Rethinking ^{13}C-Metabolic Flux Analysis – The Bayesian Way of Flux Inference

Axel Theorell[1], Johann F. Jadebeck[1,2], Wolfgang Wiechert[1,2], Johnjoe McFadden[3,#], Katharina Nöh[1,#,*]

[1] Institute of Bio- and Geosciences, IBG-1: Biotechnology, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany [2] Computational Systems Biotechnology (AVT.CSB), RWTH Aachen University, 52062 Aachen, Germany

[3] Department of Microbial and Cellular Sciences, University of Surrey, GU2 7XH Guildford, United Kingdom

* Corresponding author email: k.noeh@fz-juelich.de

[#] These authors contributed equally to the work.

Journal: Metabolic Engineering

Short title: Bayesian ^{13}C-Metabolic Flux Analysis

Keywords: Bayesian ^{13}C-Metabolic Flux Analysis; Multi-model inference; Model selection uncertainty; Bayesian model averaging; Bidirectional reaction steps; MCMC

Abstract

Metabolic reaction rates (fluxes) play a crucial role in comprehending cellular phenotypes and are essential in areas such as metabolic engineering, biotechnology, and biomedical research. The state-of-the-art technique for estimating fluxes is metabolic flux analysis using isotopic labelling (^{13}C-MFA), which uses a dataset-model combination to determine the fluxes. Bayesian statistical methods are gaining popularity in the field of life sciences, but the use of ^{13}C-MFA is still dominated by conventional best-fit approaches. The slow take-up of Bayesian approaches is, at least partly, due to the unfamiliarity of Bayesian methods to metabolic engineering researchers. To address this unfamiliarity, we here outline similarities and differences between the two approaches and highlight particular advantages of the Bayesian way of flux analysis. With a real-life example, re-analysing a moderately informative labelling dataset of *E. coli,* we identify situations in which Bayesian methods are advantageous and more informative, pointing to potential pitfalls of current ^{13}C-MFA evaluation approaches. We propose the use of Bayesian model averaging (BMA) for flux inference as a means of overcoming the problem of model uncertainty through its tendency to assign low probabilities to both, models that are unsupported by data, and models that are overly complex. In this capacity, BMA resembles a tempered Ockham's razor. With the tempered razor as a guide, BMA-based ^{13}C-MFA alleviates the problem of model selection uncertainty and is thereby capable of becoming a game changer for metabolic engineering by uncovering new insights and inspiring novel approaches.

1. Introduction

Intracellular metabolic reaction rates (fluxes) provide a quantitative and experimentally-anchored account of the cellular physiology that underpins phenotypes (Nielsen, 2003; Sauer, 2006). Once captured, intracellular fluxes are readily interpretable as flux maps (analogues of traffic congestion maps) that highlight metabolic bottlenecks (Stephanopoulos G. et al., 1998), pinpoint the effectiveness of genetic or bioprocess optimization efforts (Becker et al., 2011; Das et al., 2020), inform how cells regulate their metabolism (Kochanowski et al., 2021), reveal energy leaks (Zhao et al., 2012), and allow deriving otherwise indeterminable cellular parameters (Zelle et al., 2021). Metabolic fluxes also play an important role in the biomedical sector, e.g. to detect metabolic “reprogramming” induced by disorders, viruses, pathogens or other diseases as well as to study the mode of action of drugs (Borah et al., 2019; Lagziel et al., 2019; Munger et al., 2008; Murphy et al., 2013). Knowledge of metabolic fluxes has also far-reaching implications beyond flux maps as they narrow plausible intracellular metabolite concentration ranges (Beard et al., 2002; Wiechert, 2007; Xu et al., 2020), calibrate kinetic metabolic models (Foster et al., 2019), are essential for reconciling enzyme kinetic parameters from databases (Liebermeister and Noor, 2021), and provide a solid basis for machine learning approaches to study metabolism (Wu et al., 2016). Fluxes also provide the principal means by which the results of metabolic engineering can be assessed. Consequently, the accurate system-wide estimation of metabolic fluxes, which includes the reliable quantification of their uncertainties, is a vitally important technique for all basic as well as applied biosciences, such as metabolic engineering.

The state-of-the-art method for inferring metabolic fluxes is ^{13}C metabolic flux analysis (MFA) (Wiechert, 2001). In ^{13}C-MFA, data from labelling experiments conducted under metabolic (quasi-)steady state conditions are analysed in the context of a biochemical reaction model to generate detailed metabolic flux maps (Long and Antoniewicz, 2019; Niedenführ et al., 2015). Two variants are distinguished based on whether the labelling information consists of labelling time-courses (isotopically nonstationary, short INST, ^{13}C-MFA) or is acquired after reaching an isotopic steady-state in the target intermediates (steady-state ^{13}C-MFA) (Wiechert and Nöh, 2021). Regardless of which variant is used, the key feature of ^{13}C-MFA is that it not only provides an estimate of intracellular fluxes, but it also equips each flux value (and pool size in case of INST) with an uncertainty measure, either in terms of confidence or credible intervals (Theorell et al., 2017). Essentially, these intervals articulate how the inevitable experimental noise in the data propagates through the biochemical network, thereby indicating the degree of caution that is required in interpreting the flux estimates. The confidence limits of the inferred fluxes can be narrowed by increasing the quantity and/or quality of the data, e.g., by conducting extensive experimental campaigns (Leighty and Antoniewicz, 2013), or targeted selection of a (co-)tracer mixture and measurement technology (Borah Slater et al., 2023; Nöh et al., 2018). However, theoretical investigations (Kappelmann et al., 2016) as well as extensive experimental exercises (Crown et al., 2015) have demonstrated that, in any realistic study, several central metabolic fluxes will remain unresolved.

^{13}C-MFA models are built with knowledge of mass balances, formulated upon genetic and biomolecular knowledge that formalizes the incorporation of isotopically labelled substrates into the metabolic pathways of the target organism. The scope of the models – the extent of the metabolic reaction set that is to be modelled – is roughly shaped by the observed metabolites and their ability to provide information on the pathways of primary interest. Typical ^{13}C-MFA models inform on central carbon metabolism and amino acid biosynthesis, with a recent trend to increasing the model scope (McCloskey et al., 2016a), as reviewed in (Hendry et al., 2020), by more application of more advanced analytical techniques (Kappelmann et al., 2019; McCloskey et al., 2016b; Zheng et al., 2024) and performant simulation tools (Quek et al., 2009; Rahim et al., 2022; Sokol et al., 2012; Weitzel et al., 2013; Wu et al., 2023; Young, 2014).

Concerning uncertainty quantification for the metabolic fluxes, the ^{13}C-MFA field has been shaped by the statistical "frequentist" viewpoint (Antoniewicz et al., 2006; Wiechert et al., 1999). Bayesian methods were first mentioned only in 2006 (Kadirkamanathan et al., 2006), and are only rarely used in practice, even though Bayesian inference provides a consistent framework for updating prior knowledge about fluxes with new evidence, a situation that is frequently encountered in metabolic engineering. One key difference between the Bayesian and frequentist approaches is that the latter's maximum likelihood estimator gives the best fit, a so-called point estimate, for the true, but unknown flux vector (the common term "flux distribution" is ambivalent, so we here use the term "flux vector" to indicate that the outcome is a concrete flux map). In the Bayesian paradigm, in contrast, a flux vector is considered a multivariate random variable which is assigned a probability density. Thus, the unknown fluxes are interpreted as inherently uncertain quantities. The final outcome of a Bayesian ^{13}C-MFA is then represented by the so-called joint flux posterior probability distribution, which, once determined, unlocks insights into flux probabilities and correlations that go beyond the single point estimate derived using flux fitting with subsequent statistical analysis. The flux uncertainty obtained from flux posterior-based credible intervals by marginalization (integration) of the joint posterior probability distribution is straightforward to interpret, i.e., the true flux value is with, say, 95% probability contained in the interval, an interpretation that is often wrongly attributed to frequentist confidence intervals (Ellison, 2004; Morey et al., 2016). Moreover, the derivation of confidence intervals is brittle in the sense of credibility of the confidence limits. This is because the confidence intervals are sensitive to the algorithm with which they are determined (Theorell et al., 2017). For introductory texts on Bayesian concepts, we refer to the excellent textbook by Gelman et al., (2013).

We here demonstrate that the statistical rigour of the Bayesian machinery is superior to frequentist methods of ^{13}C-MFA in its ability to map data uncertainty into flux uncertainty. This is because the Bayesian framework empowers us to, besides data noise, also take the uncertainty in the model used for flux inference into account. However, the Bayesian nomenclature can quickly become overwhelming for non-statisticians. To explain the advantages of the Bayesian way of flux analyses to the metabolic engineer who has grown up in the frequentist world, we give the core principles without statistical overload. We showcase the Bayesian machinery in action, with a worked example to highlight the “Why”, “When”, and “How” of this modern approach to metabolic flux inference, by deriving a novel solution to a long-standing ^{13}C-MFA problem.

2. Theory/calculation

2.1. Bayesian ^{13}C-MFA using single models

Phrased in the language of Bayesian statistics, the goal of metabolic flux analysis is to determine the posterior probability distribution of net and exchange fluxes θ from a labelling data set D by means of an atom transition model $\mathcal{M}$[*] (Wiechert and de

[*] With the term *model* we mean a system of mathematical equations, defined over a parameter space and with an image in the observation space.

Graaf, 1997)). Here we denote the flux vector $\theta_{\mathcal{M}} = \left(\theta_{\mathcal{M}}^{net}, \theta_{\mathcal{M}}^{xch}\right)$, where the subscript indicates the parameters' model affinity. Throughout this section we consider the model structure fixed, as this is standard in ^{13}C-MFA. In the Bayesian world, any flux is considered a random variable equipped with a credibility represented by a probability. Considering all possible flux values gives a probability distribution. The central desired quantity of ^{13}C-MFA is, thus, the joint flux posterior probability distribution, denoted $p(\theta_{\mathcal{M}}|\mathcal{M}, D)$ ("the probability of the flux vector $v_{\mathcal{M}}$ given model $\mathcal{M}$ and data D", short *posterior*), which assigns a probability to each flux constellation taking into account the data. If that posterior probability distribution is narrow, there are only few flux vectors considered credible and there is low uncertainty about their values. On the contrary, if the posterior distribution is wide, there are many flux values with a weak credibility, thereby representing high uncertainty. To derive the posterior, Bayes theorem is employed (Bayes and R, 1763):

$$p(\theta_{\mathcal{M}}|\mathcal{M}, D) = \frac{p(D|\theta_{\mathcal{M}}, M) \cdot p(\theta_{\mathcal{M}}|\mathcal{M})}{p(\mathcal{M}, D)} \qquad (1)$$

which expands $p(\theta_{\mathcal{M}}|\mathcal{M}, D)$ into three ingredients: the prior probability distribution $p(\theta_{\mathcal{M}}|\mathcal{M})$, (short, the *prior*), the likelihood $p(D|\mathcal{M}, \theta_{\mathcal{M}})$, and the normalization constant $p(\mathcal{M}, D)$, that transforms the (relative) flux posterior probability distribution (enumerator in Eq. (1)) into a probability density with total probability of one.

The normalizing constant is called *evidence* that expresses the probability of the model-data combination, which is calculated by integrating over the feasible flux values

$$p(\mathcal{M}, D) = \int p(D|\theta_{\mathcal{M}}, \mathcal{M}) \cdot p(\theta_{\mathcal{M}}|\mathcal{M}) \, d\theta_{\mathcal{M}} \qquad (2)$$

The likelihood, $p(D|\theta_{\mathcal{M}}, M)$, is the probability that we observe the data D given model $\mathcal{M}$ with flux vector $\theta_{\mathcal{M}}$. The likelihood is closely related to the frequentist world, where it reflects the maximum likelihood estimation (or optimization) landscape, from which the best fit (maximum agreement between measured and model-predicted data) is derived. From the mode of the posterior $p(\theta_{\mathcal{M}}|\mathcal{M}, D)$ the particular flux vector that best explains the observed data set is the so-called maximum a posteriori (MAP) point estimate. In particular, the MAP coincides with the best maximum likelihood fit, in case of uninformative priors (Theorell et al., 2017).

The prior, $p(\theta_{\mathcal{M}}|\mathcal{M})$, expresses the knowledge about the model-specific fluxes $\theta_{\mathcal{M}}$ that is available before the labelling experiment is made. The more knowledge is encoded in the prior, the larger is its influence on the result. Since priors are formally unknown to frequentists, their formulation often gives rise to confusion. So how do we specify flux priors for a ^{13}C-MFA? In the common scenario that we study an organism under conditions that have already been studied, some knowledge about flux value ranges and a notion about their (un)certainty exists. This knowledge, formalized as an informed prior probability distribution, enters Eq. (1). Actually the same happens in traditional ^{13}C-MFA: the modeller specifies available knowledge by adding constraints, e.g., assumptions on reaction reversibility or the range of possible flux values. In a scenario, where less knowledge on the fluxes is available or should be used, any theoretically possible flux constellation is considered equally likely. In any way, subtly, the reaction stoichiometry imposes non-trivial flux boundaries that, together with upper and lower flux limits, renders the flux prior a density and makes even an uniform flux prior informative (Jadebeck et al., 2021; Theorell et al., 2022).

To exemplify Bayesian ^{13}C-MFA, we revisit a published study that features *E. coli* (Zamboni et al., 2009). In the study, a labelling experiment with a tracer mixture of 20% [U-^{13}C]-glucose and 80% naturally labelled glucose was conducted in a continuous cultivation with the MG1655 wild type at a growth rate of 0.12 h^{-1}.In total

192 independent labelling measurements were generated by mass spectrometry of 11 amino acids, as well as growth rate and glucose uptake rate measurements (Supplementary Table S1, the glucose tracer mixture was fixed). The original network model, hitherto termed $\mathcal{M}_0$, used for data analysis covers central carbon metabolism and biosynthesis routes (Figure 1), therewith representing a typical ^{13}C-MFA example. The model comprises 66 reactions connecting 37 metabolites. Of the 66 reactions, 49 were considered uni-directional and 17 bidirectional, as a consequence of assumptions imposed by thermodynamic properties. Further expert knowledge was implemented, such as the activity of malic enzyme (*ana3*) in gluconeogenetic direction or the exclusion of the Entner-Doudoroff pathway, known to be inactive under the studied conditions. Therewith, reaction stoichiometry leaves 10 adjustable net fluxes and 17 adjustable exchange fluxes. Sampling the flux solution space uniformly then reveals the flux prior introduced by the model formulation $\mathcal{M}_0$: while the marginal net flux priors exhibit informative shapes (Supplementary Fig. S5), exchange flux priors obviously remain diffuse within their boundaries. The task of single-model ^{13}C-MFA is to infer all 27 free flux parameters from the 193 independent measurements.

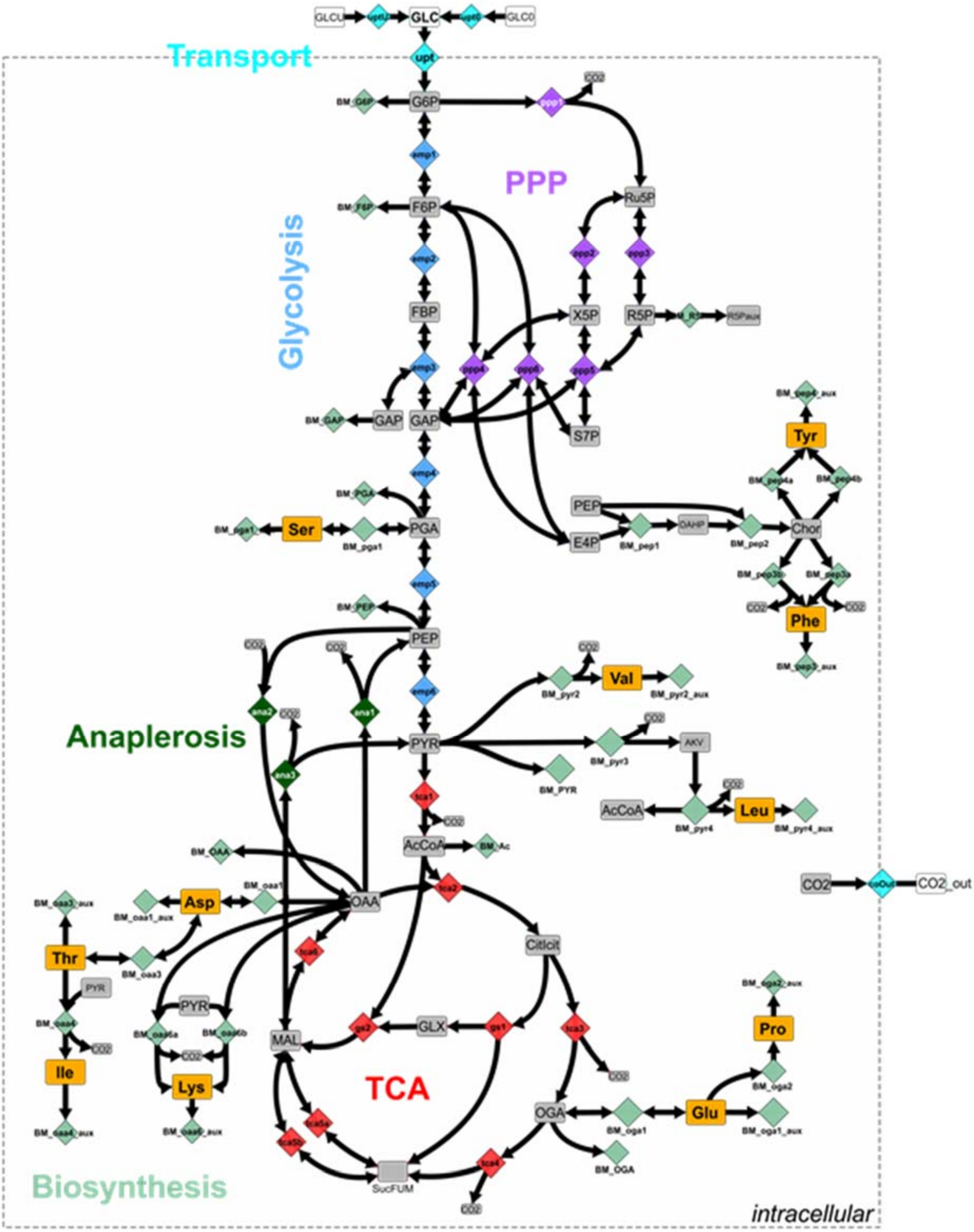

***Figure 1: Metabolic network of the central carbon metabolism of E. coli representing the $\mathcal{M}_0$ model taken from* Zamboni et al., (2009)**. *Double headed arrows indicate bidirectional reactions (diamonds), unidirectional reactions are shown single headed. Metabolic pathways are color-coded with the 11 measured amino acids (rounded rectangles) indicated in orange. The model $\mathcal{M}_0$ was formulated with the model editor Omix (v1.9.34) (Droste et al., 2013) and exported to FluxML format (v1.3), an universal language for*

specification of ^{13}C-MFA models (Beyß et al., 2019). The full FluxML model specification, including atom mapping and measurements, is given in the Supplementary Data.

Bayesian flux estimation with the $\mathcal{M}_0$ model was carried out as described previously (Theorell et al., 2017). In short, the high-performance simulator 13CFLUX2 (v2.0) (Weitzel et al., 2013) was used for likelihood calculation. Flux prior and posterior probability distributions given by Eq. (1) were calculated using Markov chain Monte Carlo (MCMC) sampling , using the highly optimized polytope sampling library HOPS (Jadebeck et al., 2021), after rounding (Theorell et al., 2022) and thinning (Jadebeck et al., 2023). The posterior distributions were checked for convergence (flux-wise potential scale reduction factor < 1.01 and effective sample sizes (ESS) >> 1,000, see Supplementary Table S.2 and Supplementary Methods for more details). Figure 2 shows the inference results for 9 out of the 10 free net fluxes selected from different metabolic pathways of *E. coli*, along with their marginal 95% credibility intervals. For reference, the values reported in the original study (Zamboni et al., 2009), using nonlinear flux fitting in combination with linearized error propagation (Wiechert et al., 1997), using 13CFLUX (Wiechert et al., 2001), are also shown.

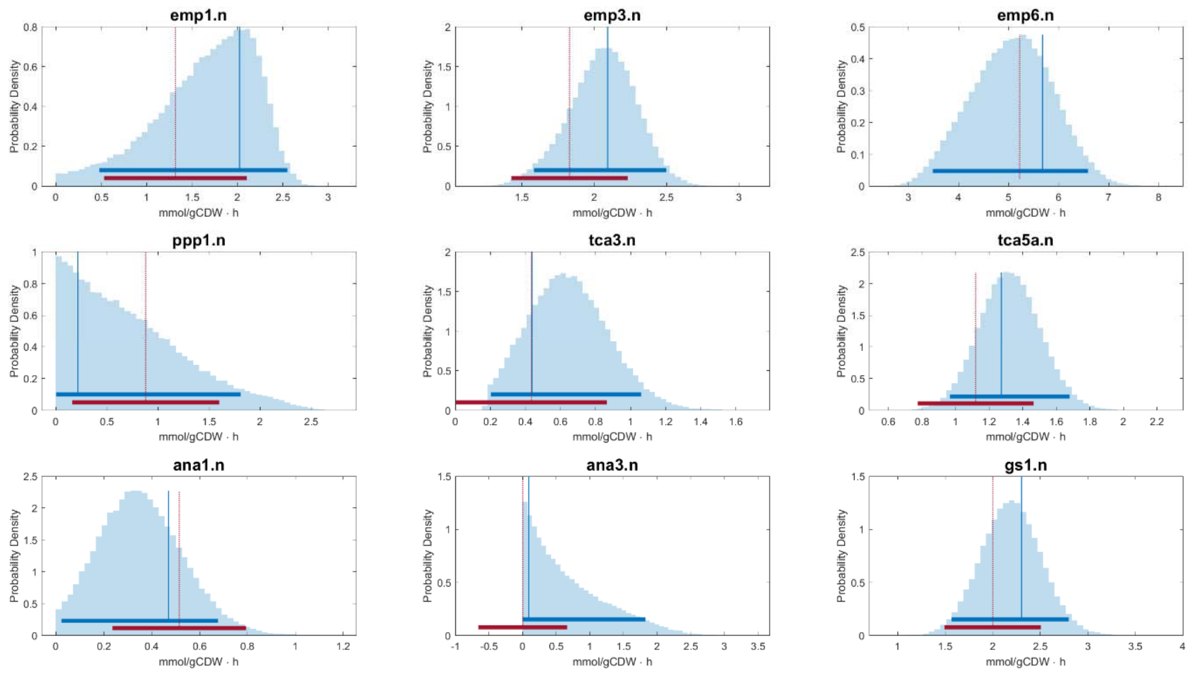

Figure 2: Single-model flux inferences using $\mathcal{M}_0$. *Marginal posterior probability distributions for net fluxes of reactions located in different metabolic pathways are shown: emp1.n, emp3.n, emp6.n (glycolysis), ppp1.n (pentose phosphate pathway), tca3.n, tca5a.n (tricarboxylic acid cycle), ana1.n, ana3.n (anaplerosis), and gs1.n (glyoxylate shunt). Marginal flux posterior probability distributions derived from Eq. (1) with the MAP and 95% credibility intervals are shown in blue, best fit values and 95% confidence intervals reported by Zamboni et al., (2009) in red. In the original study 21 of the 27 free fluxes were set constant after fitting before performing the statistical analysis, whereas the Bayesian analysis was carried out using model $\mathcal{M}_0$ with 27 free fluxes. Therefore, emp6.n (pyruvate kinase) lacks a confidence interval, as the flux was fixed to a value of 5.22 mmol/(g_{CDW} h). Credibility intervals are bounded by definition to not include values beyond the feasible flux range, unlike confidence intervals produced by linearized statistics (cf. tca3.n, ana3.n). Marginal distributions for the remaining net fluxes and their 95% credibility intervals are given in Supplementary Fig. S6.*

3. Results and discussion

3.1. Bidirectional reaction steps give rise to model selection uncertainty

Models are never perfect, but always approximate specific aspects of the system under study. For a model to be useful when operated within a particular scientific context such as flux estimation, two criteria have to be fulfilled: the model scope has to be general enough to address the questions of interest, and yet it must be verifiable experimentally. In biochemical network modelling, these two criteria are not sufficient to determine a unique model formulation (Hangos et al., 2014; Haunschild et al., 2005). In this situation, the choice of the particular model to be used can profoundly influence the conclusions drawn from the analysis of the particular dataset. ^{13}C-MFA is no exception, despite its well-defined biochemical groundings and best-practice to collect experimental evidence for major influencing factors (Zamboni et al., 2009). Modern high-throughput technologies (Heux et al., 2017) open the door to retain a subset of the data for validation purposes and model checking (Gelman et al., 2020), therewith offering rigorous tools to (in)validate candidate models, thereby curbing against model misspecification, as exemplified by Sundqvist *et al* (2022). Nonetheless, even with all these safeguards in place and within an agreed and validated model scope, there are, as in any modelling effort, specific design decisions involved, each of which has potential consequences on flux estimates.

An ever-present question in ^{13}C-MFA is whether reversible reactions[†] operate uni- or bidirectionally under *in vivo* conditions. The answer to this question determines much of the mathematical structure of the labelling systems (Wiechert and de Graaf, 1997); but, in contrast to pathway enrichment or enzyme activity testing, *in vivo* reaction bidirectionality is experimentally hardly accessible. The second law of thermodynamics provides pointers that may identify a reaction as likely to be unidirectional (Wiechert, 2007); but the vast majority of enzymatic reactions operates near thermodynamic equilibrium such that bidirectional labelling exchange can

[†] Essentially all biochemical reactions are reversible but when a strong thermodynamic driving force drives a reaction in a single reaction direction then it is said unidirectional, otherwise bidirectional.

happen even if the net flux is small. For example, even in the glycolytic pathway of extremely well-studied organisms such as *E. coli such* “surprises” can happen (Long et al., 2017), proving previous model specifications wrong. Isotope labelling data can be informative about reaction bidirectionality, as was demonstrated in the early days of ^{13}C-MFA (Wiechert and de Graaf, 1997), both *in silico* (Follstad and Stephanopoulos, 1998), and *in vivo* (Marx et al., 1996). Since the reaction bidirectionality setting is a crucial ingredient for constructing the model that is used to analyse the labelling data, the modeller here has to decide on a model formulation *before* evaluating the data set at hand despite epistemic uncertainty.

While for describing the labelling flow through a unidirectional reaction one flux parameter is sufficient, in case of a bidirectional reaction two parameters are required. Commonly net and exchange fluxes are used to describe the fluxes of a reversible reaction step (Wiechert and de Graaf, 1997), where a net flux quantifies the net transport of label between substrates and products of a reaction, while the exchange flux is the quantity of the labelling flow that goes in both, forward and backward direction. It has been argued that in the case of a reaction could be bidirectional, it is safest to allow the data to provide evidence of whether it actually is uni- or bidirectional under the tested conditions. Exchange fluxes, however, are deemed to be not well-identifiable (Wiechert and Nöh, 2021), making it desirable to reduce their number to reduce the risk of describing the noise in the data.

3.2. Single bidirectionality settings ignore model uncertainty in ^{13}C-MFA

Figure 3 visualises the bidirectionality settings of the original model $\mathcal{M}_0$ and two alternative models, $\mathcal{M}_1$ and $\mathcal{M}_2$, derived by setting some of the bidirectional reactions unidirectional. The two models $\mathcal{M}_1$ and $\mathcal{M}_2$ are simpler in that they have only 3 exchange fluxes each ($\mathcal{M}_0$ has 17 exchange parameters), whereas all models share the same set of net fluxes (10). The two models $\mathcal{M}_1$ and $\mathcal{M}_2$ were selected heuristically by inspecting flux sensitivities and correlations and ensuring that no flux or flux combination essential to achieve a good fit to the data were eliminated. All models were given the same flux bounds as $\mathcal{M}_0$, except for their bidirectionalities (Supplementary Table S1). The weighted sum of squared residuals (WSSR) obtained by maximum likelihood estimation are 137.8 and 137.9 for $\mathcal{M}_1$ and $\mathcal{M}_2$, respectively; values that are very similar to the value of the original model $\mathcal{M}_0$ (132.0).

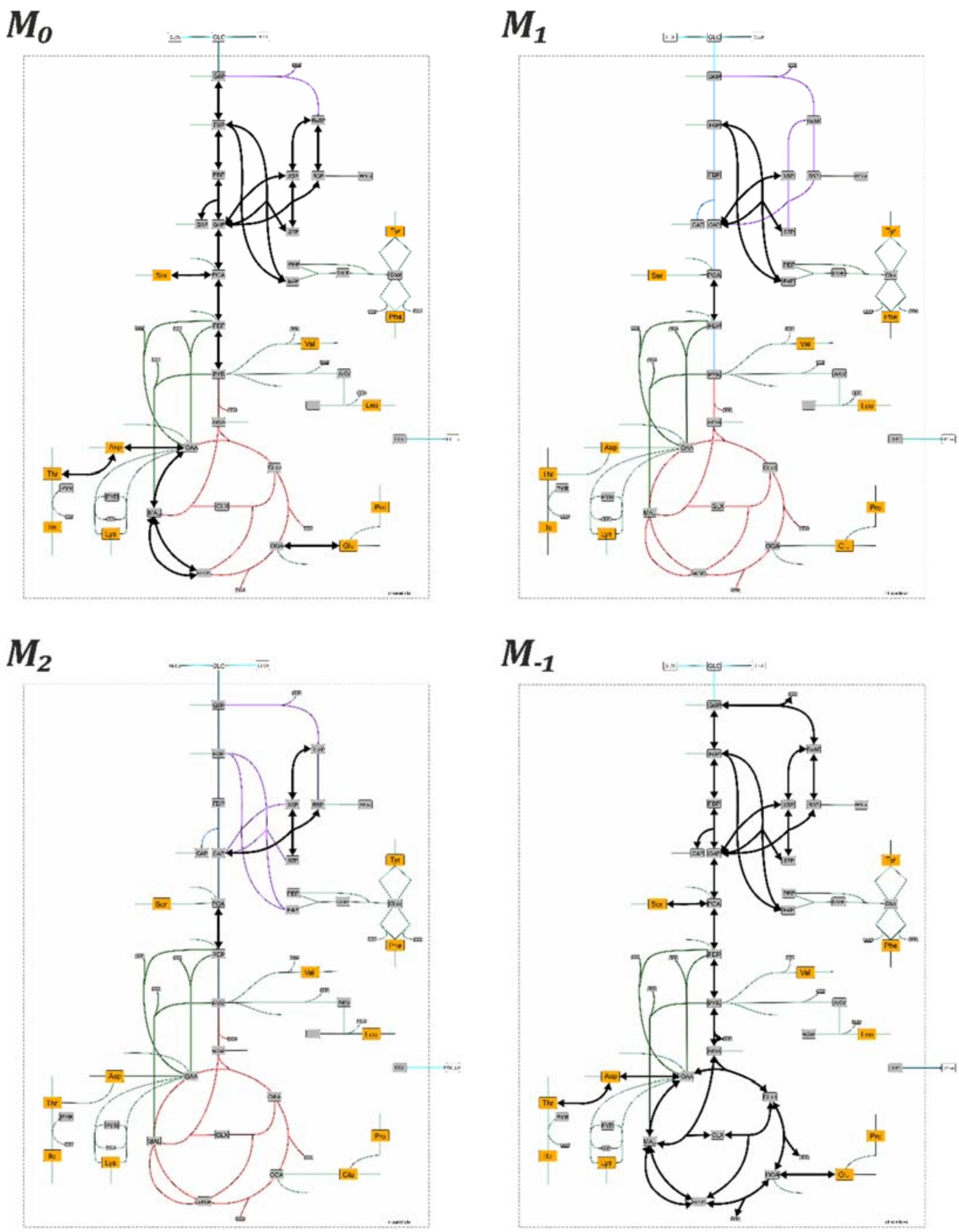

Figure 3: Alternate bidirectional reaction settings of the original model $\mathcal{M}_0$. *Bidirectional reactions are highlighted with black double headed arrows. All models share the same 10 net fluxes, while the exchange fluxes are specific to the models: 17 for $\mathcal{M}_0$ (WSSR=132.0), 3 for $\mathcal{M}_1$ (WSSR 137.8) and $\mathcal{M}_2$ (WSSR 137.9), as well as 24 for $\mathcal{M}_{-1}$ (WSSR 131.3). Specifications of the models are available in the Supplementary Data.*

That the bidirectionality setting indeed has dramatic consequences on practical net flux inference becomes evident in Figure 4, where the net flux posterior probability densities for the 9 selected net fluxes *E. coli* (remaining net flux distributions are provided in Supplementary Fig. S7), derived by Eq. (1), are shown for the three different models. The flux posterior for each of the models differs strongly in several net fluxes, in particular for the simple models $\mathcal{M}_1$ and $\mathcal{M}_2$. The comparison shows that there is also little consensus amongst the minimal models and $\mathcal{M}_0$ in the upper glycolysis (*emp1.n*) and the pentose phosphate pathway (*ppp1.n* – *ppp6.n*). The results suggest that the minimal models yield inferences with smaller variances compared to the original model, as expected, but also suffer from higher biases (*emp1.n, ppp1.n* – *ppp6.n*), induced by setting too many reactions unidirectional.

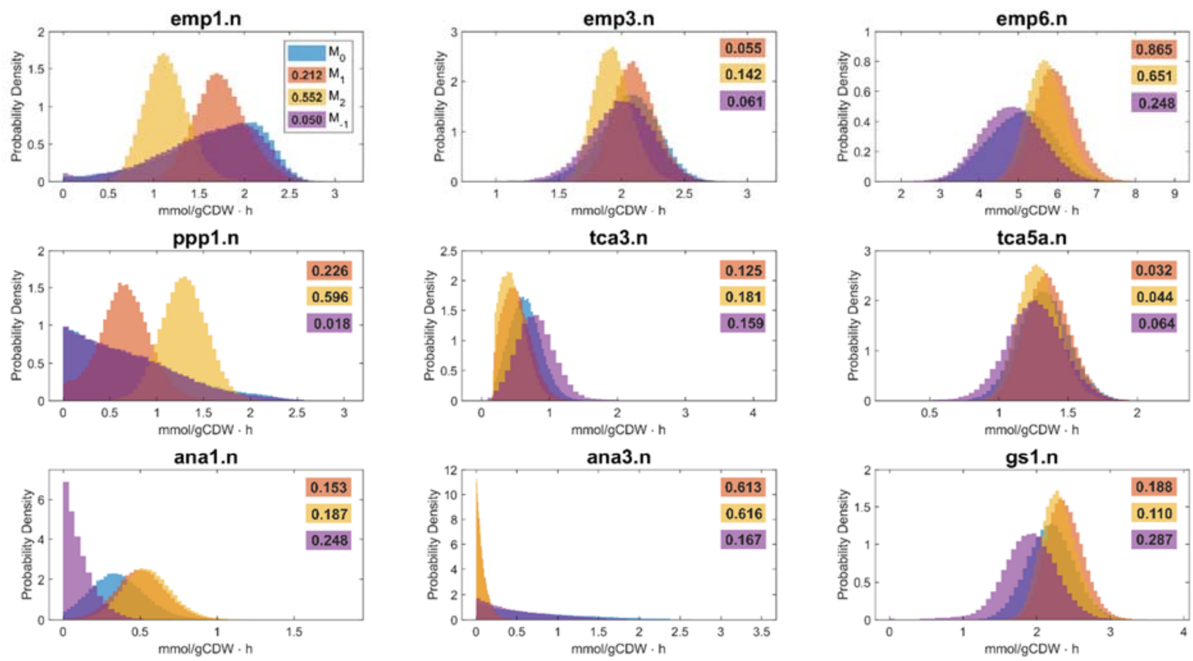

Figure 4: Flux posterior probability distributions derived from single-model inference. *Net fluxes of reactions located in different metabolic pathways are selected: emp1.n, emp3.n, emp6.n (glycolysis), ppp1.n (pentose phosphate pathway), tca3.n, tca5a.n (tricarboxylic acid cycle), ana1.n, ana3.n (anaplerosis), and gs1.n (glyoxylate shunt), derived with the original model $\mathcal{M}_0$ (blue), the simple models $\mathcal{M}_1, \mathcal{M}_2$ (rose, yellow), and $\mathcal{M}_{-1}$ (plum) a super-model of $\mathcal{M}_0$.The parsimonious models $\mathcal{M}_1$ and $\mathcal{M}_2$ yield contradictory inferences and, unsurprisingly, grossly underestimate flux uncertainty, whereas the models $\mathcal{M}_0$ and $\mathcal{M}_{-1}$ show higher similarity for most, but not all net fluxes. Marginal distributions for the remaining net fluxes and their 99% credible intervals are given in Supplementary Figs. S7 and S12. Numbers give the 1-Wasserstein distance (in mmol/g_{CDW}/h), a metric that quantifies the discrepancy between the posterior distributions related to $\mathcal{M}_0$. Values close to 0 indicate good agreement. Wasserstein distances were calculated using SciPy (1.5.4). A pairwise model comparison across all net fluxes revealing most dissimilar net flux inferences is given in Supplementary Figs. S11.*

That only 17 fluxes are considered bidirectional in the original model is the consequence of bidirectionality assumptions supported by expert knowledge. To test a model variant that has more bidirectional reactions than the original model, we relaxed some of these assumptions and considered 24 reactions to be potentially bidirectional (the relaxed reactions are indicated in Figure 6 in red). The relaxed model is hitherto denoted $\mathcal{M}_{-1}$ (Figure 3). Flux posterior distributions derived with $\mathcal{M}_{-1}$ (Figure 4 in plum) are comparable with those of model $\mathcal{M}_0$ in many, but not all net fluxes. Most prominent discrepancies occur in the anaplerotic section (*ana1.n*), where the credibility intervals produced with the complex model $\mathcal{M}_{-1}$ are smaller than those produced by the much more constrained original model $\mathcal{M}_0$, despite using the same data (Supplementary Fig. S12). An explanation for this phenomenon is that the complex model $\mathcal{M}_{-1}$ starts to fit the noise in the data, which renders the credibility limits less reliable. This shows that using an "all-inclusive super" model (a model with all potentially reversible reactions set bidirectional) does not secure reliable flux inference.

Summarizing, in ^{13}C-MFA a multitude of equally well-performing model candidates exists for all real-world problems. The choice of whether to set reactions as either uni- or bidirectional is crucial to the flux solution. The frequently used strategy of setting the reactions with uncertain bidirectionality as bidirectional is not a safe bet in terms of inferential reliability, because it can produce different MAPs and cause unpredictability in over- or under-estimation of credibility intervals. The question is how to determine a useful model from the set of possible models that gives reliable flux estimates and confidence limits?

3.3. Bayesian model selection

To identify the "best" model among many candidates, a variety of information-theoretical model selection criteria has been formalized that trade-off fit quality (how well the model explains the data) and model simplicity (often related to the number of independent model parameters) (Burnham and Anderson, 2002). A review of these selection criteria and their application in systems biology is provided by Kirk *et al*. (2013). The justification that among all model candidates with the same fit quality, the simplest one is favoured (Ockham's Razor) rests on probabilistic grounds: Simple models, with few adjustable parameters, are able to adjust to fit only a narrow range of experimental outcomes, compared to complex models with many adjustable parameters (MacKay, 2003). Therefore, in the common scenario where both a simple and a complex model can fit the data equally well, the simple model is suggested to be more credible, since it is less likely to fit the data by chance (Jefferys and Berger, 1992; McFadden, 2023).

3.3.1. Comparing ^{13}C-MFA models using likelihood ratios

The guiding principle of simplicity (or parsimony) is embodied in Bayesian model selection (MacKay, 2003); see Supplementary Methods for an educational example. In the context of potentially bidirectional reaction steps, simplicity is linked to the number of parameters in ^{13}C-MFA models. So, what is the outcome of Bayesian model selection when we compare our (complex) reference model $\mathcal{M}_0$ (27 free flux parameters) with the simpler models $\mathcal{M}_1$ or $\mathcal{M}_2$ (13 free flux parameters)? For this, we need to determine the posterior probability of the models $\mathcal{M}_i$ in view of the data D, in short the model posterior $p(\mathcal{M}_i|D)$. To derive these quantities, we first expand $p(\mathcal{M}_i|D)$ using Bayes theorem:

$$p(\mathcal{M}_i|D) = \frac{p(D|\mathcal{M}_i) \cdot p(\mathcal{M}_i)}{p(D)} \qquad (3)$$

To assess, which particular model formulation from a set of two ^{13}C-MFA model candidates, say $\mathcal{M}_0$ and $\mathcal{M}_1$, is more likely, using Eq. (3), we then compute the ratio of model posteriors:

$$\frac{p(\mathcal{M}_1|D)}{p(\mathcal{M}_0|D)} = \frac{p(D|\mathcal{M}_1) \cdot p(\mathcal{M}_1)}{p(D|\mathcal{M}_0) \cdot p(\mathcal{M}_0)} \tag{4}$$

where $p(D)$ cancels out. The likelihood of a model $\mathcal{M}_i$, $p(D|\mathcal{M}_i)$, is derived by averaging out (marginalizing) the influence of its flux parameters, $\theta_{\mathcal{M}i}$, specific to $\mathcal{M}_i$:

$$p(D|\mathcal{M}_i) = \int p\left(D\middle|\theta_{\mathcal{M}_i}, \mathcal{M}_i\right) \cdot p\left(\theta_{\mathcal{M}_i}\middle|\mathcal{M}_i\right) d\theta_{\mathcal{M}_i} \tag{5}$$

Inserting Eq. (5) into Eq. (4) gives the marginal probability likelihood ratio for the comparison of two models $\mathcal{M}_0$ and $\mathcal{M}_1$:

$$\frac{p(\mathcal{M}_1|D)}{p(\mathcal{M}_0|D)} = \frac{\int p\left(D\middle|\theta_{\mathcal{M}_1}, \mathcal{M}_1\right) \cdot p\left(\theta_{\mathcal{M}_1}\middle|\mathcal{M}_1\right) d\theta_{\mathcal{M}_1}}{\int p\left(D\middle|\theta_{\mathcal{M}_0}, \mathcal{M}_0\right) \cdot p\left(\theta_{\mathcal{M}_0}\middle|\mathcal{M}_0\right) d\theta_{\mathcal{M}_0}} \cdot \frac{p(\mathcal{M}_1)}{p(M_0)} \tag{6}$$

When, in the absence of a preference, as equal prior probability is assigned to each of the two models, i.e. $p(\mathcal{M}_0) = p(\mathcal{M}_1)$, the posterior odds in Eq. (6) reduces to the so-called Bayes factor (Kass and Raftery, 1995; Wasserman, 2000), the ratio of the marginal likelihoods (or evidences) of the two model hypotheses. A Bayes factor of 10, for instance, states that the data D are considered 10 times more likely to be produced from model $\mathcal{M}_0$ rather than from model $\mathcal{M}_1$. Hence, Eq. (6) is a rigorous means for model comparison, providing a measure for the evidence, where the model with the higher probability is preferred (MacKay, 2003; Pullen and Morris, 2014).

To answer our question, which of the three models is supported by the data most, we calculated the posterior probabilities of the models $\mathcal{M}_1$ and $\mathcal{M}_2$ relative to the original model $\mathcal{M}_0$ without preference to any of the three models (i.e., the prior odds ratios are 1.0 in Eq. (6)). Practically, Bayesian evidence approximation is performed using nested sampling (Skilling, 2006), precisely through Diffusive Nested Sampling (Brewer et al., 2011), a performant MCMC technique implemented in the software DNest4 (v4) (Brewer and Foreman-Mackey, 2018). The calculations yields:

$$\frac{p(\mathcal{M}_1|D)}{p(\mathcal{M}_0|D)} = \frac{p(D|\mathcal{M}_1)}{p(D|\mathcal{M}_0)} \approx 4.8 \cdot 10^2$$
$$\frac{p(\mathcal{M}_2|D)}{p(\mathcal{M}_0|D)} = \frac{p(D|\mathcal{M}_2)}{p(D|\mathcal{M}_0)} \approx 5.1 \cdot 10^3 \qquad (7)$$

meaning that both minimal models, $\mathcal{M}_1$ and $\mathcal{M}_2$, are at least two orders of magnitude more probable than the published model $\mathcal{M}_0$. This is a consequence of the action of Ockham's razor that is embodied in Bayesian statistics: the original model $\mathcal{M}_0$ accommodates a larger range of data with its 14 additional parameters than the simpler models; but because the simpler models explain the data almost as well as the original model, they are considered to be more likely.

3.3.2. Model probability

In the *E. coli* example, it is important to recognize that, although the $\mathcal{M}_2$ model is more likely than $\mathcal{M}_0$ and $\mathcal{M}_1$,we have derived our conclusion by comparing the posterior probabilities of only three models rather than accounting for *all* possible models. In particular, even though $\mathcal{M}_2$ is ten times more likely than $\mathcal{M}_1$, its *absolute* probability, may nevertheless be very small, taking all of the un-investigated model candidates into account. Indeed, recalling the difference of 14 bidirectionalities between the simple and the original model, it is very likely that there exists a large number of models that are simpler than the original model, and have a similar likelihood, but different flux posterior distributions (Figure 4). In that situation, selecting any single model with a high likelihood, can nevertheless seriously underestimate the flux uncertainty.

To see this, we now formulate the probability of a single model, as in Eq. (3), by relating its probability to the probability of the totality of all possible models (henceforth, *absolute model probability*). This requires us to assess the probability of the data under any possible bidirectionality hypothesis, as represented by the normalizing constant, the model evidence $p(D)$ in Eq. (3). Knowing that the sum of the probabilities of all, say N, possible models has to equal one, we can rewrite Eq. (3) to give the absolute probability density of a model $\mathcal{M}_i$ from a set of alternative models:

$$p(\mathcal{M}_i|D) = \frac{p(D|\mathcal{M}_i) \cdot p(\mathcal{M}_i)}{p(D)} = \frac{p(D|\mathcal{M}_i) \cdot p(\mathcal{M}_i)}{\sum_{j=1}^{N} p\left(D\middle|\mathcal{M}_j\right) \cdot p\left(\mathcal{M}_j\right)} \qquad (8)$$

We here consider all alternate model formulations that lead to a combinatorial set of models, precisely $N = 2^{n_{bi}}$ possible model structures, where n_{bi} is the number of reactions with uncertain bidirectionality for modelling the reaction bidirectionality. Then, in view of Eq. (8), a high *relative* model prior probability $p(\mathcal{M}_i)$ and data likelihood $p(D|\mathcal{M}_i)$, such as for $\mathcal{M}_2$, do not necessarily lead to a high (close to one) *absolute* model probability $p(\mathcal{M}_i|D)$ of models, since $p(\mathcal{M}_i|D)$ has to be normalized against the totality of all other possible $2^{n_{bi}}$ models. This speaks to common sense: If we have a great many equally good explanations, the probability of any one being the correct model is small. This apparent conclusion does, however, have far-reaching consequences for ^{13}C-MFA in general. In particular, employing model comparison approaches, such as probability likelihood ratios, to test all model variants for finding a winning model makes little sense. This is because, for example, selecting a single bidirectionality setting neglects a substantial amount of uncertainty related to the model selection process, yet the flux inferences would be oblivious to this uncertainty. This insight thus reveals a fundamental Achilles heel of the current practice of applying a single ^{13}C-MFA model for flux inference: there are likely to be many equally good, but quite different, solutions based on slightly different model structures that are consistent with the data.

3.4. Multi-model ^{13}C-MFA using Bayesian Model Averaging

A promising approach to avoiding the pitfalls of using a single model for flux inference when there is considerable epistemic model uncertainty, is to combine the inferences of all candidate models. Bayesian model averaging (BMA) adopts such an approach by including all potential models, but weighting each model according to its model posterior probability, the likelihood of the data, given the model (Hoeting et al., 1999). Despite being an established statistical tool in many fields (Fragoso et al., 2018), BMA is only rarely employed in biochemical network modelling (Borah Slater et al., 2023; Mitosch et al., 2023; Oates et al., 2014; Theorell and Nöh, 2020). To consider model uncertainty for flux inference with BMA, we extend Eq. (1) by averaging over a model set:

$$p(\theta|D) = \sum_{i=1}^{N} p\left(\theta_{\mathcal{M}_i}\middle|\mathcal{M}_i, D\right) \cdot p(\mathcal{M}_i|D) \qquad (9)$$

Herein, $p(\mathcal{M}_i|D)$ is the absolute posterior model probability of model $\mathcal{M}_i$ in view of the data D, as given in Eq. (8), which is weighted by $p\left(\theta_{\mathcal{M}_i}\middle|\mathcal{M}_i, D\right)$, i.e., the model-specific flux posterior probability distribution of the (net) flux parameters, θ, that are shared across the models contained in the model set. Loosely speaking, the calculation rule in Eq. (9) averages out the joint influences of the uncertain model structures.

Application of BMA requires the computationally challenging task of computing the flux posteriors for a set of models, which relates to a two-layered averaging process: the first averaging over the (continuous) flux space of each model candidate, and the second averaging over the (discrete) model space. Instead of calculating the posterior probabilities of all $N = 2^{n_{bi}}$ models and their flux probability distributions separately, inference is performed using Reversible Jump Markov Chain Monte Carlo (RJMCMC), a trans-dimensional sampling algorithm that samples (discrete) model structures and (continuous) parameter values simultaneously (Green and Hastie, 2009), thereby side-stepping the explicit calculation of each model weight in Eq. (9). Specifically, the relative number of times a model was sampled by RJMCMC here approximates the model's relative probability. Technical details about the construction of the RJMCMC transition densities, specific to ^{13}C-MFA, are described in the Supplementary Information (Supplementary Sec. Methods) and Theorell and Nöh (2020).

In this context, an important, but subtle point of distinction between single- and multi-model ^{13}C-MFA is that, from the perspective of probability theory, there is a fundamental difference between replacing a bidirectional reaction in a network by a unidirectional one (change in model structure), and the alternative of setting its exchange flux to zero (fixing a parameter value), despite the fact these alternative settings yield the same simulated labelling enrichment. This is because removing a bidirectionality (or exchange flux) from a model alters the models' probability, as this is proportional to the likelihood of the data, averaged over all feasible flux values. Consequently, eliminating a bidirectionality from a model lowers the dimensionality of its flux solution space that is averaged over, and therefore affects the average likelihood. In contrast, setting an exchange flux to 0 has no influence on the fit averaged over all other fluxes, since the flux solution space in this case remains the same.

3.5. BMA-based ^{13}C-MFA: More robustness with fewer assumptions

Applying BMA to the *E. coli* study, we next consider the set of all (nested) candidate models derived from model $\mathcal{M}_0$, through setting some bidirectional reactions as unidirectional. We denote this set by $\{\mathcal{M}_0\}$, and recall that the simpler models $\mathcal{M}_1$ and $\mathcal{M}_2$ are two of the total of $2^{17} = 131.072$ models that constitute the set. All models in $\{\mathcal{M}_0\}$ share the same set of net fluxes, but differ in the composition of exchange fluxes, and therefore the number of free parameters, short degrees of freedom (DOF), i.e., $\theta = \theta^{net}$ in Eq. (9). Furthermore, we consider all model candidates in the set $\{\mathcal{M}_0\}$ to be equally likely. If the labelling data is informative about the exchange fluxes, application of Bayes' theorem results in constriction of the model set, by excluding models that have both too few and too many bidirectional reactions. This automatism contrasts the sequential test-based model updating strategies that are currently in use (Hendry et al., 2020).

BMA with the model set $\{\mathcal{M}_0\}$ yields net flux credible intervals similar to those of the original model $\mathcal{M}_0$ using single-model inference, as seen for 9 net fluxes in Figure 5 (red and blue, respectively), and for the remaining fluxes in Supplementary Fig. S9. However, the shape of the posterior distributions exhibits marked differences. Most prominently, BMA locates more probability mass to low values of *emp1.n* and, consequently, much higher values of *ppp1.n,* than the conventional single-model inference with the original model. Because $\mathcal{M}_0$ is contained in the model set $\{\mathcal{M}_0\}$, the difference in shape between the single-model posterior distributions and the BMA-derived distributions can be interpreted as inability of the $\mathcal{M}_0$ model to represent the overall uncertainty in the model structure.

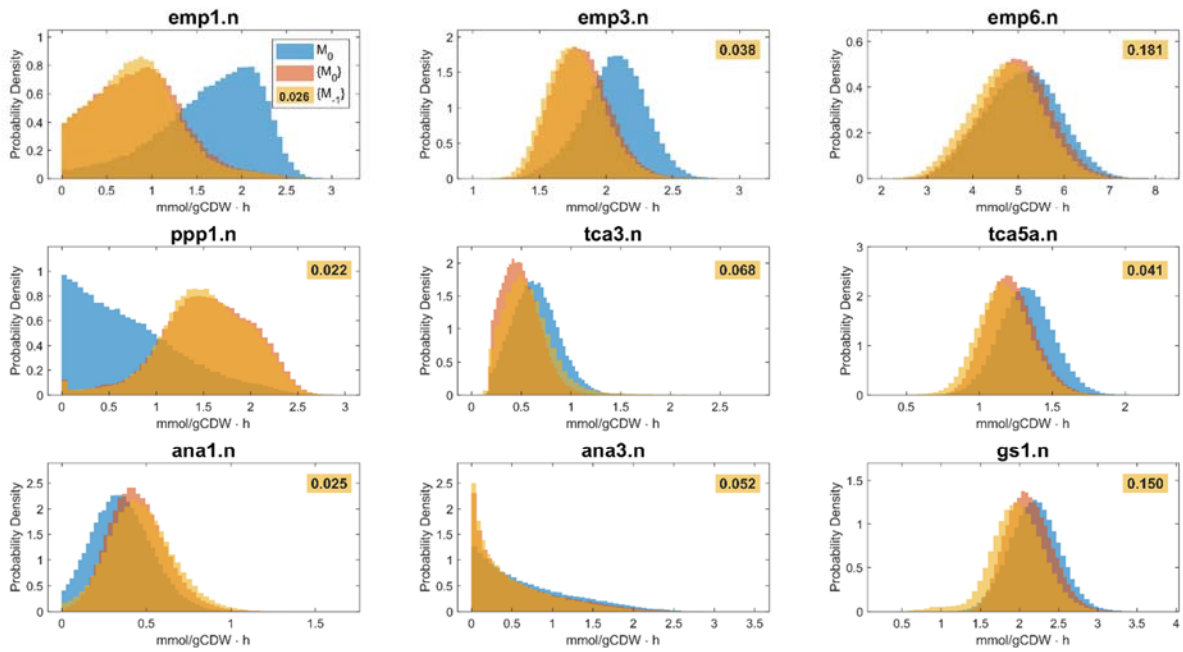

Figure 5: Comparison of flux posterior distributions using single-model and multi-model inference. *Single-model inferences are derived with the published model* $\mathcal{M}_0$ *(blue, see also Fig. 4), multi-model inferences with the model sets* $\{\mathcal{M}_0\}$ *and* $\{\mathcal{M}_{-1}\}$ *(light and dark orange), for net fluxes shown in Fig. 4. Multi-model inferences are strikingly consistent. Results for the remaining fluxes are given in Supplementary Figs. S8-S10, expected values and 99% credible intervals in Supplementary Fig. S12. Numbers give the 1-Wasserstein distance between the posterior distributions derived with* $\{\mathcal{M}_0\}$ *and* $\{\mathcal{M}_{-1}\}$ *(in mmol/*g_{CDW}*/h), respectively. The pairwise model comparison of Wasserstein distances in Supplementary Figs. S11 shows that similarity across all net fluxes inferred using multi-model sets is substantially larger than inferences derived with any two single models.*

For single-model inference we found similar incoherent net flux inferences when comparing the original model $\mathcal{M}_0$ and the model $\mathcal{M}_{-1}$ having 7 additional bidirectional reactions (Figure 5, Supplementary Fig. S7). To study whether net flux inferences (in terms of net flux estimates and credible intervals) derived with the BMA approach are sensitive with respect to a larger model set, we repeated our BMA analysis with the more complex model set $\{\mathcal{M}_{-1}\}$. This set is a 128 times larger *super-set* of $\{\mathcal{M}_0\}$ that consists of $2^{24} = 16.777.216$ model structures. In fact, we found a striking consistency for the multi-model inferences obtained for $\{\mathcal{M}_0\}$ and $\{\mathcal{M}_{-1}\}$, i.e., all net flux posteriors were entirely reproduced as can be seen for the 9 selected net fluxes in Figure 5 and the remaining ones in Supplementary Fig. S8. In addition, Supplementary Fig. S11 shows that the differences in posterior distributions between the two models sets are considerably smaller than the differences between single models, e.g. between $\mathcal{M}_0$ and $\mathcal{M}_{-1}$, as measured by 1-Wasserstein distances. Generalizing from flux inference using a single model and MCMC (Theorell et al., 2017), to flux inference using multiple models and RJMCMC did not increase the computational effort, a remarkable finding that also has previously been noted (Theorell and Nöh, 2020).

3.6. Discovering new insights into bidirectional reaction steps

In all our BMA-based inferences, the exchange fluxes are marginalized according to Eq. (9). However, from the results obtained by RJMCMC sampling, we are still able to approximate the marginal posterior probability of how likely it is that a particular reaction step is uni- or bidirectional in view of the given data. By definition, the marginal posterior probability equals the cumulative probability over all models in which the reaction is found to be uni- and bidirectional, respectively, by inspecting the fraction of samples in which the considered exchange flux is positive. We investigated again both model sets $\{\mathcal{M}_0\}$ and $\{\mathcal{M}_{-1}\}$ and the results are shown in Figure 6. Each potentially bidirectional reaction amongst the set members was classified to be either bidirectional (probability $p \sim 1.0$, black), unidirectional (probability $p \sim 0.0$, red), or inconclusive (blue). Similar to the net flux posterior distributions, the bidirectionality inferences were found to be strikingly consistent (cf. Supplementary Fig. S13).

We found strong evidence that four of the unidirectional reactions that were artificially set to be bidirectional in $\mathcal{M}_{-1}$ are indeed unidirectional: *tca1, tca3* and *tca4* (with $p \ll 0.01$) in the TCA cycle and *ppp1* in the oxidative PPP ($p \approx 0.05$). Note that all these reactions are associated with carboxylation/decarboxylation steps in which the data strongly supports the expected decarboxylation direction, as was reasoned in the reference study (Zamboni et al., 2009). On the other hand, BMA provided no evidence in favour of either uni- or bidirectionality in the further three reactions (*tca2*, *gs1*, *gs2)* that were set unidirectional in the reference study. Furthermore, one reaction, *ppp4* in the non-oxidative part of the PPP that was set bidirectional in the original model, was found to be unidirectional in the majority ($p \approx 0.94$) of solutions. Decisive evidence in favour of the bidirectionality of *emp5* ($p \approx 1.0$) was also found, implying that all models in which the enzyme enolase was set to operate unidirectionally failed to fit the data. Also interesting is that pyruvate kinase, which converts phosphoenolpyruvate (PEP) to pyruvate (PYR), *emp6*, which was set as bidirectional in the original study, shows the second highest probability of being bidirectional ($p \approx 0.9$). This result is in line with studies subsequent to the original paper, such as Long et al. (2017), which was based on an extensive series of more than a dozen labelling experiments, also supporting bidirectionality in this enzyme. The remaining reactions have neither very high, nor very low probability of being bidirectional, meaning that the dataset was uninformative with regard to uni- or bidirectionality of these reactions.

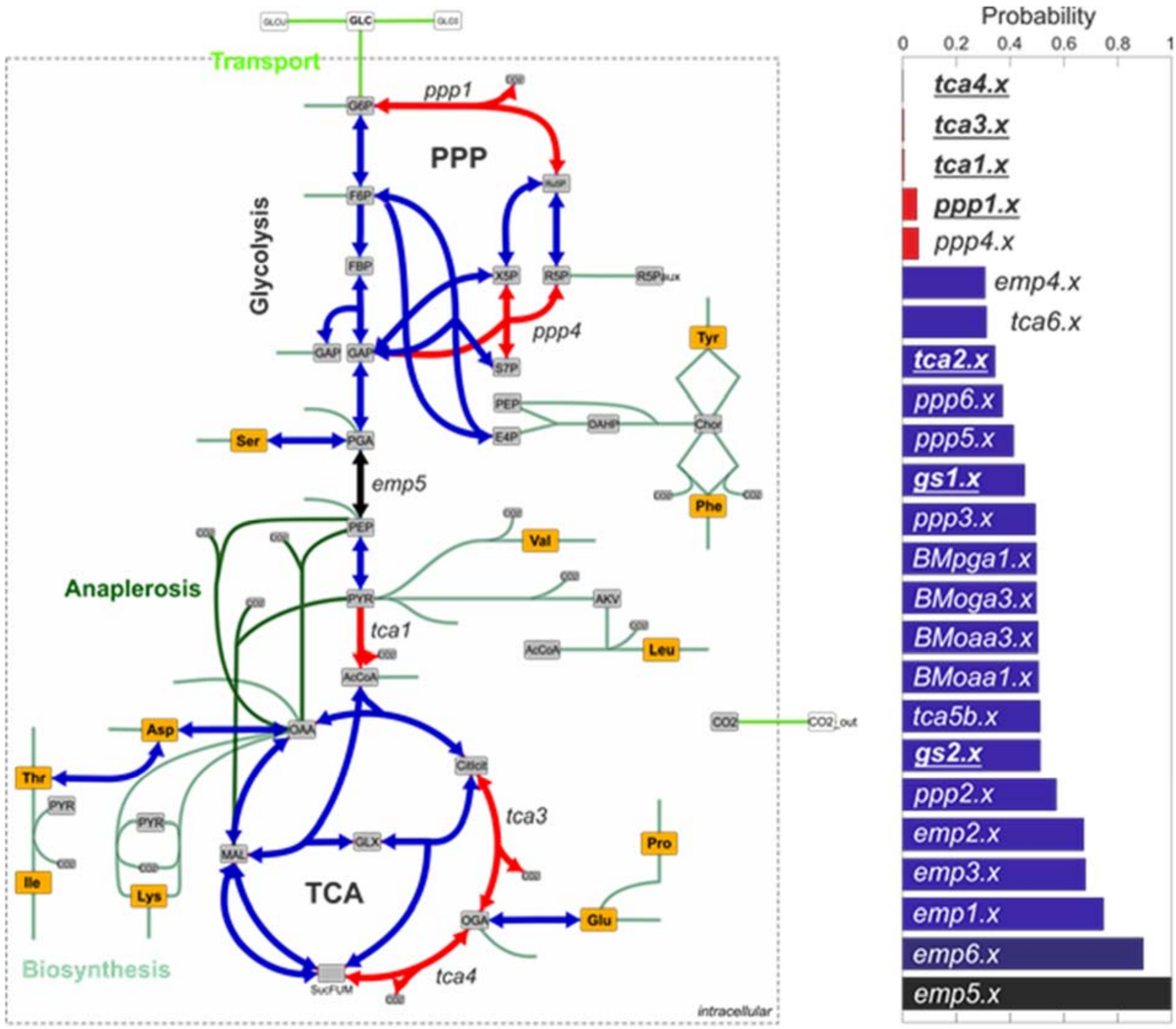

Figure 6: Structural inference of reaction bidirectionalities with model set $\{\mathcal{M}_{-1}\}$**.** *By definition, the marginal bidirectionality probability equals the summed probability of all sampled models for which a reversible reaction is assigned to be bidirectional (.x > 0), where the probability of a model is represented by the number of times it is sampled divided by the total number of samples in the sampled model ensemble. High probability values (reactions and bars printed in black) give strong evidence for bidirectionality, and low values (reactions and bars in red) give strong support for unidirectionality. Medium probabilities (blue) imply that the investigated dataset is uninformative to whether the reaction is bi- or unidirectional. The 7 reactions considered unidirectional in the reference study* (Zamboni et al., 2009)*, are printed in bold and underlined. Values for posterior probabilities are listed in Supplementary Table S1.*

3.7. BMA penalizes over-simplistic and over-complex model formulations

The marginal posterior distributions derived from the model ensemble $\{\mathcal{M}_{-1}\}$ for both, net fluxes and bidirectionalities, are nearly indistinguishable from those obtained with its subset $\{\mathcal{M}_0\}$, despite the fact that the number of considered model variants in $\{\mathcal{M}_{-1}\}$ is two orders of magnitude higher. It is surprising that the inferences derived using BMA remain remarkably robust, despite the drastic increase of the model set. We hypothesise that this is because the models with high probability, and thus high influence on the inferences, are the same regardless of whether the $\{\mathcal{M}_0\}$ or the $\{\mathcal{M}_{-1}\}$ model set is used, regardless of their difference in the underlying bidirectionality assumption sets. Since our results are calculated with RJMCMC, sampling model variants according to their probability in view of the data, the most important models should be contained in the sampled model sets, hitherto denoted the *effective model set*. To test our hypothesis, we compared the models in the effective model sets $\{\mathcal{M}_0\}^{\text{eff}}$ and $\{\mathcal{M}_{-1}\}^{\text{eff}}$, taking the number of independent fluxes (DOFs) as a proxy for model complexity.

In the Bayesian framework, the posterior DOF distribution of the effective model set is determined by the DOF distribution of the underlying "prior" (i.e. complete) model set and the data. Due to our premise that each model within the prior model set is equally conceivable, the prior DOF distribution is binomial, shifted by an offset of 10 (the number of independent net fluxes). The prior and posterior DOF distributions are shown in Figure 7. For the complex model set $\{\mathcal{M}_{-1}\}$, consisting of models with maximally 34 prior DOFs, the prior distribution is centred at 22.00 ± 3.32 DOF, whereas the posterior has a mean complexity of 20.41 ± 1.71 DOF. Thus, the data gives preference to simpler models with fewer bidirectionalities than the average model in $\{\mathcal{M}_{-1}\}$. For the set of models $\{\mathcal{M}_0\}$ with maximally 27 prior DOFs, the binomial DOF prior has a mean of 18.50 ± 3.04, 3.5 DOFs *less* than the mean model complexity of $\{\mathcal{M}_{-1}\}$. Interestingly, for $\{\mathcal{M}_0\}$ the posterior mean (19.14 ± 1.59) is located to the right compared to the prior mean, i.e. the data enforced a preference for models with *more* bidirectionalities than the prior average, showing the automatism embodied in the Bayesian framework: it does not generally select the simplest models if they fail to account for the data. The difference of only 1.23 DOFs between the two posterior DOF distribution means of the effective model ensembles $\{\mathcal{M}_0\}^{\mathrm{eff}}$ and $\{\mathcal{M}_{-1}\}^{\mathrm{eff}}$ is markedly small, despite the fact that $\{\mathcal{M}_{-1}\}$ contains more than 16 million models variants. Notably, the minimal models are located in the very left tail of the posterior DOF distributions, whereas the original model resides at the right tail.

The results show that the apparent robustness of the multi-model inferences indeed stems from the characteristics of BMA to arrive at stable posterior distributions. Practically, this means that regardless of how complex a super-model set of $\{\mathcal{M}_0\}$ is, the contributions of all unnecessarily complex models, i.e. models with additional bidirectionalities that do not have decisive support from the data, are automatically excluded so that inferences remain coherent. This tempered version of Ockham's razor contrasts the observation made for the single-model inferences and shows that BMA is a powerful remedy for dealing with uncertain bidirectionalities.

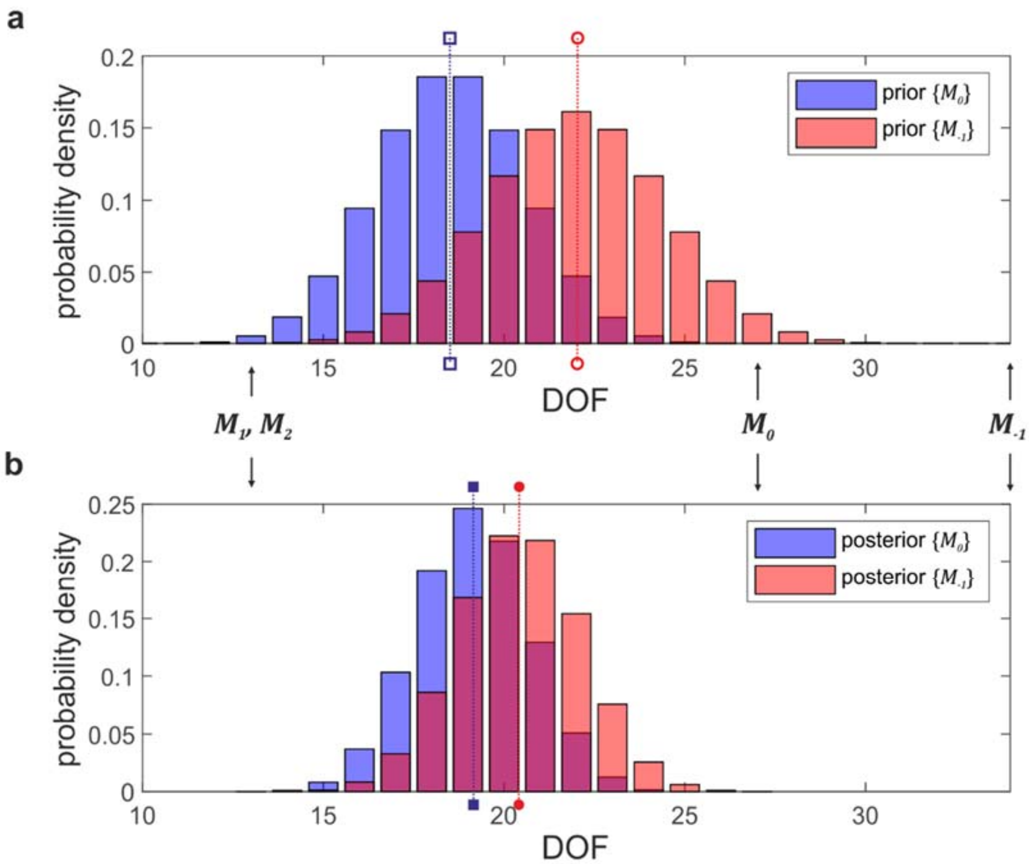

Figure 7: Necessary vs. unnecessary (effective) model complexities. *DOF posterior distributions derived by employing BMA to the model sets $\{\mathcal{M}_0\}$ and $\{\mathcal{M}_{-1}\}$. (a) Prior DOF probability densities for $\{\mathcal{M}_0\}$ (10-24 DOFs) (blue) and $\{\mathcal{M}_{-1}\}$ (10-34 DOFs) (red, overlaid) are shifted binomial distributions with means indicated by vertical lines. (b) Posterior DOF probability densities for the two effective model sets. The mean of the posterior DOF distribution for $\{\mathcal{M}_0\}^{\text{eff}}$ is shifted towards models with on average more bidirectionalities compared to the prior, whereas the mean of the posterior DOF distribution for $\{\mathcal{M}_{-1}\}^{\text{eff}}$ is shifted towards models with fewer bidirectionalities compared to the prior, yielding posterior distributions that are largely overlapping. For comparison, the DOFs of the minimal $\mathcal{M}_1, \mathcal{M}_2$ (DOF = 13), the original model $\mathcal{M}_0$ (27), and the complex model $\mathcal{M}_{-1}$ (34) are indicated.*

4. Conclusions

Modelling isotope labelling data is an inherently difficult problem, where the investigated complex biological systems invariably require the modeller to create complex models for describing the data that are nevertheless a rough approximation of the system under study. Comprehensiveness has been demanded, as the seemingly safer path so as not to overlook any possible model solution (Hendry et al., 2020), an argument that is common in systems biology (Westerhoff et al., 2009). However, the scope and granularity of models necessarily have to achieve a compromise between simplicity and comprehensiveness. In fact, in the majority of investigations, as the one scrutinized here, modellers aim at including all important model components, but expel unnecessary complexity by integrating hard-coded assumptions, as priors, about the system. However, such *a priori* assumptions may not be pertinent to a particular system or tested condition, therefore may introduce bias. For example, a biochemical reaction that acts reversible *in vitro* may not be reversible *in vivo* when actual physiological conditions differ from those applied in the test tube. This modelling assumption is therefore only reliable when they can be tested for validity, in the best case experimentally. The extent to which model-based inferences depend on untested, or even untestable, assumptions has not been rigorously investigated until now.

Adopting the Bayesian approach to tackle the longstanding problem of modelling bidirectional reaction steps in ^{13}C-MFA as a test case, we here delineate the extent to which bidirectionality assumptions, either too few or too many, are problematic. We show for a well-investigated *E. coli* test case that, by ignoring uncertainty about the model assumptions, the use of a single model, whether too simple or comprehensive, can come at severe risks of biases together with loss of reliability and robustness of the resulting flux inferences: On the one hand, we show that direct application of simplicity (Ockham’s razor) to find the simplest solution neglects model, and thereby underestimates flux, uncertainty, as there may be a great many simple models with very different flux solutions. On the other hand, a comprehensive model does not safeguard reliable inferences, because it risks attributing experimental noise to superfluous model parameters.

To overcome these problems in ^{13}C-MFA, we propose to use Bayesian multi-modelling as an alternative paradigm, where multiple models of a range of complexities are considered simultaneously, and the importance of the individual models is weighted according to the models' probability, given the data. This factor automatically penalizes too complex models, so is consistent with the principle of Ockham's razor, but also favours more complex models if these are needed to explain the data. An overview of single- and multi-model ^{13}C-MFA inference is found in Supplementary Fig. S4.

The practical utility of the multi-model approach to faithfully represent model uncertainty was demonstrated for the published study where BMA gives coherent flux inferences, even in the case of comprehensive model sets. Here, BMA unlocked insights that cannot be obtained for any single-model technique. Strikingly, the analysis confirmed some of the uni- and bidirectionalities that were fixed in the published model, whereas other published uni- and bidirectionalities remained unsupported by the data at hand. This feature of the multi-model approach, which was previously demonstrated on synthetic data (Theorell and Nöh, 2020), allowed us to not only arrive at robust flux inferences, but also to discover new evidence in the data for uni- and bidirectional reactions in the published model, but without risking the bias of *a priori* assumptions. Notably, BMA-based ^{13}C-MFA now supports the established, but previously problematic, practice of, in case of doubt, setting reversible reactions as bidirectional.

Our analysis thereby highlights the epistemological potential of multi-model inference using BMA for the kind of model inference that is fundamental for metabolic engineering. We demonstrated that rigorous inferences can still be obtained even in the frequent case of incomplete knowledge of the system. Moreover, valid inferences can still be made in the common situation in which details of some entities in the system remain unresolved. Both features are particularly useful in network biology where problems are vastly underdetermined and finding a single model that resolves all uncertainties is highly improbable. Importantly, with a proxy model that describes the relevant aspects of the system being a component of the BMA model set, the working mechanism of BMA automatically assigns low importance to unnecessarily comprehensive models, even if they outnumber the set of "useful" models by orders of magnitude. In the case of countably many mechanistic models, as in modelling bidirectionalities in ^{13}C-MFA, this pruning capability of the tempered Ockham's razor (1) counteracts the "madness of crowds" (Stumpf, 2020), (2) behaves beneficially with respect to its computational resource efficiency and scalability (Theorell and Nöh, 2020), and (3) has proved extraordinary useful in practice (Borah Slater et al., 2023; Mitosch et al., 2023).

Although we investigated the case of bidirectional reaction steps, multi-model inference with BMA is far from limited to this type of model uncertainty in metabolic flux inference, but is relevant for inference under any kind of model uncertainty. This makes BMA also pertinent to a wide range of inference problems in biology where many model assumptions are only rarely tested or are even untestable by the data. Similar problems arise for gene regulatory networks and signalling networks or indeed any problem that must deal with model uncertainty (Hangos et al., 2014; Haunschild et al., 2005; Oates et al., 2014; Timonen et al., 2019). Neglecting model uncertainty can only be justified in the cases where one model is so likely that it outweighs all other candidates, which is a very rare occurrence, particularly in biology. For the vast majority of cases in which uncertainty is embedded in models, we argue that multi-model approaches, such as BMA, should be applied to unlock the full epistemological potential of these underdetermined models.

Availability of data

The model files used for the study are available in Supplementary Data. The MCMC datasets generated and analysed in this study, as well as the Wasserstein distances are available at https://github.com/JuBiotech/Supplement-to-Theorell-et-al.-Metabolic_Engineering-2024.

References

Antoniewicz, M.R., Kelleher, J.K., Stephanopoulos, G., 2006. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. Metab. Eng. 8, 324–337. https://doi.org/10.1016/j.ymben.2006.01.004

Bayes, T., R, P., 1763. An essay towards solving a problem in the doctrine of chances. By the late Rev. Mr. Bayes, F. R. S. communicated by Mr. Price, in a letter to John Canton, A. M. F. R. S. Philos. Trans. R. Soc. London 53, 370–418. https://doi.org/10.1098/rstl.1763.0053

Beard, D. a, Liang, S., Qian, H., 2002. Energy balance for analysis of complex metabolic networks. Biophys. J. 83, 79–86. https://doi.org/10.1016/S0006-3495(02)75150-3

Becker, J., Zelder, O., Häfner, S., Schröder, H., Wittmann, C., 2011. From zero to hero - Design-based systems metabolic engineering of *Corynebacterium glutamicum* for L-lysine production. Metab. Eng. 13, 159–168. https://doi.org/10.1016/j.ymben.2011.01.003

Beyß, M., Azzouzi, S., Weitzel, M., Wiechert, W., Nöh, K., 2019. The design of FluxML: A universal modeling language for ^{13}C metabolic flux analysis. Front. Microbiol. 10. https://doi.org/10.3389/fmicb.2019.01022

Borah, K., Beyß, M., Theorell, A., Wu, H., Basu, P., Mendum, T.A., Nöh, K., Beste, D.J.V., McFadden, J., 2019. Intracellular *Mycobacterium tuberculosis* exploits multiple host nitrogen sources during growth in human macrophages. Cell Rep. 29, 3580-3591.e4. https://doi.org/10.1016/j.celrep.2019.11.037

Borah Slater, K., Beyß, M., Xu, Y., Barber, J., Costa, C., Newcombe, J., Theorell, A., Bailey, M.J., Beste, D.J. V, McFadden, J., Nöh, K., 2023. One-shot ^{13}C^{15}N-metabolic flux analysis for simultaneous quantification of carbon and nitrogen flux. Mol. Syst. Biol. 19. https://doi.org/https://doi.org/10.15252/msb.202211099

Brewer, B.J., Foreman-Mackey, D., 2018. DNest4: Diffusive nested sampling in C++ and python. J. Stat. Softw. 86, 31. https://doi.org/10.18637/jss.v086.i07

Brewer, B.J., Pártay, L.B., Csányi, G., 2011. Diffusive nested sampling. Stat. Comput. 21, 649–656. https://doi.org/10.1007/s11222-010-9198-8

Burnham, K.P., Anderson, D.R., 2002. Model selection and multimodel inference: A practical information-theoretic approach, 2nd ed. Springer Verlag, New York.

Crown, S.B., Long, C.P., Antoniewicz, M.R., 2015. Integrated ^{13}C-metabolic flux analysis of 14 parallel labeling experiments in *Escherichia coli*. Metab. Eng. 28, 151–158. https://doi.org/10.1016/j.ymben.2015.01.001

Das, M., Patra, P., Ghosh, A., 2020. Metabolic engineering for enhancing microbial biosynthesis of advanced biofuels. Renew. Sustain. Energy Rev. 119, 109562. https://doi.org/10.1016/j.rser.2019.109562

Droste, P., Nöh, K., Wiechert, W., 2013. Omix - A visualization tool for metabolic networks with highest usability and customizability in focus. Chemie Ing. Tech. 85, 849–862. https://doi.org/10.1002/cite.201200234

Ellison, A.M., 2004. Bayesian inference in ecology. Ecol. Lett. 7, 509–520. https://doi.org/10.1111/j.1461-0248.2004.00603.x

Follstad, B.D., Stephanopoulos, G., 1998. Effect of reversible reactions on isotope label redistribution - analysis of the pentose phosphate pathway. Eur. J. Biochem. 252, 360–371. https://doi.org/10.1046/j.1432-1327.1998.2520360.x

Foster, C.J., Gopalakrishnan, S., Antoniewicz, M.R., Maranas, C.D., 2019. From *Escherichia coli* mutant ^{13}C labeling data to a core kinetic model: A kinetic model parameterization pipeline. PLOS Comput. Biol. 15, e1007319. https://doi.org/10.1371/journal.pcbi.1007319

Fragoso, T.M., Bertoli, W., Louzada, F., 2018. Bayesian Model Averaging: A systematic review and conceptual classification. Int. Stat. Rev. 86, 1–28. https://doi.org/10.1111/insr.12243

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2013. Bayesian Data Analysis, 3rd ed. Chapman & Hall/CRC, Boca Raton.

Gelman, A., Vehtari, A., Simpson, D., Margossian, C.C., Carpenter, B., Yao, Y., Kennedy, L., Gabry, J., Bürkner, P.-C., Modrák, M., 2020. Bayesian workflow. arxiv. https://doi.org/10.48550/arXiv.2011.01808

Green, P.J., Hastie, D.I., 2009. Reversible Jump MCMC. Genetics 155, 1391–1403. https://doi.org/10.1093/bioinformatics/btn143

Hangos, K.M., Szederkenyi, G., Rudan, J., 2014. Efficient computation of alternative structures for large kinetic systems using Linear Programming. MATCH Commun. Math. Comput. Chem. 71, 71–92.

Haunschild, M.D., Freisleben, B., Takors, R., Wiechert, W., 2005. Investigating the dynamic behavior of biochemical networks using model families. Bioinformatics 21, 1617–1625. https://doi.org/10.1093/bioinformatics/bti225

Hendry, J.I., Dinh, H. V., Foster, C., Gopalakrishnan, S., Wang, L., Maranas, C.D., 2020. Metabolic flux analysis reaching genome wide coverage: Lessons learned and future perspectives. Curr. Opin. Chem. Eng. 30, 17–25. https://doi.org/10.1016/j.coche.2020.05.008

Heux, S., Bergès, C., Millard, P., Portais, J.-C., Létisse, F., 2017. Recent advances in high-throughput ^{13}C-fluxomics. Curr. Opin. Biotechnol. 43, 104–109. https://doi.org/10.1016/j.copbio.2016.10.010

Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: A tutorial. Stat. Sci. 14, 382–417. https://doi.org/10.1214/ss/1009212519

Jadebeck, J.F., Theorell, A., Leweke, S., Nöh, K., 2021. HOPS: High-performance library for (non-)uniform sampling of convex-constrained models. Bioinformatics 37, 1776–1777. https://doi.org/10.1093/bioinformatics/btaa872

Jadebeck, J.F., Wiechert, W., Nöh, K., 2023. Practical sampling of constraint-based models: Optimized thinning boosts CHRR performance. PLOS Comput. Biol. 19, e1011378. https://doi.org/10.1371/journal.pcbi.1011378

Jefferys, W.H., Berger, J.O., 1992. Ockham's razor and bayesian analysis. Am. Sci. 80, 64–72.

Kadirkamanathan, V., Yang, J., Billings, S.A., Wright, P.C., 2006. Markov chain Monte Carlo algorithm based metabolic flux distribution analysis on Corynebacterium glutamicum. Bioinformatics 22, 2681–2687. https://doi.org/10.1093/bioinformatics/btl445

Kappelmann, J., Beyß, M., Nöh, K., Noack, S., 2019. Separation of ^{13}C- and ^{15}N-isotopologues of amino acids with a primary amine without mass resolution by means of O-phthalaldehyde derivatization and collision induced dissociation. Anal. Chem. 91, 13407–13417. https://doi.org/10.1021/acs.analchem.9b01788

Kappelmann, J., Wiechert, W., Noack, S., 2016. Cutting the Gordian Knot: Identifiability of anaplerotic reactions in *Corynebacterium glutamicum* by means of ^{13}C-metabolic flux analysis. Biotechnol. Bioeng. 113, 661–674. https://doi.org/10.1002/bit.25833

Kass, R.E., Raftery, A.E., 1995. Bayes Factors. J. Am. Stat. Assoc. 90, 773–795. https://doi.org/10.1080/01621459.1995.10476572

Kirk, P., Thorne, T., Stumpf, M.P., 2013. Model selection in systems and synthetic biology. Curr. Opin. Biotechnol. 24, 767–774. https://doi.org/10.1016/j.copbio.2013.03.012

Kochanowski, K., Okano, H., Patsalo, V., Williamson, J., Sauer, U., Hwa, T., 2021. Global coordination of metabolic pathways in *Escherichia coli* by active and passive regulation. Mol. Syst. Biol. 17. https://doi.org/10.15252/msb.202010064

Lagziel, S., Lee, W.D., Shlomi, T., 2019. Studying metabolic flux adaptations in cancer through integrated experimental-computational approaches. BMC Biol. 17, 51. https://doi.org/10.1186/s12915-019-0669-x

Leighty, R.W., Antoniewicz, M.R., 2013. COMPLETE-MFA: Complementary parallel labeling experiments technique for metabolic flux analysis. Metab. Eng. 20, 49–55. https://doi.org/10.1016/j.ymben.2013.08.006

Liebermeister, W., Noor, E., 2021. Model balancing: A search for in-vivo kinetic constants and consistent metabolic states. Metabolites 11, 749. https://doi.org/10.3390/metabo11110749

Long, C.P., Antoniewicz, M.R., 2019. High-resolution ^{13}C metabolic flux analysis. Nat. Protoc. 14, 2856–2877. https://doi.org/10.1038/s41596-019-0204-0

Long, C.P., Au, J., Sandoval, N.R., Gebreselassie, N.A., Antoniewicz, M.R., 2017. Enzyme I facilitates reverse flux from pyruvate to phosphoenolpyruvate in *Escherichia coli*. Nat. Commun. 8, 14316. https://doi.org/10.1038/ncomms14316

MacKay, D.J.C., 2003. Information Theory, Inference, and Learning Algorithms. Cambridge University Press, Cambridge.

Marx, A., de Graaf, A.A., Wiechert, W., Eggeling, L., Sahm, H., 1996. Determination of the fluxes in the central metabolism of *Corynebacterium glutamicum* by nuclear magnetic resonance spectroscopy combined with metabolite balancing. Biotechnol. Bioeng. 49, 111–129. https://doi.org/10.1002/(SICI)1097-0290(19960120)49:2<111::AID-BIT1>3.0.CO;2-T

McCloskey, D., Young, J.D., Xu, S., Palsson, B.O., Feist, A.M., 2016a. Modeling method for increased precision and scope of directly measurable fluxes at a genome-scale. Anal. Chem. 88, 3844–3852. https://doi.org/10.1021/acs.analchem.5b04914

McCloskey, D., Young, J.D., Xu, S., Palsson, B.O., Feist, A.M., 2016b. MID Max: LC-MS/MS method for measuring the precursor and product mass isotopomer distributions of metabolic intermediates and cofactors for metabolic flux analysis applications. Anal. Chem. 88, 1362–1370. https://doi.org/10.1021/acs.analchem.5b03887

McFadden, J., 2023. Razor sharp: The role of Occam's razor in science. Ann. N. Y. Acad. Sci. 1530, 8–17. https://doi.org/10.1111/nyas.15086

Mitosch, K., Beyß, M., Phapale, P., Drotleff, B., Nöh, K., Alexandrov, T., Patil, K.R., Typas, A., 2023. A pathogen-specific isotope tracing approach reveals metabolic activities and fluxes of intracellular Salmonella. PLOS Biol. 21, e3002198. https://doi.org/10.1371/journal.pbio.3002198

Morey, R.D., Hoekstra, R., Rouder, J.N., Lee, M.D., Wagenmakers, E.-J., 2016. The fallacy of placing confidence in confidence intervals. Psychon. Bull. Rev. 23, 103–123. https://doi.org/10.3758/s13423-015-0947-8

Munger, J., Bennett, B.D., Parikh, A., Feng, X.J., McArdle, J., Rabitz, H.A., Shenk, T., Rabinowitz, J.D., 2008. Systems-level metabolic flux profiling identifies fatty acid synthesis as a target for antiviral therapy. Nat. Biotechnol. 26, 1179–1186. https://doi.org/10.1038/Nbt.1500

Murphy, T.A., Dang, C. V, Young, J.D., 2013. Isotopically nonstationary ^{13}C flux analysis of Myc-induced metabolic reprogramming in B-cells. Metab. Eng. 15, 206–217. https://doi.org/10.1016/j.ymben.2012.07.008

Niedenführ, S., Wiechert, W., Nöh, K., 2015. How to measure metabolic fluxes: A taxonomic guide for ^{13}C fluxomics. Curr. Opin. Biotechnol. 34, 82–90. https://doi.org/10.1016/j.copbio.2014.12.003

Nielsen, J., 2003. It is all about metabolic fluxes. J. Bacteriol. 185, 7031–7035. https://doi.org/10.1128/JB.185.24.7031-7035.2003

Nöh, K., Niedenführ, S., Beyß, M., Wiechert, W., 2018. A Pareto approach to resolve the conflict between information gain and experimental costs: Multiple-criteria design of carbon labeling experiments. PLoS Comput. Biol. 14, e1006533. https://doi.org/10.1371/journal.pcbi.1006533

Oates, C.J., Dondelinger, F., Bayani, N., Korkola, J., Gray, J.W., Mukherjee, S., 2014. Causal network inference using biochemical kinetics. Bioinformatics 30, i468–i474. https://doi.org/10.1093/bioinformatics/btu452

Pullen, N., Morris, R.J., 2014. Bayesian model comparison and parameter inference in systems biology using nested sampling. PLoS One 9, e88419. https://doi.org/10.1371/journal.pone.0088419

Quek, L.-E., Wittmann, C., Nielsen, L.K., Krömer, J.O., 2009. OpenFLUX: Efficient modelling software for ^{13}C-based metabolic flux analysis. Microb. Cell Fact. 8, 25.

https://doi.org/10.1186/1475-2859-8-25

Rahim, M., Ragavan, M., Deja, S., Merritt, M.E., Burgess, S.C., Young, J.D., 2022. INCA 2.0: A tool for integrated, dynamic modeling of NMR- and MS-based isotopomer measurements and rigorous metabolic flux analysis. Metab. Eng. 69, 275–285. https://doi.org/10.1016/j.ymben.2021.12.009

Sauer, U., 2006. Metabolic networks in motion: ^{13}C-based flux analysis. Mol. Syst. Biol. 2, 62. https://doi.org/10.1038/msb4100109

Skilling, J., 2006. Nested sampling for general Bayesian computation. Bayesian Anal. 1, 833–859. https://doi.org/10.1214/06-BA127

Sokol, S., Millard, P., Portais, J.-C., 2012. influx_s: Increasing numerical stability and precision for metabolic flux analysis in isotope labelling experiments. Bioinformatics 28, 687–693. https://doi.org/10.1093/bioinformatics/btr716

Stephanopoulos G., Aristidou A., J., N., 1998. Metabolic Engineering. Academic Press, San Diego, CA. https://doi.org/10.1016/B978-0-12-666260-3.X5000-6

Stumpf, M.P.H., 2020. Multi-model and network inference based on ensemble estimates: avoiding the madness of crowds. J. R. Soc. Interface 17, 20200419. https://doi.org/10.1098/rsif.2020.0419

Sundqvist, N., Grankvist, N., Watrous, J., Mohit, J., Nilsson, R., Cedersund, G., 2022. Validation-based model selection for ^{13}C metabolic flux analysis with uncertain measurement errors. PLOS Comput. Biol. 18, e1009999. https://doi.org/10.1371/journal.pcbi.1009999

Theorell, A., Jadebeck, J.F., Nöh, K., Stelling, J., 2022. PolyRound: Polytope rounding for random sampling in metabolic networks. Bioinformatics 38, 566–567. https://doi.org/10.1093/bioinformatics/btab552

Theorell, A., Leweke, S., Wiechert, W., Nöh, K., 2017. To be certain about the uncertainty: Bayesian statistics for ^{13}C metabolic flux analysis. Biotechnol. Bioeng. 114, 2668–2684. https://doi.org/10.1002/bit.26379

Theorell, A., Nöh, K., 2020. Reversible jump MCMC for multi-model inference in metabolic flux analysis. Bioinformatics 36, 232–240. https://doi.org/10.1093/bioinformatics/btz500

Timonen, J., Mannerstrom, H., Lahdesmaki, H., Intosalmi, J., 2019. A probabilistic framework for molecular network structure inference by means of mechanistic modeling. IEEE/ACM Trans. Comput. Biol. Bioinforma. 16, 1843–1854. https://doi.org/10.1109/TCBB.2018.2825327

Wasserman, L., 2000. Bayesian model selection and model averaging. J. Math. Psychol. 44, 92–107. https://doi.org/10.1006/jmps.1999.1278

Weitzel, M., Nöh, K., Dalman, T., Niedenführ, S., Stute, B., Wiechert, W., 2013. 13CFLUX2—high-performance software suite for ^{13}C-metabolic flux analysis. Bioinformatics 29, 143–145. https://doi.org/10.1093/bioinformatics/bts646

Westerhoff, H. V., Winder, C., Messiha, H., Simeonidis, E., Adamczyk, M., Verma, M., Bruggeman, F.J., Dunn, W., 2009. Systems Biology: The elements and principles of Life. FEBS Lett. 583, 3882–3890. https://doi.org/10.1016/j.febslet.2009.11.018

Wiechert, W., 2007. The thermodynamic meaning of metabolic exchange fluxes. Biophys. J. 93, 2255–2264. https://doi.org/10.1529/biophysj.106.099895

Wiechert, W., 2001. ^{13}C metabolic flux analysis. Metab. Eng. 3, 195–206. https://doi.org/10.1006/mben.2001.0187

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

Wiechert, W., Möllney, M., Isermann, N., Wurzel, M., de Graaf, A.A., 1999. Bidirectional reaction steps in metabolic networks. Part III: Explicit solution and analysis of isotopomer labeling systems. Biotechnol. Bioeng. 66, 69–85. https://doi.org/10.1002/(SICI)1097-0290(1999)66:2<69::AID-BIT1>3.0.CO;2-6

Wiechert, W., Möllney, M., Petersen, S., de Graaf, A.A., 2001. A universal framework for ^{13}C metabolic flux analysis. Metab. Engingeering 3, 265–283. https://doi.org/10.1006/mben.2001.0188

Wiechert, W., Nöh, K., 2021. Quantitative metabolic flux analysis based on isotope labeling, in: Nielsen, J., Stephanopoulos, G., Lee, S.Y. (Eds.), Metabolic Engineering: Concepts and Applications. Wiley, pp. 73–136. https://doi.org/10.1002/9783527823468.ch3

Wiechert, W., Siefke, C., de Graaf, A.A., Marx, A., 1997. Bidirectional reaction steps in metabolic networks: II. Flux estimation and statistical analysis. Biotechnol. Bioeng. 55, 118–135. https://doi.org/10.1002/(SICI)1097-0290(19970705)55:1<118::AID-BIT13>3.0.CO;2-I

Wu, C., Guarnieri, M., Xiong, W., 2023. FreeFlux: A Python package for time-efficient isotopically nonstationary metabolic flux analysis. ACS Synth. Biol. https://doi.org/10.1021/acssynbio.3c00265

Wu, S.G., Wang, Y., Jiang, W., Oyetunde, T., Yao, R., Zhang, X., Shimizu, K., Tang, Y.J., Bao, F.S., 2016. Rapid prediction of bacterial heterotrophic fluxomics using machine learning and constraint programming. PLOS Comput. Biol. 12, e1004838. https://doi.org/10.1371/journal.pcbi.1004838

Xu, J., Martien, J., Gilbertson, C., Ma, J., Amador-Noguez, D., Park, J.O., 2020. Metabolic flux analysis and fluxomics-driven determination of reaction free energy using multiple isotopes. Curr. Opin. Biotechnol. 64, 151–160. https://doi.org/10.1016/j.copbio.2020.02.018

Young, J.D., 2014. INCA: A computational platform for isotopically non-stationary metabolic flux analysis. Bioinformatics 30, 1333–1335. https://doi.org/10.1093/bioinformatics/btu015

Zamboni, N., Fendt, S.-M., Rühl, M., Sauer, U., 2009. ^{13}C-based metabolic flux analysis. Nat. Protoc. 4, 878–892. https://doi.org/10.1038/nprot.2009.58

Zelle, E., Pfelzer, N., Oldiges, M., Koch-Koerfges, A., Bott, M., Nöh, K., Wiechert, W., 2021. An energetic profile of *Corynebacterium glutamicum* underpinned by measured biomass yield on ATP. Metab. Eng. 65, 66–78. https://doi.org/10.1016/j.ymben.2021.03.006

Zhao, Z., ten Pierick, A., de Jonge, L., Heijnen, J.J., Wahl, S.A., 2012. Substrate cycles in *Penicillium chrysogenum* quantified by isotopic non-stationary flux analysis. Microb. Cell Fact. 11, 140. https://doi.org/10.1186/1475-2859-11-140

Zheng, J., Yang, J., Liang, X., Fang, M., Wang, Y., 2024. Dual strategy for ^{13}C-Metabolic flux analysis of central carbon and energy metabolism in Mammalian cells based on LC-isoMRM-MS. Talanta 266, 125074. https://doi.org/10.1016/j.talanta.2023.125074

Acknowledgements

We wish to thank Andrea Rocco and Jakob Vanhoefer for discussions in the early stages of the work. AT was supported by the ERA-NET for Industrial Biotechnology (ERA IB Grant ERA-IB-14-81) (www.era-ib.net) and JFJ received funding from the Helmholtz Association of German Research Centres in the framework of the Helmholtz School for Data Science in Life, Earth and Energy (HDS-LEE) (www.hds-lee.de). JM was supported by the UK Biotechnology and Biological Sciences Research Council (BBSRC Grant BB/J002097/1) (www.bbsrc.ac.uk).

Author Statement

A.T. Conceptualization, Methodology, Software, Formal analysis, Investigation, Visualization, Original Draft, Writing - Reviewing & Editing

J.F.J. Software, Formal analysis, Writing - Reviewing & Editing

K.N. Conceptualization, Methodology, Formal analysis, Investigation, Visualization, Supervision, Original Draft, Writing - Reviewing & Editing, Funding acquisition

J.M. Conceptualization, Original Draft, Writing - Reviewing and Editing

W.W. Writing - Reviewing and Editing

Competing Interests statement

The authors declare no competing interests.