Heinrich Heine
Universität
Düsseldorf

Enzyme-constrained genome-scale metabolic modelling of *Staphylococcus aureus*, analysis methods, and in silico drug toxicity prediction

Inaugural thesis presented to the Faculty of Mathematics and Natural Sciences of
Heinrich Heine University Düsseldorf
for the degree of
Doctor of Philosophy in Natural Sciences
by

Hester Chapman

from London, United Kingdom

16 September 2025

From the Institute of Quantitative and Theoretical Biology of
Heinrich Heine University Düsseldorf

Printed by permission of the
Faculty of Mathematics and Natural Sciences of
Heinrich Heine University Düsseldorf

Examiners:
1. Prof. Dr. Oliver Ebenhöh
2. Prof. Dr. Martin Lercher

Date of the oral defence: 20.02.2026

Declaration of independence

I declare under oath that I have produced my thesis independently and without any undue assistance by third parties under consideration of the 'Principles for the Safeguarding of Good Scientific Practice' at Heinrich Heine University Düsseldorf.

Düsseldorf, 16 September 2025

Hester Elizabeth Chapman

Abstract

Staphylococcus aureus is the leading bacterial cause of death worldwide, and has the ability to quickly develop resistance to antimicrobials. Livestock-associated strains of this pathogen have become increasingly common over the last two decades, further contributing to the global burden of disease. Finding novel drug targets is key in the fight against this major human pathogen.

Using an enzyme-constrained genome-scale model of *S. aureus*, it was found that the enzymes most highly controlling of growth are involved in glycolysis, respiration, and the pentose phosphate pathway. Using novel analysis techniques revealed that optimal growth is achieved through a combination of respiration and fermentation. This flexibility is a potential contributing factor to the ability to colonise and infect different hosts. Several essential genes were found in agreement with experimental datasets, providing a starting point for the development of novel antimicrobials. The model highlights strain-specific metabolic vulnerabilities and serves as a resource for future drug target discovery.

In developing drugs to combat *S. aureus*, the ability to predict toxic effects in humans is crucial for reducing lead-time and minimising harm during clinical trials. A computational pipeline is introduced for predicting phase I and phase II drug metabolism in the liver. A pilot study is presented on the hepatotoxicity of paracetamol, to show the future uses of this tool. This framework demonstrates the potential of in silico approaches to anticipate toxic metabolites early in drug development, improving preclinical assessment and reducing reliance on animal models.

This thesis demonstrates the versatility of systems biology, and its applicability across distinct but interconnected biological fields. The analytical methods developed here refine the interpretation of enzyme-constrained models, while applications to bacterial pathogenicity and drug-induced toxicity illustrate their possible impacts in antimicrobial research and predictive toxicology.

Acknowledgements

First and foremost I would like to thank Prof. Dr. Oliver Ebenhöh for his exceptional supervision throughout my PhD. I am grateful for the insightful discussions, and the freedom he gave me to explore and develop as a researcher. His constant positivity and encouragement always kept me motivated. Thank you to Dr. St. Elmo Wilken for the invaluable guidance, and for the very occasional funny joke. I also wish to thank Dr. Miroslav Kratochvíl for sharing his programming expertise as well as his incredibly logical approach to writing.

I thank all current and past members of QTB for the friendly working environment. Particular thanks to Marvin and Yvan for being the best office buddies anyone could ask for, and to Chilperic for all the snacks. Thank you to Mara for providing a sense of calm.

I would like to thank everyone involved in Innotargets. In particular, thank you to Nader for teaching me all I know about working in a lab, and for making it enjoyable. Thank you to Linda and Gaia for the good times during the last few years.

Finally, I thank my family and friends for the unrelenting support, both from afar and in Düsseldorf. I would not have made it this far without them.

Contents

Chapter 1

Introduction

Antimicrobial resistance is a major challenge to global health, causing one in six deaths from bacterial infections worldwide [116, 77]. The gram-positive bacterium *Staphylococcus aureus* ranks as the second leading cause for resistance-associated deaths. A major human pathogen, *S. aureus* is responsible for a wide range of diseases, from minor skin infections to severe, life-threatening conditions such as sepsis and endocarditis. The growing prevalence of antibiotic-resistance has elevated methicillin-resistant *Staphylococcus aureus* (MRSA) to high-priority on the World Health Organisation bacterial pathogens list [119]. Addressing the global burden of MRSA requires an integrated approach, combining novel research into the bacterial metabolism and drug discovery with the responsible use of antimicrobials.

1.1 *Staphylococcus aureus*: a major pathogen

Methicillin-resistant *Staphylococcus aureus* (MRSA) has traditionally been associated with healthcare-acquired (HA-MRSA) and community-acquired infections (CA-MRSA). However, since the early 2000s a distinct clade of livestock-associated MRSA (LA-MRSA) has emerged, with the first human case reported in 2004 [174]. Currently, infections from LA-MRSA strains tend to be confined to animal husbandry, particularly pig farmers [7, 65]. However, infections are increasingly reported in individuals without any contact with animals [19, 156, 169, 81, 173]. The growing prevalence of LA-MRSA is of concern, affecting both animal populations and human health.

S. aureus has the ability to rapidly develop resistance to virtually any antimicrobial agent introduced into clinical or agricultural use [29]. Over time, the emergence of distinct MRSA lineages (HA-, CA-, LA-MRSA) has led to characteristic resistance profiles. For example, LA-MRSA is frequently resistant to tetracycline, reflecting its widespread use in intensive farming practices [124]. The close interplay between antimicrobial usage patterns and resistance evolution highlights the importance of understanding MRSA not only as a clinical problem but also as a broader agricultural and societal challenge.

Although the past decade has seen the development of several new antimicrobial and antistaphylococcal agents [9, 39, 120, 74], the continued overuse and misuse of antibiotics remains a key driver of the ongoing evolution of resistance. In many cases, resistance to new therapies can occur with only a few years of clinical deployment, as seen with penicillin and methicillin [133, 80]. This challenge is further compounded by the continued overuse and misuse of antibiotics in both human medicine and agriculture, which exerts a powerful selective pressure and accelerates the spread of resistance.

Addressing the resistance challenges requires a multifaceted approach. While expanding the arsenal of available treatments remains essential, complementary strategies such as improved antimicrobial stewardship, reduced agricultural antibiotic use, and the exploration of alternative therapies are also needed. These may include host-directed interventions, or metabolic pathway inhibitors that exploit vulnerabilities in bacterial physiology. Together, such approaches aim not only to extend the lifespan of existing antibiotics but also to diversify the tools available for combating resistant pathogens.

Within this broader context, early-stage drug discovery plays a critical role in shaping the pipeline of future therapies. Identifying promising targets at the metabolic level can provide a foundation for the rational design of novel antimicrobials.

1.2 Enzyme-constrained genome-scale metabolic modelling

One promising avenue for the combatting of antibiotic resistant bacteria is genome-scale metabolic modelling. This systems biology approach reconstructs the entire metabolic network of an organism from its genomic information, enabling the prediction of metabolic fluxes, nutrient requirements, and potential vulnerabilities under different conditions. By integrating experimental data, genome-scale models can identify essential or highly sensitive pathways that may serve as novel drug targets. In the case of *S. aureus* and other antimicrobial-resistant pathogens, such models provide valuable insights into how bacteria adapt their metabolism in response to antibiotic pressure, host environments, or nutrient limitations.

Typically, a genome-scale metabolic model (GSMM) is investigated using flux balance analysis (FBA), a linear programming (LP) framework based on the assumption that cell growth is at steady state, and subject to imposed constraints on certain reaction fluxes [121]. The standard formulation of FBA can be written as the following linear programme:

$$\begin{aligned} \text{maximise} \quad & v_r \\ \text{subject to} \quad & \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0} \\ & a_i \leq v_i \leq b_i \end{aligned} \tag{1.1}$$

In FBA, the objective reaction, v_r, typically representing biomass production and used

to simulate growth, is maximised. The first constraint enforces a steady state condition, involving the stoichiometric matrix, $\boldsymbol{S}$, which describes the relationships between metabolites and reactions, and the flux vector $\boldsymbol{v}$. The second constraint imposes minimum (a_i) and maximum (b_i) bounds on each reaction v_i.

In the context of multi-drug resistant pathogens, such as MRSA, FBA can been used on a GSMM to find novel drug targets. Essential genes in the target pathogen can be predicted by performing in silico gene knockouts and using the LP (1.1) to simulate growth. On simulating a knockout, if an optimal solution exists with $v_r > 0$, the gene is categorised as non-essential, but if no non-zero optimal solution can be found the gene is essential and a potential drug target. In this way, FBA models can provide a relatively low-cost first step in drug discovery [18, 151, 2, 131].

While FBA performs well in categorising essential genes, it assumes an unlimited proteome capacity, which can lead to unrealistic flux predictions. This limitation has been address through recent advances in enzyme-constrained genome-scale metabolic models (ecGSMMs). Incorporating enzyme kinetics is the next step in advancing knowledge of metabolism. Using an ecGSMM, one can look into the sensitivity of model variables to model parameters. Further possible drug targets can then be found among the enzymes that exert high control on growth or on virulence-related metabolic processes. This approach lays the groundwork for the prioritisation of potential drug targets, bridging the gap between computational modelling and experimental validation in the early phases of antimicrobial discovery.

The remainder of this section of the introduction provides a broad setup for ecGSMMs, which are extensively used in chapter 2 and 3 of this thesis. The original draft was written by myself, and reviewed by The Economic Cell Collective (2025). It is identical up to editorial changes to section 5.4 'Enzyme constrained FBA' in the online textbook Economic Principles in Cell Biology [32].

In its most simple form, flux balance analysis requires a stoichiometric matrix, an objective function, and at least one flux constraint to ensure that the problem is bounded. Solving an FBA problem allows for the prediction of intracellular fluxes and essential gene knockouts, given measured uptake and secretion rates.

However, whilst classical FBA can capture the effects of essential gene knockouts well, it falls down when it comes to non-lethal knockouts, and the prediction of growth phenotypes. For example, overflow metabolism in *E. coli*, and similarly the Crabtree and Warburg effects in *S. cerevisiae* and cancer cells respectively, cannot be captured in FBA models without ad-hoc flux constraints being imposed. Overflow metabolism, the Crabtree effect, and the Warburg effect broadly encompass the seemingly wasteful strategy of cells at high growth rates using a combination of respiration and fermentation, despite the higher

ATP-releasing efficiency of respiration per carbon molecule.

It has been proposed that overflow metabolism results from optimal protein allocation in the cell [14, 24]. In FBA models, we may impose total proteome constraints to perform enzyme-constrained flux balance analysis (ecFBA).

The usual formulation for ecFBA can be written as follows [145]:

$$\begin{aligned} \text{maximise} \quad & v_r \\ \text{subject to} \quad & \mathsf{(C1)}\ \boldsymbol{Sv} = \mathbf{0} \\ & \mathsf{(C2)}\ v_i \geq 0 \\ & \mathsf{(C3)}\ v_i = k_{\text{cat},i} \cdot e_i \ \ \forall i \in R \\ & \mathsf{(C4)}\ \sum_{i \in R_k} e_i \leq E_k \ \ \forall k \in [1, \ldots, K] \end{aligned} \tag{1.2}$$

Here, we wish to maximise flux through the objective reaction v_r, subject to four conditions. The matrix $\boldsymbol{S}$ is our stoichiometric matrix, with all reversible reactions split into a forward and a reverse reaction, and the condition C2 ensures that all fluxes are positive. In this formulation, we give all metabolic reactions an associated catalysing enzyme, and stipulate that the flux through a reaction is equal to the concentration of this enzyme multiplied by the apparent turnover number (k_{cat}) value (C3). Finally, we constrain total proteome constraints, in the form of enzyme pools. The total concentration of enzymes in the k-th enzyme pool must not exceed the constraint E_k (C4).

Predictions using ecFBA do not rely on the input of flux constraints, but rather good estimates for the total protein in different cellular compartments (for example the membrane and the cytosol), as well as the k_{cat}-values.

The ecFBA formulation ensures that all metabolic enzymes have an associated cost, relative to the gene product molar mass and turnover number. A simplified technique to provide a proxy for these costs is parsimonious flux balance analysis (pFBA). Central to pFBA is the assumption that cells minimize their total enzyme usage. Here, an optimal objective value is calculated via standard FBA, and the sum of the gene-associated reaction fluxes is then minimized. pFBA significantly reduces flux variability compared to standard FBA, but still does not typically capture overflow metabolism [177]. Thus, despite the additional required parameterisation, ecFBA is used in this thesis as a framework to investigate the metabolism of MRSA, with a focus on novel drug targets.

1.3 Drug discovery pipelines

New therapeutics against *Staphylococcus aureus* must be both effective at killing the pathogen, and non-harmful to the human or animal host. A key step in early drug tar-

get identification is ensuring that candidate targets lack close homologues in the human genome, thereby reducing the risk of off-target effects. Once a promising compound is identified, it must undergo extensive pre-clinical and clinical evaluation before reaching the market. At every stage, the therapeutic agent must demonstrate safety, tolerability, and minimal adverse effects to be considered a viable treatment option.

Typically, drug discovery encompasses five main steps: target discovery, hit identification, lead optimisation, pre-clinical trials, and clinical trials [146]. The use of GSMMs and ecGSMMs for target discovery is briefly described above. Wet-lab experiments and high-throughput screening are then typically used for hit identification. During screening, hits are defined as compounds with the desired activity against a target. This is an expensive process, and can be improved with virtual screening methods such as molecular docking. In this approach, the target is screened against a virtual library of compounds to generate an in silico prediction of the binding. Those compounds with good binding affinity with the target are then further developed as lead compounds during lead optimisation [142].

The pre-clinical phase involves assessment of the efficacy and safety of the lead compounds, providing the critical bridge between early-stage compound development and human clinical trials. These studies are typically conducted in vitro using cell lines, followed by in vivo studies in animal models. The results are used to establish safe starting doses for first-in-human trials, as well as to identify potential toxicities that might block further development. Despite the significant resources devoted to pre-clinical testing, a large proportion of compounds still fail to progress, most often due to toxicity concerns or lack of efficacy [171].

A major challenge in the pre-clinical phase is that experimental models do not always reliably predict human responses. Animal models frequently fail to capture the complexity of human metabolism, with agreement between human and animal studies essentially random [127]. This limitation has motivated the development of more predictive in silico approaches, which can either complement or partially replace traditional animal testing. With the United States Environmental Protection Agency pledging to drastically reduce animal testing by 2035 [66], there is more interest than ever in in silico alternatives. One area where such approaches are particularly valuable is in the assessment of drug toxicity.

The understanding of drug toxicity in human tissues, particularly the liver, remains a critical bottleneck in pharmaceutical research and development. Drug hepatotoxicity is a leading cause of clinical trial failures and post-market drug withdrawals [85]. Often, toxicity results from the breakdown of drugs in the liver into harmful substances, rather than from the parent compound itself [12]. In silico tools to predict these metabolites can provide early insights into the safety of a drug, offering a cost-effective strategy to reduce reliance on animal testing and improve the accuracy of pre-clinical tests.

The metabolism of drugs typically occurs in the liver, through a series of reactions categorised into phase I and phase II metabolism. Phase I metabolism adds or exposes polar groups through oxidation, reduction, and hydrolysis reactions. Phase II involves the conjugation of drugs or their phase I metabolites to polar ionic groups, increasing the water solubility to facilitate excretion. Being able to predict the compounds resulting from this drug metabolism enables early warnings of potential toxicity, reducing the reliance on animal models.

The incorporation of in silico toxicity assessments into pre-clinical research can therefore improve both efficiency and predictive accuracy. By simulating the metabolic fate of drug candidates and identifying potential toxic metabolites, these methods provide an early warning system that can guide compound selection before expensive in vivo studies are undertaken. Ultimately, in silico approaches hold promise not only for reducing reliance on animal models but also for improving translational success rates, supporting safer drug discovery.

1.4 Thesis overview

This thesis aims to provide a starting point for a systems biology-based investigation into the metabolism of methicillin-resistant *Staphylococcus aureus*, alongside key considerations of drug toxicity.

Chapter 2 presents the development of novel analysis tools for ecGSMMs, aimed at exploring optimal pathway usage in an assortment of biological contexts. The central contribution of this chapter is the introduction of optimal flux modes (OFMs), a framework for decomposing optimal solutions into distinct growth-supporting modes. In an ecGSMM, multiple OFMs may be used in an optimal solution. For example, during overflow metabolism, a combination of one fermentative and one respiratory OFM may be used. By differentiating the relative usage of these OFMs, the method presented highlights enzymes with strong control over metabolic flux, which can be prioritised as candidates for metabolic engineering or targeted drug development.

Next, chapter 3 of the thesis focusses on the reconstruction of an enzyme-constrained genome-scale metabolic model of livestock-associated *S. aureus*, analysed using both traditional techniques and those introduced in chapter 2. Several enzymes in the model are found to be highly controlling of the optimal metabolic strategy, providing a starting point for research into drug targets of LA-MRSA.

An in silico framework for predicting drug-induced liver toxicity is presented in chapter 4. A pipeline is developed for predicting the metabolic breakdown of drugs via phase I and phase II metabolism. The results of this pipeline are then fed into an existing enzyme-substrate predictor using enzymes from the liver, to investigate potential toxic effects of

metabolised compounds. Using paracetamol as a case study, the pipeline is able to capture the known toxic metabolite of this drug. This chapter is a promising starting point for the development of new pre-clinical tools.

Together, these chapters reflect the versatility and capabilities of systems biology approaches in addressing current challenges in biological and biomedical research. The tools and models developed here contribute to the growing body of knowledge in metabolic systems biology, with a grounding in rigorous theory. Through the focus on pathogenic metabolism and hepatotoxicity, this work has the potential to enhance drug discovery and predictive toxicology.

Chapter 2

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

Hester Chapman, Miroslav Kratochvíl, Oliver Ebenhöh, St Elmo Wilken.

This chapter presents novel analysis tools for enzyme-constrained models, and up to editorial changes is identical to the published paper [31]. I developed the basic concepts, performed formal mathematical and computational analysis, proved mathematical theorems, composed the original draft, and was involved in writing the final paper.

2.1 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.

2.2 Acknowledgments

This work was supported by funding from the European Union's Horizon 2020 Programme under the Innotargets Project (`www.innotargets.ku.dk`) [956154] (HC, OE). This work was also supported by the European Union's Horizon 2020 Programme under the PerMedCoE Project (`www.permedcoe.eu`) [951773] (MK). OE and SEW were sup-

ported by Deutsche Forschungsgemeinschaft under Germany's Excellence Strategy—EXC-2048/1—project ID [390686111].

2.3 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.

2.4 Introduction

Genome-scale metabolic models (GSMMs) provide a framework for studying cellular metabolism [172, 54]. 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) [121]. Main applications include metabolic engineering [125], investigating human metabolism and disease [25], and pathogen-drug target

identification [87].

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 [14, 24]; 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) [145, 15].

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 [42]. 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 [93].

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 [176]. 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 [99, 20]. 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) [44]. 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) [82]. 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 [98], 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.

2.5 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 [176]. Detailed proofs of all lemmas and theorems are presented in the Supplementary material.

2.5.1 Model formulation

Following the model formulation from Sanchez [145], we describe the enzyme-constrained FBA (ecFBA) optimisation problem applied to a simple ecGSMM as follows:

$$\begin{aligned} \text{maximise} \quad & v_r \\ \text{subject to} \quad & \text{(C1)}\ \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0} \\ & \text{(C2)}\ v_i \geq 0 \\ & \text{(C3)}\ v_i = k_{\text{cat},i} \cdot e_i \ \ \forall i \in R \\ & \text{(C4)}\ \sum_{i \in R_k} e_i \leq E_k \ \ \forall k \in [1, \ldots, K] \end{aligned} \tag{2.1}$$

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, $k_{\text{cat},i}$. 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 (2.1). 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 [121]. This is the matrix we use to find elementary flux modes of a network.

2.5.2 Flux cones and elementary flux modes

The set of flux vectors $\boldsymbol{v}$ satisfying both $\boldsymbol{S}\boldsymbol{v} = \boldsymbol{0}$ and $v_i \geq 0$ forms a convex polyhedral cone, called the flux cone [89]:

$$\text{FC}(\boldsymbol{S}) = \{\boldsymbol{v} \in \mathbb{R}^n | \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0}, v_i \geq 0\} \tag{2.2}$$

Any non-trivial flux vector $\boldsymbol{v}$ is called an admissible flux mode if it is contained in the flux cone $\text{FC}(\boldsymbol{S})$ (2.2). We define the support of a mode, $\textbf{supp}(\boldsymbol{v}) = \{i | v_i \neq 0\}$ 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*

$\boldsymbol{supp}(\boldsymbol{m}) \supset \boldsymbol{supp}(\boldsymbol{v})$. *Thus,* $\boldsymbol{supp}(\boldsymbol{m})$ *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 $\boldsymbol{v} = 0$.

2.5.3 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 (2.1) can be done as follows:

$$\begin{aligned} \text{maximise} \quad & v_r \\ \text{subject to} \quad & \mathsf{(C1\text{-}C4)} \\ & \mathsf{(C5)}\ v_j \geq c_j > 0\ \forall j \in R_f \end{aligned} \tag{2.3}$$

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 (2.3) will no longer form a flux cone, but rather the flux polyhedron $\mathrm{P}(\boldsymbol{S})$, defined as:

$$\mathrm{P}(\boldsymbol{S}) = \{\boldsymbol{v} \in \mathbb{R}^n | \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0},\ v_i \geq 0,\ v_j \geq c_j\ \forall j \in R_f\} \tag{2.4}$$

We distinguish between the homogeneous problem (2.1) and the inhomogeneous problem (2.3) since EFMs are not defined for flux polyhedra such as (2.4) [89]. 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 [149], 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:

$$\begin{aligned} \text{maximise} \quad & \overline{v} \\ \text{subject to} \quad & \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0} \\ & x_i \geq 0 \\ & x_i = k_{\text{cat},i} \cdot e_i \;\; \forall i \in R \\ & \sum_i e_i \leq E_k \;\; \forall k \in [1, \ldots, K] \\ \text{where} \quad & \boldsymbol{A} := \begin{bmatrix} \boldsymbol{S^{(1)}} & \boldsymbol{w} \end{bmatrix}, \; \boldsymbol{x} := \begin{bmatrix} \boldsymbol{v^{(1)}} \\ \overline{v} \end{bmatrix}. \end{aligned} \tag{2.5}$$

When $\overline{v} = 1$, solutions to the homogeneous system (2.5) coincide precisely with solutions of problem (2.3).

A non-trivial vector $\boldsymbol{x}$ is called an admissible flux mode of (2.3) if it is contained in the optimal flux cone

$$\text{OFC}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}, v_i \geq 0\} \tag{2.6}$$

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 (2.3) 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* $\boldsymbol{supp}(\boldsymbol{x}) \supset \boldsymbol{supp}(\boldsymbol{y})$
- *Optimality:* $\overline{v} = 1$.

The first two conditions for an OFM coincide with Definition 1 of an EFM of the optimal flux cone (2.6); 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 [89, 168], 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.

2.5.4 Differentiating solutions

The solution of a convex optimisation problem, such as (2.5), with a unique optimal solution, may be implicitly differentiated [3]. We previously showed that one can equate these differentials to model sensitivities [176]. If there are multiple optimal solutions to problem (2.3), 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 (2.1) also apply to inhomogeneous problems (2.3).

Lemma 1. *The extreme rays of problem (2.5) with a non-zero $\overline{v}$ correspond to the optimal flux modes (OFMs) of (2.3).*

The proof follows from the result of Gagneur [59] that extreme rays of a pointed polyhedral cone $\mathrm{FC}(\boldsymbol{S}) = \{\boldsymbol{x} | \boldsymbol{S}\boldsymbol{x} = \boldsymbol{0}\ , x_i \geq \boldsymbol{0}\}$ correspond to elementary flux modes of the homogeneous problem (2.1). See the Supplementary material for details.

De Groot et al. previously proved that a homogeneous problem (2.1) with K enzymatic constraints will use at most K EFMs in an optimal solution [44]. We extend this result to inhomogeneous problems (2.3) and OFMs.

Lemma 2. *An optimal solution to problem (2.3) with K enzymatic constraints (C4), rescaled to be in the form*

$$C_{\Sigma}^{(k)} := \sum_{j=1}^{r} w_j^{(k)} e_j \leq 1 \ \ \textit{for} \ \ k \in \{1, ..., K\} \tag{2.7}$$

will use at most K optimal flux modes.

In short, we know from de Groot [44] that in the homogeneous problem (2.1), at most K EFMs will be active in an optimal solution. Combining this with Lemma 3, we prove that

an optimal solution to (2.3) 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.

2.5.5 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 (2.1), 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 [44], we define a cost matrix of the EFMs in a solution, and rewrite the optimisation problem (2.1) 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 (2.3), we find again that the pruned solution is unique:

Corollary 1. *The optimal solution of a pruned ecGSMM of the form (2.3) is always unique.*

Proof. We can rewrite problem (2.3) as problem (2.5), and then use the proof of Theorem 2, equating EFMs in Theorem 2 with OFMs in the current problem.

□

Corollary 2. *Pruned models in the form of (2.1) or (2.3) 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. [3]. Theorem 2 and Corollary 3 ensure that pruned models do indeed have unique optimal solutions. Therefore, these

solutions are also implicitly differentiable, using the method of Wilken et al. [176].

□

Corollary 4 is required to guarantee that the strategy from [176] 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 4. Below, we describe the implementation of implicit differentiation for OFM sensitivities.

2.6 Methods

2.6.1 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 (2.1) or (2.3), 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.

2.6.2 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. [159], 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 [90]. 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 [113], and other linear programming problems, provided a unique optimal solution can be guaranteed.

2.6.3 Models

Enzyme-constrained fungi models

Using 342 published fungal GSMMs [98], 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 [112] was used as the base GSMM. Enzyme turnover numbers were taken from [71]. If available, enzyme subunit stoichiometries were taken from the complex portal [109], otherwise Swiss-Prot entries [23] 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 $160\frac{\text{mg}}{\text{g}_{\text{DCW}}}$ membrane protein, and $340\frac{\text{mg}}{\text{g}_{\text{DCW}}}$ 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 [147].

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 [57, 45, 38, 102, 180, 101].

2.7 Results

2.7.1 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× more efficient than finite differencing (Table 2.1) on full models, and more than 25× 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

			CPU time consumed (s)	
Repr.	Model	#vars	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	**3.81** ± 0.15	144.90 ± 0.99
	iML1515	2	**3.87** ± 0.19	99.58 ± 0.77

Table 2.1: 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.

for each method, was around 7× in the benchmark.

2.7.2 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.

Figure 2.1 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 [128]. 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 [34], or proposed or used for other bio-processes [17, 100, 1].

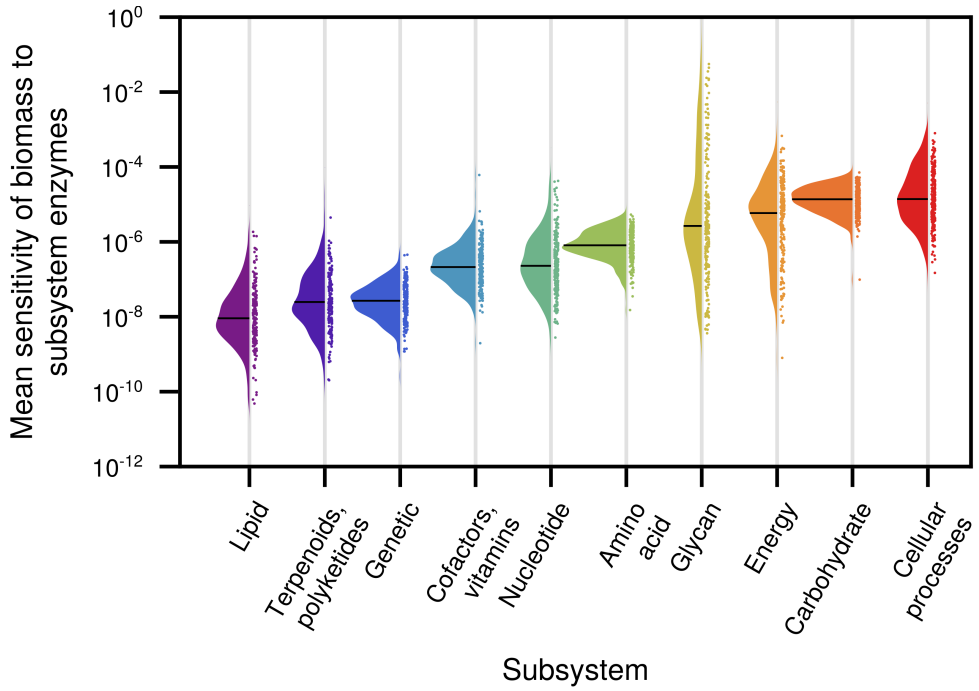


Figure 2.1: The mean sensitivity of biomass reactions across 342 fungal species is consistent across all subsystems except glycan. Each point on the plot is the mean sensitivity of the growth of a model to its enzyme turnover numbers in a subsystem. Rain clouds show the spread across sensitivity of model growth rates to kinetic parameters in one subsystem.

2.7.3 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 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 [46], since its presence can slow down growth [117] and decrease product yield [123]. 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 4, 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 2.2 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 $v = k_{\text{cat}} \cdot e$) and there will hence be more available space in the membrane for every other membrane bound enzyme.

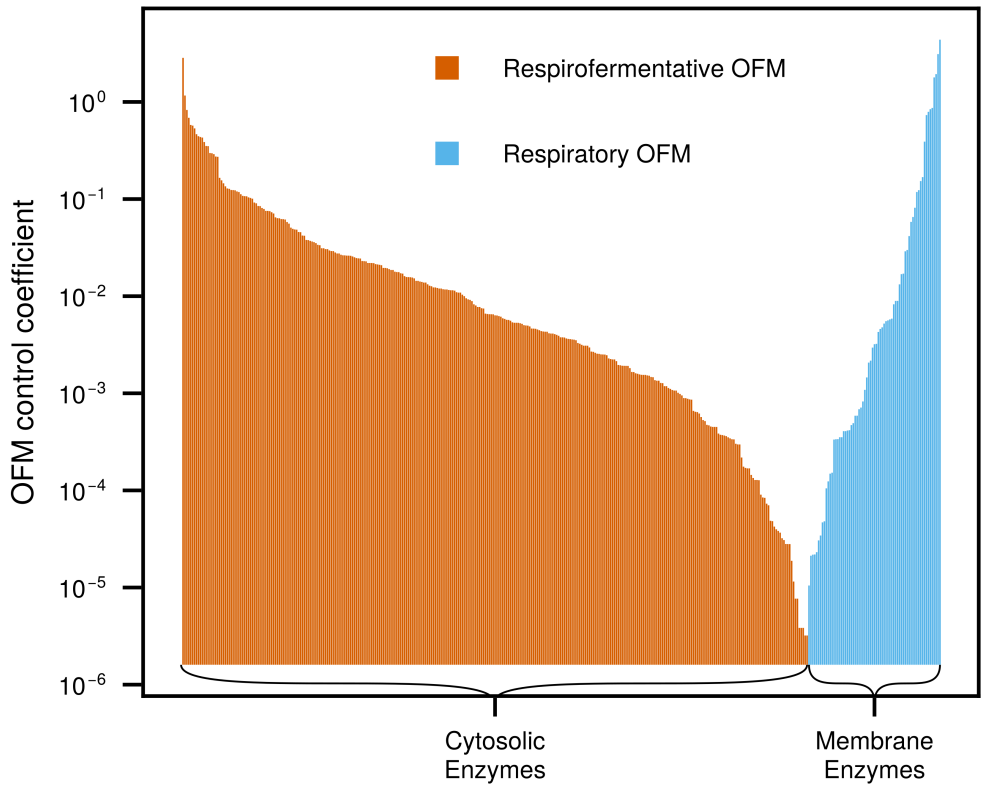


Figure 2.2: Control coefficients of OFM usage in *E. coli*, $\frac{p}{\lambda}\frac{\partial \lambda}{\partial p}$, support the assumption that space in the membrane controls respiration. Along the x-axis we have every enzyme parameter, p, in the active solution, and we plot on the y-axis whichever of $\frac{p}{\lambda_1}\frac{\partial \lambda_1}{\partial p}$ 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.

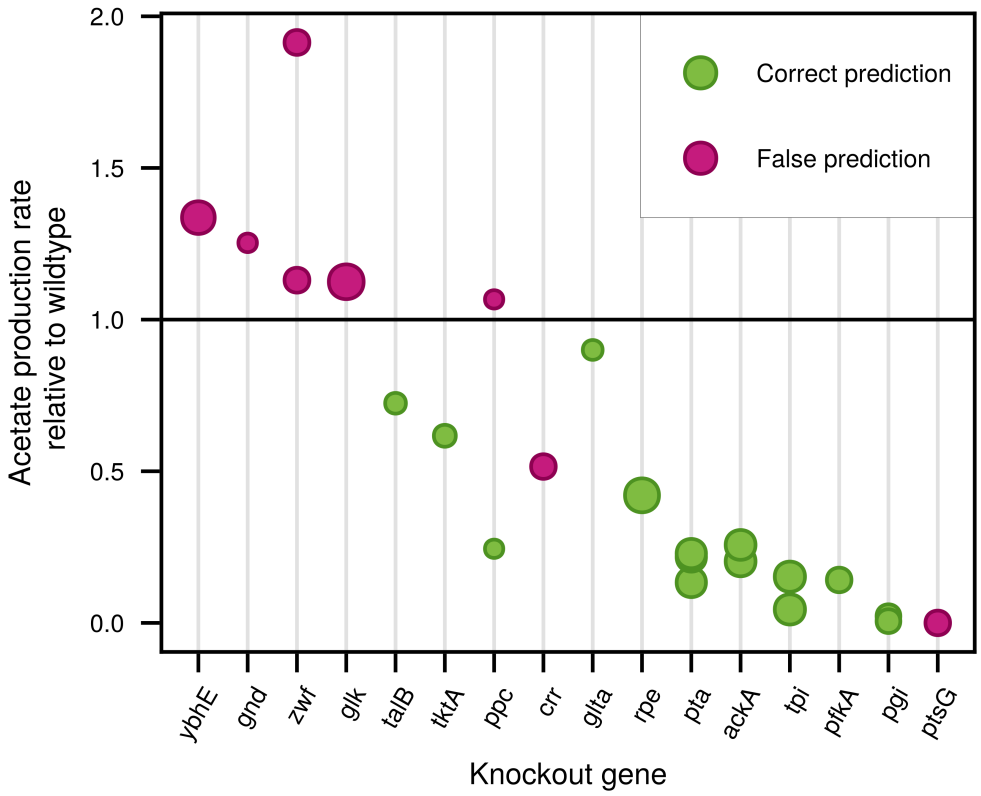


Figure 2.3: Sensitivity of the respirofermentative OFM in *E. coli* provides qualitatively accurate predictions of those gene knockouts that decrease acetate production. We plot the acetate production rate in 23 knockout mutants of 16 different genes relative to the wildtype, taken from experimental 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 2.3, 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 in Figure 2.3, 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.

2.8 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 [176], 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.

2.9 Supplementary information

2.9.1 Toy model

In this section, we provide a simplified enzyme constrained metabolic model (Figure 2.4), to demonstrate pruning optimal solutions, calculating OFMs, and to highlight the difference between EFMs and OFMs. We constrain the model with two distinct enzyme pools. We give reactions R_3 to R_5 appropriate k_{cat} values and gene product molar masses to ensure that both R_5 and reactions R_3 and R_4 are used in the optimal solution, but R_8 is not.

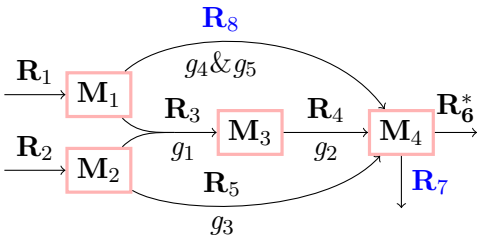
To demonstrate calculation of EFMs in an optimal solution, the LP is set up as follows:

$$\begin{aligned} \text{maximise} \quad & v_6 \\ \text{subject to} \quad & \mathsf{(C1)}\ \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0} \\ & \mathsf{(C2)}\ v_i \geq 0 \ \ \forall i \\ & \mathsf{(C3)}\ v_i = k_{\text{cat},i} \cdot e_i \ \ \forall i \in [3,4,5] \\ & \mathsf{(C4)}\ e_1 + e_2 \leq E_1 \\ & \mathsf{(C5)}\ e_3 + e_4 + e_5 \leq E_2 \end{aligned} \tag{2.8}$$

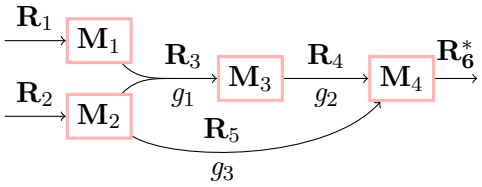
When we optimise, we get the solution in Figure 2.4e, and can see that reaction R_7 and R_8 have zero flux. Therefore, these two reactions are pruned from the model, and we are left with the network in Figure 2.4b and the stoichiometric matrix in 2.4d. When we then optimise the LP in (2.8), using this pruned stoichiometric matrix $\boldsymbol{S}$, we find the optimal solution in 2.4f. By theorem 2, we know that this is a unique optimal solution, and thus calculate the EFMs given in 2.4g. In 2.4h, we give the usage of the two EFMs in the optimal solution.

Adding the condition (C6), $v_7 = 15$, to the LP in (2.8), we ensure that reaction R_7 carries flux. This is included to demonstrate how we would treat a reaction in a genome scale model with an active inhomogeneous flux constraint or a known flux, for example a fixed glucose exchange or an ATP maintenance reaction. We now no longer have a flux cone but rather a flux polyhedron. We therefore need to calculate OFMs, rather than EFMs. First, we solve the LP, and see in Figure 2.5e that reaction R_8 is now the only reaction carrying zero flux. Pruning the model removes R_8, and when we resolve the LP with the pruned stoichiometric matrix from 2.5d, we once again find a unique optimal solution.

The difference between how to calculate OFMs and EFMs now becomes evident. After pruning, the first step is to separate the pruned stoichiometric matrix into $\boldsymbol{S}^{(1)}$ and $\boldsymbol{S}^{(2)}$, the columns corresponding to unfixed and fixed reactions, respectively. In this case, we know that $v_7 = 15$ and the objective $v_6 = 60$ in the optimal solution. Therefore, reactions



(a) Model network



(b) Pruned network

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & -1 & 1 \end{bmatrix}$$

(c) Full stoichiometric matrix

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 \end{bmatrix}$$

(d) Pruned stoichiometric matrix

$$\boldsymbol{v}^* = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{bmatrix} = \begin{bmatrix} 50 \\ 75 \\ 50 \\ 50 \\ 25 \\ 75 \\ 0 \\ 0 \end{bmatrix}$$

(e) Optimal solution (full model)

$$\boldsymbol{v}^* = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{bmatrix} = \begin{bmatrix} 50 \\ 75 \\ 50 \\ 50 \\ 25 \\ 75 \end{bmatrix}$$

(f) Optimal solution (pruned model)

$$\mathbf{EFM} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & 1 \end{bmatrix}$$

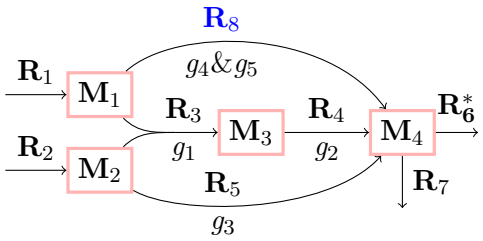
(g) EFMs

$$\boldsymbol{v}^* = 50 \cdot \mathbf{EFM}_1 + 25 \cdot \mathbf{EFM}_2$$

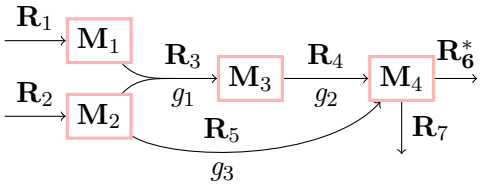
(h) Proportions of EFMs in optimal solution

Figure 2.4: Procedure to calculate the EFMs in the network given in (a). Reactions shown in blue have zero flux in the optimal solution and are pruned before calculating EFMs.

R_6 and R_7 are taken into $\boldsymbol{S}^{(2)}$. All other fluxes are variables, and will be taken into $\boldsymbol{S}^{(1)}$, as seen in Figure 2.6b and 2.6a. Next, we need to multiply $\boldsymbol{S}^{(2)}$ by $\boldsymbol{v}^{(2)} = \begin{bmatrix} v_6 & v_7 \end{bmatrix}'$ to use this in the matrix $\boldsymbol{A}$ (2.6d), creating a slack variable. Running the double description algorithm, we calculate the extreme rays of the optimal flux cone $\mathbf{OFC}(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n | \boldsymbol{A}\boldsymbol{x} = \mathbf{0}, v_i \geq 0\}$ as Figure 2.6e. The last row corresponds to the slack variable introduces into the last column of $\boldsymbol{A}$. If required, we rescale the extreme rays so that this last row is equal to $\begin{bmatrix} 1 & 1 \end{bmatrix}$, and reintroduce R_6 and R_7 into their original positions and with their original fixed values. The result of this is the final OFMs, given by the two columns in Figure 2.6f. We see that both OFMs produce the optimal objective flux, $v_6 = 60$, and use the required $v_7 = 15$. As with EFMs, linearly scaled OFMs are equivalent, so we scale to have flux through the objective R_6 equal to 1 in both OFMs. The optimal solution can



(a) Model network



(b) Pruned network

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & -1 & 1 \end{bmatrix}$$

(c) Full stoichiometric matrix

$$\boldsymbol{S} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -1 & -1 \end{bmatrix}$$

(d) Pruned stoichiometric matrix

$$\boldsymbol{v}^* = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{bmatrix} = \begin{bmatrix} 50 \\ 75 \\ 50 \\ 50 \\ 25 \\ 60 \\ 15 \\ 0 \end{bmatrix}$$

(e) Optimal solution (full model)

$$\boldsymbol{v}^* = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{bmatrix} = \begin{bmatrix} 50 \\ 75 \\ 50 \\ 50 \\ 25 \\ 60 \\ 15 \end{bmatrix}$$

(f) Optimal solution (pruned model)

Figure 2.5: Pruning a model with inhomogeneous flux constraints

be decomposed into $\boldsymbol{v}^* = 40 \cdot \mathrm{OFM}_1 + 20 \cdot \mathrm{OFM}_2$.

2.9.2 Theory

Here, we provide the mathematical proofs of all lemmas and theorems lacking a formal proof in the main text.

Optimal flux modes

Lemma 3. *The extreme rays of problem (2.5) with a non-zero $\overline{v}$ correspond to the optimal flux modes (OFMs) of (2.3).*

Proof. Previously, Gagneur [59] proved that extreme rays of a pointed polyhedral cone $\mathrm{FC}(\boldsymbol{S}) = \{\boldsymbol{x} | \boldsymbol{S}\boldsymbol{x} = \boldsymbol{0}\ , x_i \geq \boldsymbol{0}\}$ correspond to the elementary flux modes of the problem (2.1).

We define the pointed polyhedral cone $\mathrm{FC}(\boldsymbol{A}) = \{\boldsymbol{x} | \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0},\ \boldsymbol{x} \geq \boldsymbol{0}\}$, where $\boldsymbol{A} := \begin{bmatrix} \boldsymbol{S}^{(1)} & \boldsymbol{w} \end{bmatrix}$ as in problem (2.5). The cone $\mathrm{FC}(\boldsymbol{A})$ is in the same form as $\mathrm{FC}(\boldsymbol{S})$, where $\boldsymbol{S}$ is the stoichiometric matrix of (2.1). Thus, the extreme rays of $\mathrm{FC}(\boldsymbol{A})$ correspond to the elementary flux modes of $\boldsymbol{A}$ [59].

$$\boldsymbol{S}^{(1)} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

(a) Unknown flux stoichiometries

$$\boldsymbol{S}^{(2)} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ -1 & -1 \end{bmatrix}$$

(b) Fixed flux stoichiometries

$$\boldsymbol{w} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -75 \end{bmatrix}$$

(c) $\boldsymbol{S}^{(2)}\boldsymbol{v}^{(2)}$

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & -75 \end{bmatrix}$$

(d) Augmented matrix, $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{S}^{(1)} & \boldsymbol{w} \end{bmatrix}$

$$\text{extreme rays} = \begin{bmatrix} 75 & 0 \\ 75 & 75 \\ 75 & 0 \\ 75 & 0 \\ 0 & 75 \\ 1 & 1 \end{bmatrix}$$

(e) Extreme rays of OFC($\boldsymbol{A}$)

$$\mathbf{OFM} = \begin{bmatrix} 75 & 0 \\ 75 & 75 \\ 75 & 0 \\ 75 & 0 \\ 0 & 75 \\ 60 & 60 \\ 15 & 15 \end{bmatrix} = \begin{bmatrix} 1.25 & 0 \\ 1.25 & 1.25 \\ 1.25 & 0 \\ 1.25 & 0 \\ 0 & 1.25 \\ 1 & 1 \\ 0.25 & 0.25 \end{bmatrix}$$

(f) OFMs of P($\boldsymbol{S}$)

$$\boldsymbol{v}^* = \lambda_1 \cdot \mathrm{OFM}_1 + \lambda_2 \cdot \mathrm{OFM}_2 = 40 \cdot \mathrm{OFM}_1 + 20 \cdot \mathrm{OFM}_2 \qquad (2.9)$$

(g) OFMs in optimal solution

Figure 2.6: Implementation in ElementaryFluxModes.jl of calculating OFMs of a pruned optimal solution.

Elementary flux modes, and optimal flux modes, are the elementary modes of the associated polyhedral cone. An EFM or OFM producing one unit of objective flux need not adhere to enzyme constraints, since the optimal solution will use a scaled sum of the EFMs or OFMs. Thus, we can always scale an extreme ray $\boldsymbol{x} = \begin{bmatrix} \boldsymbol{v^{(1)}} & \overline{v} \end{bmatrix}$ of FC($\boldsymbol{A}$) so that $\overline{v} = 1$. Therefore, all vectors $\boldsymbol{x}$ admissible in FC($\boldsymbol{A}$), with a non-zero $\overline{v}$, must carry flux in the optimal ratio of the objective to the fixed fluxes.

To illustrate this optimal ratio, let $\boldsymbol{x}$ be the vector of fluxes $\boldsymbol{x} = [v_1, \ldots, v_r, , \overline{v}]$, where $v_1 = c_1$ is a fixed known flux and v_r is the objective flux, with optimal value c_r. All admissible vectors in FC($\boldsymbol{A}$) must obey $v_1/v_r = c_1/c_r$. Therefore, as extreme rays of FC($\boldsymbol{A}$) give the elementary flux modes of A, these extreme rays equivalently correspond to the optimal flux modes of (2.3).

□

Lemma 4. *An optimal solution to problem (2.3) with K enzymatic constraints (C4), rescaled to be in the form*

$$C_{\sum}^{(k)} := \sum_{j=1}^{r} w_j^{(k)} e_j \leq 1 \ \ for \ \ k \in \{1, ..., K\} \tag{2.10}$$

will use at most K optimal flux modes.

Proof. Firstly, we know that extreme rays of the homogeneous polyhedral cone FC($\boldsymbol{S}$) = $\{\boldsymbol{v} | \boldsymbol{S}\boldsymbol{v} = \boldsymbol{0}, \ \boldsymbol{v} \geq \boldsymbol{0}\}$ for a stoichiometric matrix $\boldsymbol{S}$ from problem (2.1) correspond to the elementary flux modes [59]. We also know that an optimal solution to (2.1) with K enzymatic constraints in the form (2.10) will use at most K EFMs [44]. Therefore, since extreme rays of FC($\boldsymbol{S}$) correspond to EFMs of (2.1), there will be at most K extreme rays used in an optimal solution to FC($\boldsymbol{S}$) with K linear enzymatic constraints.

Now we consider problem (2.3). From Lemma 3, we know that extreme rays with non-zero $\overline{v}$ of the polyhedral cone FC($\boldsymbol{A}$) where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{S^{(1)}} & \boldsymbol{w} \end{bmatrix}$ correspond to optimal flux modes of the problem (2.3).

We can transform the inhomogeneous problem (2.3) with K enzymatic constraints into the form (2.5), and we obtain a polyhedral cone FC($\boldsymbol{A}$) = $\{\boldsymbol{v} | \boldsymbol{A}\boldsymbol{v} = \boldsymbol{0}, \ \boldsymbol{v} \geq \boldsymbol{0}\}$ where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{S^{(1)}} & \boldsymbol{w} \end{bmatrix}$. We may now regard $\boldsymbol{A}$ in the same way as the above $\boldsymbol{S}$, and see that an optimal solution to FC($\boldsymbol{A}$) with K enzymatic constraints uses at most K extreme rays. The extreme rays of FC($\boldsymbol{A}$) with non-zero $\overline{v}$ correspond to the optimal flux modes of (2.3), and thus an optimal solution will use at most K OFMs.

□

A pruned optimal solution is unique

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

Theorem 2. *Given a model of the form (2.1), 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.*

Proof. Let the objective of the LP (2.1) be to maximise the flux through reaction r. We first prove the simplest case where there is only one capacity constraint.

Assume that the model has one capacity constraint in the form $C_\Sigma = \sum_{j=1}^{r} w_j e_j \leq 1$, where w_j is the fraction of the enzyme pool that one mole of the j-th enzyme uses up. In this case, if the problem is solvable then any solution will be composed of one EFM [44]. Pruning the model will thus leave only the reactions that are in this EFM. Maximising the objective will leave no flux variability, since knowing the flux through one reaction in an EFM determines the flux through every other reaction. There will also be no variation in the enzyme concentrations if the turnover numbers and masses are sufficiently different, which we can assume is the case in a full-scale ecGSMM. As required, we thus have a unique optimal solution in a pruned model with one capacity constraint.

It now remains to prove that a pruned model with $K > 1$ capacity constraints has a unique optimal solution. Following the approach of de Groot et al. [44], if we have K active enzyme constraints, then we have $C_\Sigma^{(k)} = \sum_{j=1}^{r} w_j^k e_j^k = 1$ for every constraint k, where e_j^k gives the concentration of enzyme j in EFM k. We have at most K EFMs in our optimal solution. The flux through reaction i in EFM k is denoted V_i^k. Each EFM is then rescaled such that the flux through the objective reaction, r, through each EFM k, denoted V_r^k, is equal to 1:
$\mathbf{EFM}^k = (V_1^k, \ldots, V_{r-1}^k, 1)$. The flux solution can now be written as a linear combination of these EFMs:

$$\boldsymbol{v} = \lambda_1 \mathbf{EFM}^1 + \ldots + \lambda_K \mathbf{EFM}^K, \text{ where } \lambda_k \geq 0 \quad \forall\ 1 \leq k \leq K \tag{2.11}$$

The quantity of objective flux in the optimal solution contributed by the k-th EFM is denoted by λ_k.

The fluxes V_j^k through each EFM are fixed, since knowing one flux in an EFM fully determines every other reaction flux. It now remains to show that the λ_i are also fixed.

Since $v_i = k_{cat,i} \cdot e_i$, the enzyme constraints can be written as

$$\begin{aligned} 1 = C_\Sigma^{(k)} &= \sum_{j=1}^{r} w_j^k e_j^k \\ &= \sum_{j=1}^{r} w_j^k \sum_{i=1}^{K} \lambda_i \frac{V_j^i}{k_{cat,j}} \\ &= \sum_{i=1}^{K} \lambda_i \sum_{j=1}^{r} w_j^k \frac{V_j^i}{k_{cat,j}} \\ &= \sum_{i=1}^{K} \lambda_i \sum_{j=1}^{r} d_k^i \end{aligned} \tag{2.12}$$

where $d_k^i := \sum_{j=1}^{r} w_j^k \frac{V_j^i}{k_{\mathrm{cat},j}}$ denotes the total protein cost to constraint pool k of one unit of objective flux through $\mathbf{EFM}^i$.

From equation (2.11), the objective flux v_r can be written as

$$v_r = \lambda_1 + \ldots + \lambda_K \tag{2.13}$$

meaning we can rewrite the optimisation as a maximisation of the sum of the λ_i, that must adhere to the constraint given by (2.12):

$$\max_{\boldsymbol{\lambda}} \{ \sum_{i}^{K} \lambda_i : \quad \lambda_i \geq 0, \quad \boldsymbol{D\lambda} \leq 1 \} \tag{2.14}$$

where $\boldsymbol{D} = [\boldsymbol{d^1} \ldots \boldsymbol{d^L}]$.

We now build upon the work of de Groot et al. [44] to prove that pruning the model does indeed leave a unique optimal solution.

Due to the setup of the problem, an optimal solution with $K > 1$ enzyme constraints will use $L \leq K$ EFMs, and these EFMs will always satisfy $\lambda_1 + \cdots + \lambda_L = v_r$. Therefore, we can always write the linear system of equations

$$\begin{bmatrix} \mathbf{1} \\ \boldsymbol{D} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_L \end{bmatrix} = \begin{bmatrix} v_r \\ \mathbf{1} \end{bmatrix} \tag{2.15}$$

where $\begin{bmatrix} \mathbf{1} & \boldsymbol{D} \end{bmatrix}^T$ is a $(K+1) \times L$ matrix. The system (2.15) has a unique solution when $\begin{bmatrix} \mathbf{1} & \boldsymbol{D} \end{bmatrix}^T$ has rank L, which is the case if the L cost vectors of the L EFMs are linearly independent. Where there is difference in enzyme costs, the cost vectors will be linearly independent, which implies that the solution to (2.14) will be unique. Thus, the optimal solution of a pruned model in the form (2.1) is unique.

□

In the setup where we have inhomogeneous constraints (2.3), we find again that the pruned solution is unique:

Corollary 3. *The optimal solution of a pruned ecGSMM (2.3) is always unique.*

Proof. We can rewrite problem (2.3) as problem (2.5), and then use the proof of Theorem 2, equating EFMs in Theorem 2 with OFMs in the current problem.

□

Corollary 4. *Pruned models in the form of (2.1) or (2.3) 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. [3]. Theorem 2 and Corollary 3 ensure that pruned models do indeed have unique optimal solutions. Therefore, these solutions are also implicitly differentiable, using the method of Wilken et al. [176].

□

Differentiating OFM usage

In order to differentiate OFM usage, we reformulate problem (2.5) of maximising the objective flux subject to enzyme constraints, as a problem of maximising the sum of the OFM usage subject to equivalent constraints:

Let $\boldsymbol{p}$ be the vector of model parameters, and $\boldsymbol{\lambda}(\boldsymbol{p}) = [\lambda_1(\boldsymbol{p}), \ldots, \lambda_K(\boldsymbol{p})]^T$ be the vector of the optimal scalar coefficients of the OFMs, as given in (2.11). Let $\boldsymbol{D}(\boldsymbol{p})$ be the matrix of cost vectors, as in the proof of Theorem 2. Now, we may rewrite (2.1) as the LP of maximising the sum of the OFM usage, subject to the enzymatic constraints being active, and non-negativity of the OFM usage:

$$\begin{aligned} \underset{j}{\text{maximise}} \quad & \sum_{j=1}^{r} \lambda_j(\boldsymbol{p}) \\ \text{subject to} \quad & \boldsymbol{D}(\boldsymbol{p})\boldsymbol{\lambda}(\boldsymbol{p}) = \mathbf{1} \\ & \lambda_j(\boldsymbol{p}) \geq 0 \ \ \forall j \end{aligned} \tag{2.16}$$

Since the fluxes in a pruned model are necessarily all positive, we do not in fact need the inequality constraint on $\boldsymbol{\lambda}$, and can thus use Lagrange multipliers to calculate the optimum.

We define the objective function f and the constraints function $\boldsymbol{g}$ as

$$\begin{aligned} f(\boldsymbol{\lambda}) &:= \sum_{j=1}^{r} \lambda_j(\boldsymbol{p}) \\ \boldsymbol{g}(\boldsymbol{\lambda}) &:= \boldsymbol{D}(\boldsymbol{p})\boldsymbol{\lambda}(\boldsymbol{p}) - \mathbf{1} \end{aligned} \tag{2.17}$$

Write the Lagrangian $\mathcal{L}$, where $\boldsymbol{\nu}$ is the Lagrange multiplier:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\lambda}, \boldsymbol{\nu}) &= f(\boldsymbol{\lambda}) + \boldsymbol{g}(\boldsymbol{\lambda})^T \boldsymbol{\nu} \\ &= \sum_{j=1}^{r} \lambda_j(\boldsymbol{p}) + \ (\boldsymbol{D}(\boldsymbol{p})\boldsymbol{\lambda}(\boldsymbol{p}) - \mathbf{1})^T \boldsymbol{\nu}(\boldsymbol{p}) \end{aligned} \tag{2.18}$$

Using the Lagrange multiplier method, we require the following two conditions for optimality of $\boldsymbol{\lambda}$:

$$\begin{aligned} \nabla_{\boldsymbol{\lambda}} \mathcal{L} &= \quad \mathbf{1} + \boldsymbol{D}(\boldsymbol{p})^T \boldsymbol{\nu} = \mathbf{0} \\ \nabla_{\boldsymbol{\nu}} \mathcal{L} &= \boldsymbol{D}(\boldsymbol{p})\boldsymbol{\lambda}(\boldsymbol{p}) - \mathbf{1} = \mathbf{0} \end{aligned} \tag{2.19}$$

We can incorporate the Lagrange optimality conditions (2.19) into a new function $\boldsymbol{L}$, which is constructed to obey the same conditions:

$$\boldsymbol{L}(\boldsymbol{\lambda}(\boldsymbol{p}), \boldsymbol{p}, \boldsymbol{\nu}) = \begin{bmatrix} \mathbf{1} + \boldsymbol{D}(\boldsymbol{p})^T \boldsymbol{\nu} \\ \boldsymbol{D}(\boldsymbol{p})\boldsymbol{\lambda}(\boldsymbol{p}) - \mathbf{1} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix} \tag{2.20}$$

We seek to find the derivatives of the optimal $\boldsymbol{\lambda}$ with respect to $\boldsymbol{p}$, so we define $\boldsymbol{z} = (\boldsymbol{\lambda}, \boldsymbol{\nu})$ and implicitly differentiate $\boldsymbol{L}$ at the optimum $\boldsymbol{z}^*$ to yield:

$$\begin{aligned} \frac{\partial \boldsymbol{L}}{\partial \boldsymbol{z}^*} \frac{d\boldsymbol{z}^*}{d\boldsymbol{p}} + \frac{\partial \boldsymbol{L}}{\partial \boldsymbol{p}} &= \mathbf{0} \\ \frac{d\boldsymbol{z}^*}{d\boldsymbol{p}} &= -\left(\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{z}^*}\right)^{-1} \frac{\partial \boldsymbol{L}}{\partial \boldsymbol{p}} \end{aligned} \tag{2.21}$$

We can calculate the value of $\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{z}^*}$, and we can use automatic or symbolic differentiation to efficiently calculate $\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{p}}$ [114, 137]. We may now extract the values of $\frac{d\boldsymbol{\lambda}^*}{d\boldsymbol{p}}$ from $\frac{d\boldsymbol{z}^*}{d\boldsymbol{p}}$, by taking the columns corresponding to the indices in $\boldsymbol{\lambda}$, since $\boldsymbol{z} = (\boldsymbol{\lambda}, \boldsymbol{\nu})$. The values of $\frac{d\boldsymbol{\lambda}^*}{d\boldsymbol{p}}$ are precisely the unscaled sensitivities of OFM usage to model parameters $\boldsymbol{p}$.

Metabolic control analysis

A major difference between our approach and that of classical metabolic control analysis of ODE models is that the enzyme constraints cause implicit dependencies between fluxes and turnover numbers of different reactions.

In MCA theory [82], it is assumed that reaction i is only directly affected by parameter i,

that is,

$$\frac{\partial v_i}{\partial p_i} \neq 0, \quad \frac{\partial v_i}{\partial p_j} = 0 \text{ for any } i \neq j. \tag{2.22}$$

In our model setup (2.3) this is not the case. A change in the kinetic parameter p_j affects the concentration of enzyme e_j, which in turn can indirectly affect the optimal concentration of every other enzyme in the solution, due to the total enzyme constraints. Therefore, the summation theorem from MCA for ODE models [73], which states that normalised flux control coefficients sum to 1, will not necessarily hold.

2.9.3 Non-unique optimal solutions

Our method takes an optimal solution to an enzyme-constrained genome scale metabolic model and prunes inactive reactions and enzymes to ensure that this is a unique optimal solution. Before pruning, it is in general not guaranteed that there is a unique optimal solution. Choosing which optimal solution to analyse is vital in any analysis of constraint-based models. Here we wish to discuss this in the context of sensitivity analysis.

The issue of which optimal solution sensitivities should be calculated for is a problem in any sensitivity analysis of an optimisation problem with a non-unique optimal solution. Should a parameter change cause a switch to a different optimal solution, there will be discontinuity or smoothness challenges. In any method, the quantity $\frac{\partial v}{\partial p}$ will be either undefined, or difficult to interpret (a generalised derivative), when the function $v(p)$ is discontinuous or non-smooth at p. When we prune the model, we make the assumption that the set of reactions used in an optimal solution is stable to infinitesimal parameter changes. We only allow one optimal solution, forcing $v(p)$ to be smooth and differentiable. We can therefore calculate the sensitivity of the variables in the chosen solution to the model parameters. These sensitivities are interpretable within our assumption. In frameworks without pruning, such as finite differentiation or shadow prices, the switching between optimal solutions can occur. This renders sensitivities very difficult to interpret. Investigation into switching points of metabolism is beyond the scope of this paper, but would be valuable in furthering understanding of whole cell metabolism.

In the case of classic FBA models, running flux variability analysis (FVA) often reveals highly variable optimal fluxes. This reveals that optimal solutions can vary greatly from one another. However, this does not appear to be as significant an issue with enzyme-constrained models. When running FVA on ecFBA models, we find that the variability of optimal fluxes is greatly reduced. This is due to the different costs associated with different enzymes. Using realistic k_{cat}-values and gene product molar masses causes the optimal solution space of ecFBA models to be vastly reduced compared to FBA models.

If an ecFBA model has multiple optimal solutions, deciding which optimal solution to differentiate is at the discretion of the reader. It may be of interest to take a sample of

optimal solutions. For example, by constraining one or more fluxes to their upper or lower optimal flux values and solving the new LP, one would find new optimal solutions to the original LP. These alternative optimal solutions could all be pruned and differentiated, so that the sensitivities of both OFM usage and reaction fluxes in these different optimal solutions could be compared to find some sense of an 'average' sensitivity across solutions.

Chapter 3

Enzyme-constrained genome-scale metabolic model of *Staphylococcus aureus* and its implications for drug target finding

I was responsible for establishing and curating the metabolic network, including the supervision of bachelors student Nastasia Swann who worked on gapfilling, and work students Flora Schlüter and Vincent von Häven who both helped detail the enzyme stoichiometries. The original draft was written by myself, with supervision and review from Oliver Ebenhöh, and conceptualisation from St Elmo Wilken. Finally, growth experiments were performed by Yibing Ma from the Department of Veterinary and Animal Sciences, University of Copenhagen.

The code required to reproduce all results from this chapter is available at `https://github.com/HettieC/StaphylococcusAureus/`, and as an archive from `https://doi.org/10.5281/zenodo.17116904`.

3.1 Introduction

Staphylococcus aureus is a human commensal bacteria, living in up to 30% of the population. It is an opportunistic pathogen, causing diseases such as pneumonia, bacteraemia, toxic shock syndrome and meningitis. Methicillin resistant *S. aureus* (MRSA) causes around 10 000 deaths per year in the USA alone, and is the leading cause of both hospital and community-associated infections of *S. aureus* [162, 129]. Indwelling medical devices, such as catheters, artificial heart valves, and joint prosthetics, are frequently the site of biofilm infections.

However, it is not only humans that are affected by *S. aureus*, but also livestock. Infections in pigs and cattle are common, with some pig farming regions having nearly 50% of pigs colonised by LA-MRSA, and 50-60% of farm workers [138, 153]. This can lead to animal suffering, and huge economic losses, as well as infections in the community (particularly in farm workers, veterinarians, and slaughterhouse workers). Colonisation of Dutch swine farmers with MRSA has been found to be 760 times higher than among patients admitted to hospitals in the Netherlands [174]. Direct or indirect animal contact is proposed to be the main cause for human colonisation with livestock-associated MRSA (LA-MRSA) [65, 10].

In this study, I investigate the metabolism of the most commonly found LA-MRSA sequence type, ST398 [152], through enzyme-constrained genome-scale metabolic modelling. A genome-scale metabolic model (GSMM) is a stoichiometric representation of all metabolic reactions in a cell, based on the genome of the organism of interest. Further details on enzyme-constrained metabolic models can be found in section 1.2.

Here, I provide a manually curated enzyme-constrained genome-scale metabolic model, as a computational resource for future use. I use this model to investigate the cause of overflow metabolism in *S. aureus*, as well as to employ testable predictions for potential drug targets.

3.1.1 Previous *S. aureus* models

To date, there are at least 114 published genome-scale metabolic models (GSMMs) of *S. aureus* [136]. Two of these models are for sequence type ST398, our strain of interest. These two models were automatically built with no strain-specific manual curation. The first model was built as part of a large-scale generation of GSMMs [26]. Unfortunately, the resulting model of *S. aureus* ST398 is not available to download in a usable SBML or JSON format. The second model of our strain was created as part of an automatic generation of models for 64 strains of *S. aureus* [22]. This publication presents an analysis of minimal media and predicted strain auxotrophies, however five of these auxotrophies are in disagreement with our experimental data. Generally, automatically built GSMMs

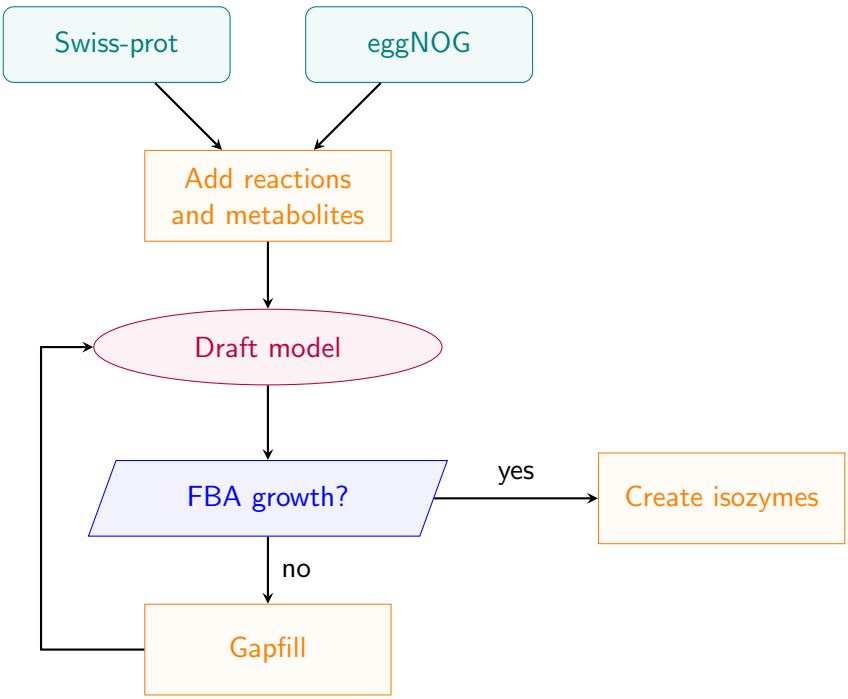


Figure 3.1: Model building workflow, with major steps of the process presented. Genome was blasted against Swiss-prot database and eggNOG mapper used to match genes to enzyme commission (EC) numbers. Reactions matching these EC numbers and their metabolites were added to a draft model. FBA was used to check for growth. If the model could not simulate growth, it was gapfilled, otherwise isozymes were created for each reaction and the initial model was considered complete.

have worse predictive capabilities than those that were manually curated [161]. Therefore, we have built a manually curated ecGSMM of *Staphylococcus aureus* ST398, to allow for strain-specific analysis and to enable further investigation into drug targets.

3.2 Model building

The initial draft of a genome-scale metabolic model of the livestock-associated methicillin-resistant *Staphylococcus aureus* (LA-MRSA) S0385, of sequence type 398 (ST398) was created following the steps presented in figure 3.1. Starting with the annotated genome with NCBI accession number AM990992.1, I performed a bidirectional blast against the Swiss-prot database [167], and used eggNOG mapper [27] to match genes to enzyme commission (EC) numbers. I then added reactions from the Rhea database [13], and their corresponding CHEBI metabolites [70], to a draft model.

Using the draft model, FBA simulations were run in order to find gaps in metabolism. Initially I checked whether the model could produce ATP, by using ATP synthase as the objective reaction. I then used the biomass reaction (see section 3.2.4), iteratively adding metabolites and checking for growth. If the model could not grow, I manually gapfilled (see section 3.2.2), and then ran this process again. When the model could produce biomass using the media described in table 3.2, I considered the model to be complete.

Reaction isozymes were then created for all reactions with associated genes. Where data was available, these were given enzyme stoichiometries according to the corresponding swiss-prot entries, otherwise enzymes were taken to be monomers.

Enzyme turnover numbers are traditionally measured through labour-intensive assays, and as such there is a lack of available data, even for model organisms such as *Escherichia coli*. Here, turnover numbers (k_{cat}-values) were predicted using the TuRNuP tool [93]. This tool creates a reaction fingerprint of the substrates and products, and uses the enzyme sequence to make a prediction for the k_{cat}-value via machine learning.

3.2.1 Exchanges and transporters

To simulate growth, the model must be able to import metabolites from the extracellular space. As such, exchange reactions were added from the media into the cell periplasm for 17 metabolites. Transporters were then added between the periplasm and the cytosol, catalysed by enzymes as far as possible. Initially, the genome annotation was used to identify transporter genes. Further, the Transporter Automated Analysis Pipelipe (transAAP) [55] was used to predict more transporter genes. This automated pipeline compares the genome to various curated databases, including the Transporter Classification Database [144], the Pfam database [111], the COG database [60], and TIGRFAMS [68].

In total, 65 transporter proteins, encoding for ABC, PTS, symport, antiport and permease reactions were identified. Proteins could not be identified for an ABC glucose transporter. Where gene-associated transporter reactions were not found for media components, a permease reaction was added.

Mixed acid fermentation products of *S. aureus* are restricted to 2,3-butanediol, ethanol, acetate, and lactate [163]. This is reflected in the model.

3.2.2 Gapfilling

Since manually curated GSMMs have better predictive capabilities, I manually gapfilled the initial draft model to ensure that it could synthesise all biomass precursors from minimal media. Initially, I focussed on the central carbon metabolism, including glycolysis, the citric acid cycle (TCA cycle), and the pentose phosphate pathway (PPP). To simulate a chemically defined minimal media, exchange reactions were added for the metabolites: water, oxygen, carbon dioxide, phosphoric acid, beta-D-glucose, ammonium, L-cysteine, L-arginine, L-proline, riboflavin, protons, acetate, lactate, ethanol.

In order to simulate growth, the network must have continuous pathways from the media metabolites to the essential biomass components. To find gaps in the pathways in the network, I created Escher maps [88]. I based these maps on the KEGG pathway maps, using the enzyme-associated reactions from the draft model. I was able to find metabolites

that needed to be connected by the addition of new, spontaneous, reactions. For example, in the PPP, the model required a reaction to convert aldehydo-D-ribose 5-phosphate into D-ribose 5-phosphate.

It is known that *S. aureus* contains a full glycolysis pathway, TCA cycle and pentose phosphate pathway (PPP) [158]. Therefore, I first gapfilled in the central carbon metabolism. The LA-MRSA strain ST398 is only missing a gene annotation for one reaction in glycolysis and the TCA cycle. Phosphoenolpyruvate carboxylase was added as a spontaneous reaction to link phosphoenolpyruvate in glycolysis to oxaloacetate in the TCA cycle. In the PPP, a spontaneous reaction for the interconversion of aldehydo-D-ribose 5-phosphate and D-ribose 5-phosphate was required. Besides these two additions, the central carbon metabolism was well annotated and did not require any further gapfilling.

Using experimental growth data on a range of media, we were able to conclude that our strain is a prototroph for all amino acids besides L-arginine and L-cysteine. Therefore, it was necessary that the model could produce all other amino acids. Here, there were a number of gaps in the network. Before gapfilling, the model was unable to synthesise L-histidine, L-methionine, L-lysine, L-serine and glycine. Respectively, the following reactions were required for synthesis of those amino acids: L-histidinol phosphate to L-histidinol; L-homocysteine to L-methionine; (S)-2,3,4,5-tetrahydrodipicolinate to (S)-2-succinylamino-6-oxoheptanedioate and (2S,6S)-2,6-diaminopimelate to meso-2,6-diaminopimelate; 3-phosphooxypyruvate to O-phospho-L-serine. These reactions were included in the model as spontaneous reactions, and the model was able to synthesise all amino acids except L-arginine and L-cysteine, in agreement with experimental observations.

In lipid metabolism, gapfilling was most notably required for the de novo biosynthesis of all fatty acids, for the reaction:

$$\text{fatty acid-[ACP]}^- + \text{H}_2\text{O} \rightarrow \text{fatty acid}^- + \text{H}^+ + \text{holo-[ACP]}^-. \tag{3.1}$$

In fatty acid degradation, spontaneous reactions were added for all fatty acids for the conversion:

$$\text{a medium-chain (2E)-enoyl-CoA} + \text{H}_2\text{O} \leftrightarrow \text{medium-chain (3S)-3-hydroxyacyl-CoA}. \tag{3.2}$$

Besides these core pathways, I gapfilled in energy, nucleotide, and vitamin metabolism. In total, 45 spontaneous metabolic reactions were added to the model, with the logic that a smaller but more carefully curated model is of more use than a large automatically curated one.

3.2.3 Respiration

S. aureus is a facultative anaerobe, and capable of growing in both oxygen-rich and oxygen-poor environments. Under aerobic conditions, it primarily uses oxidative phosphorylation for ATP production, and in the absence of oxygen it switches to fermentation pathways. The model contains reactions for both aerobic and anaerobic respiration, as well as fermentation pathways, to reflect the metabolic versatility of *S. aureus*. It is known that an impaired electron transport chain in *S. aureus* leads to biofilm formation and programmed cell lysis [106]. Thus, it was important to include a well-curated electron transport chain and respiratory pathway in the model.

In *S. aureus*, menaquinone serves as the sole quinone electron carrier, passing electrons to quinol oxidases [154]. Since *S. aureus* produces menaquinone-8 as the major quinone [37], this is the only quinone included in the model. Electrons from NADH enter the electron transport chain (ETC) and are transferred to menaquinone by NADH dehydrogenase-2 (Ndh2), which is non-proton-pumping [107, 170]. Additional electron input to the menaquinone pool is provided by succinate dehydrogenase (Sdh), L-lactate dehydrogenase (Ldh) [58], and malate quinone oxidoreductase (Mqo) [37, 62, 157]. An NAD-independent lactate dehydrogenase (Lqo) is also present, and is known to contribute to virulence [157, 141, 155].

The ETC in *S. aureus* contains three distinct cytochrome respiratory branches: cytochrome aa_3, cytochrome bo_3, and cytochrome bd [36]. Cytochrome aa_3 is unusual in that it does not translocate protons [166], which affects the overall proton motive force and ATP yield. Cytochrome bo_3 and bd branches provide alternative routes for electron transfer to oxygen, allowing adaptation to varying oxygen concentrations and supporting survival under stress conditions. The inclusion of a detailed and strain-specific respiratory network in the model enables simulation of metabolic responses to oxygen availability, prediction of growth phenotypes under different conditions, and investigation of the impact of ETC disruption on cellular physiology and pathogenicity.

3.2.4 Biomass reaction

To simulate growth with the ecGSMM, a biomass reaction was required. This reaction simulates the cell drain of biomass precursors. In an ideal world, this would be calculated based on metabolomic measurements of the dry cell biomass. However, it was not possible to find experimental data for *S. aureus* biomass measurements.

As is standard practice when organism-specific biomass composition data are not available, I have based the biomass function on a combination of previously published models. Namely, the basis of the biomass function was taken from a model of strain N315 [72]. In this model, data from a range of *S. aureus* strains were utilised to create a representative

Metabolite	mmol/gDW	Metabolite	mmol/gDW
Protein	**0.265**	**Fatty acids**	**0.0208**
L-isoleucine	0.0228	tetradecanoate	0.00346
L-threonine	0.0153	octadecanoate	0.00346
L-histidine	0.00616	decanoate	0.00346
L-cysteine	0.00166	hexadecanoate	0.00346
L-proline	0.00851	dodecanoate	0.00346
L-methionine	0.00699	octanoate	0.00346
L-phenylalanine	0.00851	**DNA**	**0.0696**
L-glutamate	0.0172	CoA	1.53e-5
L-lysine	0.0199	dGTP	0.0114
L-tyrosine	0.0103	dCTP	0.0114
L-asparagine	0.015	dTTP	0.0234
L-tryptophan	0.00197	dATP	0.0234
L-homocysteine	0.00346	**Soluble metabolites**	**4.85e-5**
L-valine	0.0178	FAD	2.32e-5
L-glutamine	0.011	NAD(+)	6.92e-6
L-serine	0.0162	acetyl-CoA	1.14e-5
glycine	0.016	NADP(+)	6.92e-6
L-leucine	0.0242	**RNA**	**1.78**
L-alanine	0.017	UTP	0.0175
L-arginine	0.00928	ATP	1.73
L-aspartate	0.0154	CTP	0.0172
		GTP	0.0172

Table 3.1: Biomass composition of the model.

average biomass composition. In order to report on the predicted growth rate and cellular yield, our biomass formula has a molecular weight of 1g/mmol.

3.2.5 Directionality

Initially, all reactions were bidirectional in order to not overly constrain the solution space of the linear programme. However, with all exchanges closed the model was able to produce ATP, which was tested by optimising for the ATP synthase reaction:

$$\mathrm{ADP}^{-3} + \mathrm{HPO}_4^{-2} + 4\mathrm{H}^+(\text{periplasm}) \rightarrow \mathrm{ATP}^{-4} + \mathrm{H_2O} + 3\mathrm{H}^+(\text{cytosol}). \tag{3.3}$$

This signified that there must have been infeasible energy-generating loops. In order to keep the model as general as possible, I aimed to fix the directions of only those reactions that were necessary. This was done by plotting the active reactions in Escher to find the ATP-generating loops. Reaction directions were checked against the MetaCyc [28] and Reactome [110] databases. Where reactions involved in infinite loops had a specified direction in either MetaCyc, Reactome or KEGG, they were fixed as such in the model. This was done iteratively until the model could no longer produce ATP with closed exchanges. Subsequently, I opened media exchange reactions, optimised for ATP synthase, and again found loops by plotting the fluxes in Escher. After fixing the necessary reaction directions, I repeated these steps with the biomass function as the objective to

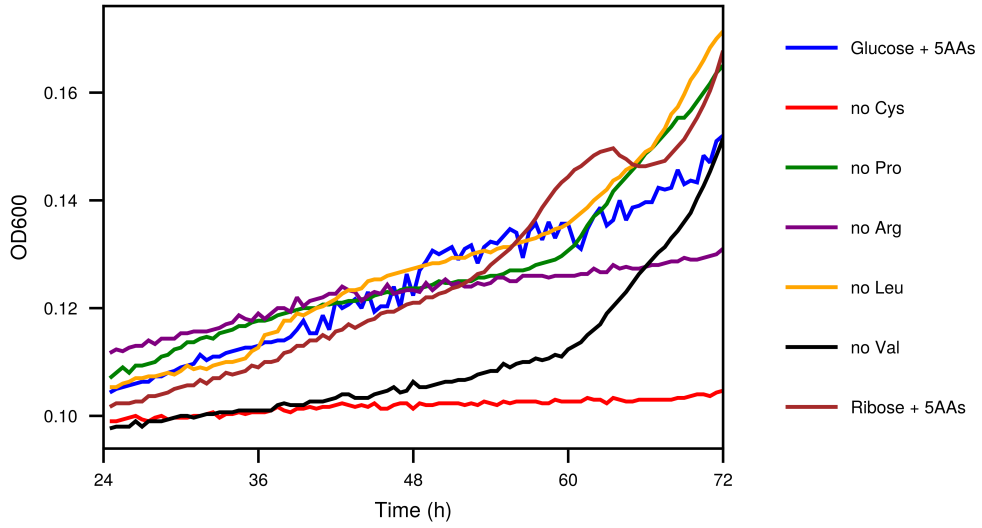


Figure 3.2: Growth curves for *S. aureus* under seven different media compositions. This data was collected by Yibing Ma from the collaborative project Innotargets. The experimental setup is as follows. The bacteria was pre-inoculated in tryptic soy broth (TSB) for 48 hour pre-cultures, followed by washing with PBS. A 96-well plate was inoculated using 100μl media per well, and pre-incubated in a 37°C shaking device for 24 hours. Finally, OD600 measurements were taken for hour 24 to hour 72 in a plate reader. The average OD600 over three technical replicates in each media is displayed. The media are described in table 3.2.

the linear programme. In total, I set 136 reactions as unidirectional. The reaction:

$$\text{succinyl-CoA}^{-5} + \text{ADP}^{-3} + \text{HPO}_4^{-2} \leftrightarrow \text{succinate}^{-2} + \text{ATP}^{-4} + \text{CoA}^{-4} \tag{3.4}$$

in the TCA cycle has no direction specified by MetaCyc, Reactome or KEGG, but led to an infinite loop when left as bidirectional. Therefore, it was constrained to the forward, ATP-releasing, direction. This unfortunately means that the TCA-cycle can also only run in the forward direction. In future, the model should be further curated to allow for the TCA cycle to run in both directions, without causing infinite loops.

3.3 Results and analyses

3.3.1 Network properties

The curated GSMM of *S. aureus* ST398 consists of 1140 total reactions, 1156 internal and 17 external metabolites. Of the total reactions, 1042 have an associated enzyme, including 22 enzyme-associated transport reactions. There are a further 15 spontaneous transport reactions, 45 spontaneous internal metabolic reactions, and 17 exchanges. Reactions are mass and charge balanced.

Ingredients	mg/L
Group 1	
$Na_2HPO_4 \cdot 2\,H_2O$	10 000
KH_2PO_4	3 000
$MgSO_4 \cdot 7\,H_2O$	500
Group 2	
Nicotinic acid	2
Calcium pantothenate	2
Pyrodoxal hydrochloride	4
Pyrodoxamin dihydrochloride	4
Riboflavin	2
Thiamin hydrochloride	2
Biotin	0.1
M9 trace	1
L-arginine	100
L-valine	750
L-leucine	750
L-proline	750
L-cysteine	250
D-glucose	5 000
D-ribose	5 000

Table 3.2: Composition of chemically-defined experimental growth media used for figure 3.2. Stock solutions were prepared for group 1, group 2, thiamin hydrochloride, biotin, M9 trace elements, D-glucose, D-ribose, and each individual amino acid. The media was prepared from these stock solutions, adjusted to pH 7.2, and sterile filtered. In figure 3.2, the first six media use D-glucose as the carbon source, Glucose + 5AAs uses D-glucose and all of these 5 amino acids, no-Cys refers to media using each amino acid in this table except L-cysteine, no-Pro uses all except L-proline, and so on. The last media, Ribose + 5AAs, uses D-ribose and all 5 amino acids.

Carbon source	Carbon yield
Acetate	0.0
N-acetyl-d-glucosamine	0.0
Alanine	0.0
L-arabinose	0.0
Pyruvate	0.1477
Glutamate	0.2031
Malate	0.2216
Fumarate	0.2216
L-lactate	0.2708
Succinate	0.3139
Glycerol	0.4924
Gluconate	0.5923
D-Ribose	0.6213
Beta-D-glucose	0.6834

Table 3.3: Carbon yield (Cmol/Cmol) of the model on different carbon sources, calculated using FBA. The highest yield is during growth on D-glucose.

3.3.2 Model validation

Experimental growth on chemically defined medium

Based on the chemically defined HHW medium [75] and the draft model network, I removed amino acids to find a minimal growth-supporting medium. Leaving only the five amino acids L-arginine, L-valine, L-leucine, L-proline and L-cysteine gave a medium showing visible growth in an incubator after 24 hours inoculation. These were then removed iteratively to find a more minimal medium, and the growth curves can be seen in figure 3.2. The media composition are detailed in table 3.2. We see that on removing L-cysteine, the strain is unable to grow, and appears to be an auxotroph for this amino acid. The model agrees with this auxotrophy, as it does not contain a full cysteine biosynthesis pathway. The medium without L-arginine also appears to have negligible growth, and the model cannot simulate growth on a medium without it. Further experiments are required to confirm whether growth in a medium with only one amino acid, L-cysteine, is experimentally possible. Unless otherwise specified, all simulations used a growth medium containing L-cysteine and L-arginine.

Carbon source utilisation

In the blood, glucose is the main source of carbon [108], and as such most simulations were run with glucose in the synthetic media. However, to investigate the possibility of growth on different carbon sources, I simulated aerobic growth on a minimal media containing one main carbon source from table 3.3, along with phosphoric acid, ammonium, L-cysteine, L-arginine, and riboflavin as the other media components.

In table 3.3, we see that the highest carbon yield is predicted to be obtained during growth on D-glucose, closely followed by D-ribose. In figure 3.2, growth on the ribose-containing

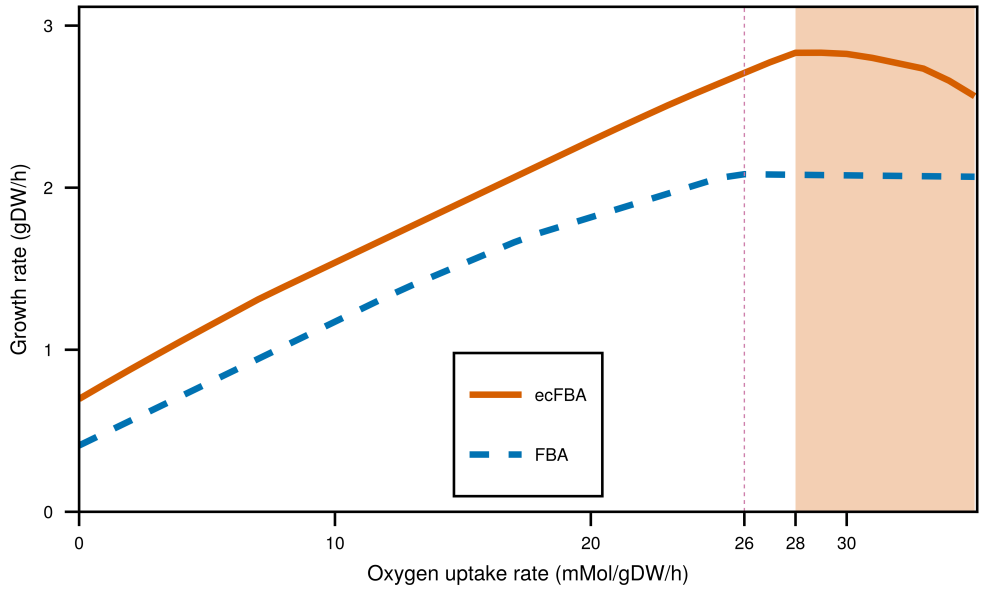


Figure 3.3: Simulated growth rate as a function of the oxygen uptake rate. In FBA simulations (dashed blue line), increasing the O2 uptake increases the growth rate until the growth remains constant at an uptake rate of 26mmol/gDW/h. In ecFBA simulations (solid orange line) the growth rate increase with the O2 uptake rate, until there is no space in the membrane for more protein, and the membrane bound becomes active, at an O2 uptake rate of 28mmol/gDW/h. After this point the growth rate decreases when the model is forced to import more oxygen than is needed.

medium does not appear to be different to growth on glucose-containing media. This provides good agreement to the model predictions. Further growth experiments on the different carbon sources are required to confirm the other predictions.

Oxygen scan

As part of the model validation, I performed an oxygen scan. Here, I iterated over an oxygen uptake rate between 0 and 35mmol/gDW/hour. The growth rates using FBA and ecFBA are then plotted, as can be seen in figure 3.3. Using FBA with no enzyme constraints (blue dashed line), we see the typical shape of oxygen scans in GSMMs: growth increases with oxygen uptake, until a second constraint starts to limit growth in conjunction with the oxygen constraint at 26mmol/gDW/h. In FBA, there are no costs associated with enzymes. Thus, one could infinitely increase the oxygen uptake rate, and the model would continue to grow at its maximum optimal rate. This is why we see that increasing oxygen uptake has no effect on growth after a rate of 26mmol/gDW/h.

Constraining the cytosolic proteins to 200mg/gDW and the membrane proteins to 120mg/gDW, and running ecFBA tells a very different story. Up to an oxygen uptake rate of 28mmol/gDW/h, growth is increasing. After this point, the membrane protein bound becomes active (shaded orange), and increasing the oxygen uptake rate does not increase the growth rate. Instead, the model is forced to expend some of its limited proteome to metabolise this excess oxygen, leading to a decrease in the growth rate. This effect was not captured running FBA without enzyme constraints, since there are no costs associated

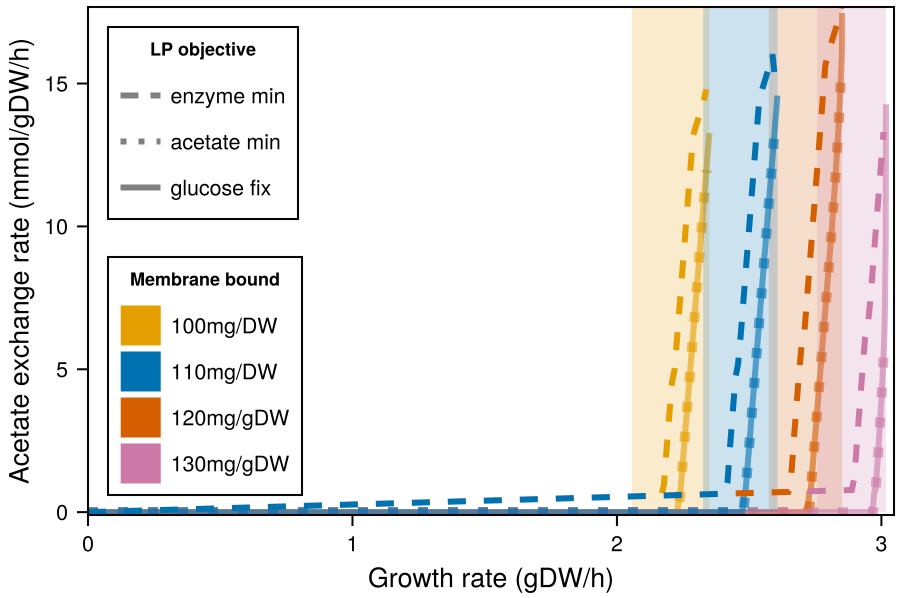


Figure 3.4: The acetate exchange rate is plotted as a function of the growth rate, showing that the metabolic switch from respiration to acetate overflow metabolism in the model is dependent on the membrane protein capacity. Three separate LP objectives and four different membrane bounds were used, giving twelve different simulations. The acetate secretion rate is plotted on the y-axis. The three line styles represent the three different objectives used. The LP objective was either total enzyme usage minimisation (dashed line), acetate secretion minimisation (dotted line), or fixing the glucose uptake rate and maximising growth (solid line). The simulations used a cytosolic enzyme bound of 200mg/gDW, and a membrane bound of 100mg/gDW (yellow), 110mg/gDW (blue), 120mg/gDW (orange), or 130mg/gDW (pink). The shaded areas show the growth rate at which the protein in the membrane reaches the corresponding membrane capacity bound. Using the three different LP objectives, there was minimal difference between when this bound becomes active (see supplementary figure 3.11), so only the simulations with fixed glucose uptake are displayed here.

to reactions. In ecFBA, the cost of metabolising the excess oxygen can be seen to have a direct impact on the maximum growth.

Overflow metabolism

Staphylococcus aureus is known to exhibit overflow metabolism under excess oxygen and glucose conditions [143]. Here, glucose is catabolised via glycolysis, resulting in two molecules of pyruvate and two molecules of ATP per molecule of glucose. The pyruvate is decarboxylated to acetyl-coenzyme A (acetyl-CoA) and carbon dioxide via pyruvate dehydrogenase. Next, the acetyl-CoA is either oxidised by the TCA cycle, is hydrolysed by the phosphotransacertylase-acetate kinase (Pta-AckA) pathway, or a combination of the two. This simultaneous utilisation of the TCA cycle and Pta-AckA pathway is what we refer to as overflow metabolism. The resulting overall metabolic strategy is a mixture of respiration and fermentation, and results in the secretion of acetate.

In figure 3.4, overflow metabolism in *S. aureus* is seen to follow the same qualitative behaviour as in *E. coli*: at low growth rates, respiration predominates, but beyond a

threshold, acetate secretion increases sharply and approximately linearly with growth rate [14]. When the model is optimised for minimisation of total enzyme usage (dashed line), a low baseline level of acetate secretion is observed across all growth rates. For simplicity, however, the onset of overflow metabolism is defined here as the point at which acetate secretion begins its sharp increase rather than the low-level secretion observed at lower growth rates.

The shaded regions in the figure indicate where the simulations encounter membrane protein capacity constraints. Interestingly, the transition to overflow metabolism does not coincide with the point at which the membrane bound becomes active. Instead, overflow consistently emerges only at higher growth rates, after the membrane constraint has already been filled. This lag between the onset of the constraint and the switch to overflow metabolism remains unexplained. The cytosolic protein constraint is not responsible for this behaviour, since this only becomes active at the maximum growth rate (not shown in figure). In contrast, figure 3.3 illustrates a scenario where growth rate is immediately affected when the membrane bound becomes active, highlighting that the effect of constraints on metabolic behaviour depends strongly on the specific metabolic context.

Minimising acetate secretion, or fixing glucose uptake and maximising growth rate give almost identical results, as can be seen by the dotted line overlapping with the solid line in figure 3.4. Minimising enzyme usage gives a slightly different result to these two LP objectives, as can be seen in 3.4. The main difference is that there is no acetate leakage until the onset of overflow metabolism when minimising for acetate or fixing glucose. Further, for all membrane bounds shown, acetate secretion starts at a slightly lower growth rate in the enzyme minimisation simulations.

In the enzyme minimisation simulations, the membrane constraint becomes active at a higher growth rate compared to both the acetate secretion minimisation and fixed glucose uptake simulations (see supplementary figure 3.11). Importantly, minimising acetate secretion does not coincide with minimal enzyme usage. Under enzyme minimisation, the cytosolic constraint is only reached at the maximum achievable growth rate across all membrane bounds tested. By contrast, in the acetate minimisation simulations the cytosolic bound becomes active much earlier (not displayed), though still at growth rates between the initial activation of the membrane constraint and the onset of overflow acetate secretion.

A posited reason for the switch to overflow metabolism in *E. coli* is the lack of proteome space in the membrane [14]. We suggest that it is reasonable for this hypothesis to also hold true in *S. aureus*, and the results in figure 3.4 support this hypothesis.

Amino Acids			
L-leucine	L-arginine	L-valine	L-proline
L-cysteine	Glycine	L-serine	L-aspartate
L-alanine	L-glutamine	L-glutamate	L-lysine
L-isoleucine	Glycine	L-threonine	L-phenylalanine
L-tyrosine	L-tryptophan	L-histidine	L-homocysteine
L-asparagine	L-methionine		
Vitamins and Cofactors			
Biotin	Pantothenate	Lipoate	Nicotinate
4-aminobenzoate	Pyridoxamine	Pyridoxine	Thiamine
Nucleotides and Derivatives			
Adenine	Guanine	Uracil	Inosine
Orotate	Thymidine	Xanthine	Cytosine
Thymine			

Table 3.4: Composition of rich medium for the simulations

3.3.3 Gene essentiality

The aim of this chapter was to find novel drug targets in the LA-MRSA strain ST398. To this end, an analysis of essential genes was performed. Here, a gene is deemed essential if upon its removal the model can no longer simulate growth. It is important to note that gene essentiality predictions are context-dependent and can vary with the composition of the biomass reaction, and the completeness of the metabolic model. In this study, essentiality was initially assessed under minimal medium conditions, reflecting a nutrient-limited environment, potentially relevant to infection scenarios.

Simulating on the minimal growth medium, the model has 85 essential genes, details of which can be seen in the supplementary table 3.8. The predicted essential genes were compared to both published results [35] from Transposon Directed Insertion site Sequencing (TraDIS) and unpublished experimental results on the same library, using a refined protocol [178]. TraDIS is a high-throughput assay for the identification of bacterial essential genes [94]. The technique involves making insertions into each gene in the genome, disrupting or altering gene function. In the resulting transposon mutant library, genes with zero or very few insertions are deemed essential, and those with multiple insertions are deemed not essential. The justification for essential gene categorisation is that if an insertion can be made that disrupts the function of a gene, and the mutant survives, this gene must not have been essential.

Combining the two experimental datasets results in a total of 75 essential genes. Of the predicted essential genes on minimal media, only 4 matched those identified by the previously published experimental study [35]. Including the unpublished results on the same library allowed for the identification of a further 7 matching essential genes. The library used in the TraDIS analysis was cultured on brain heart infusion (BHI) broth, a

very rich undefined medium with glucose as the main carbon source, supplemented with 18 amino acids. Experimentally, as well as in-silico, gene essentiality is heavily dependent on growth media [52]. To better reflect the BHI medium in the transposon study, a richer synthetic medium was used. This medium included all 20 amino acids, all nucleotide bases, and an increased number of vitamins, and is detailed in table 3.4.

Gene ID	Name	Rhea	Reaction name
SAPIG0548	tmk	13520	ATP:dTMP phosphotransferase
		30658	ATP:dUMP phosphotransferase
SAPIG0566	prs	15612	ATP:D-ribose-5-phosphate diphosphotransferase
SAPIG1121	coaD	19804	ATP:pantetheine-4'-phosphate adenylyltransferase
SAPIG1207	gmk	12697	guanylate kinase
		20783	ATP:GMP phosphotransferase
SAPIG1209	coaBC	19400	(R)-4'-phosphopantothenate:L-cysteine ligase
		16796	N-[(R)-4'-phosphopantothenoyl]-L-cysteine carboxy-lyase
SAPIG1260	pyrH	24403	ATP:UMP phosphotransferase
SAPIG1274	ribF	14360	ATP:riboflavin 5'-phosphotransferase
		17240	ATP:FMN adenylyltransferase
SAPIG2141	fabZ	13097	3-hydroxyglutaryl-[acp]-methyl-ester hydro-lyase
		41811	(3R)-3-hydroxyhexanoyl-[acyl-carrier-protein] hydro-lyase
		39159	(3R)-3-hydroxypalmitoyl-[acyl-carrier-protein] hydro-lyase
		41879	(3R)-3-hydroxydodecanoyl-[acyl-carrier-protein] hydro-lyase
		41831	(3R)-3-hydroxyhexanoyl-[acyl-carrier-protein] hydro-lyase
		41847	(3R)-3-hydroxyoctanoyl-[acyl-carrier-protein] hydro-lyase
		41895	(3R)-3-hydroxytetradecanoyl-[acyl-carrier-protein] hydro-lyase
		41863	(3R)-3-hydroxydecanoyl-[acyl-carrier-protein] hydro-lyase
		41908	(3R)-3-hydroxypalmitoyl-[acyl-carrier-protein] hydro-lyase

Table 3.5: List of essential genes and their associated reactions. These 8 genes are deemed essential in the model in rich medium and at least one of the TraDIS datasets [35, 178].

Simulating gene essentiality on the rich medium reduced the number of predicted essential genes from 85 to 17. Of these, 8 genes were also identified as essential in at least one of the experimental TraDIS datasets (see table 3.5). This improved agreement highlights the importance of accurately representing experimental conditions in silico when validating model predictions. The remaining discrepancies may be attributed to differences in biomass composition, incomplete gene annotations, or limitations in the experimental approach, such as false positives or negatives inherent to transposon mutagenesis studies [178]. Overall, the model demonstrates reasonable predictive power for gene essentiality when simulated under conditions that closely match those used experimentally.

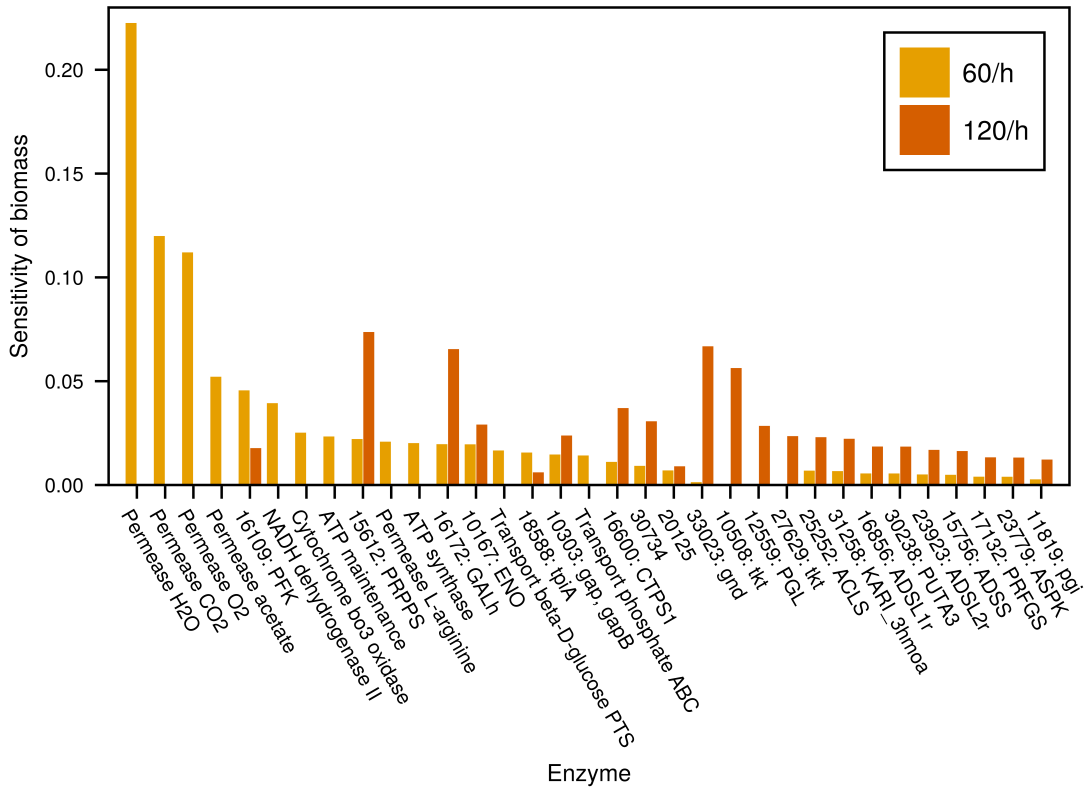


Figure 3.5: Sensitivity of the growth rate to changes in enzyme turnover numbers, when the permease k_{cat}-values are 60/h (yellow) or 120/h (orange). The 20 most controlling enzymes in both cases are plotted. Sensitivity is defined as $\frac{p}{v_r}\frac{\partial v_r}{\partial p}$, where v_r is the biomass reaction flux, and p are the enzyme turnover numbers.

The resulting list of essential genes represents promising candidates for antimicrobial drug targeting, as their inhibition is predicted to halt bacterial growth. However, further validation against experimental gene essentiality in defined medium is required.

3.3.4 Growth sensitivity analysis

Gene knockout simulations are useful in finding novel drug targets when a drug is able to fully inhibit an enzyme. However, this analysis will not capture potential targets for partially inhibiting anti-microbial compounds. Sensitivity analysis of the growth rate can elucidate these targets enzymes.

In the model, all reactions with a non-trivial associated gene have a turnover number that was predicted using TuRNuP [93]. Associated genes were not found for the permease reactions, and they were all given the same catalysing enzyme and a turnover number of 60/h. Performing a growth sensitivity analysis on the *S. aureus* model, we see in figure 3.5 that doubling the turnover number of the permease reactions has a notable effect on the growth sensitivities. When the permease turnover numbers are lower, the four most controlling enzymes catalyse permease reactions. However, when the permease turnover numbers are doubled to 120/h, displayed in orange, no transport reactions are among the

20 most controlling enzymes. In this case, the model is limited by other reactions, and the permease enzymes are no longer growth-limiting.

Besides the transport reactions, the most controlling enzymes when the permease k_{cat}-values are 60/h includes four reactions involved in glycolysis - 16109: PFK (Phosphofructokinase), 10167: ENO (enolase), 18588: tpiA (D-glyceraldehyde-3-phosphate aldose-ketose-isomerase), 10303: gap (D-glyceraldehyde-3-phosphate:NAD+ oxidoreductase (phosphorylating)). Further highly controlling reactions are NADH dehydrogenase, cytochrome bo3 oxidase, 15612: PRPPS (ATP:D-ribose-5-phosphate diphosphotransferase), and ATP synthase. Three of these are part of oxidative phosphorylation, and the fourth is involved in the pentose phosphate pathway.

In the case of higher permease k_{cat}-values, there are no transport reactions among the 20 most controlling enzymes. Instead, five of the most controlling enzymes are involved in the pentose phosphate pathway - 15612: PRPPS, 33023: gnd (6-phospho-D-gluconate:NAD+ 2-oxidoreductase (decarboxylating)), 12559: PGL (6-phospho-D-glucono-1,5-lactone lactonohydrolase), 10508 and 27629: tkt transketolase. This highlights the importance of accurate parameterisation of ecGSMMs.

3.3.5 Whole solution sensitivity analysis

A sensitivity analysis of the active optimal solution was performed, under minimal media aerobic conditions, with D-glucose as the carbon source. This approach quantifies how changes in enzyme turnover numbers (k_{cat}) affect fluxes through all reactions in the active optimal solution, rather than focusing solely on growth rate. In figure 3.6, it can be seen, similarly to the growth sensitivity in figure 3.5, that permease reactions are highly controlling. The only differences between the top 15 growth-controlling enzymes and the top 15 overall most controlling enzymes are that the PTS-transport for D-glucose is only among the highest growth-controlling enzymes, and has a lesser effect on the overall solution. It can also be seen in figure 3.6 that 80% (200 out of 247 active reactions) of the fluxes behave in the same way (the lower 80% of the y-axis). A decrease in the turnover number of any enzyme would reduce the flux through these reactions by the same proportion. In section 3.3.7, I perform an analysis on the optimal flux modes of the system. These 200 reactions are in fact shared by the two active OFMs, and used in the same proportion in each OFM. The remaining 47 active reactions (top 20% of y-axis) are affected differently by perturbations in turnover numbers. This means that upon an the inhibition of an enzyme, the model predicts that the flux through some of these 47 reactions will decrease, and through others will increase.

In aerobic growth conditions, respiration is key for fast growth. Concentrating on the active respiratory reactions, we see in figure 3.7 that the fluxes through succinate dehydrogenase (Sdh) and Malate:menaquinone oxidoreductase (Mqo) are the most sensitive

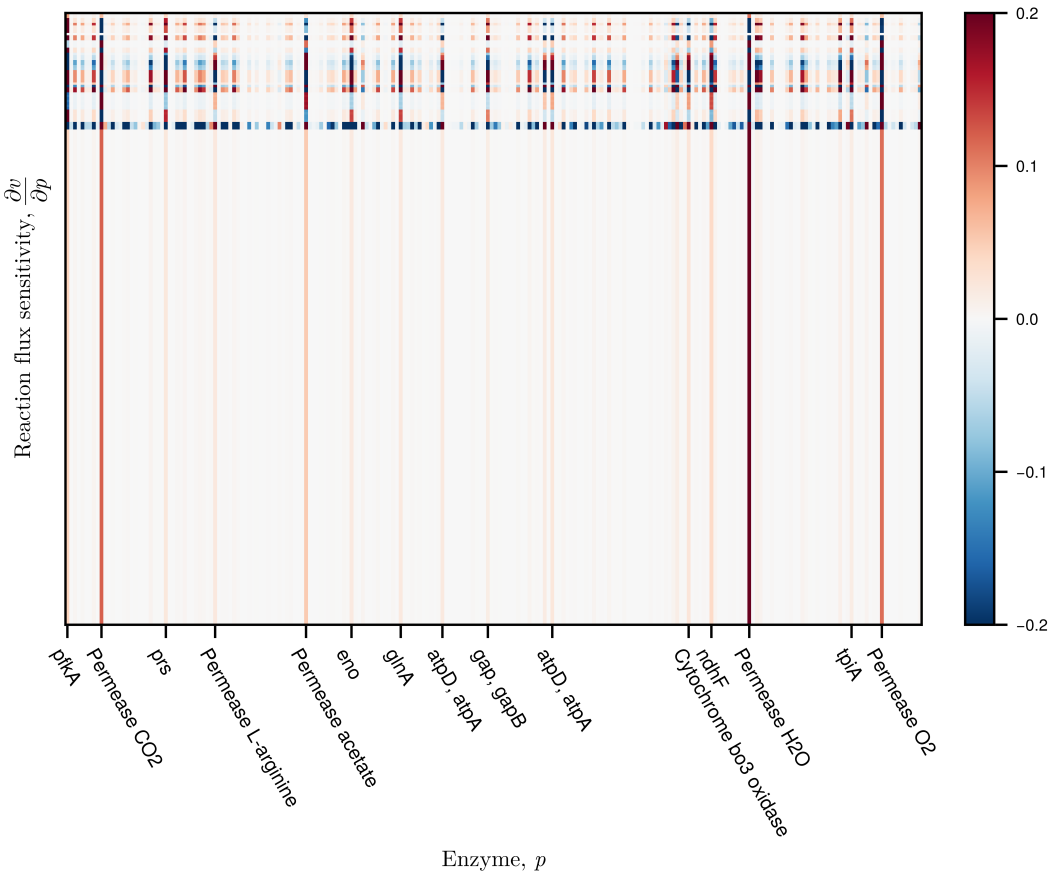


Figure 3.6: Sensitivities of reaction fluxes (y-axis) to changes in enzyme turnover numbers (x-axis), with the 15 enzymes exerting the highest average absolute control labelled. Sensitivity is defined as $\frac{p}{v_i}\frac{\partial v_i}{\partial p}$, where v_i is the flux through reaction i, and p are the enzyme turnover numbers.

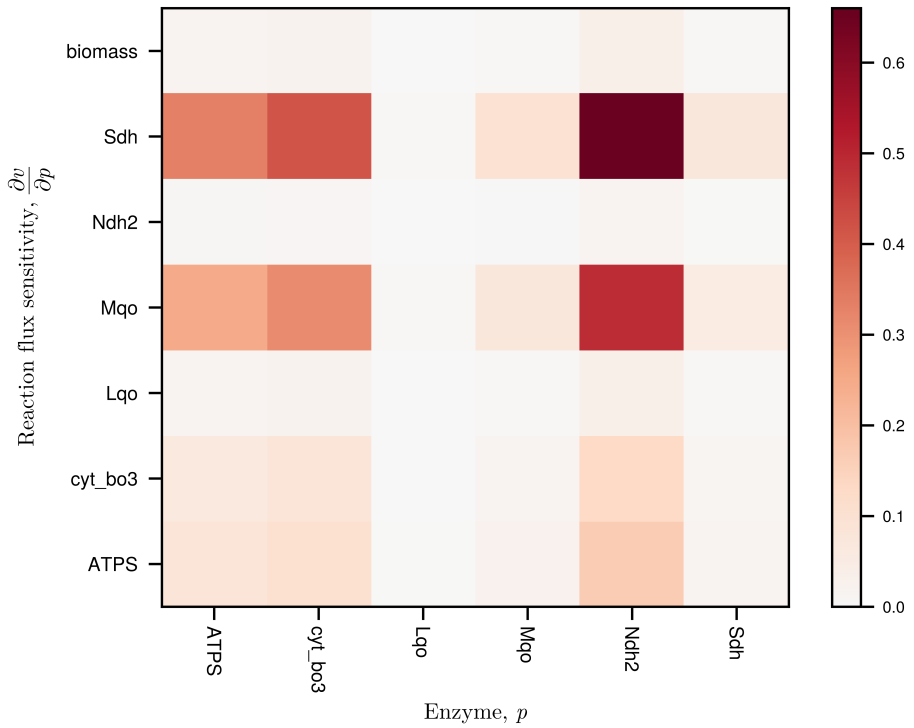


Figure 3.7: Sensitivity of flux through active repiratory reactions (y-axis) to changes in enzyme turnover numbers (x-axis), during growth on D-glucose. Sensitivity is defined as previously.

to changes in turnover numbers of other respiratory enzymes. Biomass reaction flux, ie the growth rate, and NAD-independent lactate dehydrogenase are the least sensitive to changes in respiratory parameters.

3.3.6 Growth on L-lactate

Once a patient is infected, *Staphylococcus aureus* can be incredibly difficult to treat, partly due to its ability to subvert innate host immunity. One important aspect of the innate immune response is the production of nitric oxide (NO·) radicals, which act as antimicrobial agents. In growth medium, *S. aureus* is resistant to high concentrations of NO· [140], a property that contributes to its persistence during infection.

NO· stress in *S. aureus* leads to changes in the metabolism, caused by regulatory effects [67]. Since a GSMM does not capture gene regulation, the resulting metabolic effects are studied in isolation. It is known that upon exposure to host-derived NO·, *S. aureus* undergoes a metabolic shift characterised by increased secretion of L-lactate. This L-lactate is subsequently reassimilated and metabolised via the NAD-independent lactate dehydrogenase, Lqo [58]. Utilisation of L-lactate as a carbon source during NO· stress is essential for survival and virulence.

The model cannot fully capture the regulatory and physiological effects of NO· stress, such as the damage to respiratory enzymes [139], but the usage of L-lactate as the primary carbon source can be investigated. To reflect the increased proteome investment in

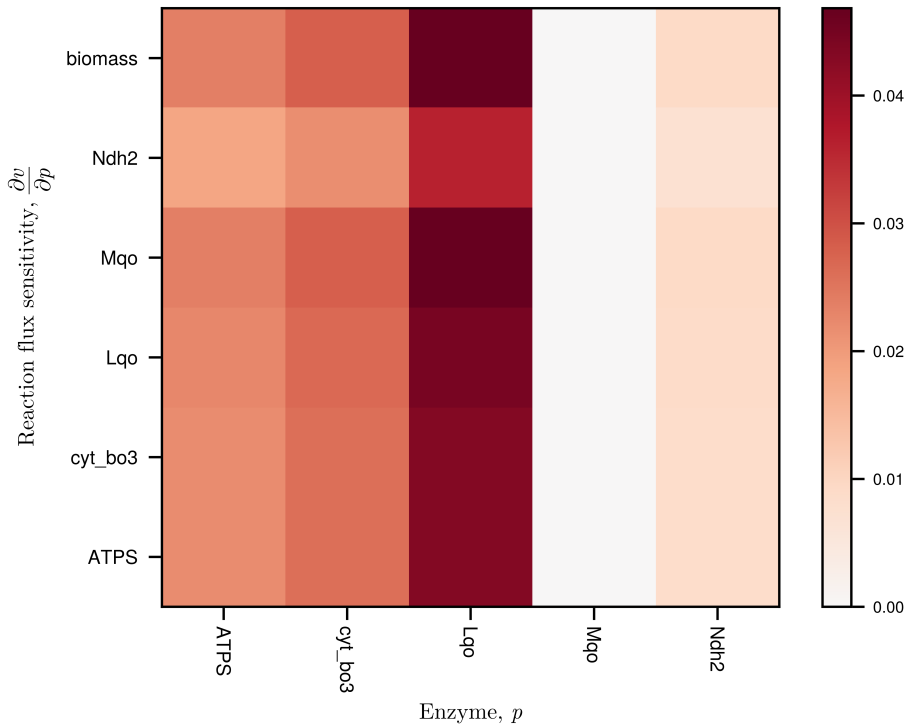


Figure 3.8: Sensitivities of fluxes through the respiratory and biomass reactions (y-axis) to changes in enzyme turnover numbers (x-axis), when grown on L-lactate as the carbon source. NAD-independent lactate dehydrogenase, Lqo, exerts the most control on respiratory enzyme.

NO· detoxification and the associated reduction in available metabolic capacity, I reduced the enzyme constraints to 120mg/gDW in both the cytosol and the membrane. This adjustment is to simulate the proteome reallocation that occurs during stress responses, where resources are diverted from growth-related processes to defense mechanisms [140].

I performed a sensitivity analysis of the whole model solution, grown on L-lactate. The results, shown in figure 3.8, demonstrate that Lqo exerts the greatest control over the respiratory fluxes and the growth rate under these conditions. This finding is consistent with experimental observations that Lqo is essential for survival during NO· stress, as it enables the utilisation of L-lactate when canonical NAD-dependent pathways are compromised [58].

3.3.7 Active optimal flux mode analysis

In an optimal solution to an ecGSMM, the active optimal flux modes provide the building blocks of metabolism [31]. The active solution of an ecGSMM is highly affected by the enzyme constraints used in the optimisation problem. I simulated growth on minimal media, using capacity constraints of 200mg/gDW and 130mg/gDW for the cytosol and membrane, respectively. Since there are two constraints, the optimal solution uses two OFMs. These OFMs have both been scaled to produce one unit of flux through the biomass reactions.

Exchange metabolite	OFM 1	OFM 2
riboflavin	0.000232	0.000232
L-cysteine	0.121	0.121
L-arginine	1.92	1.92
phosphate	3.65	3.65
beta-D-glucose	6.28	5.38
O2	9.98	11.4
H(+)	-5.35	-1.93
acetate	-5.61	-2.2
CO2	-10.7	-12.1
H2O	-20.2	-21.5

Table 3.6: Simulated exchange rates (mmol/gDW/h) of metabolites in the two OFMs. A negative rate signifies that the metabolite is secreted, and a positive rate is an uptake. Riboflavin, L-cysteine, L-arginine and phosphate are imported at equivalent rates by both OFMs, the remaining metabolites (highlighted in grey) differ.

The fluxes of the active exchange reactions for the two OFMs are presented in Table 3.6. Riboflavin, L-cysteine, L-arginine, and phosphate are imported at identical rates in both OFMs. The strain is auxotrophic for L-cysteine and L-arginine, which are therefore directly imported and incorporated into the biomass reaction. Phosphate and riboflavin are likewise required in the same ratio across both OFMs. The remaining four exchange metabolites differ between the two modes. Notably, acetate secretion is 250% higher and O2 uptake is 14% lower in the first OFM than in the second. While both modes employ a combination of respiration and fermentation, for clarity the first will be referred to as the fermentative OFM and the second as the respiratory OFM.

Table 3.7 presents the reaction rates with the greatest difference between the two OFMs. The fermentative OFM (OFM 1) lacks flux through some TCA cycle reactions (isocitrate oxidoreductase, citrate hydroxymutase, and citrate synthase), while the respiratory OFM (OFM 2) shows a more active TCA cycle. The fermentative OFM exhibits higher acetate production (seen in table 3.6), with a higher flux through acetyl-CoA:phosphate acetyltransferase and ATP:acetate phosphotransferase. This is consistent with the first OFM being referred to as the fermentative, and the second with higher flux through the TCA cycle as the respiratory. The two reactions that are only active in the fermentative OFM (ATP:deoxycitidine 5'-phosphotransferase, deoxycytidine kinase) are both involved in pyrimidine metabolism. Both OFMs share flux through central metabolism, but OFM 1 exhibits higher acetate secretion (via acetyl-CoA:phosphate acetyltransferase and ATP:acetate phosphotransferase), indicating a more fermentative phenotype, whereas OFM 2 is more respiratory with higher TCA cycle activity. These results suggest that *S. aureus* can flexibly switch between a fermentative and a respiratory metabolic strategy.

The increased flux through pyrimidine metabolism in the fermentative OFM (OFM 1) likely reflects a greater demand for nucleotide turnover or salvage under fermentative conditions. During fermentation, the ATP yield per glucose is lower, and the cells are running

Reaction	Name	OFM 1	OFM 2
19632	isocitrate:NADP+ oxidoreductase (decarboxylating)	0.0	1.6
10339	citrate hydroxymutase	0.0	1.6
16848	citrate synthase	0.0	1.6
46043	ATP:deoxycitidine 5'-phosphotransferase	0.907	0.0
46036	deoxycytidine kinase	0.907	0.0
17664	succinate:CoA ligase (ADP-forming)	1.16	2.76
56948	UTP:pyruvate 2-O-phosphotransferase	1.37	0.461
Sdh	Succinate dehydrogenase	1.46	3.06
17297	2-oxoglutarate:ferredoxin oxidoreductase (decarboxylating)	1.46	3.06
Mqo	Malate:quinone oxidoreductase	2.12	3.72
12463	(S)-malate hydro-lyase (fumarate-forming)	2.28	3.88
14108	L-proline:NAD+ 5-oxidoreductase	3.91	5.3
14112	L-proline:NADP+ 5-oxidoreductase	3.99	5.38
19524	acetyl-CoA:phosphate acetyltransferase	5.61	2.2
11355	ATP:acetate phosphotransferase	5.61	2.2
12768	pyruvate:ferredoxin 2-oxidoreductase (CoA-acetylating)	7.21	5.39

Table 3.7: Reaction rates (mmol/gDW/h) in the two OFMs with the greatest difference. The reactions that are only used by one of the OFMs are highlighted in grey.

ATP:deoxycitidine 5'-phosphotransferase in reverse to release more ATP. In contrast, in the respiratory OFM (OFM 2), higher ATP yields from the TCA cycle and oxidative phosphorylation may reduce the need for such activity, shifting metabolic resources toward biosynthesis. This suggests that *S. aureus* adapts its nucleotide metabolism according to the prevailing metabolic strategy, with increased pyrimidine metabolism supporting growth during fermentation.

The optimal solution is composed of a unique optimal combination of the two OFMs, for which the theory was described in chapter 2. Here, the optimal solution is made up of 2.3 units of the fermentative and 0.72 units of the respiratory OFM, corresponding to a ratio of roughly 3 : 1. This optimal ratio of the two OFMs is dependent on the model parameters, in particular the enzyme turnover numbers. If one were to partially inhibit or speed up an active enzyme, this would affect the optimal ratio of OFM usage. We may write that $\lambda_1 \boldsymbol{OFM_1} + \lambda_2 \boldsymbol{OFM_2} = \boldsymbol{v}$, where $\boldsymbol{v}$ is the whole optimal solution of fluxes and $\lambda_1 = 2.3, \lambda_2 = 0.72$ give the proportion of the OFMs required for the optimal solution. Therefore, since OFMs are scaled for one unit through the objective reaction, we can write $\lambda_1 + \lambda_2 = v_r$, where v_r is the objective reaction. Sensitivity of OFMs is therefore the sensitivity of these λ, corresponding to the optimal usage of each OFM.

Investigating the OFM sensitivity, I found that for every enzyme one OFM has a positive and the other a negative sensitivity. That is, increasing the turnover number of an enzyme will always increase the optimal usage of one OFM and decrease the optimal usage of the other. We can see this in figure 3.9, where the sensitivities of the usage of the two OFMs are always opposite in sign. The left hand side of the figure are negative sensitivities, and

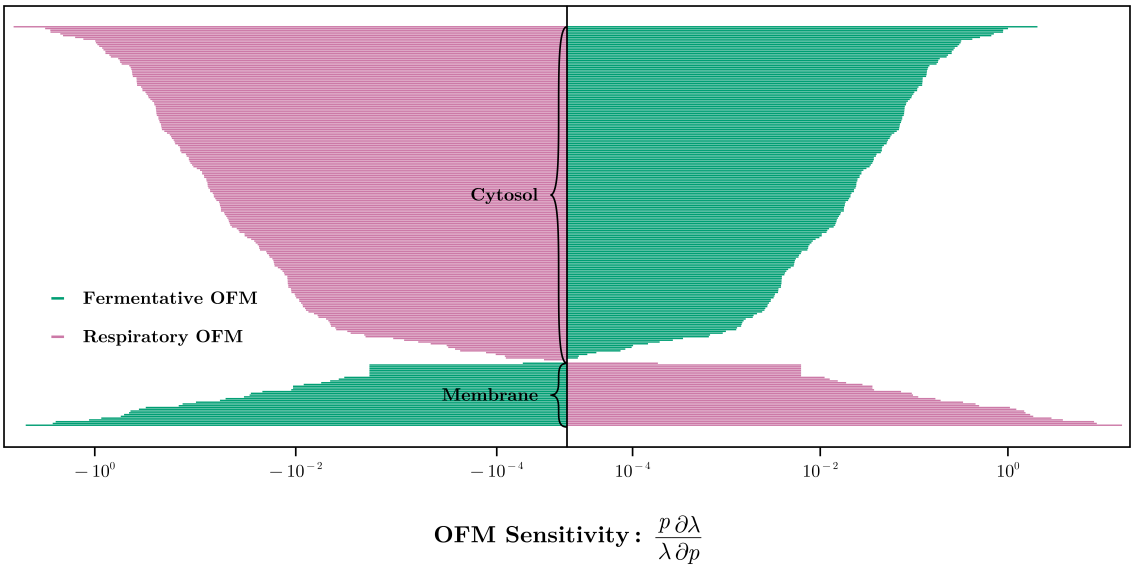


Figure 3.9: Sensitivities of the OFM usage (x-axis) to enzyme turnover numbers (y-axis). Green shows the sensitivity of the fermentative OFM (OFM 1) while pink shows the sensitivity of the respiratory OFM (OFM 2). The sensitivities are sorted in the y-axis by size, and are defined by $\frac{p}{\lambda}\frac{\partial \lambda}{\partial p}$, where $\lambda_1 \boldsymbol{OFM_1} + \lambda_2 \boldsymbol{OFM_2} = \boldsymbol{v}$, where $\boldsymbol{v}$ is the optimal solution.

the right hand side are positive. The fermentative OFM (green) has a positive sensitivity to enzymes in the cytosol, whereas the respiratory OFM (pink) has a positive sensitivity to enzymes in the membrane.

For clarity, the enzyme names are not displayed in figure 3.9. The five enzymes that are the most controlling in the membrane, corresponding to the bottom five lines on the y-axis, are permease enzymes for H2O, CO2, O2, and acetate, as well as NADH dehydrogenase, from most to least controlling. In the cytosol, the five most controlling enzymes, corresponding to the top five enzymes on the y-axis, are 6-phosphofructokinase (pfkA), ribose-phosphate diphosphokinase (prs), glutamine synthase (glnA), enolase (eno), and triose phosphate isomerase (tpiA). These enzymes belong to glycolysis (pfkA, eno, tpiA), the pentose phosphate pathway (prs) and amino acid biosynthesis (glnA).

Just below the top of the membrane enzymes in figure 3.9, there are 7 enzymes exerting the same control on OFM usage. These are highlighted in red in figure 3.10, and all catalyse the conversion of a fatty acid-[ACP] into the fatty acid, for example:

$$\text{octanoyl-[ACP]}^{-1} + \text{H}_2\text{O} \rightarrow \text{octanoate}^{-1} + \text{H}^{+} + \text{holo-[ACP]}^{-1}. \tag{3.5}$$

The fatty acids are required by the biomass reaction in the same proportions (see table 3.1), and these reactions were gapfilled, as described in section 3.2.2. Therefore, the enzymes are all identical, and the sensitivity of the OFM usage to each of these enzymes is also identical.

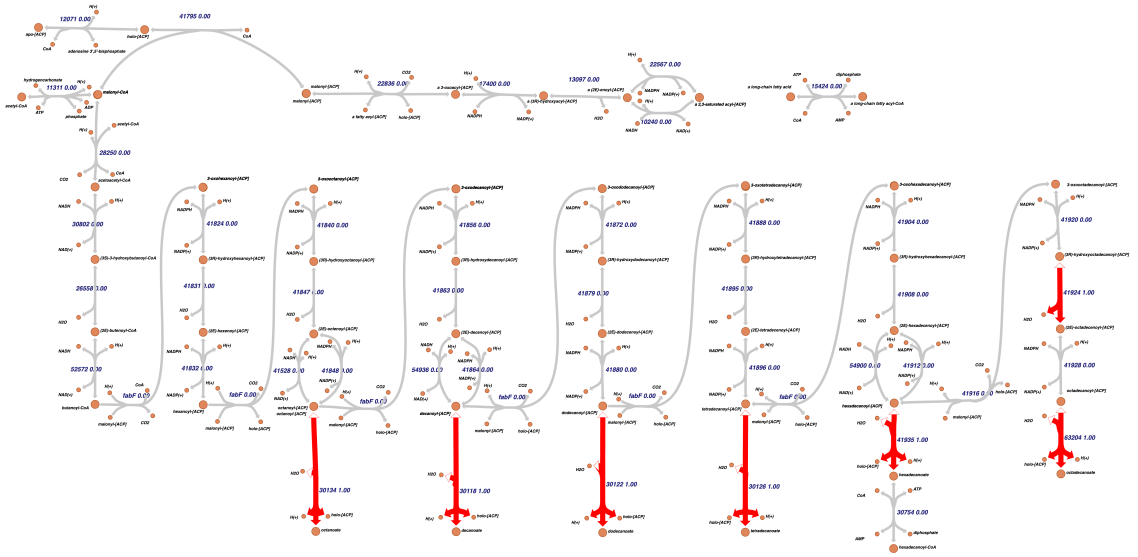


Figure 3.10: Escher map of fatty acid biosynthesis. Highlighted in red are seven reactions to which the OFM sensitivity is the same, meaning $\frac{p_i}{\lambda}\frac{\partial\lambda}{\partial p_i} = \frac{p_j}{\lambda}\frac{\partial\lambda}{\partial p_j}$ for the i, j among these seven reactions. A higher resolution image is available in the supplementary figure 3.12.

3.4 Discussion

The current chapter has detailed the construction, curation, and validation of the first enzyme-constrained genome-scale metabolic model (ecGSMM) for the livestock-associated methicillin-resistant *Staphylococcus aureus* (LA-MRSA) ST398. In this section, I discuss the implications of the modelling results, the strengths and limitations of the approach, and the broader significance of the findings for understanding *S. aureus* metabolism and identifying potential antimicrobial targets. I also reflect on the predictive capabilities of the model, its agreement with experimental data, and avenues for future research and model improvement.

I presented simulations for the onset of acetate overflow in figure 3.4. FBA models with the biomass reaction as the objective function tend to predict metabolic phenotypes reasonably well [76, 96, 69]. Adding enzyme constraints, and optimising again for biomass production, improves on the predictive capabilities of a model [145, 177]. However, in order to simulate overflow metabolism at different growth rates, the biomass reaction needed to run at a fixed rate, and the ecFBA problem therefore required a different objective.

The first approach in modelling overflow metabolism was to fix the biomass and minimise the enzyme usage. Cost-benefit theory suggests that cells evolve in a constant environment towards minimal enzymes expression [47], and indeed maximising biomass production whilst minimising enzyme usage does fit somewhat well with measure flux distributions [148]. Minimising enzyme usage predicted that the membrane bound becomes active, and then the model rearranges metabolically before starting overflow metabolism. In a cell, it is not possible to instantaneously change the composition of the proteome, so this seems to be a reasonable prediction. However, the mehcanistic reasons for this lag in the model

are still to be uncovered.

Fixing the growth rate and minimising enzyme usage as the objective may not be reflective of true cellular behaviour. It has previously been predicted that, during growth on glucose, 30% of the *E. coli* proteome is taken up by idle enzymes [122]. Hence, interpreting the results of enzyme minimisation simulations must be done with caution, and different optimisation strategies must be investigated.

The second optimisation approach to model the onset of overflow metabolism was minimising acetate secretion. There is no obvious evolutionary reason for a cell to minimise acetate secretion, but it can be useful in understanding the metabolic switch between respiration and overflow metabolism. It is encouraging that the model predicts the onset of overflow metabolism at almost the same growth rate as using enzyme minimisation (approximately 2% higher for all membrane bounds). This leads one to conclude that whilst the choice of objective is pivotal in certain ecGSMM analyses, for instance gene essentiality or sensitivity analyses, it does not appear to be as important here. The size of the membrane is directly correlated with the onset of overflow metabolism, and this seems to be the key determinant in this model.

The final approach to model the onset of overflow metabolism was performed by fixing the glucose uptake rate, and maximising biomass production. It is known that overflow metabolism occurs at high growth rates under excess glucose conditions, and experiments have shown that it is possible to reduce overflow metabolism by adjusting the glucose uptake rate in *Bacillus subtilis* [179]. The simulations shown in 3.4 suggest that a decreased glucose uptake rate in *S. aureus* would also reduce overflow metabolism.

The three linear programming objectives used to simulate overflow metabolism gave similar predictions for the growth rate at which the metabolic switch occurs. Specifically, minimising acetate secretion or fixing glucose uptake and maximising growth rate gave almost identical predictions. This suggests that the metabolic strategy employed is not highly dependent on the LP objective. Ribosomal capacity is a key limiting factor for determining the growth rate of bacterial cultures [16, 21], which is not explicitly accounted for in ecGSMMs. Future work is needed to investigate the extent to which incorporating ribosomal constraints into the *S. aureus* model would affect predictive accuracy, but it is encouraging to see that three different LP objectives can be used to give the same metabolic predictions at low growth rates.

Comparing model predictions to TraDIS datasets provides an indication of how well the model captures gene essentiality. However, TraDIS data must be interpreted with caution. The classification of genes as essential or non-essential depends critically on the presence of insertions across the genome during library preparation. Genes with no insertions in the resulting library are designated as essential, as are genes with very few insertions.

The output of a TraDIS analysis is a frequency distribution of the number of insertions. This frequency distribution is typically bimodal, with the first peak at zero insertions. Likelihood ratios (LRs) are then calculated to assess the likelihood that a gene belongs to this first peak, and essentiality is assigned based on the log(LR), relative to an arbitrary threshold. Although TraDIS pipelines enable high-throughput assessment of gene essentiality, the outcomes should not be treated as definitive.

Besides gene essentiality predictions, sensitivity analysis of the growth rate can provide insights into potential drug targets. When growth is simulated using the original permease turnover numbers, the growth rate has a high sensitivity to these enzymes, implying that the ability of the cell to import and export metabolites limits the growth rate. To confirm whether or not this holds when permease reactions have a higher turnover number, I performed a growth sensitivity analysis in which permease turnover numbers were doubled. I found that they were no longer highly controlling, demonstrating that the choice of turnover number for permease reactions can have a substantial impact on the predicted control points in the metabolic network. When permease turnover is low, transport across the membrane becomes a bottleneck for growth, highlighting these reactions as potential drug targets. Increasing the turnover number alleviates this bottleneck, shifting control to other metabolic reactions. This underlines the need for accurate estimation of enzyme kinetics, especially for transporters, in order to obtain reliable sensitivity predictions in enzyme-constrained models.

Sensitivity analysis of model growth rates with respect to enzyme parameters is a standard approach in metabolic modelling. This analysis identifies enzymes whose partial inhibition impacts growth, but it does not reveal the broader effects on other metabolic fluxes. Certain enzymes are known to affect and increase virulence in *S. aureus*. Extending the sensitivity analysis to the whole active flux solution allows for predictions on the influence of turnover perturbations on all flux rates in the network. This provides a more comprehensive view of metabolic control, and can elucidate the potential impacts on virulence-associated pathways.

The whole solution sensitivity analysis highlights the enzymes that have a disproportionate impact on the entire metabolic state of the cell. Such enzymes represent critical control points whose inhibition or modulation could have broad effects on cellular physiology, beyond simply reducing growth. These findings underscore the importance of considering whole-network effects when prioritising drug targets or interpreting the consequences of enzyme perturbations in genome-scale models.

The NAD-independent lactate dehydrogenase Lqo is known to be essential for virulence. However, in the sensitivity analysis of respiratory enzymes during growth on glucose, Lqo was neither strongly influenced by changes in respiratory parameters nor exhibited high

control over fluxes (figure 3.7). This discrepancy suggests that the conditions encountered during infection differ substantially from those used in the growth simulations, as closer agreement between the sensitivity analysis and experimental observations would otherwise be expected.

When growth on L-lactate is simulated, the model captures the essential role of Lqo. This highlights the need to incorporate stress-specific conditions and alternative metabolic pathways into genome-scale models to improve the accuracy of pathogen behaviour predictions under host-relevant environments. Targeting Lqo may represent a promising strategy for sensitising *S. aureus* to immune-mediated killing.

Studying optimal flux modes in the model revealed two distinct metabolic strategies during growth. The presence of both a fermentative and a respiratory OFM in the optimal solution suggests that *S. aureus* can maintain two partially distinct growth-maximising modes. This flexibility may contribute to adaptability and pathogenicity across diverse environments. The predominance of the fermentative OFM suggests that *S. aureus* ST398 preferentially adopts a fermentative strategy under the simulated conditions.

The dominant fermentative OFM makes greater use of the Pta-Ack pathway, which is essential for growth under aerobic, excess glucose conditions. Inactivation of this pathway drasticially decreases the growth rate of an *S. aureus* strain grown on rich media [143]. To find further target enzymes for growth inhibition, a sensitivity analysis was performed on optimal OFM usage. Enzymes to which the OFMs are most sensitive may represent novel candidates for growth inhibition, although strain-specific experimental validation will be required to confirm these predictions.

This metabolic flexibility may provide a survival advantage during infection, supporting rapid ATP generation and adaptation to fluctuating oxygen levels in host tissues. The ability to switch between fermentative and respiratory modes may therefore facilitate both persistence and virulence. The relationship between the two OFMs highlights a fundamental trade-off in proteome allocation. Increased catalytic efficiency of cytosolic enzymes, such as those in glycolysis and fermentation, promotes the usage of the fermentative OFM, whereas enhanced turnover of membrane-associated enzymes, particularly those involved in oxidative phosphorylation and transport, shifts metabolism toward the respiratory OFM.

Enzymes located in the cytosol, such as those involved in glycolysis and the Pta-Ack pathway, promote the usage of the fermentative OFM when their catalytic efficiency increases. Conversely, improvements in the turnover of membrane-associated enzymes, particularly those involved in oxidative phosphorylation and transport, promote usage of the respiratory OFM. This is consistent with experimentally observed overflow behaviour in other bacteria [14]. From a physiological perspective, these findings imply that environmental

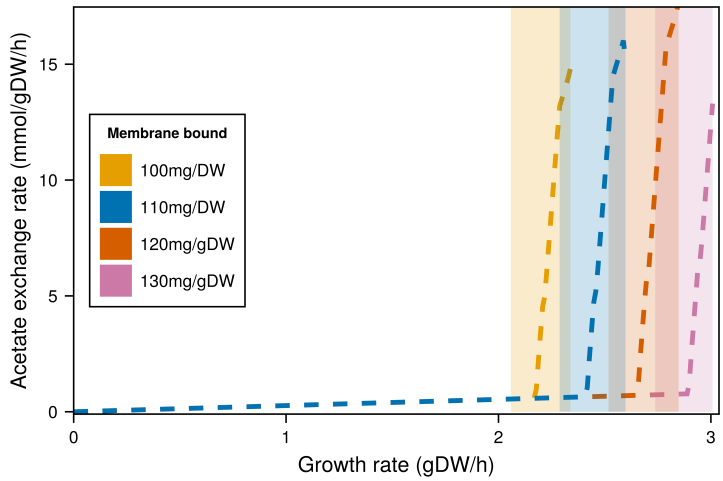
or genetic perturbations that alter the abundance or activity of key enzymes can shift the balance between fermentation and respiration.

The findings in this chapter underscore the importance of enzyme allocation constraints in shaping metabolic strategies and provide a mechanistic explanation for the observed flexibility of *S. aureus* metabolism. They provide a starting point for the development of novel anti-microbials for this livestock-associated strain.

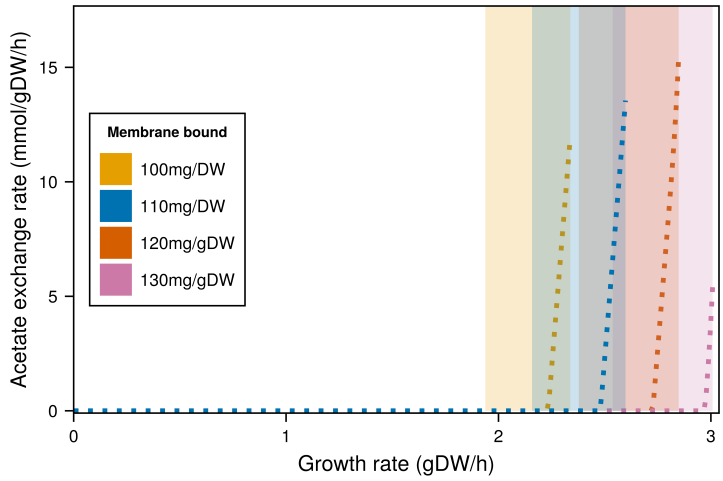
3.5 Supplementary

3.5.1 Acetate overflow

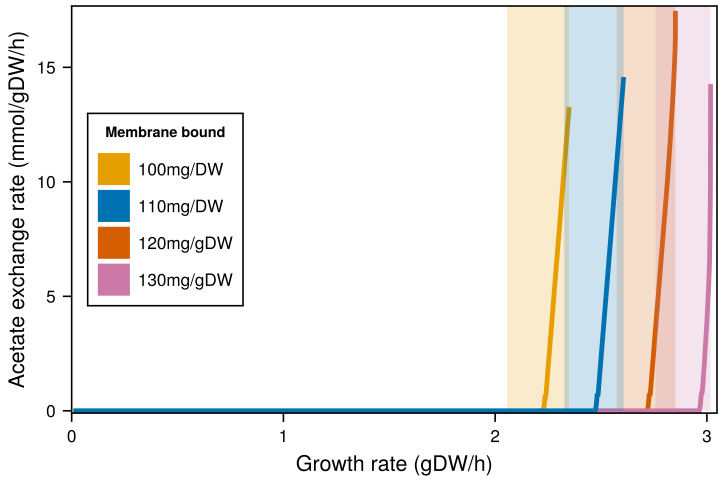
Figure 3.11 presents the acetate overflow simulations in three separate plots. This data is identical to that in figure 3.4. The shaded areas correspond to the growth rates at which the membrane bound is active. We see that this occurs at a similar point for all three LP objectives.



(a) Minimise enzymes



(b) Minimise acetate



(c) Fix glucose

Figure 3.11: The point at which the membrane bound becomes active and the onset of overflow metabolism are not greatly affected by the choice of LP objective.

Gene ID	Name	Rhea	Reaction
SAPIG0242	fadN	13284	(2S,3S)-3-hydroxy-2-methylbutanoyl-CoA:NAD+ oxidoreductase
		11171	3-hydroxypimeloyl-CoA:NAD+ oxidoreductase
		31190	(S)-hydroxydecanoyl-CoA:NAD+ oxidoreductase
		34854	(S)-3-hydroxyacyl-CoA:NAD+ oxidoreductase
		30802	(S)-3-hydroxybutanoyl-CoA:NAD+ oxidoreductase
		22435	(3S)-3-hydroxyacyl-CoA:NAD+ oxidoreductase
		31170	(S)-3-hydroxytetradecanoyl-CoA:NAD+ oxidoreductase
		31198	(S)-hydroxyoctanoyl-CoA:NAD+ oxidoreductase
		31182	(S)-3-hydroxydodecanoyl-CoA:NAD+ oxidoreductase
		31146	(S)-hydroxyhexanoyl-CoA:NAD+ oxidoreductase
		31162	(S)-3-hydroxyhexadecanoyl-CoA:NAD+ oxidoreductase
SAPIG0243	-	30734	butanoyl-CoA:electron-transfer flavoprotein 2,3-oxidoreductase
		30946	octanoyl-CoA:electron-transfer flavoprotein 2-oxidoreductase
		28361	tetradecanoyl-CoA:electron-transfer flavoprotein 2,3-oxidoreductase
		31210	hexanoyl-CoA:electron-transfer flavoprotein 2,3-oxidoreductase
		31042	lauroyl-CoA:electron-transfer flavoprotein 2,3-oxidoreductase
		30750	palmitoyl-CoA:electron-transfer flavoprotein 2,3-oxidoreductase
		39135	(6Z,9Z,12Z,15Z,18Z)-tetracosapentaenoyl-CoA:oxygen 2-oxidoreductase
		30850	glutaryl-CoA:electron-transfer flavoprotein 2,3-oxidoreductase (decarboxylating)
		38959	(6Z,9Z,12Z,15Z,18Z,21Z)-tetracosahexaenoyl-CoA:oxygen 2-oxidoreductase
		28357	decanoyl-CoA:electron-transfer flavoprotein 2,3-oxidoreductase
SAPIG0548	tmk	13520	ATP:dTMP phosphotransferase
		30658	ATP:dUMP phosphotransferase
SAPIG0566	prs	15612	ATP:D-ribose-5-phosphate diphosphotransferase
SAPIG0980	fabF	41916	41916
		41903	tetradecanoyl-[acyl-carrier protein]:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
		55040	55040
		41839	hexanoyl-[acyl-carrier protein]:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
		41823	butyryl-[acyl-carrier protein]:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)

		42251	butyryl-[acyl-carrier protein]:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
		22836	acyl-[acyl-carrier-protein]:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
		41887	dodecanoyl-[acyl-carrier-protein]:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
		41855	octanoyl-[acyl-carrier protein]:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
		41871	decanoyl-[acyl-carrier protein]:malonyl-[acyl-carrier-protein] C-acyltransferase (decarboxylating)
SAPIG1006	fabI	41912	hexadecanoyl-[acp]:NADP+ trans-2-oxidoreductase
		44968	tetradecanoyl-[acp]:NADP+ trans-2-oxidoreductase
		22567	acyl-[acyl-carrier protein]:NADP+ oxidoreductase
		44964	dodecanoyl-[acp]:NADP+ trans-2-oxidoreductase
		54900	hexadecanoyl-[acp]:NAD+ trans-2-oxidoreductase
		41832	hexanoyl-[acp]:NADP+ trans-2-oxidoreductase
		41928	octadecanoyl-[acp]:NAD+ trans-2-oxidoreductase
		52572	butyryl-[acp]:NAD+ trans-2-oxidoreductase
		36143	hexadecanoyl-[acp]:NADP+ trans-2-oxidoreductase
		44952	octanoyl-[acp]:NADP+ trans-2-oxidoreductase
		10240	acyl-[acyl-carrier protein]:NADP+ oxidoreductase
		41528	octanoyl-[acp]:NAD+ trans-2-oxidoreductase
		41896	tetradecanoyl-[acp]:NADP+ trans-2-oxidoreductase
		54936	decanoyl-[acp]:NAD+ trans-2-oxidoreductase
		41880	dodecanoyl-[acp]:NADP+ trans-2-oxidoreductase
		44960	decanoyl-[acp]:NADP+ trans-2-oxidoreductase
		52576	hexanoyl-[acp]:NADP+ trans-2-oxidoreductase
		41864	decanoyl-[acp]:NADP+ trans-2-oxidoreductase
		41848	octanoyl-[acp]:NADP+ trans-2-oxidoreductase
SAPIG1121	coaD	19804	ATP:pantetheine-4'-phosphate adenylyltransferase
SAPIG1207	gmk	12697	guanylate kinase
		20783	ATP:GMP phosphotransferase
SAPIG1209	coaBC	19400	(R)-4'-phosphopantothenate:L-cysteine ligase
		16796	N-[(R)-4'-phosphopantothenoyl]-L-cysteine carboxy-lyase
SAPIG1229	fabD	41795	malonyl-CoA:[acyl-carrier-protein] S-malonyltransferase
SAPIG1260	pyrH	24403	ATP:UMP phosphotransferase
SAPIG1274	ribF	14360	ATP:riboflavin 5'-phosphotransferase
		17240	ATP:FMN adenylyltransferase
SAPIG1467	ndk	44640	44640
		27681	ATP:dCDP phosphotransferase
		27685	ATP:dTDP phosphotransferase

		25237	ATP:2'-deoxy-5-hydroxymethylcytidine-5'-diphosphate phosphotransferase
		30350	ATP:IDP phosphotransferase
		30942	ATP:dIDP phosphotransferase
		27677	ATP:dADP phosphotransferase
		18116	ATP:nucleoside-diphosphate phosphotransferase
		25098	ATP:5-fluorouridine-diphosphate phosphotransferase
		28585	ATP:dUDP phosphotransferase
		27693	ATP:dGDP phosphotransferase
		27689	ATP:GDP phosphotransferase
SAPIG1742	coaE	18248	ATP:dephospho-CoA 3'-phosphotransferase
SAPIG2141	fabZ	13097	3-hydroxyglutaryl-[acp]-methyl-ester hydro-lyase
		41811	(3R)-3-hydroxyhexanoyl-[acyl-carrier-protein] hydro-lyase
		39159	(3R)-3-hydroxypalmitoyl-[acyl-carrier-protein] hydro-lyase
		41879	(3R)-3-hydroxydodecanoyl-[acyl-carrier-protein] hydro-lyase
		41831	(3R)-3-hydroxyhexanoyl-[acyl-carrier-protein] hydro-lyase
		41847	(3R)-3-hydroxyoctanoyl-[acyl-carrier-protein] hydro-lyase
		41895	(3R)-3-hydroxytetradecanoyl-[acyl-carrier-protein] hydro-lyase
		41863	(3R)-3-hydroxydecanoyl-[acyl-carrier-protein] hydro-lyase
		41908	(3R)-3-hydroxypalmitoyl-[acyl-carrier-protein] hydro-lyase
SAPIG2161	tdk	28209	ATP:deoxyuridine 5'-phosphotransferase
		19129	ATP:5-fluorodeoxyuridine 5'-phosphotransferase
SAPIG2172	coaW	16373	ATP:pantothenate 4'-phosphotransferase
		22475	ATP:pantothenate 4'-phosphotransferase

Table 3.8: List of the 17 genes predicted to be essential in the model during growth on rich medium, with their associated reactions.

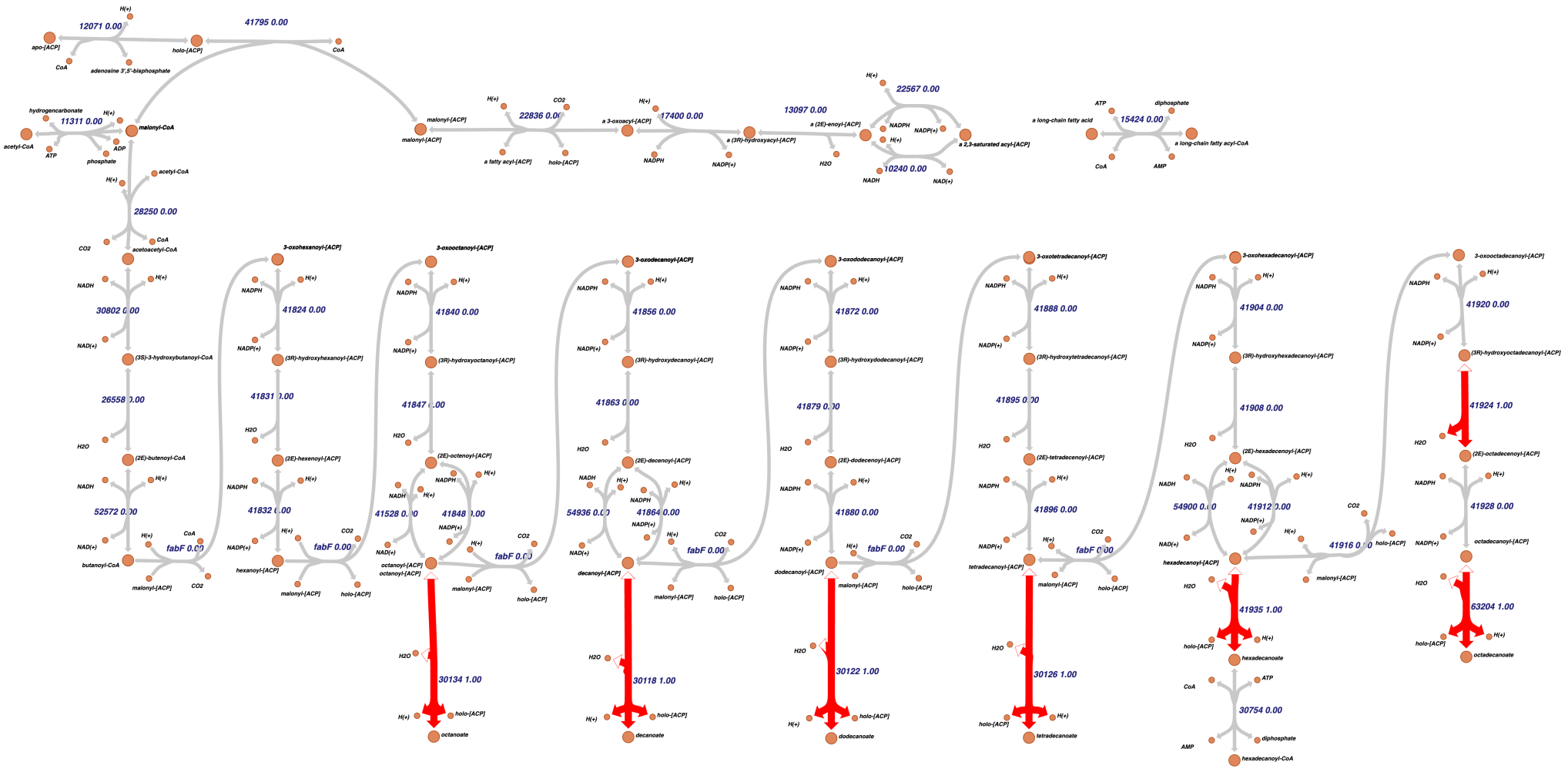


Figure 3.12: Escher map of fatty acid biosynthesis, identical to figure 3.10 at higher resolution.

Chapter 4

In silico drug toxicity prediction

This chapter was conceptualised by myself, Linda Chiappalupi and Oliver Ebenhöh. Methodology and software were a collaboration between myself, Linda Chiappalupi, Marvin van Aalst and Kevin Roß. My further contributions were: validation, formal analysis, investigation, and writing the original draft.

The software for the prediction of phase I and phase II metabolic products is available as an open source python package from `https://pypi.org/project/drug_toxicity`, and the git page is `https://github.com/HettieC/DrugToxicity`.

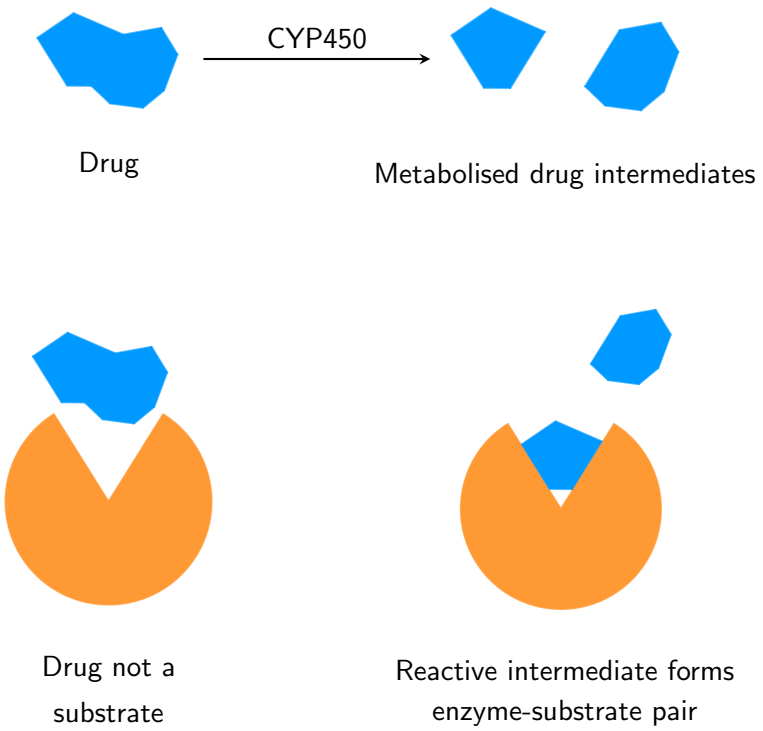


Figure 4.1: Demonstration of toxicity resulting from metabolised drug intermediates. On the left, we see that the drug (blue) does not fit the active site of a metabolic enzyme (orange), and therefore does not cause any harmful effects in itself. The drug is metabolised via CYP450 and phase I metabolism into two intermediates (top right), one of which fits reasonably well into the active sit of the enzyme (bottom right), forming an enzyme-substrate pair.

4.1 Introduction

In searching for novel therapeutics against *Staphylococcus aureus*, it is not only necessary that a potential drug is able to kill the bacteria, but it must also not harm the host. An initial filtering for potential drug targets would first check that there are no gene homologues in the first. After this, should a drug candidate be found, rigorous testing must take place before a candidate can be used in clinical trials, and then enter the market. To be viable, a therapeutic agent must also demonstrate safety, tolerability, and minimal adverse effects in the human host.

Despite substantial investments in drug development, the success rate remains exceedingly low. It is estimated that only 0.1% of drug candidates entering the pre-clinical stage ultimately reach the market [49, 8]. Notably, the primary cause of failure is not necessarily the lack of therapeutic efficacy, but rather unanticipated toxicological effects in humans. The economic burden associated with such failures is substantial: the average cost to bring a new drug to market is estimated at $2.87 million [48], with the pre-clinical phase alone requiring approximately five and a half years and accounting for nearly one-third of total development costs [126].

Improving the alignment between pre-clinical predictions and clinical outcomes is therefore a key priority. More accurate early-stage assessments of toxicity would not only reduce the likelihood of late-stage attrition but also shorten development timelines and lower financial costs. This is especially urgent in the context of multidrug-resistant infections, where delays in therapeutic innovation can have serious public health consequences.

A major contributor to drug toxicity in humans arises from the metabolic transforma-

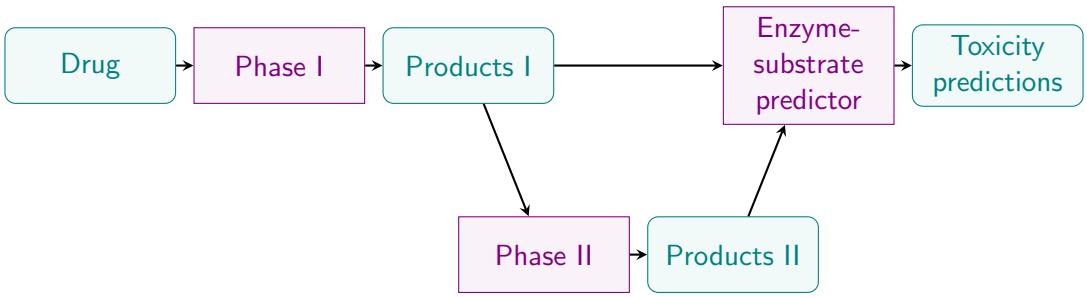


Figure 4.2: Pipeline for drug toxicity predictions. A drug is entered into the pipeline, in SMILES format, phase I reactions are applied to the drug. The products of these reactions (Products I) are then metabolised using phase II reactions to produce Products II. An enzyme-substrate predictor is used on both the phase I and phase II products with enzymes in the liver, hypotheses on toxic effects can then be made from these results.

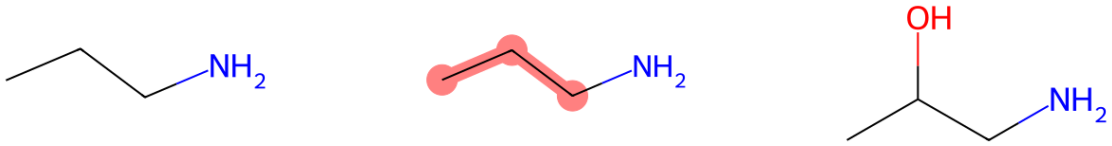
tion of drugs into reactive or harmful intermediates [12]. The liver, the main location of drug metabolism, plays a central role through a series of enzymatic reactions categorised into phase I and phase II metabolism. Phase I reactions (modification) primarily involve oxidation, reduction, or hydrolysis, and are mainly catalysed by cytochrome P450 (CYP450) enzymes. These reactions often introduce or expose functional groups, resulting in metabolites that may be more chemically reactive than the parent compound. Phase II reactions (conjugation) involve the covalent attachment of endogenous molecules, such as glucuronic acid, sulfate, or glutathione, to the drug or its Phase I metabolites. These conjugations generally increase the water solubility of the compounds, and facilitate excretion. A depiction of the toxicity of metabolised products in contrast to the parent drug compound is given in figure 4.1.

Traditionally, toxicity assessments rely on a combination of in vitro assays and in vivo animal models to evaluate safety profiles before human trials. While these approaches provide critical information, they are expensive, time-consuming, and frequently fail to capture human-specific metabolic pathways and toxicities due to species differences. The accurate determination of human toxicity is a major challenge in the development of novel drug compounds [50, 97, 63]. Moreover, the scale of modern drug discovery pipelines makes it impractical to experimentally evaluate the toxicity of every lead compound. As such, there is an increasing demand for the development of in silico approaches that can complement and enhance traditional toxicological methods.

The phase I and phase II reactions catalysed by CYP450 enzymes are well described [78], and it is known that metabolism of drugs by these enzymes often results in more reactive and more harmful products. When developing new therapeutics, the safety of the drug to human patients is of utmost importance. Here, we provide rule-based reaction templates for cytochrome P450-mediated oxidation and conjugation. The major metabolic products of novel drugs can therefore be anticipated, prior to costly in vitro and in vivo experiments.

Knowing the expected products of the metabolism of novel drugs is not sufficient to decide whether to continue research. Knowledge of the toxicity of these products is also required. If all predicted products are well-characterised, the pipeline can stop here, and decisions can be made on how to progress. However, if investigated a novel drug, it is likely that its metabolism will lead to unknown or poorly characterised products. Therefore, an indication of the potential toxicity of these products is needed.

Using an open-source machine-learning tool for the prediction of small molecule substrates of enzymes, we are able to give an initial prediction for the toxicity of a compound [92]. Enzymes catalysing reactions with their non-canonical substrates are known as promiscuous enzymes. In the context of drug toxicity, we are interested in the po-



(a) Propylamine (b) C-chain highlighted (c) 1-Amino-2-propanol

Figure 4.3: Demonstration of aliphatic hydroxylation of propylamine, using RDKit. The original compound (propylamine) is displayed in (a), with the aliphatic carbon chain highlighted in (b), this is where the reaction will take place. The result of the hydroxylation is shown in (c).

tential for metabolised drug products to bind to the active site of non-CYP450 enzymes. In many cases, broad specificity of an enzyme provides a fitness advantage. For example, glutathione S-transferases detoxify electrophilic compounds in the liver, and possessing substrate-promiscuity allows these enzymes to act on a wide range of substrates as needed [84]. However, promiscuity can also cause problems. If too high a portion of an available enzyme is bound to non-canonical substrates, this enzyme cannot sufficiently perform its metabolic role. Thus, we are interested in finding non-canonical enzyme-substrate pairs from metabolised drug products as this could uncover the causes of hepatotoxicity.

This chapter presents the development of a Python-based in silico pipeline designed to predict the metabolic products of phase I and phase II liver metabolism, and the resulting hepatotoxicity implications. A description of an earlier iteration of the methods was reported as a general exploration of its capabilities [33]. Here, an application to acetaminophen (paracetamol) metabolism is investigated as a case study. The objective is to provide early indicators of potential toxicity during drug design, and to improve the predictive accuracy of pre-clinical assessments, supporting the development of safer therapeutic agents.

4.2 Methods

To simulate the breakdown of drugs in the liver, phase I and phase II reactions were defined using RDKit [135]. This is an open source toolkit for performing chemical reactions. It identifies molecular moieties in the substrates, and performs the chemical transformation. In figure 4.3, propylamine is used as an example to demonstrate aliphatic hydroxylation. Here, a hydroxyl (-OH) group is added to an aliphatic carbon chain. An aliphatic chain is described using SMARTS notation as "$[\#6:1]-[\#6:2]-[\#6:3]$", where the "#6" refers to carbon, the ": 1", ": 2" or ": 3" give the position in the molecule, and the "-" tells us that there are single bonds. In figure 4.3b, this chain is highlighted by RDKit, and the carbon in the centre is where the hydroxylation will occur. Figure 4.3c shows the end product of aliphatic hydroxylation of propylamine.

For the pipeline, SMARTs IDs were described for 48 molecular moieties, in order to define 22 phase I and 9 phase II reactions. To perform a reaction, the required molecular moieties are matched in the chosen substrate, and RDKit generates the plausible products of the reaction.

The pipeline for predicting the metabolic products of a drug is presented in figure 4.2. Products of phase I metabolism are metabolised via phase II reactions, or are directly supplied to a machine learning model for the prediction of substrates of enzymes. A

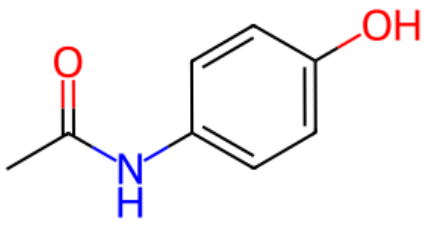


Figure 4.4: Acetaminophen chemical structure.

published tool was used for enzyme-substrate pair predictions [92]. The inputs of this tool are an enzyme sequence and a SMILES description of the molecule. A model based on over 18,000 enzyme-substrate pairs with experimental evidence then gives a prediction score between 0 and 1.

In this work we investigate the prediction of enzyme-substrate pairs from 952 liver-associated enzymes with the phase I metabolised products of acetaminophen (paracetamol). However, it is trivial to change the target enzymes of interest. For example, one could follow the same pipeline as presented here, but use proteins upregulated in the heart instead of the liver. Since cardiac disorders are responsible for almost 20% of drugs withdrawn from the market between 1990 and 2010 [40], this investigation would also be beneficial.

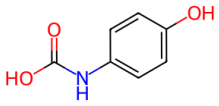
4.3 Results

To investigate the reliability of results of the pipeline, acetaminophen (APAP) was used as an initial case-study. Also known as paracetamol, APAP is one of the most widely used drugs worldwide, and the top most commonly used in the United States [86]. Although safe at therapeutic doses, overdose of APAP can cause severe liver injury, with the first cases of hepatotoxicity reported in 1966 [43]. Since its use is so prevalent, the mechanisms of toxicity have been relatively well described. The chemical structure of APAP is given in figure 4.4.

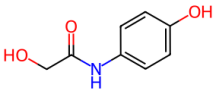
Cytochrome P450 enzymes are largely responsible for the metabolic activation of APAP [130]. N-acetyl-p-benzoquinone imine (NAPQI) is believed to be the reactive metabolite most responsible for hepatotoxicity [41]. To check whether the pipeline could reproduce this, we ran the phase I metabolic reactions on acetaminophen. In figure 4.5, we see that there are 15 total products of phase I metabolism.

Encouragingly, NAPQI is among the predicted products of phase I metabolism on APAP, as seen in figure 4.5l. To investigate the ability of our pipeline to predict hepatotoxicity, a machine learning model for the prediction of enzyme-substrate pairs was used [91]. A total of 7 enzymes were predicted to bind to NAPQI, with a likelihood of more than 50%, these are detailed in table 4.1.

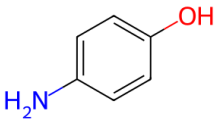
The protein with the highest prediction score, P, in table 4.1 is DCXR, responsible for L-xylulose reductase. DCXR is also involved in the detoxification of xenobiotics, and is believed to relieve carbonyl stress in the liver [95]. It has been previously found that



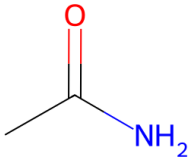
(a) N-Boc-4-hydroxyaniline



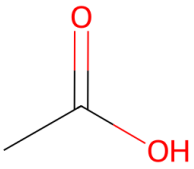
(b) 2-hydroxy-N-(4-hydroxyphenyl) acetamide



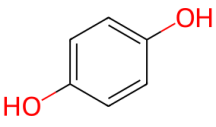
(c) 4-aminophenol



(d) acetamide



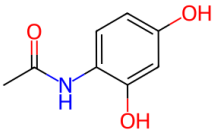
(e) acetic acid



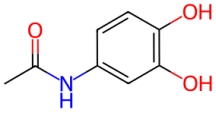
(f) benzene-1,4-diol

(g) methane

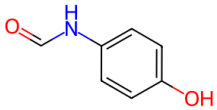
(h) methanol



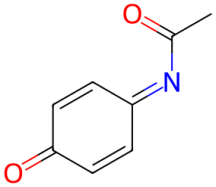
(i) N-(2,4-dihydroxyphenyl) acetamide



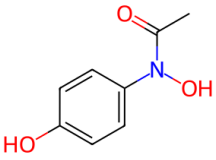
(j) N-(3,4-dihydroxyphenyl) acetamide



(k) N-(4-hydroxyphenyl) formamide



(l) Acetimidoquinone (**NAPQI**)



(m) N-hydroxy-N-(4-hydroxyphenyl) acetamide

Figure 4.5: Phase I metabolic products of acetaminophen, calculated using the pipeline described here.

Gene	Protein name	P
DCXR	Dicarbonyl/L-xylulose reductase	0.63
HSD17B2	17-beta-hydroxysteroid dehydrogenase type 2	0.53
HSD3B7	3 beta-hydroxysteroid dehydrogenase type 7	0.52
SIGMAR1	Sigma non-opioid intracellular receptor 1	0.52
LCAT	Phosphatidylcholine-sterol acyltransferase	0.51
LDHD	Probable D-lactate dehydrogenase; mitochondrial (DLD) (Lactate dehydrogenase D)	0.51
TMT1B	Thiol S-methyltransferase	0.51

Table 4.1: Enzyme-substrate predictions of NAPQI. The column P gives the prediction score from the enzyme-substrate prediction tool [92].

DCXR is downregulated in a rat model receiving a high dose of APAP, with elevated ALT and AST levels [61]. Contrastingly, DCXR is upregulated in male zebrafish on exposure to high doses (1152μg/L) of APAP [11].

4.4 Discussion

The in silico pipeline described here for the simulation of drug metabolism in the liver, followed by the prediction of enzyme inhibitory effects, is in its infant stage. Since the study started, a machine learning based model for the prediction of enzyme-substrate pairs has been published [92].

The enzyme dicarbonyl/L-xylulose reductase (DCXR) has the highest prediction score for binding to NAPQI, the toxic metabolite from acetaminophen. This enzyme acts as a defence mechanism against potentially harmful exogenous or endogenous carbonyls [53]. The prediction seems reasonably, due to the carbonyl group on NAPQI, and the up-/down-regulation of DCXR in different animal models on exposure to high doses of acetaminophen. First discovered to have enzymatic activity in 1970, DCXR is relatively well characterised [175]. The high prediction score for acting on NAPQI, as well as its differential expression in two animal models, means that our pipeline has identified a future potential avenue for research into hepatotoxicity of acetaminophen. On exposure to high doses of acetaminophen, DCXR has been found to be over- or under-expressed depending on the animal model used. This highlights the need for accurate and reliable animal and in silico models for investigating the effects of drugs on human metabolism.

Among the remaining enzymes predicted most likely to form enzyme-substrate pairs with NAPQI is HSD17B2. In a microfluid biochip of HepG2/C3A cells (epithelial-like cells isolated from a the liver of a human patient with hepatocellular carcinoma), HSD17B2 is upregulated upon treatment with APAP [132]. While HSD17B2 is primarily involved in steroid metabolism, its upregulation in this context raises the possibility of a role in the cellular response to APAP-induced hepatotoxicity.

Further investigation of HSD17B2 in the context of acetaminophen hepatotoxicity could follow several lines. Direct biochemical assays using recombinant HSD17B2 and NAPQI could determine whether covalent modification or inhibition occurs, supported by LC-MS/MS detection of adducted peptides. Cellular studies in hepatocyte models could assess whether HSD17B2 knockdown or overexpression alters APAP-induced cell death, glutathione depletion, or protein adduct formation, and whether these effects correlate with steroid hormone balance. Targeted proteomics in APAP-treated liver samples could confirm in situ modification, while humanized liver models or 3D organoids would help clarify the human relevance of any observed effects. The identification of HSD17B2 along-

side DCXR highlights the potential of the in silico pipeline to reveal less obvious enzyme targets of NAPQI and suggests that integrating reactivity filters with enzyme–substrate prediction tools may improve candidate prioritization.

The tool used for enzyme-substrate pair prediction acknowledges that predictions are most reliable when the inputted substrate was part of the training data [92]. This is clearly a downside when investigated novel drugs, as it is likely that their metabolised products will not have been used in the training data. It is necessary to acknowledge this as a drawback to our pipeline. Two new machine-learning based tools for the prediction of enzyme-substrate pairs have been published this year, but both were based on the same training dataset as the tool used here [51, 118]. Further improvements in predicting the toxic effects of drug metabolites depends on advances in enzyme-substrate pair prediction methods, particularly the broadening of training datasets.

Methanol and methane were predicted as products of phase I metabolism of APAP (see figure 4.5). Methanol is known to be present in small quantities in the blood, with a sharp increase after methanol-free ethanol [150]. Methane in a breath test is associated with non-alcoholic fatty liver disease [6] and excessive alcohol intake [164]. Both of these compounds have been found to alleviate the effects of APAP-induced liver injury [56, 103], but neither have previously been identified as products of APAP metabolism. Targeted metabolomic analysis would be required to identify whether these products are indeed produced via phase I metabolism of APAP.

The pipeline may overestimate the number of products arising from phase I and phase II metabolism. For instance, Figure 4.5 shows N-Boc-4-hydroxyaniline as a predicted product of APAP metabolism. This compound can serve as a precursor in the synthesis of photoisomerisation molecules [115], but has not been reported in human metabolism. Given the extensive study of APAP, it is unlikely that N-Boc-4-hydroxyaniline represents a genuine phase I metabolite. This highlights the need for additional filtering of prediction results based on the plausibility of their occurrence resulting from human metabolic pathways.

Chapter 5

Discussion

This thesis had two main goals: firstly, to better understand the metabolism of a livestock-associated strain of *Staphylococcus aureus*, and secondly to improve pre-clinical drug development. To this end, I have developed novel analysis tools for enzyme-constrained genome-scale metabolic models (chapter 2), alongside an ecGSMM of *S. aureus* (chapter 3). Further, I have created a pipeline for the prediction of metabolic drug toxicity effects (chapter 4).

5.1 Summary of results

Chapter 2 introduces the concept of optimal flux modes (OFMs), extending the well-established framework of elementary flux modes to genome-scale models with non-homogeneous constraints. As a natural way to decompose non-homogeneous GSMMs, OFMs capture the elementary modes that are able to achieve optimal growth. As such, they better capture the active metabolic pathways in GSMMs than EFMs that are solely based on the network stoichiometry. To study the effects of parameter changes on model variables, one may use implicit differentiation to perform a sensitivity analysis of an active solution to an ecGSMM. OFMs are central to a proof that this differentiation is always possible. Beyond this, I introduced the differentiation of OFM usage, providing a new way to analyse ecGSMMs.

An important implication of chapter 2 is the recognition that standard GSMMs possess a vast optimal solution space, since they are typically constrained only by a small number of import/export fluxes and an ATP maintenance reaction. As a result, any calculated optimal solution is not unique. To address this, optimal flux modes provide alternative minimal representations of these optimal states. Their enumeration in standard GSMMs can be achieved using pre-existing methods for elementary flux modes (EFMs) [104, 30, 160, 83], where one simply inputs the matrix $\boldsymbol{A}$ from equation 2.5 into the chosen EFM calculator.

An optimal solution can be decomposed into a small number of OFMs, no greater than the number of active constraints imposed on the ecGSMM. These OFMs can be considered as the building blocks of metabolism, and are used in some optimal ratio. Investigating how this optimal ratio changes upon parameter perturbations, via OFM differentiation, is a novel method to investigate metabolism.

Chapter 3 concerns the building and manual curation of an enzyme-constrained genome-

scale model of *Staphylococcus aureus*. Here, it was found that the optimal growth on minimal media is achieved through the simultaneous usage of a respiratory and a fermentative OFM. The presence of these two different metabolic strategies highlights the adaptability of this pathogen, and helps explain its ability to grow in different environments.

Chapter 4 presents a computational pipeline for the prediction of metabolic drug toxicity. Firstly, drugs are metabolised into their phase I and phase II products. Then, predictions for enzyme-substrate pairs are made for enzymes of interest with these metabolic products. Products with a high predicted score for making enzyme-substrate pairs are possible causes of toxicity.

5.2 Future perspectives

The usage of OFMs in a pathogenic bacterium is examined in Chapter 3, and suggests that the pathogen is metabolically flexible. However, OFM analysis should not be limited to this context. Metabolic engineering for the production of molecules of interest using microbial cell factories is a particular field in which OFMs can be utilised. In an optimal solution of an ecGSMM using n total enzyme constraints, there will be n active OFMs, under conditions specified in chapter 2. When one OFM produces a greater amount of the target molecule per gDW biomass, OFM sensitivity analysis can highlight the enzymes exerting the most control on the proportional usage of the OFMs. These enzymes provide targets for enzyme engineering. As tools for the prediction of kinetic parameters increase, so will the propensity for reliable predictions using ecGSMMs and OFMs.

The ecGSMM of *S. aureus* ST398 presented in chapter 3 provides a computational framework for investigating the metabolism of this livestock-associated strain. Progress in extending this work is currently limited by the scarcity of interpretable experimental data. Current data available for this strain is largely obtained from growth cultures using ex vivo models [165], animal models [134], brain-heart infusion (BHI) [35], or tryptic soy broth (TSB) [5]. It is not possible to simulate chemically defined media to accurately mimic any of these experimental setups, and it is thus not possible to directly analyse any produced data with a genome-scale metabolic model. Particularly in the case of BHI and TSB, the benefits of also using a chemically defined medium cannot be overstated. Knowing precisely what a bacteria is grown on allows for the analysis of active metabolic pathways from a GSMM (or an ecGSMM). With this, the precise cause for under- or over-expression of proteins and genes can be found. Future development of the ecGSMM of *S. aureus* would greatly benefit from quantitative proteomics and metabolomics taken from growth on defined media. Essential gene predictions using TraDIS on a chemically defined medium would also improve the model, and allow for further curation to make more accurate predictions of drug targets. Finally, detailed organism-specific biomass composition is difficult to find for organisms other than *E. coli*. This data is crucial in defining a biomass reaction, which greatly affects the predictive capabilities on a GSMM or ecGSMM.

Feasible drug targets in pathogenic bacteria must fulfil two main criteria. Firstly, inhibition of a given target should significantly impair bacterial growth, ideally rendering the cell non-viable under infection-relevant conditions. Secondly, the target must not have close homologues in the host, since inhibition of shared or highly conserved proteins would risk off-target effects and host toxicity.

Enzyme-constrained genome-scale metabolic models (ecGSMMs) provide a powerful framework for systematically assessing these criteria. By incorporating enzyme kinetics and proteome usage limits, ecGSMMs move beyond purely stoichiometric predictions and enable a more quantitative evaluation of enzyme inhibition effects. Candidate targets can be identified by simulating complete or partial inhibition of specific enzymes and quantifying

the resulting impact on growth. Targets predicted to be highly sensitive in this framework are more likely to represent true vulnerabilities, as the enzyme constraints capture the trade-offs faced by bacteria in balancing fermentative and respiratory strategies, as well as proteome allocation under stress.

The second criterion, minimising similarity to host proteins, can then be evaluated by cross-referencing model-identified essential enzymes against the human proteome to exclude those with close homologues. Beyond this comparative filtering, host ecGSMMs can be incorporated alongside pathogen-specific models to enable more integrative analyses. For example, constructing a community model where the two members represent the bacterium and a tissue- or cell-specific host ecGSMM would allow direct evaluation of metabolic interactions under gene knockouts or enzyme inhibitions. Such an approach would make it possible to assess not only the essentiality of bacterial targets but also the consequences of their inhibition for host metabolism. In doing so, this strategy provides a framework to bridge the gap between computational predictions and clinically relevant drug discovery, guiding the identification of targets that are both effective and selective.

The final topic of this thesis was in silico drug toxicity predictions. This presented early-stage investigations into the modelling of phase I and phase II reactions, and the prediction of toxic effects from the metabolised drug products. Further improvements on the pipeline would come from the improvement of the machine learning-based enzyme-substrate predictor. Currently, the accuracy of this tool is limited by the training data, specifically the availability of enzymes forming enzyme-substrate pairs with less common compounds. As this data becomes more available, the predictive capabilities will improve.

Building on the initial drug-toxicity prediction pipeline, an important next step would be to integrate the predictions into a genome-scale metabolic model. Several reconstructions of human hepatocytes already exist [64, 105, 79, 4], providing a foundation for this approach. Using the pipeline developed in this thesis, one can generate a list of candidate enzymes predicted to be affected by drug-derived metabolites. Two complementary strategies can be employed for modelling the hepatotoxic effects of these compounds within GSMMs.

First, the metabolic enzymes in the list of candidates can be modelled as fully inhibited in a GSMM, effectively blocking their associated reactions. This would assume that a compound forming an enzyme-substrate pair is fully inhibiting of that enzyme and prevents it from catalysing its reactions. While straightforward, this approach likely oversimplifies reality, since most enzymes have multiple substrates and may retain their primary function despite partial inhibition.

A more refined strategy would involve constructing an enzyme-constrained GSMM, by extending an existing model using experimentally measured or predicted kinetic parameters. In this context, candidate enzymes could be partially inhibited by reducing their turnover numbers, simulating diminished catalytic capacity. The impact of these perturbations on predicted growth rates and metabolic flux distributions could then serve as a proxy for quantifying potential hepatotoxic effects of drug metabolites.

5.3 Conclusions

By developing novel approaches for analysing enzyme-constrained genome-scale metabolic models, I have shown how optimal flux mode analysis can reveal metabolic flexibility and pathogenic strategies in *Staphylococcus aureus*. This work highlights the potential of ecGSMMs not only for identifying condition-specific vulnerabilities in pathogens but also for understanding fundamental trade-offs in proteome allocation. In parallel, the extension of drug-toxicity prediction pipelines towards metabolic modelling underscores the value of

integrating computational toxicology with human genome-scale models, offering a route to more predictive and mechanistic assessments of hepatotoxicity. Together, these contributions illustrate the versatility of constraint-based modelling frameworks across diverse biological contexts, reinforcing their role as powerful tools to guide both basic research and applied efforts in antimicrobial discovery and predictive medicine.

In conclusion, this thesis demonstrates how systems biology can be used in addressing questions that range from fundamental mechanisms of cellular resource allocation to applied challenges in antimicrobial discovery and predictive toxicology. By extending and refining analytical methods for ecGSMMs, this work has shown how more nuanced interpretations of metabolic trade-offs can be achieved, offering insights into the balance between fermentative and respiratory strategies and the constraints that cause them. The application of these methods to *Staphylococcus aureus* revealed how metabolic flexibility underpins the ability of this pathogen to adapt and persist in diverse environments, and highlighted strain-specific vulnerabilities that could inform the development of novel therapeutic strategies.

In parallel, the investigation into drug-induced hepatotoxicity through computational pipelines demonstrates how mechanistic models can be used not only to describe biological behaviour but also to anticipate the unintended consequences of therapeutic interventions. The integration of predictive toxicology with genome-scale models represents a promising direction towards more reliable pre-clinical assessment frameworks, with the potential to reduce reliance on animal testing and improve the accuracy of drug safety evaluation.

Together, these contributions underline the unifying theme of this work: systems biology provides powerful frameworks to address practical questions. By advancing the methodology of ecGSMM analysis, applying these tools to the pressing problem of antibiotic resistance, and laying the groundwork for model-based prediction of hepatotoxicity, this thesis contributes to a growing toolkit that supports metabolic research. Looking forward, the continued integration of high-quality omics data, robust computational methods, and close collaboration with experimental and clinical researchers will be essential for realising the full potential of systems biology in improving human health.

List of Figures

List of Tables

Bibliography

[1] Desouky AM Abd-El-Haleem, Marwa R Elkatory, and Gadallah M Abu-Elreesh. "Uncovering novel polyhydroxyalkanoate biosynthesis genes and unique pathway in yeast hanseniaspora valbyensis for sustainable bioplastic production". In: *Scientific Reports* 14.1 (2024), p. 27162.

[2] Alyaa M Abdel-Haleem et al. "Functional interrogation of Plasmodium genus metabolism identifies species-and stage-specific differences in nutrient essentiality and drug targeting". In: *PLoS computational biology* 14.1 (2018), e1005895.

[3] Akshay Agrawal et al. "Differentiating through a cone program". In: *arXiv preprint arXiv:1904.09043* (2019).

[4] Rasmus Agren et al. "Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT". In: *PLoS computational biology* 8.5 (2012), e1002518.

[5] Dina Al Nahhas et al. "Comparative proteomics analysis of MRSA under different experimental conditions". In: *Scientific Reports* 15.1 (2025), p. 28656.

[6] Sanggwon An et al. "Methane gas in breath test is associated with non-alcoholic fatty liver disease". In: *Journal of Breath Research* 18.4 (2024), p. 046005.

[7] Laurence Armand-Lefevre, Raymond Ruimy, and Antoine Andremont. "Clonal comparison of Staphylococcus aureus isolates from healthy pig farmers, human controls, and pigs". In: *Emerging infectious diseases* 11.5 (2005), p. 711.

[8] John Arrowsmith and Philip Miller. "Phase II and Phase III attrition rates 2011-2012". In: *Nature reviews Drug discovery* 12.8 (2013), pp. 569–570.

[9] Samia Arshad et al. "Ceftaroline fosamil monotherapy for methicillin-resistant Staphylococcus aureus bacteremia: a comparative clinical outcomes study". In: *International Journal of Infectious Diseases* 57 (2017), pp. 27–31.

[10] Hélène Aubry-Damon et al. "Antimicrobial resistance in commensal flora of pig farmers". In: *Emerging infectious diseases* 10.5 (2004), p. 873.

[11] Steve Uwa Ayobahan. "Proteomics as an alternative approach for hazard characterization and the identification of specific chemical targets: elucidation of potential biomarkers for differentiating endocrine disruption from hepatotoxicity". PhD thesis. Dissertation, RWTH Aachen University, 2020, 2020.

[12] Thomas A Baillie. "Metabolism and toxicity of drugs. Two decades of progress in industrial drug metabolism". In: *Chemical research in toxicology* 21.1 (2008), pp. 129–137.

[13] Parit Bansal et al. "Rhea, the reaction knowledgebase in 2022". In: *Nucleic acids research* 50.D1 (2022), pp. D693–D700.

[14] Markus Basan et al. "Overflow metabolism in Escherichia coli results from efficient proteome allocation". In: *Nature* 528.7580 (2015), pp. 99–104.

[15] Qasim K Beg et al. "Intracellular crowding defines the mode and sequence of substrate uptake by Escherichia coli and constrains its metabolic activity". In: *Proceedings of the National Academy of Sciences* 104.31 (2007), pp. 12663–12668.

[16] Nathan M Belliveau et al. "Fundamental limits on the rate of bacterial growth and their influence on proteomic composition". In: *Cell systems* 12.9 (2021), pp. 924–944.

[17] Konstantin Bellut, Kristoffer Krogerus, and Elke K Arendt. "Lachancea fermentati strains isolated from kombucha: fundamental insights, and practical application in low alcohol beer brewing". In: *Frontiers in microbiology* 11 (2020), p. 764.

[18] Dany JV Beste et al. "GSMN-TB: a web-based genome-scale network model of Mycobacterium tuberculosis metabolism". In: *Genome biology* 8.5 (2007), R89.

[19] Meera Bhat et al. "Staphylococcus aureus ST398, New York City and Dominican Republic". In: *Emerging infectious diseases* 15.2 (2009), p. 285.

[20] Samira van den Bogaard, Pedro A Saa, and Tobias B Alter. "Sensitivities in protein allocation models reveal distribution of metabolic capacity and flux control". In: *Bioinformatics* (2024), btae691.

[21] Evert Bosdriesz et al. "How fast-growing bacteria robustly tune their ribosome concentration to approximate growth-rate maximization". In: *The FEBS journal* 282.10 (2015), pp. 2029–2044.

[22] Emanuele Bosi et al. "Comparative genome-scale modelling of Staphylococcus aureus strains identifies strain-specific metabolic capabilities linked to pathogenicity". In: *Proceedings of the National Academy of Sciences* 113.26 (2016), E3801–E3809.

[23] Emmanuel Boutet et al. "UniProtKB/Swiss-Prot: the manually annotated section of the UniProt KnowledgeBase". In: *Plant bioinformatics: methods and protocols*. Springer, 2007, pp. 89–112.

[24] Frank J Bruggeman et al. “Searching for principles of microbial physiology”. In: *FEMS microbiology reviews* 44.6 (2020), pp. 821–844.

[25] Elizabeth Brunk et al. “Recon3D enables a three-dimensional view of gene variation in human metabolism”. In: *Nature biotechnology* 36.3 (2018), pp. 272–281.

[26] Finja Büchel et al. “Path2Models: large-scale generation of computational models from biochemical pathway maps”. In: *BMC systems biology* 7 (2013), pp. 1–19.

[27] Carlos P Cantalapiedra et al. “eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale”. In: *Molecular biology and evolution* 38.12 (2021), pp. 5825–5829.

[28] Ron Caspi et al. “The MetaCyc database of metabolic pathways and enzymes-a 2019 update”. In: *Nucleic acids research* 48.D1 (2020), pp. D445–D453.

[29] Henry F Chambers. “The changing epidemiology of Staphylococcus aureus?” In: *Emerging infectious diseases* 7.2 (2001), p. 178.

[30] Siu HJ Chan et al. “Accelerating flux balance calculations in genome-scale metabolic models by localizing the application of loopless constraints”. In: *Bioinformatics* 34.24 (2018), pp. 4248–4255.

[31] Hester Chapman et al. “Algebraic differentiation for fast sensitivity analysis of optimal flux modes in metabolic models”. In: *Bioinformatics* 41.6 (2025), btaf287.

[32] Hester Chapman et al. “Optimization of metabolic fluxes”. In: *Economic Principles in Cell Biology*. No commercial publisher — Online open access book, 2025. DOI: 10.5281/zenodo.8156386. URL: https://doi.org/10.5281/zenodo.8156386.

[33] Linda Chiappalupi. “Development and validation of in vitro and in silico tests to predict compounds’ toxicity”. PhD thesis.

[34] Yasunori Chiba and Hiromi Akeboshi. “Glycan engineering and production of ‘humanized’glycoprotein in yeast cells”. In: *Biological and Pharmaceutical Bulletin* 32.5 (2009), pp. 786–795.

[35] Mette T Christiansen et al. “Genome-wide high-throughput screening to investigate essential genes involved in methicillin-resistant Staphylococcus aureus sequence type 398 survival”. In: *PLoS One* 9.2 (2014), e89018.

[36] Mark O Clements et al. “CtaA of Staphylococcus aureus is required for starvation survival, recovery, and cytochrome biosynthesis”. In: *Journal of bacteriology* 181.2 (1999), pp. 501–507.

[37] Matthew D Collins and DOROTHY Jones. “Distribution of isoprenoid quinone structural types in bacteria and their taxonomic implication”. In: *Microbiological reviews* 45.2 (1981), pp. 316–354.

[38] Jonas Contiero et al. “Effects of mutations in acetate metabolism on high-cell-density growth of Escherichia coli”. In: *Journal of Industrial Microbiology and Biotechnology* 24.6 (2000), pp. 421–430.

[39] G Ralph Corey et al. “Pooled analysis of single-dose oritavancin in the treatment of acute bacterial skin and skin-structure infections caused by Gram-positive pathogens, including a large patient subset with methicillin-resistant Staphylococcus aureus”. In: *International Journal of Antimicrobial Agents* 48.5 (2016), pp. 528–534.

[40] Nuno S Craveiro et al. “Drug withdrawal due to safety: a review of the data supporting withdrawal decision”. In: *Current drug safety* 15.1 (2020), pp. 4–12.

[41] David C Dahlin et al. “N-acetyl-p-benzoquinone imine: a cytochrome P-450-mediated oxidation product of acetaminophen.” In: *Proceedings of the National academy of Sciences* 81.5 (1984), pp. 1327–1331.

[42] Dan Davidi and Ron Milo. “Lessons on enzyme kinetics from quantitative proteomics”. In: *Current opinion in biotechnology* 46 (2017), pp. 81–89.

[43] DG Davidson and WN1943529 Eastham. “Acute liver necrosis following overdose of paracetamol”. In: *British medical journal* 2.5512 (1966), p. 497.

[44] Daan H De Groot et al. “The number of active metabolic pathways is bounded by the number of cellular constraints at maximal metabolic rates”. In: *PLOS Computational Biology* 15.3 (2019), e1006858.

[45] SL De Maeseneire et al. “Metabolic characterisation of E. coli citrate synthase and phosphoenolpyruvate carboxylase mutants in aerobic cultures”. In: *Biotechnology letters* 28 (2006), pp. 1945–1953.

[46] Marjan De Mey et al. “Minimizing acetate formation in E. coli fermentations”. In: *Journal of Industrial Microbiology and Biotechnology* 34.11 (2007), pp. 689–700.

[47] Erez Dekel and Uri Alon. “Optimality and evolutionary tuning of the expression level of a protein”. In: *Nature* 436.7050 (2005), pp. 588–592.

[48] Joseph A DiMasi, Henry G Grabowski, and Ronald W Hansen. “Innovation in the pharmaceutical industry: new estimates of R&D costs”. In: *Journal of health economics* 47 (2016), pp. 20–33.

[49] Joseph A DiMasi et al. “Trends in risks associated with new drug development: success rates for investigational drugs”. In: *Clinical Pharmacology & Therapeutics* 87.3 (2010), pp. 272–277.

[50] Michael A Dorato and Lorrene A Buckley. “Toxicology testing in drug discovery and development”. In: *Current protocols in toxicology* 31.1 (2007), pp. 19–1.

[51] Zhenjiao Du et al. “FusionESP: Improved Enzyme–Substrate Pair Prediction by Fusing Protein and Chemical Knowledge”. In: *Journal of Chemical Information and Modeling* 65.6 (2025), pp. 2806–2817.

[52] Michael A D’Elia, Mark P Pereira, and Eric D Brown. “Are essential genes really essential?” In: *Trends in microbiology* 17.10 (2009), pp. 433–438.

[53] Bettina Ebert, Michael Kisiela, and Edmund Maser. “Human DCXR–another ‘moonlighting protein’involved in sugar metabolism, carbonyl detoxification, cell adhesion and male fertility?” In: *Biological Reviews* 90.1 (2015), pp. 254–278.

[54] Jeremy S Edwards and Bernhard O Palsson. “Systems properties of the Haemophilus influenzaeRd metabolic genotype”. In: *Journal of Biological Chemistry* 274.25 (1999), pp. 17410–17416.

[55] Liam DH Elbourne et al. “TransAAP: an automated annotation pipeline for membrane transporter prediction in bacterial genomes”. In: *Microbial genomics* 9.1 (2023), p. 000927.

[56] Yang Feng et al. “Methane alleviates acetaminophen-induced liver injury by inhibiting inflammation, oxidative stress, endoplasmic reticulum stress, and apoptosis through the Nrf2/HO-1/NQO1 signaling pathway”. In: *Oxidative Medicine and Cellular Longevity* 2019.1 (2019), p. 7067619.

[57] Stephen S Fong et al. “Latent pathway activation and increased pathway capacity enable Escherichia coli adaptation to loss of key metabolic enzymes”. In: *Journal of Biological Chemistry* 281.12 (2006), pp. 8024–8033.

[58] James R Fuller et al. “Identification of a lactate-quinone oxidoreductase in Staphylococcus aureus that is essential for virulence”. In: *Frontiers in cellular and infection microbiology* 1 (2011), p. 19.

[59] Julien Gagneur and Steffen Klamt. “Computation of elementary modes: a unifying framework and the new binary approach”. In: *BMC bioinformatics* 5.1 (2004), pp. 1–21.

[60] Michael Y Galperin et al. “COG database update: focus on microbial diversity, model organisms, and widespread pathogens”. In: *Nucleic acids research* 49.D1 (2021), pp. D274–D281.

[61] Yuan Gao et al. “Proteomic analysis of acetaminophen-induced hepatotoxicity and identification of heme oxygenase 1 as a potential plasma biomarker of liver injury”. In: *PROTEOMICS–Clinical Applications* 11.1-2 (2017), p. 1600123.

[62] Rosmarie Gaupp et al. “Advantage of upregulation of succinate dehydrogenase in Staphylococcus aureus biofilms”. In: *Journal of bacteriology* 192.9 (2010), pp. 2385–2394.

[63] Thijs J Giezen et al. “Safety-related regulatory actions for biologicals approved in the United States and the European Union”. In: *Jama* 300.16 (2008), pp. 1887–1896.

[64] Christoph Gille et al. “HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology”. In: *Molecular systems biology* 6.1 (2010), p. 411.

[65] Haitske Graveland et al. “Methicillin resistant Staphylococcus aureus ST398 in veal calf farming: human MRSA carriage related with animal antimicrobial usage and farm hygiene”. In: *PloS one* 5.6 (2010), e10990.

[66] N Greenfieldboye. *EPA Chief Pledges to Severely Cut Back on Animal Testing of Chemicals. NPR Health News. September 10, 2019*. 2019.

[67] Melinda R Grosser et al. “Regulatory requirements for Staphylococcus aureus nitric oxide resistance”. In: *Journal of bacteriology* 198.15 (2016), pp. 2043–2055.

[68] Daniel H Haft et al. “TIGRFAMs and genome properties in 2013”. In: *Nucleic acids research* 41.D1 (2012), pp. D387–D395.

[69] William R Harcombe et al. “The ability of flux balance analysis to predict evolution of central metabolism scales with the initial distance to the optimum”. In: *PLoS computational biology* 9.6 (2013), e1003091.

[70] Janna Hastings et al. “ChEBI in 2016: Improved services and an expanding collection of metabolites”. In: *Nucleic acids research* 44.D1 (2016), pp. D1214–D1219.

[71] David Heckmann et al. “Kinetic profiling of metabolic specialists demonstrates stability and consistency of in vivo enzyme turnover numbers”. In: *Proceedings of the National Academy of Sciences* 117.37 (2020), pp. 23182–23190.

[72] Matthias Heinemann et al. “In silico genome-scale reconstruction and validation of the Staphylococcus aureus metabolic network”. In: *Biotechnology and bioengineering* 92.7 (2005), pp. 850–864.

[73] Reinhart Heinrich and Tom A Rapoport. “A linear steady-state treatment of enzymatic chains: general properties, control and effector strength”. In: *European journal of biochemistry* 42.1 (1974), pp. 89–95.

[74] David B Huang et al. “A phase II randomized, double-blind, multicenter study to evaluate efficacy and safety of intravenous iclaprim versus vancomycin for the treatment of nosocomial pneumonia suspected or confirmed to be due to Gram-positive pathogens”. In: *Clinical Therapeutics* 39.8 (2017), pp. 1706–1718.

[75] M Hussain, JGM Hastings, and PJ White. “A chemically defined medium for slime production by coagulase-negative staphylococci”. In: *Journal of medical microbiology* 34.3 (1991), pp. 143–147.

[76] Rafael U Ibarra, Jeremy S Edwards, and Bernhard O Palsson. “Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth”. In: *Nature* 420.6912 (2002), pp. 186–189.

[77] Kevin S Ikuta et al. “Global mortality associated with 33 bacterial pathogens in 2019: a systematic analysis for the Global Burden of Disease Study 2019”. In: *The Lancet* 400.10369 (2022), pp. 2221–2248.

[78] Takashi Iyanagi. “Molecular mechanism of phase I and phase II drug-metabolizing enzymes: implications for detoxification”. In: *International review of cytology* 260 (2007), pp. 35–112.

[79] Livnat Jerby, Tomer Shlomi, and Eytan Ruppin. “Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism”. In: *Molecular systems biology* 6.1 (2010), p. 401.

[80] M Patricia Jevons. ““Celbenin”-resistant staphylococci”. In: *British medical journal* 1.5219 (1961), p. 124.

[81] J Natalia Jiménez et al. “Livestock-associated methicillin-susceptible Staphylococcus aureus ST398 infection in woman, Colombia”. In: *Emerging infectious diseases* 17.10 (2011), p. 1970.

[82] H Kacser and JA Burns. “Rate control of biological processes”. In: *Symp. Soc. Exp. Biol.* Vol. 27. 1973, pp. 65–104.

[83] Axel von Kamp and Stefan Schuster. “Metatool 5.0: fast and flexible elementary modes analysis”. In: *Bioinformatics* 22.15 (2006), pp. 1930–1931.

[84] N Kaplowitz. “Physiological significance of glutathione S-transferases”. In: *American Journal of Physiology-Gastrointestinal and Liver Physiology* 239.6 (1980), G439–G444.

[85] Neil Kaplowitz. “Idiosyncratic drug hepatotoxicity”. In: *Nature reviews Drug discovery* 4.6 (2005), pp. 489–499.

[86] David W Kaufman et al. “Recent patterns of medication use in the ambulatory adult population of the United States: the Slone survey”. In: *Jama* 287.3 (2002), pp. 337–344.

[87] Hyun Uk Kim, Tae Yong Kim, and Sang Yup Lee. “Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen Acinetobacter baumannii AYE”. In: *Molecular BioSystems* 6.2 (2010), pp. 339–348.

[88] Zachary A King et al. “Escher: a web application for building, sharing, and embedding data-rich visualizations of biological pathways”. In: *PLoS computational biology* 11.8 (2015), e1004321.

[89] Steffen Klamt et al. “From elementary flux modes to elementary flux vectors: Metabolic pathway analysis with arbitrary linear flux constraints”. In: *PLoS computational biology* 13.4 (2017), e1005409.

[90] Miroslav Kratochvíl et al. “COBREXA 2: tidy and scalable construction of complex metabolic models”. In: *Bioinformatics* (Feb. 2025), btaf056. ISSN: 1367-4811. DOI: 10.1093/bioinformatics/btaf056. URL: https://doi.org/10.1093/bioinformatics/btaf056.

[91] Alexander Kroll, Sahasra Ranjan, and Martin J Lercher. “A multimodal Transformer Network for protein-small molecule interactions enhances predictions of kinase inhibition and enzyme-substrate relationships”. In: *PLOS Computational Biology* 20.5 (2024), e1012100.

[92] Alexander Kroll et al. “A general model to predict small molecule substrates of enzymes based on machine and deep learning”. In: *Nature communications* 14.1 (2023), p. 2787.

[93] Alexander Kroll et al. “Turnover number predictions for kinetically uncharacterized enzymes using machine and deep learning”. In: *Nature Communications* 14.1 (2023), p. 4139.

[94] Gemma C Langridge et al. “Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants”. In: *Genome research* 19.12 (2009), pp. 2308–2316.

[95] Sun-Kyung Lee, Hee-Jung Choi, Joohong Ahnn, et al. “Dicarbonyl/l-xylulose reductase (DCXR): The multifunctional pentosuria enzyme”. In: *The international journal of biochemistry & cell biology* 45.11 (2013), pp. 2563–2567.

[96] Nathan E Lewis et al. “Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models”. In: *Molecular systems biology* 6.1 (2010), p. 390.

[97] Albert P Li. “Accurate prediction of human drug toxicity: a major challenge in drug development”. In: *Chemico-biological interactions* 150.1 (2004), pp. 3–7.

[98] Feiran Li et al. “Deep learning-based k cat prediction enables improved enzyme-constrained model reconstruction”. In: *Nature Catalysis* 5.8 (2022), pp. 662–672.

[99] Wolfram Liebermeister. “The value structure of metabolic states”. In: *bioRxiv* (2018), p. 483891.

[100] Guigao Liu et al. “Engineering of Saccharomyces pastorianus old yellow enzyme 1 for the synthesis of pharmacologically active (S)-profen derivatives”. In: *Molecular Catalysis* 507 (2021), p. 111568.

[101] Christopher P Long et al. “Characterization of physiological responses to 22 gene knockouts in Escherichia coli central carbon metabolism”. In: *Metabolic engineering* 37 (2016), pp. 102–113.

[102] Gema Lozano Terol et al. “Engineering protein production by rationally choosing a carbon and nitrogen source using E. coli BL21 acetate metabolism knockout strains”. In: *Microbial cell factories* 18 (2019), pp. 1–19.

[103] ND Mahmood et al. “Amelioration of paracetamol-induced hepatotoxicity in rat by the administration of methanol extract of Muntingia calabura L. leaves”. In: *BioMed research international* 2014.1 (2014), p. 695678.

[104] Maxime Mahout, Ross P Carlson, and Sabine Peres. “Answer set programming for computing constraints-based elementary flux modes: application to escherichia coli core metabolism”. In: *Processes* 8.12 (2020), p. 1649.

[105] Adil Mardinoglu et al. “Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease”. In: *Nature communications* 5.1 (2014), p. 3083.

[106] Ameya A Mashruwala, Adriana van de Guchte, and Jeffrey M Boyd. “Impaired respiration elicits SrrAB-dependent programmed cell lysis and biofilm formation in Staphylococcus aureus”. In: *Elife* 6 (2017), e23845.

[107] Sonja Mayer et al. “The Staphylococcus aureus NuoL-like protein MpsA contributes to the generation of membrane potential”. In: *Journal of Bacteriology* 197.5 (2015), pp. 794–806.

[108] J MICHAEL McMILLIN. “Blood glucose”. In: *Clinical Methods: The History, Physical, and Laboratory Examinations. 3rd edition* (1990).

[109] Birgit H M Meldal et al. “The complex portal–an encyclopaedia of macromolecular complexes”. In: *Nucleic acids research* 43.Database issue (2015), D479—84. ISSN: 0305-1048. DOI: 10.1093/nar/gku975. URL: https://europepmc.org/articles/PMC4384031.

[110] Marija Milacic et al. “The reactome pathway knowledgebase 2024”. In: *Nucleic acids research* 52.D1 (2024), pp. D672–D678.

[111] Jaina Mistry et al. “Pfam: The protein families database in 2021”. In: *Nucleic acids research* 49.D1 (2021), pp. D412–D419.

[112] Jonathan M Monk et al. “i ML1515, a knowledgebase that computes Escherichia coli traits”. In: *Nature biotechnology* 35.10 (2017), pp. 904–908.

[113] Matteo Mori et al. “Constrained allocation flux balance analysis”. In: *PLoS computational biology* 12.6 (2016), e1004913.

[114] William Moses and Valentin Churavy. "Instead of rewriting foreign code for machine learning, automatically synthesize fast gradients". In: *Advances in neural information processing systems* 33 (2020), pp. 12472–12485.

[115] Anne Müller et al. "Synthesis of bifunctional azobenzene glycoconjugates for cysteine-based photosensitive cross-linking with bioactive peptides". In: *Chemistry–A European Journal* 21.39 (2015), pp. 13723–13731.

[116] Christopher JL Murray et al. "Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis". In: *The lancet* 399.10325 (2022), pp. 629–655.

[117] K Nakano et al. "Influence of acetic acid on the growth of Escherichia coli K12 during high-cell-density cultivation in a dialysis reactor". In: *Applied microbiology and biotechnology* 48 (1997), pp. 597–601.

[118] Zhiwei Nie et al. "OmniESI: A unified framework for enzyme-substrate interaction prediction with progressive conditional deep learning". In: *arXiv preprint arXiv:2506.17963* (2025).

[119] World Health Organization. *WHO bacterial priority pathogens list, 2024: bacterial pathogens of public health importance, to guide research, development, and strategies to prevent and control antimicrobial resistance*. World Health Organization, 2024.

[120] William O'Riordan et al. "A comparison of the efficacy and safety of intravenous followed by oral delafloxacin with vancomycin plus aztreonam for the treatment of acute bacterial skin and skin structure infections: a phase 3, multinational, double-blind, randomized study". In: *Clinical Infectious Diseases* 67.5 (2018), pp. 657–666.

[121] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. "What is flux balance analysis?" In: *Nature biotechnology* 28.3 (2010), pp. 245–248.

[122] Edward J O'Brien, Jose Utrilla, and Bernhard O Palsson. "Quantification and classification of E. coli proteome utilization and unused protein costs across environments". In: *PLoS computational biology* 12.6 (2016), e1004998.

[123] Jae Gu Pan, Joon Shick Rhee, and Jean M Lebeault. "Physiological constraints in increasing biomass concentration of Escherichia coli B in fed-batch culture". In: *Biotechnology Letters* 9 (1987), pp. 89–94.

[124] Annalisa Pantosti. "Methicillin-resistant Staphylococcus aureus associated with animals and its relevance to human health". In: *Frontiers in microbiology* 3 (2012), p. 127.

[125] Jong Myoung Park, Tae Yong Kim, and Sang Yup Lee. "Constraints-based genome-scale metabolic simulation for systems metabolic engineering". In: *Biotechnology advances* 27.6 (2009), pp. 979–988.

[126] Steven M Paul et al. “How to improve R&D productivity: the pharmaceutical industry’s grand challenge”. In: *Nature reviews Drug discovery* 9.3 (2010), pp. 203–214.

[127] Pablo Perel et al. “Comparison of treatment effects between animal experiments and clinical trials: systematic review”. In: *Bmj* 334.7586 (2007), p. 197.

[128] Mari A Piirainen, Heidi Salminen, and Alexander D Frey. “Production of galactosylated complex-type N-glycans in glycoengineered Saccharomyces cerevisiae”. In: *Applied Microbiology and Biotechnology* (2022), pp. 1–15.

[129] Eric JG Pollitt et al. “Staphylococcus aureus infection dynamics”. In: *PLoS pathogens* 14.6 (2018), e1007112.

[130] WZ Potter et al. “Acetaminophen-induced hepatic necrosis. III. Cytochrome P-450-mediated covalent binding in vitro”. In: *The Journal of pharmacology and experimental therapeutics* 187.1 (1973), pp. 203–210.

[131] Luana Presta et al. “Constraint-based modeling identifies new putative targets to fight colistin-resistant A. baumannii infections”. In: *Scientific reports* 7.1 (2017), p. 3706.

[132] Jean Matthieu Prot et al. “Integrated proteomic and transcriptomic investigation of the acetaminophen toxicity in liver microfluidic biochip”. In: *PloS one* 6.8 (2011), e21268.

[133] Charles H Rammelkamp and Thelma Maxon. “Resistance of Staphylococcus aureus to the action of penicillin.” In: *Proceedings of the Society for Experimental Biology and Medicine* 51.3 (1942), pp. 386–389.

[134] Pranay R Randad et al. “Comparison of livestock-associated and community-associated Staphylococcus aureus pathogenicity in a mouse model of skin and soft tissue infection”. In: *Scientific reports* 9.1 (2019), p. 6774.

[135] *RDKit: Open-Source Cheminformatics Software*. `https://www.rdkit.org`. Accessed: 2025-08-05.

[136] Alina Renz and Andreas Dräger. “Curating and comparing 114 strain-specific genome-scale metabolic models of Staphylococcus aureus”. In: *NPJ systems biology and applications* 7.1 (2021), p. 30.

[137] Jarrett Revels, Miles Lubin, and Theodore Papamarkou. “Forward-Mode Automatic Differentiation in Julia”. In: *arXiv preprint arXiv:1607.07892* (2016).

[138] Esteban Reynaga et al. “Prevalence of colonization by methicillin-resistant Staphylococcus aureus ST398 in pigs and pig farm workers in an area of Catalonia, Spain”. In: *BMC infectious diseases* 16.1 (2016), p. 716.

[139] Anthony R Richardson. “Virulence and metabolism”. In: *Microbiology spectrum* 7.2 (2019), pp. 10–1128.

[140] Anthony R Richardson, Paul M Dunman, and Ferric C Fang. “The nitrosative stress response of Staphylococcus aureus is required for resistance to innate immunity”. In: *Molecular microbiology* 61.4 (2006), pp. 927–939.

[141] Anthony R Richardson, Stephen J Libby, and Ferric C Fang. “A nitric oxide–inducible lactate dehydrogenase enables Staphylococcus aureus to resist innate immunity”. In: *Science* 319.5870 (2008), pp. 1672–1676.

[142] Lars Ruddigkeit et al. “Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17”. In: *Journal of chemical information and modeling* 52.11 (2012), pp. 2864–2875.

[143] Marat R Sadykov et al. “Inactivation of the Pta-AckA pathway causes cell death in Staphylococcus aureus”. In: *Journal of bacteriology* 195.13 (2013), pp. 3035–3044.

[144] Milton H Saier Jr et al. “The transporter classification database (TCDB): 2021 update”. In: *Nucleic acids research* 49.D1 (2021), pp. D461–D467.

[145] Benjamín J Sánchez et al. “Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints”. In: *Molecular systems biology* 13.8 (2017), p. 935.

[146] Nalini Schaduangrat et al. “Towards reproducible computational drug discovery”. In: *Journal of cheminformatics* 12.1 (2020), p. 9.

[147] Alexander Schmidt et al. “The quantitative and condition-dependent Escherichia coli proteome”. In: *Nature biotechnology* 34.1 (2016), pp. 104–110.

[148] Robert Schuetz et al. “Multidimensional optimality of microbial metabolism”. In: *Science* 336.6081 (2012), pp. 601–604.

[149] Ronny Schuster and Stefan 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”. In: *Bioinformatics* 9.1 (1993), pp. 79–85.

[150] Anastasia V Shindyapina et al. “Dietary methanol regulates human gene activity”. In: *PloS one* 9.7 (2014), e102837.

[151] Gunnar Sigurdsson et al. “A systems biology approach to drug targets in Pseudomonas aeruginosa biofilm”. In: *PLoS One* 7.4 (2012), e34337.

[152] Vanessa Silva et al. “Staphylococcus aureus and MRSA in livestock: antimicrobial resistance and genetic lineages”. In: *Microorganisms* 11.1 (2023), p. 124.

[153] Tara C Smith et al. “Methicillin-resistant Staphylococcus aureus (MRSA) strain ST398 is present in midwestern US swine and swine workers”. In: *Plos one* 4.1 (2009), e4258.

[154] Greg A Somerville and Richard A Proctor. “At the crossroads of bacterial metabolism and virulence factor synthesis in Staphylococci”. In: *Microbiology and Molecular Biology Reviews* 73.2 (2009), pp. 233–248.

[155] Nicole A Spahich et al. “S taphylococcus aureus lactate-and malate-quinone oxidoreductases contribute to nitric oxide resistance and virulence”. In: *Molecular microbiology* 100.5 (2016), pp. 759–773.

[156] M Stegger et al. “Genetic diversity in CC398 methicillin-resistant Staphylococcus aureus isolates of different geographical origin”. In: *Clinical microbiology and infection* 16.7 (2010), pp. 1017–1019.

[157] AE Stockland and CL San Clemente. “Multiple forms of lactate dehydrogenase in Staphylococcus aureus”. In: *Journal of bacteriology* 100.1 (1969), pp. 347–353.

[158] KC Strasters and KC Winkler. “Carbohydrate metabolism of Staphylococcus aureus”. In: *Microbiology* 33.2 (1963), pp. 213–229.

[159] Marco Terzer. “Large scale methods to enumerate extreme rays and elementary modes”. PhD thesis. ETH Zurich, 2009.

[160] Marco Terzer and Jörg Stelling. “Large-scale computation of elementary flux modes with bit pattern trees”. In: *Bioinformatics* 24.19 (2008), pp. 2229–2235.

[161] Ines Thiele and Bernhard Ø Palsson. “A protocol for generating a high-quality genome-scale metabolic reconstruction”. In: *Nature protocols* 5.1 (2010), pp. 93–121.

[162] Steven YC Tong et al. “Staphylococcus aureus infections: epidemiology, pathophysiology, clinical manifestations, and management”. In: *Clinical microbiology reviews* 28.3 (2015), pp. 603–661.

[163] Anne Troitzsch et al. “Carbon source-dependent reprogramming of anaerobic metabolism in Staphylococcus aureus”. In: *Journal of Bacteriology* 203.8 (2021), pp. 10–1128.

[164] E Tuboly et al. “Excessive alcohol consumption induces methane production in humans and rats”. In: *Scientific reports* 7.1 (2017), p. 7329.

[165] Pawel Tulinski et al. “Staphylococcus aureus ST398 gene expression profiling during ex vivo colonization of porcine nasal epithelium”. In: *BMC genomics* 15.1 (2014), p. 915.

[166] Zofia Tynecka et al. “Energy conservation in aerobically grown Staphylococcus aureus”. In: *Research in microbiology* 150.8 (1999), pp. 555–566.

[167] “UniProt: the Universal protein knowledgebase in 2025”. In: *Nucleic Acids Research* 53.D1 (2025), pp. D609–D617.

[168] Robert Urbanczik. “Enumerating constrained elementary flux vectors of metabolic networks”. In: *IET systems biology* 1.5 (2007), pp. 274–279.

[169] Anne-Sophie Valentin-Domelier et al. “Methicillin-susceptible ST398 Staphylococcus aureus responsible for bloodstream infections: an emerging human-adapted subclone?” In: *PLoS One* 6.12 (2011), e28369.

[170] K Vamshi Krishna and S Venkata Mohan. “Purification and characterization of NDH-2 protein and elucidating its role in extracellular electron transport and bioelectrogenic activity”. In: *Frontiers in microbiology* 10 (2019), p. 880.

[171] Gail A Van Norman. “Phase II trials in drug development and adaptive trial design”. In: *JACC: Basic to Translational Science* 4.3 (2019), pp. 428–437.

[172] Amit Varma and Bernhard O Palsson. “Metabolic flux balancing: basic concepts, scientific and practical use”. In: *Bio/technology* 12.10 (1994), pp. 994–998.

[173] Erwin Verkade et al. “Recent emergence of Staphylococcus aureus clonal complex 398 in human blood cultures”. In: (2012).

[174] Andreas Voss et al. “Methicillin-resistant Staphylococcus aureus in pig farming”. In: *Emerging infectious diseases* 11.12 (2005), p. 1965.

[175] YM Wang and J Van Eys. “The enzymatic defect in essential pentosuria”. In: *New England Journal of Medicine* 282.16 (1970), pp. 892–896.

[176] St E Wilken et al. “Interrogating the effect of enzyme kinetics on metabolism using differentiable constraint-based models”. In: *Metabolic engineering* 74 (2022), pp. 72–82.

[177] Hock Chuan Yeo et al. “Enzyme capacity-based genome scale modelling of CHO cells”. In: *Metabolic Engineering* 60 (2020), pp. 138–147.

[178] Sally W Yousief, Nader Abdelmalek, and Bianca Paglietti. “Optimizing phage-based mutant recovery and minimizing heat effect in the construction of transposon libraries in Staphylococcus aureus”. In: *Scientific Reports* 14.1 (2024), p. 22831.

[179] Panhong Yuan et al. “Dynamically regulating glucose uptake to reduce overflow metabolism with a quorum-sensing circuit for the efficient synthesis of d-pantothenic acid in Bacillus subtilis”. In: *ACS Synthetic Biology* 12.10 (2023), pp. 2983–2995.

[180] Jiao Zhao et al. “Effect of zwf gene knockout on the metabolism of Escherichia coli grown on glucose or acetate”. In: *Metabolic Engineering* 6.2 (2004), pp. 164–174.