Metabolic Engineering 74 (2022) 72–82

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

Interrogating the effect of enzyme kinetics on metabolism using differentiable constraint-based models

St. Elmo Wilken [a,b,*], Mathieu Besançon [c], Miroslav Kratochvíl [d], Chilperic Armel Foko Kuate [a], Christophe Trefois [d], Wei Gu [d], Oliver Ebenhöh [a,b]

[a] Institute of Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225, Düsseldorf, Germany
[b] Cluster of Excellence on Plant Sciences, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225, Düsseldorf, Germany
[c] Department for AI in Society, Science, and Technology, Zuse Institute Berlin, Takustraße 7, 14195, Berlin, Germany
[d] Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Campus Belval, L-4367, Belvaux, Luxembourg

ARTICLE INFO

Keywords:
Metabolic control analysis
Flux balance analysis
Constraint-based models
Differentiation
Optimization
Enzyme constrained

ABSTRACT

Metabolic models are typically characterized by a large number of parameters. Traditionally, metabolic control analysis is applied to differential equation-based models to investigate the sensitivity of predictions to parameters. A corresponding theory for constraint-based models is lacking, due to their formulation as optimization problems. Here, we show that optimal solutions of optimization problems can be efficiently differentiated using constrained optimization duality and implicit differentiation. We use this to calculate the sensitivities of predicted reaction fluxes and enzyme concentrations to turnover numbers in an enzyme-constrained metabolic model of *Escherichia coli*. The sensitivities quantitatively identify rate limiting enzymes and are mathematically precise, unlike current finite difference based approaches used for sensitivity analysis. Further, efficient differentiation of constraint-based models unlocks the ability to use gradient information for parameter estimation. We demonstrate this by improving, genome-wide, the state-of-the-art turnover number estimates for *E. coli*. Finally, we show that this technique can be generalized to arbitrarily complex models. By differentiating the optimal solution of a model incorporating both thermodynamic and kinetic rate equations, the effect of metabolite concentrations on biomass growth can be elucidated. We benchmark these metabolite sensitivities against a large experimental gene knockdown study, and find good alignment between the predicted sensitivities and *in vivo* metabolome changes. In sum, we demonstrate several applications of differentiating optimal solutions of constraint-based metabolic models, and show how it connects to classic metabolic control analysis.

1. Introduction

Biological processes are typically characterized by a large number of parameters. In the context of constraint-based models, these can include enzyme kinetics and thermodynamic constants. Databases of *in vitro* measurements (Flamholz et al., 2012; Chang et al., 2021) organize and catalogue decades of research, allowing these parameters to be included in modern models (Noor et al., 2014; Adadi et al., 2012; Henry et al., 2007). Incorporating parameters into constraint-based models has been shown to increase their predictive capabilities (Sánchez et al., 2017; Zhou et al., 2021; Wilken et al., 2021; Li et al., 2021). For example, classic flux balance analysis is incapable of modeling overflow metabolism without ad hoc assumptions (Beg et al., 2007), but if enzyme turnover numbers and capacity limitations are incorporated, this phenomenon can be mechanistically modeled (de Groot et al., 2020).

Recently, *in vivo* estimates of enzyme turnover numbers have become available through integrating omics measurements with models (Davidi et al., 2016). These estimates substantially increase the quantitative accuracy of predictions afforded by enzyme-constrained metabolic models (Chen and Nielsen, 2021a). Machine learning techniques have also made use of these omics driven measurements to extend kinetic parameter estimates, including turnover numbers and Michaelis constants, to the genome-scale (Heckmann et al., 2018; Kroll et al., 2021). With these advancements it is becoming feasible to construct increasingly detailed metabolic models (O'Brien et al., 2013; Goelzer et al., 2015). However, an unaddressed question is the sensitivity of model

* Corresponding author. Institute of Quantitative and Theoretical Biology, Heinrich-Heine-Universität Düsseldorf, Universitätsstraße 1, 40225, Düsseldorf, Germany.
E-mail address: wilkenst@hhu.de (St.E. Wilken).

https://doi.org/10.1016/j.ymben.2022.09.002
Received 11 July 2022; Received in revised form 8 September 2022; Accepted 10 September 2022
Available online 21 September 2022

predictions to parameters.

Metabolic control analysis (MCA) encompasses a rich theory for estimating parameter sensitivities in ordinary differential equation (ODE) based models (Heinrich and Schuster, 2012). However, a corresponding theory for constraint-based models is lacking. In the latter case, a finite difference based procedure is typically used to estimate the effect of a perturbed parameter on an optimal solution relative to a reference solution(Nilsson and Nielsen, 2016; Tsouka et al., 2021; Domenzain et al., 2022). Consequently, for each parameter, a new optimization problem needs to be solved. This suggests that the time required to map the complete sensitivity grows with the product of the optimizer solve time and the number of parameters. In contrast, MCA implicitly differentiates the ODE-based model, yielding all the sensitivities directly after solving a single system of linear algebraic equations.

Here, we show that an optimal solution of a constraint-based model can be likewise implicitly differentiated, and demonstrate several applications. Calculating the so-called flux control coefficients (Nilsson and Nielsen, 2016) simplifies to finding the derivatives of the optimization variables to parameters, analogously to classic MCA. Moreover, the ability to differentiate a constraint-based metabolic model allows it to be efficiently embedded in more complex optimization schemes. We use this to solve a bilevel optimization problem, which seeks to fit apparent turnover numbers to an enzyme-constrained model by minimizing the difference between model predictions and observations. Finally, we show that the method is easily extendable to problems of arbitrary complexity by comparing the predicted sensitivity of intracellular metabolite concentrations to biomass growth in a model that incorporates both thermodynamic constraints, as well as Michaelis-Menten reaction kinetics.

2. Results

A brief note on nomenclature: one dimensional variables are denoted like x, vectors like $\mathbf{x}$, and matrices like $\mathbf{X}$. Further, qualitative models are shown in Section 3, but the full models are shown in Section 5.

2.1. Extending metabolic control analysis to constraint-based models

Metabolic control analysis (MCA) is a quantitative technique widely used to gauge the sensitivity of variables to parameters in dynamical systems (Hatzimanikatis and Bailey, 1996; Heinrich and Schuster, 2012). Briefly, suppose a biochemical process is governed by a system of differential equations,

$$\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}t} = \mathbf{S}\cdot\mathbf{v}(\mathbf{s}(\mathbf{p}), \mathbf{p}), \tag{1}$$

which relate reaction fluxes ($\mathbf{v}$) to metabolite concentrations ($\mathbf{s}$) through parameters ($\mathbf{p}$), using the stoichiometric matrix ($\boldsymbol{S}$). We can define the function $\mathbf{f}_{\mathrm{SS}}$,

$$\mathbf{f}_{\mathrm{SS}}(\mathbf{s}(\mathbf{p}), \mathbf{p}) := \mathbf{S}\cdot\mathbf{v}(\mathbf{s}(\mathbf{p}), \mathbf{p}), \tag{2}$$

and at steady state,

$$\mathbf{f}_{\mathrm{SS}}(\mathbf{s}(\mathbf{p}), \mathbf{p}) = \mathbf{0}, \tag{3}$$

which implies an implicit relationship between $\mathbf{p}$, and $\mathbf{s}$. Consequently, implicit differentiation of Equation (3) yields,

$$\begin{aligned} &\frac{\partial \mathbf{f}_{\mathrm{SS}}}{\partial \mathbf{s}}\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\mathbf{p}} + \frac{\partial \mathbf{f}_{\mathrm{SS}}}{\partial \mathbf{p}} = \mathbf{0} \\ &\mathbf{S}\frac{\partial \mathbf{v}}{\partial \mathbf{s}}\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\mathbf{p}} + \mathbf{S}\frac{\partial \mathbf{v}}{\partial \mathbf{p}} = \mathbf{0} \\ &\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\mathbf{p}} = -\left(\mathbf{S}\frac{\partial \mathbf{v}}{\partial \mathbf{s}}\right)^{-1}\mathbf{S}\frac{\partial \mathbf{v}}{\partial \mathbf{p}}. \end{aligned} \tag{4}$$

Thus, $\frac{\mathrm{d}\mathbf{s}}{\mathrm{d}\mathbf{p}}$ represents the sensitivity of the variables to parameters at steady state, and from here the classic flux and concentration control coefficients, as well as elasticities, can be calculated (Villadsen et al., 2011). Taken together, these sensitivities are calculated by applying the implicit function theorem to a steady state solution of some governing differential equation. Moreover, if the derivatives of the constituent functions are known, then finding the sensitivities only require the solution of a single system of linear algebraic equations.

2.1.1. Constraint-based metabolic control analysis

Typically, constraint-based models without integer variables can be expressed as either convex linear or convex quadratic programs (Orth et al., 2010), which have the general form shown in Problem (P1),

$$\begin{aligned} \mathbf{z}^* = \quad & \underset{\mathbf{z}}{\operatorname{argmin}} && \frac{1}{2}\mathbf{z}^T\mathbf{Q}(\mathbf{p})\mathbf{z} + \mathbf{q}(\mathbf{p})^T\mathbf{z} \\ & \text{s. t.} && \mathbf{S}(\mathbf{p})\mathbf{z} = \mathbf{d}(\mathbf{p}) \\ & \mathbf{M}(\mathbf{p})\mathbf{z} \leq \mathbf{h}(\mathbf{p}), \end{aligned} \tag{P1}$$

where $\mathbf{Q}$ is a positive semidefinite matrix. Here, $\mathbf{z}^*$ represents an optimum of the variables (often fluxes) given parameters ($\mathbf{p}$), which can include enzyme turnover numbers, Michaelis constants, etc. The dependence of the optimization program on parameters is usually not shown, but here it has been made explicit. Normally, $\mathbf{S}$ represents the stoichiometric matrix, $\mathbf{d} = \mathbf{0}$, and $\mathbf{M}$, $\mathbf{h}$ are bounds placed on the variables to enforce physiological constraints (like reaction directions, flux bounds, etc.). For convenience, define,

$$\begin{aligned} f(\mathbf{z}, \mathbf{p}) &:= \frac{1}{2}\mathbf{z}^T\mathbf{Q}(\mathbf{p})\mathbf{z} + \mathbf{q}(\mathbf{p})^T\mathbf{z} \\ \mathbf{g}(\mathbf{z}, \mathbf{p}) &:= \mathbf{S}(\mathbf{p})\mathbf{z} - \mathbf{d}(\mathbf{p}) \\ \mathbf{k}(\mathbf{z}, \mathbf{p}) &:= \mathbf{M}(\mathbf{p})\mathbf{z} - \mathbf{h}(\mathbf{p}). \end{aligned} \tag{5}$$

Optimization theory holds that for a convex problem, like Problem (P1), the stationarity, primal feasibility and complementary slackness conditions (first-order Karush-Kuhn-Tucker conditions), shown in Equation (6), are both necessary and sufficient for optimality (Boyd et al., 2004),

$$\begin{aligned} \partial_z f(\mathbf{z}^*, \mathbf{p})^T + \partial_z \mathbf{g}(\mathbf{z}^*, \mathbf{p})^T\boldsymbol{\nu} + \partial_z \mathbf{k}(\mathbf{z}^*, \mathbf{p})^T\boldsymbol{\lambda} &= \mathbf{0} \\ \mathbf{g}(\mathbf{z}^*, \mathbf{p}) &= \mathbf{0} \\ \operatorname{diag}(\boldsymbol{\lambda})\mathbf{k}(\mathbf{z}^*, \mathbf{p}) &= \mathbf{0}. \end{aligned} \tag{6}$$

Here, $\boldsymbol{\nu}$ and $\boldsymbol{\lambda}$ are the equality and inequality dual variables (generalizations of Lagrange multipliers), and diag(·) denotes the diagonal matrix formed from a vector. To highlight the implicit dependence of the optimum variables on the parameters, one may rewrite the optimality conditions by substituting the definitions of Equation (5) into Equation (6), and define a new function, $\mathbf{f}$, which resembles $\mathbf{f}_{\mathrm{SS}}$,

$$\mathbf{f}(\mathbf{z}(\mathbf{p})^*, \mathbf{p}) := \begin{cases} \mathbf{Q}(\mathbf{p})\mathbf{z}^* + \mathbf{q} + \mathbf{S}(\mathbf{p})^T\boldsymbol{\nu} + \mathbf{M}(\mathbf{p})^T\boldsymbol{\lambda} \\ \mathbf{S}(\mathbf{p})\mathbf{z}^* - \mathbf{d}(\mathbf{p}) \\ \operatorname{diag}(\boldsymbol{\lambda})(\mathbf{M}(\mathbf{p})\mathbf{z}^* - \mathbf{h}(\mathbf{p})). \end{cases} \tag{7}$$

Thus, at the optimum,

$$\mathbf{f}(\mathbf{z}(\mathbf{p})^*, \mathbf{p}) = \mathbf{0}. \tag{8}$$

This suggests that implicitly differentiating $\mathbf{f}$ will yield the derivatives of the optimization variables to their parameters (Gould et al., 2016; Amos and Kolter, 2017; Blondel et al., 2021; Sharma et al., 2022). Specifically, the similarity between $\mathbf{f}_{\mathrm{SS}}$ in Equation (3) and $\mathbf{f}$ in Equation (8), reveals the connection point between classic MCA and its constraint-based model counterpart. By implicit differentiation of the optimality condition function,

$$\frac{\partial \mathbf{f}}{\partial \mathbf{z}^*}\frac{\mathrm{d}\mathbf{z}^*}{\mathrm{d}\mathbf{p}}+\frac{\partial \mathbf{f}}{\partial \mathbf{p}}=\mathbf{0}$$
$$\frac{\mathrm{d}\mathbf{z}^*}{\mathrm{d}\mathbf{p}}=-\left(\frac{\partial \mathbf{f}}{\partial \mathbf{z}^*}\right)^{-1}\frac{\partial \mathbf{f}}{\partial \mathbf{p}}, \tag{9}$$

the sensitivity of the variables at the optimum to model parameters, $\frac{\mathrm{d}\mathbf{z}^*}{\mathrm{d}\mathbf{p}}$, can be found. Moreover, similar to the classic MCA approach, if the derivatives of the constituent functions are known, then the constraint-based metabolic control analysis (CB-MCA) procedure is computationally efficient, because only a single linear system of equations needs to be solved (subsequent to the optimization) to calculate *all* the derivatives. Encouragingly, recent developments in automatic and symbolic differentiation allow these constituent function derivatives to be calculated efficiently for arbitrary problems automatically (Moses et al., 2020; Revels et al., 2016; Gowda et al., 2022).

2.2. Metabolic control analysis of an enzyme-constrained metabolic model

Constraint-based models represent a scalable framework to interrogate metabolism. Incorporating enzyme capacity and rate limitations into these models unlocks the ability to mechanistically explain resource allocation phenomena, including overflow metabolism (de Groot et al., 2020). Various frameworks have been proposed to model enzyme limitations (Adadi et al., 2012; Sánchez et al., 2017; Bekiaris and Klamt, 2020). Qualitatively, Problem (P2) summarizes their main features,

$$\begin{aligned}\max_{\mathbf{v},\mathbf{e}}\quad & \mu(\mathbf{v},\mathbf{e})\\ \text{s. t.}\quad & \mathbf{Sv}=0\\ & v_n=k_{\mathrm{cat},n}\cdot e_n\ \forall\ n \text{ metabolic reactions}\\ & \sum_n e_n\le E_{\mathrm{total}}.\end{aligned} \tag{P2}$$

Importantly, these models hold that the flux, v, through each reaction is proportional to the enzyme concentration, e, that catalyzes the reaction, and its associated turnover number, k_{cat}. Additionally, the total proteome capacity is limited by E_{total}, and μ represents the biomass objective function. Typically, the turnover numbers are taken as constants, often inferred from databases (Domenzain et al., 2022) or estimated (Davidi et al., 2016; Heckmann et al., 2018). With the CB-MCA framework introduced earlier, it is possible to efficiently estimate the sensitivity of the predicted intracellular fluxes and enzyme concentrations to the turnover numbers.

Fig. 1 shows the predicted reaction fluxes and enzyme concentrations of *Escherichia coli* growing aerobically on glucose in minimal media using the GECKO formulation (Sánchez et al., 2017) of Problem (P2), together with turnover numbers for all the metabolic enzymes, generated using a machine learning based approach (Heckmann et al., 2020).

The sensitivities of both the reaction fluxes (Fig. 2 A), as well as the enzyme concentrations (Fig. 2 B), to the turnover numbers can be found by differentiating the optimal solution, shown in Fig. 1. To prevent differentiability issues caused by the potential non-invertibility of $\frac{\partial \mathbf{f}}{\partial \mathbf{z}^*}$, only active reactions (those that are not dashed in Fig. 1) and expressed enzymes are differentiated (see Section 4 for more details).

In Fig. 2 A, the flux sensitivities indicate that cytochrome oxidase (*CYTBO3_4pp*), pyruvate dehydrogenase (*PDH*), and ATP synthase (*ATPS4rpp*) exert the most control over the optimal solution.

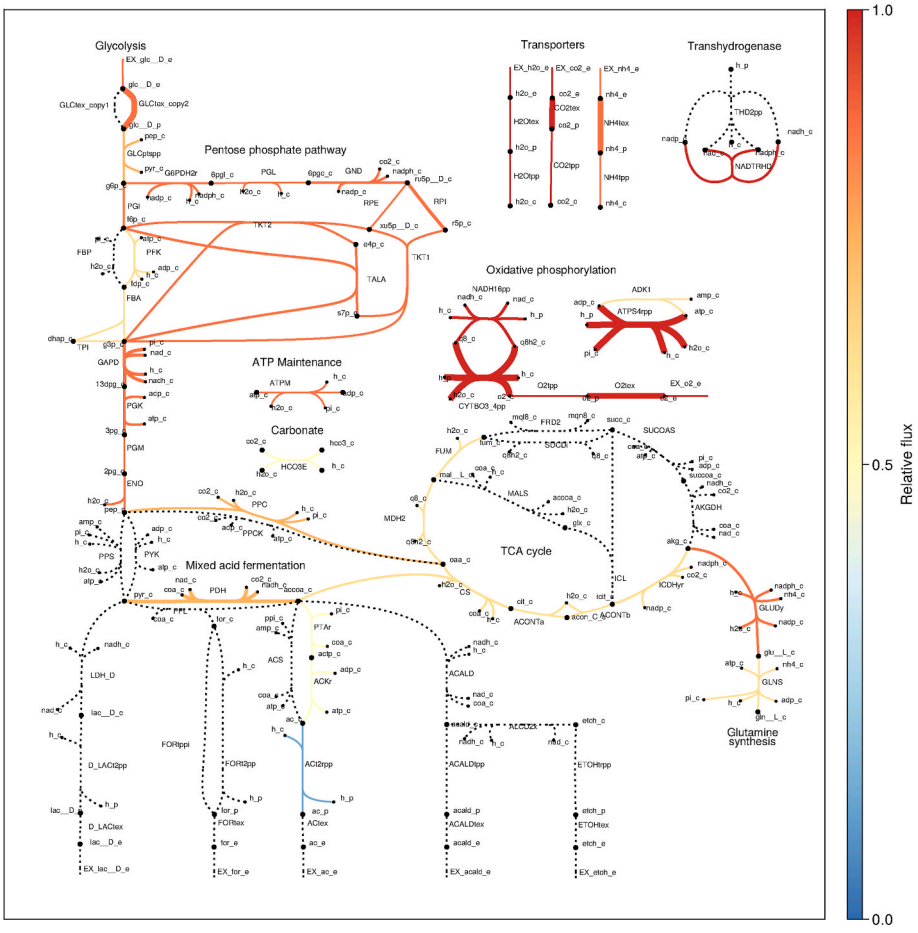

Fig. 1. Enzyme-constrained metabolic models can predict intracellular fluxes and protein concentrations. The core carbon metabolism of *E. coli* growing aerobically on glucose in minimal media is shown here. Enzyme limitations were incorporated into its latest metabolic model (Monk et al., 2017) using the GECKO formulation (Sánchez et al., 2017), and turnover numbers were taken from a genome-wide dataset (Heckmann et al., 2020). Reaction fluxes are colored according to their relative absolute magnitude, and reaction edge width is scaled relative to the predicted enzyme concentration. Metabolically inactive reactions are dashed. This metabolic realization (fluxes and enzyme concentrations) is differentiated in Fig. 2.

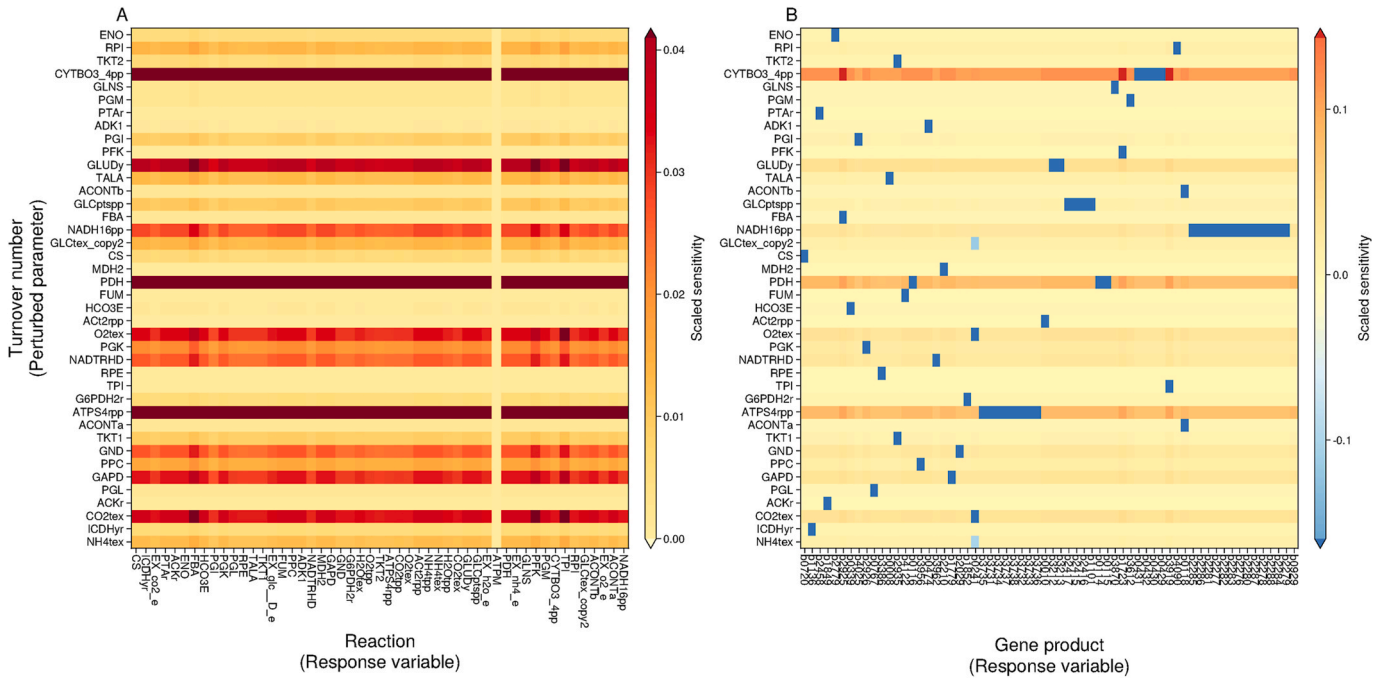

Fig. 2. Scaled sensitivities of predicted reaction fluxes and gene product concentrations to enzyme turnover numbers. *E. coli* model iML1515 was simulated under aerobic, glucose fed conditions, and subsequently the active solution, shown in Fig. 1, was differentiated to yield these sensitivities. The sensitivity of variable z_i to parameter p_j shown here is scaled, $\frac{\mathrm{dln}(z_i)}{\mathrm{dln}(p_j)}$. For simplicity, only enzymes and reaction fluxes in the core carbon metabolism are shown here.

Quantitatively, increasing their associated turnover numbers will increase flux across all active reactions, except the maintenance reaction (*ATPM*). Fig. 2 B, which shows the sensitivities of the predicted gene product concentrations to the turnover numbers, illustrates why this is expected. Taking cytochrome oxidase as an example, the sensitivities of its subunits to its turnover number are negative, while all other gene products have a positive sensitivity to the turnover number of cytochrome oxidase. If the turnover number for cytochrome oxidase increases, less enzyme is required to catalyze the same amount of flux, consequently, more enzyme can be distributed to other reactions while satisfying the enzyme capacity bound. Consequently, the flux through metabolism increases. In contrast, the maintenance reaction is an ATP sink, and increasing flux through it can only decrease the objective function. Hence, when biomass is maximized it is forced to its lower bound, and thus its sensitivity is zero to all turnover numbers. Comparing the model-based flux sensitivity of an enzyme to its turnover number against the measured absolute protein abundance of the respective enzyme under the same conditions, we find a positive linear relationship ($R^2 = 0.51$, see Fig. S1). Interestingly, an enzyme constrained model focussing on the energy metabolism of *E. coli* predicted that doubling the turnover numbers of pyruvate dehydrogenase (PDH), enolase (ENO), glyceraldehyde phosphate dehydrogenase (GAPD), or ATP synthase (ATPS4rpp) would have the most impact on growth rate (Chen and Nielsen, 2019). Experimental evidence was found for PDH (Miles et al., 1988), GAPD (Cho et al., 2012), and ATPS4rpp (Ahmad and Brudecki, 2010), which correspond to enzymes with high flux control coefficients as identified in Fig. 2. In contrast, enolase has a comparatively low flux control coefficient, indicating some disagreement in predictions.

Beyond identifying flux controlling enzymes, the sensitivities shown in Fig. 2 also indicate the relative flux increase a reaction would experience if the associated turnover number were to be increased. However, unlike classic MCA, the summation theorem does not have a clear analog in the constraint-based case. Despite this difference, the sensitivities shown here are *exactly* the same as the finite difference based sensitivities, called flux control coefficients, introduced in other work (Nilsson and Nielsen, 2016; Domenzain et al., 2022) (see Fig. S2). Finally, inspecting all the sensitivities (see Fig. S3) shows that most parameters have a small scaled flux sensitivity ($\leq 10^{-2}$), suggesting that precise turnover numbers for all metabolic enzymes are likely not critically important.

2.3. Differentiating an enzyme-constrained metabolic model to fit turnover numbers to measured data

Incorporating enzyme turnover numbers and capacity limitations into flux constraints has been shown to dramatically increase the predictive capabilities of constraint-based models (Chen and Nielsen, 2021a). Unfortunately, it is experimentally infeasible to clone and express each enzyme a cell possesses for individual characterization, in part due to the large number of protein coding genes contained in even minimal-sized microbes. Recent approaches to tackle this characterization problem for enzyme turnover numbers include machine learning (Heckmann et al., 2018, 2020), and constraint-based metabolic modeling (Davidi et al., 2016) techniques. The former method makes use of a variety of enzyme specific training data (crucially, including measured kinetic data), and is capable of predicting enzyme turnover numbers at the genome-scale for a specific organism, albeit through a black-box technique. The latter method couples omics measurements to a mechanistic model, and can directly estimate the apparent turnover numbers of expressed enzymes. Despite the usefulness of these *in vivo* estimated turnover numbers, only *Saccharomyces cerevisiae* (Chen and Nielsen, 2021a) and *Escherichia coli* (Heckmann et al., 2018) have been subject to such genome-scale based approaches.

In the constraint-based model approach to estimate apparent turnover numbers, flux balance analysis (FBA) *without enzyme constraints* is used to estimate the internal reaction fluxes. Measured fluxes are used to constrain the model to increase the accuracy of the predicted fluxes. Apparent turnover numbers are then estimated by,

$$k_{\mathrm{cat,\ app}} = \frac{v}{\hat{e}}, \tag{10}$$

where $\hat{e}$ is the intracellular measured protein concentration of the enzyme catalyzing the relevant reaction. By sampling many metabolic

conditions, the turnover number (the theoretical maximum catalytic rate of the enzyme) can be estimated by taking the maximum of all the apparent turnover numbers. The primary drawback of this approach is that quantitative proteomic measurements do not span the entire metabolism, limiting which reaction turnover numbers (typically homomeric enzymes) can be estimated in this way.

Fundamentally, conventional FBA is used instead of an enzyme-constrained variant, because once enzyme kinetics, intracellular fluxes, and enzyme concentrations are incorporated into a model as variables, it becomes a nonlinear optimization problem, which is challenging to solve at the genome-scale. A practical workaround for this issue is to set the kinetic constants, which need to be estimated, as parameters, and only optimize over the fluxes and enzyme concentrations. For a given set of turnover numbers, $\mathbf{k}_{\text{cat}}$, the mean squared relative error between the model predicted intracellular fluxes and enzyme concentrations, and their measured counterparts, can be minimized, as shown qualitatively in Problem (P3),

$$
\begin{aligned}
L(\mathbf{k}_{\text{cat}}) = \min_{\mathbf{v},\mathbf{e}} \quad & \frac{1}{|I|}\sum_{i\in I}\left(1-\frac{v_i}{\widehat{v}_i}\right)^2 + \frac{1}{|J|}\sum_{j\in J}\left(1-\frac{e_j}{\widehat{e}_j}\right)^2 \\
\text{s. t.} \quad & \mathbf{S}\mathbf{v} = 0 \\
& v_n = k_{\text{cat},n}\cdot e_n \;\forall\; n \text{ metabolic reactions} \\
& \sum_n e_n \le E_{\text{total}},
\end{aligned}
\tag{P3}
$$

where index sets I and J indicate measured variables. The benefit of this approach is that both missing fluxes *and* missing enzyme concentrations can be imputed by solving the model. This broadens the scope of apparent turnover numbers that can be estimated by using the model-based approach. Additionally, this optimization problem is a convex quadratic program with only diagonal Hessian terms, for which highly efficient methods and corresponding solvers exist (IBM ILOG Cplex, 2021). All that remains is to change $\mathbf{k}_{\text{cat}}$ to minimize the error function, L, in Problem (P3), to find the turnover numbers that best fit the measured data.

A computationally tractable approach to reduce the error induced by inaccurate turnover number estimates is to perform gradient descent on L, which requires that its derivatives can be efficiently calculated. Thus, by using the differentiation techniques introduced earlier, the bilevel optimization problem of finding the set of turnover numbers minimizing L subject to Problem (P3) can be solved. The optimum corresponds to the turnover numbers that minimize the discrepancy between model predictions and measured data.

Using measured intracellular fluxes and protein concentrations of *E. coli* under various culture conditions (Heckmann et al., 2020), this technique can be tested. In Panel A of Fig. 3, the mean squared relative error, L, is shown over multiple gradient descent iterations. Starting from the turnover numbers estimated by the machine learning approach, the algorithm successfully changes them to reduce the error. Panel B shows the resultant turnover numbers for enzymes in glycolysis, and Panel C their derivatives. We observe that both derivatives and errors became smaller over iterations, indicating that the turnover numbers converge.

Subsequently, for each culture condition dataset, the turnover numbers associated with the lowest loss gradient descent iterate was taken as the dataset specific turnover numbers estimate. Subsequently, in keeping with previous work (Davidi et al., 2016), the maximum over all the dataset specific turnover number estimates was taken as the best estimate of the underlying enzyme turnover numbers. Panel A in Fig. 4 shows that using these improved turnover number estimates, compared to the state-of-the-art machine learning generated estimates, increases the accuracy of model predictions against experimental measurements by 35 ± 2%. Comparing the improved turnover number estimates to a subset of those found in BRENDA (Davidi et al., 2016), the coefficient of determination (R^2) is 0.37. The final set of improved turnover number estimates is listed in Supplementary Dataset 1.

2.4. Differentiable metabolic models can account for enzyme saturation and thermodynamic effects

While the aforementioned enzyme constraints represent physiologically realistic simplifications, enzyme kinetics are typically much more complex *in vivo*. Regulatory, thermodynamic, and saturation effects can dramatically alter the maximum rate at which an enzyme can metabolize its substrate (Noor et al., 2013). In particular, metabolite concentrations directly affect the thermodynamic driving force, and enzyme active site occupancy (Bennett et al., 2008). Recently, CRISPRi was used

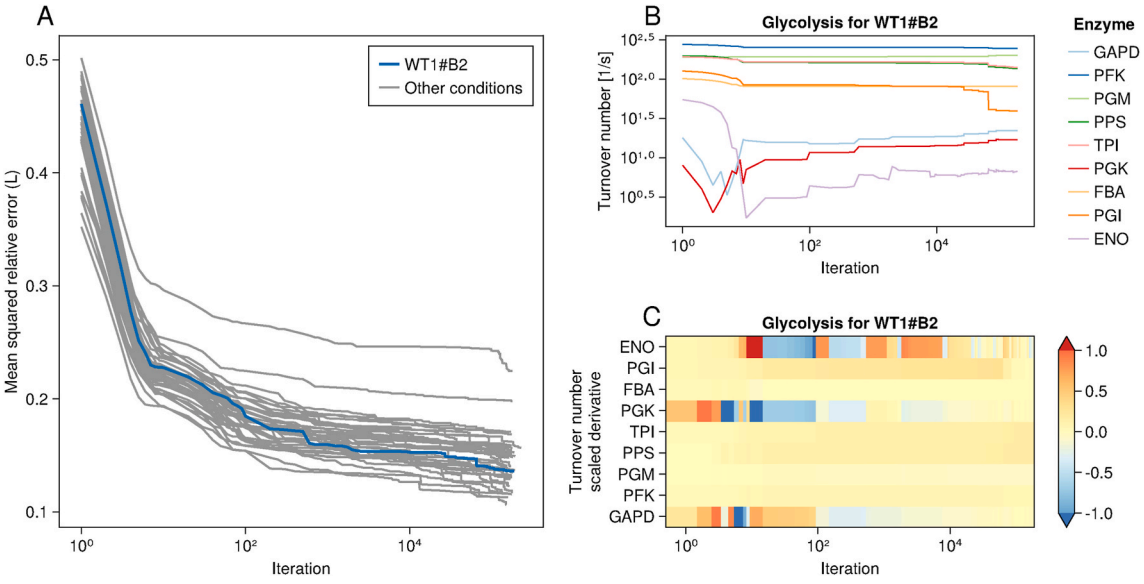

Fig. 3. Turnover number estimates can be improved by minimizing the difference between observations and model predictions. Panel A highlights the mean squared relative error of a specific culture condition in a published dataset (WT1#B2: wild type cell, condition replicate 1, technical replicate 2), the other conditions are shown in grey. Panel B shows how the turnover number estimates in glycolysis changes over gradient descent iterations, subject to the derivatives in Panel C.

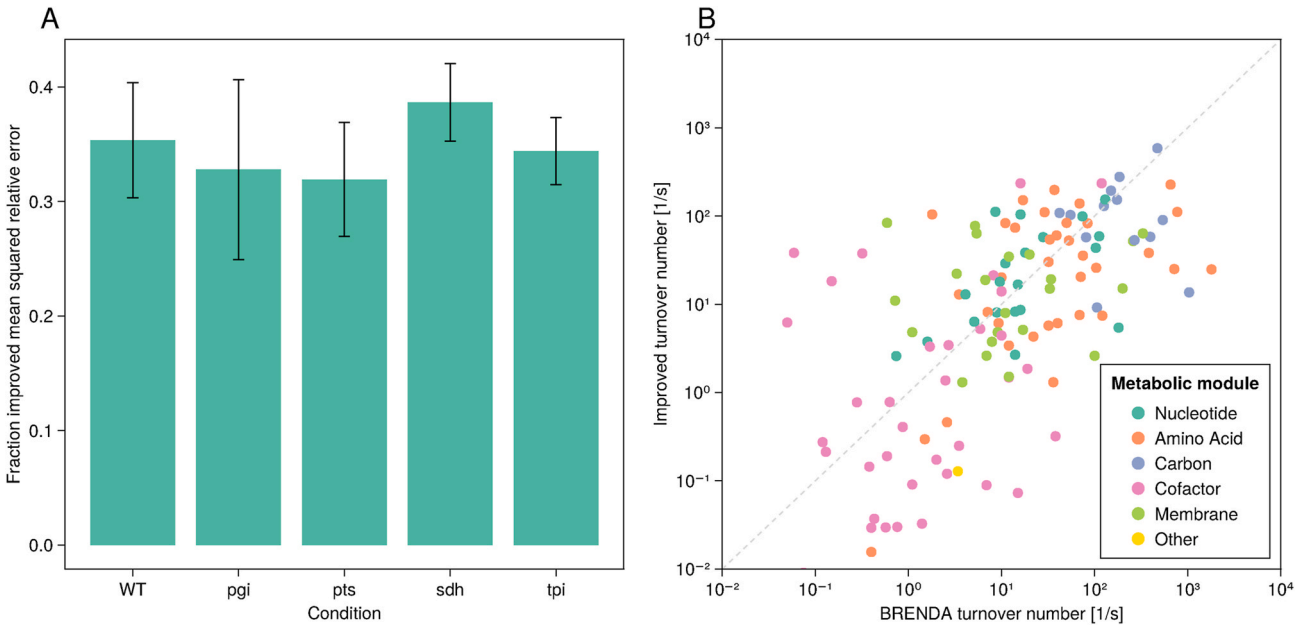

Fig. 4. Improving the machine learning generated turnover number estimates with the differentiable modeling approach causes the accuracy of enzyme-constrained model predictions to increase. Panel A shows the fractional increase in accuracy achieved when using the improved turnover number estimates compared to their state-of-the-art machine learning generated counterparts to predict intracellular fluxes and enzyme concentrations using the latest *E. coli* metabolic model under various genetic knockout conditions (datasets) (WT: wild type, pgi: glucose-6-phosphate isomerase, pts: glucose phosphotransferase system, sdh: succinate dehydrogenase, tpi: triose-phosphate isomerase). To make the comparison fair, for each test condition its data is left out of the dataset used by the gradient descent technique (hold-out method). However, the published machine learning data is used as is, which incorporates all the available data, making the test more conservative against the gradient descent technique. Error bars reflect the standard deviation of the error function, L, for all the experimental replicates in each condition. Panel B shows a comparison of the improved turnover numbers to those found in BRENDA for a curated list of enzymes (Davidi et al., 2016), grouped by the metabolic module they occur in.

to knockdown gene expression, reducing the concentration of various enzymes in *E. coli* with the goal of understanding how the metabolome responds to enzyme limitations. The experimental data showed that the substrates of the enzyme that was throttled tended to increase in concentration, unless another form of regulation (e.g. allosteric, etc.) was available (Donati et al., 2021). Thus, by altering its metabolome, *E. coli* compensated for the knockdowns, and the deleterious effect on growth rate was reduced.

Incorporating metabolite concentrations as variables into constraint-based models is challenging, because thermodynamic and saturation effects are inherently nonlinear (Noor et al., 2016). Typically, either fluxes (e.g., in max-min driving force analysis (Noor et al., 2014)) or metabolite concentrations (e.g., in flux balance analysis (Orth et al., 2010)) are abstracted away into the model. This approach precludes the ability to investigate the sensitivity of metabolite concentration on flux and enzyme concentration predictions. To address this, we extend the simple enzyme-constrained model given in Problem (P2) to include metabolite concentrations as parameters, as shown in Problem (P4),

$$
\begin{aligned}
&\min_{\mathbf{v},\mathbf{e}} \quad \mu(\mathbf{v},\mathbf{e}) \\
&\text{s.t.} \quad \mathbf{S}\mathbf{v} = 0 \\
&v_n = k_{\text{cat},n} \cdot e_n \cdot \frac{\prod_i \left(\frac{s_{s,i}}{K_{\text{M},n,i}}\right)^{\nu_{n,i}}}{1 + \prod_i \left(\frac{s_{s,i}}{K_{\text{M},n,i}}\right)^{\nu_{n,i}} + \prod_i \left(\frac{s_{p,i}}{K_{\text{M},n,i}}\right)^{\nu_{n,i}}} \cdot \left(1 - \exp\left(\frac{\Delta_r G_n}{RT}\right)\right) \\
&\Delta_r G_n = \Delta_r G_n^0 + RT\left(\sum_i \log s_{s,i}^{\nu_{n,i}} - \sum_i \log s_{p,i}^{\nu_{n,i}}\right) \\
&\sum_n e_n \le E_{\text{total}}.
\end{aligned}
\tag{P4}
$$

Here, $K_{\text{M},n,i}$ and $\nu_{n,i}$ denote the Michaelis constant and the stoichiometric coefficient of reaction n and metabolite i, respectively. Further, $s_{s,i}$ or $s_{p,i}$ indicate the substrate or product concentration of metabolite i relative to the associated reaction, and $\Delta_r G_n^0$ is the standard Gibbs free energy change of reaction n (Noor et al., 2013). This model accounts for both saturation and thermodynamic effects, but ignores regulation. Since the metabolite concentrations are taken as parameters (estimated as described in Section 5), the problem remains convex, and is computationally tractable.

Through the addition of Michaelis-Menten like kinetics in Problem (P4), we hypothesize that the experimentally observed metabolome changes can be understood by investigating the sensitivity of the biomass function to intracellular metabolite concentrations under knockdown conditions. Specifically, metabolites with high sensitivity are likely to exert larger control on the biomass function, suggesting that they could be used to counteract the effect of the knockdowns. Thus, we test if the model can recapitulate the observed trend that substrates to the throttled enzyme increase in concentration, by evaluating if the substrate metabolites have high sensitivity.

Fig. 5 shows these sensitivities for a representative selection of knockdown conditions (all other conditions are shown in Fig. S4, but are broadly similar to the ones shown here). Each simulated knockdown condition constrains the corresponding gene concentration to be five-fold less than it would be under wild-type conditions. Subsequently, the model in Problem (P4) is simulated, and differentiated. The resulting sensitivities of the biomass function to intracellular metabolites are compared to experimentally measured metabolite concentration fold changes subject to the same knockdowns (Donati et al., 2021).

The simulated sensitivity results, shown in Fig. 5, indicate that the substrate and product metabolites of the throttled enzyme have the largest sensitivities. The sensitivity of most other intracellular metabolites is small ($\le 10^{-6}$). This suggests that by increasing the substrate concentration (positive sensitivity), or decreasing the product

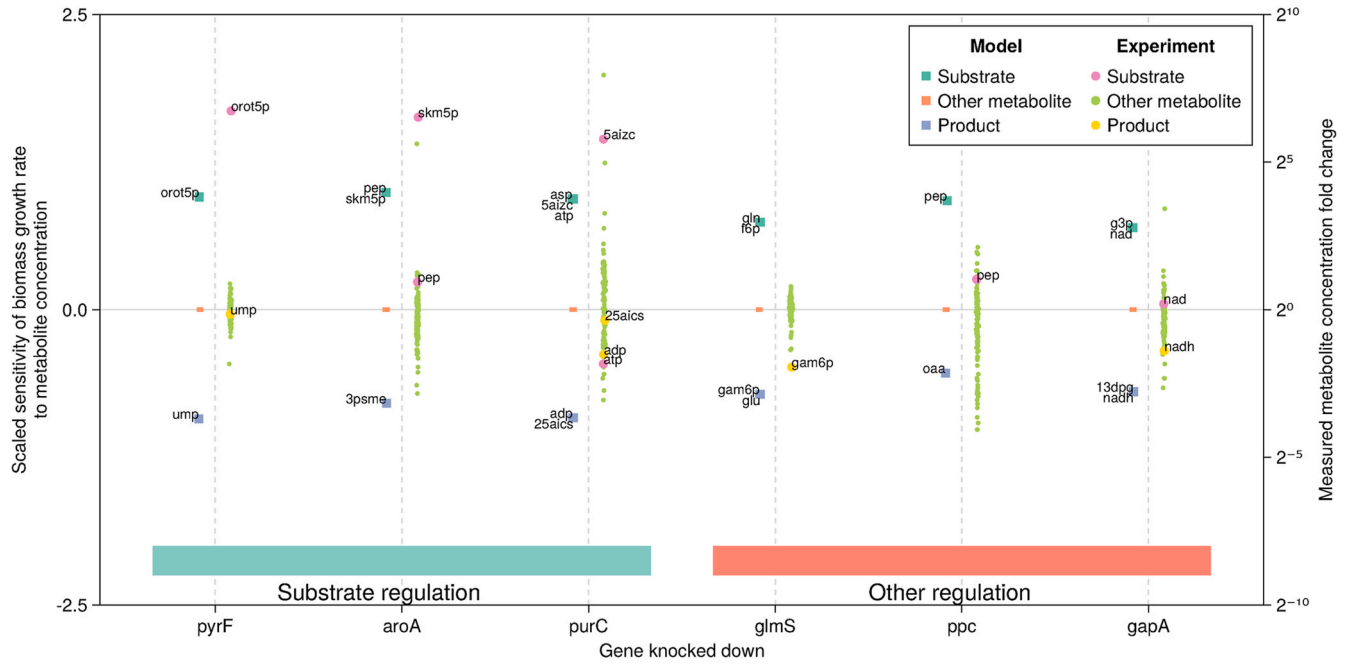

Fig. 5. Predicted biomass sensitivity to metabolites aligns with experimentally measured metabolome changes under genetic perturbations. Each gene knockdown has two datasets associated with it, separated by a vertical dotted line. On the left y-axis, the simulated sensitivities of the biomass function to metabolite concentrations is shown after the associated gene is constrained to be five-fold less abundant than the wild type system. Substrate and product metabolite sensitivities are highlighted. On the right y-axis, experimental data of intracellular metabolite concentration fold changes observed after CRISPRi knockdowns of the associated genes are shown (Donati et al., 2021). Substrate and product metabolites are also highlighted if measured. The gene abundance constraints experimentally observed match those simulated *in silico*. The groupings denoted by the horizontal blocks near the bottom of the figure separate the primary compensating mechanism observed *in vivo*. Briefly, 'Substrate regulation" denotes enzymes where the measured substrate metabolites played a large role in controlling the kinetics. "Other regulation" denotes cases where substrate concentration played a smaller role, e.g. allosteric regulation compensated for the knockdown, etc. Each gene catalyzes the following reactions: Orotidine-5'-phosphate decarboxylase (*pyrF*), 3-phosphoshikimate 1-carboxyvinyltransferase (*aroA*), Phosphoribosylaminoimidazolesuccinocarboxamide synthase (*purC*), Glutamine-fructose-6-phosphate transaminase (*glmS*), Phosphoenolpyruvate carboxylase (*ppc*), Glyceraldehyde-3-phosphate dehydrogenase (*gapA*).

concentration (negative sensitivity), of the throttled enzyme, the flux through the biomass reaction can be increased. Thus, the model recapitulates the experimental observations, although it is agnostic between product or substrate metabolites. Specifically, for enzymes where reagent metabolites compensate for enzyme limitations (e.g., *pyrF*, *aroA*, *purC*), the sign of the sensitivities aligns with the observed metabolome changes.

However, since Problem (P4) does not include regulatory mechanisms, it is incapable of accounting for allosteric control, as observed in *glmS* and *ppc* (allosterically regulated through glucosamine-phosphate, and aspartate and malate, respectively) (Donati et al., 2021). Instead, the model predicts that the substrate and product metabolites would compensate for the knockdown. Similarly, for *gapA*, where the compensation mechanism is unclear (Donati et al., 2021), the model predicts saturation compensation, because this is the only explanatory mechanism built into the mathematical structure of the model.

3. Discussion

3.1. Constraint-based metabolic control analysis

Classic metabolic control analysis provides a precise description of the sensitivities of variables to parameters in ODE-based models. However, these models are exceedingly challenging to parameterize at the genome-scale, limiting their applicability to smaller systems (Foster et al., 2021). Constraint-based models make use of substantially fewer parameters, which widens their scope. However, they are cast as optimization problems, superficially occluding the direct application of MCA to them. Here, we show that precisely the same mathematical technique, as is used in ODE-based models, can be used to calculate the sensitivities of variables in constraint-based models to their parameters. This dramatically improves upon the finite difference based approach currently used (a single parameter is perturbed and the difference between a reference and response solution is used to estimate sensitivities). Moreover, by making the definition of the derivatives precise, metabolic degeneracy must be dealt with explicitly, clarifying the meaning of the resultant sensitivities.

In particular, models making use of flux balance analysis, including their enzyme-constrained variants, invariably need to sample different solutions to form a complete picture of the metabolic potential of an organism (Schellenberger and Palsson, 2009). Fundamentally, this is because they are typically solved through linear programming, which does not guarantee a unique solution. This presents problems for finding meaningful derivatives, as the variables used in the derivative calculations are not guaranteed to be unique. In the finite difference based approach, this effect is ignored. However, the implicit differentiation technique introduced here reveals that the Jacobian, $\frac{\partial \mathbf{f}}{\partial \mathbf{z}}$, is singular in these cases, because there are linearly dependent (degenerate) metabolic states. Unless special steps are taken to ensure a unique basis forms the optimum objective, the Jacobian will not be invertible, and thus the sensitivities are not well defined. In general, a small quadratic regularizing term can be added to the linear program, turning it into a quadratic program, with uniqueness guarantees on both the objective and variables. While this transforms it into a strongly convex program that can be differentiated (Parikh and Boyd, 2014), the cost is solving a more complicated problem. Alternatively, by solving the model and removing metabolically inactive variables (i.e. removing the metabolic degeneracy), a unique optimum basis can be found for enzyme constrained problems (de Groot et al., 2020; Müller et al., 2014). Subsequently, the Jacobian is full rank, and the solution may be differentiated. In sum, either procedure results in well defined derivatives, making their subsequent interpretation simpler (see Section S1 for more details).

In light of the foregoing, different metabolic realizations (satisfying the same optimality condition) can yield different sensitivities (demonstrated in Fig. S5). While most reactions display small variations in sensitivity, some metabolic modules, including the tricarboxylic acid cycle and nucleotide metabolism, can be substantially different. This highlights the importance of preferring the implicit differentiation approach: it ensures that the resulting sensitivities are well-defined by forcing the reference solution to be defined prior to differentiation. The clear connection to classic MCA and the associated derivatives suggests deeper connections between theoretical advances and specific models are possible. For example, marginal fitness cost, the derivative of protein cost to metabolite concentration as introduced in growth balance analysis (Dourado and Lercher, 2020), can be simply extended to arbitrary metabolic models within this framework, as the necessary derivatives can now be calculated automatically.

3.2. Estimating enzyme turnover numbers

Enzyme kinetic databases collect data gathered over decades, and can be used for model parameterization (Chang et al., 2021; Wittig et al., 2012). However, these *in vitro* estimates are typically noisy (Bar-Even et al., 2011) and *in vivo* generated data tends to produce better predictions when used in models (Heckmann et al., 2018; Chen and Nielsen, 2021a). Somewhat paradoxically, while Fig. 4 confirms that significant improvements in model predictive power can be achieved by using tuned, organism specific turnover number estimates, the magnitude of most of the flux control coefficients is relatively small (see Fig. S3). This suggests that the effect of most individual turnover numbers on overall predictions is also relatively small, but their total contribution is significant. Despite the clear benefits of using *in vivo* data to estimate enzyme kinetics, challenges remain in generating accurate parameter estimates across all metabolically active enzymes.

Conceptually, it is possible to *only* use measured intracellular fluxes and protein abundances to estimate turnover numbers. However, the primary drawback of this approach is that proteomic and intracellular flux measurements do not span the entire metabolism. Thus, measurement gaps need to be filled, motivating the use of a model. In previous work, FBA was used to find missing fluxes, with the shortcoming that missing enzyme concentrations could not be imputed (Davidi et al., 2016). Subsequently, machine learning was used to estimate the remaining turnover numbers. In this work, the differentiability of an enzyme-constrained model was leveraged to refine the machine learning estimates (which were used as the initial conditions for the gradient descent algorithm). Measurement gaps were imputed by the model for all metabolically active enzymes. A caveat to this approach is that the bilevel optimization problem being solved is nonconvex, suggesting multiple solutions may exist. Different initial conditions may give rise to different sets of improved turnover number estimates, but the effect size seems to be moderate, possibly because the space of turnover numbers associated with low model error is relatively small (see Fig. 6A). Likewise, the variability in improved turnover numbers estimated from different datasets are relatively small (see Fig. S6B). While randomly sampled starting points also converged to low errors, using the machine learning estimated turnover numbers as starting points tended to yield the lowest error over the gradient descent iterations (see Fig. S7). This suggests that the machine learning approach generated good initial estimates. Comparing the final improved turnover number estimates to their machine learning counterparts reveals that the former are higher on average (see Fig. S8). Since the improved turnover numbers increased the predictive accuracy of the model, it indicates that the machine learning approach underestimated the true turnover numbers.

The bilevel algorithm introduced here imputes missing enzyme concentrations while descending on the associated turnover numbers. Consequently, there is no clear upper bound for the associated turnover numbers of the missing enzymes. Thus, it is possible for the algorithm to increase the turnover numbers of these enzymes to physiologically unrealistic values to minimize the amount of enzyme required to catalyze the predicted fluxes. Practically, this is not too concerning because the objective function of the inner problem merely seeks to fit data, and not e.g., maximize growth (the latter objective could more easily lead to unrealistically large turnover number estimates). Moreover, the gradient descent algorithm changes the turnover numbers of all the enzymes simultaneously, thus the imputed enzyme concentrations effectively get changed relative to the initial estimates, attenuating dramatic changes in turnover number estimates of single reactions (see Fig. S9). Thus, given the increasing number of machine learning generated turnover number estimates (Heckmann et al., 2018; Li et al., 2022), this algorithm is well suited as a refinement step to increase their accuracy.

While data-intensive, this approach promises to unlock the ability to rapidly estimate enzyme kinetics at the genome-scale for any organism in a relatively short time. This addresses the pressing need for kinetic parameters spanning a wider selection of organisms (Chen and Nielsen, 2021b). The model based technique has the advantage of being a transparent process, making the fitting procedure conceptually easier to understand, although it cannot estimate turnover numbers of enzymes that are not expressed, unlike the machine learning-based approach. Consequently, only around half of the metabolic reactions were assigned turnover number estimates using the gradient descent-based approach (see Supplementary Dataset 1).

3.3. Sensitivity analysis of a complex model

Given the broad applicability of constraint-based models, we demonstrate that calculating their sensitivities is straightforward, even when incorporating thermodynamic and kinetic constraints. Such a model was used to determine the sensitivities of intracellular metabolite concentrations to biomass growth. This model was partially able to recapitulate measured trends in metabolite changes after genetic perturbations, likely hampered by regulatory details that were not modeled. Indeed, the model was incapable of predicting allosteric regulation, because the kinetic rate equations did not incorporate this phenomenon. However, saturation effects were captured by this model, suggesting that the approach taken here revealed physiologically realistic control levers cells use to modulate intracellular fluxes.

The technique used to simplify the nonlinear model shown in Problem (P4) can also be applied to metabolic and expression (ME) models (O'Brien et al., 2013). These detailed models are usually non-convex, and bilinear in fluxes and growth rate (Yang et al., 2016). By fixing growth rate at a specific value, these models become convex and linear, unlocking the ability to differentiate them as demonstrated in this work.

In light of the procedure used to estimate enzyme turnover numbers *in vivo*, the foregoing results raise the question: can a similar gradient descent-based technique be used to find Michaelis constants or standard Gibbs free energy of reactions at the genome-scale? Recent strides in metabolomics allow broad intracellular metabolite concentrations to be measured in addition to fluxes and protein abundances (Ishii et al., 2007; Park et al., 2016). This suggests that a similar model-based fitting procedure could be used to estimate kinetic parameters beyond just turnover numbers. A potential stumbling block for this would be the nonlinearity introduced by setting metabolite concentrations as variables (opposed to parameters as used in this work), which could make

optimization challenging. Nevertheless, the ability to efficiently differentiate the resultant solution could be leveraged to estimate elusive kinetic parameters.

In sum, differentiable constraint-based models provide a systematic basis to explore the effect parameters have on metabolic phenotypes in constraint-based models. We have shown the link between classic MCA, and its constraint-based counterpart, CB-MCA. Further, we demonstrate several applications where the differentiability of optimization problems can help refine parameter estimates, or elucidate the response of their perturbation on models. Increasingly complex metabolic models are being developed, which usually depend on a wide variety of parameters. Sensitivity analysis could help justify the inclusion of theoretical mechanisms in these models, accelerating the pace of systematically understanding cellular phenomena.

4. Materials and methods

All software and worflows used to generate the results in this work can be accessed here https://gitlab.com/qtb-hhu/differentiablemetabolismcode. In this section a high level summary is presented; see the readme in the software repository for more detailed information.

4.1. Model construction, parameter sources, and software used

The latest *E. coli* metabolic model, iML1515 (Monk et al., 2017), was used in all simulations. The core biomass function was optimized, unless otherwise noted (e.g. for the gradient descent in Section 3.3). The total enzyme mass capacity limitation was determined by counting the number of metabolically active reactions in a simulation (under the relevant conditions), and extrapolating a linear fit of the cumulative enzyme mass from measured data vs. the number of enzymes measured. This yielded capacity bounds in the range $0.31 \pm 0.03 \frac{g}{g_{DW}}$ depending on the number of active reactions in the system. To estimate the robustness of this approach, we also used a coarser capacity bound of $0.32 \frac{g}{g_{DW}}$ independent of the number of active reactions (Schmidt et al., 2016), but the results did not meaningfully change. Enzyme subunit stoichiometry was determined through Uniprot (UniProt Consortium, 2019) and Complex Portal (Meldal et al., 2022) annotations. Molar masses of proteins were taken from Uniprot. Turnover numbers were taken from (Heckmann et al., 2020), Michaelis constants from (Kroll et al., 2021), and thermodynamic data from (Flamholz et al., 2012). All simulations were performed using COBREXA.jl (Kratochvíl et al., 2022), a modern Julia Language (Bezanson et al., 2017) constraint-based analysis package. CPLEX (IBM ILOG Cplex, 2021) was used as the optimization solver. The models of Problems (P2), (P3), and (P4) were differentiated as described in the text. The specific model format is described below. Gradient descent was performed on Problem (P3) using an approximate, backtracking line search algorithm to find an appropriate step size. For Problem (P4), metabolite concentrations were estimated by minimizing the effect thermodynamics and saturation has on each enzyme, as described later. Visualizations were done using Makie (Danisch and Krumbiegel, 2021).

4.2. Enzyme-constrained metabolic model formulation

Enzyme capacity and rate limitations represent a physiological constraint that shapes the resource allocation in a cell. Briefly, assuming that saturation and thermodynamic factors are negligible, the flux (v) through an enzyme catalyzed reaction may be modeled by,

$$v = k_{\text{cat}} \cdot e, \tag{11}$$

where k_{cat} is the turnover number of the associated enzyme, with concentration e. Multiple (related) algorithms incorporate this idea into constraint-based models (Adadi et al., 2012; Sánchez et al., 2017; Bekiaris and Klamt, 2020). Here, we focus on the GECKO formulation (Sánchez et al., 2017), but the results are generalizable to the other model variants.

For GECKO, the reactions of a constraint-based model are split into their forward and reverse components, so that all fluxes are positive. Additionally, each isozyme is also modeled as a distinct reaction. These unidirectional reactions are coupled in the optimization problem, so that the original reaction fluxes can be reconstructed from the newly created variables. By largely following the existing formulation, the resultant model is,

$$\begin{aligned} \max_{\mathbf{v},\mathbf{e}} \quad & \mu(\mathbf{v},\mathbf{e}) \\ \text{s. t.} \quad & \begin{bmatrix} \mathbf{S} & \mathbf{0} \\ \mathbf{E}(\mathbf{p}) & -\mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \mathbf{0} \\ & \mathbf{c}_{\text{LB}} \le \mathbf{C} \begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix} \le \mathbf{c}_{\text{UB}} \\ & \begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix}_{\text{LB}} \le \begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix} \le \begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix}_{\text{UB}}, \end{aligned} \tag{P5}$$

where the enzyme turnover numbers enter $\mathbf{E}$ as parameters ($\mathbf{p}$), and the coupling matrix $\mathbf{C}$ allows the original bounds of the model to constrain the split reactions through $\mathbf{c}_{\text{LB}}$ and $\mathbf{c}_{\text{UB}}$. Additionally, enzyme capacity bounds can also be incorporated through $\mathbf{C}$. Typically, the objective is some linear combination of fluxes ($\mathbf{v}$) and enzyme concentrations ($\mathbf{e}$), expressed through the function μ. Although each term in Equation (P5) could be a function of the parameters, for simplicity we assume that only $\mathbf{E}$ takes parameters.

For models that are purely enzyme-constrained, $\mathbf{E}$ is a sparse matrix, composed of terms,

$$E_{i,j} = \frac{a_i}{k_{\text{cat},j}}, \tag{12}$$

where the columns (j) correspond to enzyme-constrained reactions, and rows (i) are pseudo-protein mass balances (exactly like the GECKO formulation). Here a_i is the stoichiometric coefficient of the protein subunit catalyzing the associated reaction.

For models that incorporate thermodynamic and enzyme constraints, $\mathbf{E}$ is again a sparse matrix, but now composed of terms,

$$E_{i,j} = \frac{a_i}{k_{\text{cat},j} \cdot \frac{\prod_n \left(\frac{s_{s,n}}{K_{\text{M},j,n}}\right)^{\nu_{j,n}}}{1+\prod_n \left(\frac{s_{s,n}}{K_{\text{M},j,n}}\right)^{\nu_{j,n}} + \prod_n \left(\frac{s_{p,n}}{K_{\text{M},j,n}}\right)^{\nu_{j,n}}} \cdot \left(1 - \exp\left(\frac{\Delta_r G_j}{RT}\right)\right)} \tag{13}$$

where the meaning of all the terms are the same as used throughout this work.

These problems can be converted into the standard form of Problem (P1), and then differentiated with respect to the associated parameters (most simply using symbolic or automatic differentiation).

4.3. Estimating intracellular concentrations

In Problem (P4), intracellular metabolite concentrations need to be

supplied as parameters. In this work, we assumed that intracellular metabolite concentrations would be adjusted by the cell to ensure that enzymes work at maximum efficiency relative to saturation limitations. Briefly, the metabolite concentrations were estimated by minimizing the maximum impact of the saturation terms on enzyme velocity for all enzymes with Michaelis constant data, as shown in Problem (P6),

$$\begin{aligned} \max_{\mathbf{s}_s,\mathbf{s}_p,\alpha} \quad & \alpha \\ \text{s. t.} \quad & \alpha \le \frac{\prod_i \left(\frac{s_{s,i}}{K_{\mathrm{M},n,i}}\right)^{\nu_{n,i}}}{1+\prod_i \left(\frac{s_{s,i}}{K_{\mathrm{M},n,i}}\right)^{\nu_{n,i}} + \prod_i \left(\frac{s_{p,i}}{K_{\mathrm{M},n,i}}\right)^{\nu_{n,i}}} \ \forall\ n \text{ reactions} \\ & \beta \ge \Delta_r G_n^0 + RT\left(\sum_i \log s_{s,i}^{\nu_{n,i}} - \sum_i \log s_{p,i}^{\nu_{n,i}}\right) \\ & s_{\mathrm{LB}} \le \left(\mathbf{s}_s, \mathbf{s}_p\right) \le s_{\mathrm{UB}} \end{aligned} \tag{P6}$$

where β was set to $-2\,\frac{\mathrm{kJ}}{\mathrm{mol}}$ and represents a lower bound on the driving force of each reaction. Additionally, $\mathbf{s}_s$, $\mathbf{s}_p$ are the substrate and product concentrations relative to each reaction. All metabolite concentrations were bounded by $s_{\mathrm{LB}} = 10^{-9}$ M and $s_{\mathrm{UB}} = 0.1$ M. This problem is nonlinear and nonconvex, a local solution was computed using KNITRO (Byrd et al., 2006), using the metabolite concentrations from max-min driving force analysis (Noor et al., 2014) as starting points.

Author contributions

St. Elmo Wilken: Conceptualization, Data curation, Software, Methodology, Investigation, Writing - Original draft preparation. **Mathieu Besançon**: Methodology, Writing - review & editing. **Miroslav Kratochvíl**: Software, Writing - review & editing. **Chilperic Armel Foko Kuate**: Validation, Writing - review & editing. **Christophe Trefois**: Funding acquisition, Writing - review & editing. **Wei Gu**: Funding acquisition, Writing - review & editing. **Oliver Ebenhöh**: Validation, Funding acquisition, Writing - review & editing.

Conflicts of interest

None declared.

Data availability

Data will be made available on request.

Acknowledgements

This work was partially supported through the Research Campus Modal funded by the German Federal Ministry of Education and Research (fund numbers 05M14ZAM, 05M20ZBM) (MB). This work was also partially funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy — EXC-2048/1 — project ID 390686111 (SEW, OE). This work was also partially supported by the European Union's Horizon 2020 Programme under the PerMedCoE Project (www.permedcoe.eu) [951773] (MK, CT, WG). This work was also partially funded by the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie Actions Grant Agreement PoLiMeR, No 812616(CAFK).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ymben.2022.09.002.

References

Adadi, Roi, Volkmer, Benjamin, Milo, Ron, Heinemann, Matthias, Shlomi, Tomer, 2012. Prediction of microbial growth rate versus biomass yield by a metabolic network with kinetic parameters. PLoS Comput. Biol. 8 (7), e1002575.

Ahmad, Zulfiqar, Brudecki, Laura E., 2010. Molecular modulation of the alpha-subunit visit-dg sequence in the catalytic sites of escherichia coli atp synthase. Faseb. J. 24, 463-1.

Amos, Brandon, Kolter, J Zico, 2017. Optnet: differentiable optimization as a layer in neural networks. International Conference on Machine Learning. PMLR, pp. 136–145.

Bar-Even, Arren, Noor, Elad, Savir, Yonatan, Liebermeister, Wolfram, Davidi, Dan, Tawfik, Dan S., Milo, Ron, 2011. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. Biochemistry 50 (21), 4402–4410.

Beg, Qasim K., Vazquez, Alexei, Ernst, Jason, Menezes, Marcio A de, Bar-Joseph, Ziv, L Barabási, A.-, Oltvai, Zoltán N., 2007. Intracellular crowding defines the mode and sequence of substrate uptake by escherichia coli and constrains its metabolic activity. Proc. Natl. Acad. Sci. 104 (31), 12663–12668.

Bennett, Bryson D., Yuan, Jie, Kimball, Elizabeth H., Rabinowitz, Joshua D., 2008. Absolute quantitation of intracellular metabolite concentrations by an isotope ratio-based approach. Nat. Protoc. 3 (8), 1299–1311.

Bezanson, Jeff, Alan Edelman, Karpinski, Stefan, Shah, Viral B., 2017. Julia: a fresh approach to numerical computing. SIAM Rev. 59 (1), 65–98.

Blondel, Mathieu, Berthet, Quentin, Cuturi, Marco, Roy, Frostig, Hoyer, Stephan, Llinares- López, Felipe, Pedregosa, Fabian, Vert, Jean-Philippe, 2021. Efficient and Modular Implicit Differentiation. Arxiv preprint arxiv:2105.15183.

Boyd, Stephen, Boyd, Stephen P., Vandenberghe, Lieven, 2004. Convex Optimization. Cambridge university press.

Byrd, Richard H., Nocedal, Jorge, Waltz, Richard A., 2006. Knitro: an integrated package for nonlinear optimization. Large-scale Nonlinear Optimization. Springer, pp. 35–59.

Chang, Antje, Jeske, Lisa, Ulbrich, Sandra, Hofmann, Julia, Koblitz, Julia, Ida, Schomburg, Neumann-Schaal, Meina, Jahn, Dieter, Schomburg, Dietmar, 2021. Brenda, the elixir core data resource in 2021: new developments and updates. Nucleic Acids Res. 49 (D1), D498–D508.

Chen, Yu, Nielsen, Jens, 2019. Energy metabolism controls phenotypes by protein efficiency and allocation. Proc. Natl. Acad. Sci. 116 (35), 17592–17597.

Yu Chen and Jens Nielsen, 2021. In vitro turnover numbers do not reflect in vivo activities of yeast enzymes. Proc. Natl. Acad. Sci., 118(32).

Chen, Yu, Nielsen, Jens, 2021b. Mathematical modeling of proteome constraints within metabolism. Curr. Opin. Struct. Biol. 25, 50–56.

Cho, Han-Saem, Seo, SangWoo, Kim, Young Mi, Gyoo Yeol Jung, Jong Moon Park, 2012. Engineering glyceraldehyde-3-phosphate dehydrogenase for switching control of glycolysis in escherichia coli. Biotechnol. Bioeng. 109 (10), 2612–2619.

UniProt Consortium, 2019. Uniprot: a worldwide hub of protein knowledge. Nucleic Acids Res. 47 (D1), D506–D515.

Danisch, Simon, Krumbiegel, Julius, 2021. Makie.jl: flexible high-performance data visualization for julia. Journal of open source software 6 (65), 3349. https://doi.org/10.21105/joss.03349.

Davidi, Dan, Noor, Elad, Liebermeister, Wolfram, Bar-Even, Arren, Flamholz, Avi, Tummler, Katja, Barenholz, Uri, Goldenfeld, Miki, Shlomi, Tomer, Milo, Ron, 2016. Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements. Proc. Natl. Acad. Sci. 113 (12), 3401–3406.

de Groot, Daan H., Julia Lischke, Muolo, Riccardo, Planqué, Robert, Frank, J Bruggeman, Bas, Teusink, 2020. The common message of constraint-based optimization approaches: overflow metabolism is caused by two growth-limiting constraints. Cell. Mol. Life Sci. 77 (3), 441–453.

Iván Domenzain, Benjamın Sánchez, Mihail Anton, Eduard J Kerkhoven, Aarón Millán-Oropeza, Céline Henry, Verena Siewers, John P Morrissey, Nikolaus Sonnenschein, and Jens Nielsen, 2022. Reconstruction of a catalogue of genome-scale metabolic models with enzymatic constraints using gecko 2.0. Nat. Commun., 13.

Donati, Stefano, Kuntz, Michelle, Pahl, Vanessa, Farke, Niklas, Beuter, Dominik, Glatter, Timo, Jose Vicente Gomes-Filho, Randau, Lennart, Wang, Chun-Ying, Link, Hannes, 2021. Multi-omics analysis of crispri-knockdowns identifies mechanisms that buffer decreases of enzymes in e. coli metabolism. Cell systems 12 (1), 56–67.

Dourado, Hugo, Lercher, Martin J., 2020. An analytical theory of balanced cellular growth. Nat. Commun. 11 (1), 1–14.

Flamholz, Avi, Noor, Elad, Bar-Even, Arren, Milo, Ron, 2012. Equilibrator-the biochemical thermodynamics calculator. Nucleic Acids Res. 40 (D1), D770–D775.

Foster, Charles J., Wang, Lin, V Dinh, Hoang, Suthers, Patrick F., Maranas, Costas D., 2021. Building kinetic models for metabolic engineering. Curr. Opin. Biotechnol. 67, 35–41.

Goelzer, Anne, Jan, Muntel, Chubukov, Victor, Jules, Matthieu, Prestel, Eric, Rolf, Nölker, Mariadassou, Mahendra, Aymerich, Stéphane, Hecker, Michael, Noirot, Philippe, et al., 2015. Quantitative prediction of genome-wide resource allocation in bacteria. Metab. Eng. 32, 232–243.

Gould, Stephen, Fernando, Basura, Cherian, Anoop, Anderson, Peter, Cruz, Rodrigo Santa, Guo, Edison, 2016. On Differentiating Parameterized Argmin and Argmax Problems with Application to Bi-level Optimization. Arxiv preprint arxiv: 1607.05447.

Gowda, Shashi, Ma, Yingbo, Cheli, Alessandro, Gwózzdz, Maja, Shah, Viral B., Edelman, Alan, Rackauckas, Christopher, 2022. High-performance symbolic-numerics via multiple dispatch. issn: 1932-2240 ACM Commun. Comput. Algebra 55 (3), 92–96. https://doi.org/10.1145/3511528.3511535.

de Groot, Daan H, Hulshof, Josephus, Bas, Teusink, Frank, J Bruggeman, Planqué, Robert, 2020. Elementary growth modes provide a molecular description of cellular self-fabrication. PLoS Comput. Biol. 16 (1), e1007559.
Hatzimanikatis, Vassily, Bailey, James E., 1996. Mca has more to say. J. Theor. Biol. 182 (3), 233–242.
Heckmann, David, Lloyd, Colton J., Nathan, Mih, Ha, Yuanchi, Zielinski, Daniel C., Haiman, Zachary B., Desouki, Abdelmoneim Amer, Lercher, Martin J., Palsson, Bernhard O., 2018. Machine learning applied to enzyme turnover numbers reveals protein structural correlates and improves metabolic models. Nat. Commun. 9 (1), 1–10.
Heckmann, David, Campeau, Anaamika, Lloyd, Colton J., Patrick, V Phaneuf, Ying, Hefner, Carrillo-Terrazas, Marvic, Feist, Adam M., Gonzalez, David J., Palsson, Bernhard O., 2020. Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers. Proc. Natl. Acad. Sci. 117 (37), 23182–23190.
Heinrich, Reinhart, Schuster, Stefan, 2012. The Regulation of Cellular Systems. Springer Science & Business Media.
Henry, Christopher S., Broadbelt, Linda J., Hatzimanikatis, Vassily, 2007. Thermodynamics-based metabolic flux analysis. Biophys. J. 92 (5), 1792–1805.
IBM ILOG Cplex, 2021. V20.1.0: User's Manual for Cplex. International business machines corporation.
Ishii, Nobuyoshi, Nakahigashi, Kenji, Baba, Tomoya, Robert, Martin, Soga, Tomoyoshi, Kanai, Akio, Hirasawa, Takashi, Miki, Naba, Hirai, Kenta, Hoque, Aminul, et al., 2007. Multiple high-throughput analyses monitor the response of e. coli to perturbations. Science 316 (5824), 593–597.
Schellenberger, Jan, Palsson, Bernhard Ø., 2009. Use of randomized sampling for analysis of metabolic networks. J. Biol. Chem. 284 (9), 5457–5461.
Kratochvıl, Miroslav, Heirendt, Laurent, Wilken, St. Elmo, Pusa, Taneli, Arreckx, Sylvain, Noronha, Alberto, Marvin van Aalst, Satagopam, Venkata P., Oliver, Ebenhöh, Schneider, Reinhard, et al., 2022. Cobrexa. jl: constraint-based reconstruction and exascale analysis. Bioinformatics 38 (4), 1171–1172.
Kroll, Alexander, Engqvist, Martin KM., Heckmann, David, Lercher, Martin J., 2021. Deep learning allows genome-scale prediction of michaelis constants from structural features. PLoS Biol. 19 (10), e3001402.
Li, Gang, Hu, Yating, Jan, Zrimec, Luo, Hao, Wang, Hao, Zelezniak, Aleksej, Ji, Boyang, Nielsen, Jens, 2021. Bayesian genome scale modelling identifies thermal determinants of yeast metabolism. Nat. Commun. 12 (1), 1–12.
Li, Feiran, Yuan, Le, Lu, Hongzhong, Li, Gang, Chen, Yu, Engqvist, Martin KM., Kerkhoven, Eduard J., Nielsen, Jens, 2022. Deep learning-based kcat prediction enables improved enzyme-constrained model reconstruction. Nat. Catal. 1–11.
Meldal, Birgit HM., Perfetto, Livia, Combe, Colin, Lubiana, Tiago, , João Vitor Ferreira Cavalcante, Bye-A-Jee, Hema, Waagmeester, Andra, Del-Toro, Noemi, Shrivastava, Anjali, Barrera, Elisabeth, et al., 2022. Complex portal 2022: new curation frontiers. Nucleic Acids Res. 50 (D1), D578–D586.
Miles, John S., Guest, John R., Radford, Sheena E., Perham, Richard N., 1988. Investigation of the mechanism of active site coupling in the pyruvate dehydrogenase multienzyme complex of escherichia coli by protein engineering. J. Mol. Biol. 202 (1), 97–106.
Monk, Jonathan M., Lloyd, Colton J., Brunk, Elizabeth, Nathan, Mih, Sastry, Anand, King, Zachary, Takeuchi, Rikiya, Nomura, Wataru, Zhang, Zhen, Mori, Hirotada, et al., 2017. iml1515, a knowledgebase that computes escherichia coli traits. Nat. Biotechnol. 35 (10), 904–908.
Moses, William, Churavy, Valentin, 2020. Instead of rewriting foreign code for machine learning, automatically synthesize fast gradients. In: Advances in Neural Information Processing Systems. Volume 33. Curran Associates, Inc., pp. 12472–12485 url: https://proceedings.neurips.cc/paper/2020/file/9332c513ef44b682e9347822c2e457ac-Paper.pdf
Müller, Stefan, Regensburger, Georg, Steuer, Ralf, 2014. Enzyme allocation problems in kinetic metabolic networks: optimal solutions are elementary flux modes. J. Theor. Biol. 347, 182–190.
Nilsson, Avlant, Nielsen, Jens, 2016. Metabolic trade-offs in yeast are caused by f1f0-atp synthase. Sci. Rep. 6 (1), 1–11.
Noor, Elad, Flamholz, Avi, Liebermeister, Wolfram, Bar-Even, Arren, Milo, Ron, 2013. A note on the kinetics of enzyme action: a decomposition that highlights thermodynamic effects. FEBS Lett. 587 (17), 2772–2777.
Noor, Elad, Bar-Even, Arren, Flamholz, Avi, Reznik, Ed, Liebermeister, Wolfram, Milo, Ron, 2014. Pathway thermodynamics highlights kinetic obstacles in central metabolism. PLoS Comput. Biol. 10 (2), e1003483.
Noor, Elad, Flamholz, Avi, Bar-Even, Arren, Davidi, Dan, Milo, Ron, Liebermeister, Wolfram, 2016. The protein cost of metabolic fluxes: prediction from enzymatic rate laws and cost minimization. PLoS Comput. Biol. 12 (11), e1005167.
Orth, Jeffrey D., Thiele, Ines, Palsson, Bernhard Ø., 2010. What is flux balance analysis? Nat. Biotechnol. 28 (3), 245–248.
O'Brien, Edward J., Lerman, Joshua A., Chang, Roger L., Hyduke, Daniel R., Palsson, Bernhard Ø., 2013. Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction. Mol. Syst. Biol. 9 (1), 693.
Parikh, Neal, Boyd, Stephen, 2014. Proximal algorithms. Foundations Trends Optimiz. 1 (3), 127–239.
Park, Junyoung O., Rubin, Sara A., , Yi-Fan Xu, Amador-Noguez, Daniel, Fan, Jing, Shlomi, Tomer, Rabinowitz, Joshua D., 2016. Metabolite concentrations, fluxes and free energies imply efficient enzyme usage. Nat. Chem. Biol. 12 (7), 482–489.
Bekiaris, Pavlos Stephanos, Klamt, Steffen, 2020. Automatic construction of metabolic models with enzyme constraints. BMC Bioinf. 21 (1), 1–13.
Revels, J., Lubin, M., Papamarkou, T., 2016. Forward-mode Automatic Differentiation in Julia. Arxiv:1607.07892 [cs.ms]. url: https://arxiv.org/abs/1607.07892.
Sánchez, Benjamın J., Zhang, Cheng, Nilsson, Avlant, , Petri-Jaan Lahtvee, Kerkhoven, Eduard J., Nielsen, Jens, 2017. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. Mol. Syst. Biol. 13 (8), 935.
Schmidt, Alexander, Kochanowski, Karl, Vedelaar, Silke, Ahrné, Erik, Volkmer, Benjamin, Callipo, Luciano, Knoops, Kevin, Bauer, Manuel, Aebersold, Ruedi, Heinemann, Matthias, 2016. The quantitative and condition-dependent escherichia coli proteome. Nat. Biotechnol. 34 (1), 104–110.
Sharma, Akshay, Besançon, Mathieu, Garcia, Joaquim Dias, Legat, Benoît, 2022. Flexible differentiable optimization via model transformations. url: https://arxiv.org/abs/2206. 06135.
Tsouka, Sophia, Ataman, Meric, Hameri, Tuure, Miskovic, Ljubisa, Hatzimanikatis, Vassily, 2021. Constraint-based metabolic control analysis for rational strain engineering. Metab. Eng. 66, 191–203.
Villadsen, John, Nielsen, Jens, Lidén, Gunnar, 2011. Bioreaction Engineering Principles. Springer Science & Business Media.
Wilken, St. Elmo, Vera Frazão, Victor, Saadat, Nima P., Oliver, Ebenhöh, 2021. The view of microbes as energy converters illustrates the trade-off between growth rate and yield. Biochem. Soc. Trans. 49 (4), 1663–1674.
Wittig, Ulrike, Kania, Renate, Golebiewski, Martin, Rey, Maja, Shi, Lei, Lenneke Jong, Algaa, Enkhjargal, Weidemann, Andreas, Sauer-Danzwith, Heidrun, Mir, Saqib, et al., 2012. Sabio-rk-database for biochemical reaction kinetics. Nucleic Acids Res. 40 (D1), D790–D796.
Yang, Laurence, Ding, Ma, Ali, Ebrahim, Lloyd, Colton J., Saunders, Michael A., Palsson, Bernhard O., 2016. Solveme: fast and reliable solution of nonlinear me models. BMC Bioinf. 17 (1), 1–10.
Zhou, Jingru, Zhuang, Yingping, Xia, Jianye, 2021. Integration of enzyme constraints in a genome scale metabolic model of aspergillus Niger improves phenotype predictions. Microb. Cell Factories 20 (1), 1–16.