REVIEW

BIOTECHNOLOGY PROGRESS

Applied Cellular Physiology and Metabolic Engineering

Model validation and selection in metabolic flux analysis and flux balance analysis

Joshua A. M. Kaste[1,2] 🔘 | Yair Shachar-Hill[2]

[1]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan, USA

[2]Department of Plant Biology, Michigan State University, East Lansing, Michigan, USA

Correspondence
Joshua A. M. Kaste, Department of Biochemistry and Molecular Biology, Michigan State University, 603 Wilson Rd, East Lansing, MI 48823, USA.
Email: kastejos@msu.edu

Abstract

13C-Metabolic Flux Analysis (13C-MFA) and Flux Balance Analysis (FBA) are widely used to investigate the operation of biochemical networks in both biological and biotechnological research. Both methods use metabolic reaction network models of metabolism operating at steady state so that reaction rates (fluxes) and the levels of metabolic intermediates are constrained to be invariant. They provide estimated (MFA) or predicted (FBA) values of the fluxes through the network in vivo, which cannot be measured directly. These fluxes can shed light on basic biology and have been successfully used to inform metabolic engineering strategies. Several approaches have been taken to test the reliability of estimates and predictions from constraint-based methods and to compare alternative model architectures. Despite advances in other areas of the statistical evaluation of metabolic models, such as the quantification of flux estimate uncertainty, validation and model selection methods have been underappreciated and underexplored. We review the history and state-of-the-art in constraint-based metabolic model validation and model selection. Applications and limitations of the χ^2-test of goodness-of-fit, the most widely used quantitative validation and selection approach in 13C-MFA, are discussed, and complementary and alternative forms of validation and selection are proposed. A combined model validation and selection framework for 13C-MFA incorporating metabolite pool size information that leverages new developments in the field is presented and advocated for. Finally, we discuss how adopting robust validation and selection procedures can enhance confidence in constraint-based modeling as a whole and ultimately facilitate more widespread use of FBA in biotechnology.

KEYWORDS
constraint-based modeling, flux balance analysis, metabolic flux analysis, metabolic modeling, model selection, model validation

1 | INTRODUCTION

The set of biochemical reaction rates in the metabolic network of a living system (its flux map) represents an integrated functional phenotype that emerges from multiple layers of biological organization and

regulation, including the genome, transcriptome, and proteome.[1] The study of metabolic fluxes is therefore important for systems biology, rational metabolic engineering, and synthetic biology. A grand challenge of systems biology is building an integrated mechanistic understanding of the operation of living organisms across these levels of

regulation[2] – an understanding that goes beyond statistical or correlative descriptions, however useful these can be. Meeting this challenge requires fluxes to be accurately predicted from network structure using explicit rules or hypotheses and reliably estimated using experimental data. Fluxes are also critical to many biotechnological and metabolic engineering applications. Examples such as the development of lysine hyper-producing strains of *Corynebacterium glutamicum*[3–5] and the rewiring of *E. coli*'s metabolism to make it grow chemoautotrophically[6] attest to the usefulness of these techniques. As the scale and complexity of integrative systems biology and biological engineering efforts increase, so too will the need for reliable and robust estimates of fluxes.

In vivo fluxes cannot be directly measured, necessitating modeling approaches to estimate or predict them. The most commonly used approaches for metabolic modeling is the constraint-based modeling frameworks of 13C-Metabolic Flux Analysis (13C-MFA) and Flux Balance Analysis (FBA). Both require a metabolic network consisting of metabolites linked by biochemical reactions to be defined using the biochemical literature, knowledge of the enzymes and transporters expressed from the genome, and physico-chemical rules. In 13C-MFA, atom mappings describing the positions and interconversions of the carbon atoms in reactants and products are also included in the model. These methods assume that the system is at metabolic steady-state, such that the concentrations of all metabolic intermediates and reaction rates are constant.[7] External fluxes, such as the uptake of a substrate or the rate of production of new cells or a product, are also measured and used to constrain the possible flux ranges. These assumptions and constraints define a "solution space" containing all flux maps consistent with them but are typically insufficient to pinpoint a unique flux map.

In 13C-MFA, isotopic labeling data is used to identify a particular solution within the solution space. 13C-labeled substrates are fed to the system under investigation and the endpoint labeling, or time-course labeling in Isotopically Nonstationary Metabolic Flux Analysis (INST-MFA), of metabolites is measured using mass spectrometry and/or NMR techniques.[7,8] Given a metabolic network, a flux map, and information about the labeled substrate fed into the system, the label distribution through all the metabolites in a network can be solved analytically. However, 13C-MFA works backwards from measured label distributions to flux maps by minimizing the differences between measured and estimated Mass Isotopomer Distribution (MID) values by varying flux estimates.[9] For INST MFA pool size measurements can also be included in the minimization process.

In FBA, linear optimization is used to identify a flux map (or set of flux maps) from the solution space.[10] This is the map(s) for which the sum of one or more fluxes (the objective function) is maximized or minimized. Objective functions frequently represent measures of efficiency, including the maximization of growth rate or product formation or the minimization of total flux.[11] Such functions may embody hypotheses about what the in vivo system has been evolutionarily tuned to optimize, or questions about the operational capacity of that system under particular conditions. Since the objective function, together with the network architecture and empirical and/or theoretical constraints introduced by the modeler, is a key determinant of the flux maps generated

by FBA, careful selection, justification, and, ideally, validation of objective functions is crucial. As shown in Ref. 12, alternative objective functions can, and should, be evaluated to identify those that result in the best agreement with experimental data. In many cases, the constraints – typically on external fluxes – imposed during an FBA optimization result in a set of viable flux maps (a solution space) rather than a single map. In such cases, related techniques, including Flux Variability Analysis[13] and random sampling[14–17] can be used to characterize the set of flux maps consistent with the set constraints. The computational tractability and small amount of experimental data necessary to perform FBA allow the analysis of Genome-Scale Stoichiometric Models (GSSMs). These models incorporate all known reactions believed to occur in an organism based on a combination of genome annotation and manual curation. Additional linear-optimization-based methods for solving GSSMs using the FBA framework have been developed and are sometimes used together with FBA. These include Minimization of Metabolic Adjustment (MOMA),[18] and Regulatory On/Off Minimization (ROOM),[19] as well as a host of methods that incorporate omic data into the optimization process (e.g.,[20–24]). FBA and its related methods are sometimes used to analyze models other than true GSSMs, such as "core" models that focus on central metabolic processes that conduct the large majority of flux.[25] When discussing validation, however, the same principles apply to all of these linear optimization methods and across the different model scales. For the sake of simplicity, we will be using "FBA" to refer to this family of methods generally and will refer to the medium- to large-scale models used with these methods as "FBA models."

Progress has been made in improving the statistical rigor and reliability of flux estimates and characterizing uncertainty in estimates and predictions. For example, in MFA, the development of effective methods for flux uncertainty estimation[26] allows researchers to better quantify confidence in flux predictions and, where appropriate, to gather additional data to better support their conclusions. Bayesian techniques for the characterization of uncertainties in flux estimates derived from isotopic labeling have also been presented.[27] On the experimental side of MFA, there have been advances in designing and implementing parallel labeling experiments, wherein multiple tracers are employed in parallel labeling experiments and the results are simultaneously fit to generate a single 13C-MFA flux map. This enables more precise estimation of fluxes than experiments with individual tracers or tracer combinations allow.[28–35] Greater resolution in isotopic labeling data through the use of tandem mass spectrometry techniques, which allow for the quantification of positional labeling, can also improve the precision of modeled fluxes, as described in Ref. 36,37. Recent years have also seen developments in FBA meant to improve the reliability of its predictions. For example studies have characterized the impact of departures from metabolic steady state and devised methods to account for uncertainties in biomass compositions (e.g.,[38,39]). The many sources of uncertainty when working with FBA and genome-scale models, and attempts to characterize and mitigate this uncertainty, have been reviewed elsewhere.[40]

In this review, we specifically focus on the validation of flux predictions and estimates from constraint-based modeling studies and

the selection of well-supported model architectures, which have received less attention and specific treatment in the literature. How can MFA and FBA researchers validate the accuracy of their estimates and predictions? These flux analysis methods also require researchers to make choices about the network structure of the model to be used. This leads to questions of model selection; that is, how do we select the most statistically justified model from among the alternatives? Validation and model selection are key to improving the fidelity of model-derived fluxes to the real in vivo ones. The fields of systems and synthetic biology have seen substantial development of model selection and validation practices,[41,42] but these topics are not frequently discussed in the metabolic modeling literature. Previous reviews and methods papers have touched on the use of tools like the χ^2-test of goodness-of-fit for the validation of MFA models.[43,44] However, to our knowledge, no reviews covering the various methods for validating FBA predictions exist, nor have previous reviews discussed the various limitations of the χ^2-test. Moreover, previous reviews have not addressed the most recent improvements in model selection in 13C-MFA, which have not been adequately incorporated into routine practice. Addressing these topics explicitly is important for practitioners as they carry out their work. It is also important for readers of the flux analysis literature, who must understand the assumptions, tests of validity, and model selection techniques underlying what they are reading.

Although only a subset of research groups conduct both FBA and MFA modeling, we believe most metabolic modeling practitioners and consumers read literature containing both modeling paradigms. As we highlight in this review, some similar themes emerge when examining the validation of both FBA and MFA flux maps. Finally, one of the most robust validations that can be conducted for FBA predictions is comparison against MFA estimated fluxes, which makes simultaneously considering the validity of both FBA and MFA flux maps crucial. For these reasons, we consider both modeling approaches in this review.

We review and provide our perspective on these areas and prospects for future development, highlighting: (1) Validation methods applicable to FBA flux maps; (2) approaches for validating 13C-MFA flux maps; and (3) developments and prospects for model selection in 13C-MFA; (4) How validation and model selection practices in 13C-MFA could benefit from a greater emphasis on the isolation of training and validation datasets and; (5) the importance of corroborating flux mapping results using independent modeling and experimental techniques.

2 | VALIDATION TECHNIQUES IN FBA AND 13C-MFA

Flux Balance Analysis and 13C-MFA studies commonly validate the model(s) used, though there is great variation in their nature and extent. We summarize these validation strategies in Figure 1.

2.1 | Validation in FBA

The COnstraint-Based Reconstruction and Analysis (COBRA) framework, implemented in software solutions such as the COBRA Toolbox[45] and *cobrapy*[46] and widely used for FBA studies, features functions and pipelines that can be used to ensure basic functionality of models including balancing of charge, pH, and cofactors/cosubstrates, thermodynamic feasibility, and connectivity of all metabolites. Model characteristics evaluated include the inability to generate ATP without an external source of energy and the inability to synthesize biomass without adding substrates not known to be needed. Additionally, the MEMOTE (MEtabolic MOdel TEsts) pipeline contains tests to ensure, for example, that biomass precursors can be successfully synthesized in a model in a variety of growth media.[47] MEMOTE has been used to ensure appropriate stoichiometry and consistency with accepted format standards in models entered into the BiGG[48] model database. These forms of Quality Control are an important first step in ensuring that models are behaving appropriately and generating useful predictions. However, following these initial checks on functionality, the techniques used to validate actual model predictions are varied and not standardized. Indeed, even in the BiGG database, which is highly curated and focuses primarily on models of microbial systems, models vary in the type and extent of validation performed. Given the variety of validation procedures that appear in the literature, it is important when using an FBA model to be aware of what specific validations were used, what their limitations are, and consequently, what inferences or downstream applications are appropriate (summarized in Table 1).

Perhaps the most common validation in FBA is comparison between FBA-predicted and empirically measured rates of growth (e.g.,[49–55]). One may similarly evaluate growth/no-growth in different media and/or with different carbon sources (e.g.,[51,54–57]). A related approach is the comparison of in silico metabolite uptake/secretion with experimental measurements.[54,57,58] Such evaluations give confidence in the model's basic predictions. To ensure that the accuracy of growth-rate predictions generalizes well, we strongly recommend validating growth rates on substrates or in media conditions from which biomass composition and parameters like Growth-Associated Maintenance (GAM) and Non-Growth Associated Maintenance (NGAM) costs were not experimentally derived, as done in Ref. 51. GAM represents the energy expenditure needed to support a certain rate of biomass growth and NGAM represents the energy expenditure required for a cell or organism to survive without any net growth.[59] These values may vary depending on growth conditions, so testing whether the values measured in one set of conditions generalize to others is important. Otherwise, future users may use a model with, for example, another common media composition and find – or worse yet, simply not notice – that the resulting predictions do not accurately reflect essential characteristics of the organism's actual metabolism.

A related approach involves comparing growth/no-growth of gene knockout strains to FBA predictions to address whether the metabolic pathways used in the model mirror the biological system. Experimentally verified lethal knockouts that appear nonlethal in silico point to alternative routes the model can use to grow. Conversely, in silico lethality predictions not confirmed by experiment suggest the model is missing isoforms or alternative reaction routes. Collecting the true positive, true negative, false positive, and false negative predictions from the in silico versus in vivo lethality predictions into a
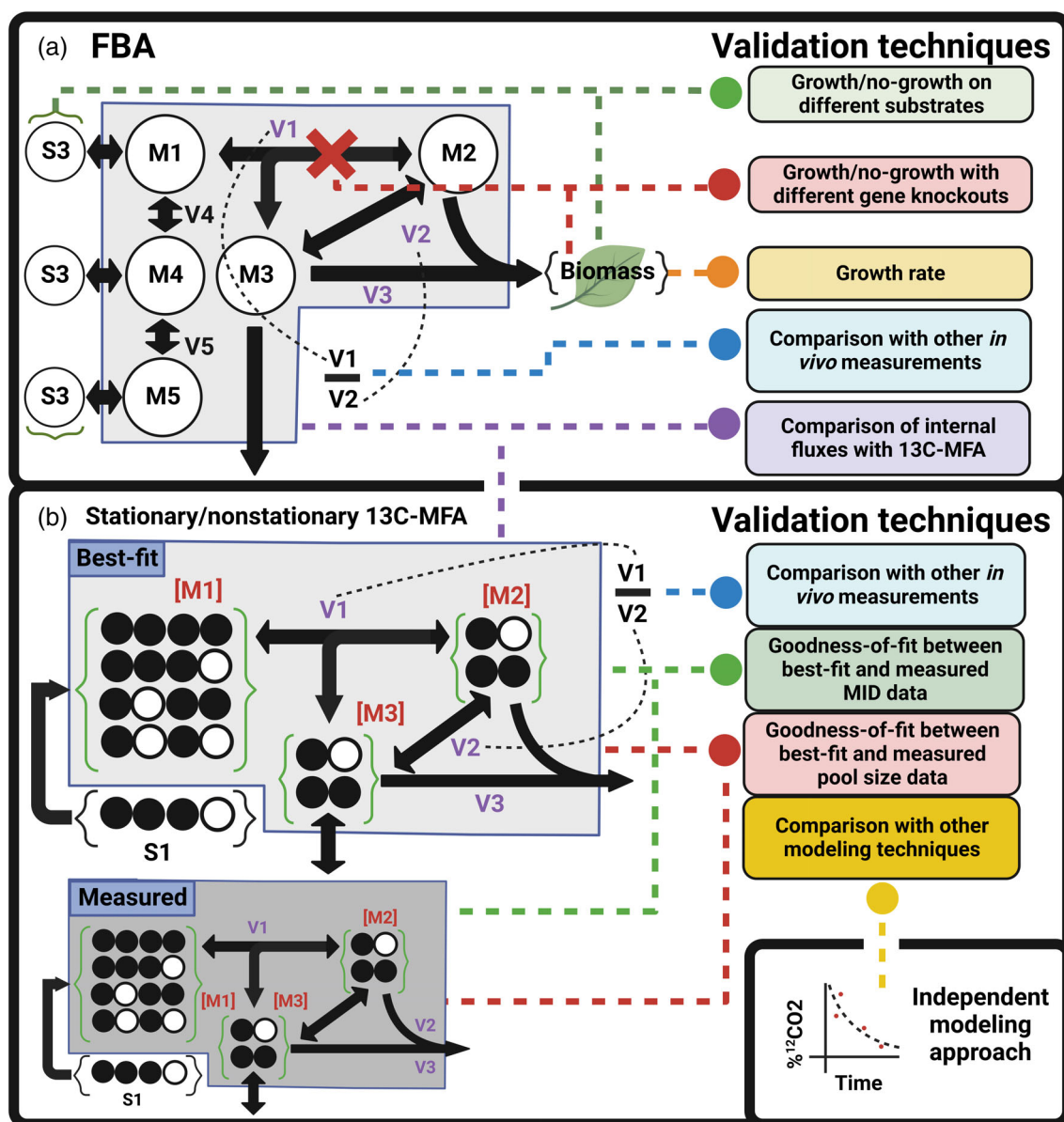
FIGURE 1   Graphical summary of validation strategies in (a) FBA and (b) 13C-MFA. Dotted lines connect inputs with the associated validation technique(s). (a) FBA predictions can be validated by comparing growth rate or growth/no-growth phenotypes across different substrates, growth conditions, or sets of gene knockouts in silico and in vivo. Values can be calculated from flux maps and compared with experimental measurements. FBA internal flux predictions can be compared with 13C-MFA fluxes. (b) Values can be calculated from 13C-MFA flux maps and compared with an independent experimental measurement from the in vivo system. Goodness-of-fit can be assessed between simulated and measured MIDs, and simulated and measured metabolite pool sizes in INST-MFA. Flux maps can be compared with the results of independent modeling exercises. Molecules are schematically shown as connected circles of atomic positions: open circles are unlabeled, and filled circles are isotopically labeled. Mn, metabolites in the metabolic network; S_n, exogenous substrates; V_i, Fluxes; $[M_n]$, metabolite concentrations.

confusion matrix allows for an at-a-glance evaluation of overall model accuracy and for the comparison of alternative model architectures.[68] Researchers sometimes use algorithms to identify knockouts that couple biomass accumulation to flux through a reaction for biotechnological applications.[63–65] This requires that models accurately predict growth/no-growth phenotypes for gene knockouts, but previous work in a model of *Saccharomyces cerevisiae*, for example, shows that FBA performs poorly at predicting the synthetic lethality of double-

knockouts, making this a serious concern.[66] When performing such validations, one must keep in mind that imposed constraints and decisions made during the model construction or optimization process may implicitly or explicitly add the predictions one is trying to validate into the model, rendering the exercise meaningless. This makes clear and transparent documentation of the assumptions used in the modeling process key for reviewers and readers to assess the epistemic value of the validations that are reported.

TABLE 1 The most common model validation strategies in Flux Balance Analysis, what these methods tell us, limitations, and important considerations for researchers and/or readers, and examples of these methods' implementation in the literature.

Method	Information content	Limitations	Use case	Examples
Comparison of growth/no-growth on one or more substrates	Presence/absence of reactions necessary for substrate utilization and biomass synthesis.	Validation is qualitative, only indicating the existence of metabolic routes. Does not test the accuracy of predicted internal flux values	Useful when viability/nonviability of different growth conditions is of interest. Unlike a growth-rate comparison, does not indicate whether the efficiency of biomass synthesis is realistic.	51,55,56,60,61
Comparison of growth rates on one or more substrates	Consistency of metabolic network, biomass composition, and maintenance costs with observed efficiency of substrate-to-biomass conversion.	Provides quantitative information on the overall efficiency of substrate conversion to biomass, but is uninformative with respect to the accuracy of internal flux predictions.	When done across multiple substrates and conditions, this validation gives confidence in the predicted efficiency with which the model produces biomass. Useful when identifying growth-limiting factors.	51,55,60,62
Comparison of in vivo and in silico knockout lethality	Presence/absence of biosynthetic reactions necessary for substrate use and growth.	Care is needed to reduce incorrect predictions from many different factors, including optimization method and biomass composition changes in response to knockout.	Critically important to perform when designing growth-coupled knockout strategies.63–65	62,66–68
Comparison of FBA predictions with MFA fluxes	Accuracy of internal flux predictions.	Few MFA flux maps exist for most organisms, making this validation impossible or requiring comparison with an MFA flux map taken for very different experimental conditions.	Important when the intended use of FBA modeling requires that the predictions of specific internal flux values be accurate.	60,69–71

It is crucial to note that the methods discussed above do not validate the internal flux predictions made by FBA. Due to the underdetermined nature of FBA, many radically different flux maps may be compatible with, for example, the optimization of growth-rate,[13] making validations using growth-rate or any other individual external flux uninformative with respect to internal flux distributions. In well-characterized systems, there may be a wealth of known metabolic functionalities that an organism can carry out and evaluating whether the model can reproduce them can give some assurance of realistic model behavior. In Ref. [72,73], 288 metabolic processes known to take place in mammalian cells were evaluated in models of human and mouse models, though it was only the ability to carry out the processes at all, and not the actual flux values, that were evaluated. In favorable cases, individual internal fluxes can be quantitatively estimated in vivo using independent methods and compared directly to ones from a predicted flux map to provide a powerful form of validation. For example, in a study from our group[74] the ratio of the cyclic electron flow (CEF) to linear electron flow (LEF) fluxes in photosynthesis predicted by FBA was evaluated against CEF/LEF ratios from fluorescence measurements for validation purposes. Though less specific, the sum of FBA-predicted values for fluxes that produce and/or consume a product (such as CO_2) can also be compared to

experimental measurements. In addition to these approaches, there is the possibility going forward of integrating metabolomics data into the FBA prediction process (e.g.,[75]) and/or comparison of FBA results against metabolomic datasets. Although, it should be noted that metabolite levels and changes in those levels in the steady-state cannot be directly interpreted in terms of fluxes, so any attempts to validate FBA results using observations in metabolomics datasets should be done with caution.

However, validations of internal flux predictions across the network require comparing FBA flux maps with high-quality ones from 13C-MFA. Such validations are the most information-rich of all the methods surveyed so far and tell us the most about how well the FBA flux maps generated by a particular combination of network architecture, constraints, and objective function line up with experimental data. Unfortunately, 13C-MFA flux maps are time-consuming to generate, making this "gold-standard" validation rare. To compare FBA-predicted and MFA-estimated fluxes, the model architectures must be the same, or the MFA must at least be a subnetwork of the model used for the FBA. Additionally, the empirical constraints (e.g., substrate uptake and biomass accumulation) must be the same in both cases. In cases where the growth rates predicted or constrained for an FBA flux map do not perfectly line up with those from an MFA

flux map, normalization of fluxes to account for this discrepancy can be used to get an apples-to-apples comparison.[69] The imposition of identical external flux constraints on both the FBA and MFA models may preclude validation of the accuracy of certain external flux predictions by the FBA. However, such comparisons can be done afterwards by removing the relevant constraints. Comparison is also complicated by the underdetermined nature of most FBA optimizations, which can result in large feasible ranges for the individual fluxes being compared against the corresponding flux values obtained from 13C-MFA, making the validation less stringent. FBA optimizations that assume parsimony[11,76] tend to yield narrower flux ranges, but this advantage may come at the cost of neglecting other plausible objective functions that might be more accurate.

Finally, when FBA-predicted and MFA-estimated flux maps disagree, assuming the experimental constraints are consistent between the two and that the person doing the comparison is confident in the MFA estimates, either the FBA network architecture or objective function could be to blame. There is not, to our knowledge, a consistent strategy for disambiguating disagreements due to architecture or objective function. If the biological/biochemical accuracy of the objective function is in question, methods for inferring objective functions using isotopic labeling data can be employed (e.g.,[77]) the resulting objective functions can be compared with the one being used, and discrepancies can be considered. All objective functions that relate to growth will be affected by the accuracy of the biomass composition used in the model, although in some systems central metabolic fluxes may be relatively robust to variability in the exact values of this composition.[78] In systems for which extensive biomass composition data is available, known variability in biomass composition can be incorporated during the optimization process.[39] Despite these various limitations and difficulties when validating FBA using 13C-MFA fluxes, some studies have evaluated the accuracy of FBA against 13C-MFA-estimated flux maps (e.g.,[22,54,60,70,79–81]) with mixed results.

A consistent challenge when validating FBA fluxes using any method is the need to compare the FBA flux map against empirical fluxes or other measurements that were generated under similar conditions to those being simulated. For organisms or systems whose metabolic models are undergoing continual refinement, thus requiring repeated validation, community-curated and updated validation datasets generated under well-defined and carefully reported conditions may be useful. Standards on what metabolic phenotypes and responses need to be captured by these models (e.g., the 288 known metabolic functions in human cells used in Ref. [72]) may also help ensure that reconstructions maintain essential biological features as they grow larger and more detailed.

To summarize, we make the following recommendations for the validation of FBA-predicted flux maps:

1. When possible, comparisons between FBA-predicted and 13C-MFA-estimated flux maps should be performed to validate the accuracy of FBA-predicted internal fluxes. This provides a greater wealth of information about where and to what extent the model is, and is not, lining up with experimental evidence. When performing such validations, care should be taken to ensure that the conditions under which the FBA-predictions and MFA-estimates are generated are as similar as possible and that any necessary normalizations to account for differences have been made. For an example of thorough FBA-to-MFA comparisons, see Ref. [69,82].
   - **Note**: FBA-predicted flux maps require definition not just of the network architecture and constraints, but also an objective function for optimization. Validation of the FBA-predicted flux maps is therefore also a validation of the selected objective function. It is possible for a poorly selected objective function to generate flux predictions that do not align with MFA-estimated fluxes; in such cases, alternative objective functions can be explored.

2. As highlighted in Table 1, different validation methods evaluate different aspects of the model's predictions. Therefore, employing a number of different validations allows for a fuller and more detailed analysis of model performance and increases the likelihood that other users of the model may be able to appropriately apply it to their research question. For an example of a study employing several different validation techniques, see Ref. [57].

3. Validations of model predictions are only valuable when the data the predictions are validated against has not already been used in the training or construction of the model. The complexity of the metabolic model reconstruction and analysis process can make it difficult to notice when contamination of the validation dataset by training data has occurred. In order to identify contamination, one must consider the source of all data used for validation and consider whether it or a value derived from it was used at any stage of the FBA modeling process. For an example of a study that clearly and systematically validates FBA predictions while avoiding such contamination, see Ref. [51].

Improving confidence in the accuracy of FBA flux maps is valuable because generating validated 13C-MFA flux maps for all systems and conditions of interest is impractical. 13C-MFA requires substantial experimental work for each set of conditions and is unsuitable for many multicellular tissues and organisms where the required combination of extended periods of metabolic steady state, controlled provision of informative, non-perturbing labeled substrates, and obtaining enough labeling data cannot be achieved. This FBA-empowered future for systems biology and biotechnology requires well-validated MFA flux maps, so we turn our attention to model validation and selection in MFA.

2.2 | Validation in 13C-MFA

13C-MFA flux estimates are typically validated based on the goodness-of-fit between measured labeling data and the corresponding values generated by the network model after the optimization of model parameters. The goodness-of-fit is represented by the sum of squared residuals (SSR) where each residual is weighted by dividing it

by its experimental variance. The χ^2-test of goodness-of-fit, which is built into commonly used 13C-MFA software,[83-85] is then used to test whether the SSR falls within the 95% confidence interval expected for the defined number of degrees of freedom (DOF). Since its development as a validation method in 13C-MFA,[26] the χ^2-test has been widely used and has been useful in the validation of 13C-MFA metabolic models inferred from genome annotations.[86-92]

However, as described in Ref. 93,27 the use of the χ^2-test can be problematic in 13C-MFA for several reasons. When upper- and lower-bounds are imposed on estimated flux parameter values, this makes accurate estimation of the effective DOF for the χ^2-test difficult.[27] It can also be difficult to accurately determine errors in the MID measurements made for 13C-MFA, resulting in distortion of the variance-weighted SSR values that are being compared against the 95% Confidence Interval.[93]

In addition to these technical difficulties with properly applying the χ^2-test, problems arise from how the test is implemented into the model development process during a typical 13C-MFA study. Especially for eukaryotic systems, 13C-MFA flux modeling generally involves making iterative changes to the model based on how well it can explain the data – as assessed informally and by the χ^2-test – followed by refinement and assessment of the data based on this agreement. For example, if the data do not allow the fluxes between the same metabolite in different compartments to be determined, they may be merged in the model or additional measurements may be made to resolve them. Metabolites may also be excluded from the model due to inconsistency between their simulated versus measured MIDs causing the model to fail the χ^2-test, on the assumption that biological, model-structural, or analytical uncertainties underlie these unexplained divergences.[94*] The difficulty of accurately quantifying MID measurement errors, mentioned earlier, may be addressed by arbitrarily increasing the assumed measurement error, which reduces the deduced precision of flux estimates to take into account the potential for error sources not accounted for by experimentally observed scatter.[93-95*] This process is a natural consequence of the diversity and uncertainty of the metabolic architecture of different systems and is a valid form of exploratory data analysis and model building. However, altering the model by excluding specific data points and adding additional fluxes or metabolites until the χ^2-test passes, and then relying on this very same test as validation is statistically dubious from a rigorous perspective. As in the case of an FBA model validation in which the prediction being validated has been implicitly introduced to the model itself, a final validation of a 13C-MFA model with the same data used to make it acceptable, as quantified by the χ^2-test, does not constitute a real validation. It also can naturally lead to over- or under-fit models, which we discuss below in the section on model selection.

Due to these difficulties, we propose that the χ^2-test, as it is currently used, should be used as one of multiple lines of evidence to consider when validating a 13C-MFA model, especially for less defined and/or more complex eukaryotic systems such as plants. One way to address the issue of using the χ^2-test for both model development and validation is to reserve a portion of the dataset only for final model validation. This practice of holding out a subset of the data to be used exclusively for validation is standard statistical practice[41] in other areas of systems biology and, conveniently, can also be used for model selection.[93]

In the absence of direct experimentally measurable fluxes, independent measurements that can be measured or inferred from empirical measurements in vivo provide an important ground-truth value to compare with flux estimates and can complement the use of the χ^2-test for validation. An example of this can be found in the plant 13C-MFA literature, where independent measurements of the relative rates of oxygenation and carboxylation by the enzyme RuBisCO can be compared with 13C-MFA flux estimates.[94-96] In Ref. 95 for example, our group compared predicted values for the relative rates of oxygenation and carboxylation by the enzyme RuBisCO in photosynthesis versus inferred values from stomatal conductance and other empirical measurements. This led us to conclude that labeling data from whole tissue extracts was insufficient to accurately estimate photorespiratory fluxes without information on the compartmentation of certain metabolites. Despite the strength of this form of validation, it is infrequently practiced.

Another little-used but potentially valuable approach to validation is the corroboration of key features of 13C-MFA models using independent modeling methods. In Ref. 94, simplified compartmental kinetic models yielded analytical solutions predicting that overall labeling time courses should take the form of sums of exponential rate components. Fitting labeling data to these exponential models and applying statistical model selection techniques provided independent corroboration of the overall architecture of the 13C-MFA model that was used to obtain a detailed flux map.

Returning to goodness-of-fit, one must also keep in mind what information is taken into consideration and the effect of the assumed network architecture. In INST-MFA, where time-course labeling data is used, metabolite pool sizes are both estimable parameters and constrainable modeling inputs. When pool sizes are not provided as empirical measurements, pool size estimates are typically imprecise and inaccurate.[97] The inaccuracy of these estimates is not usually interpreted as an impediment to publishing 13C-MFA results and according to Ref. 97, leaving out pool size information does not adversely affect flux estimate accuracy. Flux estimates are not, however, always robust against misspecifications of the network model.[93] The exclusion of pool size information provides greater flexibility in fitting experimental data, allowing robustness against model misspecifications at the expense of not detecting them.[97] A useful next step for this field would be to routinely measure and include pool size estimates to improve the detection of incorrect model architectures. Measurement of all metabolites in a way that allows discrimination of pools for identical metabolites in different cellular compartments requires a method like Non-Aqueous Fractionation (e.g.,[98]) which may be prohibitively difficult to implement in many studies. In such cases,

---

*Here we primarily cite our own work because, as discussed, there are a number of sound reasons for leaving out metabolites and/or increasing MID measurement errors. We have chosen not to highlight other studies that have employed the same practices since we do not know all of the experimental and analytical details underlying them and would not want their inclusion here to be interpreted as implicit criticism.

use of a strategically selected set of metabolite levels may be used to allow for improved detection of incorrect model architectures. This introduces the matter of model selection.

2.3 | Model Selection in 13C-MFA

As discussed earlier, model development in 13C-MFA is an iterative process. Alternate models developed during this process may differ in their numbers of reactions and metabolites, resulting in different DOF. Adding model parameters can result in overfitting when these extra DOF lead the 13C-MFA optimization to fit noise rather than biological signal. Model selection techniques can be used to avoid this overfitting and to select the most statistically supported model among alternatives. The development of FBA models can also involve deciding between alternative architectures. However, comparison and selection of such models from sets of alternatives based on their predictions' deviations from empirical measurements is uncommon, so we focus our attention on 13C-MFA.

Model misspecification can result in missing important fluxes, incorrectly estimating the rates of modeled fluxes, or incorrectly estimating the precision of flux estimates. In a study our group performed of central metabolic fluxes in the oilseed crop *Camelina sativa*,[94] previously published model architectures that passed the χ^2-test of goodness-of-fit[95] were nonetheless shown to be missing an important set of metabolic reactions involving the movement of carbohydrates to and from the vacuole. In Ref. 93, in silico examples of sub-optimal model selection resulting in flux estimates that fall outside of the 95% confidence intervals for those same fluxes generated using the correct model architecture are provided, showing the potential for biased flux estimates when model selection is not properly performed. Finally, the literature on "Genome-scale-13C-MFA" has provided evidence that the exclusion of many reactions peripheral to the metabolic network under consideration (typically core metabolism) in 13C-MFA can result in artificially narrow confidence intervals. Genome-scale-13C-MFA involves estimating a flux map by minimizing deviation between predicted and measured isotopic labeling but using the kind of genome-scale metabolic network more typically used for FBA analyses.[99,100] In studies on the *cyanobacterium Synechococcus elongatus*,[101,102] it has been shown that the substantially larger genome-scale 13C-MFA models achieved better fits to the labeling data, that these reductions in SSR were statistically justified, and that the original models of core metabolism underestimated the uncertainty in a number of flux estimates by ignoring alternative metabolic pathways that could also explain patterns in the labeling data.[100] The examples above demonstrate that rather than being a statistical curiosity, model selection (or the lack thereof) can have serious implications for the accuracy and reliability of flux modeling results.

Several approaches to model selection can be found in the 13C-MFA literature, with different approaches being taken in different studies. The simplest is selecting the model with the smallest SSR. This method does not work when the DOF of the compared models are different, as increasing the DOF in a model inevitably allows it to fit a given data set better. This may be accounted for informally by noting the change in DOF (e.g.,[94]) or in a more statistically rigorous way using the extra-sum-of-squares test[103,104] or information criteria.[105,106] The most common model selection approach used in 13C-MFA is an informal method using the χ^2-test, wherein models are iteratively modified until a model and dataset pass the test, or where several alternative models are evaluated and the one that passes the test by the widest margin is selected.[43,44,93,107] These approaches have been used, for example, to demonstrate that the isotopic labeling data of co-culture systems cannot be adequately described by modeling with a single-culture 13C-MFA model,[108,109] to provide evidence for the operation of previously undescribed fluxes in mammalian cells,[110] and to detect missing reactions in metabolic network reconstructions from genome annotations or that are needed to describe the metabolism of mutant *E. coli* strains.[80,86]

However, the previously mentioned limitations of the χ^2-test for model validation also affect its usefulness for model selection and models failing the test due to these limitations can lead to the addition of statistically unjustified metabolites or reactions to the model until it passes.[93] We refer to the χ^2-test-based methods as "informal" model selection because when multiple models are evaluated, they are not directly or formally compared to determine whether the additional parameters in more complex models are statistically justified, which can naturally lead to the selection of overfit models.

The general approach of avoiding overfitting by evaluating models based on their performance on a set of data not used during the fitting process is widely used in statistics (e.g., cross-validation techniques[111]). The validation-based approach taken in Ref. 93 implements this best practice, separating fitting and testing data sets to avoid the pitfalls discussed above. In our view, this represents a substantial advancement in model selection in 13C-MFA. This method divides the labeling dataset into training and validation subsets and then estimates fluxes in alternative models using the training data. These alternative models' flux maps, and their accompanying predicted MIDs, are then compared based on their agreement with the validation MID data. The model whose flux map results in the smallest SSR when compared with this validation data is selected. The authors generated synthetic labeling data from a predefined "correct" model and assessed the ability of their new method and other model selection techniques to identify this correct model from a set of alternatives. The validation-based approach accomplishes this more consistently than existing model selection methods, including χ^2-test-based methods, and does so irrespective of the value of the measurement error in the labeling datasets. The incorrect models selected by other methods contain flux estimates that fall outside the 95% confidence intervals of the fluxes from the correct model, highlighting the importance of model selection for obtaining accurate flux estimates.[93] The generation of MID data in additional labeling experiments to precisely measure all fluxes in a network[28–35] provides the reserved validation datasets needed for the method described in Ref. 93. This means that for 13C-MFA studies that already require a parallel labeling approach, implementation of this more rigorous model selection approach is simply a matter of setting aside a subset of data to evaluate alternative model architectures.
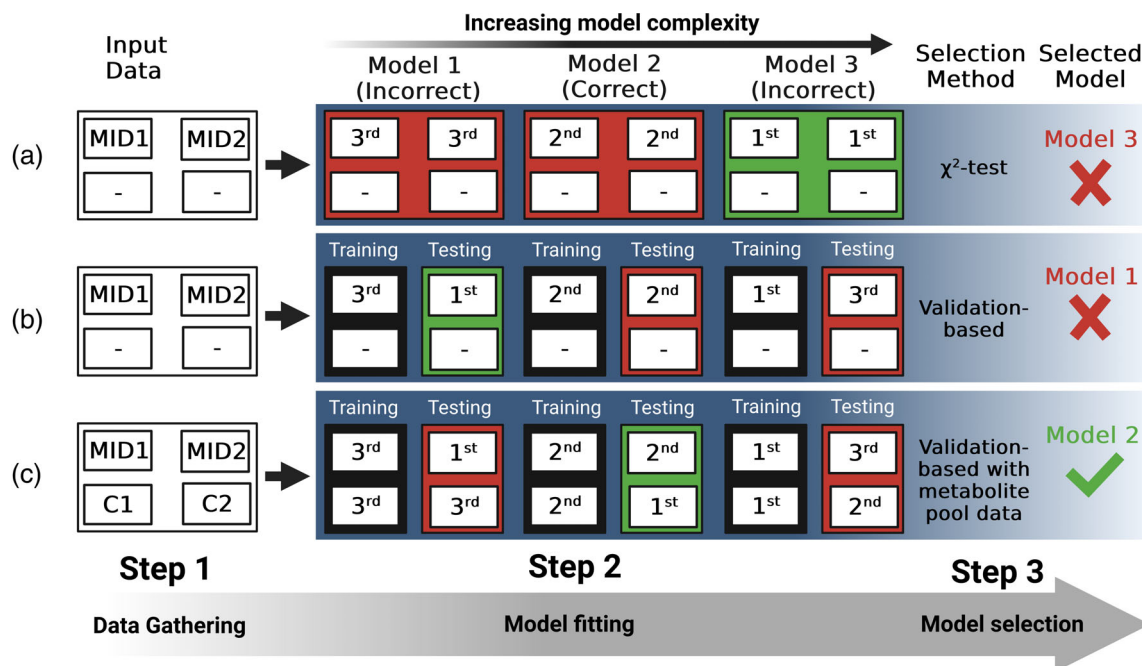
FIGURE 2   Approaches to model selection for 13C-MFA. Metabolic network models 1–3, which have increasing complexity, are compared. Model 2 in this example is the correct description of the network. (a) Labeling data (MID1 & MID2) are gathered and, for each model, agreement between model output and these data is optimized. The χ^2-test of goodness-of-fit is used to assess each model fit and these model fits are ranked 1st, 2nd, or 3rd, with the 1st passing the test by the widest margin and being selected as the most statistically well-supported model. (b) Labeling data are split into "training" and "testing" subsets and agreement between model output and the "training" data is optimized. The Sum-of-Squared Residuals (SSR) is then calculated for each model from the deviation between its output and the "testing" data. The model fits are then ranked 1st, 2nd, and 3rd, with the 1st having the lowest SSR and being selected. (c) Labeling data and metabolite pool data (C1 and C2) are gathered and split into "training" and "testing" subsets. For each model, agreement between model output and these data is optimized. The Sum-of-Squared Residuals (SSR) is then calculated for each model from the deviation between its output and the "testing" data. The model fits are then ranked 1st, 2nd, and 3rd, with the 1st having the lowest SSR and being selected. The inclusion of metabolite pool size data into both the "fitting" and "testing" datasets provides more data to go off when evaluating goodness-of-fit, potentially increasing the likelihood of identifying the correct model from a set of alternatives.

This approach can be extended in INST-MFA by using metabolite pool size measurements in the selection process. Individual pool sizes are sensitive to the local kinetic parameters and will fit poorly when reaction networks are incompletely specified.[97] We therefore suggest that validation-based model selection using pool size measurements as input measurements is a promising prospective model selection approach for INST MFA (Figure 2). Indeed, although not referred to explicitly as model selection, in Ref. 97 the authors show that inclusion of pool size information results in an incorrectly specified network architecture failing to pass the χ^2-test of goodness-of-fit, whereas a correctly specified network does pass. This corresponds to the "first to pass χ^2" method of model selection discussed in Ref. 93 and is subject to the various limitations of the χ^2-test as a model selection technique covered earlier. By incorporating these metabolite pool sizes into the formalized model selection framework described in Ref. 93, we may arrive at a more robust form of model selection that is better at detecting misspecified networks. As the authors of Ref. 93 note, the optimal model selected by their method should be subjected to a final validation to assess model quality. A model architecture may be selected by the model selection process but result in a substantial deviation of some metric from independently measured values. For this final validation, a combination of the χ^2-test, independent experimental measurements, and alternative modeling approaches can be used. Keeping in mind both the trade-off between goodness-of-fit and model complexity and the multiple ways in which 13C-MFA model predictions can be validated will ensure that flux estimates are as accurate and robust as possible.

Model validation and selection are an integral part of the 13C-MFA process. Notably, model selection practices like the use of validation-based model selection[93] and the use of the extra-sum-of-squares test[104] to compare alternative model architectures represent, in our view, a major improvement over exclusive use of the χ^2-test of goodness-of-fit test for both purposes, but are seldom practiced in the literature. We encourage the use of these techniques and believe they hold promise for improving confidence in both the fluxes and network architectures reported in studies.

With respect to validation and model selection in MFA, we recommend the following:

1. As highlighted in Ref. 44, transparency is key in 13C-MFA, given the assumptions that must be satisfied for 13C-MFA modeling as well as the sensitivity of flux estimates to model architecture.

As an example of a transparently reported 13C-MFA study, see Ref. 112.

2. The validation and selection of MFA-estimated fluxes, like the validation of any model output, benefits from multiple lines of corroborating evidence. When possible, the use of alternative modeling approaches of isotopic labeling data can be a powerful tool for arriving at well-supported model architectures, as in Ref. 94.

3. In INST-MFA, metabolite pool size measurements can be used to provide additional confidence in model validity and tighten flux confidence intervals,[113] as well as provide additional measurements for validation-based model selection. However, practitioners should be aware that these measurements can make model fits highly sensitive to incorrectly specified network models in ways that may or may not affect the accuracy of flux estimates.[97] Additionally determination of subcellular compartmentation of certain metabolites may be prohibitively difficult in some cases. In such cases, key metabolites with known subcellular compartmentation may be measured.

4. We recommend the use of a proper model selection framework to compare alternative, biochemically reasonable model architectures when performing 13C-MFA modeling. The framework outlined in Ref. 93 represents the state-of-the-art in this area. Barring the application of that method, a more traditional model selection approach, such as the extra-sum-of-squares approach used in Ref. 104 can be employed.

3 | FUTURE DIRECTIONS

We believe that validation and selection deserve greater attention from the flux analysis community and suggest that implementing the approaches highlighted in this review will improve the accuracy and reliability of constraint-based metabolic modeling and flux estimates. However, we also recognize that some approaches suggested here, such as the use of pool size measurements, can be extremely difficult to implement in practice. A recent publication on isotopically non-stationary MFA of *Arabidopsis thaliana* heterotrophic cell culture metabolism highlighted that although pool size data could potentially be used to improve the accuracy and precision of flux predictions, the experimental difficulty of measuring the concentrations of metabolites distributed across multiple subcellular compartments made this prohibitively difficult.[114] As in all areas of science, then, the development of consensus best practices in the evaluation of and inference from data and models must arise at the intersection of rigorous statistical theory and experimental practicalities. However, we believe that researchers engaged in constraint-based metabolic modeling as well as readers of modeling studies benefit when the limitations of present validation and selection practices are clarified.

Several matters call for investigation before definitive recommendations can be made on best practices. At present, it is not clear how to appropriately weight the contributions to flux estimation of unambiguous direct flux measurements like substrate uptake, which typically have relatively large standard deviations, against MIDs, which frequently have much smaller standard deviations but whose relationship to fluxes depends on model structure and whose measured values may be offset by unknown analytical effects. Likewise, it is unclear how best to deal with those not infrequent MID measurements that have extremely small, but imprecisely measured, standard deviations, which can exert too much control over the fitting process.

Finally, we would like to conclude by emphasizing that the process of careful validation and model selection can lead to the generation of models that are not only more quantitatively sound, but that yield exciting scientific insights (e.g.,[109,110]).

AUTHOR CONTRIBUTIONS
Joshua A. M. Kaste: Conceptualization; writing – original draft; investigation. **Yair Shachar-Hill:** Conceptualization; writing – review and editing; supervision.

CONFLICT OF INTEREST STATEMENT
The authors have no conflicts of interest to disclose.

PEER REVIEW
The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1002/btpr.3413.

DATA AVAILABILITY STATEMENT
Data sharing not applicable to this article as no datasets were generated or analysed during the current study.

ORCID
Joshua A. M. Kaste https://orcid.org/0000-0003-1942-0315

REFERENCES
1. Nielsen J. It is all about metabolic fluxes. *J Bacteriol*. 2003;185(24): 7031-7035.
2. Spivey A. Systems biology: the big picture. *Environ Health Perspect*. 2004;112(16):938-943.
3. Koffas MAG, Jung GY, Stephanopoulos G. Engineering metabolism and product formation in Corynebacterium glutamicum by coordinated gene overexpression. *Metab Eng*. 2003;5(1):32-41.
4. Koffas MAG, Stephanopoulos G. Strain improvement by metabolic engineering: lysine production as a case study for systems biology. *Curr Opin Biotechnol*. 2005;16(3):361-366.
5. Becker J, Zelder O, Häfner S, Schröder H, Wittmann C. From zero to hero-Design-based systems metabolic engineering of Corynebacterium glutamicum for l-lysine production. *Metab Eng*. 2011;13(2):159-168.

6. Gleizer S, Ben-Nissan R, Bar-On YM, et al. Conversion of Escherichia coli to generate all biomass carbon from CO_2. *Cell*. 2019;179(6): 1255-1263.e12.

7. Antoniewicz MR. Methods and advances in metabolic flux analysis: a mini-review. *J Ind microbiol biotech*. 2015;42(3):317-325.

8. Cheah YE, Young JD. Isotopically nonstationary metabolic flux analysis (INST-MFA): putting theory into practice. *Curr Opin Biotechnol*. 2018;54:80-87.

9. Jazmin LJ, Beckers V, Young JD. *User Manual for INCA*. 2014.

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

11. Holzhütter HG. The principle of flux minimization and its application to estimate stationary fluxes in metabolic networks. *Eur J Biochem*. 2004;271(14):2905-2922.

12. Schnitzer B, Österberg L, Cvijovic M. The choice of the objective function in flux balance analysis is crucial for predicting replicative lifespans in yeast. *PLoS One*. 2022;17(10):1-15.

13. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*. 2003;5(4):264-276.

14. Haraldsdóttir HS, Cousins B, Thiele I, Fleming RMT, Vempala S. CHRR: coordinate hit-and-run with rounding for uniform sampling of constraint-based models. *Bioinformatics*. 2017;33(11): 1741-1743.

15. Megchelenbrink W, Huynen M, Marchiori E. optGpSampler: an improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS One*. 2014;9(2):e86587.

16. Schellenberger J, Palsson B. Use of randomized sampling for analysis of metabolic networks. *J Biol Chem*. 2009;284(9):5457-5461.

17. Bordel S, Agren R, Nielsen J. Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput Biol*. 2010;6(7):16.

18. Segrè D, Vitkup D, Church GM. Analysis of optimality in natural and perturbed metabolic networks. *Proc Natl Acad Sci U S A*. 2002; 99(23):15112-15117.

19. Shlomi T, Berkman O, Ruppin E. Regulatory on/off minimization of metabolic flux changes after genetic perturbations. *Proc Natl Acad Sci U S A*. 2005;102(21):7695-7700.

20. Åkesson M, Förster J, Nielsen J. Integration of gene expression data into genome-scale metabolic models. *Metab Eng*. 2004;6(4):285-293.

21. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol*. 2008;4(5):e1000082.

22. Tian M, Reed JL. Integrating proteomic or transcriptomic data into metabolic models using linear bound flux balance analysis. *Bioinformatics*. 2018;34(22):3882-3888.

23. Pandey V, Hadadi N, Hatzimanikatis V. Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models. *PLoS Comput Biol*. 2019;15(5):1-23.

24. Ravi S, Gunawan R. ΔFBA—predicting metabolic flux alterations using genome-scale metabolic models and differential transcriptomic data. *PLoS Comput Biol*. 2021;17(11):e1009589.

25. Orth JD, Fleming RMT, Palsson BØ. Reconstruction and use of microbial metabolic networks: the Core Escherichia coli metabolic model as an educational guide. *EcoSal Plus*. 2010;4(1). doi:10.1128/ecosalplus.10.2.1

26. Antoniewicz MR, Kelleher JK, Stephanopoulos G. Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements. *Metab Eng*. 2006;8(4):324-337.

27. Theorell A, Leweke S, Wiechert W, Nöh K. To be certain about the uncertainty: Bayesian statistics for 13C metabolic flux analysis. *Biotechnol Bioeng*. 2017;114(11):2668-2684.

28. Crown SB, Long CP, Antoniewicz MR. Optimal tracers for parallel labeling experiments and (13)C metabolic flux analysis: a new precision and synergy scoring system. *Metab Eng*. 2016;38:10-18.

29. Crown SB, Long CP, Antoniewicz MR. Integrated 13C-metabolic flux analysis of 14 parallel labeling experiments in Escherichia coli. *Metab Eng*. 2015;28:151-158.

30. Leighty RW, Antoniewicz MR. COMPLETE-MFA: complementary parallel labeling experiments technique for metabolic flux analysis. *Metab Eng*. 2013;20:49-55.

31. Crown SB, Antoniewicz MR. Selection of tracers for 13C-metabolic flux analysis using elementary metabolite units (EMU) basis vector methodology. *Metab Eng*. 2012;14(2):150-161.

32. Crown SB, Ahn WS, Antoniewicz MR. Rational design of 13C-labeling experiments for metabolic flux analysis in mammalian cells. *BMC Syst Biol*. 2012;6(1):43.

33. Chang Y, Suthers PF, Maranas CD. Identification of optimal measurement sets for complete flux elucidation in metabolic flux analysis experiments. *Biotechnol Bioeng*. 2008;100(6):1039-1049.

34. Beyß M, Parra-Peña VD, Ramirez-Malule H, Nöh K. Robustifying experimental tracer design for13C-metabolic flux analysis. *Front Bioeng Biotechnol*. 2021;9:9.

35. Millard P, Sokol S, Letisse F, Portais JC. IsoDesign: a software for optimizing the design of 13C-metabolic flux analysis experiments. *Biotechnol Bioeng*. 2014;111(1):202-208.

36. Wang Y, Hui S, Wondisford FE, Su X. Utilizing tandem mass spectrometry for metabolic flux analysis. *Lab Invest*. 2021;101(4): 423-429.

37. Choi J, Antoniewicz MR. Tandem mass spectrometry for 13C metabolic flux analysis: methods and algorithms based on EMU framework. *Front Microbiol*. 2019;10:31.

38. Dinh HV, Sarkar D, Maranas CD. Quantifying the propagation of parametric uncertainty on flux balance analysis. *Metab Eng*. 2022;69: 26-39.

39. Choi YM, Choi DH, Lee YQ, et al. Mitigating biomass composition uncertainties in flux balance analysis using ensemble representations. *Comput Struct Biotechnol J*. 2023;21:3736-3745.

40. Bernstein DB, Sulheim S, Almaas E, Segrè D. Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biol*. 2021;22(1):64.

41. Gross F, MacLeod M. Prospects and problems for standardizing model validation in systems biology. *Prog Biophys Mol Biol*. 2017; 129:3-12.

42. Kirk P, Thorne T, Stumpf MPH. Model selection in systems and synthetic biology. *Curr Opin Biotechnol*. 2013;24(4):767-774.

43. Long CP, Antoniewicz MR. High-resolution (13)C metabolic flux analysis. *Nat Protoc*. 2019;14(10):2856-2877.

44. Antoniewicz MR. A guide to 13C metabolic flux analysis for the cancer biologist. *Exp Mol Med*. 2018;50(4):1-13. doi:10.1038/s12276-018-0060-y

45. Heirendt L, Arreckx S, Pfau T, et al. Creation and analysis of biochemical constraint-based models using the COBRA toolbox v.3.0. *Nat Protoc*. 2019;14(3):639-702.

46. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-based reconstruction and analysis for python. *BMC Syst Biol*. 2013;7:74.

47. Lieven C, Beber ME, Olivier BG, et al. MEMOTE for standardized genome-scale metabolic model testing. *Nat Biotechnol*. 2020;38(3): 272-276.

48. Norsigian CJ, Pusarla N, McConn JL, et al. BiGG models 2020: multistrain genome-scale models and expansion across the phylogenetic tree. *Nucleic Acids Res*. 2020;48(D1):D402-D406.

49. Varma A, Palsson BO. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110. *Appl Environ Microbiol*. 1994;60(10): 3724-3731.

50. Schroeder WL, Saha R. Introducing an optimization- and explicit Runge-Kutta- based approach to perform dynamic flux balance analysis. *Sci Rep*. 2020;10(1):1-28.

51. Arion IS, Hiroyuki O, Matti G, et al. A genome-scale metabolic model of marine heterotroph Vibrio splendidus strain 1A01. *mSystems*. 2023;28:e00377-22.

52. Feierabend M, Renz A, Zelle E, Nöh K, Wiechert W, Dräger A. High-quality genome-scale reconstruction of Corynebacterium glutamicum ATCC 13032. *Front Microbiol*. 2021;12:750206.

53. Noecker C, Sanchez J, Bisanz JE, et al. Systems biology elucidates the distinctive metabolic niche filled by the human gut microbe Eggerthella lenta. *PLoS Biol*. 2023;21(5):e3002125.

54. Blázquez B, San León D, Rojas A, Tortajada M, Nogales J. New insights on metabolic features of Bacillus subtilis based on multi-strain genome-scale metabolic modeling. *Int J Mol Sci*. 2023;24(8):7091.

55. Tec-Campos D, Posadas C, Tibocha-Bonilla JD, et al. The genome-scale metabolic model for the purple non-sulfur bacterium Rhodopseudomonas palustris bis A53 accurately predicts phenotypes under chemoheterotrophic, chemoautotrophic, photoheterotrophic, and photoautotrophic growth conditions. *PLoS Comput Biol*. 2023;19(8):e1011371.

56. Ong W, Vu TT, Lovendahl KN, et al. Comparisons of Shewanella strains based on genome annotations, modeling, and experiments. *BMC Syst Biol*. 2014;8(1):31.

57. Heinken A, Hertel J, Acharya G, et al. Genome-scale metabolic reconstruction of 7,302 human microorganisms for personalized medicine. *Nature Biotechnol*. 2023;41:1320-1331. doi:10.1038/s41587-022-01628-0

58. Heinken A, Magnúsdóttir S, Fleming RMT, Thiele I. DEMETER: efficient simultaneous curation of genome-scale reconstructions guided by experimental data and refined gene annotations. *Bioinformatics*. 2021;37(21):3974-3975.

59. Thiele I, Palsson B. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc*. 2010;5(1):93-121.

60. Coppens L, Tschirhart T, Leary DH, et al. Vibrio natriegens genome-scale modeling reveals insights into halophilic adaptations and resource allocation. *Mol Syst Biol*. 2023;27(19):e10523.

61. Pinchuk GE, Hill EA, Geydebrekht OV, et al. Constraint-based model of Shewanella oneidensis MR-1 metabolism: a tool for data analysis and hypothesis generation. *PLoS Comput Biol*. 2010;6(6):1-8.

62. Oftadeh O, Salvy P, Masid M, Curvat M, Miskovic L, Hatzimanikatis V. A genome-scale metabolic model of Saccharomyces cerevisiae that integrates expression constraints and reaction thermodynamics. *Nat Commun*. 2021;12(1):4790.

63. Burgard AP, Pharkya P, Maranas CD. OptKnock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol Bioeng*. 2003;84(6):647-657.

64. Tepper N, Shlomi T. Predicting metabolic engineering knockout strategies for chemical production: accounting for competing pathways. *Bioinformatics*. 2009;26(4):536-543.

65. Stanford NJ, Millard P, Swainston N. RobOKoD: microbial strain design for (over)production of target compounds. *Front Cell Dev Biol*. 2015;3:1-12.

66. Alzoubi D, Desouki AA, Lercher MJ. Flux balance analysis with or without molecular crowding fails to predict two thirds of experimentally observed epistasis in yeast. *Sci Rep*. 2019;9(1):1-9.

67. Gatto F, Miess H, Schulze A, Nielsen J. Flux balance analysis predicts essential genes in clear cell renal cell carcinoma metabolism. *Sci Rep*. 2015;5(1):10738.

68. Santos-Merino M, Gargantilla-Becerra Á, de la Cruz F, Nogales J. Highlighting the potential of Synechococcus elongatus PCC 7942 as platform to produce α-linolenic acid through an updated genome-scale metabolic modeling. *Front Microbiol*. 2023;14. doi:10.3389/fmicb.2023.1126030

69. Broddrick JT, Welkie DG, Jallet D, Golden SS, Peers G, Palsson BO. Predicting the metabolic capabilities of Synechococcus elongatus PCC 7942 adapted to different light regimes. *Metab Eng*. 2019;52:42-56.

70. Machado D, Herrgård M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput Biol*. 2014;10(4):e1003580.

71. Shinfuku Y, Sorpitiporn N, Sono M, Furusawa C, Hirasawa T, Shimizu H. Development and experimental verification of a genome-scale metabolic model for Corynebacterium glutamicum. *Microb Cell Fact*. 2009;8(1):43.

72. Duarte NC, Becker SA, Jamshidi N, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*. 2007;104(6):1777-1782.

73. Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BØ. A detailed genome-wide reconstruction of mouse metabolism based on human recon 1. *BMC Syst Biol*. 2010;4:140.

74. Kaste JAM, Shachar-Hill Y. Accurate flux predictions using tissue-specific gene expression in plant metabolic modeling. *Bioinformatics*. 2023;39:btad186.

75. Lee JM, Gianchandani EP, Papin JA. Flux balance analysis in the era of metabolomics. *Brief Bioinform*. 2006;7(2):140-150.

76. Lewis NE, Hixson KK, Conrad TM, et al. Omic data from evolved E. Coli are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol*. 2010;6(390).

77. Gianchandani EP, Oberhardt MA, Burgard AP, Maranas CD, Papin JA. Predicting biological system objectives de novo from internal state measurements. *BMC Bioinform*. 2008;24(9):43.

78. Yuan H, Cheung CYM, Hilbers PAJ, van Riel NAW. Flux balance analysis of plant metabolism: the effect of biomass composition and model structure on model predictions. *Front Plant Sci*. 2016;7. doi:10.3389/fpls.2016.00537

79. Chen X, Alonso AP, Allen DK, Reed JL, Shachar-Hill Y. Synergy between 13C-metabolic flux analysis and flux balance analysis for understanding metabolic adaption to anaerobiosis in E. Coli. *Metab Eng*. 2011;13(1):38-48.

80. Long CP, Antoniewicz MR. Metabolic flux responses to deletion of 20 core enzymes reveal flexibility and limits of E. Coli metabolism. *Metab Eng*. 2019;55:249-257.

81. Schuetz R, Kuepfer L, Sauer U. Systematic evaluation of objective functions for predicting intracellular fluxes inEscherichia coli. *Mol Syst Biol*. 2007;3:119.

82. Roell GW, Schenk C, Anthony WE, et al. A high-quality genome-scale model for Rhodococcus opacus metabolism. *ACS Synth Biol*. 2023;12(6):1632-1644.

83. Young JD. INCA: a computational platform for isotopically non-stationary metabolic flux analysis. *Bioinformatics*. 2014;30(9):1333-1335.

84. Weitzel M, Nöh K, Dalman T, Niedenführ S, Stute B, Wiechert W. 13CFLUX2 – high-performance software suite for 13C-metabolic flux analysis. *Bioinformatics*. 2013;29(1):143-145.

85. Shupletsov MS, Golubeva LI, Rubina SS, Podvyaznikov DA, Iwatani S, Mashko SV. OpenFLUX2: 13C-MFA modeling software package adjusted for the comprehensive analysis of single and parallel labeling experiments. *Microb Cell Fact*. 2014;13(1):1-25.

86. Au J, Choi J, Jones SW, Venkataramanan KP, Antoniewicz MR. Parallel labeling experiments validate clostridium acetobutylicum metabolic network model for (13)C metabolic flux analysis. *Metab Eng*. 2014;26:23-33.

87. Cordova LT, Antoniewicz MR. (13)C metabolic flux analysis of the extremely thermophilic, fast growing, xylose-utilizing Geobacillus strain LC300. *Metab Eng*. 2016;33:148-157.

88. Cordova LT, Cipolla RM, Swarup A, Long CP, Antoniewicz MR. (13)C metabolic flux analysis of three divergent extremely thermophilic bacteria: Geobacillus sp. LC300, Thermus thermophilus HB8, and Rhodothermus marinus DSM 4252. *Metab Eng*. 2017;44:182-190.

89. Dahle ML, Papoutsakis ET, Antoniewicz MR. 13C-metabolic flux analysis of clostridium ljungdahlii illuminates its core metabolism under mixotrophic culture conditions. *Metab Eng*. 2022;72:161-170.

90. Mitosch K, Beyß M, Phapale P, et al. A pathogen-specific isotope tracing approach reveals metabolic activities and fluxes of intracellular salmonella. *PLoS Biol*. 2023;21(8):e3002198.

91. Imada T, Yamamoto C, Toyoshima M, Toya Y, Shimizu H. Effect of light fluctuations on photosynthesis and metabolic flux in Synechocystis sp. PCC 6803. *Biotechnol Prog*. 2023;39(3):e3326.

92. Yu King Hing N, Aryal UK, Morgan JA. Probing light-dependent regulation of the Calvin cycle using a multi-omics approach. *Front Plant Sci*. 2021;12. doi:10.3389/fpls.2021.733122

93. Sundqvist N, Grankvist N, Watrous J, Mohit J, Nilsson R, Cedersund G. Validation-based model selection for 13C metabolic flux analysis with uncertain measurement errors. *PLoS Comput Biol*. 2022;18(4):e1009999.

94. Xu Y, Wieloch T, Kaste JAM, Shachar-Hill Y, Sharkey TD. Reimport of carbon from cytosolic and vacuolar sugar pools into the Calvin-Benson cycle explains photosynthesis labeling anomalies. *Proc Natl Acad Sci*. 2022;119(11):e2121531119.

95. Xu Y, Fu X, Sharkey TD, Shachar-Hill Y, Walker BJ. The metabolic origins of non-photorespiratory CO2 release during photosynthesis: a metabolic flux analysis. *Plant Physiol*. 2021;186(1):297-314.

96. Ma F, Jazmin LJ, Young JD, Allen DK. Isotopically nonstationary 13C flux analysis of changes in Arabidopsis thaliana leaf metabolism due to high light acclimation. *Proc Natl Acad Sci U S A*. 2014;111(47):16967-16972.

97. Zheng AO, Sher A, Fridman D, Musante CJ, Young JD. Pool size measurements improve precision of flux estimates but increase sensitivity to unmodeled reactions outside the core network in isotopically nonstationary metabolic flux analysis (INST-MFA). *Biotechnol J*. 2022;17(3):1-17.

98. Krueger S, Giavalisco P, Krall L, et al. A topological map of the compartmentalized Arabidopsis thaliana leaf metabolome. *PLoS One*. 2011;6(3):e17806.

99. Gopalakrishnan S, Maranas CD. 13C metabolic flux analysis at a genome-scale. *Metab Eng*. 2015;32:12-22.

100. Hendry JI, Dinh HV, Foster C, Gopalakrishnan S, Wang L, Maranas CD. Metabolic flux analysis reaching genome wide coverage: lessons learned and future perspectives. *Curr Opin Chem Eng*. 2020;30:17-25.

101. Gopalakrishnan S, Pakrasi HB, Maranas CD. Elucidation of photoautotrophic carbon flux topology in Synechocystis PCC 6803 using genome-scale carbon mapping models. *Metab Eng*. 2018;47:190-199.

102. Hendry JI, Gopalakrishnan S, Ungerer J, Pakrasi HB, Tang YJ, Maranas CD. Genome-scale Fluxome of Synechococcus elongatus UTEX 2973 using transient 13C-labeling data. *Plant Physiol*. 2019;179(2):761-769.

103. Draper NR, Smith H. Extra sums of squares and tests for several parameters being zero. *Applied Regression Analysis*. John Wiley & Sons, Ltd; 1998:149-177. doi:10.1002/9781118625590.ch6

104. Boyle NR, Sengupta N, Morgan JA. Metabolic flux analysis of heterotrophic growth in Chlamydomonas reinhardtii. *PLoS One*. 2017;12(5):1-23.

105. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Parzen E, Tanabe K, Kitagawa G, eds. *Selected Papers of Hirotugu Akaike*. Springer New York; 1998:199-213. doi:10.1007/978-1-4612-1694-0_15

106. Schwarz G. Estimating the dimension of a model. *Ann Statist*. 1978;6(2):461-464.

107. Dalman T, Wiechert W, Nöh K. A scientific workflow framework for 13C metabolic flux analysis. *J Biotechnol*. 2016;232:12-24.

108. Gebreselassie NA, Antoniewicz MR. (13)C-metabolic flux analysis of co-cultures: a novel approach. *Metab Eng*. 2015;31:132-139.

109. Wolfsberg E, Long CP, Antoniewicz MR. Metabolism in dense microbial colonies: (13)C metabolic flux analysis of E. Coli grown on agar identifies two distinct cell populations with acetate cross-feeding. *Metab Eng*. 2018;49:242-247.

110. Ahn WS, Crown SB, Antoniewicz MR. Evidence for transketolase-like TKTL1 flux in CHO cells based on parallel labeling experiments and (13)C-metabolic flux analysis. *Metab Eng*. 2016;37:72-78.

111. Hastie T, Tibshirani R, Friedman J. Model Assessment and Selection. *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. Springer; 2009:219-259.

112. Nicolae A, Wahrheit J, Bahnemann J, Zeng AP, Heinzle E. Non-stationary 13C metabolic flux analysis of Chinese hamster ovary cells in batch culture using extracellular labeling highlights metabolic reversibility and compartmentation. *BMC Syst Biol*. 2014;8(1):50.

113. Nöh K, Grönke K, Luo B, Takors R, Oldiges M, Wiechert W. Metabolic flux analysis at ultra short time scale: isotopically non-stationary 13C labeling experiments. *J Biotechnol*. 2007;129(2):249-267.

114. Smith EN, Ratcliffe RG, Kruger NJ. Isotopically non-stationary metabolic flux analysis of heterotrophic Arabidopsis thaliana cell cultures. *Front Plant Sci*. 2022;13:1049559.