PAPER

Algebraic differentiation for fast sensitivity analysis of optimal flux modes in metabolic models

Hester Chapman,[1] Miroslav Kratochvíl,[2] Oliver Ebenhöh[1] and St. Elmo Wilken[1,*]

[1]Institute of Quantitative and Theoretical Biology, Heinrich Heine University, 40255 Düsseldorf, Germany and [2]Luxembourg Centre of Systems Biomedicine, University of Luxembourg, Belval Campus, 6 avenue du Swing, Belvaux, L-4367, Esch-sur-Alzette, Luxembourg

*Corresponding author. wilkenst@hhu.de

Abstract

Motivation: Sensitivity analysis is a useful tool to identify key parameters in metabolic models. It is typically only applied to the growth rate, disregarding the sensitivity of other solution variables to parameters. Further, sensitivity analysis of elementary flux modes could provide low-dimensional insights into optimal solutions, but they are not defined when a model is subject to inhomogeneous flux constraints, such as the frequently used ATP maintenance reaction.

Results: We introduce optimal flux modes (OFMs), an analogue to EFMs, but specifically applied to optimal solutions of constraint-based models. Further, we prove that implicit differentiation can always be used to efficiently calculate the sensitivities of both whole-model solutions and OFM-based solutions to model parameters. This allows for fine-grained sensitivity analysis of the optimal solution, and investigation of how these parameters exert control on the optimal composition of OFMs. This novel framework is implemented in DifferentiableMetabolism.jl, a software package designed to efficiently differentiate solutions of constraint-based models. To demonstrate scalability, we differentiate solutions of 342 yeast models; additionally we show that sensitivities of specific subsystems can guide metabolic engineering. Applying our scheme to an *Escherichia coli* model, we find that OFM sensitivities predict the effect of knockout experiments on waste product accumulation. Sensitivity analysis of OFMs also provides key insights into metabolic changes resulting from parameter perturbations.

Availability and Implementation: Software introduced here is available as open-source Julia packages DifferentiableMetabolism.jl (`https://github.com/stelmo/DifferentiableMetabolism.jl`) and ElementaryFluxModes.jl (`https://github.com/HettieC/ElementaryFluxModes.jl`), which both work on all major operating systems and computer architectures. Code to reproduce all results is available from `https://github.com/HettieC/DifferentiableOFMPaper`, and as an archive from `https://doi.org/10.5281/zenodo.15183208`.

Contact: wilkenst@hhu.de

Supplementary information: Supplementary data are available at Bioinformatics online.

Key words: enzyme-constrained, flux balance analysis, differentiable, elementary flux modes, sensitivity analysis

Introduction

Genome-scale metabolic models (GSMMs) provide a framework for studying cellular metabolism (Varma and Palsson, 1994; Edwards and Palsson, 1999). The annotated genome of an organism is used to identify metabolic reactions with an associated catalysing enzyme. In the simplest case, GSMMs are simulated using flux balance analysis (FBA) (Orth et al., 2010). Main applications include metabolic engineering (Park et al., 2009), investigating human metabolism and disease (Brunk et al., 2018), and pathogen-drug target identification (Kim et al., 2010).

While FBA simulations often capture the function of essential genes, they are less successful in predicting quantitative changes in phenotypic behaviour. The seemingly wasteful strategies of overflow metabolism in bacteria, and the Crabtree effect in yeast, cannot be reproduced through FBA without imposing ad-hoc flux constraints on selected reactions. It has been posited that overflow metabolism in *Escherichia coli* results from constrained protein allocation in the cell (Basan et al., 2015; Bruggeman et al., 2020); to account for this phenomenon in GSMMs, total proteome constraints and enzyme kinetics can be added to create an enzyme-constrained genome scale metabolic model (ecGSMM) (Sánchez et al., 2017; Beg et al., 2007).

Parameterisation of an ecGSMM requires knowledge of both the molar mass and turnover number (k_{cat}) of each enzyme. While the molar masses of enzymes are typically available from databases, the k_{cat}-values are poorly characterised (Davidi and Milo, 2017). Due to this lack of reliable experimental data, researchers may have to turn to non-mechanistic approximations via machine learning, for example that of Kroll (Kroll et al., 2023).

Despite these parameterisation issues, ecGSMMs are still able to effectively capture a much wider variety of phenotypes than GSMM. It is thus of interest to analyse which parameters exert the most control over an optimal flux distribution, as this can be used to guide metabolic engineering efforts or uncover novel drug targets. Previous work has detailed how to implicitly differentiate an optimal solution to an ecGSMM to find the sensitivity of a solution to parameter changes (Wilken et al., 2022). However, well-defined criteria for when this is possible were not provided.

Alternatively, sensitivity analysis of ecGSMMs can also be performed through finite differentiation, where a parameter is perturbed and the change in the values of the model variables is used to approximate the sensitivity of those variables to the perturbed parameter. The main drawback is inefficiency, with n parameters requiring at least n additional linear programs to be solved, depending on the finite differencing scheme. Furthermore, numerical errors can arise, with accuracy dependent on the size of the perturbation. Another method, based on the shadow prices of linear optimisation problems, has also been extensively studied (Liebermeister, 2018; van den Bogaard et al., 2024). A major drawback of the shadow price approach is the inability to calculate sensitivities of variables other than the objective function.

Here, we prove that it is generally possible to implicitly differentiate a flux solution of an ecGSMM, and show that the only criterion necessary for this is the removal of inactive reactions from the model before differentiation. Moreover, when the inactive reactions are removed, we show that the solution is unique and the derivatives are well-defined.

Further, one can always decompose an optimal solution of an ecGSMM with homogeneous constraints into a linear combination of a small number of elementary flux modes (EFMs) (De Groot et al., 2019). However, when inhomogeneous constraints are included, such as an ATP maintenance reaction, EFMs are not defined. We introduce optimal flux modes (OFMs), an analogue to EFMs, which resolve the inhomogeneity issue. We show that OFMs provide a similar low-dimensional view of metabolic pathways as EFMs, but restricted to modes fulfilling the optimal objective value.

To demonstrate the utility of this approach, we investigate the OFMs associated with an enzyme-constrained model exhibiting overflow metabolism. We find both a respiratory and a fermentative OFM in the optimal solution. It may be of interest to find out to what extent the enzyme kinetic parameters control the phenotypic ratio of respiration and fermentation; for example, how would a knock down of a glycolytic enzyme affect the optimal proportion of the two OFMs? In this work, we prove that the sensitivity of OFM usage to model parameters can be calculated using the same implicit algebraic differentiation technique.

To enable efficient differentiation of constraint-based models, we introduce DifferentiableMetabolism.jl, a Julia package for the fast calculation of sensitivities of all model variables to parameters. The framework uses implicit differentiation, analogous to classic metabolic control analysis (MCA) (Kacser and Burns, 1973). We also present ElementaryFluxModes.jl to compute and implicitly differentiate EFMs or OFMs. We use these packages to differentiate the optimal solutions of 342 previously published fungal models (Li et al., 2022), showing that the method is readily applicable on a large-scale. Finally, our investigation into the minimisation of acetate production in *E. coli* highlights one possible application of our framework in metabolic engineering.

Theory

Here, we describe the ecGSMM model formulation, introduce and define optimal flux modes, and summarise the established method of implicit differentiation to find sensitivities of optimal solutions (Wilken et al., 2022). Detailed proofs of all lemmas and theorems are presented in the Supplementary material.

Model formulation

Following the model formulation from Sanchez (Sánchez et al., 2017), we describe the enzyme-constrained FBA (ecFBA) optimisation problem applied to a simple ecGSMM as follows:

$$
\begin{aligned}
\text{maximise} \quad & v_r \\
\text{subject to} \quad & \text{(C1)} \ \boldsymbol{Sv} = \boldsymbol{0} \\
& \text{(C2)} \ v_i \geq 0 \\
& \text{(C3)} \ v_i = k_{cat,i} \cdot e_i \ \ \forall i \in R \\
& \text{(C4)} \ \sum_{i \in R_k} e_i \leq E_k \ \ \forall k \in [1, \dots, K]
\end{aligned}
\tag{1}
$$

There are r reactions in the model, and we reorder them such that v_r is the objective reaction. The set $R \subset \mathbb{N}^n$ is the set of indices of reactions with associated k_{cat}-values. The flux, v_i, through reaction i is determined by the concentration, e_i, of the enzyme catalysing the reaction, multiplied by its associated turnover number, $k_{cat,i}$. The number of enzyme pools is given by K, and R_k gives the set of indices of enzymes in the k-th pool. Reversible reactions are split into a forward and backward irreversible reaction. Enzyme complexes and isozymes may be included in the model formulation, but for brevity are not included in equation (1). We constrain the proteome capacity of the model, with the total usage of enzymes from the k-th pool bounded by the capacity E_k.

In the above formulation, one places bounds E_k on a subset of model enzymes in a particular pool k. This pool could be a compartment, say the membrane, a specific pathway, or an individual enzyme. The stoichiometric matrix, $\boldsymbol{S} \in \mathbb{R}^{m \times r}$, contains a row for each of the m metabolites and a column for each of the r reactions, with entries in each column giving the stoichiometric coefficients of the metabolites involved in that column's reaction (Orth et al., 2010). This is the matrix we use to find elementary flux modes of a network.

Flux cones and elementary flux modes

The set of flux vectors $\boldsymbol{v}$ satisfying both $\boldsymbol{Sv} = \boldsymbol{0}$ and $v_i \geq 0$ forms a convex polyhedral cone, called the flux cone (Klamt et al., 2017):

$$
\mathrm{FC}(\boldsymbol{S}) = \{\boldsymbol{v} \in \mathbb{R}^n | \boldsymbol{Sv} = \boldsymbol{0}, v_i \geq 0\}
\tag{2}
$$

Any non-trivial flux vector $\boldsymbol{v}$ is called an admissible flux mode if it is contained in the flux cone $\mathrm{FC}(\boldsymbol{S})$ (2). We define the

support of a mode, $\mathbf{supp}(\boldsymbol{v}) = \{i|v_i \neq 0\}$ as the set of indices of non-zero elements in the flux mode $\boldsymbol{v}$.

Definition 1 (Elementary flux mode (EFM)). *An admissible flux mode $\boldsymbol{v} \in FC(\boldsymbol{S})$ is called an* elementary flux mode, *$\boldsymbol{m}$, if there exists no other admissible flux mode $\boldsymbol{v}$, where $\mathbf{supp}(\boldsymbol{m}) \supset \mathbf{supp}(\boldsymbol{v})$. Thus, $\mathbf{supp}(\boldsymbol{m})$ cannot be written as a proper superset of any other mode $\boldsymbol{v}$.*

From this definition, if a flux mode is elementary, it has minimal support, and any attempt to reduce its support by deleting a reaction will imply $\boldsymbol{v} = 0$.

Inhomogeneous constraints and optimal flux modes

In genome-scale metabolic modelling, there are often not only homogeneous ($v \geq 0$) but also inhomogeneous constraints ($v \geq c > 0$ for some constant c). For example, uptake and secretion rates of metabolites may be constrained from exometabolomics data, ^{13}C-metabolic flux analysis allows for quantification of intracellular fluxes, and proteomics or transcriptomics may be used to constrain enzyme concentrations. It is important that we analyse these types of model formulations. Incorporating inhomogeneous equality constraints into (1) can be done as follows:

$$\begin{aligned}
\text{maximise} \quad & v_r \\
\text{subject to} \quad & \text{(C1-C4)} \\
& \text{(C5)} \ v_j \geq c_j > 0 \ \forall j \in R_f
\end{aligned} \quad (3)$$

where R_f is the set of reactions with fixed non-zero flux, and c_j is the known constant flux of reaction j.

The solution space of a model with the setup given by problem (3) will no longer form a flux cone, but rather the flux polyhedron $P(\boldsymbol{S})$, defined as:

$$P(\boldsymbol{S}) = \{\boldsymbol{v} \in \mathbb{R}^n | \boldsymbol{Sv} = \boldsymbol{0}, \ v_i \geq 0, \ v_j \geq c_j \ \forall j \in R_f\} \quad (4)$$

We distinguish between the homogeneous problem (1) and the inhomogeneous problem (3) since EFMs are not defined for flux polyhedra such as (4) (Klamt et al., 2017). We can, however, incorporate inhomogeneous equality constraints into a modified stoichiometric matrix. Any inhomogeneous constraints that are active (fulfilled with equality) in an optimal solution may be treated as equality constraints, and we treat inhomogeneous constraints that are inactive as homogeneous constraints.

Following the approach of Schuster (Schuster and Schuster, 1993), we may introduce a slack variable $\overline{v}$ to account for inhomogeneous constraints. Let $\boldsymbol{v}^{(1)}$ and $\boldsymbol{S}^{(1)}$ contain the flux rates and stoichiometries of reactions with unknown fluxes, and let $\boldsymbol{w} := \boldsymbol{S}^{(2)}\boldsymbol{v}^{(2)}$ contain those of the reactions with fixed known fluxes. We may now write an optimisation problem involving only homogeneous constraints:

$$\begin{aligned}
\text{maximise} \quad & \overline{v} \\
\text{subject to} \quad & \boldsymbol{Ax} = \boldsymbol{0} \\
& x_i \geq 0 \\
& x_i = k_{\text{cat},i} \cdot e_i \quad \forall i \in R \\
& \sum_i e_i \leq E_k \quad \forall k \in [1,\dots,K] \\
\text{where} \quad & \boldsymbol{A} := \begin{bmatrix} \boldsymbol{S}^{(1)} & \boldsymbol{w} \end{bmatrix}, \ \boldsymbol{x} := \begin{bmatrix} \boldsymbol{v}^{(1)} \\ \overline{v} \end{bmatrix}.
\end{aligned} \quad (5)$$

When $\overline{v} = 1$, solutions to the homogeneous system (5) coincide precisely with solutions of problem (3).

A non-trivial vector $\boldsymbol{x}$ is called an admissible flux mode of (3) if it is contained in the optimal flux cone

$$\text{OFC}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{Ax} = \boldsymbol{0}, v_i \geq 0\} \quad (6)$$

Definition 2 (Optimal flux mode (OFM)). *A flux mode $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{v}^{(1)} \\ \overline{v} \end{bmatrix}$ is called an* optimal flux mode *of (3) if it fulfils the following conditions:*

- *Admissibility: $\boldsymbol{x} \in OFC(\boldsymbol{A})$*
- *Minimal support: There exists no other flux mode $\boldsymbol{y}$ satisfying the above conditions, where $\mathbf{supp}(\boldsymbol{x}) \supset \mathbf{supp}(\boldsymbol{y})$*
- *Optimality: $\overline{v} = 1$.*

The first two conditions for an OFM coincide with Definition 1 of an EFM of the optimal flux cone (6); it is the extra condition of optimality that distinguishes OFMs from EFMs. From Definition 2, an OFM is an optimal, minimal, flux-carrying mode. Should a reaction be deleted from an OFM, either the mode would no longer be able to carry flux, or it would do so at a sub-optimal rate. A mode with a sub-optimal flux is one in which the ratio of v_r to fixed fluxes $\boldsymbol{v}^{(2)}$ is lower than that in the optimal solution. A similar construction, elementary flux vectors, has previously been reported (Klamt et al., 2017; Urbanczik, 2007), but while OFMs are support minimal, the set of elementary flux vectors need not be.

To clearly highly the differences between EFMs and OFMs, we have included a toy model in the Supplementary material. Here we provide an example of whether EFMs or OFMs are the appropriate modes for two different model setups, and detail how to calculate them in either scenario.

Differentiating solutions

The solution of a convex optimisation problem, such as (5), with a unique optimal solution, may be implicitly differentiated (Agrawal et al., 2019). We previously showed that one can equate these differentials to model sensitivities (Wilken et al., 2022). If there are multiple optimal solutions to problem (3), a single solution must be chosen for analysis. See the Supplementary material for a discussion on this. To ensure that differentiation is well-defined, we describe a technique for obtaining unique optimal solutions.

Definition 3 (Inactive reaction/enzyme). *In an optimal solution, an* inactive reaction *is one carrying zero flux; an* inactive enzyme *has zero concentration.*

Definition 4 (Pruned model). *A model produced by removing all reactions and enzymes of an ecGSMM that are inactive in some optimal solution of that ecGSMM is called a* pruned model. *Removing these reactions and enzymes is called* pruning.

Note that we also reverse any backwards reactions so that a solution to a pruned model will only have forward fluxes. In order to prove the main result of this section, that EFMs and OFMs can be implicitly differentiated, we must first discuss several necessary lemmas and show that the results for homogeneous problems (1) also apply to inhomogeneous problems (3).

Lemma 1. *The extreme rays of problem (5) with a non-zero $\overline{v}$ correspond to the optimal flux modes (OFMs) of (3).*

The proof follows from the result of Gagneur (Gagneur and Klamt, 2004) that extreme rays of a pointed polyhedral cone $FC(\mathbf{S}) = \{\mathbf{x} | \mathbf{Sx} = \mathbf{0}, x_i \geq \mathbf{0}\}$ correspond to elementary flux modes of the homogeneous problem (1). See the Supplementary material for details.

De Groot et al. previously proved that a homogeneous problem (1) with K enzymatic constraints will use at most K EFMs in an optimal solution (De Groot et al., 2019). We extend this result to inhomogeneous problems (3) and OFMs.

Lemma 2. *An optimal solution to problem (3) with K enzymatic constraints (**C4**), rescaled to be in the form*

$$C_{\Sigma}^{(k)} := \sum_{j=1}^{r} w_j^{(k)} e_j \leq 1 \ \ for \ \ k \in \{1, ..., K\} \qquad (7)$$

will use at most K optimal flux modes.

In short, we know from de Groot (De Groot et al., 2019) that in the homogeneous problem (1), at most K EFMs will be active in an optimal solution. Combining this with Lemma 1, we prove that an optimal solution to (3) will use at most K OFMs in the Supplementary material.

In practice, GSMMs always incorporate fixed fluxes, whether these are an ATP maintenance reaction or the measured exchange of a carbon source. We have proven that, equivalently to EFMs, the number of OFMs in an optimal solution is bounded by the number of active enzyme constraints. Therefore, all following results apply to EFMs and OFMs, but for brevity we use only the term OFMs unless specifying is necessary.

A pruned optimal solution is unique

To assert the validity of pruning models to calculate sensitivities, we state and prove conditions under which implicit differentiation is possible.

Theorem 1. *Given a model of the form (1), let us assume that all metabolic reactions have an associated enzyme, and thus an associated enzyme cost. Then, pruning an optimal solution will give a model with a unique optimal solution.*

Following the approach of de Groot (De Groot et al., 2019), we define a cost matrix of the EFMs in a solution, and rewrite the optimisation problem (1) as a maximisation of EFM usage, subject to constraints relating to the cost matrix. In a genome scale metabolic model, the enzyme turnover numbers and gene product molar masses can be assumed to have enough variation that the vectors of the cost matrix are linearly independent. This linear independence is exploited to prove that there is a unique optimal solution to a pruned model. See Supplementary material for details.

In the setup where we have inhomogeneous constraints (3), we find again that the pruned solution is unique:

Corollary 1. *The optimal solution of a pruned ecGSMM of the form (3) is always unique.*

Proof : We can rewrite problem (3) as problem (5), and then use the proof of Theorem 1, equating EFMs in Theorem 1 with OFMs in the current problem. □

Corollary 2. *Pruned models in the form of (1) or (3) can be implicitly differentiated to calculate the sensitivity of all solution variables to all model parameters.*

Proof : The solution map of a convex optimisation problem, with a unique optimal solution, may be implicitly differentiated, as proven by Agrawal et al. (Agrawal et al., 2019). Theorem 1 and Corollary 1 ensure that pruned models do indeed have unique optimal solutions. Therefore, these solutions are also implicitly differentiable, using the method of Wilken et al. (Wilken et al., 2022). □

Corollary 2 is required to guarantee that the strategy from (Wilken et al., 2022) is indeed generally applicable. Implicit differentiation provides a robust, efficient method for calculating solution sensitivities. Specifically, we differentiate through the Karush-Kuhn-Tucker (KKT) conditions of the optimisation problem to calculate $\frac{\partial v_i}{\partial p_j}$, the sensitivity of flux through reaction i to parameter j.

Beyond the sensitivities of variables in the optimal solution, we wish to investigate the sensitivity of OFM usage to model parameters. Since OFM usage can be taken as another model variable, we may apply Corollary 2. Below, we describe the implementation of implicit differentiation for OFM sensitivities.

Methods

Sensitivity of OFM usage to model parameters

We use implicit differentiation to robustly and efficiently calculate the sensitivities of OFM usage to model parameters. In a problem of the form (1) or (3), with K enzymatic constraints, we can use Lagrange multipliers to implicitly differentiate the usage of EFMs or OFMs. The full details are provided in the Supplementary material.

Software implementation

To investigate OFM sensitivities, we must first calculate the OFMs in an optimal solution. To this end, we implemented the Double Description algorithm described by Terzer et al. (Terzer, 2009), as well as the implicit differentiation of OFMs, in the Julia package ElementaryFluxModes.jl.

Whole solution sensitivity analysis is provided in the software package DifferentiableMetabolism.jl. This was written to be used alongside COBREXA.jl (Kratochvíl et al., 2025). The implementation is not limited to the ecGSMM model formulation presented in this work; it can also be applied to community models, constrained allocation flux balance analysis models (Mori et al., 2016), and other linear programming problems, provided a unique optimal solution can be guaranteed.

Models

Enzyme-constrained fungi models

Using 342 published fungal GSMMs (Li et al., 2022), we made ecGSMMs to demonstrate the applicability of our software to a large number of models with no individual adjustments required. We allowed for unlimited uptake of glucose and oxygen, but constrained the total protein in the model to 35% of the cell dry weight, so that the models were each subject to a single enzyme constraint. We used the k_{cat} prediction results provided by the same study as the models, and enzyme complexes were assumed to contain a 1:1 ratio of subunits. All metabolic reactions without a gene-reaction-rule were given a proxy enzyme with the average k_{cat} and gene product molar mass of the model.

E. coli enzyme-constrained metabolic model

The iML1515 model of *E. coli* MG1655 (Monk et al., 2017) was used as the base GSMM. Enzyme turnover numbers were taken from (Heckmann et al., 2020). If available, enzyme subunit stoichiometries were taken from the complex portal (Meldal et al., 2015), otherwise Swiss-Prot entries (Boutet et al., 2007) were used. If no information was available, enzyme complexes were assumed to contain a 1:1 ratio of subunits. The model was given the enzyme capacity constraints of $160 \frac{\mathrm{mg}}{\mathrm{g_{DCW}}}$ membrane protein, and $340 \frac{\mathrm{mg}}{\mathrm{g_{DCW}}}$ in the rest of the cell. The numbers used for these bounds were taken from absolute quantitative proteomics measurements with glucose as the sole carbon source (Schmidt et al., 2016).

Acetate production was investigated by using the sensitivities of two OFMs in the optimal solution of the ecGSMM of iML1515. Experimental effects of gene knockouts on acetate production were taken from the measurements in (Fong et al., 2006; De Maeseneire et al., 2006; Contiero et al., 2000; Lozano Terol et al., 2019; Zhao et al., 2004; Long et al., 2016).

Results

Software performance

The software presented in this paper allows for fast differentiation of model solutions to model parameters. We have evaluated the computational resources required for DifferentiableMetabolism.jl to calculate the sensitivity of all model variables with respect to all model parameters, and compared this to the time taken by a trivial implementation of central finite differencing. Our implementation of central finite differencing involved perturbing a parameter 0.1% above and 0.1% below the baseline value, and calculating the slope of the change in model variables as an approximation for the sensitivity of variables to parameters. Conversely, implicit differentiation is mathematically exact, and accuracy only relies only on the robustness of numerically solving a system of linear equations.

In the benchmark settings, our software was around $4\times$ more efficient than finite differencing (Table 1) on full models, and more than $25\times$ more efficient with OFM-based sensitivity calculations. Notably, with finite differencing the OFMs of the pruned models need to be recalculated for each parameter perturbation. This is required to compensate for changes in the optimal objective caused by the perturbations, which invalidate the calculated OFM. DifferentiableMetabolism.jl only requires that the OFMs are calculated once; which is followed by solving a single system of linear equations that yields all sensitivities. The practically applicable efficiency improvement, comparing the better settings for each method, was around $7\times$ in the benchmark.

Sensitivity analysis of 342 fungal models

In order to demonstrate the robustness and scalability of DifferentiableMetabolism.jl, we simulated growth of 342 fungal models and calculated the sensitivity of their growth rates to the model parameters. Using traditional finite difference techniques would require solving at least $342 \times n$ linear programs, where n denotes the number of parameters in each model. Using our software, we only needed to solve one system of linear equations per model (after the initial LP was solved). In turn, our routine was more efficient than using finite differencing.

Repr.	Model	#vars	CPU time consumed (s)	
			DiffMet	CDiff
Full	yeastGEM	436	7.48 ± 0.11	32.45 ± 0.32
	iML1515	394	6.24 ± 0.08	27.06 ± 0.31
OFMs	yeastGEM	2	$\mathbf{3.81} \pm 0.15$	144.90 ± 0.99
	iML1515	2	$\mathbf{3.87} \pm 0.19$	99.58 ± 0.77

Table 1. Calculation of sensitivities more efficient using DifferentiableMetabolism.jl (labeled DiffMet in the table) than using central finite differencing (CDiff). We report the single-CPU-core time as the geometric mean ± standard deviation of 10 replicate simulations, for the sensitivity of all variables in the optimal model solutions, as well as of the OFM usage in the optimal solutions, to all parameters. All simulations in this paper were run on an AMD Ryzen 9 5950X with 32 GB main memory.
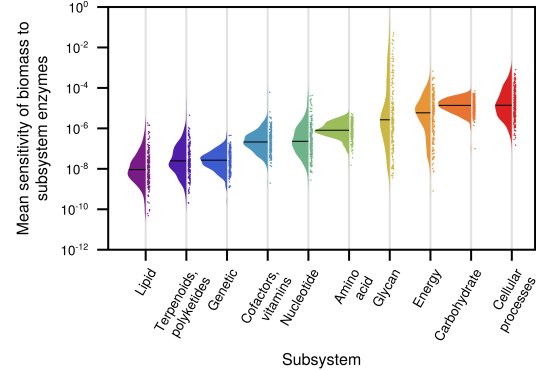

Figure 1. The mean sensitivity of biomass reactions across 342 fungal species is consistent across all subsystems except glycan. Each point on the plot is the mean sensitivity of the growth of a model to its enzyme turnover numbers in a subsystem. Rain clouds show the spread across sensitivity of model growth rates to kinetic parameters in one subsystem.

Figure 1 summarises the variation of growth sensitivity to model reaction subsystems. Each point gives the mean sensitivity of the growth of a model to the k_{cats}-values within a subsystem. A possible application of our framework is in metabolic engineering: interestingly, we observed a large spread in the sensitivity of growth to glycan metabolism across the models. Glycan engineering in yeast species focuses on the production of proteins carring human-compatible glycosylation (Piirainen et al., 2022). Deciding on which organisms to focus engineering efforts requires, amongst other factors, knowledge of the impact of metabolic alterations on the growth.

Our results indicate that the response of yeast species to changes in k_{cat}-values in glycan metabolism varies widely. The models with growth predicted to be least sensitive to glycan metabolism include *Saccharomyces cerevisiae, Lachancea fermentati, Hanseniaspora valbyensis,* and *Kazachstania naganishii,* which are all organisms that have been successfully used in glycosylation production (Chiba and Akeboshi, 2009), or proposed or used for other bio-processes (Bellut et al., 2020; Liu et al., 2021; Abd-El-Haleem et al., 2024).

OFM sensitivity predictions of acetate production in E. coli

We observed that sensitivities of OFMs captured gene knockout effects better than performing standard in-silico gene knockouts. Since *E. coli* is one of the best characterised and most successfully utilised organisms in biotechnology, we attempted
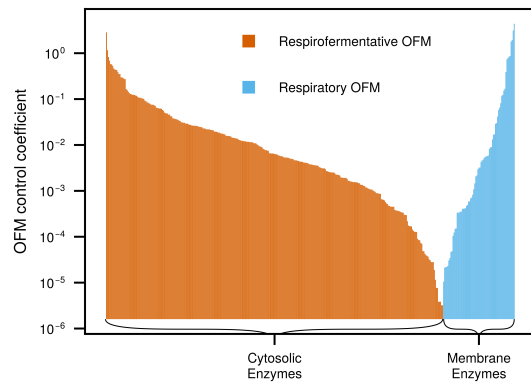
Figure 2. Control coefficients of OFM usage in *E. coli*, $\frac{p}{\lambda}\frac{\partial\lambda}{\partial p}$, support the assumption that space in the membrane controls respiration. Along the x-axis we have every enzyme parameter, p, in the active solution, and we plot on the y-axis whichever of $\frac{p}{\lambda_1}\frac{\partial\lambda_1}{\partial p}$ and $\frac{p}{\lambda_2}\frac{\partial\lambda_2}{\partial p}$ is greater than zero. Increasing any kinetic parameter in the cytosol will increase the use of the respirofermentative OFM (orange), and decrease the use of the respiratitive OFM (blue), in the optimal solution. The opposite is true for any membrane bound kinetic parameter.

to validate the OFM sensitivities with 23 known *E. coli* knockouts of 16 different genes, where some genes have been knocked out in more than one study. It is a recurring theme in *E. coli* engineering to reduce the formation of acetate (De Mey et al., 2007), since its presence can slow down growth (Nakano et al., 1997) and decrease product yield (Pan et al., 1987). We compared the phenotype in 23 *E. coli* knockout mutants to the sensitivity of an acetate producing OFM to the associated parameters.

As expected from Theorem 2, we obtained a superposition of two OFMs; an OFM composed of respiratory pathways that produces 0.05 mol of acetate per mol of glucose, and an aero-fermentative OFM producing 0.36 mol acetate per mol glucose.

Figure 2 shows that an increase in the turnover number of any membrane enzyme also increases the usage of the respiratory OFM, which in turn decreases the usage of the aero-fermentative OFM. The opposite effect is observed with cytosolic enzymes: major components of the respiratory pathway, including oxidative phosphorylation, are located in the cytoplasmic membrane. Therefore, if any enzyme in the membrane, including non-respiratory enzymes, has its turnover number increased, a smaller concentration of this enzyme will be required to catalyse the same flux (recall that $v = k_{\mathrm{cat}} \cdot e$) and there will hence be more available space in the membrane for every other membrane bound enzyme.

In Figure 3, we see that for 15 of a total 23 gene knockouts, experimental increases in acetate production were correctly predicted by the sensitivity of the acetate producing OFM (shown in green). We considered a prediction correct either if, upon knocking out the gene, acetate-production decreases and the sensitivity of the OFM to the k_{cat} is positive, or if acetate production increases and the sensitivity is negative. A positive sensitivity implies that OFM use increases or decreases together with the parameter change, negative sensitivity implies opposite direction of changes.

We further analysed the reasons for prediction failures in the 8 knockouts. The majority (6 out of 8) of the incorrect predictions were from knockout mutants where the acetate production increased compared to the wildtype. A possible additional discrepancy in two experiments has caused a prediction failure of sensitivity for phosphoenolpyruvate carboxylase (ppc) as seen
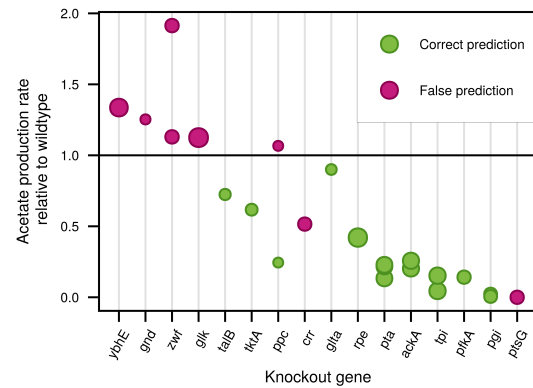

Figure 3. Sensitivity of the respirofermentative OFM in *E. coli* provides qualitatively accurate predictions of those gene knockouts that decrease acetate production. We plot the acetate production rate in 23 knockout mutants of 16 different genes relative to the wildtype, taken from experimental data (see Methods). The size of the marker corresponds to the size of the control coefficient $\frac{p_{ko}}{\lambda_{ac}}\frac{\partial\lambda_{ac}}{\partial p_{ko}}$ of the acetate producing OFM, λ_{ac}, to the turnover number, p_{ko}, of the enzyme of the knockout gene. The colour indicates whether this prediction matches the experimentally observed change in acetate in a knockout mutant.

in Figure 3, because the experiments measured opposite effects. Clarification of this problem might require additional reproduction of the experiments with ppc.

We hypothesise that pruning of the unused pathways before calculating the sensitivities might have caused some of the incorrect sensitivities. Pruned models capture local (infinitesimal) changes in turnover numbers, but can not capture larger metabolic switches that might occur in the greater neighbourhood of the given flux mode. It is noteworthy that despite this drawback, only one third of the knockouts actually had an effect different to the sensitivity of the OFM.

A further potential reason for OFM sensitivity failing to predict knockout effects is that the iML1515 model could simply be missing (or erroneously including) some reactions. To test this hypothesis, we performed in-silico gene knockouts with the standard GSM iML1515 without any enzyme constraints. We found that for 6 of the total 16 genes, the model either predicted no growth or an incorrect change in acetate production, thus performing worse than our OFM analysis.

Discussion

In this study, we have proven that pruning models leaves a unique optimal solution, which enables implicit differentiation to calculate sensitivities of the typical variables (flux rates, enzyme concentrations) and the usage of the OFMs. Many methods for the sensitivity analysis of constraint-based models analyse only the sensitivity of growth to parameters. This is of course an important factor, but neglects a fine-grained view that is prospectively useful, such as by-product sensitivities in metabolic engineering. The method introduced by Wilken (Wilken et al., 2022), and implemented here with DifferentiableMetabolism.jl, robustly and efficiently calculates the sensitivity of all model variables (fluxes and enzyme concentrations) to all model parameters (kinetic constants, upper and lower bounds on fluxes, and capacity constraints). This allows for a much wider analysis of the metabolism of an organism, as demonstrated by our analysis of acetate production in *E. coli* knockout mutants. Moreover, control of model parameters can be holistically summarised by changes in OFMs, which provide

a more interpretable overview of the active metabolic routes than a full-scale sensitivity analysis of every model variable.

The sensitivity of growth to model parameters is used to highlight the subsystems whose reactions have a large effect on growth in published fungal models. We found that the models whose growth was least sensitive to parameters of glycan metabolism have indeed been used either in glycosylation processes, or for other general metabolic engineering purposes. This validates our method as a useful tool in screening for potential microbial cell factories, and showcases DifferentiableMetabolism.jl as a scalable and robust tool to be used in future for general-purpose sensitivity analysis.

Our use of the OFM sensitivity to predict gene knockout effects in *E. coli* gave qualitatively more accurate predictions than in-silico gene knockout experiments of the iML1515 GSMM. However, the predictions did not fully match the experimental data. We assume that such analysis would be better suited to simulation of knockdowns or over-expression, but such validation is challenging due to a current lack of experimental data, and the high cost of these measurements.

Our work highlights that novel analysis techniques, such as OFM sensitivities, are of benefit when investigating model behaviour. We provide a robust, fast, and generally applicable scheme through which to analyse the control of model parameters on the whole solution, in contrast to previous efforts that only considered a simplified view of growth sensitivity. The analysis of OFMs in an optimal solution helps to provide a holistic understanding of the effects of local changes on global phenotypes. The pruning method assumes that infinitesimal parameter changes do not cause switching between optimal solutions. An analysis into the switching points of metabolism would be invaluable in future work.

Our performance measurements have provided a strong indication that the method is sufficiently efficient to scale to larger or community-wide models. Since both DifferentiableMetabolism.jl and the finite differencing algorithm offer possibilities for acceleration via parallel hardware, we aim to evaluate and compare their optimized parallel implementations in the future, giving a suitable guideline for high-performance analysis of the model sensitivities.

Supported by the new software packages presented in this paper, we expect the differentiation of model solutions to be increasingly used to investigate microbial metabolism.

Competing interests

No competing interest is declared.

Author contributions statement

HC: Conceptualisation, Methodology, Software, Validation, Formal analysis, Investigation, Writing - Original Draft. MK: Software, Validation, Investigation, Writing - Review & Editing. OE: Resources, Writing - Review, Supervision, Funding acquisition. SEW: Conceptualisation, Software, Validation, Investigation, Writing - Review & Editing, Supervision.

Acknowledgments

References

D. A. Abd-El-Haleem, M. R. Elkatory, and G. M. Abu-Elreesh. Uncovering novel polyhydroxyalkanoate biosynthesis genes and unique pathway in yeast hanseniaspora valbyensis for sustainable bioplastic production. *Scientific Reports*, 14(1): 27162, 2024.

A. Agrawal, S. Barratt, S. Boyd, E. Busseti, and W. M. Moursi. Differentiating through a cone program. *arXiv preprint arXiv:1904.09043*, 2019.

M. Basan, S. Hui, H. Okano, Z. Zhang, Y. Shen, J. R. Williamson, and T. Hwa. Overflow metabolism in escherichia coli results from efficient proteome allocation. *Nature*, 528 (7580):99–104, 2015.

Q. K. Beg, A. Vazquez, J. Ernst, M. A. de Menezes, Z. Bar-Joseph, A.-L. Barabási, and Z. N. Oltvai. Intracellular crowding defines the mode and sequence of substrate uptake by escherichia coli and constrains its metabolic activity. *Proceedings of the National Academy of Sciences*, 104(31): 12663–12668, 2007.

K. Bellut, K. Krogerus, and E. K. Arendt. Lachancea fermentati strains isolated from kombucha: fundamental insights, and practical application in low alcohol beer brewing. *Frontiers in microbiology*, 11:764, 2020.

E. Boutet, D. Lieberherr, M. Tognolli, M. Schneider, and A. Bairoch. Uniprotkb/swiss-prot: the manually annotated section of the uniprot knowledgebase. In *Plant bioinformatics: methods and protocols*, pages 89–112. Springer, 2007.

F. J. Bruggeman, R. Planqué, D. Molenaar, and B. Teusink. Searching for principles of microbial physiology. *FEMS microbiology reviews*, 44(6):821–844, 2020.

E. Brunk, S. Sahoo, D. C. Zielinski, A. Altunkaya, A. Dräger, N. Mih, F. Gatto, A. Nilsson, G. A. Preciat Gonzalez, M. K. Aurich, et al. Recon3d enables a three-dimensional view of gene variation in human metabolism. *Nature biotechnology*, 36(3):272–281, 2018.

Y. Chiba and H. Akeboshi. Glycan engineering and production of 'humanized'glycoprotein in yeast cells. *Biological and Pharmaceutical Bulletin*, 32(5):786–795, 2009.

J. Contiero, C. Beatty, S. Kumari, C. DeSanti, W. Strohl, and A. Wolfe. Effects of mutations in acetate metabolism on high-cell-density growth of escherichia coli. *Journal of Industrial Microbiology and Biotechnology*, 24(6):421–430, 2000.

D. Davidi and R. Milo. Lessons on enzyme kinetics from quantitative proteomics. *Current opinion in biotechnology*, 46: 81–89, 2017.

D. H. De Groot, C. Van Boxtel, R. Planqué, F. J. Bruggeman, and B. Teusink. The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates. *PLOS Computational Biology*, 15(3): e1006858, 2019.

S. De Maeseneire, M. De Mey, S. Vandedrinck, and E. Vandamme. Metabolic characterisation of e. coli citrate synthase and phosphoenolpyruvate carboxylase mutants in aerobic cultures. *Biotechnology letters*, 28:1945–1953, 2006.

M. De Mey, S. De Maeseneire, W. Soetaert, and E. Vandamme. Minimizing acetate formation in e. coli fermentations. *Journal of Industrial Microbiology and Biotechnology*, 34(11):689–700, 2007.

J. S. Edwards and B. O. Palsson. Systems properties of the haemophilus influenzaerd metabolic genotype. *Journal of Biological Chemistry*, 274(25):17410–17416, 1999.

S. S. Fong, A. Nanchen, B. O. Palsson, and U. Sauer. Latent pathway activation and increased pathway capacity enable escherichia coli adaptation to loss of key metabolic enzymes. *Journal of Biological Chemistry*, 281(12):8024–8033, 2006.

J. Gagneur and S. Klamt. Computation of elementary modes: a unifying framework and the new binary approach. *BMC bioinformatics*, 5(1):1–21, 2004.

D. Heckmann, A. Campeau, C. J. Lloyd, P. V. Phaneuf, Y. Hefner, M. Carrillo-Terrazas, A. M. Feist, D. J. Gonzalez, and B. O. Palsson. Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers. *Proceedings of the National Academy of Sciences*, 117(37):23182–23190, 2020.

H. Kacser and J. Burns. Rate control of biological processes. In *Symp. Soc. Exp. Biol*, volume 27, pages 65–104, 1973.

H. U. Kim, T. Y. Kim, and S. Y. Lee. Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen acinetobacter baumannii aye. *Molecular BioSystems*, 6(2):339–348, 2010.

S. Klamt, G. Regensburger, M. P. Gerstl, C. Jungreuthmayer, S. Schuster, R. Mahadevan, J. Zanghellini, and S. Müller. From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints. *PLoS computational biology*, 13(4):e1005409, 2017.

M. Kratochvíl, S. E. Wilken, O. Ebenhöh, R. Schneider, and V. P. Satagopam. Cobrexa 2: tidy and scalable construction of complex metabolic models. *Bioinformatics*, page btaf056, 02 2025. ISSN 1367-4811. doi: 10.1093/bioinformatics/btaf056.

A. Kroll, Y. Rousset, X.-P. Hu, N. A. Liebrand, and M. J. Lercher. Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning. *Nature Communications*, 14(1):4139, 2023.

F. Li, L. Yuan, H. Lu, G. Li, Y. Chen, M. K. Engqvist, E. J. Kerkhoven, and J. Nielsen. Deep learning-based k cat prediction enables improved enzyme-constrained model reconstruction. *Nature Catalysis*, 5(8):662–672, 2022.

W. Liebermeister. The value structure of metabolic states. *bioRxiv*, page 483891, 2018.

G. Liu, S. Li, Q. Shi, H. Li, J. Guo, J. Ouyang, X. Jia, L. Zhang, S. You, and B. Qin. Engineering of saccharomyces pastorianus old yellow enzyme 1 for the synthesis of pharmacologically active (s)-profen derivatives. *Molecular Catalysis*, 507:111568, 2021.

C. P. Long, J. E. Gonzalez, N. R. Sandoval, and M. R. Antoniewicz. Characterization of physiological responses to 22 gene knockouts in escherichia coli central carbon metabolism. *Metabolic engineering*, 37:102–113, 2016.

G. Lozano Terol, J. Gallego-Jara, R. A. Sola Martínez, M. Cánovas Díaz, and T. de Diego Puente. Engineering protein production by rationally choosing a carbon and nitrogen source using e. coli bl21 acetate metabolism knockout strains. *Microbial cell factories*, 18:1–19, 2019.

B. H. M. Meldal, O. Forner-Martinez, M. C. Costanzo, J. Dana, J. Demeter, M. Dumousseau, S. S. Dwight, A. Gaulton, L. Licata, A. N. Melidoni, S. Ricard-Blum, B. Roechert, M. S.

Skyzypek, M. Tiwari, S. Velankar, E. D. Wong, H. Hermjakob, and S. Orchard. The complex portal–an encyclopaedia of macromolecular complexes. *Nucleic acids research*, 43 (Database issue):D479—84, January 2015. ISSN 0305-1048. doi: 10.1093/nar/gku975.

J. M. Monk, C. J. Lloyd, E. Brunk, N. Mih, A. Sastry, Z. King, R. Takeuchi, W. Nomura, Z. Zhang, H. Mori, et al. i ml1515, a knowledgebase that computes escherichia coli traits. *Nature biotechnology*, 35(10):904–908, 2017.

M. Mori, T. Hwa, O. C. Martin, A. De Martino, and E. Marinari. Constrained allocation flux balance analysis. *PLoS computational biology*, 12(6):e1004913, 2016.

K. Nakano, M. Rischke, S. Sato, and H. Märkl. Influence of acetic acid on the growth of escherichia coli k12 during high-cell-density cultivation in a dialysis reactor. *Applied microbiology and biotechnology*, 48:597–601, 1997.

J. D. Orth, I. Thiele, and B. Ø. Palsson. What is flux balance analysis? *Nature biotechnology*, 28(3):245–248, 2010.

J. G. Pan, J. S. Rhee, and J. M. Lebeault. Physiological constraints in increasing biomass concentration of escherichia coli b in fed-batch culture. *Biotechnology Letters*, 9:89–94, 1987.

J. M. Park, T. Y. Kim, and S. Y. Lee. Constraints-based genome-scale metabolic simulation for systems metabolic engineering. *Biotechnology advances*, 27(6):979–988, 2009.

M. A. Piirainen, H. Salminen, and A. D. Frey. Production of galactosylated complex-type n-glycans in glycoengineered saccharomyces cerevisiae. *Applied Microbiology and Biotechnology*, pages 1–15, 2022.

B. J. Sánchez, C. Zhang, A. Nilsson, P.-J. Lahtvee, E. J. Kerkhoven, and J. Nielsen. Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints. *Molecular systems biology*, 13(8):935, 2017.

A. Schmidt, K. Kochanowski, S. Vedelaar, E. Ahrné, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold, and M. Heinemann. The quantitative and condition-dependent escherichia coli proteome. *Nature biotechnology*, 34(1):104–110, 2016.

R. Schuster and S. Schuster. Refined algorithm and computer program for calculating all non–negative fluxes admissible in steady states of biochemical reaction systems with or without some flux rates fixed. *Bioinformatics*, 9(1):79–85, 1993.

M. Terzer. *Large scale methods to enumerate extreme rays and elementary modes*. PhD thesis, ETH Zurich, 2009.

R. Urbanczik. Enumerating constrained elementary flux vectors of metabolic networks. *IET systems biology*, 1(5):274–279, 2007.

S. van den Bogaard, P. A. Saa, and T. B. Alter. Sensitivities in protein allocation models reveal distribution of metabolic capacity and flux control. *Bioinformatics*, page btae691, 2024.

A. Varma and B. O. Palsson. Metabolic flux balancing: basic concepts, scientific and practical use. *Bio/technology*, 12 (10):994–998, 1994.

S. E. Wilken, M. Besancon, M. Kratochvíl, C. A. F. Kuate, C. Trefois, W. Gu, O. Ebenhöh, et al. Interrogating the effect of enzyme kinetics on metabolism using differentiable constraint-based models. *Metabolic engineering*, 74:72–82, 2022.

J. Zhao, T. Baba, H. Mori, and K. Shimizu. Effect of zwf gene knockout on the metabolism of escherichia coli grown on glucose or acetate. *Metabolic Engineering*, 6(2):164–174, 2004.