknown (see Fig 3C for an example). Essentially, using the previously obtained genome-scale flux solution as a constraint, we allow the labeling for the inflow metabolites to vary and, for each metabolite with measured labeling, use each element of the mass distribution vector (MDV) as the objective function to be maximized or minimized. By this method we obtain a confidence interval that represents the maximum possible difference in labeling that could be attributed to non-core reactions for the current solution. The reactions that contribute an unacceptable amount of uncertainty are then added to the core set and the procedure can be repeated as necessary, until a core set of reactions is found which fully justifies the two-scale approximation.

In the case of the illustrative model, the use of the ELVA would gauge the impact of non-core reactions as described in Fig 3C. The ELVA would point out that the reaction transforming metabolite F to D and M needs to be included in the core: this reaction strongly influences the labeling of metabolite D, which in turn influences the labeling of metabolite C, which is being measured. This effect would be shown in large computational error bars for metabolite C through the influence of inflow metabolite D. After the reaction taking metabolite F to D and M is added to the core, a new fit would be in order (as shown in Fig 2), and the subsequent ELVA would show zero computational error, since none of the remaining reactions flow back into the core and can impact the labeling of the measured metabolites A, C and E.

To further illustrate this procedure with experimental data, we fit the labeling data from Toya *et al* [47] for wild-type *E. coli* during mid-exponential phase growth on glucose. Using an initial set of 94 central metabolic reactions (see S3 Text), we calculated the flux profiles and ELVA confidence intervals as described above. As seen in Fig 4 (left panel), several points had unacceptably wide confidence intervals (vertical error bars), in many cases over an order of magnitude greater than the experimental uncertainty (horizontal error bars), illustrating that reactions not included in the core could significantly influence the measured labeling. All of the most pronounced effects corresponded to the MDV of the TCA cycle intermediate malate (green dots on Fig 4). These effects were found (by inspection) to be due to reactions involved in glutamate, arginine, histidine metabolism and nucleotide biosynthesis, many of which use aspartate as a nitrogen donor, releasing fumarate. Since fumarate is converted to malate in the TCA cycle, this aspartate \rightarrow fumarate flux, often ignored in traditional ^{13}C MFA approaches [59–62], has a potentially important impact on the labeling of malate and must be considered as part of the core set, as some other studies have done [63]. Inclusion of these pathways in the core set dramatically narrowed the confidence intervals of the ELVA analysis (right panel of Fig 4), of the same order of magnitude now as the experimental error. This implies that the revised core (with a total of 126 reactions) satisfies the two-scale approximation, meaning that peripheral reactions not included in the core set cannot significantly influence observed labeling patterns. Hence, this method is self consistent and, furthermore, points out which reactions need to be added to the core set to make the approximation valid.

Confidence intervals provide all fluxes compatible with experimental data. ^{13}C labeling information is generally insufficient to completely constrain all fluxes in the network, meaning that a space of distinct flux solutions is consistent with the observed labeling. As such, the goal of a good flux estimation algorithm should be to provide the range of flux values that are consistent with the observed data, rather than a single “best” solution. This is particularly important in the case when flux estimation is expanded to the entire genome-scale model, with the corresponding dramatic increase in the degrees of freedom. $2\text{S-}^{13}\text{C}$ MFA addresses this directly through ^{13}C Flux Variability Analysis (^{13}C FVA): as shown in Eqs 16–23, each flux in the genome-scale model is maximized and minimized to find the range of fluxes producing

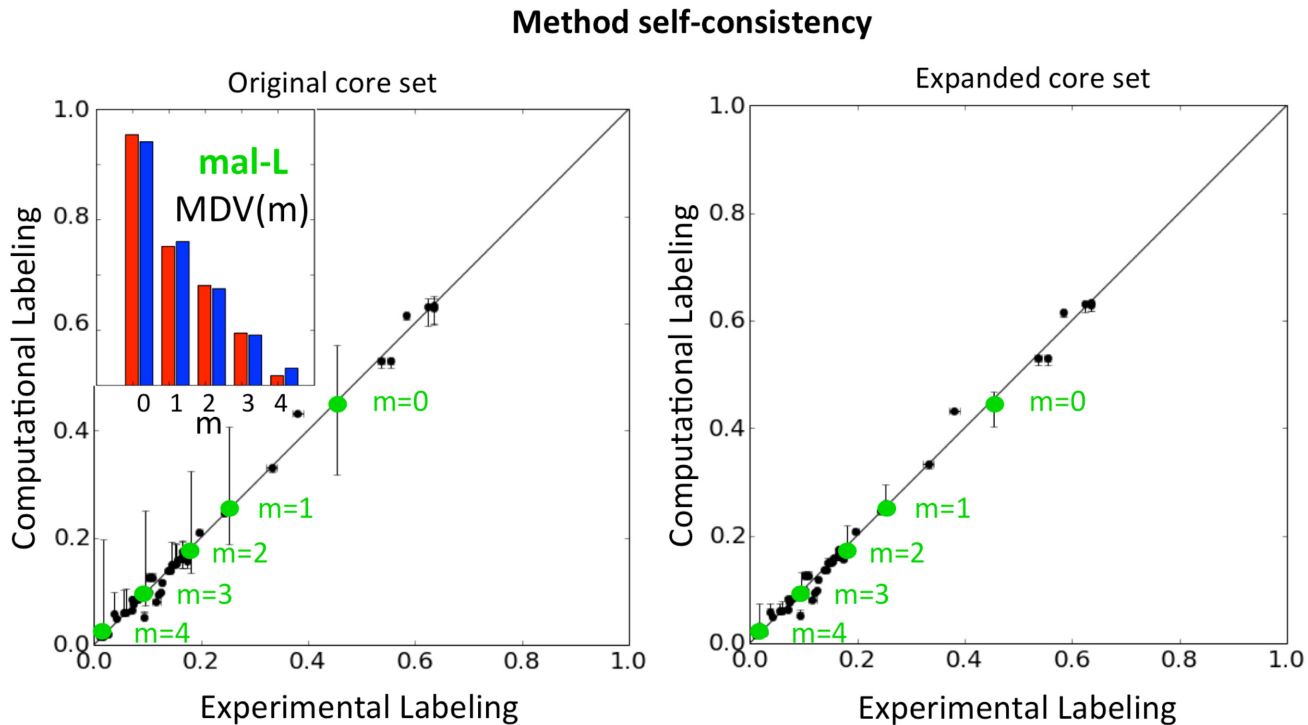


Fig 4. Method self-consistency. External Labeling Variability Analysis (ELVA, see [methods](#)) shows how the impact of ignored reactions diminishes by expanding the core set of reactions. Each point corresponds to an m value of the Mass Distribution Vector (MDV) for each of the metabolites considered. The inset provides the same information for malate in a more intuitive form (red for experimental data, blue for computational fits), see [S1–S3 Figs](#). Horizontal error bars indicate experimental CE-TOFMS error obtained from the instrument. Vertical error bars indicate computational errors obtained from the ELVA. These computational error bars indicate the maximum effect that non-core reactions (whose contribution to the carbon labeling is being ignored) could possibly have. The initial core set (left) shows a large computational error for malate (mal-L, green dots). By expanding the core set, the computational errors collapse to levels comparable with the experimental error as can be seen in the right panel. Hence, the method is self-consistent by ensuring that the final result meets the approximation used to calculate it.

doi:10.1371/journal.pcbi.1004363.g004

labeling patterns within the experimental error of the measured values (and in the context of the two-scale approximation).

The capability of the ^{13}C labeling data to constrain fluxes is very significant, as can be seen in [Fig 5](#): the ranges of fluxes compatible with the ^{13}C labeling data and the measured extracellular flux data is much smaller than those compatible with only the measured extracellular flux data. In fact, the average flux confidence interval for $2\text{S-}^{13}\text{C}$ MFA relative to the same interval for fluxes constrained only by extracellular fluxes is $\sim 2\%$ for the pentose phosphate pathway and $\sim 12\%$ for glycolysis. Hence, $2\text{S-}^{13}\text{C}$ MFA constrains fluxes 8–50 fold more effectively for these cases. Furthermore, we can see that the “limit flux to core” step does not introduce very strong constraints on the allowable fluxes.

Not all fluxes in the genome-scale model are effectively constrained by the ^{13}C labeling experiment data. This phenomenon is a general characteristic of nonlinear fits of underdetermined (“sloppy”) systems, and is expected. We see that, for example, flux values for most reactions in the core set are effectively constrained (i.e., the confidence intervals are narrow, as for GAPD in lower left panel of [Fig 6](#)). These core constraints are propagated to the rest of the fluxes through stoichiometry, and some non-core fluxes are also well determined (e.g., C160SN in upper left panel in [Fig 6](#)) while others are not (e.g., THD2 in upper left panel of [Fig 6](#)). As is often the case in fitting problems, we expect the degree to which this method can constrain fluxes to depend heavily on the available data: labeling distribution for more intracellular

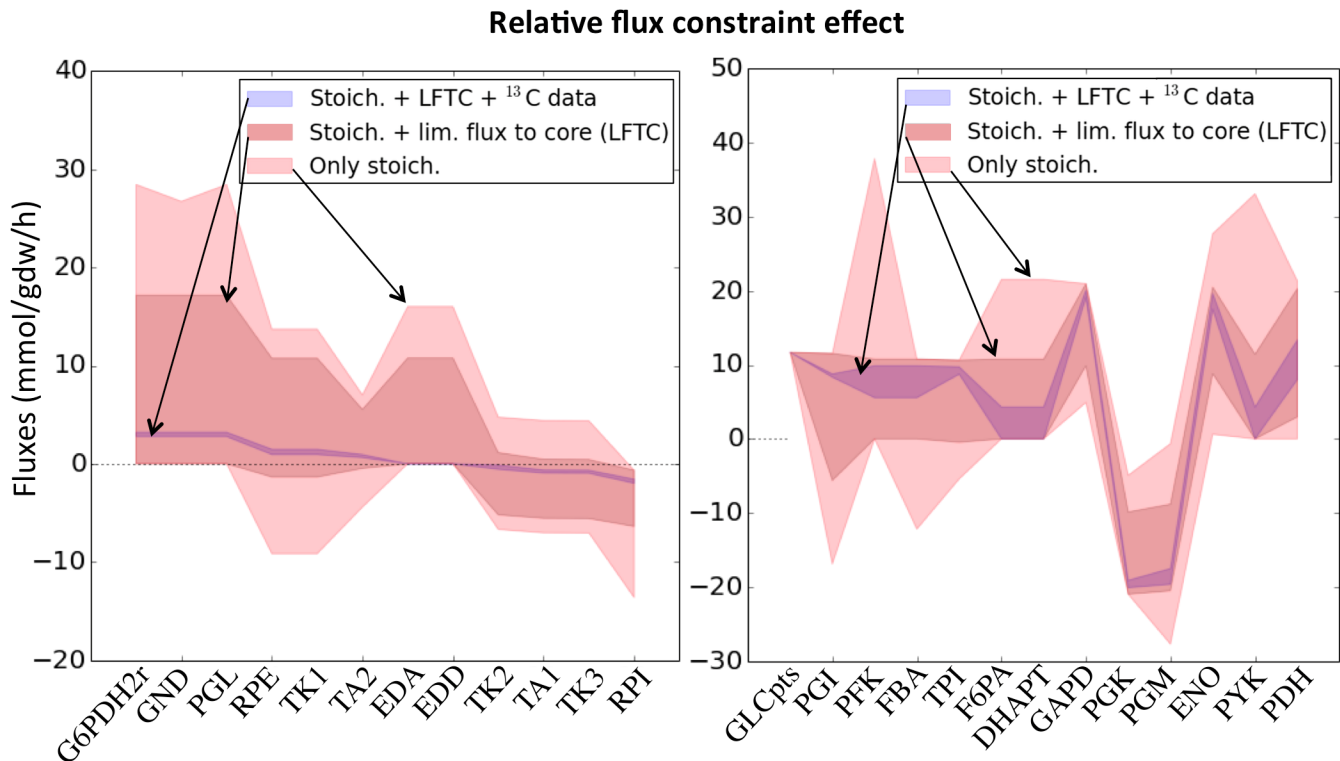


Fig 5. Relative effect of constraints. Confidence intervals for pentose phosphate pathway (left panel) and glycolysis (right panel) fluxes calculated using FBA constrained by measured extracellular fluxes (through Flux Variability Analysis, FVA [64], in red), for FBA with constraints derived from the two-scale approximation through the “Limit flux to core” step in Fig 3 (FVA, in grey), and for $2\text{S-}^{13}\text{C}$ MFA derived through Eqs 16–23 are shown (blue). Constraints induced by the two-scale approximation are not strong, hence justifying the use of this approximation. However, constraints induced by the ^{13}C labeling data are dominant. A similar pattern can be observed in S13 Fig.

doi:10.1371/journal.pcbi.1004363.g005

metabolites or measurements for more extracellular fluxes will increase the flux resolution (i.e., decrease the confidence intervals).

For the exemplary model, confidence intervals would be obtained through the optimization problem described in Eqs 16–23. If metabolite labeling data for new metabolites (e.g. D) became available, that information would decrease the confidence interval of the fluxes (and the core would have to be redefined).

Robustness with respect to data input

$2\text{S-}^{13}\text{C}$ MFA is supported by two independent data sets: ^{13}C labeling experimental data and genome-scale stoichiometry. We will now show that the method is robust to errors in the ^{13}C labeling experimental data and more robust than FBA to stoichiometric errors.

Fluxes calculated through $2\text{S-}^{13}\text{C}$ MFA are robust with respect to the experimental accuracy of the ^{13}C labeling data. We show this by generating new sets of ^{13}C data where the new labeling is randomly chosen within the experimental error. The calculated profiles do not change for this new labeling significantly from the initial profiles, as can be observed in Fig 7.

Cofactor ambiguities in the reconstruction of genome-scale models introduce errors in determining fluxes which are much smaller than for FBA. Reconstruction errors are still possible in genome-scale models, in spite of the care used to develop them and the continuous improvement in the reconstruction for each new release [13]. We studied the resulting flux profiles after a reconstruction error was simulated (by changing NADPH to NADH

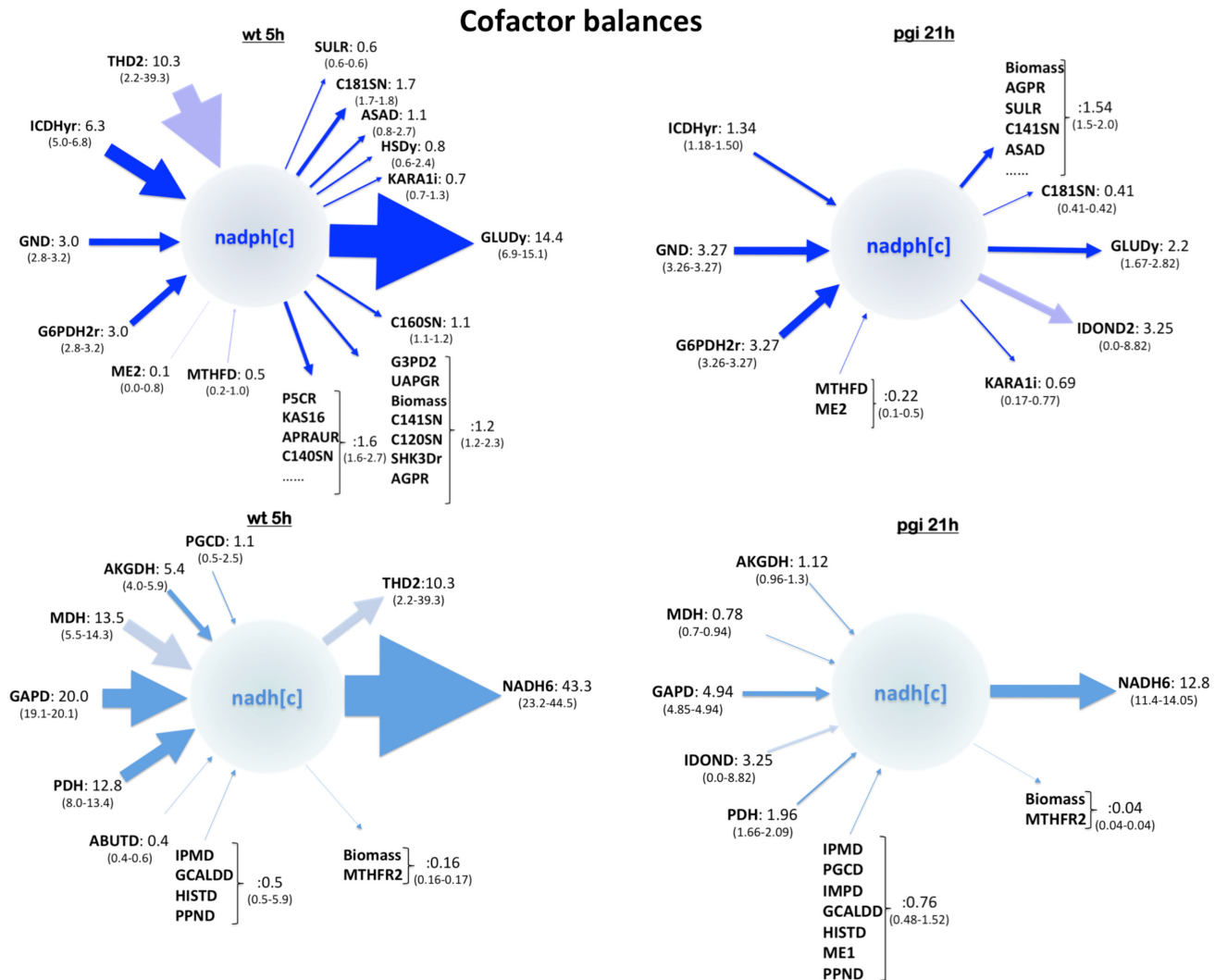


Fig 6. Cofactor balances. Cofactor balances show how NADPH and NADH production and consumption change after *pgi* is knocked out. Arrows pointing inwards on the left indicate fluxes that produce the indicated metabolite and fluxes pointing outwards on the right indicate fluxes that consume it (in units of mMol/gdw/hr). Reaction names are per iJR904 model. Upper panels show NADPH balances for wild type (left) and *pgi* KO (right) at 5 and 21 hr (equivalent growth points due to a lower growth rate in the *pgi* KO). Lower panels show NADH balances for wild type (left) and *pgi* KO (right) at the same time points. Note that, unlike FBA, 2S-¹³C MFA can provide confidence intervals bounded by the data from ¹³C labeling experiments. These are shown below the reaction name. For some cases (e.g. GND) the experimental data can very effectively constrain the flux value, even if the reaction is not in the core set over which labeling is being tracked (e.g. C181SN). For some others (e.g. THD2), the data can only constrain the flux value in a very limited fashion. Knocking out *pgi* radically changes NADPH and NADH supply and consumption patterns.

doi:10.1371/journal.pcbi.1004363.g006

dependence for a high flux reaction, G6PDH2r) and found that the change in fluxes was much less severe than for FBA (Fig 8). We found similar results for other reactions (S16 Fig). We attribute this difference to FBA relying heavily on stoichiometric information whereas 2S-¹³C MFA can additionally count on ¹³C labeling data to constrain fluxes. Furthermore, changes in the biomass composition reaction do not affect calculated fluxes significantly (S17 Fig).

Cofactor balances

2S-¹³C MFA can produce cofactor balance information which cannot be produced using either FBA (does not provide confidence intervals bounded by ¹³C data) or ¹³C MFA (does not

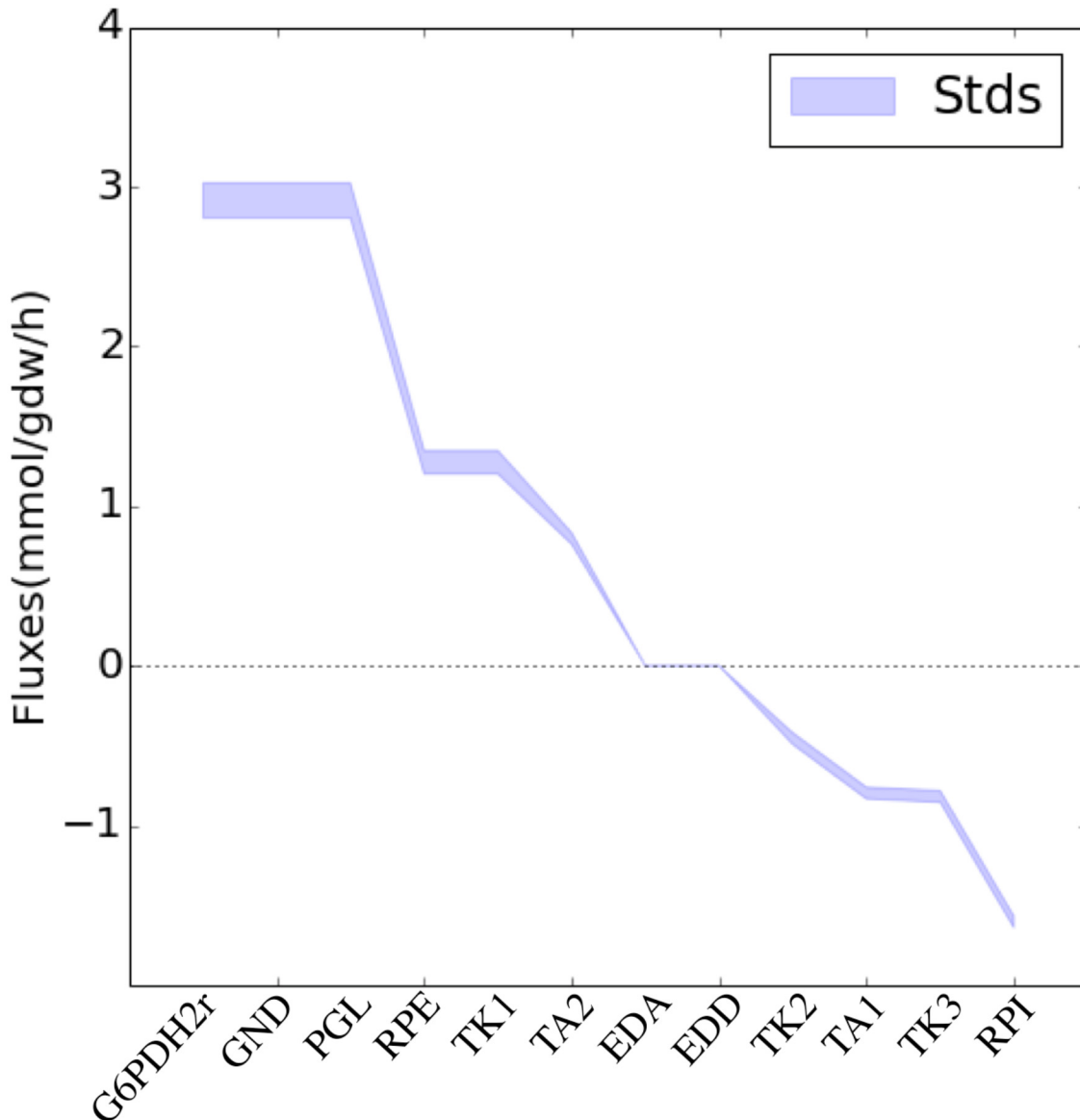


Fig 7. Robustness with respect to measurement error in labeling profile. 30 different new labeling data sets were generated by randomly choosing new labeling values within the experimental error (see equation 13 in [S2 Text](#)). Fluxes were calculated through $2\text{S-}^{13}\text{C}$ MFA for these new data sets and the standard deviation is shown for the PPP. Hence, the method is robust with respect to experiment accuracy in ^{13}C labeling.

doi:10.1371/journal.pcbi.1004363.g007

consider all possible cofactor-producing reactions in the cell metabolism). Discerning cofactor production and consumption can be crucial to understand cell behavior. Not only does cofactor balancing have an essential bearing on tumor cell death [65], but it can also have a significant impact on final production for bioengineered cells [33, 55, 66, 67]. Core metabolism and biosynthetic pathways are linked through cofactor balancing, so an expansion of ^{13}C MFA to include genome-scale stoichiometry such as that provided by $2\text{S-}^{13}\text{C}$ MFA is of great interest for bioengineering. While some previous ^{13}C MFA efforts [27, 28, 37] did take cofactor balances into account, this has not been done for all metabolites in a genome-scale network.

This capability is illustrated (Fig 6) with the main electron carriers in metabolism, NADPH and NADH, responsible for redox balances and, ultimately, for respiration. A majority of

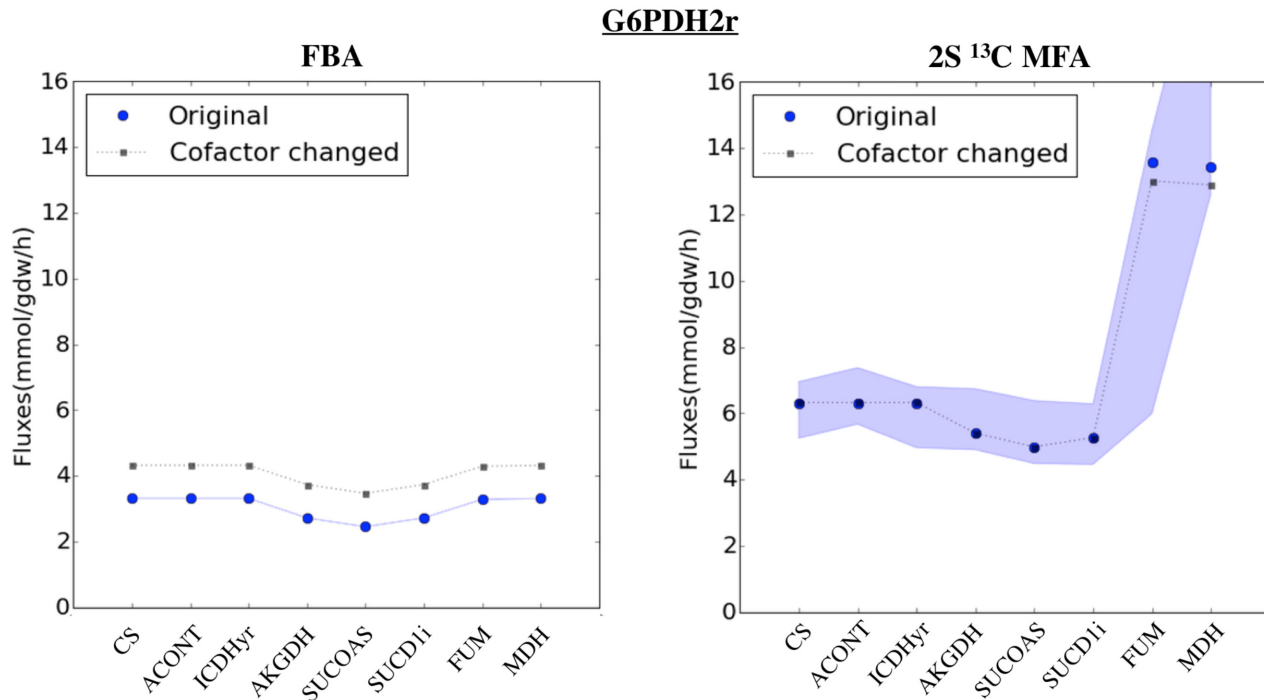


Fig 8. Robustness with respect to genome reconstruction errors. A reconstruction error was simulated by changing the NADPH dependence to NADH dependence for the glucose-6-phosphate dehydrogenase (with a large flux value of 2.9 mMol/gdw/hr in the initial 2S- ^{13}C MFA calculation). We calculated fluxes again through 2S- ^{13}C MFA and FBA (constrained by extracellular flux measurements), and the new fluxes are plotted for the TCA cycle. As can be observed, the change is much larger for FBA than 2S- ^{13}C MFA, showing that it is less robust to reconstruction errors (note that squares and circles are almost on top of each other for the 2S- ^{13}C MFA case). The transparency in the original flux profile for 2S- ^{13}C MFA indicates the confidence intervals. For the 2S- ^{13}C MFA case, the NADPH production shift is compensated entirely by the THD2 transhydrogenase. Since the flux value for SUCD1i is negative, the absolute value has been plotted. Similar figures for the glutamate dehydrogenase (GLUDy) and isocitrate dehydrogenase (ICDHyr) reactions are available as S16 Fig.

doi:10.1371/journal.pcbi.1004363.g008

NADPH production (47–55%, obtained by normalizing data from Fig 6) in wild type *E. coli* is produced by the pentose phosphate pathway (reactions G6PDH2r and GND: 25–27%) and the TCA cycle (reaction ICDHyr: 22–29%). Notice that, unlike FBA, 2S- ^{13}C MFA produces confidence intervals bounded by ^{13}C data: for example, if the flux through reaction GND were to be more than 3.2 mMol/gdw/hr (30% of total NADPH creation according to upper left part of Fig 6), the computationally derived labeling for (e.g.) $m = 0,4$ for malate (mal-L), for $m = 1$ for phosphoenolpyruvate (pep) and for $m = 1,2$ for ribose 5-phosphate (r5p) would differ from the measured value by more than the CEMS measuring error (see Eq 23 and S1–S3 Figs). The transhydrogenase reaction THD2, transferring electrons from NADH to NADP, also contributes significantly ($\sim 44\%$) but, unfortunately, the flux through this reaction is only very loosely constrained by the current data set. The main NADPH sink involves glutamate production (reaction GLUDy: 29–65%), with fatty acid biosynthesis (C181SN) in a distant second place (7%) and a collection of other reactions at equally low levels ($< 12\%$).

NADPH consumption and production are radically altered when the gene encoding the initial enzyme in glycolysis (PGI) is knocked out (Fig 6, right upper panel). This genetic manipulation forces the rerouting of all glycolytic flux into the PPP and a lower growth rate (0.17–0.23 vs 0.83–0.9 h $^{-1}$). 2S- ^{13}C MFA allows us to study the impact of this flux rerouting on the rest of the cell metabolism. For example, the overall NADPH demand falls three fold (vs a \sim four fold decrease in growth rate). Furthermore, the demand is not equally met by the pentose

phosphate pathway and the TCA cycle anymore: in the Δpgi , the PPP enzymes G6PDH2r and GND produce most of the NADPH (81%) with a much less important role played by the TCA cycle (13–19%). Interestingly, the absolute flux through the PPP for the *pgi* KO and the wild type is very similar. NADH is generated mostly by the glyceraldehyde-3-phosphate (36–37%, GAPD), malate (10–27%, MDH), pyruvate (15–25%, PDH) and 2-Oxoglutarate dehydrogenases (7–11%, AKGDH), which provide 79–98% of all NADH. Consumption is dominated by the ubiquinone-8 NADH dehydrogenase (NADH6, 43–83%). After *pgi* is knocked out, NADH demand decreased ~ 4 fold, with relative production by MDH undergoing the biggest drop (from 10–27% to 5–7%).

These results are consistent with previous ^{13}C MFA calculations for NADPH production [68], which validates this approach. Furthermore, they also provide information on secondary metabolism and NADH balances in a systematic manner. A similar analysis can be done for acetyl-CoA, a key metabolite and common bottleneck in metabolic engineering [69, 70] or any other relevant metabolite such as ATP, AMP, Coenzyme A or FADH, which would provide detailed information on the organism's underlying physiology.

Extracellular metabolite prediction

The 2S- ^{13}C MFA results pinpoint which (non measured) metabolites are expected to be detected in the extracellular medium based on the values of exchange fluxes as constrained by the ^{13}C labeling experiments (S13 Fig). Knowledge and quantification of the full range of excreted metabolites (also known as exometabolome or metabolic footprint [71]) is desirable for a full understanding of the biochemical impact of the cell on its environment. For metabolic engineering purposes, this knowledge provides important clues as to how to close the carbon balance and whether toxic compounds are being produced in the fermentation. For human metabolism, better prediction of extracellular fluxes can yield improved metabolic predictions when integrated with physiologically-based pharmacokinetic models [72].

The exchange fluxes predictions typically have large confidence intervals, but these intervals are much smaller than for FBA (see S13 Fig). Metabolites expected to be detected in the medium are those whose exchange fluxes have net positive maximum and minimum values for a period of time sufficiently long so as to reach detection limits. For the *E. coli* strains considered here, urea, glycolate (glyc, S14 Fig), fumarate (fum, S15 Fig) and acetaldehyde (acald) are the non-typical metabolites expected to be present in the medium (acetate is already measured). This prediction of atypical metabolites is of particular interest in light of the recent discovery of extended overflow metabolism [73]. A full prioritized list can be obtained from the exchange flux information and used to direct mass spectrometry, NMR or vibrational spectroscopy efforts to find the missing metabolites until carbon balance is met. For future improvement of intracellular metabolic flux predictions, constraints introduced by extracellular metabolite measurements are very effective and usually easy to measure, but it is necessary to know which metabolites to look for and 2S- ^{13}C MFA provides precisely that type of insights.

Comparison with ^{13}C MFA

2S- ^{13}C MFA produces nearly the same results as ^{13}C MFA for central carbon metabolism (see e.g. S4 Fig and S18 Fig). This similarity is not surprising since 2S- ^{13}C MFA is designed to mimic ^{13}C MFA for this part of metabolism (see “Limiting flux to core reactions” section). The only difference for the current data set can be found in the TCA cycle flux, as described below. These differences arise because genome-scale models account for fluxes to biomass in a more detailed and realistic manner and because they do not rule out unexpected metabolic routes compatible with the available data.

Flux through the TCA cycle is lower in the ^{13}C MFA solution (S4 Fig vs S18 Fig) because of an inaccurate account of fluxes to biomass in the ^{13}C MFA model. Specifically, the large fluxes draining acetyl-CoA into biomass and to the exterior of the cell as acetate (each a third of the pyruvate dehydrogenase flux, PDH) imposed in the original publication [47] have been overestimated in this particular case. While the typical biomass function used in ^{13}C MFA involves a specific stoichiometry of central carbon intermediates converted directly to biomass, this is only an approximation since some of the metabolites required for biomass growth are not represented in the minimal network model and need to be substituted by their requirements in terms of intermediates present in the minimal network (acetyl-CoA, in this case). These effects require significant effort to be accurately incorporated into a small-scale model, but they are elegantly handled by the genome-scale model. In this case a large flux of acetyl-CoA to biomass is assumed in the ^{13}C MFA solution, while the 2S- ^{13}C MFA solution determines the requirements on core metabolism by setting the biomass flux to the measured growth rate, and automatically determining the fluxes needed to provide the metabolites present in the biomass equation through the comprehensive stoichiometry network reflected in Eq 2. Hence, genome-scale models provide an automatic and detailed accounting of biomass requirements.

The difference in TCA cycle flux extends to the glyoxylate shunt in some cases in which the isocitrate lyase (ICL) is active for the 2S- ^{13}C MFA (S5 and S6 Figs) solution, but not in the ^{13}C MFA solution due to a limited ^{13}C MFA model. In the 2S- ^{13}C MFA solution, the objective function (Eq 1) becomes slightly lower by diverting flux into the ICL and then shuttling it out of the core set from glyoxylate (glx) into glycolate. This route is not included in the ^{13}C MFA model and activating the ICL for this model would produce a large amount of flux through the malate synthase (MLS). This MLS flux would significantly deteriorate the fit to malate labeling, so the glyoxylate shunt remains inactive for ^{13}C MFA case. The shuttling of glyoxylate into glycolate is unexpected and, in fact, could be the result of a numerical artifact since the only labeling data available in the full TCA cycle is that of malate (mal-L) and small errors in the labeling measurements (or their confidence intervals) for this metabolite can lead to erroneous solutions. The glyoxylate shunt is known to be inactive under the given conditions and one could use this information and constrain its flux to zero. However, we aim to produce a general method; of use under conditions where this information may not be available (e.g. exotic feeds or bioengineered cells). Hence, we decided not to constrain the glyoxylate flux in order to show how the genome-scale model constrained by measured data can produce testable and falsifiable consequences to detect this type of errors. In this case, the shuttling of glyoxylate through glycolate results in glycolate being exported out to the medium (S14 and S15 Figs) and detectable through (e.g.) MS methods. If no glycolate is found, that extracellular flux can be set to zero and the glyoxylate shunt flux will decrease to zero, so as not to deteriorate the fit of the malate labeling pattern. Alternatively, the availability of labeling patterns for additional metabolites (fumarate, fum and glyoxylate, glx) would confirm or deny the ICL activity. In this way, 2S- ^{13}C MFA can fruitfully use available data to suggest and test unexpected metabolic activity; such as the surprising heme degradation in cancer cells with a non-functional TCA cycle [26].

Using 2S- ^{13}C MFA to test flux prediction algorithms

Our goal in developing 2S- ^{13}C MFA is to make a clear distinction between highly reliable constraints such as those induced by ^{13}C labeling data, measured extracellular fluxes, carbon transition information and reaction stoichiometry, and reasonable hypotheses (such as growth rate optimization) that may not be universally applicable. 2S- ^{13}C MFA has been developed to provide a self-consistent and unbiased determination of the range of metabolic fluxes compatible with available experimental data. Unlike other COBRA methods, the fluxes obtained by 2S- ^{13}C

MFA are backed up by the extra validation provided by the correct fit of 48 relative labeling measurements (see [S1–S3 Figs](#)). Hence, we use it here as a reference to compare various COBRA predictive methods based on different hypotheses. This procedure permits us to determine which method predicts fluxes most accurately. We compared flux profile predictions for *pyk* and *pgi* gene knock outs calculated at three different time points (5, 6 and 7 hrs for wild type and *pyk* KO, and 16, 21 and 23 hrs for *pgi* KO) using six different methods: FBA maximizing either growth and ATP production [[74](#)], Minimization of Metabolic Adjustment (MOMA [[16](#)]), Regulatory On/Off Minimization (ROOM [[75](#)]) and two new methods developed in this paper ([Fig 9](#) and [S19](#) and [S20 Figs](#)). These new methods are ^{13}C MOMA and ^{13}C ROOM, which leverage $2\text{S-}^{13}\text{C}$ MFA flux profiles obtained for the wild type strain to improve flux predictions (see [S3 Text](#)), similar in spirit to the approach by Kuepfer *et al* [[39](#)].

Growth maximizing FBA provides predictions that are not quantitatively accurate, but seem approximately right for most cases ([S19](#) and [S20 Figs](#)). This fact probably explains its success in metabolic engineering applications. Since FBA is typically used to make partial flux predictions by using some measured flux information for the predicted experiment, we tested this variation as well. Once fluxes for growth rate and excreted metabolites were constrained to the measured values, *pgi* predictions became extremely accurate ([S20 Fig](#)). However, when a full prediction (i.e., not using any data from the predicted experiment) was sought, FBA noticeably failed for this KO due to its inability to predict the drop in growth rate. ATP maximizing FBA fails most noticeably for the *pyk* KO, probably because the PYK reaction is involved in ATP production and its elimination significantly changes the ATP balance when ATP is to be maximized.

MOMA, ROOM, ^{13}C MOMA and ^{13}C ROOM flux predictions fail most blatantly for the *pgi* knockout strain because growth rates change radically and the method tries to maintain the previous flux levels ([S19 Fig](#)). Hence, we can expect these methods to offer good results only when changes expected in glucose intake and growth rates are relatively small. An improvement of these methods could be obtained if only the relative flux profiles are used for the prediction in the algorithm and the growth rate is separately obtained.

Since ^{13}C MOMA and ^{13}C ROOM use $2\text{S-}^{13}\text{C}$ MFA flux profiles from the wild type strain, we expected that they would more accurately predict fluxes than would MOMA and ROOM. We find that is the case for fluxes in glycolysis and PPP, but the TCA cycle flux values are less accurate (see [S19 Fig](#)). This phenomena is most likely due to our initial flux profile not being very accurate for the TCA cycle (large confidence intervals, see [S4–S6 Figs](#)) since many fewer labeling measurements are available for TCA metabolites (only malate vs eight metabolites available for glycolysis and PPP). Labeling data for more metabolites next to branch points (e.g. fumarate, glyoxylate, isocitrate, alphaketoglutarate) would help reduce the flux confidence intervals for this area of metabolism.

These comparisons illustrate how $2\text{S-}^{13}\text{C}$ MFA can be used to test COBRA methods, see where they fail and why, and use this information to improve them.

Quantitative prediction of direct metabolite labeling measurements

The combination of ^{13}C labeling data and genome-scale models with COBRA prediction methods allows us to make some predictions we cannot do using either ^{13}C MFA or FBA. The ability to make accurate predictions of metabolism is of fundamental importance to make metabolic engineering a more predictive discipline.

The ^{13}C MOMA predictions of glycolysis and PPP fluxes for the *pyk* KO at the 5-hr time point surpass those of all other methods ([Fig 9](#) and [S19 Fig](#)). In fact, the flux predictions are precise enough that we can even predict with reasonable accuracy the metabolite labeling to be expected from that strain at that time point ([Fig 10](#)). This is a prediction of directly measured

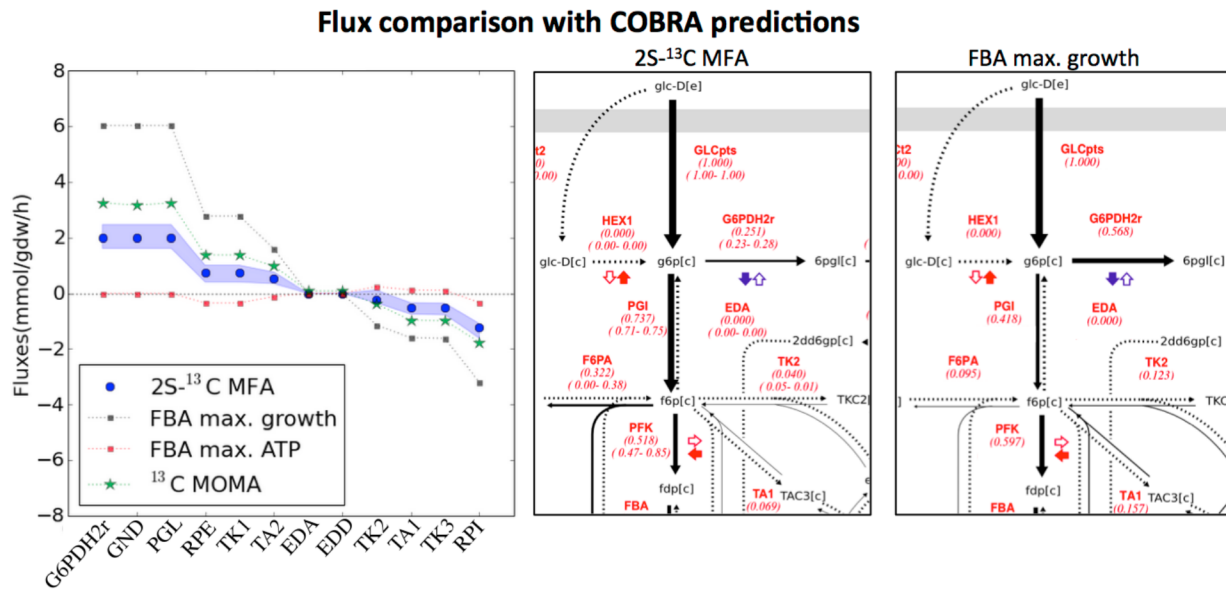


Fig 9. Flux comparison with COBRA predictions for *pyk* KO at 5 hours. The comparison of flux predictions through COBRA methods with fluxes measured through 2S-¹³C MFA shows how and why predictive methods fail. Left panel: Predicted fluxes for reactions in the pentose phosphate pathway (PPP) through FBA using maximum growth, FBA using maximum ATP production and ¹³C MOMA are compared with measurements through 2S-¹³C MFA (shading indicates confidence intervals). Fluxes predicted through maximum growth FBA offer a good qualitative description of fluxes but are quantitatively erroneous. ¹³C MOMA is a variation of MOMA that leverages 2S-¹³C MFA flux profiles and predicts fluxes more accurately. Center and right panels: Flux maps for two special cases (FBA and 2S-¹³C MFA). Maximum growth FBA overestimates flux into the PPP. Flux values are depicted in red: the upper value is the flux for the best fit and the lower values are the range of values compatible with data from ¹³C labeling experiments.

doi:10.1371/journal.pcbi.1004363.g009

data instead of a derived measurement such as flux. No information from the *pyk* strain at 5 hrs was used. We used the 2S-¹³C MFA flux calculations from the wild type strain at 5 hrs to obtain fluxes compatible with those measurements and then, through ¹³C MOMA, obtain the predicted fluxes if *pyk* were to be knocked out. The corresponding labeling patterns were then derived. Notice that this is very different from the partial predictions of labeling that ¹³C MFA routinely produces, where the labeling for the present experiment is predicted. Hence, this constitutes a full quantitative prediction of metabolite labeling from a ¹³C labeling experiment, which has not been reported before, to our knowledge.

The prediction of metabolite labeling is particularly relevant as validation because they refer to a directly measured quantity (i.e. the MDV obtained from the CE-TOFMS). Fluxes are derived quantities relying on a variety of implicit assumptions (i.e. the two-scale approximation, metabolic pseudo-steady state, no accumulation of intermediate metabolites, genome-scale model completeness and accuracy, cell homogeneity. . .). The real test that these assumptions are not severely violated and that the method provides reliable flux profiles is to use them to predict directly measured quantities for other experiments (in this case, labeling patterns). Moreover, this example shows that coupling ¹³C labeling data with COBRA methods opens the possibility to go beyond qualitative predictions (e.g. grow/no grow).

Conclusion

In this manuscript, we have shown how to maximize the information obtained from ¹³C data to constrain genome-scale models, and that once core metabolism is set by ¹³C labeling data information, the rest of metabolism is generally highly constricted. As is a usual behavior in “sloppy” nonlinear fitting problems [40], some fluxes are very effectively constrained and some

Metabolite labeling prediction

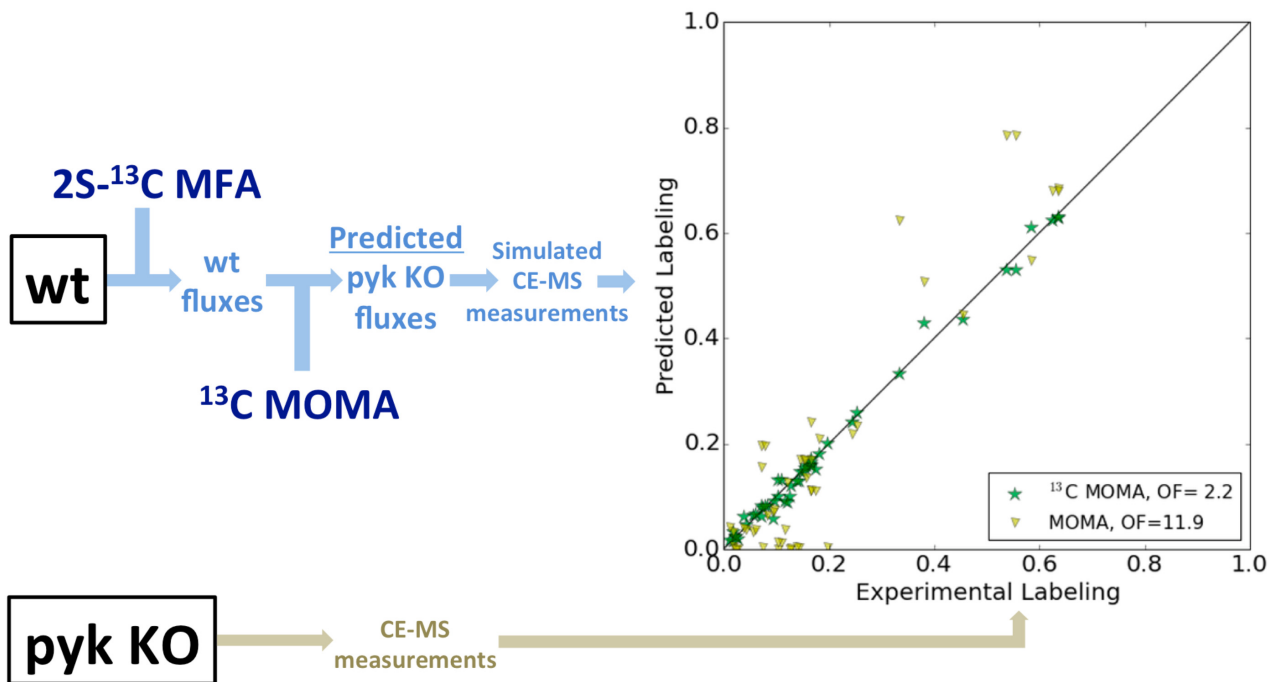


Fig 10. Metabolite Labeling prediction. The combination of ^{13}C labeling data and genome-scale models with COBRA prediction methods produces predictions that cannot be obtained through either ^{13}C MFA or FBA. The ^{13}C MOMA predictions of fluxes shown in Fig 9 and S19 Fig are accurate enough that they can be used to predict the metabolite labeling (MDV, see Fig 4) for the *pyk* strain at 5 hours without using any data from the experiment on the *pyk* strain. A genome-scale model is needed to use ^{13}C MOMA, and the accuracy provided by the ^{13}C data is necessary to produce an accurate initial flux profile for ^{13}C MOMA (see S2 Text). OF denotes the objective function: the average deviation of predicted labeling from the experimental value, measured in units of the experimental error. The prediction through standard MOMA, based on a FBA flux profile, is much less accurate (OF = 11.9) than the one obtained through ^{13}C MOMA (OF = 2.2).

doi:10.1371/journal.pcbi.1004363.g010

others only loosely. Confidence intervals obtained through 2S- ^{13}C MFA immediately identify these two types of fluxes and show that the use of ^{13}C labeling experimental data produce much narrower confidence intervals than those produced by FBA. The method is generally applicable to any genome-scale model or feed, and can be used to expand the use of ^{13}C -based flux analysis beyond the customary cases to tackle non-standard feeds, exotic organisms and systems described by large stoichiometric matrices such as the human metabolic network [76], those derived from adding macromolecular synthesis [77] or microbial communities [24].

2S- ^{13}C MFA produces similar results as ^{13}C MFA for the region where the latter is valid: central carbon metabolism. 2S- ^{13}C MFA, however, extrapolates the constraints induced by the ^{13}C labeling data to a genome-scale model, providing fluxes not only central carbon metabolism but also for peripheral metabolism. This was illustrated firstly by a detailed description of NADPH and NADH production and consumption and, secondly, by predicting unmeasured metabolites expected in the extracellular medium.

2S- ^{13}C MFA does not use an evolutionary optimization principle (such as growth rate optimization) but, rather characterizes all flux profiles compatible with the experimental data. The extra validation gained by matching the measured labeling values is used to test the validity of the maximization hypothesis by comparing these results with predictions obtained through FBA and other COBRA methods. The comparison of flux profiles predicted with COBRA methods and those obtained through 2S- ^{13}C MFA provided not only a ranking of accuracy for predictions but

also insight as to how to improve predictive methods. An improved version of MOMA using $2\text{S-}^{13}\text{C}$ MFA profiles as a starting point is able to predict the outcome of 48 direct measurements of metabolite labeling for a *pyk* KO experiment using only data from a different experiment involving the wild type. This capability shows that using ^{13}C labeling experimental data enables accurate predictions beyond qualitative cases (e.g. grow/no grow). This method represents another step in the effort to make bioengineering a more predictable endeavour.

The new method hinges on a simple assumption: flux flows from the core set to peripheral metabolism and does not flow back. This assumption is supported by the good fits obtained in general by ^{13}C MFA methods thus far. There might be situations where this two-scale assumption is not applicable and these will be pinpointed by unacceptable fits to the labeling data. The core set of reactions, however, is flexible and can be enlarged as needed to provide acceptable fits to the labeling data. Hence, phenomena such as protein turnover [48, 78] or cell scavenging in stationary phase can be included by adding the appropriate reactions to the core. The availability of carbon transitions for genome-scale models [79] facilitates a systematic core enlargement.

In spite of its simplicity, FBA-based modelling has already exhibited significant success: only three elements are used in this modelling scheme (genome-scale stoichiometry, measured extracellular fluxes and an optimization principle) but they have been successfully used to rationally engineer strains used for large-scale industrial production [21–23]. However, certainly not every single flux in a flux profile obtained through FBA can be trusted. $2\text{S-}^{13}\text{C}$ MFA unites the informative constraints of ^{13}C labeling experiments with genome-scale stoichiometry to improve the determination of internal metabolic fluxes and set confidence intervals based on experimental data. $2\text{S-}^{13}\text{C}$ MFA completes and improves ^{13}C MFA by enforcing a global balance of metabolites instead of balancing only a few chosen metabolites. We believe it will be a tool of extreme utility in bioengineering, at a time when a variety of different frameworks for flux prediction for genome-scale models are becoming available [80, 81]. Furthermore, we think that its widespread use to determine metabolic fluxes will affect our understanding of fundamental biological problems [82] beyond bioengineering.

Materials and Methods

Mathematical details for the optimization problems in each of the different phases depicted in Fig 2 can be found below. The full procedure was scripted in python and uses the CONOPT solver version 3.15D to solve the nonlinear problems and CPLEX version 12.4.0 for the linear programming problems within the GAMS modeling environment.

All experimental data were obtained from Toya *et al* [47]. This study includes intracellular metabolite labeling, incoming and outgoing extracellular fluxes for glucose, acetate and growth rate for three strains of *E. coli* (wild type BW25113 and *pgi* and *pyk* knockouts) at three different time points each (5, 6 and 7 hrs for wild type and *pyk* KO, and 16, 21 and 23 hrs for *pgi* KO). Measured metabolites are 3-phospho-D-glycerate (3pg), dihydroxyacetone phosphate (dhap), D-fructose 1,6-bisphosphate (fdp), L-malate (mal-L), phosphoenolpyruvate (pep), pyruvate (pyr), alpha-D-Ribose 5-phosphate (r5p), ribulose 5-phosphate (ru5p) and Sedoheptulose 7-phosphate (s7p). Throughout the text, metabolites and reactions are named according to the iJR904 model notation [57, 83](bigg.ucsd.edu). The comparison with ^{13}C MFA was done through the use of equations 4–9 in S1 Text. Equations for ROOM and MOMA can be found in S2 Text. While the iJR904 model was used in this manuscript, the method is equally applicable to iAF1260 or iJO1366. iJR904 was used because the compartmentalization in iAF1260 and iJO1366 (e.g. transport to periplasm) complicate the recursive procedure to generate a core set, without obviously improving the fit to the data for this case.

Protein turnover reactions were not included in the core set of reactions since it has been shown that protein degradation effects on intracellular metabolites are negligible once the labeling has reached steady state [84], which is the case for the current data set (see Fig 1 in [47]).

Limiting flux to core

The two-scale approximation assumes that non-core reactions do not contribute directly to the labeling of core metabolites, since carbon precursors flow from core metabolism into peripheral metabolism and do not flow back. This is represented in terms of a genome-scale model by limiting to zero the flux of reactions flowing into core metabolism (see Algorithm 1). The first step in $2\text{S-}^{13}\text{C}$ MFA (Fig 2) hence consists in taking each reaction that has a product in core metabolism and setting the upper bound to zero. However, it may be the case that this extra constraint makes it impossible to meet the measured growth rate (we check this by solving the corresponding FBA problem). In that case, setting the upper bound to a fraction of the glucose uptake rate is tested (first 0.05 and then 0.2 for this case). Since the labeling of core metabolism can be impacted by reversible reactions with reactants included in the core set as well, we cover this case by limiting the *lower* bound of the reaction to zero or the lowest value that permits growth. The impact of the reactions that could not be set to zero will be checked later through External Labeling Variability Analysis (ELVA, Figs 2 and 4). The input for the first step (Fig 2) is the genome-scale model with the carbon transitions for the core set integrated in it. The output consists of the genome-scale model with lower and upper bounds modified by this “Limiting flux to core” procedure. A detailed description of this first step in the diagram shown in Fig 2 can be found in the pseudo code in Algorithm 1.

Algorithm 1. “Limiting flux to core” pseudo code

```

for each reaction j flowing into core:
  limits = [ 0, 0.05, 0.2 ] * glucose_uptake
  limit = limits.next()
  goOn = True
  while goOn:
    if reaction j has forward flux:
      ub[ j ] = min(ub[ j ], limit)
    else if reaction j has backward flux:
      lb[ j ] = sign(lb[ j ]) * min(abs(lb[ j ]), limit)
    solve FBA problem
    goOn = (FBA problem has no solution) and (limit is not the last value in
limits)
    limit = limits.next()

```

Where `limits.next()` obtains next value in list `limits` (the first one if uninitiated), and `glucose_uptake` is the value of the glucose uptake rate. `Solve FBA problem` refers to finding the solution to the problem given by equations 1–3 in S1 Text. `has forward flux` refers to the reaction having a possible positive flux (i.e. positive upper bound, `ub`) flowing into the core set and `has backward flux` refers to the reaction having a possible negative flux (i.e. negative lower bound, `lb`) flowing into the core set for reversible reactions. `sign(lb[j])` is the sign of the lower bound `lb[j]`. `ub[j]` and `lb[j]` denote upper and lower bounds for reaction `j`, respectively.

$2\text{S-}^{13}\text{C}$ MFA data fit

The second step in $2\text{S-}^{13}\text{C}$ MFA (Fig 2) involves fitting the measured metabolite labeling by solving the optimization problem in Eqs 1–7, where the upper (ub_j) and lower bounds (lb_j)

have been limited by the previous step (“Limiting flux to core”). 2S-¹³C MFA is a hybrid of FBA and ¹³C MFA (see Fig 1 and S1 Text) where the stoichiometry constraint is applied to the full genome-scale network, as in the case of FBA, and the labeling constraints [58] are applied only to the core set of reactions and metabolites, as is the case for ¹³C MFA.

Unlike previous efforts [36, 39], these constraints are enforced simultaneously, instead of sequentially using the results of ¹³C MFA to constrain the FBA problem. This simultaneous approach is more rigorous than using slack coefficients (δ in Kuepfer *et al* [39]), does not need to invoke an optimization principle to calculate fluxes and allows for the global metabolite balance to affect core metabolism fluxes. Furthermore, it is also self-consistent whereas in the sequential approach one might find that fluxes that flow into core metabolism are active, even though they were not taken into account to do the initial ¹³C MFA fit.

In the notation of Suthers *et al* [58] (GAMS files available in S1 Code):

$$\text{Minimize } OF = \sqrt{\left(\sum_{\substack{e \in E_{meas} \\ m \in M_e}} \left(\frac{f_{em}^{exp} - f_{em}}{\Delta_{em}} \right)^2 / |M_e| \right) / |E_{meas}|} \quad (1)$$

Subject to:

$$\sum_j S_{ij} v_j = 0 \quad \forall i \in I^N, j \in J \quad (2)$$

$$lb_j \leq v_j \leq ub_j \quad \forall j \in J \quad (3)$$

$$\sum_{m \in M_e} f_{em} = 1 \quad \forall e \in E_{co} \quad (4)$$

$$\sum_{e' \in E_{co}} \left((\sum_{|EMM^l_{e' \rightarrow e} > 0} EMM^l_{e' \rightarrow e} V_l) f_{e'm} \right) + (\sum_{|S_{il}^* < 0} S_{il}^* V_l) f_{em} = 0 \quad \forall m \in M_e, e \in E_i, i \in I_{co}^N \quad (5)$$

$$f_{em} = \sum_{w \in W_{em}} \prod_{n=1}^{|E_e|} f_{nm_n} \quad \forall m \in M_e, e \in E_{co}^c \quad (6)$$

$$v_j = \sum_{l \in J_{co}^B} map_{jl} V_l \quad \forall j \in J \quad (7)$$

where:

Sets

- $I \equiv \{i\}$: Set of all metabolites.
- $I_{co} \subset I$: Set of core metabolites.
- $I_{co}^N \subset I_{co}$: Set of non – exchange core metabolites.
- J : Set of fluxes.
- $J_{co} \subset J$: Set of core fluxes.
- J^B : Set of fluxes with backward and forward fluxes differentiated, e.g. PGL_f, PGL_b, PGL... etc.
- $J_{co}^B \subset J^B$: Set of core fluxes for J^B .
- $E = \{e\}$: Elementary Metabolite Units (EMUs).
- $E^c \subset E$: Combined EMUs.
- $E_i \subset E$: EMUs from metabolite $i \in I$.
- $E_{co}^c \subset E^c$: Core combined EMUs.
- $E_e \subset E$: EMUs that produce combined EMU e .
- $E_{co} \subset E$: EMUs corresponding to core metabolites.
- $E_{meas} \subset E$: EMUs corresponding to measured EMUs.
- W_{em} : Set of every possible mass isotopomer multiplet of E_e that produce the mass isotopomer m of e .
- M_e : m values for MDV of emu e : $0, 1, \dots$, #of carbons in e .

Parameters

- $EMM_{e' \rightarrow e}^l = \frac{1}{k}$ if e' produces e through reaction $l \in J_{co}^B$, 0 otherwise. See Suthers *et al*[58].
- S_{ij} : Stoichiometry matrix.
- S_{il}^* : Stoichiometry matrix with backward and forward fluxes differentiated.
- ub_j, lb_j : Upper and lower bounds for reaction j .
- $f_{em}^{exp} \in [0, 1]$: Experimentally measured MDV for emu e from metabolite m .
- Δ_{em} : Measurement error for f_{em}^{exp} .
- $map_{jl} = 1 * glucupt$ if l corresponds to forward flux of j .
- $map_{jl} = -1 * glucupt$ if l corresponds to backward flux of j .

Variables

- v_j : Flux value of reaction $j \in J$, in mmol/gdw/h.
- V_l : Flux value of reaction $l \in J_{co}^B$, normalized to glucose input rate.
- $f_{em} \in [0, 1]$: Mass isotopomer fraction (MDV) for emu e from metabolite m .

Notice that S_{il}^* is not the same as S_{ij} , since J and J^B are slightly different sets of fluxes. In fact:

$$\begin{aligned} S_{il}^* &= S_{ij} && \text{if } l \text{ is the forward version of } j. \\ S_{il}^* &= -S_{ij} && \text{if } l \text{ is the backward version of } j. \end{aligned} \tag{8}$$

Notice that Eqs 2 and 3 are the traditional FBA constraints and that Eqs 4–6 are the ¹³C labeling constraints, but they have been limited to core metabolites and reactions. The mapping

(Eq 7) converts the fluxes from the ^{13}C MFA description (higher resolution) where fluxes are normalized to the glucose uptake rate (*glucupt*) into the FBA description (lower resolution). We describe the ^{13}C MFA description as of higher resolution because the ^{13}C labeling data can pick up differences in forward and backward fluxes (J^B set), whereas the purely stoichiometric approach of FBA can only constrain net fluxes (J set).

For all 2S- ^{13}C MFA calculations all input flux was supposed to be routed through GLCpts (glucose transport via PEP:pyr pts [57]). Optimization problems were run $N = 30$ times and the one with lowest objective function picked. The value of N was chosen so as to avoid relative minima of OF. A plot of how the OF saturated as N increased can be found in S21 Fig in the supplementary material.

In addition to the OF, the more typical sum-of-squares residuals (SSR) for the fits have been included in the legends of S1–S3 Figs. However, standard good-of-fitness metrics, such as those proposed by Antoniewicz *et al* [85], are not applicable to 2S- ^{13}C MFA. By using genome-scale models, the number of estimated free fluxes (p) is higher than the number of independent measurements ($n = 48 - 9 = 39$ in this case) and the $\chi^2(n - p)$ distribution of a negative number of degrees of freedom is then not properly defined. This apparent paradox arises because of the implicit assumptions in the χ^2 statistics approach to these types of fits. The null model assumes that each of the terms in the SSR are independent random variables, hence the degrees of freedom are the number of terms in the SSR [86]. Nonetheless, we know that, for ^{13}C MFA, the terms in the SSR are typically not independent: the labeling pattern (MDV) of related compounds (i.e. amino acids arising from the same precursors) are very similar. Hence, Antoniewicz *et al* [85], decided to use as degrees of freedom the number of individual measurements minus the number of estimated free fluxes. This proposal seems reasonable for standard ^{13}C MFA, but breaks down for genome-scale models with a much larger number of degrees of freedom.

One could choose a different null model for the χ^2 statistics (and previous approaches have shown that the standard χ^2 goodness-of-fit approach is probably too conservative, see Fig 3A in [27]) but the crux of the matter is that we believe that using p-values < 0.05 as an absolute arbitrary threshold for significant vs insignificant results is too simplistic, as do other biological researchers [87] or R.A. Fisher himself [88]. Hence, what we report here (in S1–S3 Figs) is what we believe is a better (and more intuitive) way to estimate how good a fit is: the average objective function normalized to the measurement error (Eq 1). This is the answer to the question: how different are my fits from the experimentally measured values, measured in units of the experimental error? This is in accordance in spirit with the suggestion of presenting measures of significance without arbitrary thresholds [87]. Notice that none of the objective functions (OF) in S1–S3 Figs is smaller than one, indicating that the difference between experiment and theory cannot be explained through experimental error. This may be because the experimental error for the labeling pattern was underestimated or because the model fails to explain the full labeling pattern. These results are in line with the general trend that fits to intracellular metabolites tend to be worse [89] than fits to proteogenic amino acids [49].

The discarding of the 0.05 p-value criterium does not imply that the considerable effort employed in obtaining excellent fits to data [49, 51] goes unrewarded. Under the 2S- ^{13}C MFA method, a bad fit to the experimental data results in a larger value for $\delta_e m$ in Eq 23 and wider confidence intervals. Hence, worse fits beget less flux resolution.

Finally, we think that the best validation of the of a flux fit is using the flux distribution to predict the results of another experiment, as we did in Figs 9 and 10.

External Labeling Variability Analysis (ELVA)

Once a core set is chosen, ELVA establishes the maximum impact of non-core metabolism in the labeling of the measured metabolites. In order to do so, only core metabolism is considered and the impact of non-core metabolism is represented through “inflow” metabolites and reactions (see Fig 3C for an example). Inflow reactions agglomerate all non-core reactions flowing into (or out of) a particular core metabolite (see S22 Fig in supplementary material) and are assigned trivial carbon transitions (e.g. abc \rightarrow abc for a three carbon metabolite). Inflow metabolites are dummy metabolites with the same number of carbons as the involved core metabolite. ELVA constrains fluxes to the solution obtained in the “Fit data” step (Eq 10) and maximizes and minimizes the computational MDV for each m in each of the labeled metabolites. Since fluxes are fixed for all reactions and labeling is fixed for all metabolites except inflow metabolites, the optimization problem in Eqs 9–15 quantifies the maximum and minimum effect that this unknown labeling (since it comes from non-core metabolism) could have on the labeling pattern for the measured metabolites ($f_{em} \forall m, e \in E_{meas}$):

$$\text{Min/max } f_{em} \quad \forall m \in M_e, e \in E_{meas} \quad (9)$$

Subject to:

$$V_j = \bar{V}_j \quad j \in J_{coext}^B \quad (10)$$

$$\sum_j S_{ij}^* V_j = 0 \quad \forall i \in I_{coext}^N, j \in J_{coext}^B \quad (11)$$

$$lb_j \leq V_j \leq ub_j \quad \forall j \in J_{coext}^B \quad (12)$$

$$\sum_{m \in M_e} f_{em} = 1 \quad \forall e \in E \quad (13)$$

$$\sum_{e' \in E} \left(\left(\sum_{j | EMM_{e' \rightarrow e}^j > 0} EMM_{e' \rightarrow e}^j V_j \right) f_{e'm} \right) + \left(\sum_{j | S_{ij}^* < 0} S_{ij}^* V_j \right) f_{em} = 0 \quad (14)$$

$$\forall m \in M_e, e \in E_i, i \in I^N$$

$$f_{em} = \sum_{w \in W_{em}} \prod_{n=1}^{|E_e|} f_{\epsilon_n m_n} \quad \forall m \in M_e, e \in E_{coext}^c \quad (15)$$

where symbols are as explained before, with the addition of:

Sets

I_{coext}^B : Set of extended metabolites.

J_{coext}^B : Set of extended fluxes with backward and forward fluxes differentiated.

Parameters

\bar{V}_j : solutions to the problem given by equations 1 – 7

J_{coext}^B and I_{coext} are the set of reactions and metabolites obtained after expanding the core set to meet stoichiometry requirements (Eq 11) as is explained in supplementary S22 Fig. The input for this step is the flux profile obtained from the data fit and the output is an ELVA plot (Fig 4) used to decide whether the solution is self-consistent or not.

¹³C flux variability analysis

Once a self-consistent core set has been determined through the recursive procedure in Fig 2, flux ranges compatible with the experimental data are obtained through the following optimization problem:

$$\text{Min/max } V_j \quad \forall j \in J \quad (16)$$

Subject to:

$$\sum_j S_{ij} v_j = 0 \quad \forall i \in I^N, j \in J \quad (17)$$

$$lb_j \leq v_j \leq ub_j \quad \forall j \in J \quad (18)$$

$$\sum_{m \in M_e} f_{em} = 1 \quad \forall e \in E_{co} \quad (19)$$

$$\sum_{e' \in E_{co}} ((\sum_{i|EMM_{e' \rightarrow e}^i > 0} EMM_{e' \rightarrow e}^i V_i) f_{e'm}) + (\sum_{i|S_{ij}^* < 0} S_{ij}^* V_i) f_{em} = 0 \quad (20)$$

$$\forall m \in M_e, e \in E_i, i \in I_{co}^N$$

$$f_{em} = \sum_{w \in W_{em}} \prod_{n=1}^{|E_c|} f_{e_n m_n} \quad \forall m \in M_e, e \in E^c \quad (21)$$

$$v_j = \sum_{l \in J_{co}^B} map_{jl} V_l \quad \forall j \in J \quad (22)$$

$$(f_{em} - f_{em}^{exp})^2 \leq \delta_{em}^2 \quad \forall e \in E_{meas}, m \in M_e \quad (23)$$

where symbols are as explained before, with the addition of:

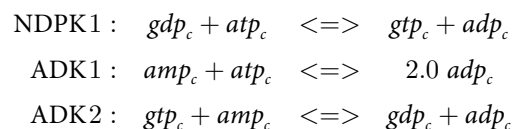
Parameters

δ_{em} : Maximum error allowed for f_{em} .

$\delta_{em} = \Delta_{em}$ if $\Delta_{em} > \epsilon_{em}$ and $\delta_{em} = 1.1 * \epsilon_{em}$ if $\Delta_{em} \leq \epsilon_{em}$

where $\epsilon_{em} = f_{em}^{fit} - f_{em}^{exp}$ from the solution to equations 1 – 7.

The fluxes v_j and V_j are initialized to the values obtained from solving Eqs 1–7. The results of the minimization and maximization give the flux upper and lower bound compatible with the experimental data from the ¹³C labeling experiments. This procedure is similar to the *FluxRange* procedure in Suthers *et al* [27], with the exception that we use set constraints for each measured data point (Eq 23) instead of only for the objective function. This approach guarantees that fluxes produce labeling patterns within the experimental error in a much more efficient way than a monte carlo approach. For example, consider these reactions which conform a futile cycle:



and for which the ^{13}C labeling data cannot constrain the flux values in any fashion. This is shown correctly in the confidence intervals obtained with the method expressed through Eqs 16–23:

$$\begin{aligned} \text{NDPK1} &: [-497.4 \quad - \quad 500 \quad] \\ \text{ADK1} &: [-497.4 \quad - \quad 500 \quad] \\ \text{ADK2} &: [-498.0 \quad - \quad 499.3] \end{aligned}$$

whereas if we perform a Monte Carlo, where labeling data is randomly chosen within the experimental error (100 instances), we do not obtain the proper range:

$$\begin{aligned} \text{NDPK1} &: [-490.8 \quad - \quad -489.7] \\ \text{ADK1} &: [500.0 \quad - \quad 500.0] \\ \text{ADK2} &: [-498.0 \quad - \quad -498.0] \end{aligned}$$

The input for this step is the flux profile obtained in the “Fit data” step and the outputs are the flux profile with corresponding confidence intervals, as shown in S4–S12 Figs.

Supporting Information

S1 Text. FBA and ^{13}C MFA equations.
(PDF)

S2 Text. ^{13}C MOMA and ^{13}C ROOM.
(PDF)

S3 Text. Reactions and carbon transitions.
(PDF)

S1 Code. GAMS files. GAMS files, input files and expected outputs for the nine strains considered (wild type at 5, 6 and 7 hrs, pyk KO at 5, 6 and 7 hrs and pgi KO at 16, 21 and 23 hrs) for the data fits.
(GZ)

S1 Fig. Labeling data fit for wild type strain. Red denotes the MDV for experimentally measured data, blue columns are the fit. The MDV is the fraction of molecules with $m = 1, 2, 3, 4 \dots$ ^{13}C incorporated atoms. Sum of square residuals (SSR, [85]) are 216.5, 282.2 and 467.7 for each strain from top to bottom.
(TIF)

S2 Fig. Labeling data fit for pyk KO. Red denotes the MDV for experimentally measured data, blue columns are the fit. SSRs are 773.2, 1195.9 and 817.36 for each strain from top to bottom.
(TIF)

S3 Fig. Labeling data fit for pgi KO. Red denotes the MDV for experimentally measured data, blue columns are the fit. SSRs are 7244.8, 6377.4 and 581.3 for each strain from top to bottom.
(TIF)

S4 Fig. 2S- ^{13}C MFA flux map for wild type at 5 hours. Best fit for flux is given on top red number for each reaction and confidence interval at the bottom. Cofactors and common metabolites are indicated by small arrows. Reversible reactions are indicated by double arrows.
(TIF)

S5 Fig. 2S- ^{13}C MFA flux map for wild type at 6 hours.
(TIF)

S6 Fig. $2\text{S-}^{13}\text{C}$ MFA flux map for wild type at 7 hours.
(TIF)

S7 Fig. $2\text{S-}^{13}\text{C}$ MFA flux map for *pyk* KO at 5 hours.
(TIF)

S8 Fig. $2\text{S-}^{13}\text{C}$ MFA flux map for *pyk* KO at 6 hours.
(TIF)

S9 Fig. $2\text{S-}^{13}\text{C}$ MFA flux map for *pyk* KO at 7 hours.
(TIF)

S10 Fig. $2\text{S-}^{13}\text{C}$ MFA flux map for *pgi* KO at 16 hours.
(TIF)

S11 Fig. $2\text{S-}^{13}\text{C}$ MFA flux map for *pgi* KO at 21 hours.
(TIF)

S12 Fig. $2\text{S-}^{13}\text{C}$ MFA flux map for *pgi* KO at 23 hours.
(TIF)

S13 Fig. Extracellular metabolite prediction. Maximum (dark bar) and minimum (light bar) values of the exchange fluxes obtained by $2\text{S-}^{13}\text{C}$ MFA show how ^{13}C experimental data can effectively constrain exchange fluxes that have not been measured (in blue). For comparison, maximum and minimum values for the model constrained by extracellular flux measurements (through FVA, in red) are included as well, as are maximum and minimum values obtained through FVA for a model constrained by extracellular flux measurements along with constraints induced by the two-scale approximation (black, see “Limiting flux to core” section). Fluxes are for the wild type at 5 hrs, and exchanged metabolites are indicated in the x axis (iJR904 notation, [57, 83]). A positive exchange flux (excreted metabolite) that remains positive for long enough should produce a detectable pool of the corresponding metabolite. Acetate and glucose are used as constraints for the flux determination, hence the confidence intervals are very narrow. For this particular case, glycolate and urea are expected in the media.
(TIF)

S14 Fig. Extracellular flux prediction for *pyk* KO at 6 hours. Expected metabolites in the medium include alpha-Ketoglutarate (akg) and glycolate (glyclt).
(TIF)

S15 Fig. Extracellular flux prediction for *pgi* KO at 21 hours. Expected metabolites in the medium include fumarate (fum) and acetaldehyde (acald).
(TIF)

S16 Fig. Version of Fig 8 for glutatamate dehydrogenase (GLUDy) and isocitrate dehydrogenase (ICDHyr) reactions. Since the flux value for SUCD1i is negative, the absolute value has been plotted.
(TIF)

S17 Fig. Comparison between original flux profiles for the *pgi* KO and flux profile after changing biomass composition according to previously reported results [90, 91] Changes in flux profiles in central carbon metabolism are minimal. The ^{13}C labeling data constrains fluxes strongly to a particular solution whereas changes in biomass requirements can be easily accommodated by the increased degrees of freedom found in genome-scale models.
(TIF)

S18 Fig. ^{13}C MFA flux map for wild type at 5 hours.
(TIF)

S19 Fig. Flux profile comparison of full predictions with $2\text{S-}^{13}\text{C}$ MFA. Full prediction means that no data from the target experiment was used to constrain fluxes: all predictions were derived from data from a different experiment. $2\text{S-}^{13}\text{C}$ MFA profiles are found by solving Eqs 1–7. Maximum growth and ATP profiles are found by solving equations 1–3 in [S1 Text](#). MOMA, ^{13}C MOMA, ROOM and ^{13}C ROOM flux profiles are obtained as explained in [S3 Text](#). Fluxes are sorted according to the following order; (PPP): G6PDH2r, GND, PGL, RPE, TK1, TA2, EDA, EDD, TK2, TA1, TK3, RPI; (GG): GAPD, ENO, PDH, TPI, PGI, FBA, PFK, F6PA, GLCS1, GLGC, FBP, G1PP, GLCP, HEX1, PPS, PYK, PGM, PGK; (CAC): FUM, MDH, ACONT, CS, ICDHyr, SUCD1i, AKGDH, CITL, FRD2, FRD3, MDH2, MDH3, SUCOAS; (OUT): EX_ac(e), BiomassEcoli, EX_glc(e). All reaction names according to iJR906 [\[57\]](#).
(TIF)

S20 Fig. Flux profile comparison of partial predictions with $2\text{S-}^{13}\text{C}$ MFA. Partial prediction means data from the target experiment was used to constrain fluxes, in this case the values for the growth rate, glucose intake and acetate excretion rate. $2\text{S-}^{13}\text{C}$ MFA profiles are found by solving Eqs 1–7 in the main paper. Maximum growth and ATP profiles are found by solving equations 1–3 in [S1 Text](#). Transparencies indicate confidence intervals for the $2\text{S-}^{13}\text{C}$ MFA.
(TIF)

S21 Fig. Average objective Function (OF) scaling with number of processes run (N) for wild type at 5 hours. OF plateaus at $N \approx 15$. We chose $N = 30$ for our simulations.
(TIF)

S22 Fig. Obtaining the network for External Labeling Variability Analysis (ELVA). The purpose of ELVA is to determine if the reactions left out of the core metabolism significantly affect core metabolite labeling. In order to do so, only the core metabolism network is used and non-core metabolism is represented through inflow reactions and metabolites. Inflow reactions and metabolites are dummy reactions and metabolites that aggregate the non-core effects. In the figure, black denotes core metabolites and reactions and blue denotes noncore metabolites and reactions. Inflow reactions and metabolites are added to the core set to meet stoichiometric requirements (since core fluxes are fixed to the values obtained in the previous “Fit data” step). For example, the upper figures show how reactions ARGSL and ADSL1r are combined into ROfum leading into the dummy metabolite OUTfum, while keeping the same net flux out of fumarate. In the case of the lower figures ARGSL and ADSL1r have a net flux into the core set and are substituted by a inward flowing dummy reaction RIfum and a dummy metabolite OUTfum. The point of the ELVA is to elucidate the impact of the metabolites in the noncore set (see [materials and methods](#)). Outflowing reactions have no effect (upper panels) but inflowing reactions do (lower panels). The labeling of incoming dummy metabolites is left unconstrained since its value is not being tracked and our goal is to determine what is the maximum effect they may have on the measured metabolite labeling.
(TIF)

Acknowledgments

We thank Harvey Blanch for valuable discussions.

Author Contributions

Analyzed the data: HGM. Contributed reagents/materials/analysis tools: VSK DW AG. Wrote the paper: HGM AA JDK VC AM.

References

1. Heinemann M, Sauer U. Systems biology of microbial metabolism. *Current opinion in microbiology*. 2010 Jun; 13(3):337–43. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20219420>. doi: [10.1016/j.mib.2010.02.005](https://doi.org/10.1016/j.mib.2010.02.005) PMID: [20219420](https://pubmed.ncbi.nlm.nih.gov/20219420/)
2. Joyce AR, Palsson BO. The model organism as a system: integrating ‘omics’ data sets. *Nature Reviews Molecular Cell Biology*. 2006; 7(3):198–210. doi: [10.1038/nrm1857](https://doi.org/10.1038/nrm1857) PMID: [16496022](https://pubmed.ncbi.nlm.nih.gov/16496022/)
3. Kitano H. Computational systems biology. *Nature*. 2002; 420(6912):206–210. doi: [10.1038/nature01254](https://doi.org/10.1038/nature01254) PMID: [12432404](https://pubmed.ncbi.nlm.nih.gov/12432404/)
4. Sauer U. Metabolic networks in motion: 13C-based flux analysis. *Molecular Systems Biology*. 2006; 2. doi: [10.1038/msb4100109](https://doi.org/10.1038/msb4100109) PMID: [17102807](https://pubmed.ncbi.nlm.nih.gov/17102807/)
5. Tang YJ, Martin HG, Myers S, Rodriguez S, Baidoo EEK, Keasling JD. Advances in analysis of microbial metabolic fluxes via 13C isotopic labeling. *Mass spectrometry reviews*. 2009; 28(2):362–375. doi: [10.1002/mas.20191](https://doi.org/10.1002/mas.20191) PMID: [19025966](https://pubmed.ncbi.nlm.nih.gov/19025966/)
6. G N Stephanopoulos, A A Aristidiou, J Nielsen. *Metabolic Engineering, Principles and Methodologies*; 1998.
7. Llaneras F, Sala A, Picó J. A possibilistic framework for constraint-based metabolic flux analysis. *BMC systems biology*. 2009 Jan; 3(1):79. doi: [10.1186/1752-0509-3-79](https://doi.org/10.1186/1752-0509-3-79) PMID: [19646223](https://pubmed.ncbi.nlm.nih.gov/19646223/)
8. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nature reviews Microbiology*. 2012 Feb; 10(4):291–305. doi: [10.1038/nrmicro2737](https://doi.org/10.1038/nrmicro2737) PMID: [22367118](https://pubmed.ncbi.nlm.nih.gov/22367118/)
9. Wiechert W. 13C metabolic flux analysis. *Metabolic engineering*. 2001 Jul; 3(3):195–206. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3530911&tool=pmcentrez&rendertype=abstract>. doi: [10.1006/mben.2001.0187](https://doi.org/10.1006/mben.2001.0187) PMID: [11461141](https://pubmed.ncbi.nlm.nih.gov/11461141/)
10. Nyberg GB, Balcarcel RR, Follstad BD, Stephanopoulos G, Wang DI. Metabolism of peptide amino acids by Chinese hamster ovary cells grown in a complex medium. *Biotechnology and bioengineering*. 1999 Feb; 62(3):324–35. doi: [10.1002/\(SICI\)1097-0290\(19990205\)62:3%3C324::AID-BIT9%3E3.0.CO;2-C](https://doi.org/10.1002/(SICI)1097-0290(19990205)62:3%3C324::AID-BIT9%3E3.0.CO;2-C) PMID: [10099544](https://pubmed.ncbi.nlm.nih.gov/10099544/)
11. Herwig C, von Stockar U. A small metabolic flux model to identify transient metabolic regulations in *Saccharomyces cerevisiae*. *Bioprocess and Biosystems Engineering*. 2002 Mar; 24(6):395–403. doi: [10.1007/s00449-001-0277-2](https://doi.org/10.1007/s00449-001-0277-2)
12. Bonarius HP, Hatzimanikatis V, Meesters KP, de Gooijer CD, Schmid G, Tramper J. Metabolic flux analysis of hybridoma cells in different culture media using mass balances. *Biotechnology and bioengineering*. 1996 May; 50(3):299–318. doi: [10.1002/\(SICI\)1097-0290\(19960505\)50:3%3C299::AID-BIT9%3E3.0.CO;2-B](https://doi.org/10.1002/(SICI)1097-0290(19960505)50:3%3C299::AID-BIT9%3E3.0.CO;2-B) PMID: [18626958](https://pubmed.ncbi.nlm.nih.gov/18626958/)
13. Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*. 2010; 5(1):93–121. doi: [10.1038/nprot.2009.203](https://doi.org/10.1038/nprot.2009.203) PMID: [20057383](https://pubmed.ncbi.nlm.nih.gov/20057383/)
14. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*. 2007; 3. doi: [10.1038/msb4100162](https://doi.org/10.1038/msb4100162) PMID: [17625511](https://pubmed.ncbi.nlm.nih.gov/17625511/)
15. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protocols*. 2011; 6(9):1290–1307. doi: [10.1038/nprot.2011.308](https://doi.org/10.1038/nprot.2011.308) PMID: [21886097](https://pubmed.ncbi.nlm.nih.gov/21886097/)
16. Segrè D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*. 2002 Nov; 99(23):15112–7. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=137552&tool=pmcentrez&rendertype=abstract>. doi: [10.1073/pnas.232349399](https://doi.org/10.1073/pnas.232349399) PMID: [12415116](https://pubmed.ncbi.nlm.nih.gov/12415116/)
17. Fong SS, Palsson BO. Metabolic gene deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nature Genetics*. 2004; 36(10):1056–1058. doi: [10.1038/ng1432](https://doi.org/10.1038/ng1432) PMID: [15448692](https://pubmed.ncbi.nlm.nih.gov/15448692/)
18. Kumar VS, Maranas CD. GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. *PLoS computational biology*. 2009 Mar; 5(3):e1000308. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2645679&tool=pmcentrez&rendertype=abstract>. doi: [10.1371/journal.pcbi.1000308](https://doi.org/10.1371/journal.pcbi.1000308) PMID: [19282964](https://pubmed.ncbi.nlm.nih.gov/19282964/)

19. 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). 2010; 26(12):i255–i260. doi: [10.1093/bioinformatics/btq183](https://doi.org/10.1093/bioinformatics/btq183)
20. Park JH, Lee KH, Kim TY, Lee SY. Metabolic engineering of *Escherichia coli* for the production of L-valine based on transcriptome analysis and in silico gene knockout simulation. *Proceedings of the National Academy of Sciences of the United States of America*. 2007 May; 104(19):7797–802. Available from: <http://www.pnas.org/cgi/content/abstract/104/19/7797>. doi: [10.1073/pnas.0702609104](https://doi.org/10.1073/pnas.0702609104) PMID: [17463081](https://pubmed.ncbi.nlm.nih.gov/17463081/)
21. Yim H, Haselbeck R, Niu W, Pujol-Baxley C, Burgard A, Boldt J, et al. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. *Nature Chemical Biology*. 2011; 7(7):445–452. doi: [10.1038/nchembio.580](https://doi.org/10.1038/nchembio.580) PMID: [21602812](https://pubmed.ncbi.nlm.nih.gov/21602812/)
22. Lane, J. Genomatica and the art of big wave surfing; 2013. Available from: <http://www.biofuelsdigest.com/bdigest/2013/02/12/genomatica-and-the-art-of-big-wave-surfing/>.
23. Lane, J. The Greening of BASF: #1 chemco commits to biobased BDO; 2013. Available from: <http://www.biofuelsdigest.com/bdigest/2013/05/13/the-greening-of-basf-1-chemco-commits-to-biobased-bdo/>.
24. Stolyar S, Van Dien S, Hillesland KL, Pinel N, Lie TJ, Leigh JA, et al. Metabolic modeling of a mutualistic microbial community. *Molecular systems biology*. 2007 Jan; 3(92):92. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1847946&tool=pmcentrez&rendertype=abstract>. doi: [10.1038/msb4100131](https://doi.org/10.1038/msb4100131) PMID: [17353934](https://pubmed.ncbi.nlm.nih.gov/17353934/)
25. Keibler Ma, Fendt SM, Stephanopoulos G. Expanding the concepts and tools of metabolic engineering to elucidate cancer metabolism. *Biotechnology progress*. 2012; 28(6):1409–18. doi: [10.1002/btpr.1629](https://doi.org/10.1002/btpr.1629) PMID: [22961737](https://pubmed.ncbi.nlm.nih.gov/22961737/)
26. Frezza C, Zheng L, Folger O, Rajagopalan KN, MacKenzie ED, Jerby L, et al. Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature*. 2011 Sep; 477(7363):225–8. doi: [10.1038/nature10363](https://doi.org/10.1038/nature10363) PMID: [21849978](https://pubmed.ncbi.nlm.nih.gov/21849978/)
27. Suthers PF, Burgard AP, Dasika MS, Nowroozi F, Van Dien S, Keasling JD, et al. Metabolic flux elucidation for large-scale models using ¹³C labeled isotopes. *Metabolic Engineering*. 2007; 9(5–6):387–405. doi: [10.1016/j.ymben.2007.05.005](https://doi.org/10.1016/j.ymben.2007.05.005) PMID: [17632026](https://pubmed.ncbi.nlm.nih.gov/17632026/)
28. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Elementary metabolite units (EMU): A novel framework for modeling isotopic distributions. *Metabolic Engineering*. 2007; 9(1):68–86. doi: [10.1016/j.ymben.2006.09.001](https://doi.org/10.1016/j.ymben.2006.09.001) PMID: [17088092](https://pubmed.ncbi.nlm.nih.gov/17088092/)
29. He L, Xiao Y, Gebreselassie N, Zhang F, Antoniewicz MR, Tang YJ, et al. Central metabolic responses to the overproduction of fatty acids in *Escherichia coli* based on (¹³C)-metabolic flux analysis. *Biotechnology and bioengineering*. 2013 Oct; 111(3):575–585. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/24122357>. doi: [10.1002/bit.25124](https://doi.org/10.1002/bit.25124)
30. Reed JL, Senger RS, Antoniewicz MR, Young JD. Computational approaches in metabolic engineering. *Journal of biomedicine & biotechnology*. 2010 Jan; 2010:207414. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3092504&tool=pmcentrez&rendertype=abstract>.
31. Ahn WS, Antoniewicz MR. Towards dynamic metabolic flux analysis in CHO cell cultures. *Biotechnology journal*. 2012 Jan; 7(1):61–74. doi: [10.1002/biot.201100052](https://doi.org/10.1002/biot.201100052) PMID: [22102428](https://pubmed.ncbi.nlm.nih.gov/22102428/)
32. Müller S, Hiller K, Metallo CM. Profiling metabolic networks to study cancer metabolism. *Current Opinion in Biotechnology*. 2013; 24(1):60–68. Available from: <http://www.sciencedirect.com/science/article/pii/S0958166912001772>. doi: [10.1016/j.copbio.2012.11.001](https://doi.org/10.1016/j.copbio.2012.11.001)
33. Ghosh A, Zhao H, Price ND. Genome-scale consequences of cofactor balancing in engineered pentose utilization pathways in *Saccharomyces cerevisiae*. *PloS one*. 2011 Jan; 6(11):e27316. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3208632&tool=pmcentrez&rendertype=abstract>. doi: [10.1371/journal.pone.0027316](https://doi.org/10.1371/journal.pone.0027316) PMID: [22076150](https://pubmed.ncbi.nlm.nih.gov/22076150/)
34. Fischer E, Sauer U. Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nature genetics*. 2005 Jun; 37(6):636–40. doi: [10.1038/ng1555](https://doi.org/10.1038/ng1555) PMID: [15880104](https://pubmed.ncbi.nlm.nih.gov/15880104/)
35. Schuetz R, Zamboni N, Zampieri M, Heinemann M, Sauer U. Multidimensional Optimality of Microbial Metabolism. *Science*. 2012 May; 336(6081):601–604. doi: [10.1126/science.1216882](https://doi.org/10.1126/science.1216882) PMID: [22556256](https://pubmed.ncbi.nlm.nih.gov/22556256/)
36. Choi HS, Kim TY, Lee DY, Lee SY. Incorporating metabolic flux ratios into constraint-based flux analysis by using artificial metabolites and converging ratio determinants. *Journal of Biotechnology*. 2007; 129(4):696–705. doi: [10.1016/j.jbiotec.2007.02.026](https://doi.org/10.1016/j.jbiotec.2007.02.026) PMID: [17408794](https://pubmed.ncbi.nlm.nih.gov/17408794/)
37. Quek LE, Wittmann C, Nielsen LK, Krömer JO. OpenFLUX: efficient modelling software for ¹³C-based metabolic flux analysis. *Microbial Cell Factories*. 2009; 8(1):25. doi: [10.1186/1475-2859-8-25](https://doi.org/10.1186/1475-2859-8-25) PMID: [19409084](https://pubmed.ncbi.nlm.nih.gov/19409084/)

38. Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y. Synergy between ¹³C-metabolic flux analysis and flux balance analysis for understanding metabolic adaptation to anaerobiosis in *E. coli*. *Metabolic Engineering*. 2011; 13(1):38–48. doi: [10.1016/j.ymben.2010.11.004](https://doi.org/10.1016/j.ymben.2010.11.004) PMID: [21129495](https://pubmed.ncbi.nlm.nih.gov/21129495/)
39. Kuepfer L, Sauer U, Blank LM. Metabolic functions of duplicate genes in *Saccharomyces cerevisiae*. *Genome research*. 2005 Oct; 15(10):1421–30. Available from: <http://genome.cshlp.org/content/15/10/1421.long>. doi: [10.1101/gr.3992505](https://doi.org/10.1101/gr.3992505) PMID: [16204195](https://pubmed.ncbi.nlm.nih.gov/16204195/)
40. Brown KS, Hill CC, Calero GA, Myers CR, Lee KH, Sethna JP, et al. The statistical mechanics of complex signaling networks: nerve growth factor signaling. *Physical biology*. 2004 Dec; 1(3–4):184–95. Available from: <http://iopscience.iop.org/1478-3975/1/3/006/fulltext/>. doi: [10.1088/1478-3967/1/3/006](https://doi.org/10.1088/1478-3967/1/3/006) PMID: [16204838](https://pubmed.ncbi.nlm.nih.gov/16204838/)
41. Brown KS, Sethna JP. Statistical mechanical approaches to models with many poorly known parameters. *Physical review E, Statistical, nonlinear, and soft matter physics*. 2003 Aug; 68(2 Pt 1):021904. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/14525003>. doi: [10.1103/PhysRevE.68.021904](https://doi.org/10.1103/PhysRevE.68.021904) PMID: [14525003](https://pubmed.ncbi.nlm.nih.gov/14525003/)
42. Transtrum MK, Machta BB, Sethna JP. Why are Nonlinear Fits to Data so Challenging? *Physical Review Letters*. 2010 Feb; 104(6):060201. doi: [10.1103/PhysRevLett.104.060201](https://doi.org/10.1103/PhysRevLett.104.060201) PMID: [20366807](https://pubmed.ncbi.nlm.nih.gov/20366807/)
43. Transtrum MK, Machta BB, Sethna JP. Geometry of nonlinear least squares with applications to sloppy models and optimization. *Physical Review E*. 2011 Mar; 83(3):036701. doi: [10.1103/PhysRevE.83.036701](https://doi.org/10.1103/PhysRevE.83.036701)
44. Gutenkunst RN, Casey FP, Waterfall JJ, Myers CR, Sethna JP. Extracting falsifiable predictions from sloppy models. *Annals of the New York Academy of Sciences*. 2007 Dec; 1115:203–11. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/17925353>. doi: [10.1196/annals.1407.003](https://doi.org/10.1196/annals.1407.003) PMID: [17925353](https://pubmed.ncbi.nlm.nih.gov/17925353/)
45. Waterfall J, Casey F, Gutenkunst R, Brown K, Myers C, Brouwer P, et al. Sloppy-Model Universality Class and the Vandermonde Matrix. *Physical Review Letters*. 2006 Oct; 97(15):150601. doi: [10.1103/PhysRevLett.97.150601](https://doi.org/10.1103/PhysRevLett.97.150601) PMID: [17155311](https://pubmed.ncbi.nlm.nih.gov/17155311/)
46. Frederiksen Sr, Jacobsen K, Brown K, Sethna J. Bayesian Ensemble Approach to Error Estimation of Interatomic Potentials. *Physical Review Letters*. 2004 Oct; 93(16):165501. doi: [10.1103/PhysRevLett.93.165501](https://doi.org/10.1103/PhysRevLett.93.165501) PMID: [15525000](https://pubmed.ncbi.nlm.nih.gov/15525000/)
47. Toya Y, Ishii N, Nakahigashi K, Hirasawa T, Soga T, Tomita M, et al. ¹³C-metabolic flux analysis for batch culture of *Escherichia coli* and its *pyk* and *pgi* gene knockout mutants based on mass isotopomer distribution of intracellular metabolites. *Biotechnology progress*. 2010; 26(4):975–992. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20730757http://doi.wiley.com/10.1002/btpr.420papers2://publication/doi/10.1002/btpr.420>. PMID: [20730757](https://pubmed.ncbi.nlm.nih.gov/20730757/)
48. Iwatani S, Van Dien S, Shimbo K, Kubota K, Kageyama N, Iwahata D, et al. Determination of metabolic flux changes during fed-batch cultivation from measurements of intracellular amino acids by LC-MS/MS. *Journal of biotechnology*. 2007 Jan; 128(1):93–111. doi: [10.1016/j.jbiotec.2006.09.004](https://doi.org/10.1016/j.jbiotec.2006.09.004) PMID: [17055605](https://pubmed.ncbi.nlm.nih.gov/17055605/)
49. Antoniewicz MR, Krainie DF, Laffend LA, González-Lergier J, Kelleher JK, Stephanopoulos G. Metabolic flux analysis in a nonstationary system: fed-batch fermentation of a high yielding strain of *E. coli* producing 1,3-propanediol. *Metabolic engineering*. 2007 May; 9(3):277–92. doi: [10.1016/j.ymben.2007.01.003](https://doi.org/10.1016/j.ymben.2007.01.003) PMID: [17400499](https://pubmed.ncbi.nlm.nih.gov/17400499/)
50. Schaub J, Mauch K, Reuss M. Metabolic flux analysis in *Escherichia coli* by integrating isotopic dynamic and isotopic stationary ¹³C labeling data. *Biotechnology and bioengineering*. 2008 Apr; 99(5):1170–85. doi: [10.1002/bit.21675](https://doi.org/10.1002/bit.21675) PMID: [17972325](https://pubmed.ncbi.nlm.nih.gov/17972325/)
51. Moxley J. Linking high-resolution metabolic flux phenotypes and transcriptional regulation in yeast modulated by the global regulator *Gcn4p*. *Proceedings of the ...* 2009; 106(16):6477–6482. Available from: <http://www.pnas.org/content/106/16/6477.short>.
52. Kajihata S, Matsuda F, Yoshimi M, Hayakawa K, Furusawa C, Kanda A, et al. (¹³C)-based metabolic flux analysis of *Saccharomyces cerevisiae* with a reduced Crabtree effect. *Journal of bioscience and bioengineering*. 2015 Jan; Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25634548>. doi: [10.1016/j.jbiosc.2014.12.014](https://doi.org/10.1016/j.jbiosc.2014.12.014)
53. Oberhardt MA, Palsson BO, Papin JA. Applications of genome-scale metabolic reconstructions. *Molecular systems biology*. 2009 Jan; 5:320. doi: [10.1038/msb.2009.77](https://doi.org/10.1038/msb.2009.77) PMID: [19888215](https://pubmed.ncbi.nlm.nih.gov/19888215/)
54. Pramanik J, Keasling JD. Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnology and bioengineering*. 1998 Oct; 60(2):230–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10099424>. doi: [10.1002/\(SICI\)1097-0290\(19981020\)60:2%3C230::AID-BIT10%3E3.0.CO;2-Q](https://doi.org/10.1002/(SICI)1097-0290(19981020)60:2%3C230::AID-BIT10%3E3.0.CO;2-Q) PMID: [10099424](https://pubmed.ncbi.nlm.nih.gov/10099424/)
55. Javidpour P, Pereira JH, Goh EB, McAndrew RP, Ma SM, Friedland GD, et al. Biochemical and structural studies of NADH-dependent FabG used to increase the bacterial production of fatty acids under

- anaerobic conditions. *Applied and environmental microbiology*. 2014 Jan; 80(2):497–505. Available from: <http://aem.asm.org/content/80/2/497.long>. doi: [10.1128/AEM.03194-13](https://doi.org/10.1128/AEM.03194-13) PMID: [24212572](https://pubmed.ncbi.nlm.nih.gov/24212572/)
56. Goldenfeld N, Athreya BP, Dantzig JA. Renormalization Group Approach to Multiscale Modelling in Materials Science. *Journal of Statistical Physics*. 2006 Mar; 125(5–6):1015–1023. doi: [10.1007/s10955-005-9013-7](https://doi.org/10.1007/s10955-005-9013-7)
 57. Reed JL, Vo TD, Schilling CH, Palsson BO. An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome biology*. 2003 Jan; 4(9):R54. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=193654&tool=pmcentrez&rendertype=abstract>. doi: [10.1186/gb-2003-4-9-r54](https://doi.org/10.1186/gb-2003-4-9-r54) PMID: [12952533](https://pubmed.ncbi.nlm.nih.gov/12952533/)
 58. Suthers PF, Chang YJ, Maranas CD. Improved computational performance of MFA using elementary metabolite units and flux coupling. *Metabolic Engineering*. 2009;p. 1–6. Available from: <http://dx.doi.org/10.1016/j.ymben.2009.10.002papers2://publication/doi/10.1016/j.ymben.2009.10.002>.
 59. Wahl A, ElMassaoudi M, Schipper D, Wiechert W, Takors R. Serial 13C-Based Flux Analysis of an L-Phenylalanine-Producing *E. coli* Strain Using the Sensor Reactor. *Biotechnology progress*. 2004; 20(3):706–714. doi: [10.1021/bp0342755](https://doi.org/10.1021/bp0342755) PMID: [15176872](https://pubmed.ncbi.nlm.nih.gov/15176872/)
 60. van Rijsewijk Bart R B H, Nanchen A, Nallet S, Kleijn RJ, Sauer U. Large-scale 13C-flux analysis reveals distinct transcriptional control of respiratory and fermentative metabolism in *Escherichia coli*. *Molecular Systems Biology*. 2011; 7:1–12. Available from: <http://dx.doi.org/10.1038/msb.2011.9papers2://publication/doi/10.1038/msb.2011.9>.
 61. Ranganathan S, Wei Tee T, Chowdhury A, Zomorodi AR, Moon Yoon J, Fu Y, et al. An integrated computational and experimental study for overproducing fatty acids in *Escherichia coli*. *Metabolic engineering*. 2012 Oct;p. 1–18. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23036703>.
 62. Okahashi N, Kajihata S, Furusawa C, Shimizu H. Reliable Metabolic Flux Estimation in *Escherichia coli* Central Carbon Metabolism Using Intracellular Free Amino Acids. *Metabolites*. 2014 Jan; 4(2):408–20. Available from: <http://www.mdpi.com/2218-1989/4/2/408/htm>. doi: [10.3390/metabo4020408](https://doi.org/10.3390/metabo4020408) PMID: [24957033](https://pubmed.ncbi.nlm.nih.gov/24957033/)
 63. Crown SB, Long CP, Antoniewicz MR. Integrated (13)C-metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. *Metabolic engineering*. 2015 Jan; doi: [10.1016/j.ymben.2015.01.001](https://doi.org/10.1016/j.ymben.2015.01.001) PMID: [25596508](https://pubmed.ncbi.nlm.nih.gov/25596508/)
 64. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic Engineering*. 2003 Oct; 5(4):264–276. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S1096717603000582>. doi: [10.1016/j.ymben.2003.09.002](https://doi.org/10.1016/j.ymben.2003.09.002) PMID: [14642354](https://pubmed.ncbi.nlm.nih.gov/14642354/)
 65. Jeon SM, Chandel NS, Hay N. AMPK regulates NADPH homeostasis to promote tumour cell survival during energy stress. *Nature*. 2012 May; 485(7400):661–5. doi: [10.1038/nature11066](https://doi.org/10.1038/nature11066) PMID: [22660331](https://pubmed.ncbi.nlm.nih.gov/22660331/)
 66. Bruinenberg PM, Bot PHM, Dijken JP, Scheffers WA. The role of redox balances in the anaerobic fermentation of xylose by yeasts. *European Journal of Applied Microbiology and Biotechnology*. 1983; 18(5):287–292. doi: [10.1007/BF00500493](https://doi.org/10.1007/BF00500493)
 67. Van Dien S. From the first drop to the first truckload: commercialization of microbial processes for renewable chemicals. *Current Opinion in Biotechnology*. 2013 Mar; 24:1–8. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0958166913000566>. doi: [10.1016/j.copbio.2013.03.002](https://doi.org/10.1016/j.copbio.2013.03.002)
 68. Sauer U, Canonaco F, Heri S, Perrenoud A, Fischer E. The soluble and membrane-bound transhydrogenases UdhA and PntAB have divergent functions in NADPH metabolism of *Escherichia coli*. *The Journal of biological chemistry*. 2004 Feb; 279(8):6613–9. Available from: <http://www.jbc.org/content/279/8/6613.long>. doi: [10.1074/jbc.M311657200](https://doi.org/10.1074/jbc.M311657200) PMID: [14660605](https://pubmed.ncbi.nlm.nih.gov/14660605/)
 69. Shiba Y, Paradise EM, Kirby J, Ro DK, Keasling JD. Engineering of the pyruvate dehydrogenase bypass in *Saccharomyces cerevisiae* for high-level production of isoprenoids. *Metabolic Engineering*. 2007; 9(2):160–168. doi: [10.1016/j.ymben.2006.10.005](https://doi.org/10.1016/j.ymben.2006.10.005) PMID: [17196416](https://pubmed.ncbi.nlm.nih.gov/17196416/)
 70. Shen CR, Lan E I, Dekishima Y, Baez A, Cho KM, Liao JC. Driving Forces Enable High-Titer Anaerobic 1-Butanol Synthesis in *Escherichia coli*. *Applied and environmental microbiology*. 2011; 77(9):2905. doi: [10.1128/AEM.03034-10](https://doi.org/10.1128/AEM.03034-10) PMID: [21398484](https://pubmed.ncbi.nlm.nih.gov/21398484/)
 71. Mapelli V, Olsson L, Nielsen J. Metabolic footprinting in microbiology: methods and applications in functional genomics and biotechnology. *Trends in biotechnology*. 2008 Sep; 26(9):490–7. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/18675480>. doi: [10.1016/j.tibtech.2008.05.008](https://doi.org/10.1016/j.tibtech.2008.05.008) PMID: [18675480](https://pubmed.ncbi.nlm.nih.gov/18675480/)
 72. Krauss M, Schaller S, Borchers S, Findeisen R, Lippert J, Kuepfer L. Integrating Cellular Metabolism into a Multiscale Whole-Body Model. *PLoS Computational Biology*. 2012 Oct; 8(10):e1002750. Available from: <http://dx.plos.org/10.1371/journal.pcbi.1002750>. doi: [10.1371/journal.pcbi.1002750](https://doi.org/10.1371/journal.pcbi.1002750) PMID: [23133351](https://pubmed.ncbi.nlm.nih.gov/23133351/)
 73. Paczia N, Nilgen A, Lehmann T, Gätgens J, Wiechert W, Noack S. Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms. *Microbial cell factories*. 2012 Jan;

- 11:122. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3526501&tool=pmcentrez&rendertype=abstract>. doi: [10.1186/1475-2859-11-122](https://doi.org/10.1186/1475-2859-11-122) PMID: [22963408](https://pubmed.ncbi.nlm.nih.gov/22963408/)
74. Ramakrishna R, Edwards JS, McCulloch A, Palsson BO. Flux-balance analysis of mitochondrial energy metabolism: consequences of systemic stoichiometric constraints. *Am J Physiol Regulatory Integrative Comp Physiol*. 2001 Mar; 280(3):R695–704. Available from: <http://ajpregu.physiology.org/content/280/3/R695.long>.
 75. Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proceedings of the National Academy of Sciences of the United States of America*. 2005 May; 102(21):7695–700. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1140402&tool=pmcentrez&rendertype=abstract>. doi: [10.1073/pnas.0406346102](https://doi.org/10.1073/pnas.0406346102) PMID: [15897462](https://pubmed.ncbi.nlm.nih.gov/15897462/)
 76. Duarte NC, Becker Sa, Jamshidi N, Thiele I, Mo ML, Vo TD, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proceedings of the National Academy of Sciences of the United States of America*. 2007 Feb; 104(6):1777–82. Available from: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1794290&tool=pmcentrez&rendertype=abstract>. doi: [10.1073/pnas.0610772104](https://doi.org/10.1073/pnas.0610772104) PMID: [17267599](https://pubmed.ncbi.nlm.nih.gov/17267599/)
 77. Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, Palsson BO. Multiscale Modeling of Metabolism and Macromolecular Synthesis in *E. coli* and Its Application to the Evolution of Codon Usage. *PLoS one*. 2012 Jan; 7(9):e45635. doi: [10.1371/journal.pone.0045635](https://doi.org/10.1371/journal.pone.0045635) PMID: [23029152](https://pubmed.ncbi.nlm.nih.gov/23029152/)
 78. Grotkjær T, Kesson M, Christensen B, Gombert AK, Nielsen J. Impact of transamination reactions and protein turnover on labeling dynamics in ¹³C-labeling experiments. *Biotechnology and bioengineering*. 2004; 86(2):209–216. doi: [10.1002/bit.20036](https://doi.org/10.1002/bit.20036)
 79. Ravikirithi P, Suthers PF, Maranas CD. Construction of an *E. coli* genome-scale atom mapping model for MFA calculations. *Biotechnology and bioengineering*. 2011 Jun; 108(6):1372–82. doi: [10.1002/bit.23070](https://doi.org/10.1002/bit.23070) PMID: [21328316](https://pubmed.ncbi.nlm.nih.gov/21328316/)
 80. Karr J, Sanghvi J, Macklin D, Gutschow M, Jacobs J, Bolival B, et al. A Whole-Cell Computational Model Predicts Phenotype from Genotype. *Cell*. 2012 Jul; 150(2):389–401. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0092867412007763>. doi: [10.1016/j.cell.2012.05.044](https://doi.org/10.1016/j.cell.2012.05.044) PMID: [22817898](https://pubmed.ncbi.nlm.nih.gov/22817898/)
 81. O'Brien EJ, Lerman JA, Chang RL, Hyduke DR, Palsson BO. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. *Molecular Systems Biology*. 2013 Oct; 9. Available from: <http://dx.doi.org/10.1038/msb.2013.52>.
 82. Di Ventura B, Lemerle C, Michalodimitrakis K, Serrano L. From in vivo to in silico biology and back. *Nature*. 2006 Oct; 443(7111):527–33. doi: [10.1038/nature05127](https://doi.org/10.1038/nature05127) PMID: [17024084](https://pubmed.ncbi.nlm.nih.gov/17024084/)
 83. Schellenberger J, Park JO, Conrad TM, Palsson BO. BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions. *BMC bioinformatics*. 2010 Jan; 11(1):213. doi: [10.1186/1471-2105-11-213](https://doi.org/10.1186/1471-2105-11-213) PMID: [20426874](https://pubmed.ncbi.nlm.nih.gov/20426874/)
 84. Mori E, Furusawa C, Kajihata S, Shirai T, Shimizu H. Evaluating (¹³C) enrichment data of free amino acids for precise metabolic flux analysis. *Biotechnology Journal*. 2011; 6(11):1377–1387. doi: [10.1002/biot.201000446](https://doi.org/10.1002/biot.201000446) PMID: [22069095](https://pubmed.ncbi.nlm.nih.gov/22069095/)
 85. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metabolic Engineering*. 2006; 8(4):324–337. doi: [10.1016/j.ymben.2006.01.004](https://doi.org/10.1016/j.ymben.2006.01.004) PMID: [16631402](https://pubmed.ncbi.nlm.nih.gov/16631402/)
 86. Eric Weisstein. Chi-Squared Distribution;. Available from: <http://mathworld.wolfram.com/Chi-SquaredDistribution.html>.
 87. Sterne JAC, Smith GD, Cox DR. Sifting the evidence—what's wrong with significance tests? *Physical Therapy*. 2001 Aug; 81(8):1464–1469. Available from: <http://ptjournal.apta.org/content/81/8/1464.full>.
 88. Fisher RA. *Statistical methods for research workers*; 1950.
 89. Toya Y, Ishii N, Hirasawa T, Naba M, Hirai K, Sugawara K, et al. Direct measurement of isotopomer of intracellular metabolites using capillary electrophoresis time-of-flight mass spectrometry for efficient metabolic flux analysis. *Journal of chromatography A*. 2007 Aug; 1159(1–2):134–41. Available from: <http://www.sciencedirect.com/science/article/pii/S0021967307006760>. doi: [10.1016/j.chroma.2007.04.011](https://doi.org/10.1016/j.chroma.2007.04.011) PMID: [17462663](https://pubmed.ncbi.nlm.nih.gov/17462663/)
 90. Long, CP, Antoniewicz, MR. Quantifying Biomass Composition by Gas Chromatography/Mass Spectrometry. *Analytical chemistry*. 2014 Sep; Available from: <http://pubs.acs.org/doi/abs/10.1021/ac502734e>.
 91. Usui Y, Hirasawa T, Furusawa C, Shirai T, Yamamoto N, Mori H, et al. Investigating the effects of perturbations to *pgi* and *eno* gene expression on central carbon metabolism in *Escherichia coli* using ¹³C metabolic flux analysis. *Microbial cell factories*. 2012 Jun; 11(1):87. doi: [10.1186/1475-2859-11-87](https://doi.org/10.1186/1475-2859-11-87) PMID: [22721472](https://pubmed.ncbi.nlm.nih.gov/22721472/)