KU LEUVEN
FACULTEIT INGENIEURSWETENSCHAPPEN

Improving ammonia secretion by *Bacillus subtilis* via model-driven genome network analysis and optimisation

Lucas De Vrieze

Thesis submitted for the degree of
Master of Science in
Chemical Engineering, option
Chemical and biochemical process
engineering

Thesis supervisors:
Prof. dr. K. Bernaerts
Prof. dr. S. Waldherr

Assessors:
Prof. dr. I. Smets
Prof. dr. K. Faust

Mentor:
K. Simoens

Academic year 2020 – 2021

Preface

“I bet that one day you will revolutionise the global ammonia business”, stated one of my friends three years ago. Such blatant bets – a favourite pastime of engineering students musing about their future in the bicycle shed after a long day of classes – are one of such memories that can pop up unasked in your mind when brainstorming about a possible research topic. Previous year, I had got the chance to be part of the iGEM team of the KU Leuven, who backed this idea far into our project selection process. Unluckily, the COVID crisis erupted and the team was forced to break up. I wish all members good luck and the ones who reengage this year a lot of success.

This brings me seamlessly to my promotors, Prof. dr. Kristel Bernaerts and Prof. dr. Steffen Waldherr. I already thanked them last year for their promises to back our project idea and for accepting this topic as my thesis subject, but I also want to show now my gratitude for the support they gave me this year. The final result has substantially improved because of their feedback during the regular meetings and their proofreading. Next, I would like to thank Kenneth Simoens, who mainly helped me during the experimental part of this thesis and was keen to share his great technical and practical know-how when I ran into doubts. I also want to thank my assessors Prof. dr. Ilse Smets and Prof. dr. Karoline Faust for critically reading my thesis.

Furthermore, I would like to acknowledge Stephen and Ruben for the casual conversations in the lab. You sometimes were the only ones of my fellow students I saw face-to-face in weeks this year. I was more at home than expected, whereby I would like to thank my family for the warm welcome and the moral support. Lastly, I am grateful for that pleasant group of friends that slowly was formed back at the Kulak five years ago. We definitely deserve a full reunion instead of our online chatserver.

Lucas De Vrieze

Contents

Abstract

Protein deamination is a relatively new bioprocessing concept which recovers fixed nitrogen from protein side streams. It aims to alleviate our reliance on the energy-intensive Haber-Bosch process and to reduce environmental problems caused by nitrogen losses. It relies on the microbial secretion of ammonia when processing nitrogenous substrates. *Bacillus subtilis* is deemed an attractive organism in this context because of its powerful pallet of proteases. A systematic screening of its metabolism using genome-scale metabolic modelling and flux balance analysis provides directions on which genetic modifications can improve its ammonia secretion rate in a protein medium.

First, after selecting a model from literature, the sensitivity of the secretion flux to the medium composition is assessed using dynamic flux balance analysis. It is found that only a couple of amino acids have a large impact on the final ammonium concentration, with special attention to glutamate, arginine and alanine. As a side effect, plant-based proteins have the greatest potential for high ammonia yields.

Then, for the case of LB medium, two metabolic optimisation algorithms, sequential scanning and OptGene, determine the best genetic deletions and/or amplifications in the carbohydrate and the amino acid subsystems for a couple of closely related optimisation objectives. It is observed that yield increases mainly come at the expense of the growth rate, but that acting on the arginine degradation pathway results in the best flux increases. To reduce the practical disadvantages of low growth rates, the idea of inducible knockouts using metabolic toggle switches is proposed.

Finally, some batch bioreactor runs provide data with which the model is adjusted for better qualitative predictions. A couple of metabolic shifts result in a smooth, gradually flattening growth curve. The main metabolic products are ammonium, isobutyrate, isovalerate and acetate, which is remetabolised. Hence, the implemented model alterations are scaling the exchange flux constraints according to the observed growth rate, enabling the reuptake of acetate, introducing a dissolved oxygen balance and adding new pathways to facilitate the secretion of isobutyrate and isovalerate. Due to an insufficient number of offline samples within each growth phase, a further quantitative refinement would entail a high uncertainty and is therefore omitted.

List of Figures

List of Tables

Nomenclature

2MB	2-Methylbutyrate
AA	Amino acid
AC	Acetate
ALA	L-Alanine
ARG	L-Arginine
ASN	L-Asparagine
ASP	L-Aspartate
BCAA	Branched chain amino acid
BCFA	Branched chain fatty acid
CAS	Casein hydrolysate
CYS	L-Cysteine
ddFBA	Dynamic optimisation approach dynamic flux balance analysis
dFBA	Dynamic flux balance analysis
DNA	Desoxyribonucleic acid
DOA dFBA	Dynamic optimisation approach dynamic flux balance analysis
DW	Dry weight
EX-...	Exchange reaction of ...
FBA	Flux balance analysis
GECKO	Genome network model with Enzymatic Constraints using Kinetic and Omics data
GLC	D-Glucose
GLN	L-Glutamine
GLU	L-Glutamate
GLY	Glycine
GSMN	Genome-scale metabolic network
HIS	L-Histidine
IB	Isobutyrate
ILE	L-Isoleucine
IV	Isovalerate
LEU	L-Leucine
LP	Linear programming
LYS	L-Lysine
MET	L-Methionine
MFA	Metabolic flux analysis
MILP	Mixed-integer linear programming

MOMA	Minimisation of metabolic adjustment
NLP	Non-linear programming
NRMSE	Normalised root mean square error
OSD	Oxygen saturation difference $C_{O_2}^{sat} - C_{O_2}$
pFBA	Parsimonious flux balance analysis
PHE	L-Phenylalanine
PPA	Propionic acid
PRO	L-Proline
PYR	Pyruvate
QP	Quadratic programming
rFBA	Regulatory Flux Balance Analysis
RNA	Ribonucleic acid
sdFBA	Static optimisation approach dynamic flux balance analysis
SER	L-Serine
SOA dFBA	Static optimisation approach dynamic flux balance analysis
TCA	Tricarboxylic acid
THR	L-Threonine
TRP	L-Tryptophan
TYR	L-Tyrosine
VAL	L-Valine

Chapter 1

Introduction

Ammonia is a commodity chemical that is indispensable for our modern life with a global production amounting up to 146 million tons in 2015 [1]. It is predominantly used to produce fertilisers and consequently sustains the global food production, but also fulfils needs in the nitrogen and inorganics industry to manufacture e.g. explosives and nitrogenous chemicals like nitric acid, acrylonitrile and hydrazine [2]. Next to that, newly emerging applications comprise functions as a refrigerant (R717) [2], as a nitrogen source for the bioconversion industry or as a novel fuel or energy carrier to compete with hydrogen gas [3].

Nowadays, the Haber-Bosch process covers most of the industrial production by fixating nitrogen from air, but it requires much energy to break the triple bond of nitrogen gas. Our reliance on ammonia is such that the process is responsible for an annual global energy consumption of $2.5 \cdot 10^6$ TJ, 1 % of the global CO_2 emissions and 3-5 % of the global natural gas demand [4]. Moreover, significant parts of the applied fertilisers are lost by run-off or volatilisation, resulting in NOx emissions and eutrophication [5]. To continue achieving adequately high agricultural yields and to keep on benefiting of its chemical versatility without a high environmental impact, ammonia production has to be rechallenged as a whole.

There is definitely a need to reduce the energy intensity of the ammonia production and to restrict losses by recovering fixated nitrogen from other sources, including waste streams. Much research is nowadays carried out in different fields to formulate a solution to this issue. One of these approaches is situated within the field of bioprocessing: **microbial deamination**. This is a relatively new concept which relies on the excretion of excess amino groups by microbes when metabolising nitrogenous substrates such as amino acids and proteins [1]. As the concept is rather new, research has focused on proof-of-concept genetic modification experiments lacking a wider view. This thesis, however, attempts to systematically screen an entire microbial metabolism for genetic modifications improving the ammonia yield, by applying genome-scale metabolic network modelling.

Chapter 2 provides more information on microbial deamination, describes the other most important ammonia production techniques and introduces basic and some more advanced computational metabolic modelling concepts. Using the data and the procedures clarified in Chapter 3, Chapter 4 assesses, remediates and selects some metabolic models reported in literature and explores the capabilities and the sensitivities of the chosen wild-type model regarding the composition of the bioreactor medium. Chapter 5 conducts a systematic network optimisation based on this model and reports the most promising genetic modifications. Chapter 6 reviews possible issues and improvements of this model and introduces some *in vitro* experiments of which the data are used to align its predictions qualitatively.

Chapter 2

Literature study

2.1 Ammonia production processes

In this thesis, the secreted ammonia is produced microbially by deaminating amino acids and proteins. This section defines the concepts of this process and provides an overview of the major production alternatives, both conventional and novel ones.

2.1.1 Microbial protein deamination

Microbes produce proteins from a variety of substrates. Thereby they often interconvert amino acids or extract their carbon backbones by trans- or deamination. Both reaction types involve handling an amino group, which is excreted as $NH_4{}^+$ if it is in excess. The processing of these amino groups can be intensified by providing solely amino acids or proteins as substrates [1]. An overview of the typical microbial deamination products is provided in Figure 2.1.

If the substrate compositions and the metabolic driving forces are managed well, this metabolism can be turned into an ammonia production process. However, thermodynamics and biological evolution constrain the rate at which free ammonia can be synthesised. Trans- and deamination reactions often are highly reversible reactions, which are biologically favoured towards amino acid anabolism instead of the desired catabolism. As a consequence, excreted free ammonia tends to be reassimilated quickly. Moreover, Darwinian evolution forces organisms to grow and reproduce as fast as possible, which spurs free ammonia assimilation [5, 6].

Anyhow, producing ammonia by deamination has some interesting properties. First, the production of ammonia from proteins is connected to biofuel production. The trans- and deamination pathways yield, next to the amino group, α-keto acids such as 2-oxoglutarate, oxaloacetate and pyruvate, which normally are routed to the TCA cycle to serve as carbon source. As mentioned, much trans- and deamination pathways obey to a thermodynamical equilibrium, but driving deaminations using heterologous pathways and/or throttling ammonia assimilation could force it to keep producing ammonia and α-keto acids. A heterologous Ehrlich pathway, for example, converts the latter into biofuels like ethanol, isobutanol and 3-methylbutanol, and drives the equilibria towards catabolism [5, 6, 7].

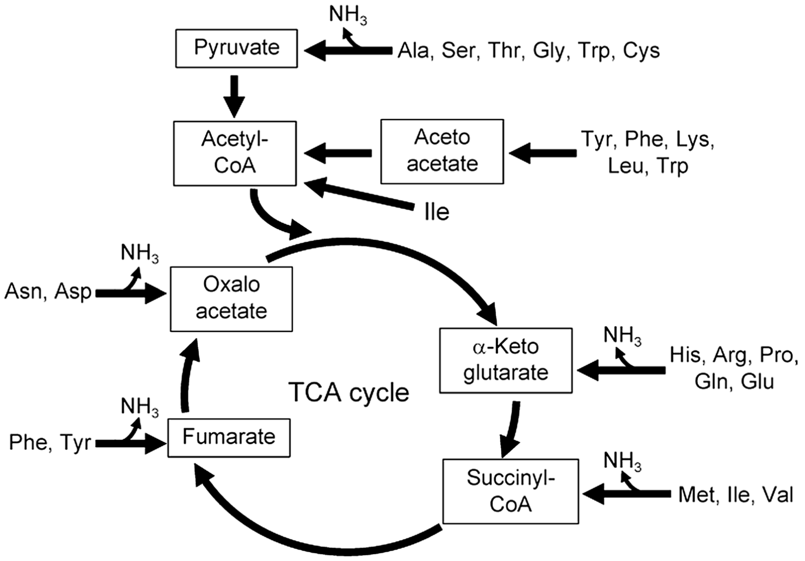


Figure 2.1: Overview of the typical microbial deamination products and their positions in the central carbon metabolism of *E. coli*. Figure taken from [1].

Another consequence regards the environmental impact of ammonia usage and its role in recycling nitrogen compounds. Ammonia is nowadays mainly used to produce fertilisers, which are indispensable to achieve high agricultural yields. However, producing ammonia is highly energy and capital intensive due to the Haber-Bosch process (see Section 2.1.2). Moreover, a considerable share of the applied fertilisers is lost by run-off and volatilisation, causing eutrophication and nitrogen oxide emissions respectively. Much energy and resources are thus invested in something that is only partly used and of which the losses cause environmental problems. As ammonia will remain an important commodity because of its newly emerging applications, it is necessary to reduce the energy intensity of fresh ammonia production and to recycle fixated nitrogen [5]. Possible solutions include curative ones like run-off water collection basins or canals connected to a treatment plant, and preventive ones like microbial protein production from waste nitrogen (sewers, waste proteins, waste biomass) to use as direct or animal feed [8]. Microbial deamination of waste proteins could be considered such a preventive measure.

Microbial protein deamination is a fairly new concept and has so far only been applied in *E. coli* [1, 6, 9] and *B. subtilis* [7], both because of their well-characterised metabolisms. However, the Gram-positive *B. subtilis* has an advantage over the Gram-negative *E. coli*: it is able to digest polypeptides using its own naturally secreted proteases. This avoids a more expensive protein hydrolysis pretreatment, simplifying industrial upscaling [7]. Moreover, the wild-type is able to grow with eight individual amino acids as the sole carbon source [10], while *E. coli* wild-types can only do this for four amino acids [6], indicating that *B. subtilis* has a richer amino acid metabolism.

2.1.2 Other ammonia production processes

Ammonia production processes mainly exist in two kinds depending on the source of the nitrogen. Firstly, there are the nitrogen fixation processes which convert gaseous N_2 from air into the more tangible ammonia. However, these processes require relatively large amounts of energy to break the triple bond of N_2. On the other hand, nitrogen recovery processes utilise feeds with already fixated nitrogen, avoiding breaking triple bonds. However, these feeds mostly are impure waste streams, hampering easy processing and consistent product quality.

Haber-Bosch The established ammonia production process combines N_2 from air with H_2, according to the chemical equilibrium reaction below.

$$N_2 + 3\,H_2 \rightleftharpoons 2\,NH_3$$

A trade-off between the thermodynamics (conversion) and the kinetics (reaction rate) has resulted in operating the reactor at high pressure (150-300 bar) and high temperature (>400 °C), applying large recycle streams and Fe- or Ru-based catalysts. The process itself stayed somewhat the same over the years, because most advances in energy economics and sustainability are done at the upstream H_2 generation facilities [11].

Traditionally, H_2 is generated by steam reforming natural gas to syngas or by gasifying coal. Current research includes producing H_2 by electrolysing water with wind or solar power [11]. The use of these renewable power sources imposes to develop small-scale ammonia plants for local markets, as this energy is not available anywhere at any time in any quantity [12]. Other recent research covers the gasification of lignocellulosic waste biomass into biosyngas. New approaches in this field include gasifying nitrogen-rich biomass, like waste proteins, resulting in additional NH_3 next to the biosyngas. Such applications could serve as a stand-alone ammonia production facility as well when equipped with the appropriate purification apparatus [4].

Electrochemical catalysis [13] Instead of driving an equilibrium reaction under harsh – and thus expensive – conditions, electrochemical ammonia production utilises electrical power at milder temperatures. It creates voltage gradients between an anode and a cathode over a ceramic or metallic electrolyte, being a liquid, a molten salt, a composite membrane or a solid. The nitrogen gas is sourced from air, while any compound from which hydrogen atoms can be extracted electrochemically (H_2, H_2O, CH_4, ethanol, ...), can supply the hydrogen either in a separate reaction unit or *in situ*.

However, current production rates are too low to be competitive. Next to conversion-rate trade-offs similar as for the Haber-Bosch process, electrochemical ammonia synthesis is plagued by low transport rates in the electrolytes and a high reaction competition by H^+-ions recombining into H_2.

Nitrogenase biocatalysis [14] Nature itself also has suitable machinery to synthesise ammonia from nitrogen gas, the so-called nitrogenase enzyme. It produces NH_3 in the absence of oxygen by converting nitrogen gas and protons into ammonia and hydrogen gas, as in the chemical reaction below.

$$\mathrm{N_2 + 8\,H^+ + 16\,ATP + 8\,e^- \longrightarrow 2\,NH_3 + H_2 + 16\,ADP + 16\,P}_i$$

So-called diazotrophic bacteria employ nitrogenases at ambient temperatures and pressures. These are free living species such as the *Azotobacter*, which are commonly found in soil. Others, like the *Rhizobium*, thrive in symbiotic relations with the root cells of legumes and are known as root nodule bacteria.

There are two approaches to apply nitrogenase catalysis for ammonia production: *in vitro* and *in vivo*. The *in vitro* way is to isolate the enzyme. Problems with this approach are the high costs of protein purification and the instability of the enzyme. This can be resolved by immobilisation and/or encapsulation, but the costs would then even rise further.

The *in vivo* way employs the microorganism itself. By meaningful metabolic engineering, the expression and activity of the microbial nitrogen fixing complexes could be ameliorated. This approach, however, is thwarted by the fact that this metabolism is not fully understood yet.

However, in both cases the production rate is limited due to nitrogenase itself, which is a rather bulky enzyme (molar mass: 230000 g/mol) with only two active sites. Using biotechnological techniques such as directed evolution and protein engineering, the number of active sites could be increased.

Slude digestate stripping [15] This process originates from the waste water treatment sector and, more specifically, from anaerobic digestion processes. In view of the economic performance of such units, the focus of these units is on the primary digestion product, high-value biogas, which can be further refined into methane. However, the liquid remnants of the waste stream, the digestate, contain a high amount of ammonium. The most efficient method nowadays to obtain this ammonia is by stripping the digestate with air to release ammonia gas, which is neutralised in an acid scrubber and crystallised to an ammonium salt.

However, the use of acids implies corrosion risks and increases the material costs. The counterions originating from the scrubbing acid, often have little value as well if the ammonium salt is not the desired product (fertilisers). These salts can even be dangerously reactive, cfr. ammonium nitrates.

2.2 Metabolic modelling tools

Modelling in general has some clear advantages over hands-on experimental tryouts. Instead of wasting time and resources in large experiment series, an engineer tries to extract a mathematical model of a certain process or phenomenon from some well-chosen experimental setups. This model immediately serves as a framework to predict the response of the system to a new situation, and as a consequence, allows to optimise it with respect to some desired criteria.

Modelling bioprocesses has evolved over the years and is rising in complexity more and more to capture all aspects of nature's own production unit: the cell. Coming from simple kinetic models only valid for a specific range of conditions, bioprocess modelling has reached the genomics scale and all reactions that a cell is capable to carry out. In the future, this rise in complexity probably will continue and models will include other omics scales as well.

These biological models serve different purposes in all development stages of a production process. In the experimental stage for example, they are used for characterising and optimising the behaviour of the micro-organism, by changing conditions such as temperature, pH, substrate feed profile, oxygen transfer, reactor operation mode etc. At larger scales, these models enable process monitoring and control by adjusting the appropriate heat and material streams. They also facilitate process troubleshooting by predicting the consequences of possible corrective actions.

This section will explain the basic concepts of genome-scale metabolic network modelling and some network optimisation strategies. A novel way of implementing proteomic data will be discussed as well, next to modelling genetic modifications.

2.2.1 Metabolic network models

A cell maintains several chemical reactions constituting its metabolism. The classical way of writing these down, is in a chemical reaction equation, indicating which reagentia are converted to which products in which ratios. However, a metabolism can consist of hundreds of reactions. That is why these reactions are ordered in a mathematical graph, as for example in Figure 2.2. Metabolites are drawn as the nodes of the network, while reactions are the arrows. Biological evolution has selected a couple of patterns, as can be seen in their frequent appearance in metabolic networks. These patterns are the subsystems which are responsible for specific metabolic functions, e.g. respiration, breakdown of sugars, synthesis of biomass etc.

As the synthesis of macromolecules like DNA, a protein or the cell wall, is not always completely understood, such reactions often are included in a lumped form, i.e. a black-box reaction [16].

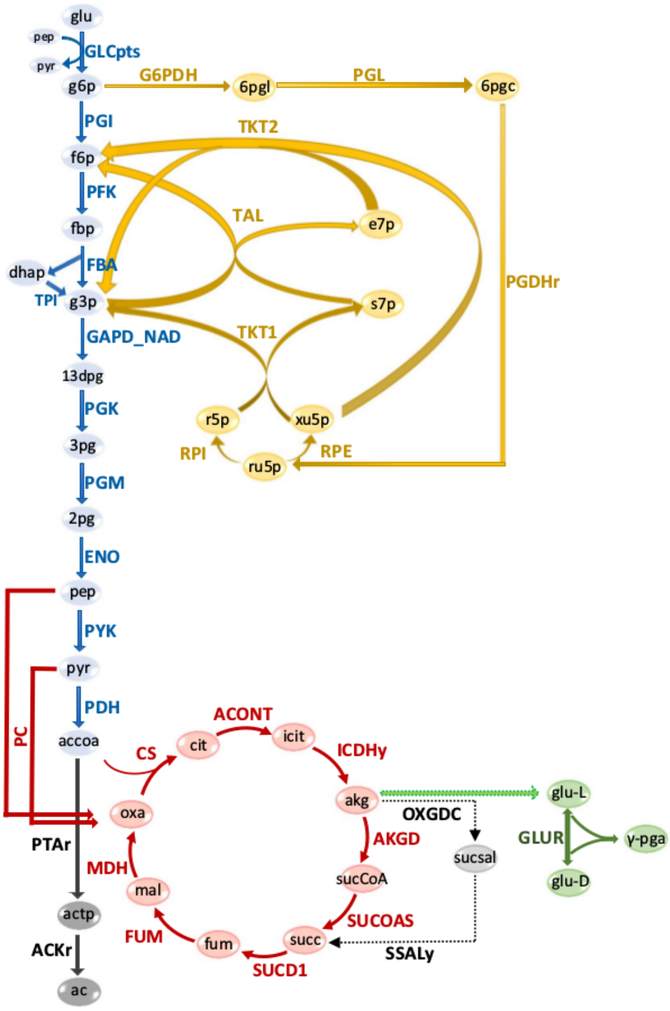


Figure 2.2: Example network map of the metabolic model for the central carbon and γ-polyglutamic acid (γPGA) subsystems of *B. subtilis*. Figure taken from [17].

A metabolic network can also be mathematically represented by a stoichiometric matrix S of size $m \times r$, in which m is the number of metabolites in the network and r the number of reactions [16]. By writing a mass balance for each metabolite in the network, one obtains a linear differential system which can be written in matrix form using S. The mass balance for a metabolite x_i is as following in equation (2.1), with v_{ij} the flux of metabolite i in reaction j and S_{ij} the stoichiometric coefficient of that metabolite in that reaction. It is assumed, however, that there are no external feeds or effluents and consequently, this is the mass balance for batch systems [16].

$$\frac{dx_i}{dt} = \sum_{j=1}^{r} S_{ij} v_{ij} = \begin{pmatrix} S_{i1} & S_{i2} & ... & S_{ir} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ ... \\ v_r \end{pmatrix} \tag{2.1}$$

Assembling the mass balances for all m metabolites gives the general matrix form in equation (2.2)

$$\frac{d\boldsymbol{x}}{dt} = \begin{pmatrix} S_{11} & S_{12} & ... & S_{1r} \\ S_{21} & S_{22} & ... & S_{2r} \\ ... & ... & ... & ... \\ S_{m1} & S_{m2} & ... & S_{mr} \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ ... \\ v_r \end{pmatrix} = \boldsymbol{S} \cdot \boldsymbol{v} \tag{2.2}$$

Note that most metabolites are involved in only a few reactions and that consequently most elements of S are 0, rendering S a sparse matrix.

By including exchange pseudoreactions and an extracellular counterpart for the metabolites that the cell can exchange with its environment, an extended form of S is obtained. The part that only takes the internal metabolites and reactions into account, is now denoted as S_{int}. This extended matrix has some clearly distinguishable compartments, as depicted in Figure 2.3 [16].

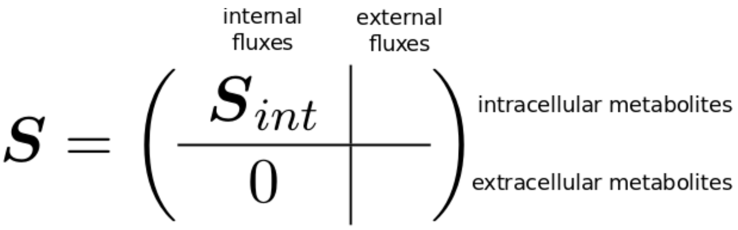


Figure 2.3: Compartments in an extended stoichiometric matrix

2.2.2 Model analysis tools

Metabolic Flux Analysis (MFA)

MFA is one of the oldest and simplest analysis methods and only relies on the metabolic network by solving the system in equation (2.2) in a slightly different form [18]. The difference is the additional term accounting for the biomass growth μ [1/h], which has a diluting effect, as stated in equation (2.3).

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{S} \cdot \boldsymbol{v} - \mu \boldsymbol{x} \tag{2.3}$$

$\boldsymbol{x}$ is the vector of metabolite concentrations in mmol/g_{DW}, $\boldsymbol{S}$ the stoichiometric matrix as defined earlier, and $\boldsymbol{v}$ a vector containing the metabolite fluxes in mmol/(g_{DW}h). The dilution term is often small and omitted. The fluxes $\boldsymbol{v}$ depend on the concentration of the involved metabolites via their reaction kinetics and, as a result, equation (2.2) is a first order system of m coupled differential equations in x.

To reduce calculation efforts – especially in case of large metabolic models –, a simple approximation is called in: the pseudo-steady state approximation. It assumes that the dynamics of the metabolism are much faster than those of the control mechanisms of the cell (gene expression level adjustments etc.). As a consequence, the differential term is neglected, resulting in equation (2.4).

$$0 = \boldsymbol{S} \cdot \boldsymbol{v} \tag{2.4}$$

This is a linear algebraic system with m equations for r variables. Usually, there are more reactions than metabolites ($r > m$) and the system is thus underdetermined with $r - m$ degrees of freedom, coupled to a solution space of dimension $r - m$. The remaining degrees of freedom are reduced by a number of experimental flux measurements a. This turns the system either determined and thus solvable ($a = r - m$) or overdetermined ($a > r - m$). In the latter case, the solution is obtained by minimising the sum of squared errors between the model and the experimental values [19].

Alternative constrains can be imposed via ^{13}C-MFA, which utilises ^{13}C-labelled substrates to incorporate marked C-atoms in the intracellular metabolites. Via mass spectometry measurements, a ratio of the corresponding reaction fluxes is determined and included as an additional equation in the MFA equation system [20].

The advantage of MFA is its simplicity, as it only relies on linear algebra to solve the system. However, the larger the network, the more measurements are required to make the system at least determined.

Flux Balance Analysis (FBA)

Contrary to MFA, FBA is able to solve underdetermined systems. The two enabling key features of FBA are the flux constraints and the objective function.
FBA treats the same system as MFA (equation (2.4)), exploiting the pseudo-steady state approximation too. However, the solution space is now restricted by *flux constraints*, which are written in a general form in equation (2.5) [21].

$$\boldsymbol{v_{min}} \leq \boldsymbol{v} \leq \boldsymbol{v_{max}} \tag{2.5}$$

Non-limiting reactions are constrained by an arbitrary high (or low for $v_{i,min}$) boundary, e.g. 1000 mmol/(g_{DW}h). Thermodynamically irreversible reactions have a lower boundary at zero, while directly measurable fluxes like the exchange reactions are constrained by experimental values if available. Constraining upper and lower bounds of a flux to zero mimics a genetic knock-out, while setting a positive lower bound forces flux through black-box reactions, such as the biomass formation reaction or the 'ATP maintenance reaction'. This last reaction is often added to a model to include non-growth related energy consumption [21].
On the other hand, the *objective function Z* eliminates the degrees of freedom that were not covered by the systems equations. It relates some criteria set by the modeller to the fluxes within the solution subspace via a vector of weights c, as stated in equation (2.6). The final solution is then acquired by maximising the objective function. A graphical summary of the FBA procedure is shown in Figure 2.4.

$$Z = \boldsymbol{c}^T \boldsymbol{v} \tag{2.6}$$

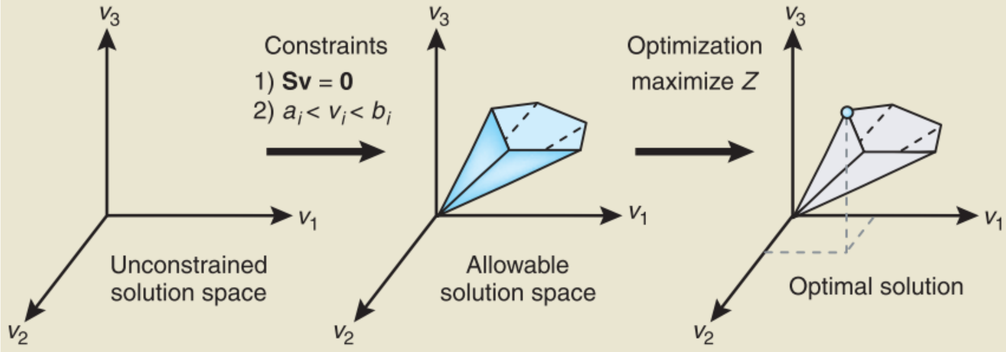


Figure 2.4: Graphical summary of the FBA method. Starting from an underdetermined solution space with three dimensions (*left*), the solution space is reduced by imposing the system equations and the flux constraints (*center*). Finally, an LP solver determines the optimal solution with respect to a certain objective function (*right*). Figure taken from [21].

There are many biologically meaningful objective functions. In favourable circumstances, for example, most organisms tend to maximise their biomass formation due to their evolutionarily induced striving for survival. To implement this in an FBA model, $\boldsymbol{c}$ is an all-but-one vector of zeros with a single 1 on the position of the biomass formation flux in $\boldsymbol{v}$. Other objective functions include maximisation of ATP generation, minimisation of carbon usage, maximisation of oxygen usage etc.

As the system equations, its constraints and the objective function are all linear, the resulting problem is a so-called linear programming problem (LP), as defined in equation (2.7). Such problems are frequently encountered in many fields including economics, industrial operational research and mathematics [22].

$$\begin{aligned} \max\ & Z = \boldsymbol{c}^T \boldsymbol{v} \\ \text{s.t.}\ & \boldsymbol{S} \cdot \boldsymbol{v} = 0 \\ & \boldsymbol{v_{min}} \leq \boldsymbol{v} \leq \boldsymbol{v_{max}} \end{aligned} \tag{2.7}$$

The advantages of FBA are clear: no kinetic parameters required, quick computations and versatile model tweaking. However, its advantages have a downside too. As kinetic parameters are not included in the system equations, it is impossible to directly relate metabolite fluxes v to metabolite concentrations x. Consequently, FBA cannot predict concentrations, but only fluxes. Moreover, they have been computed while assuming pseudo-steady state. Thus, they are only valid during exponential growth phases with a constant growth rate. Other relevant biological phenomena such as gene regulation and the size and activity of the enzyme pool, are not taken into account either.

A last point of attention is the relation between model redundancy and the LP solver. The larger the metabolic network, the more chance that two distinct flux distributions perform equally according to the objective function. A usual LP solver algorithm returns the first optimal value it encounters, starting from an initial guess. To tackle this issue, one should apply a Mixed Integer Linear Programming (MILP) algorithm, which searches all optimal values [21, 23].

Parsimonious Flux Balance Analysis (pFBA)

Over the years, it was often observed that the fastest growing micro-organisms express few enzymes that are not used for their metabolism. Except for some stand-by enzymes mediating sudden environmental changes, a micro-organism appears to minimise the total amount of enzymes, resulting in less maintenance efforts to keep the enzyme abundances at appropriate levels.
The consequence of this phenomenon is that micro-organisms tend to obey to another minimalist flux solution than predicted by FBA. A way to implement this feature, is to minimise the sum of all fluxes v obtained from the preceding FBA problem while fixing the objective function Z at its optimum Z_{opt}. To account for negative fluxes, either the absolute value or the squared value is taken for the calculations. This is translated in the mathematical problem in equation (2.8) [24].

$$
\begin{aligned}
\max\ & Z = \boldsymbol{c}^T \boldsymbol{v} \\
\text{s.t.}\ & S\boldsymbol{v} = 0 \\
& \boldsymbol{v_{min}} \leq \boldsymbol{v} \leq \boldsymbol{v_{max}} \\
\text{Then}\ & \\
\min\ & \sum_i v_i^2 \text{ or } \sum_i |v_i| \\
\text{s.t.}\ & Z = Z_{opt} \\
& S\boldsymbol{v} = 0 \\
& \boldsymbol{v_{min}} \leq \boldsymbol{v} \leq \boldsymbol{v_{max}}
\end{aligned}
\tag{2.8}
$$

It is clear that the computational work increases in comparison with FBA. Also, when using squared values, a quadratic programming solver (QP) is required.

Regulatory Flux Balance Analysis (rFBA)

As mentioned earlier, FBA does not include any regulatory information. However, biological evolution has equipped cells with multi-level regulatory mechanisms to be able to tweak certain fluxes or subsystems depending on the experienced conditions. FBA tends to metabolise everything suitable instead of only substrates that are allowed by these regulation mechanisms. This artefact is attributable to the FBA optimisation procedure. For example, when maximising biomass formation, FBA will utilise all metabolic pathways that result in an increase of the objective value.

Regulatory interactions can be included in a qualitative way by using Boolean logic [25]. By the pseudo-steady state approximation, it is assumed that the network behaves equally over time, including the regulatory network. This determines which reactions of the model should be throttled, which is implemented by multiplying the existing flux constraints with a Boolean expression. If for example, a reaction v_i can only take place if the necessary enzyme E_i is present and if its inhibitor I_i is absent, its flux constraint would be stated as in equation (2.9).

$$v_{i,min} \cdot (\mathrm{E}_i \text{ and } \mathrm{not}(\mathrm{I}_i)) \leq v_i \leq v_{i,max} \cdot (\mathrm{E}_i \text{ and } \mathrm{not}(\mathrm{I}_i)) \tag{2.9}$$

The state of the regulatory network of the situation at hand is determined separately by building a static probabilistic network which is validated using omics techniques, after which it is implemented by constraining the appropriate reactions [26]. Finally, an ordinary FBA is performed for the throttled metabolic model [25].
The impact of these regulations is not always recognisable in the simulation results, as visualised in Figure 2.5.

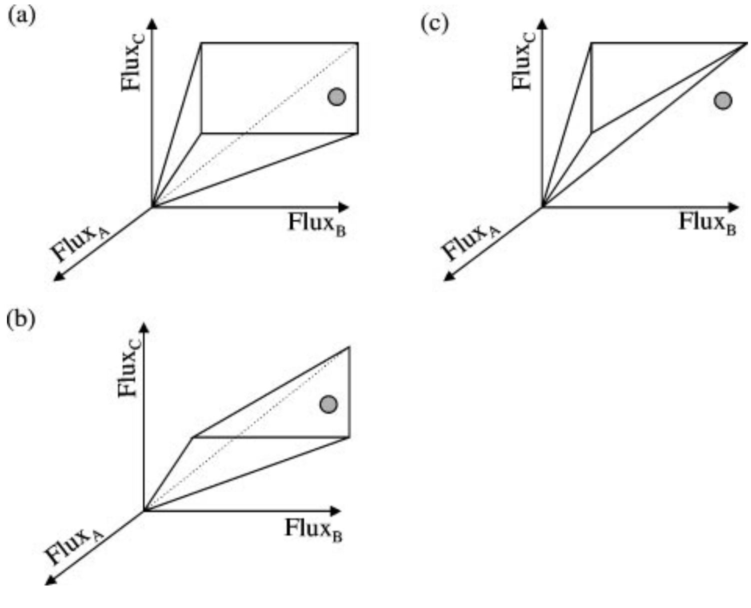


Figure 2.5: Graphical representation of the impact of regulatory constraints on the solution subspace. A hypothetical solution subspace with optimal solution is shown in figure a). In figure b), the optimal solution still fulfils the new regulatory constraint. In case it does not, like in figure c), a new optimal solution will be obtained. Figure taken from [25].

Dynamic Flux Balance Analysis (dFBA)

As mentioned in Section 2.2.2, FBA cannot directly predict metabolite concentrations. dFBA circumvents this issue by integrating FBA flux distributions in the mass balance equations of the extracellular metabolites. The results are metabolite concentration and biomass time profiles, from which shifts in the metabolism due to the availability of metabolites can be captured.

The governing mass balance equations of FBA problems are valid for the biomass itself taking abstraction of the bioreactor environment and are, hence, normalised by the biomass (mind the units of equation (2.3)). To obtain the extracellular metabolite concentrations $\boldsymbol{c}$ [mmol/L], the exchange fluxes [mmol/(g$_{DW}$h)] thus have to be multiplied with the biomass X [g$_{DW}$/L], resulting in equation (2.10) [27].

$$\frac{d\boldsymbol{c}}{dt} = \boldsymbol{S}\boldsymbol{v}X \tag{2.10}$$

Extracellular metabolites c_i are produced or consumed until depleted. The latter case must be implemented in such a way that negative concentrations are impossible. Besides, most organisms consume a substrate less if it is scarce. There are some methods to implement such additional constraints for all K extracellular metabolites. A common one are Michaelis-Menten uptake kinetics, which is written down mathematically in equation (2.11).

$$v_i \leq \frac{v_{max,i} c_i}{K_{M,i} + c_i} \text{ for all } i\text{=1, ..., } K \tag{2.11}$$

$K_{M,i}$ represents the concentration of c_i at which its uptake rate v_i is half its maximal rate $v_{max,i}$. For concentrations much lower than $K_{M,i}$, c_i can be neglected in the denominator and hence, the uptake flux scales approximately linearly with c_i. For much higher concentrations, the constraint approximately reduces to $v_{max,i}$.
The disadvantage of this method is that it introduces two (or one in case of extreme c_i) parameters per considered exchange flux, which counteracts the parameterless flexibility of FBA. Moreover, these parameters must be measured experimentally, introducing some uncertainty errors. Besides, they are not always independently measurable, as exchange fluxes might be metabolically coupled, especially in case of multiple substrates.

A simple way to avoid this, was proposed by Varma and Palsson [28]. Here, the uptake rate constraint is defined as the flux v_i that can be consumed by the present biomass X within one simulation time unit Δt, i.e. the rate v_i at which the current metabolite concentration c_i is completely depleted right at the end of a time period Δt. This can be mathematically derived from an elementary mass balance as presented in equation (2.12).

$$\left.\frac{dc_i}{dt}\right|_{max} \approx \frac{\Delta c_{max,i}}{\Delta t} \geq v_i \cdot X \xLeftrightarrow{\Delta c_{max,i}=c_i} v_i \leq \frac{c_i}{X \cdot \Delta t} \tag{2.12}$$

This type of constraint is not as invasive as Michaelis-Menten constraints, but only intervenes when the concentrations are that low that they might turn negative within the simulation time period at hand. The disadvantage is that it is a generic constraint, neglecting the often specialised uptake behaviour of cells.

Another important aspect is the way of coupling FBA to the time-dependant mass balances. There are two main approaches: the static one and the dynamic one.

Static optimisation approach dFBA (sdFBA, SOA dFBA) discretises the time derivative and solves an FBA problem (or a derived procedure like pFBA) for each time point. The results are either numerically integrated by the forward Euler method or analytically by solving the differential equation for the time step at hand. In the first case, the governing equations are discretised in equation (2.13), with Δt the discretisation time period so that $t_{i+1} = t_i + \Delta t$ [27].

$$\boldsymbol{c}(t_{i+1}) = \boldsymbol{c}(t_i) + \boldsymbol{S}\boldsymbol{v}(t_i)X(t_i)\Delta t \tag{2.13}$$

In the second case, an expression for $X(t)$ can be easily obtained by integrating the biomass balance (2.14) by separation of variables, assuming a constant μ over the discretisation interval $[t_i, t_{i+1}]$, i.e. the pseudo-steady state approximation. Note that equation (2.14) is a specific case of equation (2.10) selecting the biomass formation reaction of $\boldsymbol{S}\boldsymbol{v}$ ($\boldsymbol{S}\boldsymbol{v}|_X = \mu$).

$$\frac{dX}{dt} = \mu X \iff X(t) = X_0 e^{\mu t} \implies X(t_{i+1}) = X(t_i)e^{\mu \Delta t} \tag{2.14}$$

By introducing this expression into equation (2.10) and integrating it over time, the following iteration formula is obtained [28]. And indeed, when rewriting the exponential term as a first order Maclaurin series, equation (2.15) reduces to (2.13). The `dynamicFBA` COBRA toolbox function utilises this analytical approach of sdFBA.

$$\boldsymbol{c}(t_{i+1}) = \boldsymbol{c}(t_i) - \frac{\boldsymbol{S}\boldsymbol{v}(t_i)}{\mu(t_i)} X(t_i)\left(1 - e^{\mu \Delta t}\right) \tag{2.15}$$

sdFBA is a sequential, trajectory-dependant procedure. First, an FBA optimisation gives the initial flux distribution $\boldsymbol{v}(t_i)$. The exchange flux constraints are obtained from expressions like equations (2.11) and (2.12), and depend on the provided initial concentration of the extracellular metabolites $\boldsymbol{c}(t_i)$. The resulting flux distribution is then inserted in equation (2.13) or (2.15) to obtain the metabolite concentrations of the next time point $c(t_{i+1})$. The flux constraints are updated based on these new metabolite concentrations and a new FBA optimisation is conducted to get the concentrations of a third time point, as graphically depicted in Figure 2.6. This loop continues until a given time period has passed [27, 28]. To speed up the computations, the time step Δt may vary if nothing indicates a near metabolic shift [29].

This combination of integrating an ODE system and FBA is a rather intuitive and computationally economical method, as the underlying problems are a sequence of linear FBA problems. It is flexible as well: the use of adaptable uptake flux constraints still allows to implement other (internal) constraints and/or features into the problem without affecting the solving procedure. However, it comes at a cost. If the time period is chosen too wide, the numerical solutions can be inaccurate because possible metabolic shifts are captured too late. As the modelled microbes often grow exponentially, so would do an inaccuracy, implying that these systems are stiff [30].

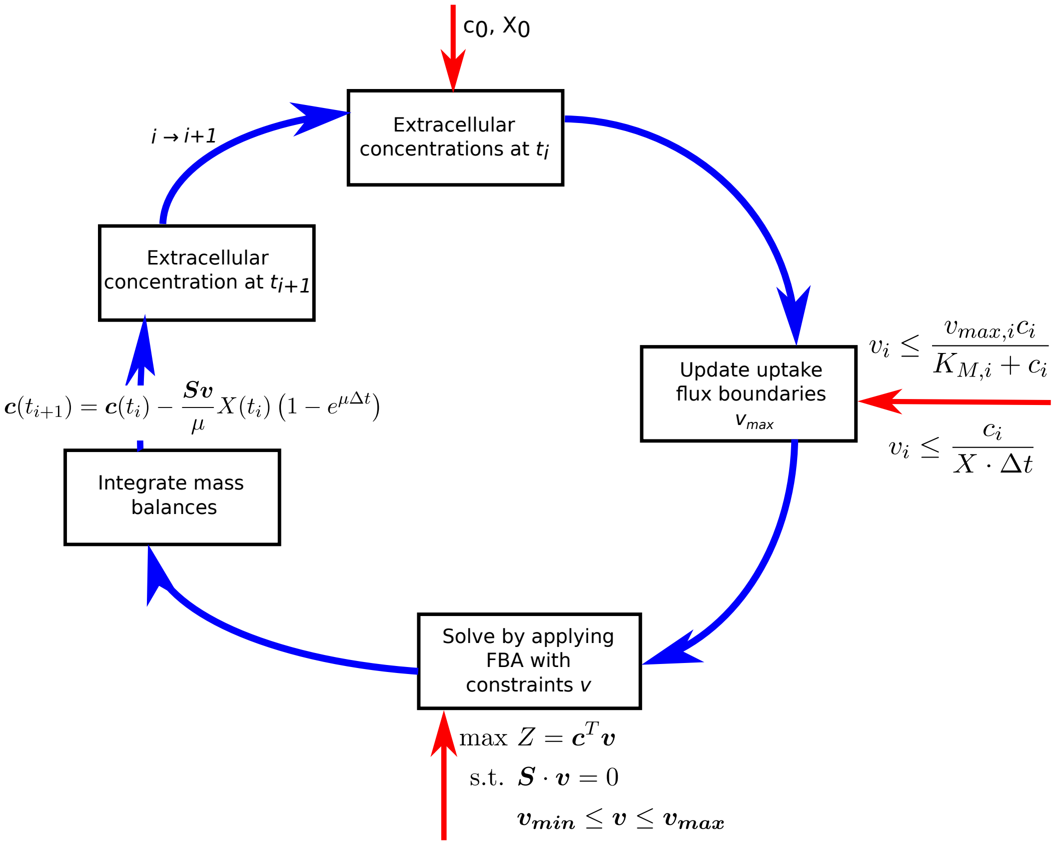


Figure 2.6: Graphical summary of the static approach dFBA procedure. User inputs are depicted with red arrows. Figure inspired on [27].

Dynamic optimisation approach dFBA (ddFBA, DOA dFBA) optimises the flux distributions over the entire time period at once. The general problem, as stated in equation (2.16), is a rather complex non-linear optimisation problem (NLP).

$$\begin{aligned}
&\max\ \hat{w}_{end}\Psi(\boldsymbol{z},\boldsymbol{v},X)|_{t=t_f} + \hat{w}_{ins}\sum_{j=0}^{M}\left(\int_{t_0}^{t_f} L(\boldsymbol{z},\boldsymbol{v},X(t))\delta(t-t_j)dt\right)\\
&\text{s.t. } \frac{d\boldsymbol{c}}{dt} = \boldsymbol{S}\boldsymbol{v}X \qquad\qquad (2.16)\\
&\qquad t_j = t_0 + j\frac{t_f-t_0}{M},\ j\in[0,M]\cap\mathbb{N}\\
&\qquad \boldsymbol{C}(c(t),v(t))\le 0,\ c(t)\ge 0,\ \forall t\in[t_0,t_f]
\end{aligned}$$

The objective function contains two terms: one around the terminal objective function Ψ and one around the instantaneous objective function L. The terminal one optimises Ψ at the final time point t_f, while the instantaneous one optimises L for one particular time point t_j. $\hat{w}_{end}$ and $\hat{w}_{ins}$ are the weights that are assigned to each part respectively. δ is the Dirac-delta function and M the number of time intervals in which the considered time horizon between t_0 and t_f is divided. $\boldsymbol{C}$ is an unspecified function that sets additional constraints on c and v if required, such as uptake kinetics. Optimising equation (2.16) happens by parametrising it at the roots of orthogonal polynomials (e.g. Legendre polynomials) within a finite elements framework [27].

Mahadevan et al. propose two optimisation functions, a terminal one and a instantaneous one. The latter maximises the scaled sum of biomass concentrations at the N_s roots of the polynomials. This scaling is necessary to treat the increasing biomass concentrations equally over time. μ^* is the experimentally measured growth rate that scales the biomass concentrations by an exponential profile [27].

$$\max \sum_{i=1}^{N_s}\frac{X(i)}{X_0e^{\mu^* t_i}} \qquad (2.17)$$

The second one is a terminal one and is very intuitive: maximising the final biomass.

$$\max\ X(t_f) \qquad (2.18)$$

Although dynamic approach dFBA accounts for the entire considered time period at once for better accuracy, it is computationally more demanding due to the necessarily fine time discretisation and its non-linear nature. Therefore, its uses were limited to small metabolic models, until Gomez et al. developed a new simulation framework able to handle large models in a fast and reliable way [30].

2.2.3 Integration of omics data

So far, modelling metabolism involved only the basic metabolic network structure. The sole exception was rFBA, in which a Boolean layer was added to the model. However, the other levels of the complex cell machinery should not be neglected too easily as these can influence all layers below. This multi-level system and its interconnectivity are illustrated in Figure 2.7.
This subsection presents a fairly novel approach integrating experimental proteomic data into the model. Attempts to integrate omics data are not limited to the one described below, but are far more numerous, illustrating the vastness of the rapidly developing scenery of omics analysis [26, 31, 32, 33, 34].

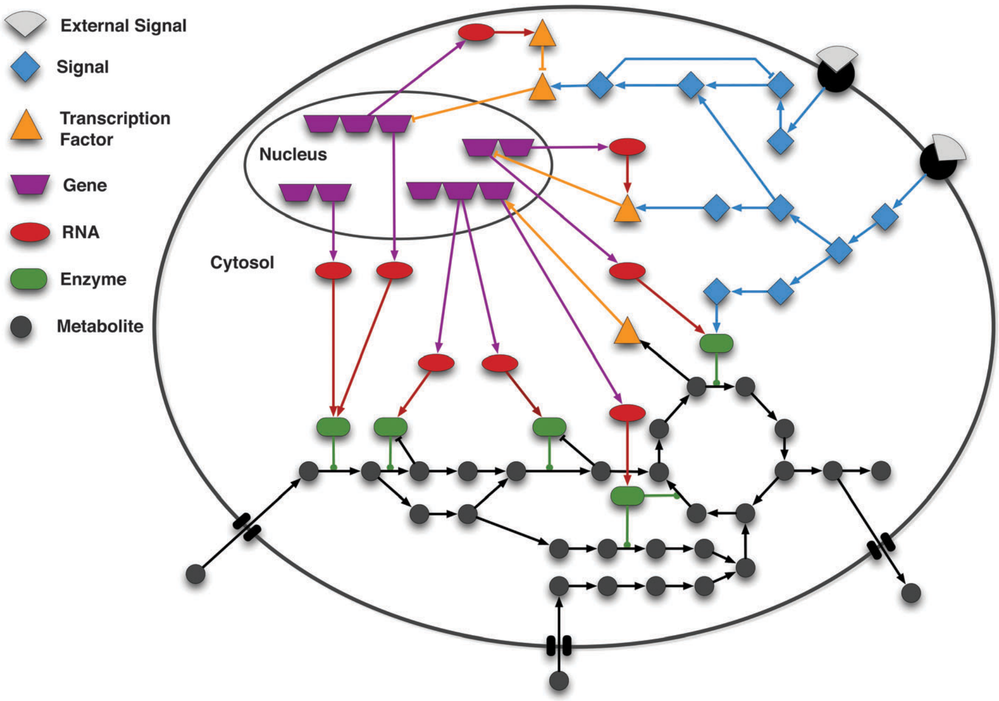


Figure 2.7: Graphical representation of the different regulatory and signalling networks connected to the metabolism of a micro-organism. Mind that FBA only takes the black metabolite network into account. Figure taken from [35].

GECKO

Up till now, it was assumed that the substrate uptake fluxes are the sole limiting factor of metabolic production. This may be a too strict assumption, because cellular reactions are facilitated by an appropriate but not unlimited number of enzymes. As enzymes are biological catalysts and consequently do not appear in any stoichiometric matrix, they are not included in any traditional metabolic network model. Simulation results thus may be violating the maximal capacity of a particular enzyme.

A straightforward way to include these enzyme capacities is by incorporating them in the stoichiometric matrix, an approach called GECKO [36]. It has been successfully implemented and validated in *S. cerevisiae* and *B. subtilis* [17, 36]. Massaiu et al., for example, applied the GECKO concept to improve TCA flux predictions by the basic *B. subtilis* iYO844 model for growth on glucose [17]. The flux prediction error for the wild-type dropped by 43 % and for a couple of knockouts in the TCA cycle, the average prediction error decreased by 36 % [17].
In this thesis, this method is slightly adapted to allow reversible reactions as well using the concept of enzyme usage, a hypothetical entity that reflects the amount of enzyme employed for a certain reaction by its absolute value and the sense of the reaction flux by its sign. A positive enzyme usage means that a reversible reaction with flux v_j is carried out in the forward direction by an amount of active enzyme e_j, while a negative one implies that the amount of active enzyme ($|e_j|$ in this case) is facilitating reaction j in the reverse direction.
More specifically, the stoichiometric matrix is extended with enzyme usage balances. Inspired by a simple kinetic definition, the reaction rate v_j increases linearly with the amount of active enzyme e_j by a kinetic factor $k_{cat,j}$, also known as the turnover number. When enough substrate available, the entire enzyme pool $[E]_j$ is busy, resulting in a maximal rate in either one of the reaction senses.

$$-k_{cat,j} \cdot [E]_j \leq v_j = k_{cat,j} \cdot e_j \leq k_{cat,j} \cdot [E]_j \tag{2.19}$$

Equation (2.19) can be rewritten as following.

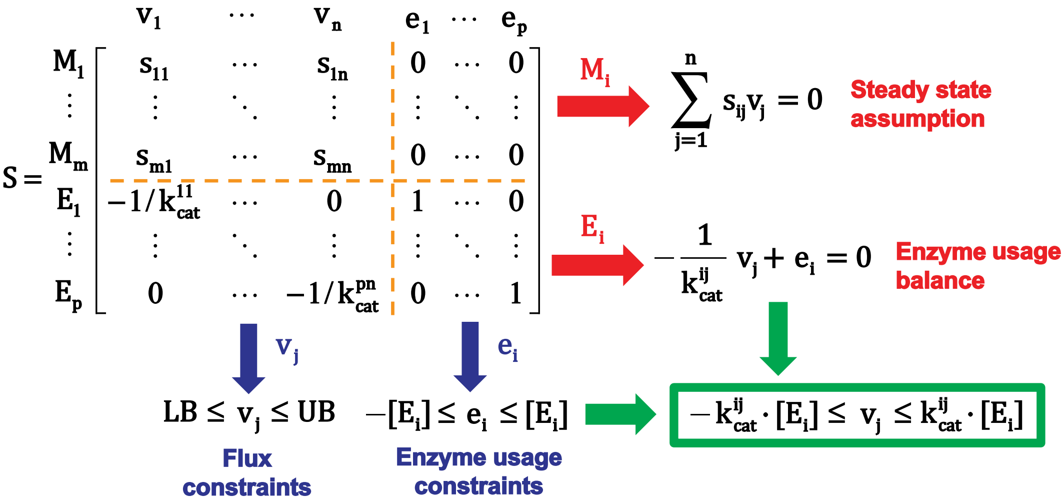
$$-[E]_j \leq \frac{v_j}{k_{cat,j}} = e_j \leq [E]_j \tag{2.20}$$

Rewriting the middle part of (2.20) as a mass balance in terms of enzyme usage e_j like in equation (2.4), the following is obtained.

$$\frac{-v_j}{k_{cat,j}} + e_j = 0 \tag{2.21}$$

As every row in a stoichiometric matrix represents the mass balance of a metabolite, a row is added for 'metabolite' e_j. The only non-zero numbers appear in column j (-$k_{cat,j}^{-1}$) and at the end (1). Mind that this new all-but-one zeros column is added at the end because a new system variable is introduced (enzyme usage e_j). Every column requires an upper and a lower constraint. For e_j, this is rather simple: it can be used as much as its capacity allows as defined in equation (2.20). This means that $[E]_j$ is the upper boundary, while -$[E]_j$ is the lower one. For irreversible reactions, the lower boundary equals 0. For clarity, the architecture of these GECKO constraints is depicted in Figure 2.8.

Both parameters (enzyme pool size $[E]$ and turnover number k_{cat}) are respectively derived from proteomics and kinetic flux measurements. The former means measuring the amount of present proteins/enzymes via e.g. mass spectrometry [37]. The latter are performed *in vitro* in such a way that the enzyme can employ its full



$$\sum_{j=1}^{n} s_{ij} v_j = 0$$

$$-\frac{1}{k_{cat}^{ij}} v_j + e_i = 0$$

Figure 2.8: Schematic representation of the GECKO model enhancement. The stoichiometric matrix is extended with enzyme usage e_i balances with boundaries at the enzyme pool size $[E]_i$. Figure adapted from [36].

capacity. However, the *in vitro* conditions are not completely comparable with the intracellular conditions that the enzyme normally experiences. Therefore, more and more assessments happen *in vivo*, but in that case it is not always guaranteed that the enzyme is performing optimal because of the cellular regulations [38].
Via equation (2.22), an apparent turnover number k_{app} is obtained. Assuming that the enzyme is optimally functional ($\eta = 1$), k_{cat} is estimated from the measured flux rate v and the applied amount of enzyme $[E]$ [38].

$$v = [E] \cdot k_{app} = [E] \cdot k_{cat} \cdot \eta \iff k_{cat} = \frac{v}{\eta \cdot [E]} \tag{2.22}$$

To avoid duplicate measurements, a couple of biochemical kinetics databases have been set up [39, 40]. Databases for proteome data are gaining popularity as well. These are mostly organism- and substrate-specific [41]. Ultimately, these databases can prelude a new omics level, the kinetome [42].

2.2.4 Implementing genetic edits *in silico*

Metabolic models so far were only representing natural wild-type networks. Nevertheless, modelling genetically perturbed networks is very interesting because it could predict the response to a genetic modification.

Complete genetic knockouts are the most popular one by far because these are the simplest to implement: it is only a matter of fixing all affected flux constraints at zero.
Implementing a genetic upregulation is not that straightforward because it is difficult to predict how much a flux would change after increasing the gene activity. Factors

beyond the scope of FBA such as the regulatory network or the enzymatic capacities of the adjacent reactions, could limit a flux change as well [43]. As a result, changes in gene activity are usually modelled in one of the following two ways. The first one is the closest to reality and only relaxes the affected flux constraints to reflect the envisaged increase of enzyme capacity. This allows the optimisation solver to rearrange the flux distribution, which does not necessarily result in a linear upscaling [44]. However, non-arbitrary flux constraints often are not available for each reaction. The second way is thus to force the upregulated flux to increase by setting both upper and lower flux constraints at the wild-type value multiplied with a desired amplification factor [45].
Both approaches have already been applied. Bro et al. modelled gene upregulations by removing existing wild-type flux constraints or by multiplying them with a specified upregulation factor [44]. On the other hand, Wang et al. forced the upregulated flux itself to be two-fold the wild-type flux as determined by pFBA [45].

The next issue is how to determine the mutant flux distribution from these altered constraints. At first sight, one would simply adjust the constraints of the reactions affected by the genetic edit and redo FBA analysis as normally. This is, however, only valid to some extent. One of the underlying premises of FBA is that an organism always strives to optimise a certain objective function, because long-term biological evolution has shaped it in that way. For genetically modified organisms, this long-term evolutionary pressure is absent, which makes it suboptimal in that view. When time passes, evolution will favour the mutations that make the biological system approach optimality. Only then, FBA is applicable again [46].

As biological evolution might be (too) slow to wait, new modelling methods have been developed to coop with this suboptimal state. All of them try to preserve the flexibility of FBA optimisation problems by switching objective functions and/or flux constraints. The most important one is MOMA.

MOMA (Minimisation Of Metabolic Adjustment)

The MOMA method assumes that a genetically perturbed metabolism will try to get as close as possible to the wild-type optimum. This is not necessarily the optimal solution of the new system, as illustrated in Figure 2.9.
The mathematical translation of this assumption is a new objective function, which minimises the Euclidian distance between the wild-type optimum and the perturbed solution subspace. The Euclidian distance D is defined in equation (2.23). The new optimisation problem is quadratic (QP) and is written down in equation (2.24), with w the wild-type solution and v the MOMA flux [46]. $v_{deleted}$ is the vector of knockout fluxes, which takes a dimension equal to the number of deleted reactions.

$$D(v,w) = \sqrt{\sum_{i}^{r}(v_i - w_i)^2} \tag{2.23}$$

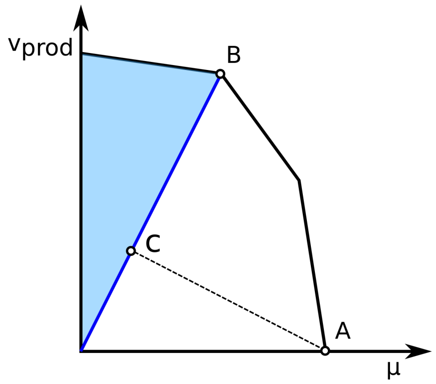


Figure 2.9: Graphical comparison of the MOMA method with the FBA optimisation method. The wild-type solution subspace is bounded by the wild-type constraints in black. The model is optimised for biomass growth, yielding point A. A knockout results in the perturbed solution subspace in blue. Optimising for biomass again gives point B, while MOMA returns point C. Figure inspired on [46].

$$\begin{aligned} \min\ & \sum_{i}^{r}(v_i - w_i)^2 \\ \text{s.t. } & Sv = 0 \\ & v_{min} \leq v \leq v_{max} \\ & v_{deleted} = 0 \end{aligned} \tag{2.24}$$

The differences between individual MOMA and FBA fluxes are often relatively small, but not absent, which neglects the efforts of a cell to reshape its regulatory network to support such a large number of small changes. A small number of large changes would be more realistic [47].

Because of its absolute nature, larger fluxes are also weighted more heavily, causing an optimisation bias. Its impact can be reduced by employing relative flux changes [48], or by calculating the Taxicab distance in equation (2.25), which considers the absolute values of the differences instead of the squared ones. In this last case, the problem has become linear and is consequently called linear MOMA or lMOMA [49].

$$D_T(v, w) = \sum_{i}^{r} |v_i - w_i| \tag{2.25}$$

dynamic MOMA

dMOMA is an alteration of the dFBA method of Section 2.2.2. While dFBA models the time evolution of biologically optimal systems by subsequent FBA calculations, dynamic MOMA does the same for suboptimal systems. For each time point, a dFBA-iteration is executed including a MOMA step which minimises the Euclidian distance between the mutant system space and the newly acquired dFBA fluxes. The MOMA fluxes are then integrated in the mass balances to get the new metabolite concentrations for the next iteration. This procedure thus adds an additional step to the dFBA procedure of Figure 2.6, as graphically depicted in Figure 2.10 [45].

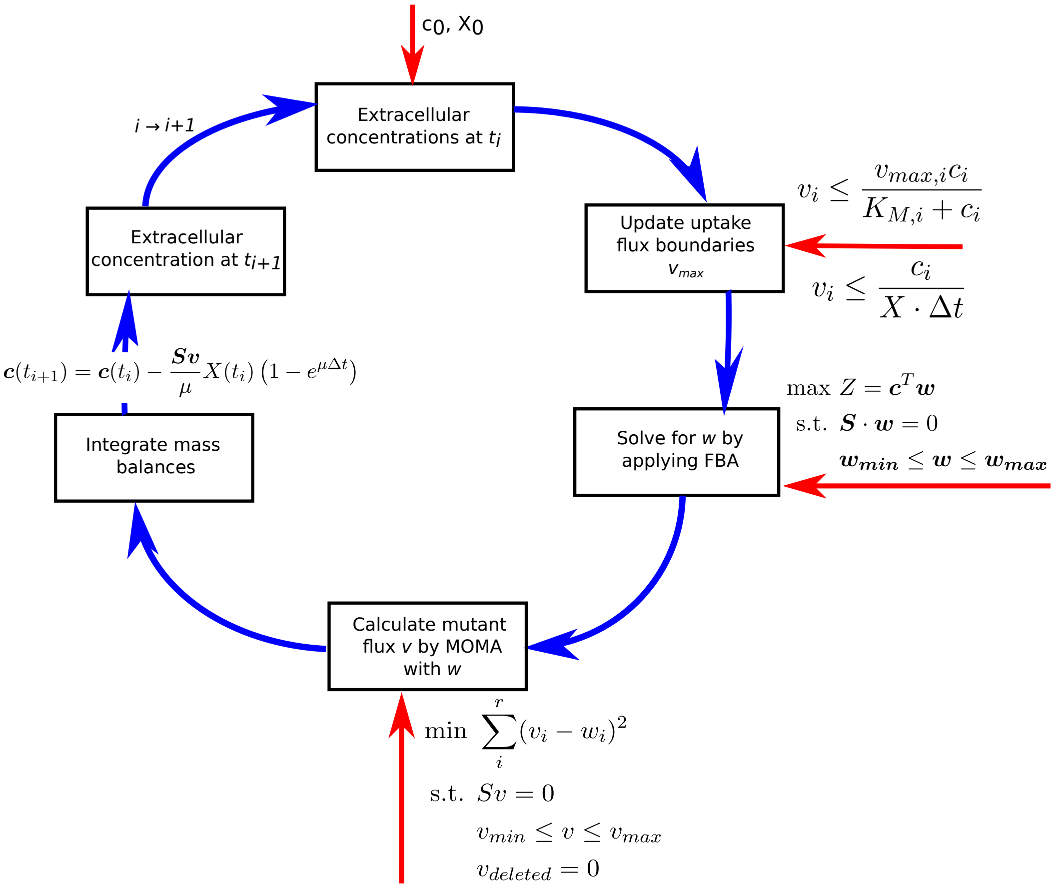


Figure 2.10: Graphical summary of the dynamic MOMA procedure. User inputs are depicted with red arrows.

2.2.5 Network optimisation algorithms

As illustrated in the previous subsection, there are tools to model genetic edits, but these do not determine which ones result in the best performance with respect to some product or process criteria. It is possible to assess all combinations, but this is not a systematic approach without redundant computational work.

In this subsection, some network optimisation algorithms are introduced. The full scope of such algorithms is not limited to these. On the contrary, a plethora of procedures exist, all with their own nuances and working principles [31, 50].
These algorithms usually return a high-potential candidate list of genes or reactions to implement *in vivo*. The majority returns knockouts, the most applied genetic modification, but other ones return partial knockdowns and/or overexpressions too. However, modelling overexpressions may be not that straightforward as knockouts, as discussed in Section 2.2.4.

Sequential scanning

A simple but naive method is ranking all possible knockout combinations, calculated by brute force. Computational work can be limited by considering only relevant subsystems or easily applicable deletions such as genes with a low number of isozymes. Alper et al. have elaborated an iterative procedure which reduces brute force computational efforts to a minimum: *sequential scanning* [51].
To select the first knockout, each deletion candidate is scored for a specified target, e.g. a production rate. Filters such as a growth rate threshold can then be applied, after which the best knockout is selected and fixed in the model. Additional ones are determined by another scanning loop. The number of knockouts increases until a specified maximum is met or when the scores do not improve anymore. The procedure is depicted in Figure 2.11.

Although a sequential approach drastically reduces the brute force computational efforts, it assumes that metabolic networks behave as convex non-linear optimisation problems. Typically, these problems can be solved by the steepest descent method, which is emulated here by selecting the knockout with the highest yield increase. This non-linear behaviour, however, has not been proven and, consequently, it is not guaranteed that this method returns global optima [51].
For example, a single knockout within a parallel pathway (e.g. isozymes) is such a problematic situation. Knocking out all branches of that parallel pathway potentially results in better yields, but this possibility has already been ruled out due to the bad performance of the single incomplete deletion. Also the use of MOMA inheres caveats as discussed in Section 2.2.4.
This procedure is easily extendable to other genetic edits such as knockdowns and upregulations. It is also easy to enlarge its searching space by making it select more than one genotype when incrementing the number of edits. Both extensions require more computational efforts though.

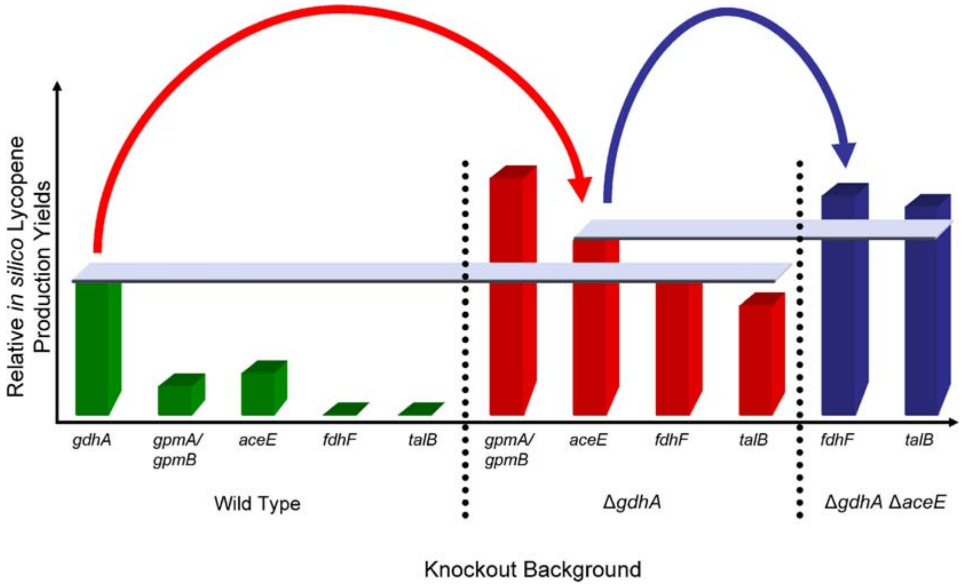


Figure 2.11: Graphical example of the sequential scanning procedure. The best genotype of the first scanning loop (green) was permanently implemented. The second scanning loop (red) resulted in another yield increase, although the best genotype was neglected as it violated the growth rate threshold. Finally, this double-knockout model was improved further by a third scanning loop (blue). Concept graph taken from a lycopene case study by Alper et al. [51].

OptKnock

Biological evolution thrives organisms to keep increasing their growth rate. If the production of the desired product is coupled to the biomass formation, it is guaranteed that its production will be optimised. Therefore, a knockout selection procedure that shapes the metabolic network accounting for these two objectives (growth rate and product yield) simultaneously, would be a valuable asset. This kind of optimisation problems are called bi-level optimisation problems. By applying linear programming duality theory, the problem can be transformed to a MILP problem, reducing the computational efforts [52].
OptKnock is the oldest procedure of this kind and employs this bi-level structure via a nested problem architecture. The inner problem is a classical FBA optimisation, in which additional constraints such as a minimal growth rate or a fixed substrate uptake, can be implemented. Also MOMA has already been applied as inner objective, but this requires other solvers and associated problem setups (cfr. biMOMA [53] & MOMAKnock [54]), making OptKnock rather inflexible to other evaluation criteria.

In more detail, the OptKnock optimisation problem can be mathematically stated as in equation (2.26), with μ_{min} a minimal desired growth rate, K the maximal number of knockouts allowed, M the knockout candidate indices and y_i a binary variable indicating if flux v_i is not knocked out in the inner problem.

$$\begin{aligned}
&\max_{y_i} \; v_{prod} \\
&\text{s.t.} \;\; \max_{v_j} \; \mu \\
&\qquad \text{s.t.} \quad S \cdot v = 0 \\
&\qquad\qquad \mu \geq \mu_{min} \\
&\qquad\qquad v_j^{min} \cdot y_i \leq v_j \leq v_j^{max} \cdot y_i \\
&\qquad y_i \in \{0,1\} \\
&\qquad \sum_{j \in M} (1 - y_i) \leq K
\end{aligned} \tag{2.26}$$

Due to high network redundancies, it is possible that multiple flux distributions score equally with respect to the cellular objective, but not to the production objective. The inner problem returns the first optimum it encounters, possibly the one with the lower production flux. Therefore, Shlomi et al. introduced a new layer between the two existing ones, which selects the smallest possible target flux. This more robust method, RobustKnock, thus maximises the minimal possible production rate [55].

OptGene

OptGene introduces concepts of evolutionary computing and enhances the target production by biologically inspired trial-and-error knockouts. In this case, the outer problem of OptKnock is replaced by a 'genetic algorithm'.

First, a population of wild-type genotypes is set up. Then, the fitness score of each individual of the current generation is evaluated for a certain objective by a method such as FBA or MOMA. Then, genetic cross-overs between individuals are performed based on their fitness scores and knockout mutations are introduced randomly, after which the fitness scores are reevaluated. This loop continues until a specified termination criterion is met. Possible criteria include a maximal computation time, a maximal number of generations or a minimal target yield. The fitness scores of the last loop determine the best performing genotype [56].

The OptGene method has three major advantages. First, fitness scoring shows high flexibility to implement any evaluation function, which contrasts with the rather rigid OptKnock algorithm. Not only plain FBA or MOMA can be used, but also derived functions such as product yields with additional filtering for a minimal growth rate. Second, it reduces computational efforts without narrowing its searching window because it does not evaluate all possible knockout combinations, neither in an exhaustive nor in a sequential way. Third, a genetic algorithm is able to optimise non-convex non-linear objective functions as its optimisation path is not fixed, but rather stochastic due to the random mutations. If run long enough, it will always approach the optimal state, although it can take some time to get away of a local optimum [56].

Chapter 3

Materials and methods

This chapter describes how the computational and experimental work in this thesis was performed. In a first part, the applied computational procedures are listed in more detail, while the experimental protocols are included in a second part.

3.1 Computational procedures

For *B. subtilis*, two GSMNs are readily available: iBsu1103 and iYO844 [10, 57]. Both were sourced from the BioModels repository [58].
All FBA calculations and derived procedures were implemented in MATLAB® R2020a/R2019b installed on a Linux Mint 18.1 Serena/Windows 10 dual-boot system. All code utilised functions of the COBRA toolbox v3 [59] and/or the IBM Cplex 12.10 solver package and/or self-developed functions.

Simulation setup of the model assessments

To ensure a proper functioning of the GSMN, the lower boundary of all exchange reactions was initially set at zero, after which the appropriate uptakes were reopened depending on the composition of the simulated medium. This was done by setting the lower flux constraint at experimental exchange rates determined in a couple of studies [17, 33]. The constraints of sulphate, phosphate, ammonium and oxygen exchanges were set at arbitrarily values in analogy with these studies. The simulated media are minimal pyruvate (PYR) and glucose (GLC) medium and complex casein hydrolysate (CAS) medium. An overview of the minimum constraints is provided in Table 3.1.

The wild-type metabolism was simulated using pFBA and validated by comparing its ability to reproduce experimental observations. These empirical data were sourced from the study by Goelzer et al. in the form of measured exchange fluxes and derived estimated TCA flux maps [33]. To ease the comparison, similar flux maps were drawn for the simulated TCA flux values using the Escher flux visualisation web-applet [60].

Table 3.1: Applied metabolite uptake rate constraints for the model assessments. All values are implemented as lower flux constraints. These values [mmol/(g_{DW}h)] were sourced from experimental data of studies as referred in the table.

Pyruvate		**Glucose**		**Casein hydrolysate**			
O_2	-100 [17]	O_2	-100 [17]	O_2	-100 [17]	HIS	-0.11 [33]
NH_4^+	-5 [17]	NH_4^+	-5 [17]	NH_4^+	-5 [17]	ILE	-0.32 [33]
SO_4^{2-}	-5 [17]	SO_4^{2-}	-5 [17]	SO_4^{2-}	-5 [17]	LEU	-0.32 [33]
p_i	-5 [17]	p_i	-5 [17]	p_i	-5 [17]	LYS	-0.04 [33]
PYR	-12.17 [33]	GLC	-7.71 [17]	ALA	-4.47 [33]	MET	-0.07 [33]
				ARG	-0.86 [33]	PHE	-0.11 [33]
				ASN	-26.81 [33]	PRO	-0.26 [33]
				ASP	-0.12 [33]	SER	-1.15 [33]
				CYS	-0.13 [33]	THR	-0.60 [33]
				GLU	-2.47 [33]	TRP	0.00 [33]
				GLN	0.00 [33]	TYR	-0.11 [33]
				GLY	0.00 [33]	VAL	-0.31 [33]

Integration of GECKO constraints

Constructing GECKO constraints for the glycolysis and TCA cycle requires appropriate kinetic and proteomic data, as described in Section 2.2.3. The stoichiometric matrix of both GSMNs was extended as depicted in Figure 2.8 by editing the model reactions using the COBRA functions. The enzyme usage variable was introduced in the affected reactions as a pseudo-metabolite $A_{counter}$ like below, which keeps the existing regular flux constraints intact. Additionally, a newly added exchange reaction functions as sink for $A_{counter}$ and obeys to the newly added enzyme usage constraint.

$$\mathrm{aA} + \mathrm{bB} + \ldots + k_{cat}^{-1}\mathrm{A}_{counter} \rightleftharpoons \mathrm{cC} + \mathrm{dD} + \ldots$$

The kinetic data were manually retrieved from the BRENDA database in the form of turnover numbers k_{cat} [39], while the protein concentrations $[E]$ were collected from a LC/MS study by Goelzer et al. [33]. As these proteomic data depend on the environmental conditions that the cell experiences, GECKO constraints have to be constructed for each considered medium (GLC and CAS).

The k_{cat}-data can be obtained in two ways from the kinetic databases: either directly, or indirectly via the specific activity SA and the molar mass MW of one enzyme subunit. SA is defined as the number of micromoles of substrate converted per milligram enzyme subunit per minute. By introducing the molar mass of the enzyme, this specific activity can be converted into a turnover number via formula (3.1) as formulated by Massaiu et al. [17].

$$k_{cat}\ \left[h^{-1}\right] = SA\ \left[\frac{\mu\mathrm{mol}}{\mathrm{mg}\cdot\mathrm{min}}\right] \cdot N\ [\text{-}] \cdot MW\ [\mathrm{kDa}] \cdot 60\ \left[\frac{\mathrm{min}}{\mathrm{h}}\right] \cdot 1\ \left[\frac{\mathrm{mg}/\mu\mathrm{mol}}{\mathrm{kDa}}\right] \qquad (3.1)$$

It is, however, assumed that the enzyme is pure and that the number of subunits N equals the number of active catalytic sites. If data for *B. subtilis* were unavailable, those for *E. coli* were used instead.

The proteomic data $[E]$ are available in the form of numbers of enzyme molecules per cell n. The upper boundary of the 95%-confidence interval was taken to avoid overconstraining. If no enzyme was detected, the minimal value in the data set was used ($6.8 \cdot 10^{-8}$ mmol/g_{DW}). To convert these data into cellular concentrations $[E]$ with units of mmol/g_{DW}, the following formula (3.2) was applied. Thereby it was assumed that per optical density unit at 600 nm (OD_{600}) $6.3 \cdot 10^8$ cells per mL are present and that there is 0.48 g_{DW}/L per OD_{600} for GLC medium and 0.47 g_{DW}/L per OD_{600} for CAS medium, in analogy with the study by Goelzer et al. [33].

$$[E]\left[\frac{\text{mmol}}{g_{DW}}\right] = \frac{n\ [\text{mlcs./cell}] \times 6.3 \cdot 10^8\ [\text{cells/(mL}\cdot\text{OD}_{600})] \times 10^3\ [\text{mL/L}]}{N_A\ [\text{mlcs./mol}] \times 0.48\ [g_{DW}/(\text{L}\cdot\text{OD}_{600}] \times 10^{-3}\ [\text{mol/mmol}]} \tag{3.2}$$

Table 3.2 reports all data required to construct GECKO constraints for both media.

Table 3.2: Kinetic and proteomic data for constructing the GECKO constraints in iBsu1103 and iYO844. Reaction names are labelled as in iYO844. Proteomic data depend on the available substrates and are measured for both media, as reported by the indicated references [17, 33]. Kinetic data data are taken from the same studies or sourced from the BRENDA database [39].

Reaction	GLC medium		Reaction	CAS medium	
	k_{cat}	[E]		k_{cat}	[E]
$\frac{mmol}{g_{DW}h}$	s^{-1}	$\frac{mmol}{g_{DW}}$	$\frac{mmol}{g_{DW}h}$	s^{-1}	$\frac{mmol}{g_{DW}}$
pgi	126 [17]	1.56×10^{-5} [17]	acont	3 [39]	7.44×10^{-5} [33]
tpi	150 [17]	1.28×10^{-5} [17]	sucoas	2343 [39]	1.37×10^{-4} [33]
gapd	70 [17]	5.77×10^{-5} [17]	gapd	70 [17]	2.68×10^{-5} [33]
pgk	329 [17]	3.61×10^{-5} [17]	pgm	766 [17]	6.12×10^{-6} [33]
pgm	766 [17]	8.86×10^{-6} [17]	eno	130 [17]	3.04×10^{-5} [33]
eno	130 [17]	3.81×10^{-5} [17]	g6pdh	174 [17]	1.10×10^{-5} [33]
g6pdh	174 [17]	8.06×10^{-6} [17]	cs	49 [17]	6.28×10^{-5} [33]
cs	49 [17]	2.52×10^{-5} [17]	icdhy	82 [17]	2.35×10^{-4} [33]
icdhy	82 [17]	1.10×10^{-4} [17]	fum	283 [17]	3.02×10^{-5} [33]
fum	283 [17]	7.30×10^{-6} [17]	mdh	177 [17]	1.78×10^{-4} [33]
mdh	177 [17]	1.10×10^{-4} [17]	ptar	652 [17]	2.27×10^{-5} [33]
ptar	652 [17]	8.49×10^{-6} [17]	ldh	6417 [17]	6.80×10^{-8} [33]
ldh	6417 [17]	6.80×10^{-8} [17]	pgcdr	15 [17]	3.76×10^{-5} [33]
pgcdr	15 [17]	1.90×10^{-5} [17]	oxadc	59 [17]	6.80×10^{-8} [33]
oxadc	59 [17]	6.80×10^{-8} [17]	micitl	19 [17]	6.80×10^{-8} [33]
micitl	19 [17]	6.80×10^{-8} [17]	oxgdc	0.2 [17]	6.80×10^{-8} [33]
oxgdc	0.2 [17]	6.80×10^{-8} [17]			

Verification of knockout-GECKO combos

The simultaneous implementation of GECKO and knockouts was validated by comparing simulation data with experimental ^{13}C flux data for a glucose medium from a study by Fischer and Sauer [61]. A representative sample was taken from this data set, including glycolysis and TCA flux deletions on one hand, and random other ones on the other hand. The MOMA method determined the knockout flux distributions using a pFBA wild-type reference. To ensure proper validation, the wild-type lower constraint of the GLC uptake was replaced by the one measured for that particular knockout. Assuming that enzyme levels are directly proportional to fluxes, the GECKO constraints $[E]$ were scaled linearly.

Effect of the medium composition

The total AA composition of three commercially available protein isolates from different sources and two laboratory media from mixed sources (LB and casein hydrolysate) was set as the initial metabolite concentrations for an sdFBA. The BD Bionutrients^{TM} Technical Manual reports the compositions of the laboratory media [62], while the commercial protein isolate compositions are available in a study by Gorissen et al. [63]. Concerning the latter, Gorissen et al. distinguish plant-based protein isolates from animal-based ones by the share of AA that are essential for human consumption (Thr, Met, Phe, His, Lys, Val, Ile, Leu). Plant-based isolates contain relatively few of these, in contrast with animal-based ones. Eventually, three media (15 g/L in total) were defined. For plant-based medium, the concentration of each of these essential AA was set at the smallest value in the data set of Gorissen et al., while the largest one was assigned to animal-based medium. Also, an average medium was defined using the average values. The share of non-essential AA was set at these averages too or, in case no data was available, its share in LB medium. Table 3.3 reports the eventual composition of these five media.

The `dynamicFBA` function of the COBRA Toolbox was modified to carry out sdFBA cycli with pFBA instead of plain FBA. This function integrates the metabolite mass balances using the general analytical solution for a static integration time step (equation (2.15)) and applies Varma-Palsson exchange flux constraints (equation (2.12)). Mineral and off-gas exchange constraints were set constant over time and the initial biomass was 0.1 g_{DW}/L. The integration time step was 1.5 min., while the simulated time interval lasted up to 40 h or until the optimisation problem became trivial or infeasible due to the absence of nutrients. The lower exchange flux boundaries are those applied earlier (see Table 3.1). Only products with a maximal concentration larger than 10 μM were plotted versus time.

Dividing these molar concentrations c_i by the total initial molar concentration of substrates c_s gives the relative molar concentrations z_i, as stated in Equation (3.3). The z_i for the twenty AA substrates is reported Table 3.3 as well.

$$z_i \left[\frac{\text{mol}}{\text{mol}_s}\right] = \frac{c_i \text{ [mM]}}{c_s \text{ [mM]}} = \frac{c_i}{\Sigma_{j=1}^{20} c_{j,s}} \tag{3.3}$$

Table 3.3: Absolute and relative molar amino acid composition of the five protein media for the analysis of the effect of the media composition. Absolute composition data were obtained from a study by Gorissen et al. [63] or from the BD Bionutrients^{TM} Technical Manual [62]. Dividing these by the total molar amount of substrates gives the relative compositions.

Amino acid conc. [mmol/L]	**Medium type**									
	Plant-based		***Animal-based***		***Average***		***LB***		***Casein hydrolysate***	
Abs. [mM] *Rel.* [mol%]	*Abs.*	*Rel.*	*Abs.*	*Rel.*	*Abs.*	*Rel.*	*Abs.*	*Rel.*	*Abs.*	*Rel.*
Alanine	5.06	9.09	5.06	5.78	5.06	7.28	6.73	8.35	4.55	4.45
Arginine	3.06	5.49	3.06	3.49	3.06	4.39	3.60	4.46	4.37	4.27
Asparagine	0.83	1.50	0.83	0.95	0.83	1.20	0.83	1.03	1.82	1.78
Aspartate	5.94	10.67	5.94	6.78	5.94	8.54	5.94	7.37	6.81	6.67
Cysteine	0.19	0.34	0.19	0.21	0.19	0.27	0.17	0.21	1.50	1.47
Glutamate	12.96	23.27	12.96	14.79	12.96	18.63	13.55	16.81	17.25	16.88
Glutamine	0.14	0.25	0.14	0.16	0.14	0.20	0.14	0.17	0.10	0.10
Glycine	4.34	7.80	4.34	4.96	4.34	6.25	4.26	5.29	3.40	3.32
Histidine	0.68	1.22	2.71	3.09	1.41	2.02	1.64	2.04	2.13	2.08
Isoleucine	1.14	2.05	4.34	4.96	2.50	3.60	5.34	6.62	4.46	4.36
Leucine	2.97	5.34	11.48	9.22	6.41	10.06	7.28	9.03	8.92	8.73
Lysine	1.02	1.83	7.24	8.26	3.70	5.32	5.77	7.16	6.83	6.68
Methionine	0.00	0.00	2.21	2.52	1.12	1.61	1.68	2.08	2.71	2.66
Phenylalanine	1.63	2.94	3.81	4.35	2.82	4.06	3.93	4.88	3.63	3.55
Proline	5.28	9.48	5.28	6.02	5.28	7.59	6.60	8.19	9.64	9.44
Serine	4.50	8.09	4.50	5.14	4.50	6.48	2.85	3.54	5.99	5.87
Threonine	1.64	2.94	6.80	7.76	3.24	4.66	2.18	2.71	2.77	2.71
Tryptophan	0.51	0.92	0.51	0.59	0.51	0.74	0.51	0.64	5.29	5.18
Tyrosine	2.12	3.80	2.12	2.42	2.12	3.04	1.05	1.30	2.98	2.92
Valine	1.67	2.99	5.51	6.29	3.43	4.93	6.53	8.11	7.05	6.89
Total contents	**55.68**	**100**	**87.61**	**100**	**69.56**	**100**	**80.58**	**100**	**102.20**	**100**

The bacterial dry weight was converted to C-moles via the molar mass of the elemental composition of *B. subtilis* biomass. Dauner et al. estimated this molar mass by directly measuring its elemental composition and by calculating it from measured concentrations of its constituents in cell lysate [64]. The average of their estimates was taken here as the molar mass of the biomass (24.415 $g_{DW}/\text{C-mol}_{DW}$).

Assessing the impact of each amino acid on the total ammonium productivity separately was done by doubling the initial concentration of each amino acid in LB medium and feeding it to the sdFBA procedure one by one. This returns new ammonium time profiles which were compared to the original LB case. For the ones with a significant impact, a complete metabolite time profile was drawn to assess the impact on metabolites different from ammonium.

Knockout optimisation algorithms

The OptGene and sequential scanning algorithms were applied to select genetic modifications. The **OptGene** algorithm only assesses deletions. An implementation of OptGene and an example overhead runner script were available in the COBRA Toolbox. The runner script was rewritten for this particular case, while the algorithm itself was adapted to optimise for a custom target by altering the embedded fitness evaluation function. The target at hand was maximising the knockout ammonia-biomass yield [$g_{NH_4^+}/g_{DW}$] as calculated via MOMA with an additional growth rate threshold of 25 % of the wild-type growth rate. The eventual OptGene procedure utilised a dual termination criterion: 4 h of computation time or 500 generations. Only knockouts in the carbohydrate and the amino acid metabolisms were considered. The procedure was rerun for each increase in number of knockouts.
The **sequential scanning** algorithm is not included in the COBRA Toolbox and was consequently built from scratch. It can assess both deletions and upregulations via MOMA using the approach of Wang et al. [45] and uses the same metabolic model, optimisation target and gene candidate set as OptGene. It was set to preserve the best 10 % of genotypes each time it incremented the number of deletions, unless otherwise specified. The code of the sequential scanning algorithm and some explanatory comments are available in Appendix C.

Finally, the feasibility of the genetic modifications suggested using both algorithms, was verified by consulting viability data in BsubCyc, a part of the BioCyc database dedicated to *B. subtilis* [65].

3.2 Experimental procedures

Strains and media compositions

The strain used for all *in vivo* experiments is the *Bacillus subtilis* 168 (ATCC 23857), which was stored in a glycerol solution at -80 °C.

The medium employed for the preculture and the reactor cultivation is the complex Luria-Bertani (LB) medium, consisting of 10 g/L tryptone (Sigma-Aldrich), 5 g/L yeast extract (Oxoid™ (Thermo Fisher Scientific)) and 10 g/L NaCl (Acros™ (Thermo Fischer Scientific)). The medium was prepared with distilled water.

Bioreactor setup and cultivation conditions

The BioFlo® 320 bioreactor (Eppendorf) with a total volume of 5 L was used for all experiments. The reactor is equipped with a pH sensor, an online OD sensor, a temperature probe, a DO sensor, two plate impellers, two peristaltic dosing pumps and baffles. Air is supplied through a sparger and a 0.2 μm PTFE air filter (Sartorius Midisart® 2000) and withdrawn through a water-cooled condenser and a similar PTFE air filter.

DASware® control software processes all the data of the sensors and controls the reactor conditions. The pH is kept constant at 7 ± 0.05 by dosing 2 M H_2SO_4 or 4 M KOH. Temperature is controlled at 37 °C, while the impeller rotation speed is set at 500 RPM. The inlet air flow rate is set at approximately 1 vvm referenced to the amount of medium present (initially 3 L). 0.5 mL of anti-foaming solution was added at the start of the experiment and when foam started to accumulate.

Analytical measurements

Biomass The biomass concentration was represented in three ways, namely via online optical density, offline optical density and dry weights. The OD was measured relative to dH_2O and is expressed in A.U., while dry weight measurements (DW) are absolute and expressed in g_{DW}.

Dry weights were measured offline by centrifuging 10 mL of reactor broth, resuspending the pellet and filtering it using a 0.2 μm cellulose ester filter, which was dried at 105 °C overnight and weighed. The initial filter weight was subtracted to obtain the bacterial dry weight. By dividing the DW by the sample volume, one gets an estimation of the biomass concentration in g_{DW}/L.

Optical density measurements rely on the absorption of monochromatic light by the sample as dictated by the Lambert-Beer absorbance law (equation (3.4)). By measuring with which intensity I light with an initial intensity I_0 transmits through the sample, the concentration of suspended (bio)mass C_x is obtained.

$$OD = \log\left(\frac{I_0}{I}\right) = \epsilon L C_x \tag{3.4}$$

The attenuation coefficient ϵ heavily depends on the absorbance characteristics of the cuvette, the medium and the biomass, and thus on the applied wave length. Also, the Lambert-Beer law strictly is only valid for simple suspensions while light scattering prevents light from being detected. Therefore, the correlation is considered valid up to OD values around 0.6. More concentrated samples are diluted after which the dilution factor is multiplied with the obtained OD value.

Both offline and online OD data were calibrated. For the offline data, this was achieved by subtracting the OD value of sterile medium. The initial online OD values contain some natural noise and were therefore calibrated by subtracting the minimal OD value of the experimental run. The online OD probe emitted another wave length than the offline OD sensor (880 nm vs. 600 nm resp.), which results in OD values different from their offline counterparts. Also, the online samples were not diluted and thus online OD values higher than 0.6 had a questionable validity.

Organics The concentrations of organic molecules such as acetic acid, were analysed by high pressure liquid chromatography (HPLC) (Agilent Technologies 1200 series). About 2 mL sample from the bioreactor was filter-sterilised using a 0.2 μm polyethersulfone filter, from which an aliquot was automatically taken and separated by an Aminex® HPX-87H column (Bio-Rad) using 5 mM H_2SO_4 aqueous eluent at a rate of 0.6 mL/min and a temperature of 40 °C. The compounds were detected with UV absorbance at 210 nm (DAD) and refractive index detectors (RID).

Amino acids The amino acids of samples were derivatised (and hydrolysed if necessary) using the EZ:faast® kit (Phenomenex) according to the manufacturer's protocols and subsequently analysed by a Perkin Elmer® Autosystem XL gas chromatograph equipped with a FID. Samples were diluted (20x) to avoid oversaturating the sorbent beads used for sample preparation.

Ammonium Ammonium concentrations were determined using LCK 304 cuvette tests (Hach). Samples were centrifuged at 3200 rcf for 15 minutes and the supernatans was filter-sterilised. 5 mL of the sample were introduced into the reaction cuvette after being diluted into the suitable concentration range of the test (400x). This is, however, a large and thus error-sensitive dilution and therefore each sample was analysed twice after which the average was taken. After 15 minutes of reaction, the mixture had turned yellow-green of which the intensity was determined by a UV-VIS spectrophotometer and translated to a concentration value.

Dissolved oxygen DO levels are defined as in Equation (3.5) and were measured online by the installed DO sensor. It was assumed that the value that the sensor returned for a sterile and fully aerated reactor medium in steady state, corresponded to 100 % DO. It was also assumed that the solubility of oxygen in LB medium is equal to its solubility in water. The oxygen saturation concentration $C_{O_2,sat}$ was thus assumed 6.7 mg/L, which is the $C_{O_2,sat}$ in water at 37 °C and 1 atm.

$$DO[\%] = \frac{C_{O_2}}{C_{O_2,sat}} \cdot 100 \tag{3.5}$$

Off-gas A BlueVary gas analyser determined the oxygen, carbon dioxide and water composition of the off-gas as well as its pressure and its temperature. The off-gas data were collected and processed by the BlueVis software and expressed in terms of standard gas conditions (sL and derived units).

Inoculum preparation

The *B. subtilis* 168 is placed on a Petri dish filled with LB agar (LB + 15 g/L agar) and grown in an incubator (Panasonic) at 37 °C.

To produce the preculture, a sterile erlenmeyer was filled with 50 to 150 mL of liquid LB medium and inoculated with a *B. subtilis* colony from the Petri dish. The preculture was put in a shake flask at 200 RPM and at 37 °C to grow for about 24 h. Then, the OD_{600} was measured. This determined which volume V would be taken of the preculture to prepare the inoculum. By applying a dilution rule using Equation (3.4) as in Equation (3.6) with a reference case (V_{ref}, $\mathrm{OD}_{600,ref}$), the suspended mass of the inoculum was kept roughly constant.

$$m_{susp} = C_{susp} \cdot V_{ref} = \frac{OD_{600,ref}}{\epsilon L} \cdot V_{ref} = \frac{OD_{600}}{\epsilon L} \cdot V \iff V = V_{ref} \cdot \frac{OD_{600,ref}}{OD_{600}} \tag{3.6}$$

The reference case was here 30 mL of a preculture with an OD_{600} value of 0.22. After centrifuging at 3220 rcf for 15 minutes at 20 °C, the supernatans was decanted as much as possible and the pellet was resuspended in the remaining LB medium to inject it into the bioreactor.

Experimental data processing

Biomass growth rate Equation (2.14) provides an expression for the biomass over time in a pseudo-steady state system. Assuming that μ remains constant within a growth phase, one can simply take the natural logarithm as following.

$$\ln(X) = \ln(X_0) + \mu t \tag{3.7}$$

X_0 represents the amount of biomass at the start of a growth phase, while t represents the time point within the phase. In case of exponential growth, plotting the natural logarithm of experimental data representing the biomass would return a linear segment for each growth phase. A linear regression using the MATLAB command `polyfit` determined its slope, which indicated the growth rate μ of the phase at hand. Additionally, the 95 %-confidence intervals were calculated using the command `polyparci` [66].

The online OD was chosen to represent the biomass because of its numerous data points. Equation (3.8) illustrates that this data representation has no influence on the regression of μ. The only necessary condition is that the online OD and the DW are linearly correlated ($X = \alpha \cdot \mathrm{OD}$) with an intercept in zero, which was verified using `polyfit` . If the 95 % - confidence interval of the intercept as calculated by `polyparci` contained zero, the correlation was accepted.

$$\ln(\alpha \cdot OD) = \ln(\alpha \cdot OD_0) + \mu t \iff \ln(OD) = \ln(OD_0) + \mu t \tag{3.8}$$

Lag time The lag time of the experimental runs was determined by comparing graphs of the process variables. This visual procedure always started by plotting the online OD on a logarithmic scale. Linear segments were identified manually after which the precise ends of each phase were identified by searching for clear changes in the other process variables, such as a change of slope in the DO curve.

Net proton balance The net proton balance represents the cumulative molar amount of free protons (H^+) added to or removed from the broth by supplying a strong acid (H_2SO_4) or a strong base (KOH) in view of pH control. It provides an indicative number for the amount of acids and ammonia produced or consumed during the experimental run.
The proton balance is directly calculated from the added volumes of 2 M H_2SO_4 and 4 M KOH at a specific time point as registered by the DASware® controller. Assuming that strong acids and bases dissociate completely, the protons of H_2SO_4 count positive on the balance, while the hydroxyl ion of KOH counts negative as it recombines with a H^+-ion into H_2O.

Oxygen transfer rate The oxygen transfer rate (OTR) represents the rate at which oxygen gas is supplied to the fermentation broth and can be defined using mass balances.

The mass balance for the gaseous side can be stated via the ideal gas law, with F_i and F_o the in- and outlet gas flow rate [m^3/h], $x_{O_2,i}$ and $x_{O_2,o}$ the in- and outlet oxygen gas concentration [mol%], P_i and P_o the in- and outlet gas pressure [Pa], T_i and T_o the in- and outlet gas temperature [K], R the ideal gas constant (8.31 J/(K·mol)), M_{O_2} the molar weight of oxygen gas (32 g/mol) and V_L the reactor content [m^3].

$$OTR\left[\frac{g}{Lh}\right] = \left(\frac{F_i x_{O_2,i} P_i}{T_i} - \frac{F_o x_{O_2,o} P_o}{T_o}\right) \cdot \frac{M_{O_2}}{RV_L} \tag{3.9}$$

As only F_i was measured, F_o was calculated using F_i by mass-balancing the nitrogen gas, which was assumed to be constant. Next to that, F_i was expressed in terms of standard conditions (sL/h) and, consequently, F_o had to be converted via the ideal gas law to standard conditions, resulting in the equation below. The standard conditions P_{ref} and T_{ref} are 1 bar and 273.15 K respectively.

$$F_o = F_i \frac{1 - x_{O_2,i} - x_{CO_2,i} - x_{H_2O,i}}{1 - x_{O_2,o} - x_{CO_2,o} - x_{H_2O,o}} \cdot \frac{P_{ref} T_o}{T_{ref} P_o} \tag{3.10}$$

By introducing equation (3.10) into equation (3.9), the final formula was obtained.

$$OTR = \left(x_{O_2,i} - x_{O_2,o} \cdot \frac{1 - x_{O_2,i} - x_{CO_2,i} - x_{H_2O,i}}{1 - x_{O_2,o} - x_{CO_2,o} - x_{H_2O,o}}\right) \cdot \frac{M_{O_2} F_i P_{ref}}{RV_L T_{ref}} \tag{3.11}$$

The inlet gas feed was assumed to be air with the following composition: 20.95 mol% O_2, 0.042 mol% CO_2, 1 mol% H_2O, and further only N_2 [67]. The outlet gas composition was measured by the BlueVary off-gas sensor.

Oxygen transfer coefficient The mass balance can also be set up for the liquid side, with k$_L$a [h^{-1}] the oxygen transfer coefficient, C_{O_2} [g/L] the bulk dissolved oxygen concentration as measured by the online DO sensor, and $C_{O_2}^{sat}$ [g/L] the dissolved oxygen saturation concentration (6.7 mg/L), which is assumed to be present at the water-gas interface.

$$OTR\left[\frac{g}{Lh}\right] = k_L a \left(C_{O_2}^{sat} - C_{O_2}\right) \tag{3.12}$$

By equalising equations (3.12) and (3.11), a linear correlation between the OTR and the oxygen saturation difference $C_{O_2}^{sat} - C_{O_2}$ (OSD) arises, linked by the oxygen transfer coefficient $k_L a$. Eventually, an estimate of the oxygen transfer coefficient was obtained via a linear regression of the experimental OTR as determined by the gas-side balance (3.11), and the OSD data using `polyfit`. Again, confidence intervals were calculated using `polyparci` .

3.3 Integration of experimental data into the model

The generated experimental data were integrated in the metabolic model or the computational procedures to refine its performances with respect to the observations.

Dissolved oxygen balance

In contrast with the other external metabolites (equation (2.10)), the dissolved oxygen mass balance (equation (3.13)) involves a transport term as well, complicating an analytical integration. From equation (3.7), $X(t)$ is already known, while $k_L a$ is the observed oxygen transfer coefficient. μ and q_{O_2} are the biomass growth rate and the oxygen uptake rate as determined by the underlying FBA problem ($q_{O_2} = (\boldsymbol{Sv})|_{O_2}$). μ is considered positive and q_{O_2} negative.

$$\frac{dC_{O_2}}{dt} = q_{O_2}(t)X(t) + k_L a \left(C_{O_2}^{sat} - C_{O_2}(t) \right) \tag{3.13}$$

Integrating this linear first order differential equation using integration factors gives the following result. A more detailed calculation is provided in Appendix B.

$$C_{O_2}(t) = C_{O_2,0} e^{-k_L a t} + C_{O_2}^{sat} \left(1 - e^{-k_L a t} \right) + \frac{q_{O_2} X_0}{\mu + k_L a} \cdot \left(e^{\mu t} - e^{-k_L a t} \right) \tag{3.14}$$

By the pseudo-steady state approximation, the oxygen balance of equation (3.13) is considered valid within a time period $t_i \to t_i + \Delta t = t_{i+1}$. Hence, equation (3.14) is introduced in the dFBA procedure as the forward iteration formula for the extracellular oxygen concentration. The next step in the dFBA cycle is reevaluating the FBA problem for the new metabolite concentrations of t_{i+1}, changing X_0, μ and/or q_{O_2}.

$$C_{O_2}(t_{i+1}) = C_{O_2}(t_i) e^{-k_L a \Delta t} + C_{O_2}^{sat} \left(1 - e^{-k_L a \Delta t} \right) + \frac{q_{O_2}(t_i) X(t_i)}{\mu(t_i) + k_L a} \cdot \left(e^{\mu \Delta t} - e^{-k_L a \Delta t} \right) \tag{3.15}$$

The initial dissolved oxygen concentration $C_{O_2}(0)$ was set at the oxygen saturation concentration $C_{O_2}^{sat}$.

Chapter 4

Model selection and analysis

Any design procedure obviously requires a metabolic network model to optimise. To select a suitable one, its performance must be assessed in depth and validated with experimental data, especially with regard to the wild-type. Metabolic models can be found in databases like BiGG [68] and BioModels [58] or in individual research articles.

Querying these databases and literature learned that there are two major models for *B. subtilis.* Model iYO844 is the first ever GSMN model for *B. subtilis*, including 844 genes, 988 metabolites and 1020 reactions of which 65 % are irreversible. It was reconstructed using the annotated genome sequence of Danchin et al. from 1997 [69], then available biochemical knowledge, large-scale gene essentiality datasets from literature and newly generated high-throughput phenotyping data [10]. It has an overall gene essentiality accuracy of 89.0 % when validated against a new database of gene essentiality data, gene interval knockout data and Biolog phenotyping data [57].

However, iYO844 is based on an outdated annotated genome and relies solely on biochemistry databases for its directionality and reversibility information, possibly resulting in overconstrained networks. iBsu1103 attempts to tackle this issue by using updated genome annotations and by estimating reaction free energies to determine the reaction directionality and reversibility, resulting in a more relaxed network. It contains 1103 genes, 1145 metabolites and 1143 reactions of which 45 % is irreversible. The overall gene essentiality accuracy increased to 93.2 % when validated against the same dataset [57].

This chapter evaluates both models at the accuracy of predicted wild-type flux distributions as well as knockout flux patterns. Model remediations are introduced when necessary. Finally, the capabilities of the wild-type metabolism are explored by studying the impact of the medium composition on the ammonia productivity over time using sdFBA.

4.1 Model selection

4.1.1 Assessment of the basic models

A first assessment includes setting up both models for a couple of media for which experimental exchange rate and/or growth rate data are available in literature. These empirical substrate uptake rates are set as the maximal uptake rates in the model, and via pFBA calculations the models are checked for being able to reproduce the experimental production rates.

More specifically, the behaviour for three media (pyruvate - PYR, glucose - GLC and casein hydrolysate - CAS) is simulated and compared to some reference growth rates and production rates taken from studies by Massaiu et al. and by Goelzer et al. [17, 33]. Also the simulated TCA cycle fluxes were compared with estimated TCA flux maps from this last study by visualising them in a similar flux map, as shown in Figure 4.1. Table 3.1 reported the applied flux constraints.

It immediately sticks out that iBsu1103 behaves different from iYO844. The predicted growth rates differ substantially in two cases out of three, and either the TCA cycle nor the glycolysis are activated in accordance with the experimental data. In one case of three, it even predicts to produce formate as the main secretion product instead of the observed acetate. Taking a closer look at all secretion rates (not shown), iBsu1103 also excretes moderate amounts of succinate and adenine, rendering its behaviour even odder. Closing the exchange sinks of these metabolites does not solve the problem: it produces then other unexpected compounds.

On the other hand, iYO844 does not perform that excellent either, when comparing the flux patterns in Figure 4.1. For example, the acetate exchange rates should be higher, especially in the CAS case, in which the simulated rate is absent but the observed one is the largest of the three cases (34.75 mmol/g_{DW}h)). In the glucose case, for example, 4.94 mmol/(g_{DW}h) acetate was observed while only 1.20 mmol/(g_{DW}h) was predicted. The growth rate also is too high in the CAS case.

As a consequence, these findings impose to find model enhancements to rectify the behaviour of iBsu1103 and/or to ameliorate that of iYO844. One way to achieve this is by integrating experimental data of the desired growth conditions to constrain the solution space into the right direction.

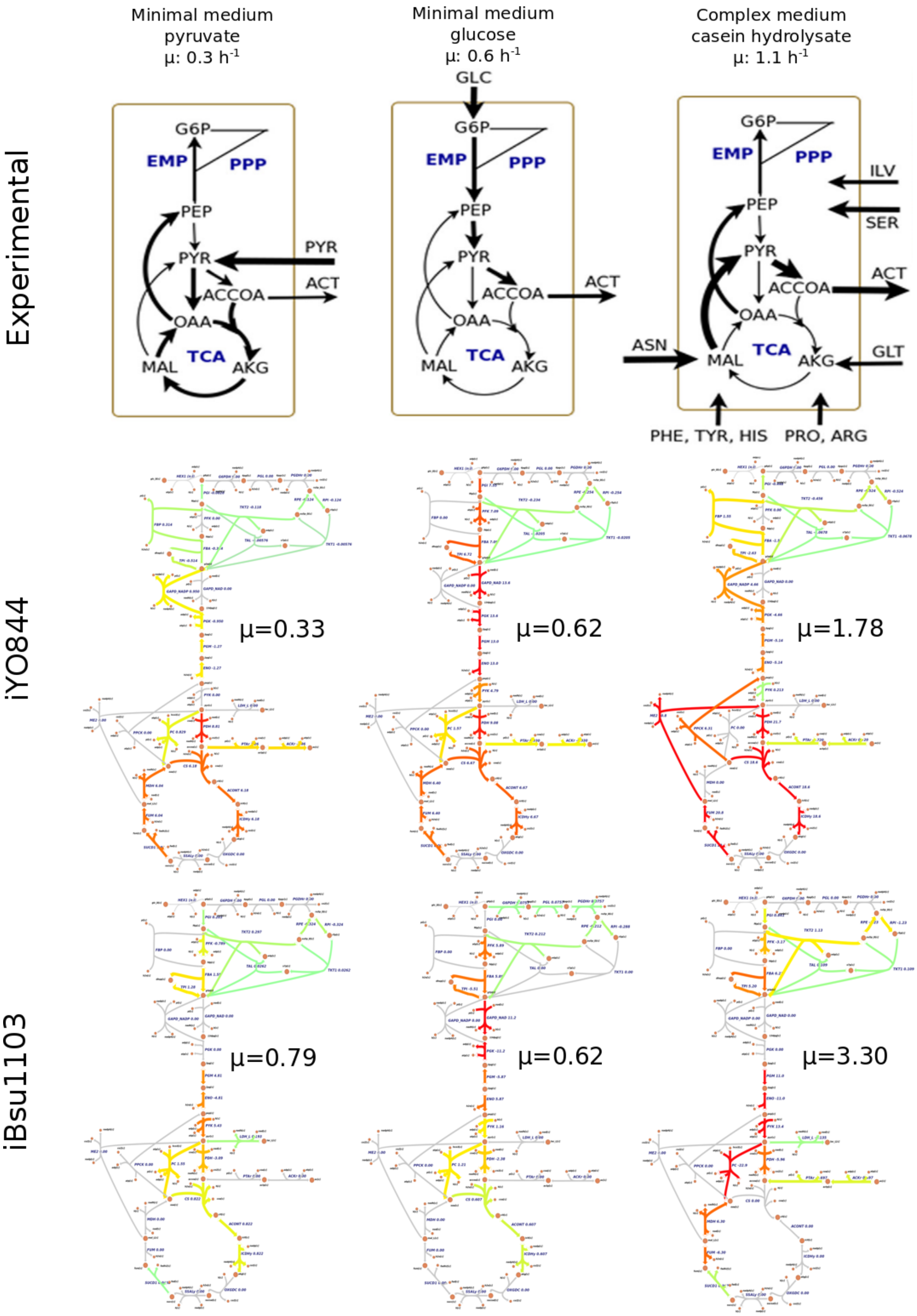


Figure 4.1: Experimental and simulated flux maps of the central carbon metabolism for a selection of growth media. The upper row of figures displays experimental flux maps from Goelzer et al., which were estimated by measuring the substrate uptake rates and the product secretion rates [33]. The arrow thickness is proportional to the flux value. The lower two rows depict simulated TCA flux values obtained from pFBA calculations for both GSMNs, which were visualised by Escher flux mapping [60] using a logarithmic colour scale from green (0) to red (30 mmol/(g_{DW}h)).

4.1.2 Assessment of the enhanced models

In analogy with [17], GECKO is applied to enhance both models because of its easily implementable but effective concept, as discussed in Section 2.2.3. This, however, requires quite some kinetic and proteomic data. For the GLC case, these data were retrieved from a proof-of-concept study by Massaiu et al. [17], while the data for the CAS case were calculated similarly using data from other studies [33] and BRENDA. The PYR case was omitted here as the Cplex solvers returned infeasibility errors after introducing GECKO constraints, which illustrates that the currently available kinetic and proteomic data are not always accurate enough for model feasibility. The data utilised for the eventual GECKO constraints are available in Table 3.2.

Again, some Escher flux maps in Figure 4.2 visualise the results, from which two conclusions are drawn. First, GECKO has visibly enhanced the iYO844 model: while it previously had difficulties to reproduce the experimental flux values, it now exhibits good similarity. The acetate secretion flux, for example, has increased for both GLC and CAS media to acceptable values without divergent growth rates. More specifically, the experimental glucose case contains molar acetate secretion fluxes of about half the size of the glucose uptake rates, similarly as in the simulation data (4.37 mmol/(g_{DW}h) AC vs. -7.71 mmol/(g_{DW}h) GLC). Moreover, the AC secretion rate (21.56 mmol/(g_{DW}h)) in the CAS case is of the same order of magnitude as the ASN rate (-26.81 mmol/(g_{DW}h)).
On the other hand, iBsu1103 has not been rectified by GECKO constraints, as its unexpected behaviour has even slightly deteriorated. While previously, parts of the glycolysis were still active, it now is almost completely inactive, just as parts of the pentose phospate pathway. The growth rate is too high in the CAS case as well.

The question then arises as to what causes this divergence and how it can be avoided or corrected. Possible reasons are that the directionality of the model is over- or underconstrained, resulting in flux pathways that are inherently feasible but very rarely active in practice. As mentioned at the beginning of this chapter, the iBsu1103 model distinguishes itself from iYO844 because of its relaxed directionality based on the reaction thermodynamics [57]. This seems interesting on its own, but loosening constraints does not imply that the cell is able to support fluxes in all thermodynamically possible senses because of its biological background. Regulatory processes for example, which are not included in these metabolic models, might prohibit this. Moreover, the necessary cofactors are not always present in the required quantities nor the enzymatic capacities are.

Because of its divergent behaviour, it might be useful to reevaluate the model. However, for the sake of time, the iBsu1103 model is abandoned in this thesis, as the iYO844 model reflects reality better and consequently has a higher potential for further applications.

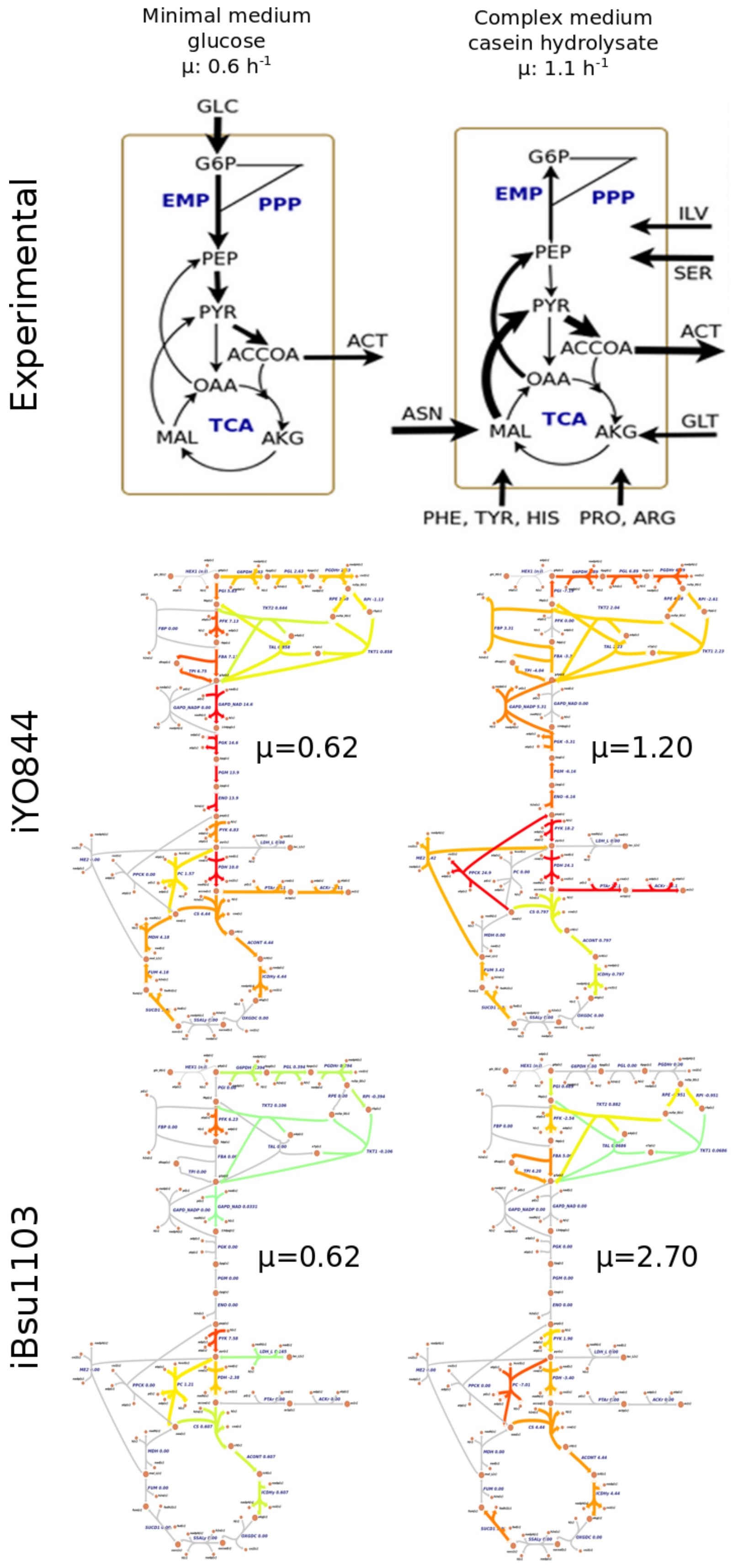


Figure 4.2: Experimental and simulated flux maps of the central carbon metabolism for a selection of growth media after GECKO enhancement. In the upper row of figures, adapted from Goelzer et al. [33], the arrow thickness is proportional to the flux value. The lower two rows of figures are simulated pFBA flux values obtained using both GECKO-enhanced models. Figures are made with Escher flux mapping [60], using a logarithmic colour scale from green (0) to red (30 mmol/(g$_{DW}$h)).

4.1.3 Verifying genetically perturbed model behaviour

GECKO-enhanced iYO844 has so far been proved a suitable model, but one aspect still must be verified. Combining genetic perturbations and GECKO appears quite straightforward as enzymatic capacities do not change in the short term. This combo has already been applied successfully as illustrated earlier in Section 2.2.3. An additional check, however, does not harm.

Via MOMA, some single deletions were simulated and compared to experimental flux data of Fischer and Sauer [61]. To have a quantitative parameter, the normalised root mean square error (NRMSE) of each reaction flux was calculated. The NRMSE is defined in Equation (4.1) with v an simulated flux, v_{exp} an experimental flux and N the number of measurements.

$$NRMSE = \sqrt{\frac{\Sigma_i (v_i - v_{i,exp})^2}{N}} \cdot \frac{N}{\Sigma_i v_{i,exp}} \tag{4.1}$$

From Table 4.1 it is clear that the match is good. Nevertheless, two remarks can be made, of which the first one concerns the systematic overprediction of the growth rate. This is attributable to the hypothesis that microbes tend to maintain unnecessary enzymes to be able to mediate sudden environmental changes, which causes a suboptimal level of enzyme maintenance that reduces metabolic yields. This hypothesis has found support in the study from which these experimental reference data were sourced [61].
Secondly, it is remarkable that extracellular fluxes (μ, glucose and acetate exchange) deviate less than intracellular fluxes (*pgi*, *g6pdh* and *cs*). This can be explained by the fact that the exchange fluxes are located at the periphery of the metabolic network, where there are almost no competing reactions. Every pathway secreting acetate, for example, contains the acetate exchange flux. On the other hand, the intracellular reactions face a higher level of connectivity and freedom and often are only one of the possible pathways to convert a metabolite into another.

Nevertheless, the general trends of the experimental measurements are predicted quite well and especially the similarities of the exchange fluxes and the growth rates are encouraging. Combining MOMA and GECKO thus appears to be compatible, passing the last checkpoint for GECKO-iYO844, hereafter called iYO844-Gk.
iYO844-Gk will now be used to explore and optimise the metabolism of ammonia secretion.

Table 4.1: Validation of combining knockouts and GECKO in model iYO844 for a selection of deletions. MOMA knockout simulations of glucose growth were compared with experimental exchange and ^{13}C-flux data from a study by Fischer and Sauer [61].

Strain	$\boldsymbol{\mu}$	**EX-GLC**	**EX-AC**	**pgi**	**g6pdh**	**cs**
Unit	h^{-1}	$\frac{mmol}{g_{DW}h}$	$\frac{mmol}{g_{DW}h}$	$\frac{mmol}{g_{DW}h}$	$\frac{mmol}{g_{DW}h}$	$\frac{mmol}{g_{DW}h}$
wild-type	0.624	-8.625	4.373	5.832	2.635	4.441
wild-type (exp.)	0.490	-8.007	4.083	5.247	2.655	2.886
Δ*acoABCL*	0.320	-4.742	2.279	2.132	2.529	2.224
Δ*acoABCL* (exp.)	0.280	-4.320	1.426	3.035	1.200	1.489
Δ*citH*	0.501	-7.079	3.153	3.186	3.767	3.313
Δ*citH* (exp.)	0.440	-6.430	2.702	4.364	1.965	2.330
Δ*fbp*	0.507	-7.163	3.185	3.224	3.812	3.352
Δ*fbp* (exp.)	0.410	-6.510	3.319	4.565	1.841	2.028
Δ*galE*	0.518	-7.307	3.239	3.289	3.888	3.419
Δ*galE* (exp.)	0.430	-6.640	3.185	4.339	2.194	1.969
Δ*gapB*	0.401	-5.791	2.669	2.605	3.085	2.713
Δ*gapB* (exp.)	0.370	-5.270	2.056	3.337	1.835	2.299
Δ*glcP*	0.492	-6.968	3.112	3.136	3.708	3.261
Δ*glcP* (exp.)	0.410	-6.330	1.898	4.429	1.796	2.281
Δ*hutM*	0.465	-6.610	2.977	2.974	3.519	3.095
Δ*hutM* (exp.)	0.400	6.010	2.523	3.923	1.982	2.152
Δ*katX*	0.505	-7.130	3.173	3.209	3.794	3.337
Δ*katX* (exp.)	0.420	-6.480	2.722	4.180	2.199	2.833
Δ*mdh*	0.287	-5.421	2.513	2.483	2.789	2.042
Δ*mdh* (exp.)	0.300	-4.760	5.000	3.001	1.672	0.408
Δ*odhA*	0.372	-6.491	6.683	2.697	3.700	0.364
Δ*odhA* (exp.)	0.330	-6.320	6.069	4.553	1.675	0.511
Δ*pckA*	0.616	-8.568	3.714	3.858	4.555	4.006
Δ*pckA* (exp.)	0.490	-7.780	3.733	4.730	2.943	2.886
Δ*pps*	0.624	-8.633	4.414	5.968	2.508	4.459
Δ*pps* (exp.)	0.530	-8.660	4.157	5.669	2.887	2.653
Δ*sdhC*	0.398	-6.886	7.058	2.863	3.923	0.389
Δ*sdhC* (exp.)	0.410	-6.700	6.632	5.078	1.520	0.580
Δ*zwf*	0.308	-8.345	3.697	6.200	0.000	3.895
Δ*zwf* (exp.)	0.340	-7.560	6.347	6.903	0.558	1.354
NRMSE [%]	**18.94**	**8.90**	**28.96**	**25.37**	**80.42**	**65.85**

4.2 Effect of the medium composition

Optimising metabolic networks often relies on assuming a specific substrate mixture, which can be subject to optimisation as well. Because complex media often consist of multiple nutrients in different quantities, it is necessary to check the sensitivity of the ammonia productivity to the nutrient concentrations. This is especially imperative for this case, as there are 20 amino acids (AAs) that can vary in concentration due the natural origin or former pretreatments of the protein sources.

A first step comprises determining which protein source would be preferable. Therefore, a couple of media are defined of which the AA compositions relate to the source of commercially available protein isolates. Those include a plant-based medium, an animal-based medium and an ‘average’ medium which reflects mixed sources [63]. Two popular laboratory polypeptide media, LB medium and casein hydrolysate medium are defined as well [62]. Compositions of these media were obtained as described in Section 3.1 and are available in Table 3.3. For each of these media, an sdFBA analysis of iYO844-Gk is executed to record a metabolite time profile for each metabolite of interest.
However, in Table 3.3, it is noticeable that the total initial amount of substrates differs quite extensively, which creates a non-negligible bias. Therefore, a relative concentration profile was also set up for each compound. This relative concentration is defined as the molar amount of the metabolite of interest divided by the initial total molar amount of substrates. The biomass in g_{DW} was converted to C-mol_{DW} via the average of the estimated molar masses of the elemental biomass composition by Dauner et al. (24.415 g_{DW}/C-mol_{DW}) [64].

The results of this analysis are depicted in Figures 4.5, 4.6 and 4.7. All simulations have three similar major exponential growth phases. The first one is short (0.5 h) and is characterised by the consumption of all amino acids, until asparagine (ASN) is depleted. In the second phase, the other amino acids are taken up till some major substrates like glutamate (GLU), arginine (ARG) and serine (SER) are used up quasi-simultaneously at 6-8 h. In the third phase the remaining AAs are converted until valine (VAL) is depleted, after which no non-zero growth rates are predicted anymore and the simulation is consequently stopped. Growth rate reductions within a major growth phase can be attributed to the depletion of a minor substrate.

In all situations a similar biomass concentration is obtained, while producing the same compounds at similar rates. However, in all cases, there is a propionate effluent (ppa), which was not mentioned in earlier studies [6, 7, 33]. New experimental observations must clarify this issue. Next, the biomass productivity of the CAS medium is slightly higher due to the more elevated amounts of GLU, SER and ARG, which are preferred substrates as reflected in the uptake rates of Table 3.1. Moreover, the larger initial amount of ASN allows to produce more biomass early as well, resulting in somewhat higher consumption rates and a lower batch time, which is nicely displayed in, for example, the isoleucine (ILE) curve of Figure 4.6.

The final ammonium concentration amounts to approximately 50 mM, or 0.9 g/L, but is expected to be positively influenced by the amount of highly aminated AAs like ARG, ASN and histidine (HIS). For the sake of completeness, an overview of the structural formulas of all AAs is provided in Figure 4.3, from which the nitrogen content of each AA can be consulted. And indeed, media with a higher share of these AAs yield higher ammonium concentrations. CAS medium, for example, which has a noticeable higher content of highly aminated AAs, has a substantially higher ammonium productivity.

Interestingly, other conclusions can be made from the relative concentration profiles. While CAS medium, for example, has the highest ammonium titer in absolute terms, plant-based medium has the highest one in relative terms, because the amount of initial substrate is larger for CAS. Secondly, the plant-based medium appears here to be the best AA blend for ammonia production. This implies that plant-based protein sources are more interesting for microbial protein deamination. However, according to the composition data of Gorissen et al. [63], plant-based protein isolates have the lowest average protein content, despite its superior AA composition. The key is thus to find a plant source with a high protein content and a favourable AA composition.

The follow-up question is what the impact of each AA on the ammonium secretion rate is. Therefore, a sensitivity analysis is set up in which the absolute content of one AA in LB medium is doubled one by one separately. Additionally, these alterations could influence the production rates of other metabolites, which hence must be checked.

Figure 4.8 displays the ammonium time profiles of these media alterations. Interestingly, not only the highly aminated AA (ARG, ASN, HIS) have a positive impact on the ammonium productivity, but also alanine (ALA), glutamate (GLU), serine (SER) and threonine (THR). Apart from the amino-rich AA, these four AA with only one amino group – hereafter called preferential AA – are consumed at the highest rates in Table 3.3 and are thus fully converted, unlike their non-preferential counterparts. These uptake rates could also explain why the influence of ALA, ARG and GLU is large, while that of HIS, SER and THR is rather small. Anyhow, because of their high consumption rates, preferential AA generate an additional ammonia flux despite having only one amino group, while the expected contribution of the amino-rich AA comes from their multiple deaminations. As an illustration, Figure 4.4 shows the total share of the mentioned amino-rich and preferential AA, which correlates with the slope of the relative ammonium profile of Figure 4.7. The medium with the highest share of these AA thus generates the fastest relative increase in ammonium. Figure 4.10 shows that the other metabolites are largely unaffected. Only in the case of ALA, the acetate productivity has almost doubled.

To conclude, changes in ammonium productivity of the wild-type metabolism are attributable to changes in the medium composition but the major growth phases remain unaffected. Increasing ammonium productivity is feasible by increasing the share of the preferential and/or the highly aminated amino acids.

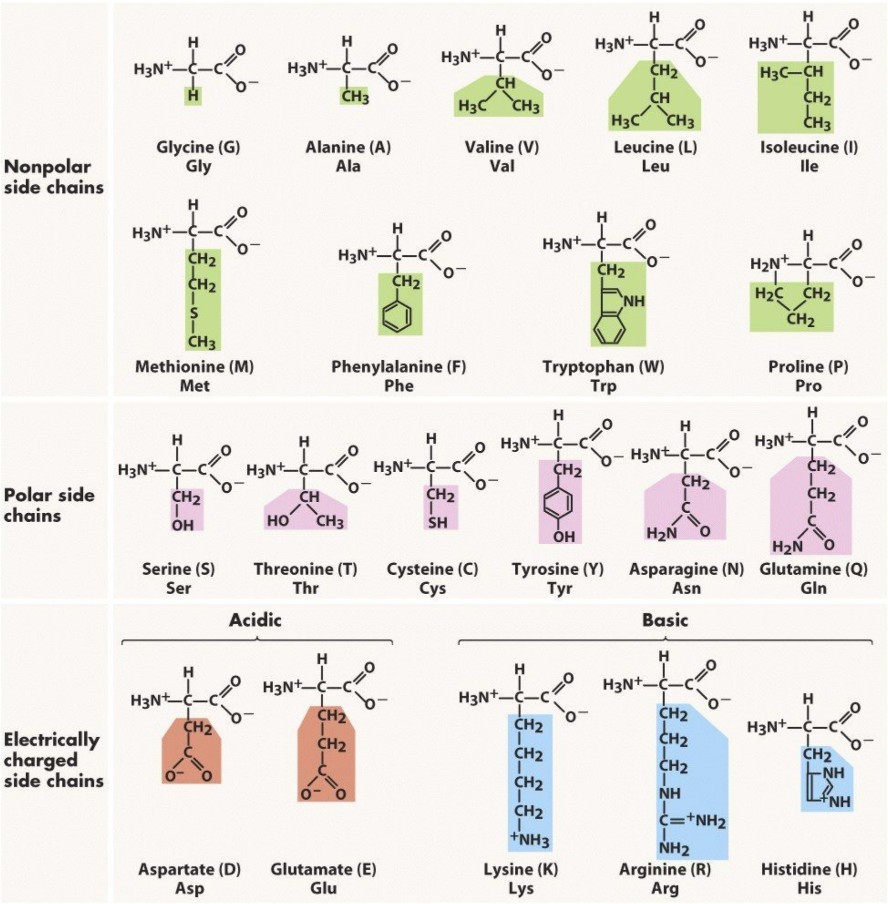


Figure 3-5 Biological Science, 2/e © 2005 Pearson Prentice Hall, Inc.

Figure 4.3: Overview of the structural formulas of all AAs. Figure taken from [70].

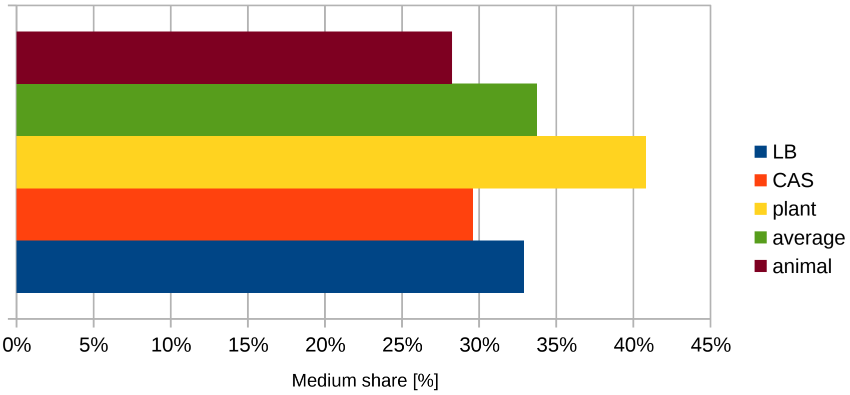
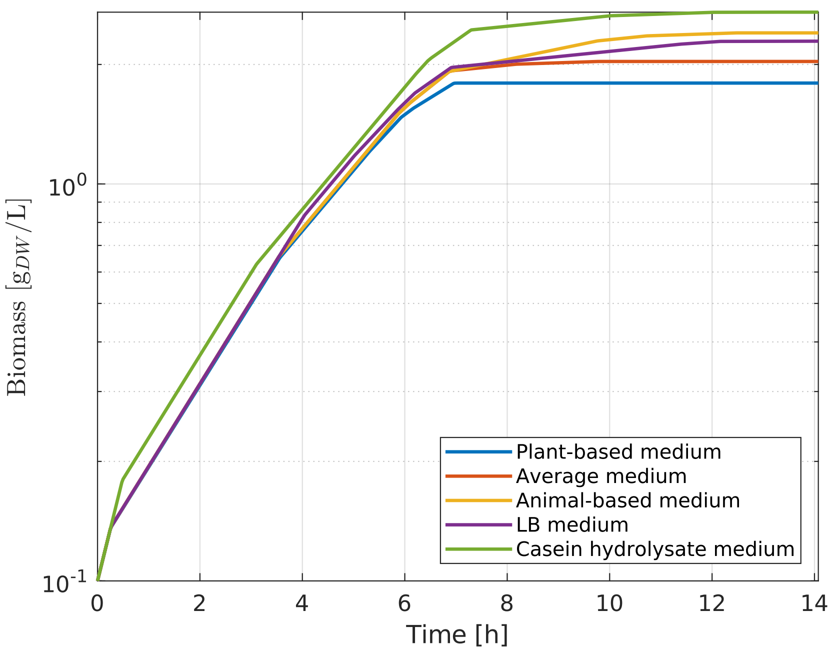
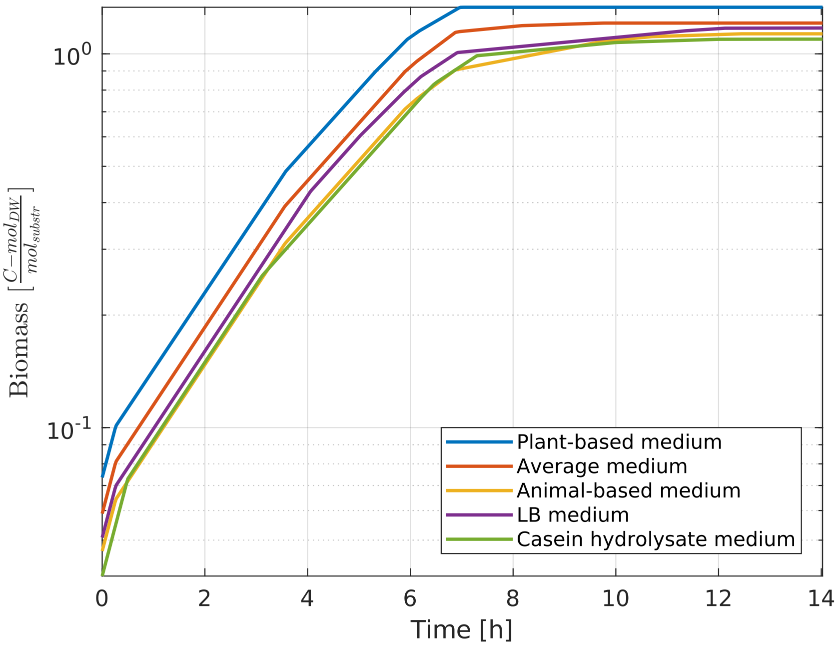


Figure 4.4: Total medium share of the amino-rich (ARG, ASN, HIS) and the preferential AA (ALA, GLU, SER, THR) in LB medium. Note that a higher share of these AA results in more ammonia per initial amount of substrates.



(a) Absolute quantities



(b) Relative molar quantities with respect to the initial molar amount of substrates

Figure 4.5: Effect of the medium composition on the biomass time profile. These logarithmic biomass concentration time profiles were obtained by executing sdFBA analyses of iYO844-Gk for five different media compositions. The relative values were converted from the absolute ones using the average of biomass molar masses estimates by Dauner et al. (24.415 g_{DW}/C-mol_{DW}) [64].

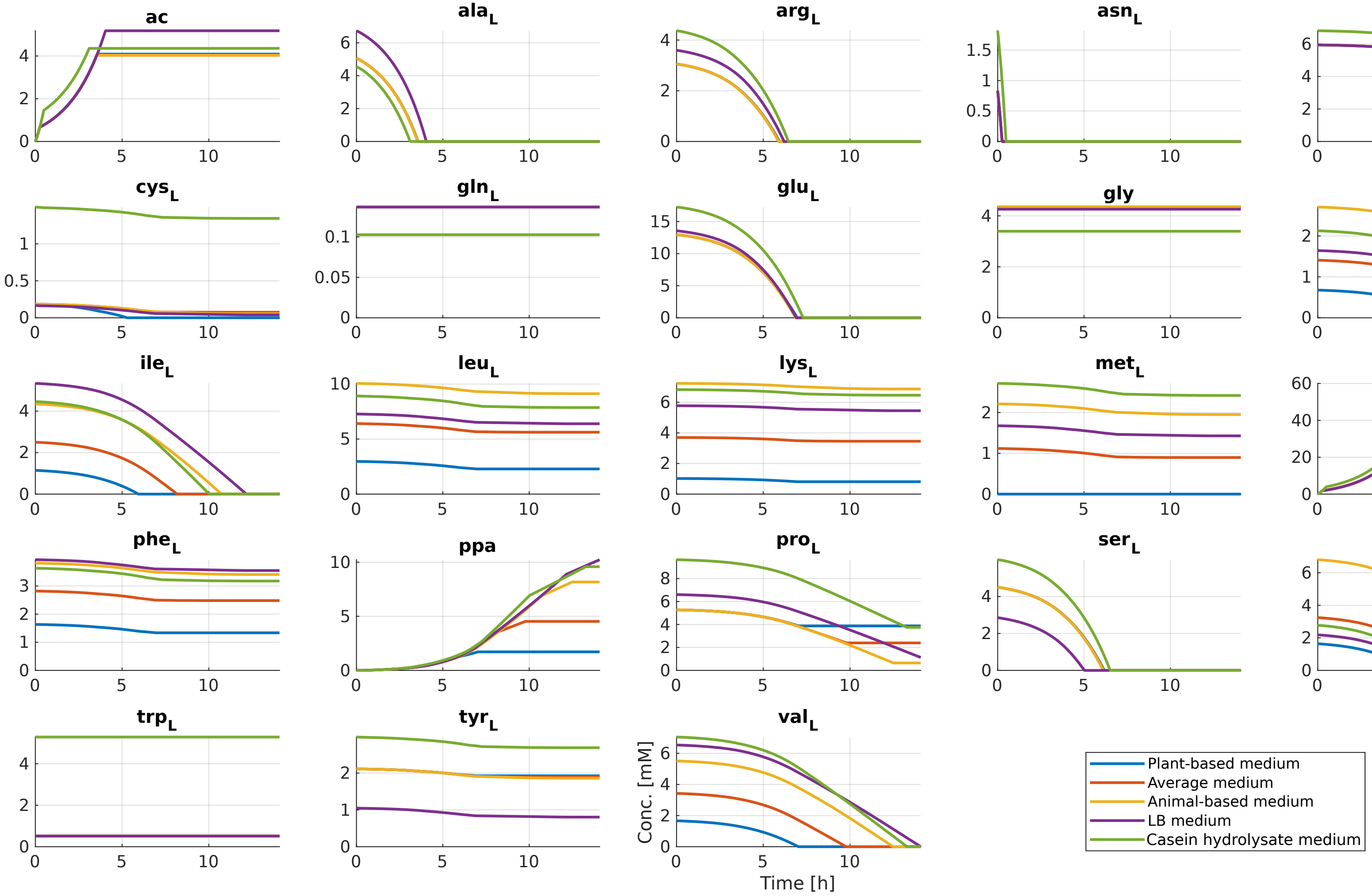


Figure 4.6: Effect of the medium composition on the metabolite time profiles. These metabolite concentration time profiles were obtained by executing sdFBA analyses of iYO844-Gk for five different media compositions. The figures include all 20 substrates and other relevant products. Product labels are as in the iYO844 model.

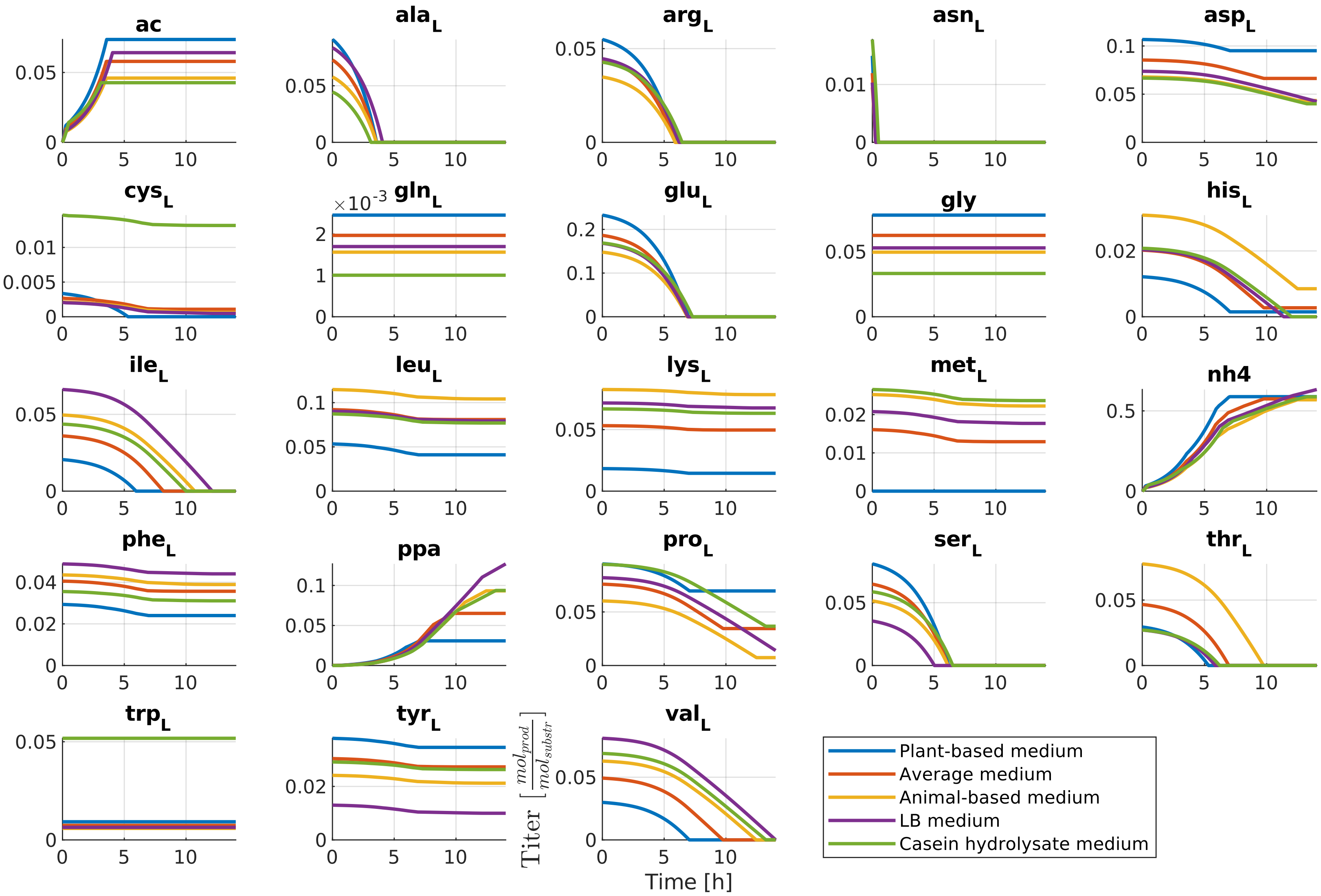


Figure 4.7: Effect of the medium composition on the metabolite time profiles with respect to the initial molar amount of substrates. These metabolite concentration time profiles were obtained by executing sdFBA analyses of iYO844-Gk for five different media compositions and converted to relative molar amounts via the appropriate molar masses. The figures include all 20 substrates and other relevant products. Product labels are as in the iYO844 model.

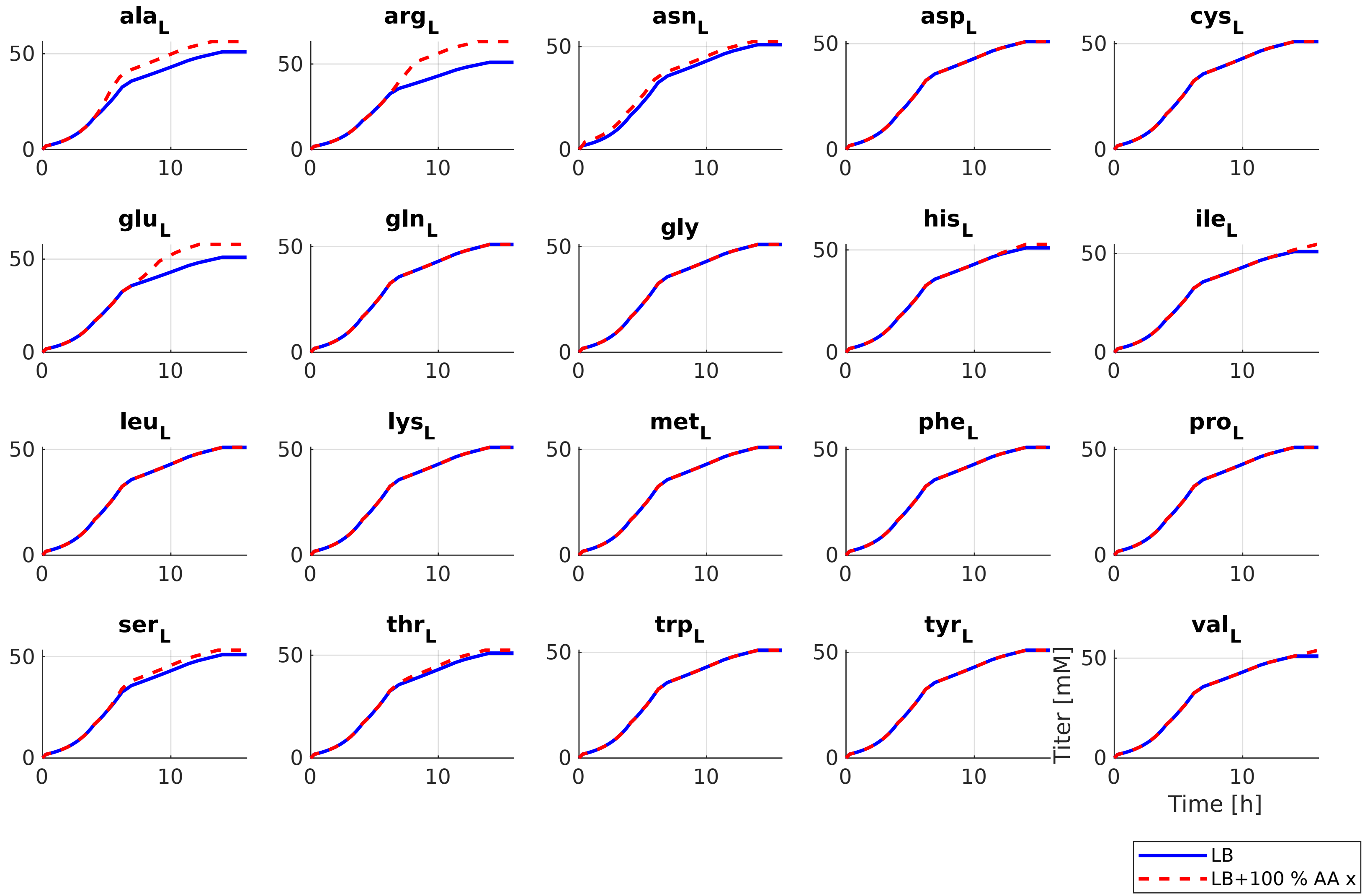


Figure 4.8: Effect of doubling the content of one specific AA on the ammonium time profile. These ammonium profiles are obtained by sdFBA analyses for media compositions in which the content of the AA mentioned above each graph was doubled (in red) with respect to the original LB media composition in Table 3.3 (in blue).

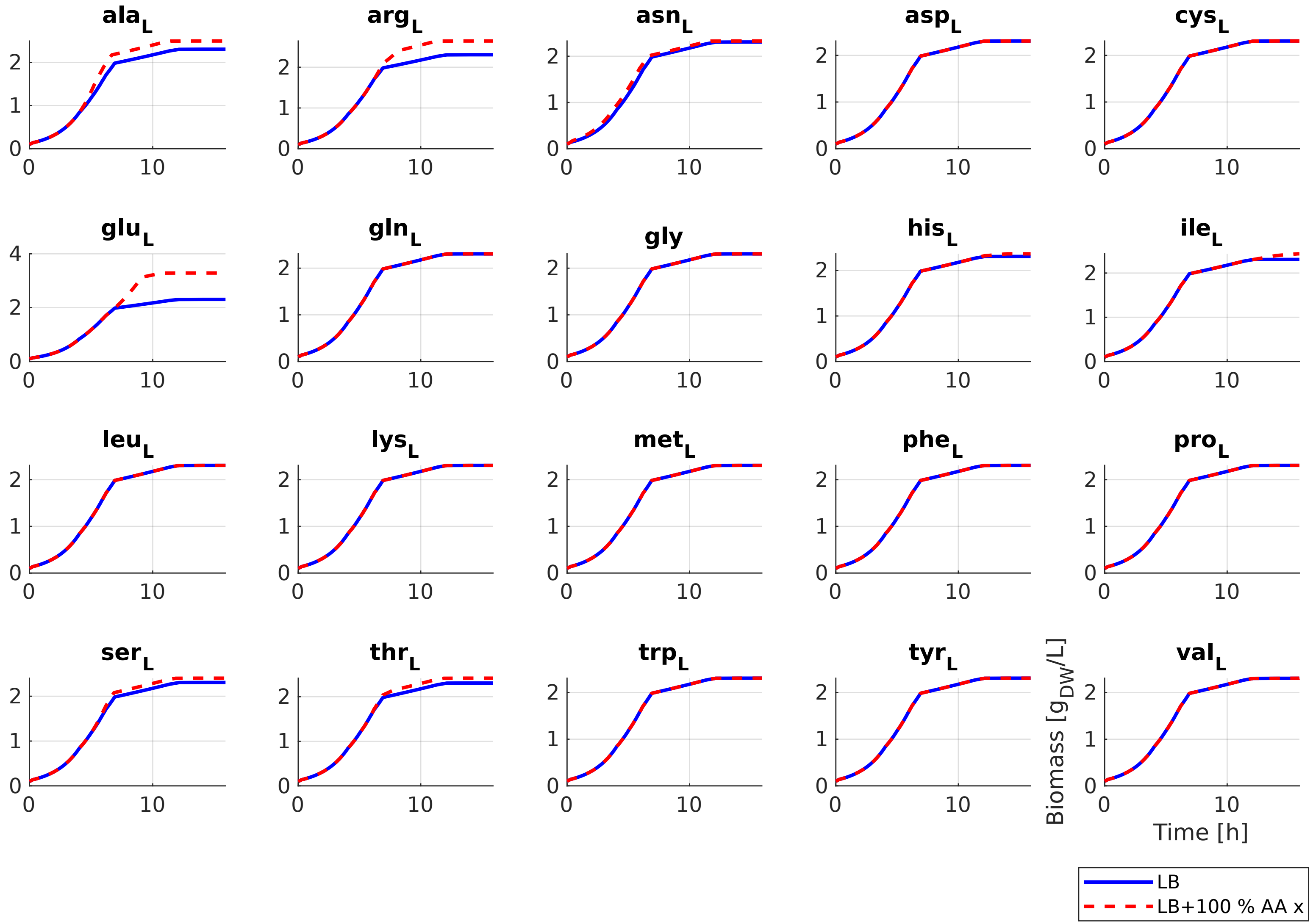


Figure 4.9: Effect of doubling the content of one specific AA on the biomass time profile. These biomass profiles are obtained by sdFBA analyses for media compositions in which the content of the AA mentioned above each graph was doubled (in red) with respect to the original LB media composition in Table 3.3 (in blue).

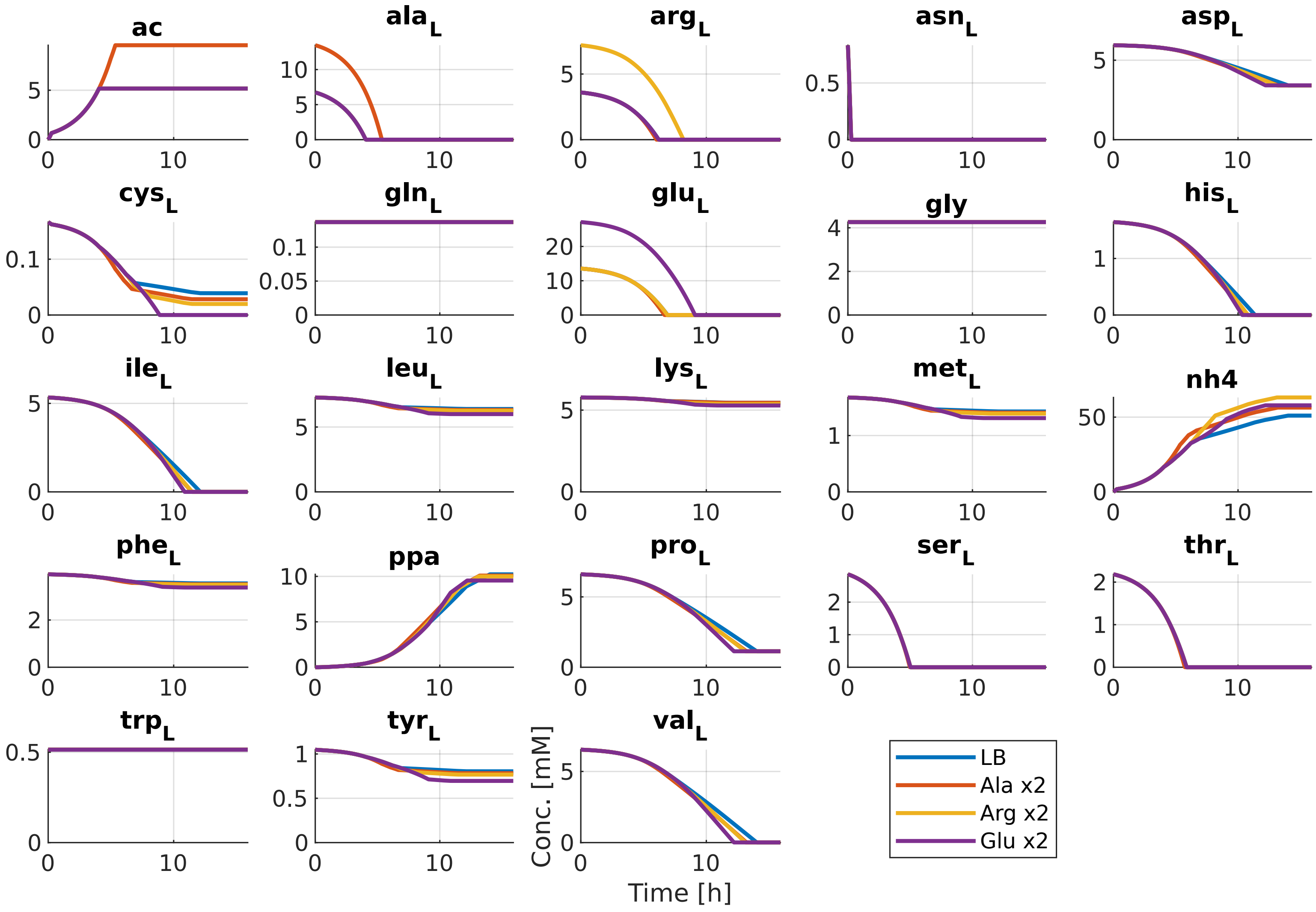


Figure 4.10: Effect of doubling the content of one specific AA on the product metabolite time profiles. These metabolite time profiles are obtained by sdFBA analyses for media compositions in which the content of one AA was doubled with respect to the original LB case with a composition as in Table 3.3. Only AA with a major positive impact on the ammonia productivity were taken into account and compared with the original LB case (ALA, ARG, GLU).

4.3 Conclusion

This chapter evaluated two GSMN models at their ability to reproduce flux values reported in literature for a couple of defined media by comparing TCA flux maps. The basic models, i.e. the form in which they can be found in databases, did not produce satisfactory similarities. While iYO844 showed qualitative agreements for the activated pathways and the produced metabolites, the quantitative flux values were not in line. The predictions by model iBsu1103 were unacceptable, for both the utilised pathways and the secreted metabolites. This underperformance might be attributable to the relaxed underlying metabolic network, because model iBsu1103 relies predominantly on the inherent reaction thermodynamics, neglecting any other biological aspect that might prohibit a reaction to take on a certain flux.
The GECKO model enhancement concept was applied as an attempt to ameliorate the prediction power of both models using empirical proteomic data. Model iYO844, hereafter called iYO844-Gk, did give satisfactory results, while model iBsu1103 remained off-target and consequently was abandoned.
In order to validate its predictive power for genetic modifications, a couple of single deletions were simulated using model iYO844-Gk via the MOMA method and compared to experimental ^{13}C-flux measurements. The model showed a good correspondence with NRMSE values ranging as low as 9 %. Consequently, iYO844-Gk was selected as the basis for further analyses.

A sensitivity analysis of the medium composition was carried out in two steps. In order to capture the evolution over time as well, sdFBA modelling was applied. First, three media were defined with AA compositions characteristic for a certain commercially available protein source, supplemented with two popular complex laboratory protein media. This analysis learned that amino-rich AA positively impact ammonia release and that, despite their lower protein content, plants provide the best mix of AA for ammonia secretion. Next, a more thorough sensitivity analysis was conducted by doubling the content of all AAs one by one. It appeared that not only amino-rich AAs (ARG, ASN, HIS) favour ammonia secretion, but also the ones that are preferentially consumed (GLU, ALA, SER, THR). The ammonia productivity thus depends on both the number of amino groups of the AAs and the rate at which the AAs are catabolised.

Chapter 5

In silico network optimisation

In the previous chapter, the GSMN model iYO844 with a GECKO enhancement for the central carbon subsystem has been shown to perform well for both natural and perturbed genetic environments. In this chapter, two different optimisation algorithms are used to suggest some genetic modifications in the carbohydrate and the amino acid subsystems to optimise the metabolic network for ammonia secretion. By using two algorithms with a distinct working principle, the final set of suggestions is expected to be more diverse.

5.1 Optimisation using knockouts only

A first step is to consider knockouts only for practical reasons: it is the simplest and most used genetic modification method and consequently is more easy to apply because of the numerous available protocols. Predicting metabolic knockout responses *in silico* is straightforward as well because knocked out fluxes are absent, no matter what the regulatory or capacity preconditions are [43].
The applied optimisation algorithms are sequential scanning and OptGene, about which more information is available in Sections 2.2.5 and 3.1. Both offer a high flexibility to adjust the optimisation objective, as this only involves modifying a fitness scoring or selection function. Both are altered to optimise the ammonium-biomass yield via MOMA with a growth rate threshold, which was arbitrarily set at 25 % of the wild-type rate. More settings were described in Section 3.1.

To maximise nitrogen recovery, the bioconversion is likely to be executed in a dynamical environment such as a batch operation. The choice of the growth phase to optimise is thus crucial, as each phase has a specific metabolism with a certain uptake flux distribution. From Figure 4.6, it is observed that most ammonia (about 70 %) is produced during the second major growth phase, the phase in which only ASN is depleted (0.5 h - 7 h). Given that the sdFBA procedure automatically updates the uptake flux constraints of the GSMN model according to the availability of the substrates, the model state of the second growth phase is the starting point of both algorithms. Consequently, the growth rate threshold is 0.1196 h^{-1}.
It is, however, possible that a knockout that optimises one growth phase, could render previous ones infeasible. This is the case if, for example, the knockout is

a reaction that is not essential for the growth phase at hand, but actually is for an earlier one, making the optimised growth phase unreachable in time for batch operations. The feasibility of the selected knockouts is therefore checked *a posteriori* via a dMOMA simulation (cfr. Section 2.2.4).
With an *in vivo* implementation in mind, it is interesting as well to have a look at the genomic properties of the knockout candidates in databases like SubtiWiki and BsubCyc [41, 65]. Isozymes, for example, complicate the deletion process as they are translated from different genes, but catalyse the same reaction [71]. Consequently, each of these genes must be deleted for a complete knockout. It would thus be convenient to remove isozymes from the selection for now. Another tricky situation are genes within a multi-gene operon. Although there is no problem in theory, if such a knockout fails in such a way that the reading frame is shifted, the entire operon is disrupted (a.k.a. a frameshift or polar mutation [71]). If it is not clear whether a frameshift has occurred, the observed phenotype may be falsely attributed to the envisaged deletion. Attention is thus required when implementing such genes *in vivo*.

Taking a look at the summarised results in Table 5.1, yields are substantially higher than the wild-type yield, while the productivity is only slightly higher. The growth rates, however, have reduced to values slightly higher than the growth rate threshold (0.1196 h^{-1}). No isozymes were found.
Comparing the number of deletions, it is remarkable that from two deletions onwards additional knockouts result in only very minor yield increases. It is questionable if such small differences would be noticeable in *in vivo* implementations because of natural noise or model errors. Consequently, in light of minimal efforts and simple designs, the less deletions to achieve approximately the same, the better. By this reasoning, knocking out *ywlF* and *pycA* would be the preference. As mentioned before, a dMOMA check was executed and showed no feasibility issues for this genotype. The rerouting of fluxes visibly results in less biomass formation and, hence, a slower bioconversion with only a minor increase in ammonia productivity. Some new excretion products such as acetoacetate and 6-phosphate-gluconate, are formed as well, as depicted in Figure 5.3.

The optimal knockout combinations are mainly situated in the same metabolic subsystems, the pentose phosphate pathway and the anaplerotic reactions, because these have a large influence on the growth rate. These subsystems each have their metabolic function in this particular situation, which is illustrated for the wild-type in Figure 5.1. The anaplerotic reactions serve as bypass between the TCA cycle and the glycolysis, as citrate synthese (CS) and pyruvate dehydrogenase (PDH) both are unidirectional and consequently block direct access to the glycolysis. Moreover, these pathways enable reaching acetate kinase, the main acetate production flux. The pentose phosphate pathway on the other hand is located at the other end of the glycolysis, and provides ribose for a.o. DNA- and RNA-synthesis.
Intervening in these two pathways affects the entire carbon household and indirectly the adjacent amino acid metabolism, using only a small number of knockouts. Figure 5.2, for example, illustrates the impact of deleting *ywlF* and *pycA* on the

central carbon fluxes in comparison with the wild-type in Figure 5.1. Adjusting the amino acid metabolism directly would require a higher number of deletions because it comprises many reactions that are specialised in converting only one or a few of the twenty amino acids. Nevertheless, some of them have potential if they intervene in the metabolism of a preferential amino acid.

The main conclusion of this analysis is that yield increases mainly come at the expense of growth rates. Although a low growth rate does not imply that it is impossible to achieve those yields, it is rather impractical. (Fed-)Batch systems face a prolonged batch time because of the modest conversion rates due to a lower biomass level. Moreover, continuous operation is limited by the wash-out phenomenon, meaning that the biomass does not replicate fast enough to compensate for the rate at which it is withdrawn via the bioreactor effluent. A retention system such as a membrane bioreactor would eliminate this constraint, but this would increase the required capital investments. On the other hand, to overcome this issue, I would like to propose the idea of inducible knockouts using metabolic toggle switches [72], which could disable the deletions when the biomass level is critically low. The idea is that this would allow the biomass to recover in a wild-type growth mode until it would have reached again a sufficiently high level. At that point, the knockouts could retrigger the high-yield production mode without excessive biomass formation. This activatable production mode could also limit the lengthening of batch conversion times, while still maintaining relatively low levels of biomass, which could cut the costs of handling the waste biomass. This could be done by switching from growth mode to production mode at a certain time point during the batch, which could be subjected to a temporal optimisation [73].

As an illustration, Figure 5.3 depicts not only the dMOMA profiles that were used as a feasibility check, but also the wild-type sdFBA case and three knockout induction scenarios. The inductions were modelled by replacing the sdFBA loop with a dMOMA loop starting from the specified induction time point.
Depending on a cost analysis between lengthening conversion times and handling more waste biomass, the induction time point can be shifted along the time interval of the second major growth phase[1]. An induction at 4 h, for example, is a mid-range solution, in which the batch conversion time lengthens 45 % yielding 38 % less biomass.

Another strategy is altering the optimisation target of the algorithms to find better genotypes. These alterations include increasing the growth rate threshold – e.g. 75 % of the wild-type – or optimising the ammonia secretion flux instead of the yield, as the latter is strongly inflated by low growth rates.
In Table 5.2, both alternate objectives are applied for the case of two deletions using both sequential scanning and OptGene. Firstly, optimising the flux has made an interesting genotype appear just outside the top 5: ΔytsJΔpckA. It is the only genotype in the top 250 (!) with a predicted growth rate higher than half the wild-type rate. On top of that, its predicted ammonium secretion flux is still 8 % higher, which is only 2 % lower than the highest value in this analysis.

---

[1] It would not make sense to trigger the optimised metabolism outside its growth phase.

Secondly, increasing the growth rate threshold returns no interesting genotypes, as it concerns growth rate reductions of about 20 % and ammonium flux increases of only about 2 %. These deletion patterns seem quite ambiguous. The ammonium secretion rates have not increased enough to classify it as a substantial productivity improvement nor the growth rates have decreased in such a way that inducible knockouts could be beneficial.

The general conclusion is that an optimisation with only genetic deletions is not very powerful. It quickly reduces to choosing between high ammonium fluxes and a low growth rate, possibly with the additional complexity of inducible knockouts to cope with that; or moderate growth rates with only a moderate increase in ammonium productivity. Therefore, the candidate set should be enlarged with another type of genetic modification to profit from more combinatorial freedom.

Table 5.1: Summary of the best yields obtained by each optimisation algorithm for an increasing number of deletions. Gene and subsystem data from the iYO844 model or SubtiWiki.

Number of knockouts	**Deleted genes**	**Affected metabolic subsystems**	μ ($\mathbf{h^{-1}}$)	**Flux** $\left(\frac{mmol}{g_{DW}h}\right)$	**Yield** $\left(\frac{g_{\mathrm{NH_4^+}}}{g_{DW}}\right)$
wild-type			0.478	10.130	0.381
Sequential scanning					
1 deletion	lysA	Lysine & threonine	0.123	10.171	1.489
2 deletions	ywlF pycA	Pentose phosphate pathway Anaplerotic reactions	0.120	11.024	1.650
3 deletions	ywlF pycA metK	Pentose phosphate pathway Anaplerotic reactions Methionine & cysteine	0.120	11.024	1.654
4 deletions	lysA gndA serC ytsJ	Lysine & threonine Pentose phosphate pathway Serine biosynthesis Anaplerotic reactions	0.123	11.313	1.656
OptGene					
1 deletion	lysA	Lysine & threonine	0.123	10.171	1.489
2 deletions	ywjH gndA	Pentose phosphate pathway Pentose phosphate pathway	0.123	11.002	1.610
3 deletions	ywlF pckA pps	Pentose phosphate pathway Anaplerotic reactions Glycolysis/Gluconeogenesis?	0.120	11.002	1.652
4 deletions	alsD gndA serC ytsJ	Acetoin metabolism Pentose phosphate pathway Serine biosynthesis Anaplerotic reactions	0.123	11.347	1.661

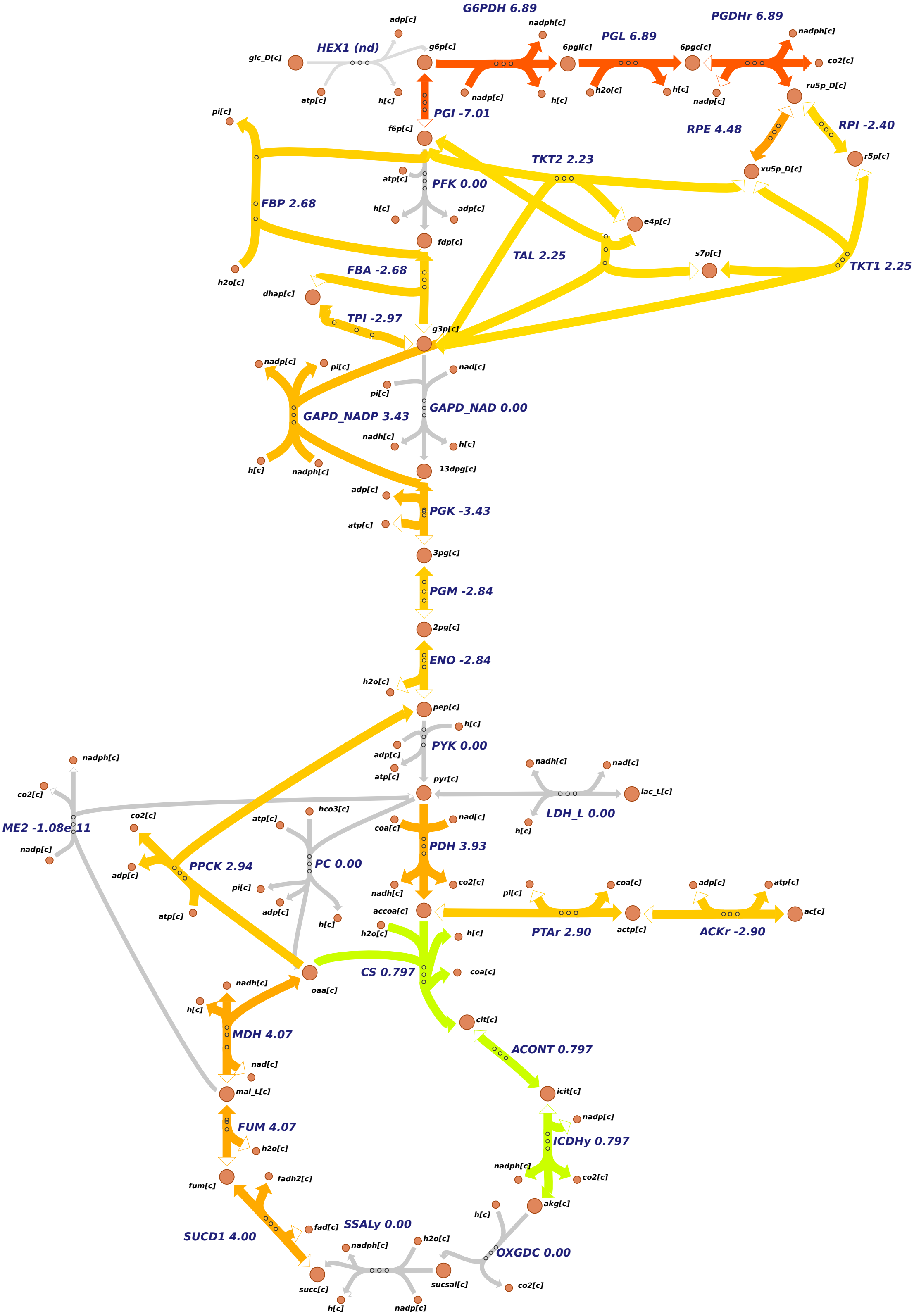


Figure 5.1: Visualisation of the central carbon fluxes for the second major growth phase of wild-type LB growth. Mind that the amino acid metabolism is connected to the central carbon metabolism via many entry points, as shown in Figure 2.1. Figure made with Escher flux mapping [60], including logarithmic colour scale (green to red; 0.1 to 10). Reactions were labelled similarly as in the iYO844 model and numbers are in units of mmol/(g_{DW}h).

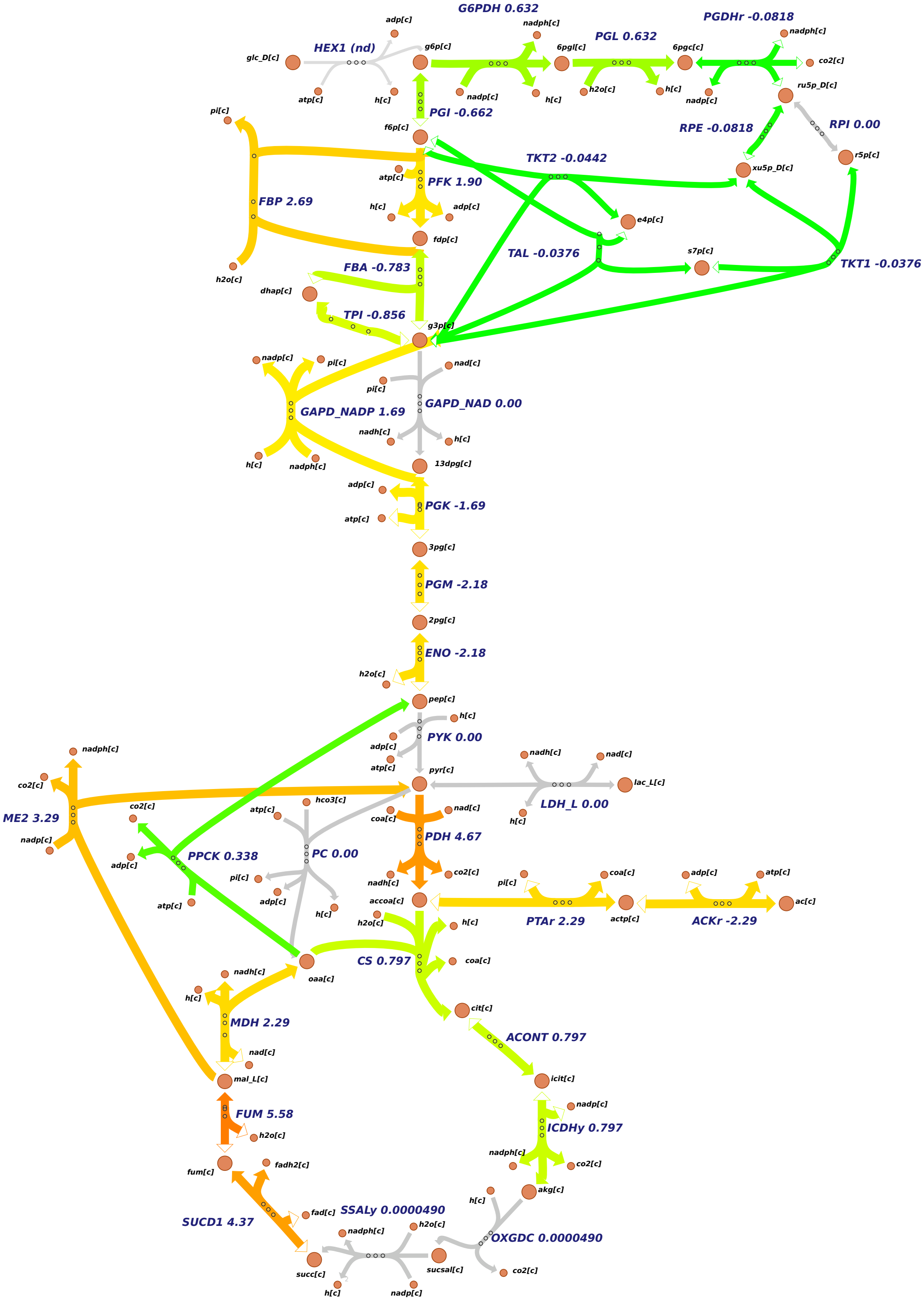


Figure 5.2: Visualisation of the central carbon fluxes for the second major growth phase of ΔywlFΔpycA LB growth. Using only two deletions, the central carbon household is largely rearranged. Mind that the amino acid metabolism is connected to the central carbon metabolism via many entry points, as shown in Figure 2.1. Consequently, these adjacent fluxes are impacted as well. Figure made with Escher flux mapping [60], including logarithmic colour scale (green to red; 0.1 to 10). Reactions were labelled similarly as in the iYO844 model and numbers are in units of mmol/(g_{DW}h).

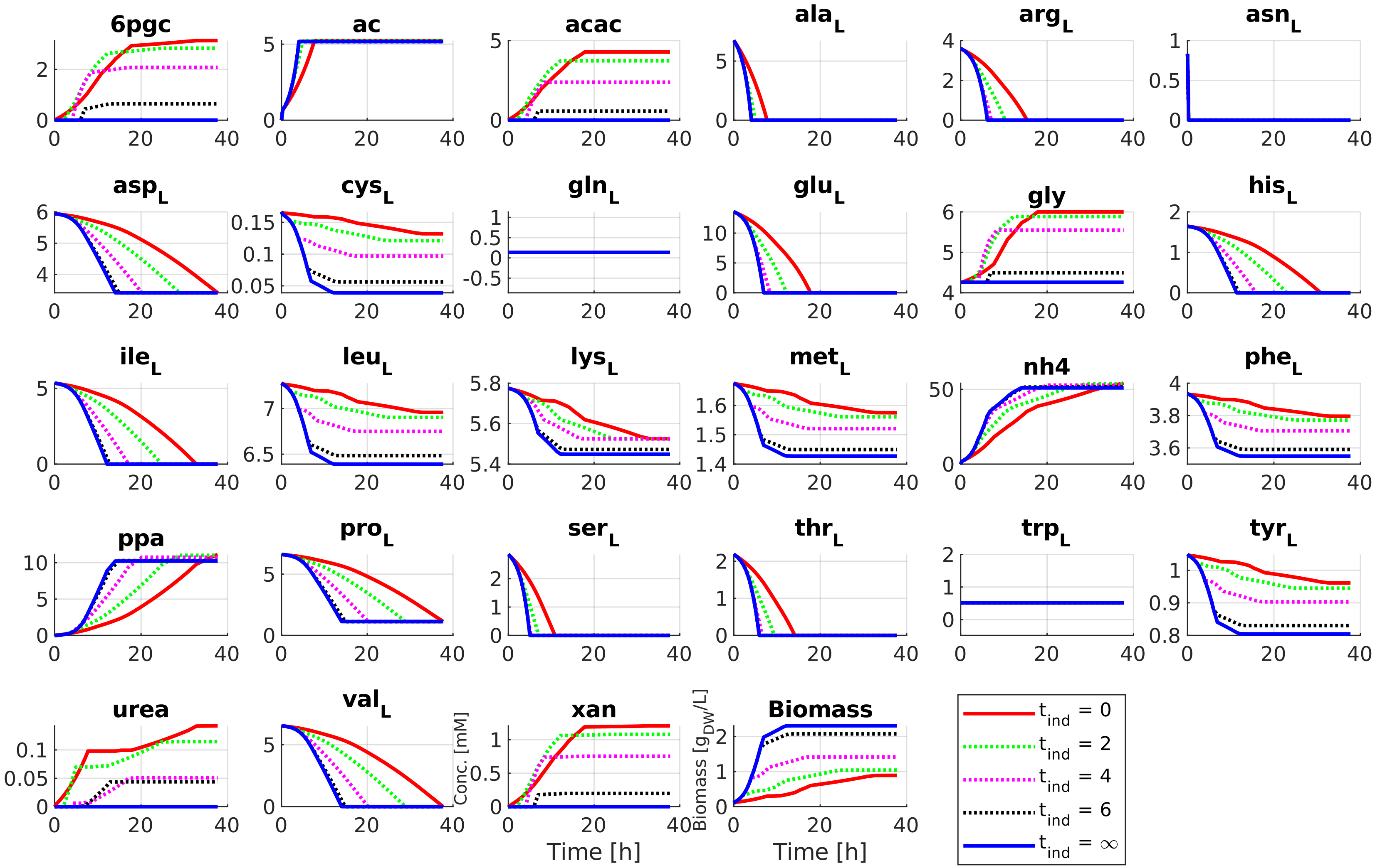


Figure 5.3: Metabolite and biomass time profiles for LB growth and a varying induction time point for the *ywlF* and *pycA* deletions. In comparison with the wild-type, some new excretion products are observed: acetoacetate (acac), 6-phopshate-gluconate (6pgc), urea and xanthine (xan). t_{ind} is the time point in h at which *ywlF* and *pycA* knockouts are induced, which is modelled by switching from sdFBA loops to dMOMA loops from t_{ind} onwards. $t_{ind} = 0$ thus reflects a permanent knockout and $t_{ind} = \infty$ the wild-type. Observe that delaying the knockout induction increases the biomass formation and decreases the batch time, while the final ammonium concentration remains approximately constant.

Table 5.2: Summary of the best five genotypes with two deletions according to the indicated altered objectives as obtained by both algorithms. Gene and subsystem data from SubtiWiki or the iYO844 model. An *S* in the column *Algorithm* means that the genotype was discovered by the sequential scanning algorithm, while an *O* denotes OptGene.

Deleted genes	**Affected metabolic subsystems**	μ ($\mathbf{h}^{-1}$)	**Flux** $\left(\frac{mmol}{g_{DW}h}\right)$	**Yield** $\left(\frac{g_{\mathrm{NH_4}^+}}{g_{DW}}\right)$	**Algorithm**
wild-type		0.478	10.130	0.381	
Growth rate threshold = 75 % of wild-type					
ureABC sucCD	Arginine degradation TCA cycle	0.364	10.755	0.532	S
sucCD pckA	TCA cycle Anaplerotic reactions	0.359	10.378	0.521	S
ald ycgM	Alanine & serine degradation Proline degradation	0.367	10.257	0.504	S, O
ackA pckA	Acetate Anaplerotic reactions	0.375	10.400	0.500	S
pta pckA	Acetate Anaplerotic reactions	0.375	10.400	0.500	S
Target = ammonium secretion flux					
serC gndA	Serine biosynthesis Pentose phosphate pathway	0.156	11.176	1.292	S, O
ywlF pckA	Pentose phosphate pathway Anaplerotic reactions	0.123	11.038	1.616	S
ywlF pycA	Pentose phosphate pathway Anaplerotic reactions	0.120	11.024	1.650	S
ywjH gndA	Pentose phosphate pathway Pentose phosphate pathway	0.123	11.002	1.611	S
lysA gndA	Lysine & threonine Pentose phosphate pathway	0.123	10.993	1.609	S
ytsJ pckA	Anaplerotic reactions Anaplerotic reactions	0.253	10.968	0.781	S

5.2 Optimisation using knockouts and upregulations

Next to knockouts, another type of genetic modification are upregulations, a forced overexpression of a gene. As already described in Section 2.2.4, modelling these is not that straightforward. Because iYO844-Gk does not contain any intracellular flux constraint (except for the GECKO constraints), the forced flux approach of Wang et al. [45] is the sole remaining possibility. Hence, upregulations are modelled by fixing both the lower and upper flux constraint of the affected reactions at 1.5 times the wild-type reaction flux as determined by pFBA. This small amplification factor was chosen to limit influences beyond the scope of metabolic network modelling that could thwart flux increases, such as the regulatory network or enzymatic capacities.

The OptGene algorithm only supports knockouts and hence is not suited for this analysis. In contrast, the sequential scanning algorithm could be modified to combine both knockouts and upregulations. More details about this final version can be found in Appendix C. Consequently, the genetic modifications suggested in this section, are formulated by the scanning algorithm. The further setup of this analysis, including aspects such as the model state, the optimisation objective and the growth rate threshold, is identical to the original case of previous section.

Table 5.3 presents the genotype ranking as obtained by the sequential scanning algorithm using both knockouts and upregulations. Accepting gene amplifications as one of two potential genetic edits has not resulted in many more further yield increases neither revealed a genotype with an acceptable growth rate. Only the couple (PdhD, UreABC) has a higher yield than the already known best double-knockout couple (ΔywlF, ΔpycA), which comes second in this ranking. Nevertheless, (PdhD, UreABC) exhibits a quite substantial flux increase caused by the overexpression of urease, which cleaves urea into two ammonia and one CO_2 molecules. The urea is released from ARG to form ornithine, which is converted to GLU [41, 74]. The overexpression of urease thus creates an additional driving force to cleave urea into ammonia (and CO_2). In Section 4.2, it appeared that the ammonium productivity is highly sensitive to the ARG share of the medium. Intensifying one of its degradation reactions indicates again that the arginine metabolism would be an interesting subsystem to exploit.
An attempt to discover more two-edit genotypes by applying the alternative optimisation objectives of Section 5.1 was fruitless as a mixture of genotypes from Table 5.2 and Table 5.3 emerged (data not shown).

Introducing three genetic edits[2] exploits the arginine metabolism further as the best genotypes involve more reactions from this subsystem, as summarised in Table 5.3. Although the growth rates are close to the threshold of 25 %, the secretion rates are high in comparison with what was encountered so far, which confirms the potential of this subsystem. However, the issue of low growth rates remains.

---

[2] To limit computational work, loop selection was set tighter, i.e. only the top 1 % of genotypes was transferred to the next loop instead of 10 % when incrementing the number of genetic edits.

Table 5.3: Summary of the five genotypes with the highest ammonium-biomass yield as obtained by the sequential scanning algorithm for one, two and three genetic edits. Edits can be both deletions and amplifications. Applying alternate optimisation criteria for two edits resulted in a mix of combinations from this table and Table 5.2. Urease and the adjacent arginine metabolism show a high potential, although the issue of low growth rates remains. Gene and subsystem data from SubtiWiki or the iYO844 model.

Number of edits	**Genetic edits**	**Affected metabolic subsystems**	μ **(h^{-1})**	**Flux** $\left(\frac{mmol}{g_{DW}h}\right)$	**Yield** $\left(\frac{g_{\mathrm{NH_4^+}}}{g_{DW}}\right)$
wild-type			0.478	10.130	0.381
1	PdhD	Central carbon	0.120	10.697	1.608
	ΔlysA	Lysine & threonine	0.123	10.171	1.489
	ΔaspB	Aspartate biosynthesis	0.156	10.267	1.182
	ΔgndA	Pentose phosphate pathway	0.176	10.849	1.113
	SdhABC	Central carbon	0.199	10.590	0.959
2	PdhD UreABC	Central carbon Arginine degradation	0.123	11.787	1.725
	ΔywlF ΔpycA	Pentose phosphate pathway Anaplerotic reactions	0.120	11.024	1.650
	ΔywlF PdhABC	Pentose phosphate pathway Central carbon	0.123	11.052	1.624
	ΔywlF ΔpckA	Pentose phosphate pathway Anaplerotic reactions	0.123	11.038	1.616
	ΔgndA ΔlysA	Pentose phosphate pathway Lysine & threonine	0.123	10.993	1.609
3	ΔpycA PdhD UreABC	Anaplerotic reactions Central carbon Arginine degradation	0.120	11.774	1.764
	PckA PdhD UreABC	Anaplerotic reactions Central carbon Arginine degradation	0.121	11.838	1.760
	RocF PdhD UreABC	Arginine degradation Central carbon Arginine degradation	0.124	12.119	1.755
	RocA PdhD UreABC	Arginine degradation Central carbon Arginine degradation	0.122	11.794	1.746
	RocD PdhD UreABC	Arginine degradation Central carbon Arginine degradation	0.122	11.794	1.743

5.3 Conclusion

This chapter employed two optimisation algorithms with a distinct working principle and a high flexibility towards altering the optimisation objective, which allowed to set a custom target. The applied objective was to determine the genotype with the highest ammonia-biomass yield as predicted by the MOMA method, using deletions only and taking an arbitrary growth rate threshold of 25 % of the wild-type into account. Furthermore, the state of the uptake flux constraints during the second major growth phase was selected as the starting point of the algorithms to maximise the impact of the optimisation.

The results of this analysis learned that the best genotypes act on the pentose phosphate pathway and/or the anaplerotic reactions in order to have a large impact with few deletions. More than two deletions, however, do not result in substantial improvements. Yield increases mainly come at the cost of lower growth rates, while the ammonia secretion flux has improved only slightly. This, however, implies that the same ammonia productivity can be achieved using less biomass, which could cut the costs of waste biomass handling. However, a lower biomass level also entails a slower bioconversion and, hence, prolonged conversion times for batch operations and more difficulties with wash-out for continuous operations.

A metabolic toggle switch was proposed to limit the impact of these drawbacks, while still achieving low biomass levels. It would allow to toggle between a wild-type high-growth mode and a mutant high-yield production mode with a low growth rate, while still maintaining the same ammonia secretion rate. A trade-off between lower conversion rates and more waste biomass would be possible instead of a dilemma.

Another strategy was altering the optimisation objective of both algorithms in order to sieve out better genotypes, e.g. by increasing the growth rate threshold or by optimising for the flux instead of the yield. These adjustments uncovered a remarkable deletion pattern, ΔytsJΔpckA, which has a predicted increase in ammonium excretion of 8 % and a growth rate of more than half the wild-type rate, which is remarkably higher than the other encountered genotypes.

However, optimising this GSMN with only deletions quickly leads to the dilemma between a high production flux with a low growth rate, and a moderate production flux and growth rate.

Therefore, genetic upregulations were included in the edit candidate set to widen the searching space. The objective alterations were also applied for this case. These analyses learned that urease, which is part of the arginine degradation pathway, is an important ammonia generator and intensifies its release when amplified.

Incrementing the number of edits to three confirmed the potential of the arginine subsystem, as the best genotypes with three edits involved an additional reaction from this subsystem and were predicted to secrete ammonium at a high rate. The growth rates, however, were again low because this genotype was discovered using a 25 % growth rate threshold.

Chapter 6

In vitro strain analysis

In the previous chapter, a couple of algorithms were applied to suggest some sets of genotypes with a high potential for ammonia secretion using a GSMN, model enhancements and media compositions sourced from literature. In this chapter, some exploratory *in vitro* runs with a wild-type *B. subtilis* strain are conducted. Rather than determining parameters, the focus of this *in vitro* strain analysis is on qualitatively assessing the behaviour of the model.

Nevertheless, these observations could allow to add or modify some model aspects. There are a couple of interesting refinements possible for which empirical data must be gathered. Firstly, oxygen was previously assumed to be abundantly present due to perfect aeration and mixing. In reality, aeration rates are limited causing the dissolved oxygen level of an operating bioreactor to drop to concentrations that are possibly throttling the desired aerobic metabolism. Therefore, the oxygen balance of the bioreactor system must be taken into account and linked to the sdFBA procedure to enable the model to react on oxygen scarcity. More specifically, this can be done by directly integrating an oxygen mass balance with a transport term. This term, however, requires an empirical oxygen transfer coefficient.
Secondly, the model now assumes that the organism secretes sufficient proteases so that all amino acids are readily degradable in a negligible time period. Measuring the concentrations of the free amino acids in the reactor broth could allow to assess whether the polypeptide hydrolysis is rate-determining.
Thirdly, it is possible that one would observe metabolic products that were not predicted by the model or reported in literature, or vice versa. Shutting the appropriate exchange sinks and/or introducing new metabolic reactions would then be the method to update the model.

This chapter first presents and discusses the obtained experimental data. From these, some modifications of the model are then proposed, after which they are implemented and qualitatively assessed for their impact.

6.1 General data overview and discussion

In total, three experimental runs using LB medium were executed, of which the technical practicalities are described in Chapter 3. The first run was rather exploratory and did not yield that much offline data as it was not known which offline sampling frequency would be adequate. The online sensors, however, were always active and did generate useful data. The next two runs utilised a two-hour offline sampling scheme by day. In order to sample the entire time horizon with these runs, the bioreactor was inoculated alternately in the morning and in the evening.
Figures 6.1, 6.2 and 6.3 display time profiles of relevant process variables for all three experimental runs such as metabolite concentrations, dry weights, online and offline OD data, the calculated OTR profile, the DO measurements, and the net proton balance. All data starting from the inoculation of the reactor is depicted, while the end of the lag phase is indicated with a dashed vertical line. The event with which this time point corresponds, is included in Table 6.1. More details about the data processing and the definition of the process variables can be consulted in Section 3.2.

All runs start with a smooth start-up to a first exponential phase that lasts about 5 h. In this phase, the biomass growth rate is the highest and the main metabolic product is acetic acid, which is corroborated by the decrease of the proton balance. Meanwhile, a similar molar amount of ammonia is secreted. Because this amount of basic ammonia is only slightly out of balance with the amount of acetic acid, changes in pH and consequently the net proton balance remain limited. The ammonia productivity in this growth phase, however, is relatively low: only about 25 % of the total ammonium is secreted in this period.
Small amounts of saccharides such as fructose, glucose and sucrose, and organic acids like formate, lactate and succinate have been detected in the initial medium (data not shown). Because their concentration amounts at most to some tens of mg/L, their influence is considered negligible.

At the end of the first phase, the DO quickly rises, the OTR drops and the online OD stagnates for about 2 h, indicating a metabolic shift. The fact that *B. subtilis* continues to grow in a polypeptide medium, confirms that it is able to secrete an adequate amount of proteases. After the stagnation, the oxygen consumption quickly rises again, reaches a maximum 10 h later and stays there for another 5 h. There is, however, a relatively small dip around 30 h of run time that recurs every run.
Meanwhile, the logarithmic online OD curve has passed through multiple linear segments with step-wise slowly decreasing slopes. The hypothesis is that this reflects a depletion of a single amino acid, similarly as in the simulations of Chapter 4. Note that it is not always clear when a linear segment starts or ends. The depletion of an amino acid would cause a gradual change in metabolism to one with only a slightly lower growth rate, which results in a flattening growth curve with smooth slope changes in which the ends of linear segments can be difficult to distinguish.

The metabolite data of run 3 indicate that acetic acid is taken up again. Combined with the continued secretion of ammonia, this results in a sharp increase of the net proton balance because of the increased addition of H_2SO_4 for pH control. Remarkably, the organism produces starting from the second growth phase isobutyrate and isovalerate as major products next to acetate. These compounds were, to my knowledge, never explicitly mentioned before in literature, either because they were effectively absent or because they were not the focus of the analysis. They can be produced from VAL and LEU degradation respectively, according to the KEGG pathway database [74]. Lastly, about 45 % of the total ammonium is secreted during this period.

The last part of the runs is characterised by a sharp decrease of the OTR and similar increase of the DO at around 35 h, which coincides with the depletion of acetate. Also note that the DW and the OD have started to decline, which indicates that the organism has reached its stationary phase. Nevertheless, it is still active as the OTR is not zero and the isobutyrate, isovalerate and ammonium concentrations continue to increase. The remaining 30 % of the final ammonium concentration (of run 3) is produced during this stationary phase. The net ammonium productivity now amounts to about 40 mM.
However, the question rises what the organism is consuming as the acetate has just been depleted. One possible explanation could be that endogenous maintenance is occurring and that the biomass is digesting itself. The famous sporulation process of *B. subtilis* could actually contribute to this. Unfavourable conditions such as nutrient scarcity trigger the cell to divide into a mother cell and an endospore, which eventually results in the spore that can survive a long time. When the endospore is ready, the mother cell releases the endospore by lysing itself [75]. The remnants of the mother cell could be a nutrient source for the remaining active biomass. Endogenous maintenance and sporulation thus allow to regain some of the nitrogen which would otherwise be lost to waste biomass.

Run 3, however, deviated from the previous ones, as illustrated in Figure 6.4. The DO did not stagnate completely after the first growth phase and the DO has a lower minimal value, namely 63 % instead of 75 %. Additionally, the OD rose faster. The profiles appear somewhat squeezed in time in comparison with the other two. For example, while the OTR drop because of the depletion of acetate occurs around 35 h in the first two runs, it already happens at 30 h in the third run. As a result, the alternating inoculation strategy to sample the entire time horizon of a reactor run has not fully succeeded and some gaps remain.
A possible cause of this divergence is that the biomass of the inoculum was not that constant as expected from the method described in Section 3.2. A higher amount of active biomass implies that more oxygen will be consumed, which reflects in a lower DO. Another factor is the variation in composition that is inherent to complex protein media, and consequently more preferential AA could have been present.

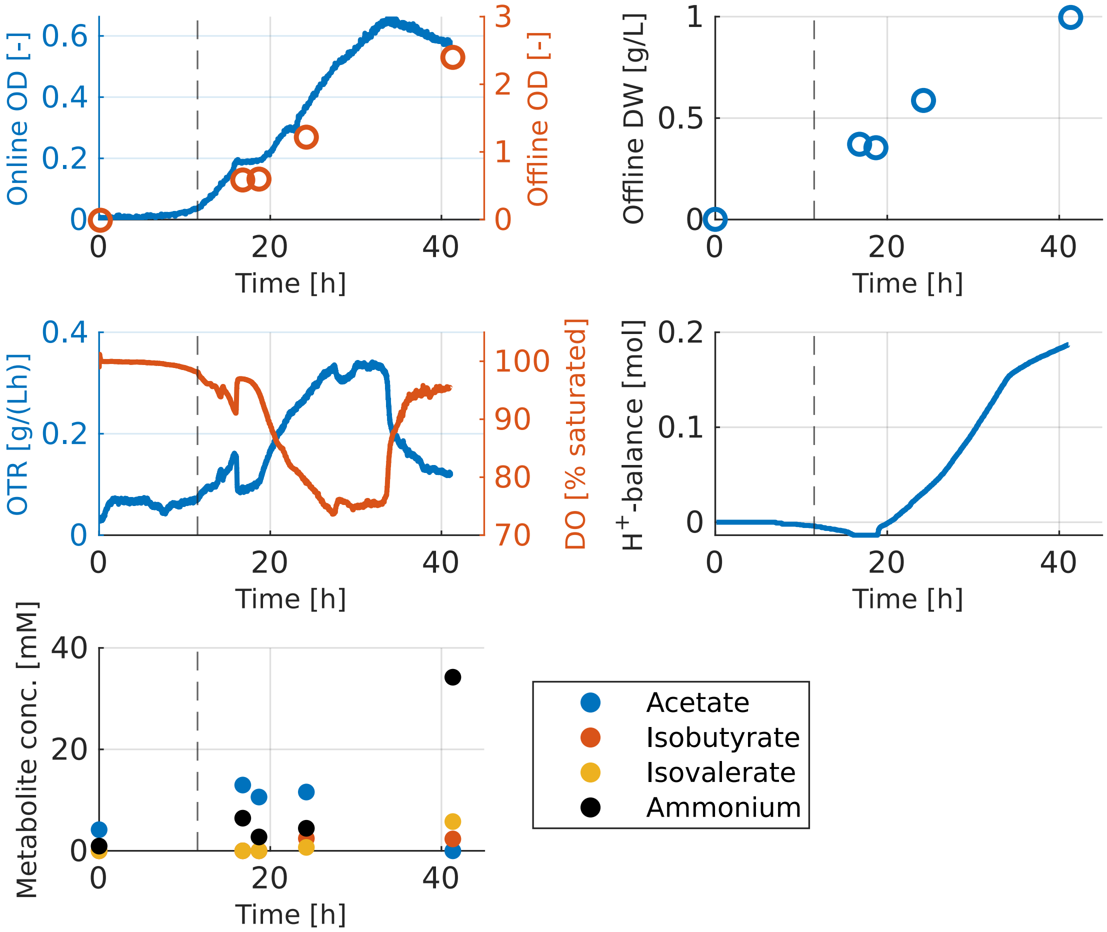


Figure 6.1: Overview figure of experimental LB run 1. The figures display the online and offline OD, the biomass, the OTR and the DO, the net proton balance and the metabolite time profiles for wild-type *B. subtilis* 168 in LB medium. The end of the lag phase is marked with a dashed line.

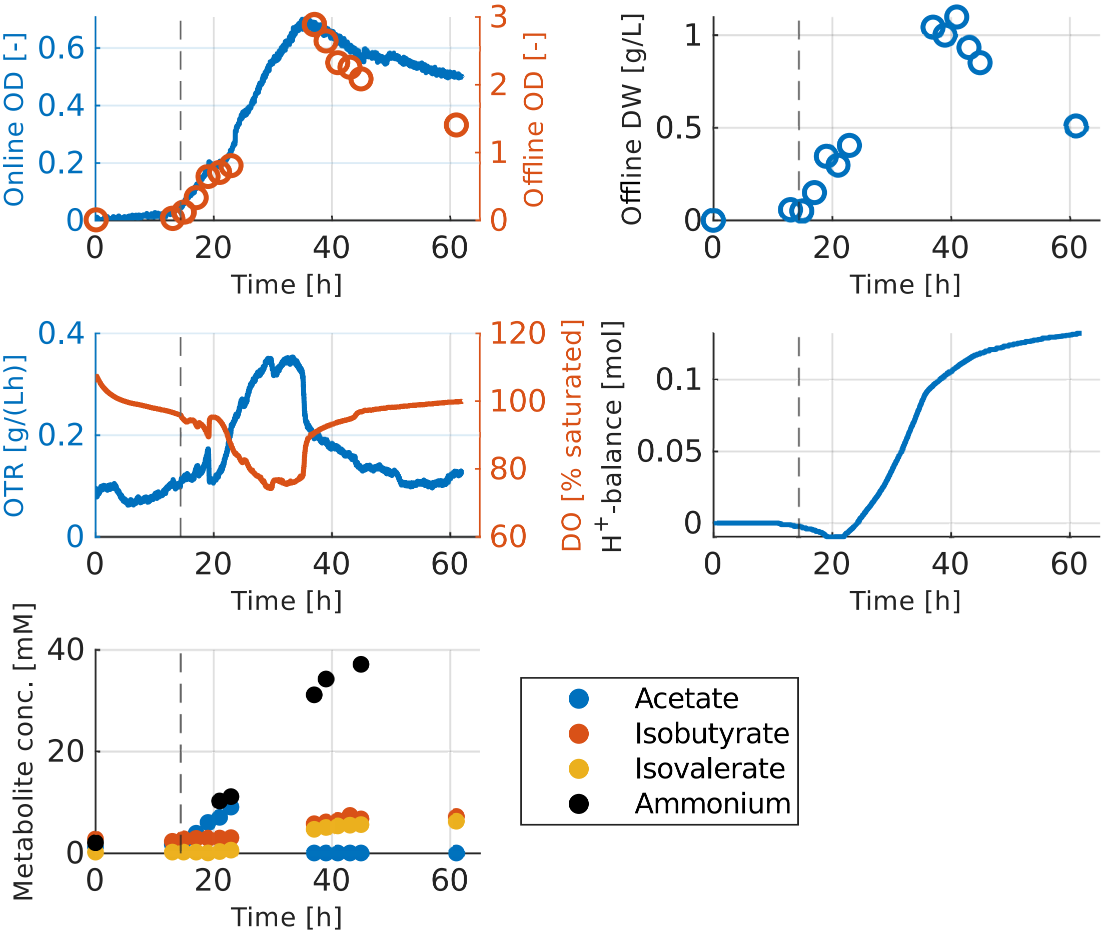


Figure 6.2: Overview figure of experimental LB run 2. The figures display the online and offline OD, the biomass, the OTR and the DO, the net proton balance and the metabolite time profiles for wild-type *B. subtilis* 168 in LB medium. The end of the lag phase is marked with a dashed line.

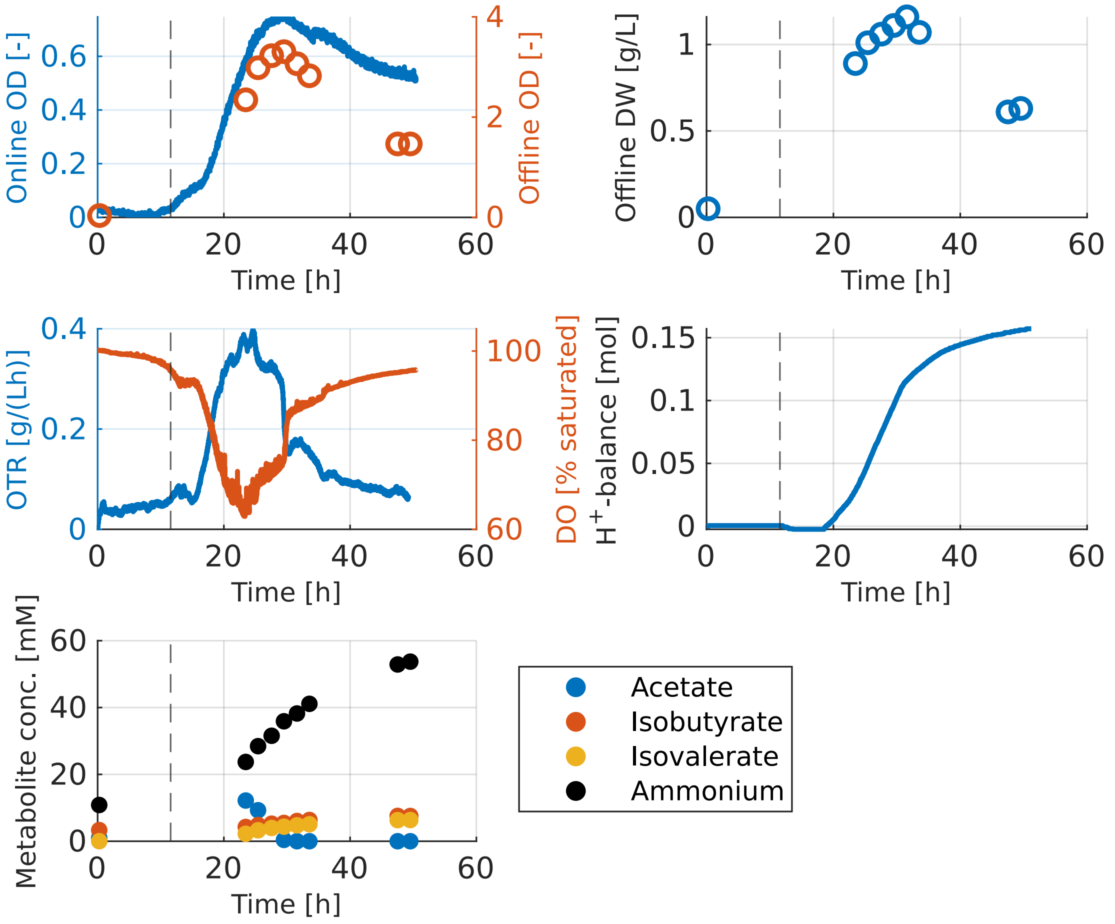
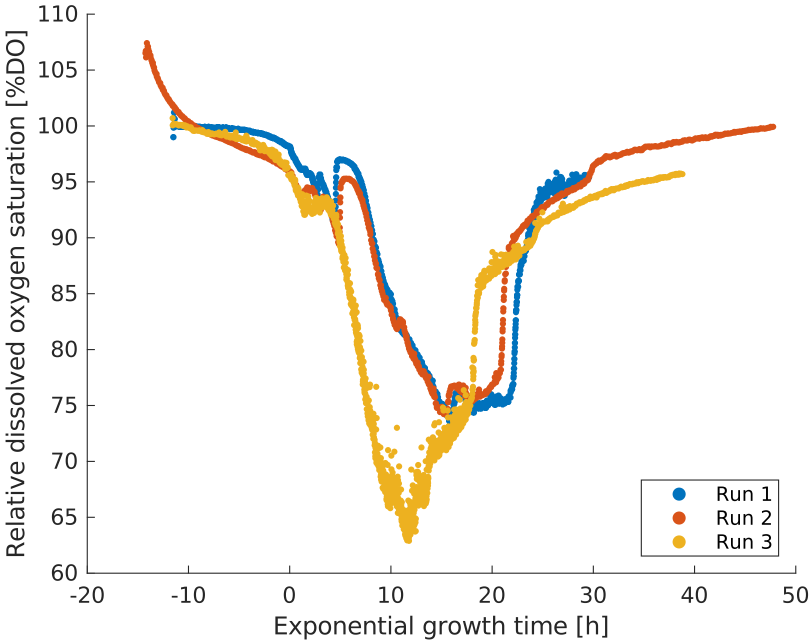
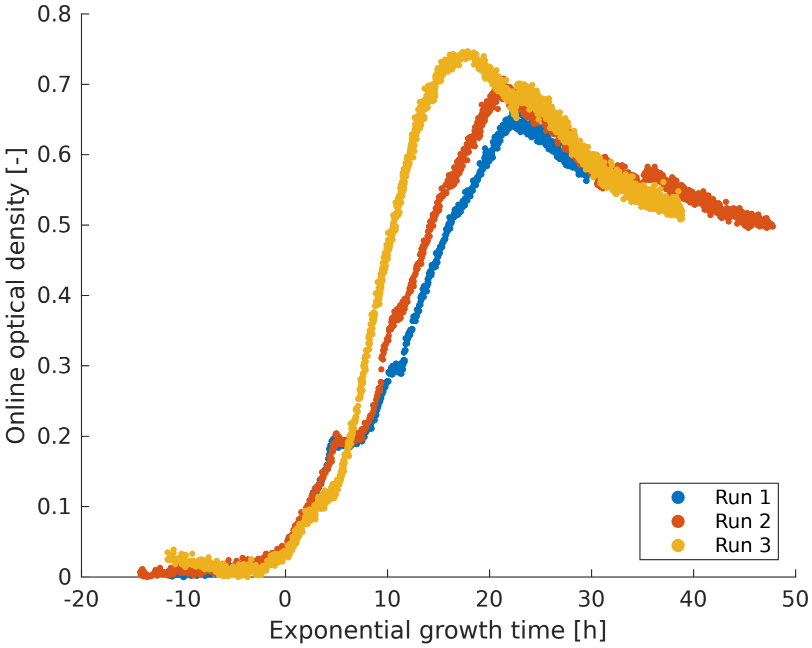


Figure 6.3: Overview figure of experimental LB run 3. The figures display the online and offline OD, the biomass, the OTR and the DO, the net proton balance and the metabolite time profiles for wild-type *B. subtilis* 168 in LB medium. The end of the lag phase is marked with a dashed line.
The initial dry weight is here obtained differently than in the other runs. Instead of determining the suspended mass of a volume of preculture that was equal to the volume used for preparing the inoculum, a sample was taken from the freshly inoculated bioreactor. Due to the low biomass concentration in a freshly inoculated bioreactor, this measurement is not considered reliable.



(a) Dissolved oxygen (DO)



(b) Online OD

Figure 6.4: Comparative illustration of the deviating behaviour of reactor run 3. For all three reactor runs, the DO profiles and the online OD data are depicted. The graphs were shifted in time so that the end of the lag phase was set at the origin. Note that the run 1 and run 2 are very comparable, while run 3 appears squeezed in the time dimension and reaches more extreme values.

Figure 6.5 depicts the measured free amino acid concentrations of samples from all three runs to get an image of the evolution over the entire time window. It appears that proteolysis occurred during the lag phase as the free amino acid concentrations had increased, while during the first exponential phase, the concentrations gradually decreased due to consumption. Around 10 h, a metabolic shift occurred, which was accompanied by a new increase of free AA. However, the ALA, PRO and PHE concentrations did not increase anymore, which could indicate other proteases were into play, targeting different AA sequences. Eventually, almost all AA were quasi-depleted when the stationary phase set in at approximately 25 h.

However, another thing to notice is which amino acids were consistently detected. According to the BD Bionutrients® technical manual [62], hereafter called the reference, every amino acid in LB medium should be partially present in a free form, which contrasts with these observations. Therefore, an initial bioreactor sample was hydrolysed prior to the AA analysis to verify the total AA content of the administered LB medium. Figure 6.6 displays the results for both the hydrolysed and non-hydrolysed initial sample next to their reference values. Remarkably, some AA remained undetected, especially glutamate, the one that should be most abundant in the hydrolysed sample. Also, according to the EZ:faast documentation, hydrolysing has its limitations. It would completely destroy CYS and TRP, convert ASN and GLN to ASP and GLU resp., and reduce the SER, THR, TYR and MET contents. Nevertheless, the remaining detected concentrations were in line with the reference, although some values were remarkably low (PRO, ASP). This hints that there might be a detection issue.

Hence, a closer look was taken at the chromatograms. By comparing these of the standards with those of the samples, it appears that some AA, including GLU, indeed were present, but their detector responses were so low that the software of the chromatography apparatus did not recognise them. Only AA with a high detector response for the standard solutions were detected properly throughout all samples, as illustrated in Figure 6.7. Those with lower ones were not reported at all or only for the early samples. Consequently, it is presumed that the sample dilution factor that was applied to avoid oversaturation of the sorbent beads (20x), was too high to assure that the detection limit was met. This could also explain why some values were abnormally low. Measurements close to the detection limit are not considered accurate, although some variation due to the complexity of the medium might also play a role. Due to time limitations, the analysis could not be repeated using a lower dilution factor.

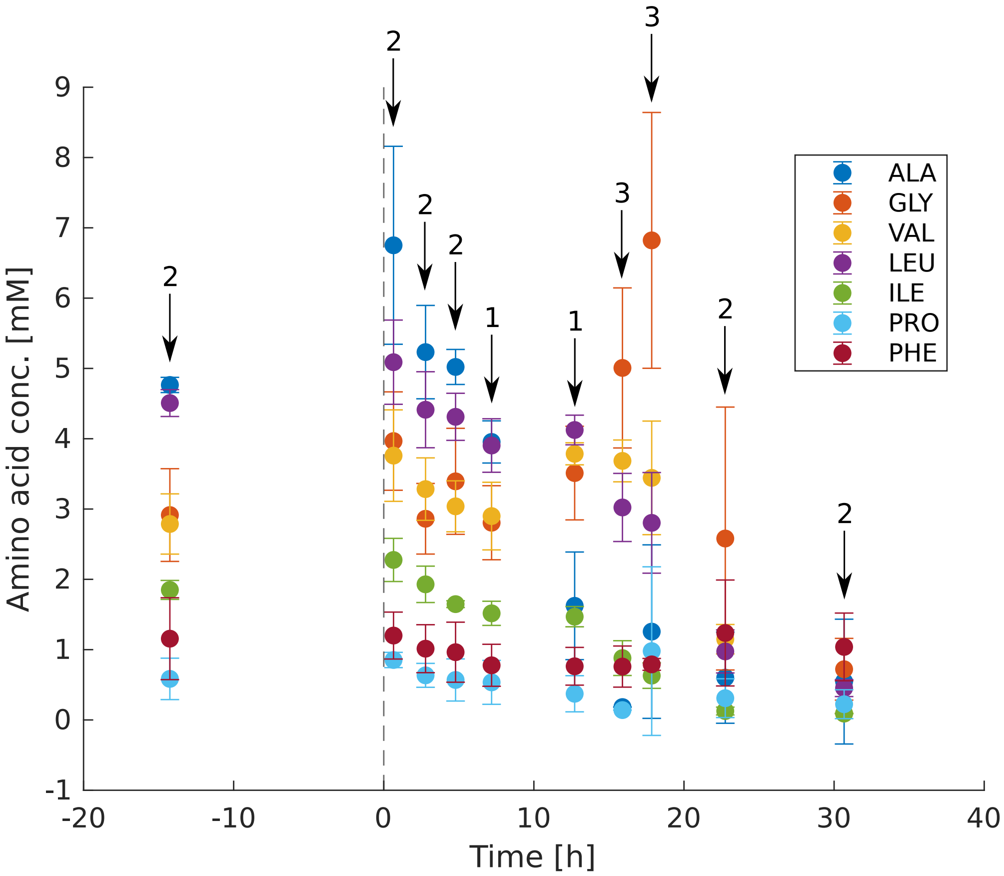
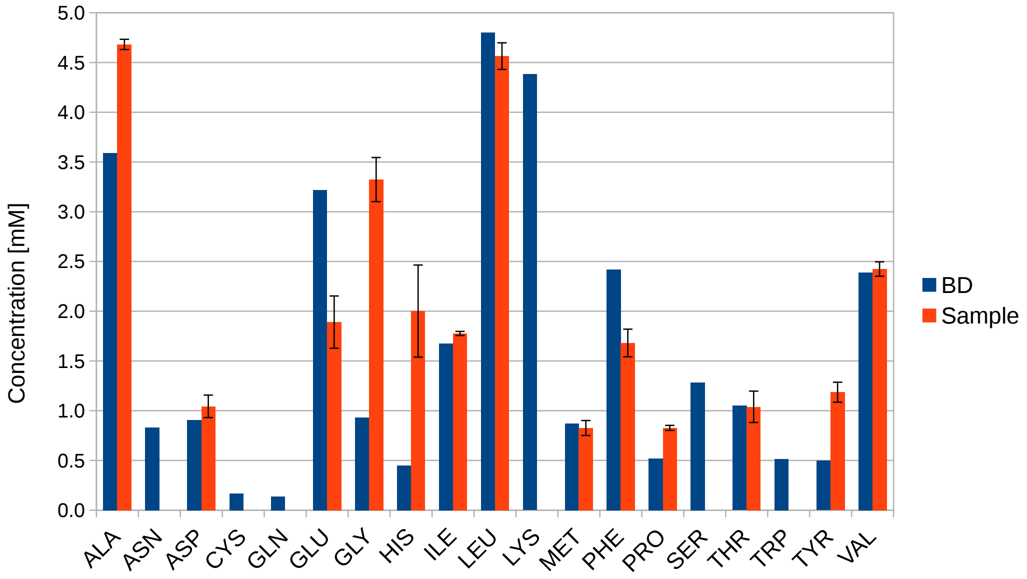
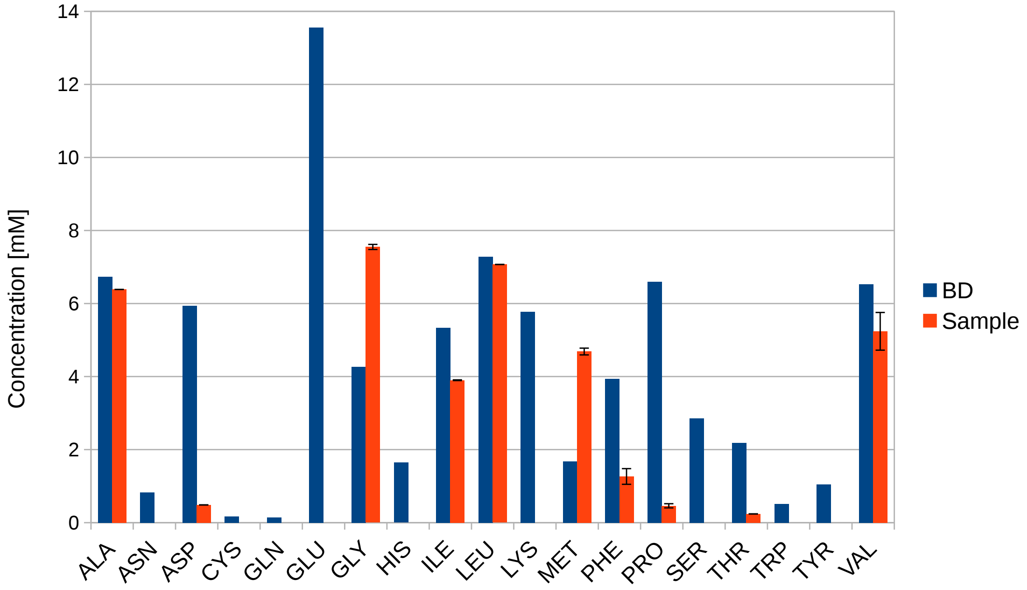


Figure 6.5: Evolution of the free amino acid concentrations over time as composed from samples from all runs. Only AA which were detected in all duplicate measurements of at least one sample, are included. Errorbars correspond with one standard deviation. The run in which the sample was taken, is indicated at the top of the graph. Mind that the lag time observed in each run was subtracted from the time data. Additionally, the time frame of run 3 was artificially stretched by 20 %, which results in a shift of 4 h. This allowed to fit the amino acid data of run 3 into the time window of the other two runs. This value was obtained by comparing the time widths of the DO valley of run 2 and run 3 (25.38 h vs. 21.07 h resp.), defined as that part of the DO curve between the end of the stagnation after the first growth phase and the last minor increase of the DO at the late end of the valley.



(a) Non-hydrolysed



(b) Hydrolysed

Figure 6.6: Empirical free amino acid concentrations for the initial sample of a bioreactor run, next to the reference values of the BD Bionutrients® technical manual [62]. A majority of the AA were fairly in line with the reference values, while others were absent or remarkably low, possibly due to the hydrolysis in case of the hydrolysed sample. From Figure 6.7, it is hypothesised that the dilution factor (20x) could have been too high as well, which resulted in inaccurate measurements on the verge of detection.

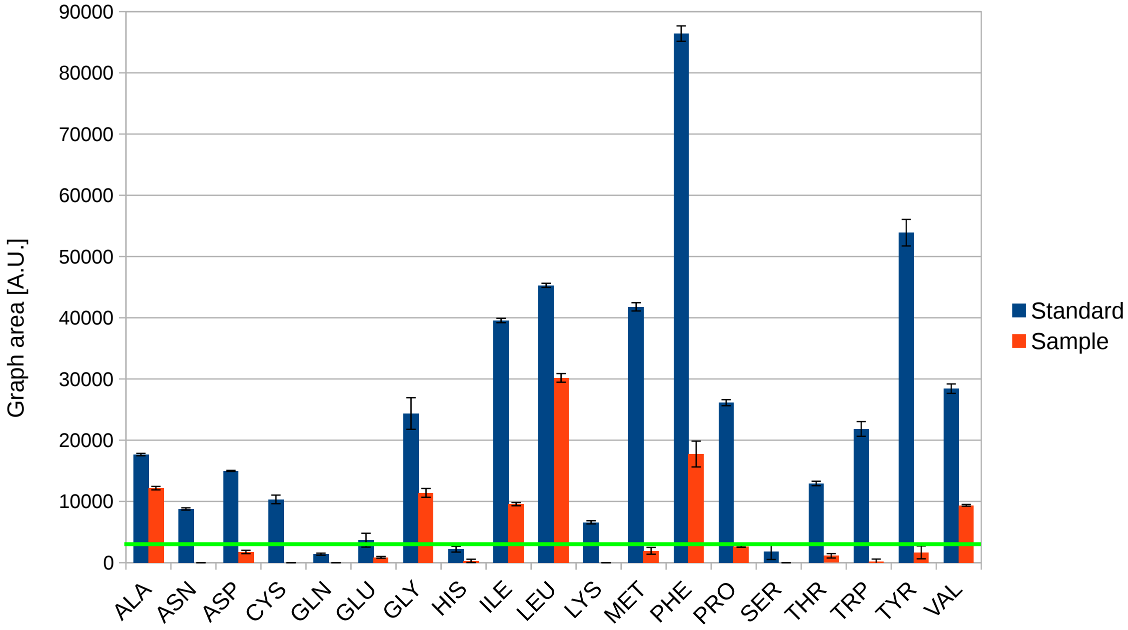


Figure 6.7: Raw detector responses for the 100 μM standard and the initial non-hydrolysed sample. The detector responses are expressed in terms of area under the graph of the detector signal. The green line indicates the lowest response at which an AA still was detected in all replicates of all samples, in this case PRO. Signal strengths around and below this value faced difficulties to be recognised consistently and are considered inaccurate. These observations could be caused by a too high sample dilution (20x), which was applied to avoid sorbent oversaturation. Some dilute standard solutions (50 μM GLN, GLU, and SER) also appear to have detection/accuracy difficulties, which reflects in their calibration curves with R^2-values below 0.8. Note that these deviations coincide with a detector response being lower than the value indicated in green.

6.2 Parameter determination

6.2.1 Growth rates and phases

First of all, it is imperative to verify the linear correlation between the dry weight data and the (calibrated) online OD data, because this is a necessary condition to represent the biomass with the online OD in the growth rate regression, as discussed in Section 3.2. Figure 6.8 proves that the DW and the online OD are linearly correlated with an intercept at zero.

Figure 6.9 displays the characterisation of the growth phases. As already mentioned in previous Section 6.1, the growth curves are very smooth with gradual changes in slope during the metabolic shifts, which complicates setting the time trajectory to determine the growth rate. Therefore, an inspection of plots of several process variables was conducted to detect events which flag a metabolic shift at the beginning or the end of a growth phase. Table 6.1 summarises which events were selected to demarcate these trajectories. A thing to notice here is that these events occur in each run, which confirms that it concerns relevant metabolic phenomena.

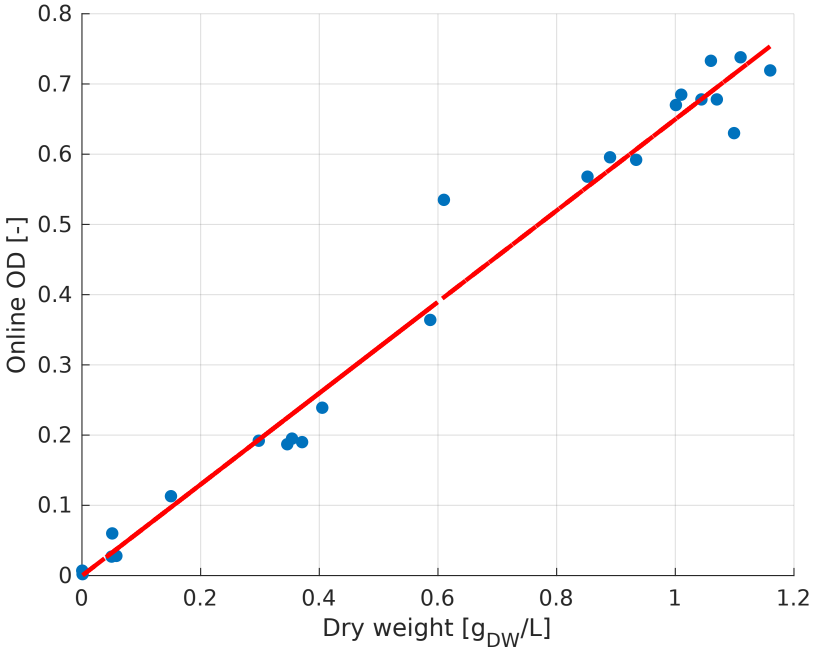
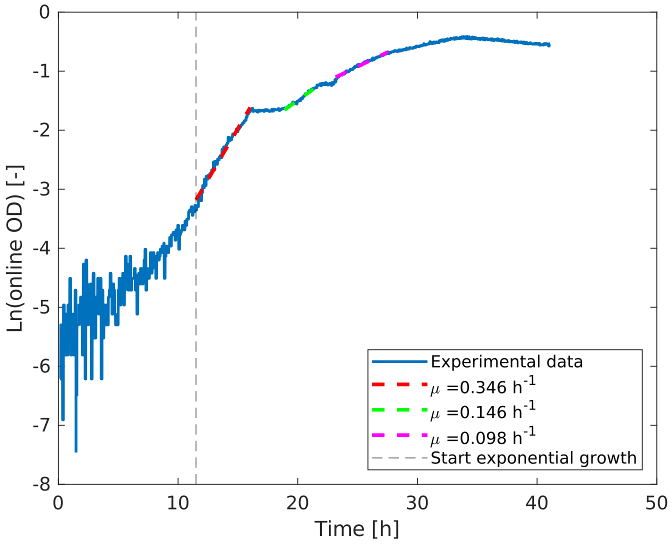
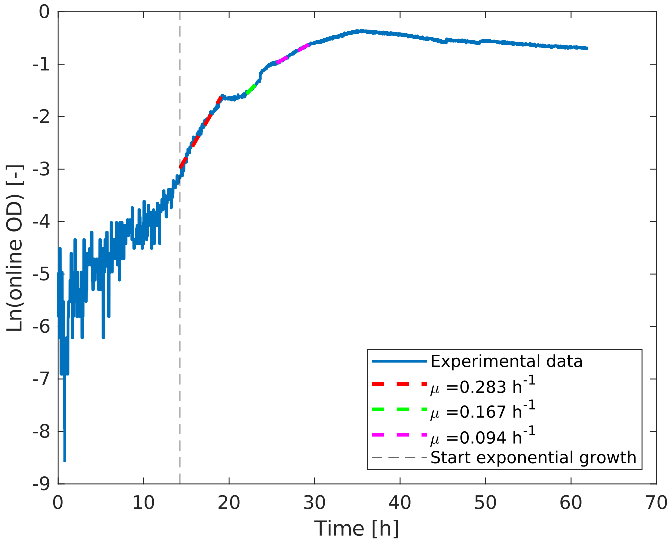


Figure 6.8: Verification of the linear correlation between the dry weight and the (calibrated) online OD. This correlation is a necessary condition for the online OD to represent the biomass in the growth rate regression as discussed in Section 3.2. Because the fitted line has a slope of 0.65 (confidence interval: [0.6145; 0.6855]) and an intercept of -2.95·10^{-4} (confidence interval: [-0.0264; 0.0258]) with R^2 = 0.978, it is accepted that the correlation is linear with the intercept at zero.



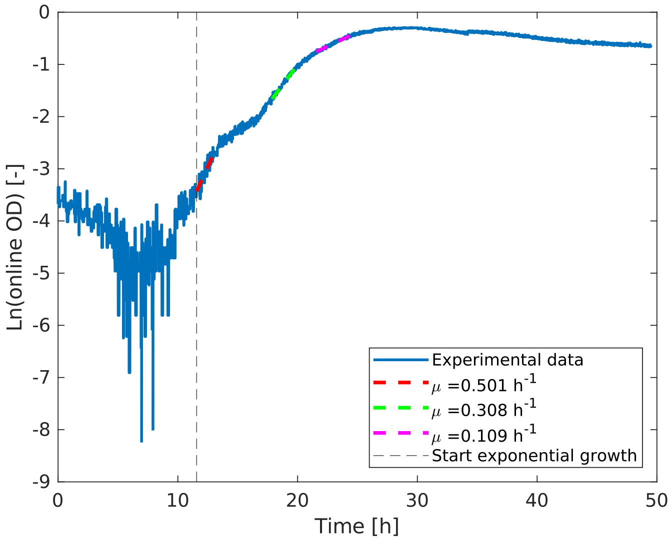
$\boldsymbol{\mu}$ [h^{-1}]	Confidence interval (α = 95 %)
0.346	0.3441 - 0.3483
0.146	0.1442 - 0.1469
0.098	0.0971 - 0.0981

(a) Run 1



$\boldsymbol{\mu}$ [h^{-1}]	Confidence interval (α = 95 %)
0.283	0.2808 - 0.2848
0.167	0.1647 - 0.1691
0.094	0.0939 - 0.0948

(b) Run 2



$\boldsymbol{\mu}$ [h^{-1}]	Confidence interval (α = 95 %)
0.501	0.4840 - 0.5188
0.308	0.3053 - 0.3106
0.109	0.1082 - 0.1099

(c) Run 3

Figure 6.9: Fitted growth rates for each reactor run separately. Note that it is possible that there are more growth phases for OD data higher than 0.6, but these were not fitted because the Lambert-Beer law is not considered valid anymore.

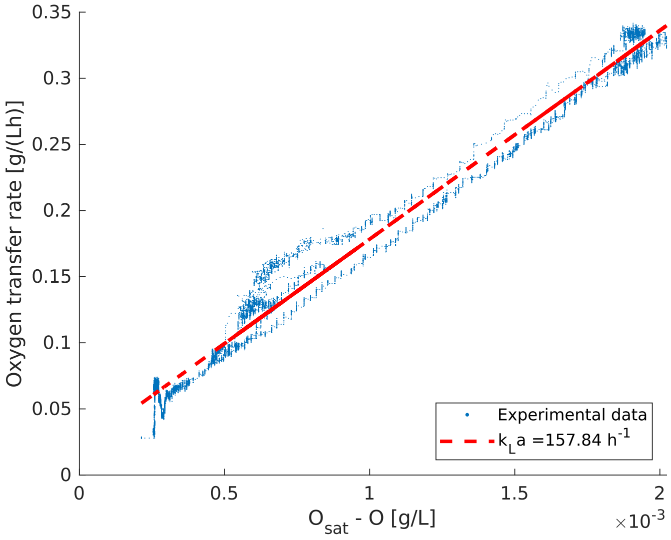
Table 6.1: Summary of the process events which demarcate the ends of the trajectories utilised to fit growth rates. Note that the start points of the first phase of each run equal its lag time.

Run	*Phase*	*Time point [h]*	*Event*
Run 1	Phase 1	11.50	Change in slope of the DO curve
		16.03	Stagnation of the net proton balance
	Phase 2	18.88	Net proton balance starts to increase.
		21.31	OffCO_2 curve starts to deviate from its exponential trajectory, as perceived by taking the logarithm of the data.
	Phase 3	23.16	Bottom of a local offCO_2 peak
		27.55	Top of a local offCO_2 peak
Run 2	Phase 1	14.25	Change in slope of the DO curve
		19.06	Stagnation of the proton balance
	Phase 2	22.05	Net proton balance starts to increase.
		23.63	Disturbance in ln(OD) coinciding with the addition of antifoam
	Phase 3	25.56	Bottom of a local offCO_2 peak
		29.65	Top of a local offCO_2 peak
Run 3	Phase 1	11.57	Net proton balance starts to decrease.
		13.17	Stagnation of the net proton balance
	Phase 2	17.88	Net proton balance starts to increase.
		19.91	OffCO_2 curve starts to deviate from its exponential trajectory, as perceived by taking the logarithm of the data.
	Phase 3	21.57	Bottom of a local offCO_2 peak
		24.57	Top of a local offCO_2 peak

The high number of growth phases has a downside in view of sampling as well. Each phase is characterised by a metabolism with a specific distribution of the available resources over its flux network. To reliably measure these, a fairly high number of measurements is required to limit the width of the confidence intervals. This can be achieved either by a high sampling frequency or by a number of experiment replicates. The alternating offline sampling strategy applied here does provide a view as wide as possible on the entire time horizon of the experiment, but a drawback is that only few data points are within the same growth phase. Characterising the metabolism quantitatively from these data, for example via metabolic yields, entails a high uncertainty and is therefore omitted.

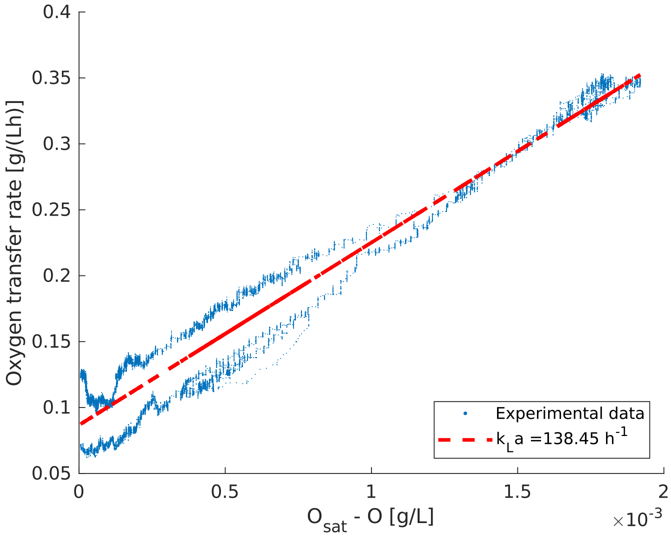
6.2.2 Oxygen transfer coefficient k_La

The oxygen transfer coefficient k_La is determined for each run as described in Section 3.2. First the OTR was calculated from the off-gas composition data, and then plotted against the difference between the oxygen saturation concentration and the measured DO. Consistent values are obtained which range from about 138 to 158 h^{-1}, as depicted in Figure 6.10.



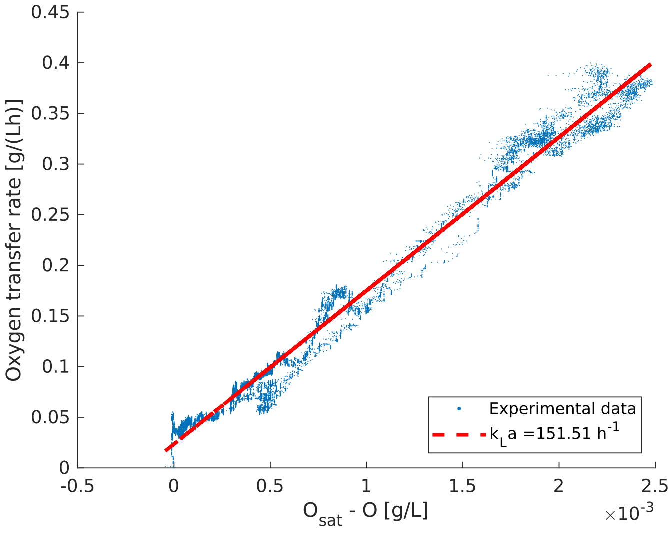
(a) Run 1

Confidence interval ($\alpha = 95$ %)
157.574 - 158.107



(b) Run 2

Confidence interval ($\alpha = 95$ %)
138.066 - 138.839



(c) Run 3

Confidence interval ($\alpha = 95$ %)
151.261 - 151.753

Figure 6.10: Estimated oxygen transfer coefficient k_La as determined for each experimental run. The data are obtained for 1 vvm of aeration with air and a stirrer speed of 500 RPM.

6.3 Aligning the model with the observations

Comparing the simulated profiles of Chapter 4, for example Figures 4.5 and 4.6, and their experimental counterparts, there are a couple of differences.

Firstly, the empirical biomass growth rates are remarkably lower and there are intermittent lag phases which indicate a metabolic shift. In the simulations, it was assumed that all AA were made readily available by an adequate amount of proteases. In fact, there are several types of proteases and peptidases required to completely breakdown a protein into readily degradable amino acids or dipeptides. *B. subtilis* is capable of producing a powerful pallet, but does not employ it fully at once. Consequently, it adapts its metabolism continuously to what is available, resulting in lower growth rates and metabolic shifts. A proper rescaling of flux constraints could be a first step towards a better model.

Secondly, it was observed that acetate was remetabolised in the later stages of the experimental run, which imposes a relaxation of the acetate exchange flux constraint.

Thirdly, an estimate of the oxygen transfer coefficient is available from the previous section. With this, the dissolved oxygen balance can be set up from which a simulated DO profile can be derived and compared with the experimental one.

Fourthly, the experimental strain secreted isobutyric and isovaleric acid, and not the predicted propionic acid. According to SubtiWiki [41] and the KEGG pathway database [74], isobutyric acid and isovaleric acid originate from the degradation of the branched chain amino acids LEU, ILE and VAL. However, both acids are not included in the iYO844-Gk metabolic model, so an extension of the model urges.

6.3.1 Model alterations

The following paragraphs explain how the metabolic model was modified for each aspect mentioned above in order to align it with the experimental observations.

Growth rate Firstly, from the experimental runs it appeared that the biomass at the onset of the exponential phase amounted to approximately 0.05 g_{DW}/L instead of the 0.1 g_{DW}/L assumed during the simulations. Consequently, this simulation parameter was modified. Secondly, it was not clear whether the simulated ASN-mediated high growth rate phase (with $\mu = 1.20\ h^{-1}$) should be present in the experiments. Due to the low ASN content of LB medium, it was unlikely that proteases would have set ASN completely free already in the beginning of a run. Consequently, it is assumed that ASN consumption occurs steadily during the run, but that a separate phase is absent. As a result, the next predicted phase (with $\mu = 0.47\ h^{-1}$) would correspond with the first observed growth phase. To align the simulated growth rates with the observed ones, the uptake flux constraints and the GECKO constraints of the central carbon fluxes were scaled by multiplying them with the ratio of the average growth rate of this first observed phase ($0.315\ h^{-1}$)[1] and the predicted one ($0.47\ h^{-1}$).

---

[1] Only the first two runs were taken into account.

Acetate reuptake The lower flux constraint of the acetate exchange reaction was arbitrarily changed from 0 to -10 mmol/(g_{DW}h) to facilitate the observed acetate reuptake. A check *a posteriori* learned that this constraint is not rate-determining.

Dissolved oxygen balance A dissolved oxygen balance was introduced and integrated analytically as described in Section 3.3 using the average of the three estimated oxygen transfer coefficients from previous section (k_La = 149.3 h^{-1}). The obtained dissolved oxygen concentrations were converted to DO percentages according to Formula (3.5).

Branched chain amino acid degradation Both isobutyric acid and isovaleric acid are branched chain fatty acids (BCFA) and are not included in the iYO844-Gk metabolic model. However, SubtiWiki and KEGG state they are products of the branched chain amino acid (BCAA) catabolism, as illustrated in Figure 6.11 [41, 74]. Propionate on the other hand is the end product of an alternative catabolic pathway for VAL and ILE only, according to KEGG. This also explains why ILE and VAL are taken up more than LEU in the simulations of Chapter 4.

The CoA-bounded form of both BCFA, however, is part of the model and is a metabolite of the long fatty acid synthesis pathways. Consequently, reactions 1 in Figure 6.11 are present in the model, but the other ones of the BCFA pathway (reactions 3) not. Neither KEGG or SubtiWiki elucidates exactly what happens next with the BCFA-CoA: they both only state that they are converted to an unspecified BCFA. Nevertheless, SubtiWiki mentions some enzymes that facilitate these reactions (Ptb, Buk). Both enzymes are known from the butyrate pathway. They exchange the CoA of butyrate-CoA for a phosphate group and, respectively, divert the phosphate to an ADP-molecule, similarly as the acetyl-CoA-acetate pathway (Pta, AckA). The question is whether these could also act on closely resembling molecules such as BCFA-CoA. In order to check that, both enzymes were queried in the BRENDA database [39], but no entries for *B. subtilis* were found. Nevertheless, the available entries confirm that both enzymes show a strong activity for BCFA-CoA substrates, albeit for other microorganisms (*Listeria monocytogenes* for Ptb and *Clostridium acetobutylicum* for Btk).

To conclude, it is opportune to add these reactions to the metabolic model. BCFA secretion is enabled by adding proton-symport exchange reactions, in analogy with butyrate, propionate and acetate secretion. Furthermore, because propionate was not observed, its exchange flux is shut. For clarity, an overview of the added reactions is provided in Table 6.2.

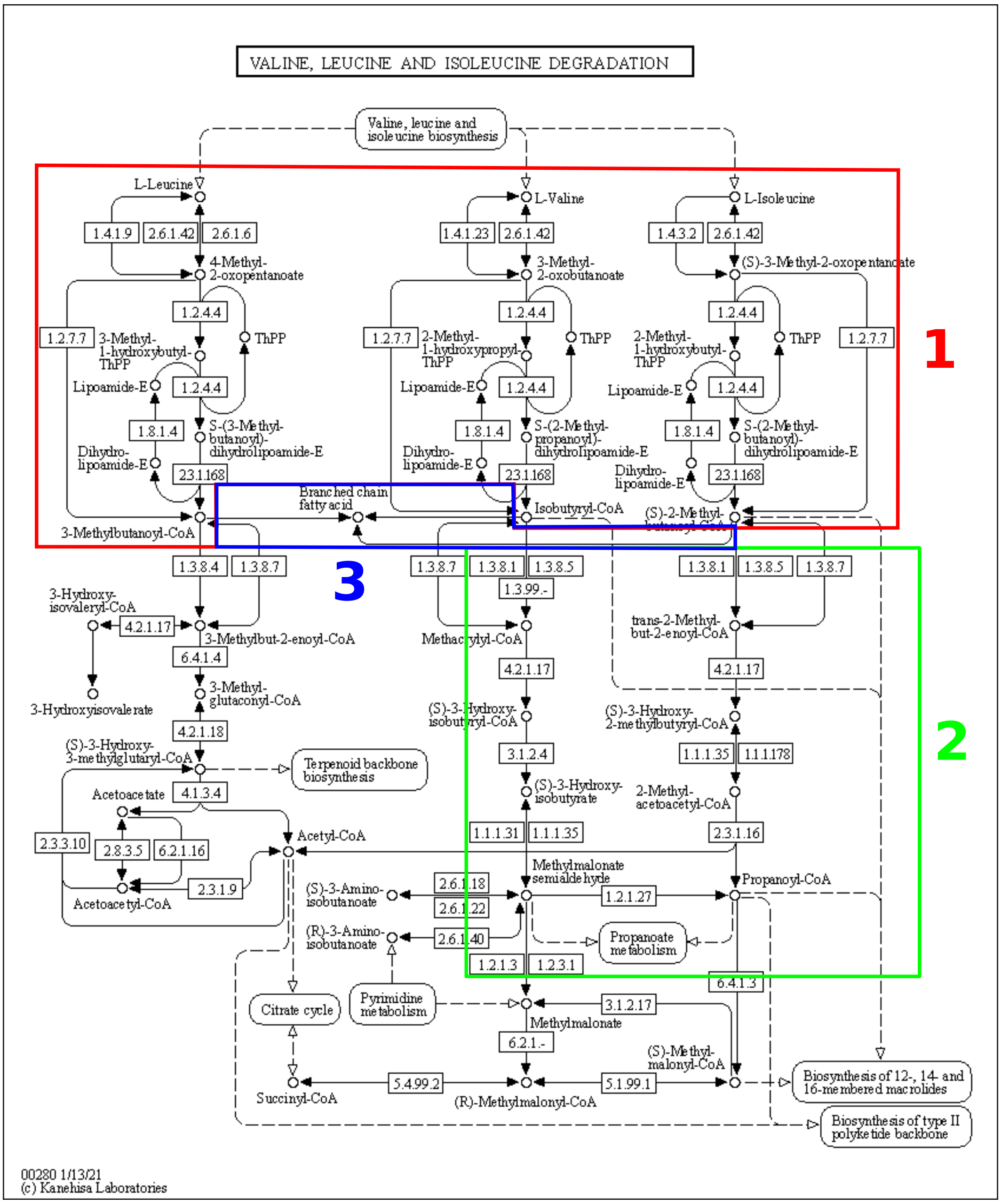


Figure 6.11: BCAA degradation pathways and the reactions of interest. Adapted pathway overview figure taken from the KEGG database [74]. Reactions 1 are already included in model iYO844-Gk as part of the long fatty acid synthesis pathways, while reactions 2 are the existing branch that currently produces propionate. Reactions 3 are missing in the model and thus added here. More information about these reactions is provided in Table 6.2.

Reaction code	***Name***	***Reaction formula***
PIVT	Phosphate isovaleryl transferase	Isovaleryl-CoA + $p_i \leftrightharpoons$ CoA + Isovaleryl phosphate
PIBT	Phosphate isobutyryl transferase	Isobutyryl-CoA + $p_i \leftrightharpoons$ CoA + Isobutyryl phosphate
P2MBT	Phosphate 2-methylbutyryl transferase	2-Methylbutyryl-CoA + $p_i \leftrightharpoons$ CoA + 2-Methylbutyryl phosphate
IVKr	Isovalerate kinase	ADP + Isovaleryl phosphate $\leftrightharpoons$ ATP + Isovalerate
IBKr	Isobutyrate kinase	ADP + Isobutyryl phosphate $\leftrightharpoons$ ATP + Isobutyrate
2MBKr	2-Methylbutyrate kinase	ADP + 2-Methylbutyrate phosphate $\leftrightharpoons$ ATP + 2-Methylbutyrate

Table 6.2: Overview of the reactions that were added to the the metabolic network of model iYO844-Gk. The following six reactions correspond with reactions 3 of Figure 6.11. They are the missing link between the BCAA degradation pathways and the observed isovalerate and isobutyrate secretion. The pathway for 2-methylbutyrate from ILE is added for the sake of completeness. The transferases and the kinases are facilitated by the Ptb and the Buk enzyme respectively. BCFA secretion was enabled by adding proton-symport exchange reactions, in analogy with butyrate, propionate and acetate secretion.

6.3.2 Qualitative validation of the altered model

After executing the model alterations discussed above, an sdFBA procedure yielded the metabolite time profiles in Figures 6.12 and 6.13. At first sight, the profiles seemed more in line with the experimental observations. There was a secretion of isobutyrate (IB) and isovalerate (IV), while acetate was retaken up, and the simulated dissolved oxygen balance provided an additional point of comparison. The following paragraphs qualitatively validate these several aspects one by one.

Firstly, the acetate profile was fairly well reproduced. It was predicted to be exponential, but this lasted up to 13.4 h instead of the observed 10 h. The predicted peak height was close to its empirical counterpart. Although the peak acetate concentration was not directly registered due to the unsuccessful sampling strategy, one could still extrapolate from run 2 and 3 that the peak would be just below 15 mM. The acetate depletion occurred at 26 h virtually and at around 19 h empirically.

Secondly, the ammonium profile was overestimated. However, both followed an exponential trajectory, which levelled off due to a metabolic shift around 13.4 h virtually and 20 h empirically. This turning point is, for example, perceivable in the combined ammonium profiles of run 2 and 3, although the number of data points is not such that a measuring error can be excluded. It is remarkable that 13.4 h in the simulation corresponded with the start of the acetate reuptake, while at 20 h in the experiments the acetate was already depleted. Furthermore, the final ammonium concentration was slightly overpredicted. Run 2 resulted in a net increase of 42 mM and run 3 39 mM, which was both lower than the predicted value of 57 mM. Defining the predicted value as a theoretical 100 % conversion within the capabilities of the microbial metabolic system, the empirical values average to a conversion of 71 %.

Thirdly, it was interesting that the metabolic model now was capable to funnel its BCAA degradation towards BCFA secretion. Nevertheless, the profiles did not agree well with the observed ones. The prediction was that 9.7 mM of IV would be produced via a steadily increasing trajectory that would accelerate at 13.4 h of simulation time, but the empirical trajectory already did so at 8 h of growth time and slowed down around 15 h to end at 6 mM. Furthermore, the IB yield was too low (only 1 mM), while there was an unexpected 4.7 mM of 2-methylbutyrate (2MB). The 2MB pathway was added for the sake of completeness as ILE degradation lacked a BCFA secretion pathway similar to LEU and VAL (see Figure 6.11).
A further analysis of which fluxes were active, learned that from 13.4 h on, the simulated metabolism was transferring VAL via a reversed VAL-aminotransferase and the normal LEU biosynthesis pathway into the degradation pathway of LEU. As a result, the VAL did not yield IB, but IV, which explains the elevated level of the latter and the shortage of the former. The reason why this pathway was employed instead of the normal VAL degradation pathway, can be attributed to the coinciding depletion of GLU. A reversed VAL-aminotransferase is a GLU source because it yields GLU next an α-keto acid, which basically is VAL stripped of its amino group. This acid is then routed to the LEU biosynthesis pathway. A visualisation of these pathways is provided in Figure 6.14. To conclude, integrating more experimental data is required to prevent such artefacts. Proteomic abundance data, for example, could exclude the LEU biosynthesis pathway via a GECKO implementation.

Fourthly, the predicted growth curve showed qualitative agreement with the observed ones, except for the first simulated phase. As already mentioned, this was the one with the ASN-mediated high growth. The next predicted phase had a growth rate of 0.296 h^{-1}, which corresponded quite well with the observed ones (0.346 h^{-1} and 0.283 h^{-1})[2]. However, it was predicted to last about 8 h, while there was already a metabolic shift at approximately 5 h of empirical run time. Nonetheless, both exhibited then a slowdown when nutrient contents were dwindling. Finally, both reached a stationary phase between 20 h and 25 h of run time. The simulation still predicted a quasi-zero growth rate of 0.006 h^{-1}, while the experimental OD and DW were already decreasing, which can be considered an artefact of the biomass optimisation objective. Eventually, empirical DW values peaked around 1.16 g_{DW}/L, while 1.658 g_{DW}/L was predicted at the start of the stationary phase. The experimental biomass concentration was 70 % of the simulated one, which nicely aligns with the conversion based on ammonium concentrations.

Lastly, the DO profiles agreed to some extent. For example, it kept track of the initial exponential trajectory, but already stalled around 7 h at a DO value about

---

[2] Again, run 3 was not taken into account because of its deviating behaviour.

79 % due to the depletion of ALA. The experimental curve, however, decreased further to 75 % at around 15 h, and lagged behind due to metabolic shifts. At 13.4 h of simulation time, GLU was depleted and the acetate reuptake started. This caused the DO to jump back to a plateau at 85 %, where it stayed until the depletion at 26 h. The empirical reuptake on the other hand, was not directly perceived in run 1 and run 2. However, in analogy with run 3 in which the acetate reuptake coincided with a plateau in the DO valley from around 12 h till 17 h, the reuptake phase in the other two runs was expected from approximately 16 h to only 21 h.
The DO profile appears thus highly dependent on the active metabolism, which reflects in, for example, the depletion of ALA. The simulation immediately consumed it completely to optimise the growth rate, while in real applications ALA possibly was only gradually released and consumed. Whereas the simulation thus optimised its metabolism for the case in which the hydrolysis was not rate-determining, the experimental strain could have had to cope with the AA mix that was available at a certain moment. As a result, changes in the DO curve could have other causes than the depletion of an AA. A more thorough refinement with quantitative data is required.

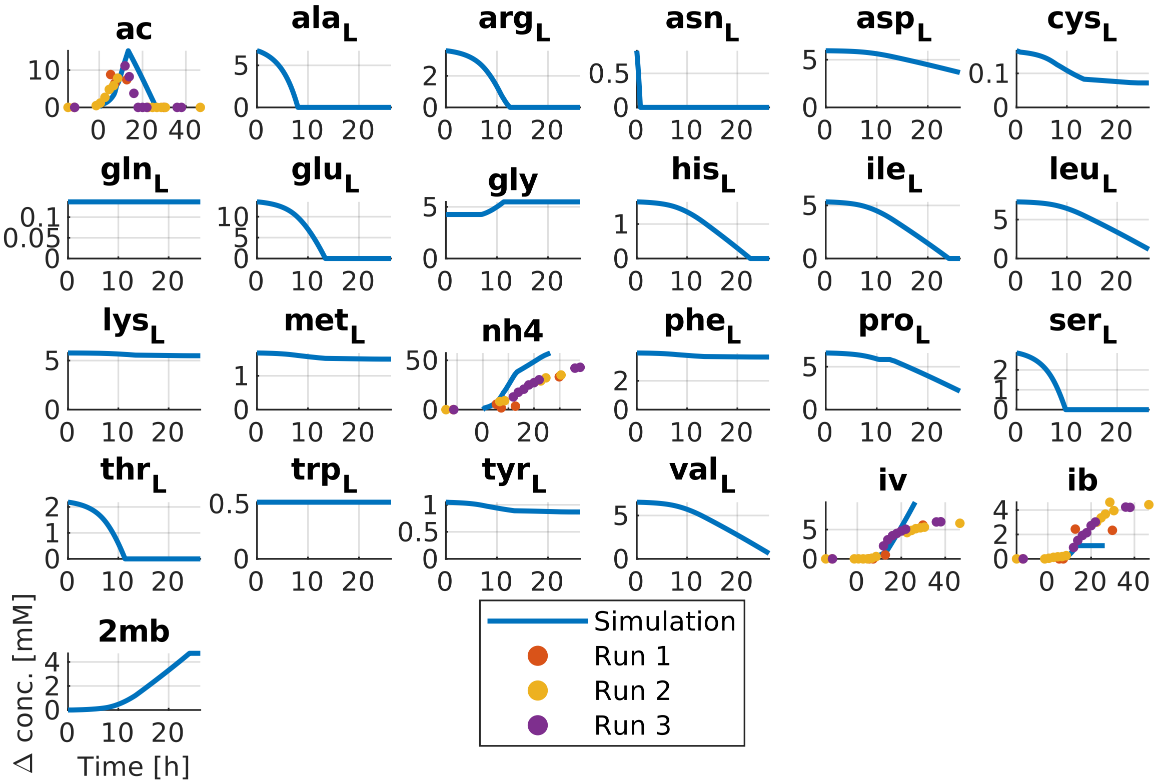
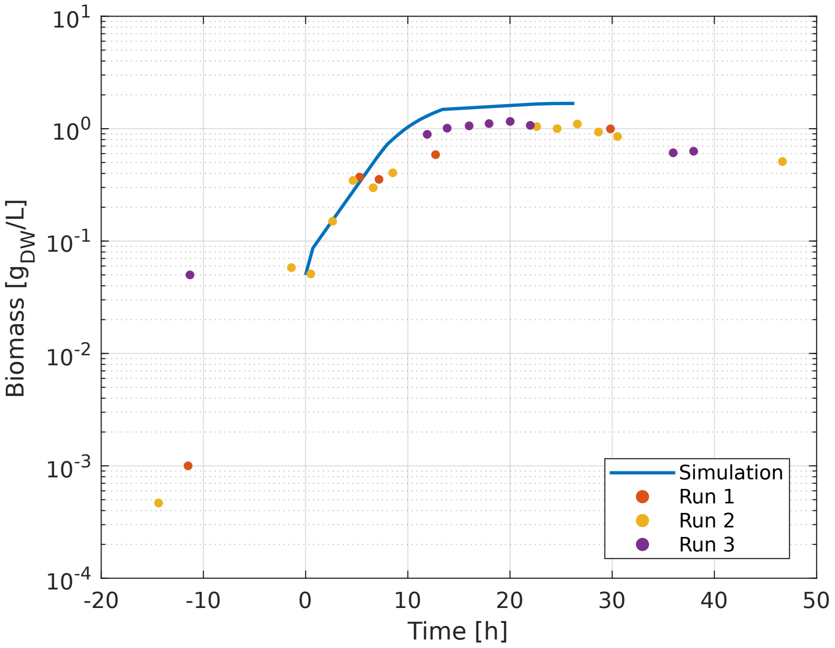
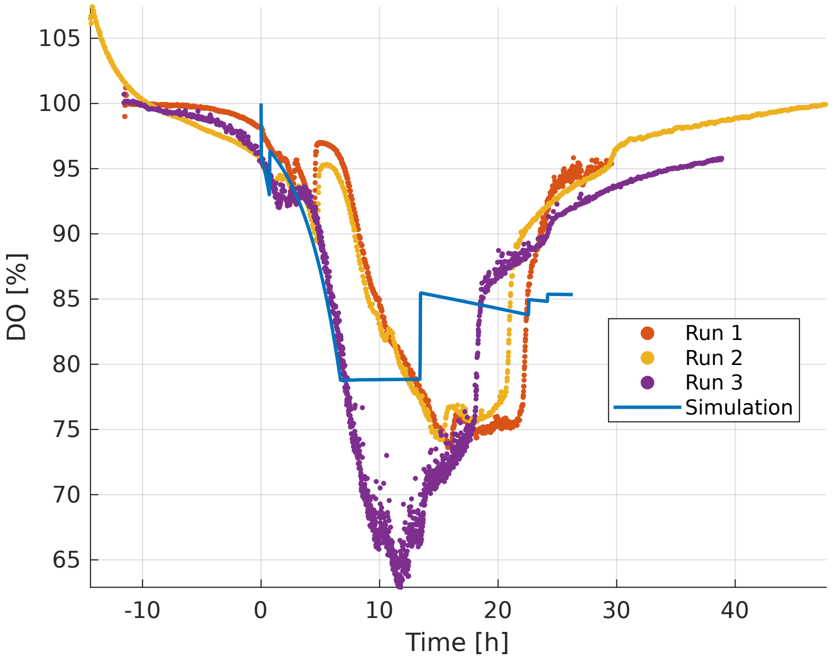


Figure 6.12: Simulated and empirical metabolite time profiles for the iYO844-Gk model after aligning it with the experimental observations. Experimental data are depicted when available. The lag times were subtracted.



(a) Biomass



(b) Dissolved oxygen

Figure 6.13: Simulated and empirical biomass and dissolved oxygen time profiles for the iYO844-Gk model after aligning it with the experimental observations. Experimental data are depicted where available. The lag times have been subtracted.

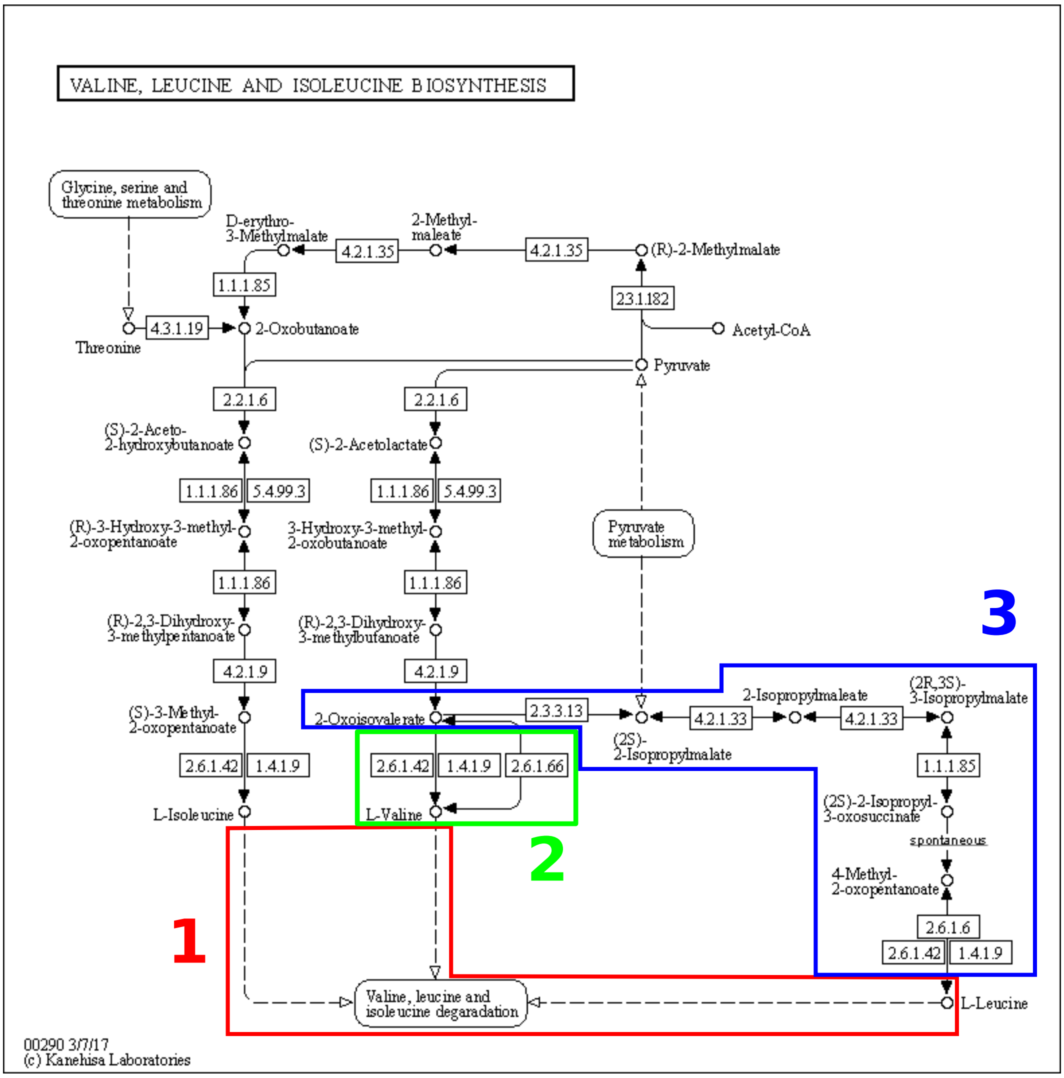


Figure 6.14: Pathway map of the conversion of VAL into LEU. When GLU is present, VAL is catabolised by the normal degradation reactions 1 to IB (see Figure 6.11). When GLU is depleted, the VAL aminotransferase reaction 2 – also part of the degradation metabolism (reactions 1) – flips sense and yields GLU, next to an α-keto acid, which is routed via the LEU biosynthesis pathway 3 to LEU. LEU on its term is broken down by reactions 1 to IV. Adapted map from KEGG [74].

6.4 Conclusion

In this chapter, a *B. subtilis* 168 strain was analysed *in vitro* in view of refining the metabolic model iYO844-Gk for better genotype selection by optimisation algorithms. Three experimental runs using complex LB protein medium provided data for several process and metabolic parameters. These data showed that lag times ranged from 11 h to 14 h, while there were several growth phases due to a couple of metabolic shifts. This was also reflected in a continuous slowdown of the exponential growth trajectories, rendering the growth curve smooth and gradually flattening. Acetate, isovalerate, isobutyrate and ammonium were the main metabolic products. While acetate mainly was a product of the early phases, isovalerate and isobutyrate were those of the later phases, in which the acetate also was remetabolised. The depletion of acetate coincided with the onset of the stationary and the death phase with accompanying decreasing DW and OD. The AA analysis did not succeed well, possibly due to detection issues. Ammonium was produced throughout the entire process, albeit at different rates.
The third experimental run was somewhat different than the other two ones. It appeared to consume more oxygen, which reflected in a lower DO and a higher OTR. The DW was also higher and the profiles seemed slightly squeezed in the time dimension, as if the substrates were converted at a higher pace.

The smooth growth curve did not make it easy to determine growth rates due to difficulties setting the ends of the data segment to be linearly regressed. The growth curve was expressed in terms of online OD because of its numerous data points, and this because the offline sampling strategy did not collect sufficient data. Nevertheless, the necessary condition that the online OD and the DW should be linear for this data representation, was fulfilled. The estimated growth rates were lower than the predicted ones, but this can be attributed to an incomplete protein hydrolysis. Furthermore, the off-gas analysis yielded consistent estimates of the oxygen transfer coefficient of the bioreactor system.

Next, the model was altered to align its predictions qualitatively with the obtained experimental data. The exchange flux and the GECKO constraints were scaled by a ratio of predicted and empirical growth rates, the acetate reuptake was enabled, a dissolved oxygen balance was introduced and new pathways were added to facilitate isobutyrate and isovalerate secretion. Concerning the latter, it was remarkable that these pathways were not present in the metabolic model, despite being fairly well documented in pathway databases.
Finally, the model was qualitatively assessed because all envisaged model alterations had been implemented successfully. It had some good agreements with the empirical data, but it was clear that it had not been refined quantitatively. Furthermore, a remarkable artefact appeared in the newly added BCFA secretion pathways. At the moment of glutamate depletion, valine aminotransferase flipped sense and turned into a glutamate generator. The result was an underestimation of IB and an excess of IV. Integrating more experimental data in the model could elucidate this.

Chapter 7

General conclusion and further research

7.1 General conclusion

This thesis applied the promising concept of protein deamination for the microorganism *Bacillus subtilis*. By systematically screening its metabolism *in silico* using genome-scale metabolic network (GSMN) modelling, genetic modifications were sought to improve its ammonia secretion rate when processing nitrogenous substrates such as proteins. Flux balance analysis (FBA) and associated procedures allow to assess such large-scale networks with only few information using constraint-based optimisation. Such a metabolic screening comprises different steps.

First, in Chapter 4, a GSMN model was selected and validated. In a dedicated database, two models were found: iYO844 and iBsu1103. Both were set up to simulate growth in three different media, but predicted distinct flux distributions in all three cases. Predictions by iYO844 qualitatively aligned with empirical fluxes, but there still was room for improvement, while those of iBsu1103 were unacceptable. Both models were enhanced by the GECKO technique using experimental proteomic data for the media at hand to tighten the solution space of the underlying optimisation problem. iBsu1103 was still off-target, which could be attributed to its loose flux constraint policy, while iYO844 showed a quantitatively good agreement. Furthermore, validating the combined powers of GECKO and MOMA did not uncover compatibility issues. Consequently, iYO844 with GECKO enhancement (iYO844-Gk) was selected for the further modelling work of this thesis.

Second, dynamic FBA provided a view on the sensitivity of the ammonia secretion flux to the composition of the growth medium. This also gave an idea of which and how much of a product to expect in which time window for a certain medium. Here, LB medium was applied, which is a widely known recipe for bacteria cultivation in laboratories. 57 mM of ammonia would be produced, while the other predicted main metabolic products were acetate and propionate. The ammonia productivity appeared sensitive to the share of only a few amino acids, with special attention to alanine, arginine and glutamate. Not only the deamination of amino acids with multiple amino groups, but also the degradation of preferential ones with only one amino group were found major ammonia sources. As a side effect, plant-based media would generally have more potential for a high ammonia yield because of their characteristic amino acid composition.

Third, the metabolic network optimisation was executed in Chapter 5. The flexibility of the selected optimisation algorithms allowed to set a custom (initial) target: optimising the ammonia-biomass yield with genetic deletions as obtained via MOMA with a growth rate threshold of 25 % of the wild-type. This analysis learned that the best deletion patterns are situated within the anaplerotic reactions and/or the pentose phosphate pathway. Yield improvements mainly came at the expense of growth rates, while the ammonia secretion fluxes increased only slightly. This implies that the same ammonia yields can be achieved with less biomass, reducing the costs of treating waste biomass. However, a low biomass growth rate is a drawback for large-scale bioprocessing operations. Therefore, inducible knockouts were proposed to reduce its impact, because these could allow to toggle between a wild-type high-growth mode and a mutant high-yield production mode while still achieving the same ammonia secretion rate. As such, a trade-off between lower conversion rates and more waste biomass could be possible instead of a dilemma.
Another strategy is to alter the optimisation objective of the algorithms to discover other genotypes. In total, three objective alterations were examined. The first two were increasing the growth rate threshold to 75 % or optimising the secretion rate instead of the yield. These adjustments uncovered ΔytsJΔpckA, a genotype with an increase of 8 % in secretion flux and more than half the wild-type growth rate, a remarkably high value. The third edit was to include genetic amplifications. This uncovered that the arginine degradation subsystem has a great impact on the ammonium release flux, which aligns with the earlier sensitivity analysis. By amplifying reactions from this metabolism, the ammonia release flux could be intensified to substantially higher levels, but the growth rate was again low.

In Chapter 6, some possible modelling issues and ameliorations requiring empirical data were formulated and some new *in vitro* experiments were conducted. In total, three batch bioreactor runs using LB medium were executed. Several growth phases were observed due to a couple of metabolic shifts, which resulted in a smooth and gradually flattening growth curve. The main metabolic products were acetate, isobutyrate, isovalerate and ammonium. While acetate mainly was a product of the first growth phases, the later phases were characterised by the reuptake of acetate and the secretion of isobutyrate and isovalerate. The latter ones were not predicted because they were not included in the iYO844-Gk model. The stationary phase started when acetate was depleted. Ammonium was produced throughout the entire run, including the stationary phase, but at varying rates. Both the ammonium and biomass concentration at the onset of the stationary phase were about 70 % of the simulated values.

The estimated growth rates were lower than predicted, but this could be attributed to an incomplete protein hydrolysis. As a result, the assumption that the proteolysis is not rate-determining, did not hold. Off-gas analysis yielded consistent estimates of the oxygen transfer coefficient. Due to the high number of growth phases, the offline sampling scheme did not generate sufficient samples for each phase. As a result, more quantitative parameters were not determined because it would entail high uncertainties.
Instead, the model was adjusted to qualitatively align its predictions with the observations. The exchange flux constraints were relaxed and scaled according to the observed growth rate and acetate reuptake, a dissolved oxygen balance was introduced using the estimated oxygen transfer coefficient, and new pathways were added to facilitate isobutyrate and isovalerate secretion. Afterwards, it showed some qualitative agreements with the observations, but it was clear that a more profound quantitative refinement is necessary. More experimental data is required as well to exclude an encountered modelling artefact.

7.2 Further research

An obvious next research step would be to validate the genotypes encountered in this thesis *in vivo* with one of the various genetic tools available. From the simulations, it is expected that most growth rates will be low, but inducible knockouts were proposed to reduce the associated drawbacks. After a successful validation, another step could thus be to design and to optimise a metabolic toggle switch framework.

The GECKO concept has been proven a useful technique, because it improved the prediction power of model iYO844 for a couple of simulated growth media. In this thesis, the technique was applied for the central carbon subsystem, because these reactions were fairly well documented. For this application, however, it would make sense to implement it in the amino acid subsystem as well. However, reliable kinetic information for each reaction in an organism is scarce and also appropriate enzyme abundance data are not plentiful. New data is needed.

In the course of this thesis, genetic amplifications were modelled as well as deletions, but genetic modifications are not limited to those two types. It is also possible to reduce the expression of a gene – a knockdown – which results in a lower enzyme abundance and, hence, a lower flux [31]. Another possibility is to add heterologous pathways, which enlarges the searching space vastly. Some research has already been executed in this area, which resulted in a.o. the optimisation algorithm OptStrain [76].

In all three experimental runs, a high number of growth phases was observed, which could not be thoroughly examined due to the low number of manual offline samples in each phase. A more intensive sampling strategy is required to quantitatively grasp the metabolism of each phase. A higher manual sampling frequency, however, is only achievable to some extent, because each sample requires some processing time.

As a result, no uptake fluxes or metabolic yields could be reliably estimated for these experimental runs, with which the uptake constraints of the model could have been refined for this specific strain. Speaking of that, the current uptake constraints might even be not valid for LB medium because it seemed that the proteolysis had an impact on the conversion rate, although *B. subtilis* disposes of a powerful pallet of proteases. The metabolic model does not contain a protein hydrolysis subsystem, which complicates further modelling. Another option is to hydrolyse the proteins beforehand to clarify whether the impact of this bottleneck is large or not. If not, perhaps it could be taken into account using a simple correction factor for the exchange constraints of the free amino acids, similarly as was done in the qualitative model alignment.

The production of isovalerate and isobutyrate from LB medium was unexpected because it was not explicitly mentioned in literature. Also the fact that both metabolites were not included in a GSMN model, raises the question whether branched chain fatty acid secretion in LB medium is common or rather exceptional. Although the pathways are known in databases and the metabolic model is able to predict their secretion after adding these pathways, a confirmation using another *B. subtilis* strain could eliminate further doubts.

LB medium often was the basis in this thesis, but Chapter 4 also assessed via simulations which kind of protein source would have the highest potential for a high ammonia yield. A comparative *in vitro* study of how *B. subtilis* copes with different media, with or without genetic modifications, and which one(s) show the highest yields, would be an interesting case. In this context, one could also examine how the presence of saccharides would influence the process, because protein sources seldom are purely proteins.

Appendix A

List of gene and reaction abbreviations

ackA	ACKr	acetate kinase
alsD	ACLDC	acetolactate decarboxylase
citB	ACONT	aconitase
ald	ALAD	L-alanine dehydrogenase
rocF	ARGN	arginase
aspB	ASPTA1	aspartate transaminase
buk	BUTKr	butyrate kinase
citAZ	CS	citrate synthase
lysA	DAPDC	diaminopimelate decarboxylase
eno	ENO	enolase
fbaAB	FBA	fructose-biphosphate aldolase
fbp	FBP	fructose-biphosphatase
citG	FUM	fumarase
zwf	G6PDH	glucose-6-phosphate dehydrogenase
gapA	GAPD_NAD	glyceraldehyde-3-phosphate dehydrogenase (NAD direction)
gapB	GAPD_NADP	glyceraldehyde-3-phosphate dehydrogenase (NADP direction)
glcK	HEX1	hexokinase (D-glucose-ATP)
rocA	HP5CD	3-hydroxy-1-pyrroline-5-carboxylate dehydrogenase
icd	ICDHy	isocitrate dehydrogenase (NADP)
ldh	LDH	L-lactate dehydrogenase
mdh	MDH	malate dehydrogenase
ytsJ	ME2	malic enzyme (NADP)
metK	METAT	methionine adenosyltransferase
yqiQ	MICITL	methylisocitrate lyase
rocD	ORNTA	ornithine transaminase
oxdC	OXADC	oxalate decarboxylase

menD	OXGDC	2-oxoglutarate decarboxylase
ptb	PBUTT	phosphate butyryltransferase
pycA	PC	pyruvate carboxylase
pdhABCD	PDH	pyruvate dehydrogenase
pfkA	PFK	phosphofructokinase
serA	PGCDr	phosphoglycerate dehydrogenase
gndA	PGDHr	phosphogluconate dehydrogenase
pgk	PGK	phosphoglycerate kinase
pgi	PGI	glucose-6-phosphate isomerase
pgl	PGL	6-phosphogluconolactonase
pgm	PGM	phosphoglycerate mutase
pckA	PPCK	phosphoenolpyruvate carboxykinase
pps	PPS	phospoenolpyruvate synthase
ycgM	PRO1x	proline oxidase (NAD)
serC	PSERTr	3-phosphoserine aminotransferase
pta	PTAr	phosphotransacetylase
pyk	PYK	pyruvate kinase
rpe	RPE	ribulose-5-phosphate 3-epimerase
ywlF	RPI	ribose-5-phosphate isomerase
gabD	SSALy	succinate-semialdehyde dehydrogenase (NADP)
sdhABC	SUCD1	succinate dehydrogenase
sucCD	SUCOAS	succinyl-coA synthetase
ywjH	TAL	transaldolase
tkt	TKT	transketolase
tpiA	TPI	triose-phosphate isomerase
ureABC	UREA2	urease

For more information about each gene and reaction, the reader is kindly referred to the supplementary information included with model iYO844 [10].

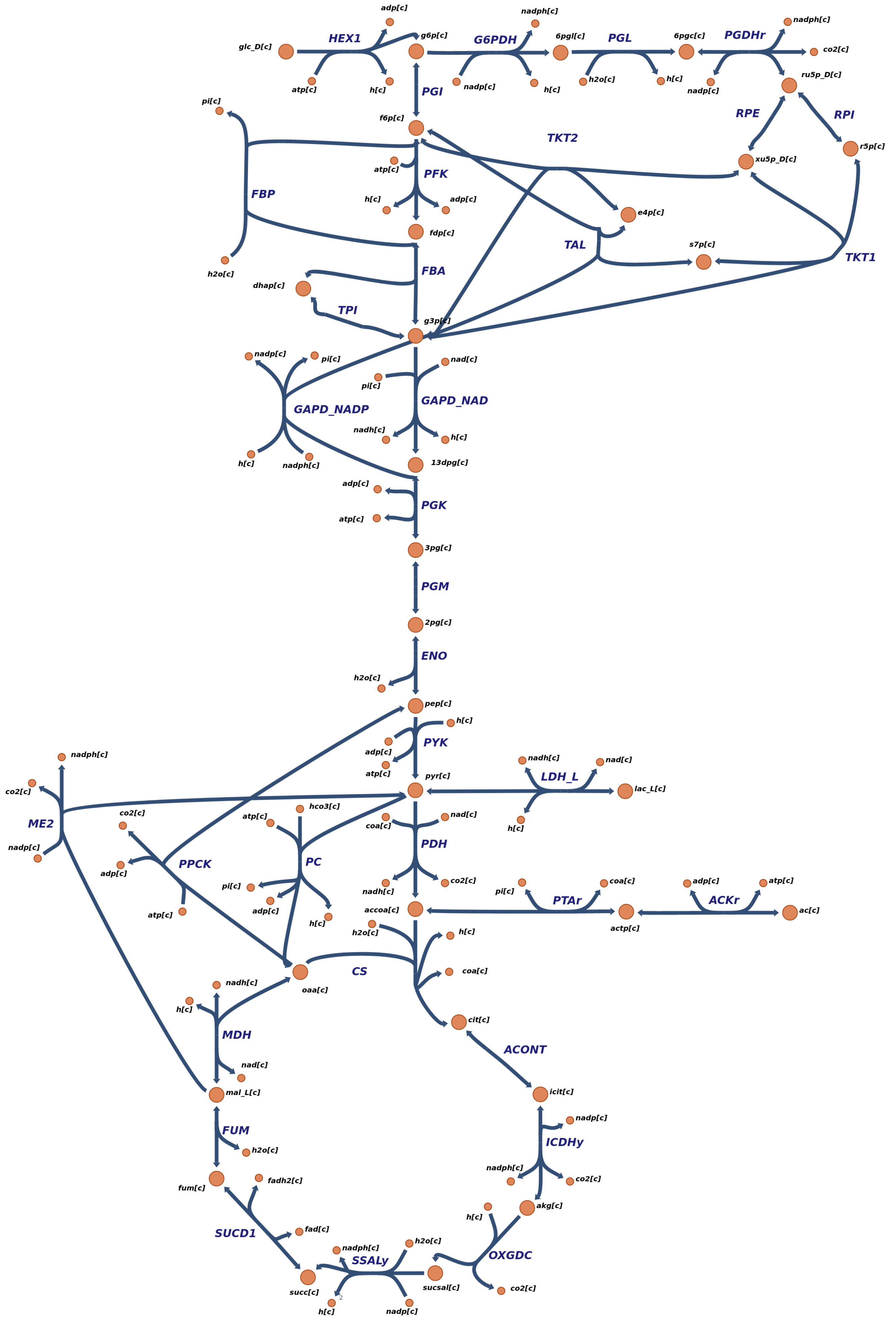


Figure A.1: Visualisation of the central carbon metabolic network of iYO844, including reaction labels as provided in the model and the entire scope of this thesis.

Appendix B

Analytical solution of the oxygen balance

The dissolved oxygen balance is stated as following (see also equation (3.13)).

$$\frac{dC_{O_2}}{dt} = q_{O_2}X_0e^{\mu t} + k_La(C_{O_2}^{sat} - C_{O_2}) \tag{B.1}$$

Rewriting this as

$$\frac{dC_{O_2}}{dt} + k_LaC_{O_2} = k_LaC_{O_2}^{sat} + q_{O_2}X_0e^{\mu t} \tag{B.2}$$

gives a linear first order differential equation with variable coefficients of the form

$$y' + p(t)y = q(t) \tag{B.3}$$

Such equations are solved by multiplying both sides of the equation with an integration factor $\chi(t)$, defined as following [77].

$$\chi(t) = e^{\int p(t)dt} = e^{k_Lat} \tag{B.4}$$

By applying the chain rule and separating the variables, this results in

$$\int d\left(C_{O_2}e^{k_Lat}\right) = \int \left(k_LaC_{O_2}^{sat}e^{k_Lat} + q_{O_2}X_0e^{(\mu+k_La)t}\right)dt \tag{B.5}$$

Integrating gives

$$C_{O_2}e^{k_Lat} + D_1 = \frac{k_LaC_{O_2}^{sat}}{k_La}\left(e^{k_Lat} - 1\right) + \frac{q_{O_2}X_0}{\mu + k_La}\left(e^{(\mu+k_La)t} - 1\right) + D_2 \tag{B.6}$$

Isolating $C_{O_2}(t)$ and setting a new integration constant returns

$$C_{O_2}(t) = C_{O_2}^{sat}\left(1 - e^{-k_Lat}\right) + \frac{q_{O_2}X_0}{\mu + k_La}\left(e^{\mu t} - e^{-k_Lat}\right) + De^{-k_Lat} \tag{B.7}$$

Introducing the initial condition $C_{O_2}(0) = C_{O_2,0}$

$$C_{O_2,0} = C_{O_2}^{sat}(1-1) + \frac{q_{O_2}X_0}{\mu + k_La}(1-1) + D \cdot 1 = D \tag{B.8}$$

gives the final solution

$$C_{O_2}(t) = C_{O_2,0} e^{-k_L a t} + C_{O_2}^{sat} \left(1 - e^{-k_L a t}\right) + \frac{q_{O_2} X_0}{\mu + k_L a} \left(e^{\mu t} - e^{-k_L a t}\right) \tag{B.9}$$

If there is no oxygen transport ($k_L a = 0$), this equation reduces back to the analytic solution of the batch mass balance in equation (2.15) (with q_{O_2} equal to $(\boldsymbol{S}\boldsymbol{v})|_{O_2}$). If the oxygen transport is unlimited ($k_L a = \infty$), it reduces to $C_{O_2}^{sat}$.

Appendix C

Sequential scanning algorithm

Listing C.1: Master function of the sequential scanning algorithm

```
%% Sequential scanner for both gene deletions and upregulations, adaptable for number of affected genes, loop selection percentile, upregulation factor & growth rate threshold
%
% INPUTS:
%   model            COBRA model structure
%   genesel          cell array containing the candidate genes with names as in model.genes
%   n                number of affected genes
%   prodRxn          char containing the reaction to be optimised with name as in model.rxns
%   MM               molar mass of the product produced by prodRxn, used to calculate the product-biomass yield
%   upRegFactor      gene upregulation factor for use as in Wang et al. (Biochem Eng J doi: 10.1016/j.bej.2017.03.017)
%   selPerc          selected percentile of the phenotype ranking when increasing the number of deletions
%   muThreshold      growth rate threshold
%   optimVar         optimisation variable, choice between product 'yield' and 'flux'
%
% OUTPUTS:
%   resPerLevel      cell array containing the phenotype ranking for each number of deletions extended with the flux distribution,
%                    the growth rate, the product flux and the product yield resp. The 6th column is for ranking purposes.
```

```matlab
%
% Author: Lucas De Vrieze (14 Apr 2021)
%
function[resPerLevel]=sequentialScanner(model,genesel,n,prodRxn,MM,upRegFactor,selPerc,muThreshold,optimVar)
    % Initialisation
    j=1;
    resPerLevel=cell(n,1);
    solWT=optimizeCbModel(model,'max','one'); % pFBA WT
    if strcmpi(optimVar,'yield')
        optimCol=5;
    elseif strcmpi(optimVar,'flux')
        optimCol=4;
    else
        error('Optimisation variable not understood!')
    end

    while j<=n
        fprintf(append('Entering loop ',num2str(j),'...\n'));

        % Full first-level scan for the first gene
        if j==1
            [res,model_c]=sequentialScannerWorker(model,genesel,prodRxn,MM,upRegFactor,solWT,muThreshold,optimCol);
            resPerLevel{j}=res;
            cutoff=ceil(size(model_c,1)*selPerc);
            topModels=model_c(1:cutoff);

        % New scan for selected percentile of previous loop
        else
            resThisLevel=cell(0,5);
            model_c_ThisLevel=cell(0,1);
            for m=1:length(topModels)

                % Determine genes already affected in previous loops and remove them from the candidate list
                geneIndices=contains(topModels{m}.genes,'_')&~contains(topModels{m}.genes,'E_');
                genesUsed=model.genes(geneIndices);
```

```
                [~,genesAlreadyUsed,~]=intersect(genesel,genesUsed);
                geneselUnused=genesel;
                geneselUnused(genesAlreadyUsed)=[];

                % Some string handling for a proper gene representation in the results cell array
                geneIndicesDels=contains(topModels{m}.genes,'_d');
                geneNamesDels=append(model.geneNames(geneIndicesDels),'-');
                geneIndicesUps=xor(geneIndices,geneIndicesDels);
                geneNamesUps=append(model.geneNames(geneIndicesUps),'+');
                geneNames=[geneNamesDels;geneNamesUps];
                geneNames=strjoin(geneNames,' & ');

                % Calling the scanner worker and storing its results
                fprintf(append('Starting scan for gene ',num2str(m),':\t',geneNames,'...\n'));
                [res,model_c]=sequentialScannerWorker(topModels{m},geneselUnused,prodRxn,MM,upRegFactor,solWT,muThreshold,optimCol);
                res(:,1)=cellfun(@(x) strjoin({x,geneNames},' & '),res(:,1),'UniformOutput',false);

                % Store all results of this loop together
                resThisLevel=[resThisLevel;res];
                model_c_ThisLevel=[model_c_ThisLevel;model_c];
            end

            % Deduplicate phenotype combinations (e.g. 'geneA+ & geneB-' and 'geneB- & geneA+')
            allCombinations=cellfun(@(x) strjoin(sort(strsplit(x,' & ')),''),resThisLevel(:,1),'UniformOutput',false);
            [~,uniqInd,~]=unique(allCombinations);
            resThisLevel=resThisLevel(uniqInd,:);
            model_c_ThisLevel=model_c_ThisLevel(uniqInd);
```

```

         % Rank the results of this loop
         [resThisLevel,ind]=sortrows(resThisLevel,6,'descend');
         model_c_ThisLevel=model_c_ThisLevel(ind);
         resPerLevel{j}=resThisLevel;

         % Determine selected percentile for the next loop
         cutoff=ceil(size(model_c_ThisLevel,1)*selPerc);
         topModels=model_c_ThisLevel(1:cutoff);
      end
      j=j+1;
   end
end
```

Listing C.2: Worker function of the sequential scanning algorithm

```
%% Worker algorithm for the sequential scanner. Executes the actual scanning loop by looping over all gene candidates for a provided model.
%
% INPUTS:
%   model              COBRA model structure
%   genesel            cell array containing the candidate genes with names as in model.genes
%   n                  number of affected genes
%   prodRxn            char containing the reaction to be optimised with name as in model.rxns
%   MM                 molar mass of the product produced by prodRxn, used to calculate the product-biomass yield
%   upRegFactor        gene upregulation factor for use as in Wang et al. (Biochem Eng J doi: 10.1016/j.bej.2017.03.017)
%   solWT              wild-type solution structure
%   muThreshold        growth rate threshold
%   optimCol           column number within res to which the optimisation variable is assigned (4 for flux, 5 for yield)
%
% OUTPUTS:
%   res                cell array containing the phenotype ranking for this number of deletions, extended with the flux distribution,
%                      the growth rate, the product flux and the product yield resp. The 6th column is for ranking purposes.
%   model_c            cell array containing the model structure of each phenotype encountered in this scan
%
% Author: Lucas De Vrieze (14 Apr 2021)
%
function[res,model_c]=sequentialScannerWorker(model,genesel,prodRxn,MM,upRegFactor,solWT,muThreshold,optimCol)

    prodRxn=find(strcmp(model.rxns,prodRxn));
    muWT=solWT.f;

    % Initialising the loop
    res=cell(length(genesel)*2,6);
```

```
    jump=length(genesel);
    model_c=cell(length(genesel)*2,1);
    fprintf(append('\tCurrent gene:\t\t',repmat(' ',1,max(cellfun('length',genesel)))));
    name='';

    % The loop itself, scanning all candidate genes
    for g=1:length(genesel)
        nu=genesel{g};

        % String handling for verbose feedback
        fprintf(repmat('\b',1,length(name)+1));
        name=model.geneNames{contains(model.genes,nu)};
        fprintf(append(name,'-'));
        t=GetGeneNameFromSysGeneName(model,{nu});

        % Assess a deletion and collect results.
        res{g,1}=append(t{1},'-');
        model_del=deleteModelGenes(model,nu);
        model_c{g}=model_del;
        sol=MOMA(model,model_del,'max',false,'one',true,solWT);
        if isempty(sol.x) % An 'infeasible' solver feedback is penalised.
            res{g,6}=-Inf;
        else
            res{g,2}=sol.x; % Flux distribution
            res{g,3}=sol.f; % Growth rate
            res{g,4}=sol.x(prodRxn); % Product flux
            res{g,5}=sol.x(prodRxn)/sol.f/1000*MM;% Yield
            if sol.f<muWT*muThreshold % A growth rate below the threshold is penalised.
                res{g,6}=0;
            else
                res{g,6}=res{g,optimCol};
            end
        end

        % Assess an upregulation and collect results. Remove GECKO constraint if encountered.
        res{g+jump,1}=append(t{1},'+');
        fprintf('\b+');
        model_up=upregulateModelGenes(model,nu,upRegFactor,solWT);
        [~,geneInfo]=findRxnsFromGenes(model,t,true);
```

```
        formsplit=strsplit(geneInfo{2});
        GECKO=find(contains(formsplit,'_ab'));
        if ~isempty(GECKO)
            model_up=removeMetabolites(model_up,formsplit(GECKO),true);
        end
        model_c{g+jump}=model_up;
        sol_up=MOMA(model,model_up,'max',false,'one',true,solWT);
        if isempty(sol_up.x)
            res{g+jump,6}=-Inf;
        else
            res{g+jump,2}=sol_up.x;
            res{g+jump,3}=sol_up.f;
            res{g+jump,4}=sol_up.x(prodRxn);
            res{g+jump,5}=sol_up.x(prodRxn)/sol_up.f/1000*MM;
            if sol_up.f<muWT*muThreshold
                res{g+jump,6}=0;
            else
                res{g+jump,6}=res{g+jump,optimCol};
            end
        end
    end

    % Rank results
    [res,ind]=sortrows(res,6,'descend');
    model_c=model_c(ind);
    fprintf('\n')
end
```

Bibliography

[1] Y. Mikami, H. Yoneda, Y. Tatsukami, W. Aoki, and M. Ueda, "Ammonia production from amino acid-based biomass-like sources by engineered Escherichia coli," *AMB Express*, vol. 7, no. 1, 2017.

[2] M. Appl, "Ammonia," in *Ullmann's Encyclopedia of Industrial Chemistry*, pp. 904–980, Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, dec 2006.

[3] T. M. Paschkewitz, *Ammonia Production at Ambient Temperature and Pressure.* PhD thesis, University of Iowa, jul 2012.

[4] N. Gil-Lalaguna, Z. Afailal, M. Aznar, and I. Fonts, "Exploring the sustainable production of ammonia by recycling N and H in biological residues: Evolution of fuel-N during glutamic acid gasification," *Journal of Cleaner Production*, p. 124417, oct 2020.

[5] D. G. Wernick and J. C. Liao, "Protein-based biorefining: metabolic engineering for production of chemicals and fuel with regeneration of nitrogen fertilizers," *Applied Microbiology and Biotechnology*, vol. 97, pp. 1397–1406, feb 2013.

[6] Y. X. Huo, K. M. Cho, J. G. Rivera, E. Monte, C. R. Shen, Y. Yan, and J. C. Liao, "Conversion of proteins into biofuels by engineering nitrogen flux," *Nature Biotechnology*, vol. 29, no. 4, pp. 346–351, 2011.

[7] K. Y. Choi, D. G. Wernick, C. A. Tat, and J. C. Liao, "Consolidated conversion of protein waste into biofuels and ammonia using Bacillus subtilis," *Metabolic Engineering*, vol. 23, pp. 53–61, 2014.

[8] I. Pikaar, S. Matassa, K. Rabaey, B. L. Bodirsky, A. Popp, M. Herrero, and W. Verstraete, "Microbes and the Next Nitrogen Revolution," *Environmental Science and Technology*, vol. 51, no. 13, pp. 7297–7303, 2017.

[9] Y. Tatemichi, K. Kuroda, T. Nakahara, and M. Ueda, "Efficient ammonia production from food by-products by engineered Escherichia coli," *AMB Express*, vol. 10, no. 1, 2020.

[10] Y. K. Oh, B. O. Palsson, S. M. Park, C. H. Schilling, and R. Mahadevan, "Genome-scale reconstruction of metabolic network in Bacillus subtilis based on

high-throughput phenotyping and gene essentiality data," *Journal of Biological Chemistry*, vol. 282, no. 39, pp. 28791–28799, 2007.

[11] C. D. Demirhan, W. W. Tso, J. B. Powell, and E. N. Pistikopoulos, "Sustainable ammonia production through process synthesis and global optimization," *AIChE Journal*, vol. 65, no. 7, 2019.

[12] M. Reese, C. Marquart, M. Malmali, K. Wagner, E. Buchanan, A. McCormick, and E. L. Cussler, "Performance of a Small-Scale Haber Process," *Industrial and Engineering Chemistry Research*, vol. 55, no. 13, pp. 3742–3750, 2016.

[13] S. Giddey, S. P. Badwal, and A. Kulkarni, "Review of electrochemical ammonia production technologies and materials," *International Journal of Hydrogen Energy*, vol. 38, no. 34, pp. 14576–14594, 2013.

[14] T. D. Rapson, C. M. Gregg, R. S. Allen, H. Ju, C. M. Doherty, X. Mulet, S. Giddey, and C. C. Wood, "Insights into Nitrogenase Bioelectrocatalysis for Green Ammonia Production," *ChemSusChem*, vol. 13, pp. 4856–4865, sep 2020.

[15] Z. Jamaludin, S. Rollings-Scattergood, K. Lutes, and C. Vaneeckhaute, "Evaluation of sustainable scrubbing agents for ammonia recovery from anaerobic digestate," *Bioresource Technology*, vol. 270, no. August, pp. 596–602, 2018.

[16] S. Bastian and F. H. Arnold, *Microbial Metabolic Engineering*, vol. 834 of *Methods in Molecular Biology.* New York, NY: Springer New York, 2012.

[17] I. Massaiu, L. Pasotti, N. Sonnenschein, E. Rama, M. Cavaletti, P. Magni, C. Calvio, and M. J. Herrgård, "Integration of enzymatic data in Bacillus subtilis genome-scale metabolic model improves phenotype predictions and enables in silico design of poly-γ-glutamic acid production strains," *Microbial Cell Factories*, vol. 18, no. 1, pp. 1–20, 2019.

[18] B. Ø. Palsson and A. Varma, "Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use," *Biotechnology*, vol. 12, no. October, pp. 994–998, 1994.

[19] M. R. Antoniewicz, "Methods and advances in metabolic flux analysis: a mini-review," *Journal of Industrial Microbiology and Biotechnology*, vol. 42, no. 3, pp. 317–325, 2015.

[20] W. Wiechert, "13C Metabolic Flux Analysis," *Metabolic Engineering*, vol. 3, pp. 195–206, jul 2001.

[21] J. D. Orth, I. Thiele, and B. O. Palsson, "What is flux balance analysis?," *Nature Biotechnology*, vol. 28, no. 3, pp. 245–248, 2010.

[22] G. Dantzig, *Linear Programming and Extensions*, vol. 48. Princeton University Press, 2016.

[23] R. Mahadevan and C. H. Schilling, “The effects of alternate optimal solutions in constraint-based genome-scale metabolic models,” *Metabolic Engineering*, vol. 5, no. 4, pp. 264–276, 2003.

[24] N. E. Lewis, K. K. Hixson, T. M. Conrad, J. A. Lerman, P. Charusanti, A. D. Polpitiya, J. N. Adkins, G. Schramm, S. O. Purvine, D. Lopez-Ferrer, K. K. Weitz, R. Eils, R. König, R. D. Smith, and B. Palsson, “Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models,” *Molecular Systems Biology*, vol. 6, no. 390, 2010.

[25] M. W. Covert, C. H. Schilling, and B. Palsson, “Regulation of gene expression in flux balance models of metabolism,” *Journal of Theoretical Biology*, vol. 213, no. 1, pp. 73–88, 2001.

[26] G. Karlebach and R. Shamir, “Modelling and analysis of gene regulatory networks,” *Nature Reviews Molecular Cell Biology*, vol. 9, no. 10, pp. 770–780, 2008.

[27] R. Mahadevan, J. S. Edwards, and F. J. Doyle, “Dynamic Flux Balance Analysis of diauxic growth in Escherichia coli,” *Biophysical Journal*, vol. 83, no. 3, pp. 1331–1340, 2002.

[28] A. Varma and B. O. Palsson, “Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110,” *Applied and Environmental Microbiology*, vol. 60, no. 10, pp. 3724–3731, 1994.

[29] K. Höffner, S. M. Harwood, and P. I. Barton, “A reliable simulator for dynamic flux balance analysis,” *Biotechnology and Bioengineering*, vol. 110, no. 3, pp. 792–802, 2013.

[30] J. A. Gomez, K. Höffner, and P. I. Barton, “DFBAlab: A fast and reliable MATLAB code for dynamic flux balance analysis,” *BMC Bioinformatics*, vol. 15, no. 1, pp. 1–10, 2014.

[31] S. Vijayakumar, M. Conway, P. Lio, and C. Angione, “Seeing the wood for the trees: A forest of methods for optimization and omic-network integration in metabolic modelling,” *Briefings in Bioinformatics*, vol. 19, no. 6, pp. 1218–1235, 2017.

[32] M. J. McAnulty, J. Y. Yen, B. G. Freedman, and R. S. Senger, “Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico,” *BMC Systems Biology*, vol. 6, no. 1, p. 42, 2012.

[33] A. Goelzer, J. Muntel, V. Chubukov, M. Jules, E. Prestel, R. Nölker, M. Mariadassou, S. Aymerich, M. Hecker, P. Noirot, D. Becher, and V. Fromion, “Quantitative prediction of genome-wide resource allocation in bacteria,” *Metabolic Engineering*, vol. 32, pp. 232–243, 2015.

[34] S. Waldherr, D. A. Oyarzún, and A. Bockmayr, "Dynamic optimization of metabolic networks coupled with gene expression," *Journal of Theoretical Biology*, vol. 365, pp. 469–485, 2015.

[35] E. Gonçalves, J. Bucher, A. Ryll, J. Niklas, K. Mauch, S. Klamt, M. Rocha, and J. Saez-Rodriguez, "Bridging the layers: Towards integration of signal transduction, regulation and metabolism into mathematical models," *Molecular BioSystems*, vol. 9, no. 7, pp. 1576–1583, 2013.

[36] B. J. Sánchez, C. Zhang, A. Nilsson, P. Lahtvee, E. J. Kerkhoven, and J. Nielsen, "Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints," *Molecular Systems Biology*, vol. 13, no. 8, p. 935, 2017.

[37] J. Muntel, V. Fromion, A. Goelzer, S. Maa, U. Mäder, K. Büttner, M. Hecker, and D. Becher, "Comprehensive absolute quantification of the cytosolic proteome of bacillus subtilis by data independent, parallel fragmentation in liquid chromatography/mass spectrometry (LC/MSE)," *Molecular and Cellular Proteomics*, vol. 13, no. 4, pp. 1008–1019, 2014.

[38] D. Davidi, E. Noor, W. Liebermeister, A. Bar-Even, A. Flamholz, K. Tummler, U. Barenholz, M. Goldenfeld, T. Shlomi, and R. Milo, "Global characterization of in vivo enzyme catalytic rates and their correspondence to in vitro kcat measurements," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 12, pp. 3401–3406, 2016.

[39] L. Jeske, S. Placzek, I. Schomburg, A. Chang, and D. Schomburg, "BRENDA in 2019: A European ELIXIR core data resource," *Nucleic Acids Research*, vol. 47, no. D1, pp. D542–D549, 2019.

[40] U. Wittig, R. Kania, M. Golebiewski, M. Rey, L. Shi, L. Jong, E. Algaa, A. Weidemann, H. Sauer-Danzwith, S. Mir, O. Krebs, M. Bittkowski, E. Wetsch, I. Rojas, and W. Müller, "SABIO-RK - Database for biochemical reaction kinetics," *Nucleic Acids Research*, vol. 40, no. D1, pp. 790–796, 2012.

[41] B. Zhu and J. Stülke, "SubtiWiki in 2018: From genes and proteins to functional network annotation of the model organism Bacillus subtilis," *Nucleic Acids Research*, vol. 46, no. D1, pp. D743–D748, 2018.

[42] A. Nilsson, J. Nielsen, and B. O. Palsson, "Metabolic Models of Protein Allocation Call for the Kinetome," *Cell Systems*, vol. 5, no. 6, pp. 538–541, 2017.

[43] H. S. Choi, S. Y. Lee, T. Y. Kim, and H. M. Woo, "In silico identification of gene amplification targets for improvement of lycopene production," *Applied and Environmental Microbiology*, vol. 76, no. 10, pp. 3097–3105, 2010.

[44] C. Bro, B. Regenberg, J. Förster, and J. Nielsen, "In silico aided metabolic engineering of Saccharomyces cerevisiae for improved bioethanol production," *Metabolic Engineering*, vol. 8, no. 2, pp. 102–111, 2006.

[45] C. Wang, J. Liu, H. Liu, J. Wang, and J. Wen, “A genome-scale dynamic flux balance analysis model of Streptomyces tsukubaensis NRRL18488 to predict the targets for increasing FK506 production,” *Biochemical Engineering Journal*, vol. 123, pp. 45–56, 2017.

[46] D. Segrè, D. Vitkup, and G. M. Church, “Analysis of optimality in natural and perturbed metabolic networks,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 23, pp. 15112–15117, 2002.

[47] T. Shlomi, O. Berkman, and E. Ruppin, “Regulatory on/off minimization of metabolic flux changes after genetic perturbations,” *Proceedings of the National Academy of Sciences*, vol. 102, pp. 7695–7700, may 2005.

[48] J. Kim and J. L. Reed, “RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations.,” *Genome biology*, vol. 13, no. 9, 2012.

[49] M. L. Mo, B. Ø. Palsson, and M. J. Herrgård, “Connecting extracellular metabolomic measurements to intracellular flux states in yeast,” *BMC Systems Biology*, vol. 3, no. 1, p. 37, 2009.

[50] M. R. Long, W. K. Ong, and J. L. Reed, “Computational methods in metabolic engineering for strain design,” *Current Opinion in Biotechnology*, vol. 34, pp. 135–141, 2015.

[51] H. Alper, Y. S. Jin, J. F. Moxley, and G. Stephanopoulos, “Identifying gene targets for the metabolic engineering of lycopene biosynthesis in Escherichia coli,” *Metabolic Engineering*, vol. 7, no. 3, pp. 155–164, 2005.

[52] A. P. Burgard, P. Pharkya, and C. D. Maranas, “OptKnock: A Bilevel Programming Framework for Identifying Gene Knockout Strategies for Microbial Strain Optimization,” *Biotechnology and Bioengineering*, vol. 84, no. 6, pp. 647–657, 2003.

[53] J. Kim, J. L. Reed, and C. T. Maravelias, “Large-Scale Bi-Level strain design approaches and Mixed-Integer programming solution techniques,” *PLoS ONE*, vol. 6, no. 9, 2011.

[54] S. Ren, B. Zeng, and X. Qian, “Adaptive bi-level programming for optimal gene knockouts for targeted overproduction under phenotypic constraints,” *BMC Bioinformatics*, vol. 14, no. Suppl 2, 2013.

[55] N. Tepper and T. Shlomi, “Predicting metabolic engineering knockout strategies for chemical production: Accounting for competing pathways,” *Bioinformatics*, vol. 26, no. 4, pp. 536–543, 2009.

[56] K. R. Patil, I. Rocha, J. Förster, and J. Nielsen, “Evolutionary programming as a platform for in silico metabolic engineering,” *BMC Bioinformatics*, vol. 6, pp. 1–12, 2005.

[57] C. S. Henry, J. F. Zinner, M. P. Cohoon, and R. L. Stevens, "iBsu1103: A new genome-scale metabolic model of Bacillus subtilis based on SEED annotations," *Genome Biology*, vol. 10, no. 6, pp. 1–15, 2009.

[58] R. S. Malik-Sheriff, M. Glont, T. V. Nguyen, K. Tiwari, M. G. Roberts, A. Xavier, M. T. Vu, J. Men, M. Maire, S. Kananathan, E. L. Fairbanks, J. P. Meyer, C. Arankalle, T. M. Varusai, V. Knight-Schrijver, L. Li, C. Dueñas-Roca, G. Dass, S. M. Keating, Y. M. Park, N. Buso, N. Rodriguez, M. Hucka, and H. Hermjakob, "BioModels-15 years of sharing computational models in life science," *Nucleic Acids Research*, vol. 48, no. D1, pp. D407–D415, 2020.

[59] L. Heirendt, S. Arreckx, T. Pfau, S. N. Mendoza, A. Richelle, A. Heinken, H. S. Haraldsdóttir, J. Wachowiak, S. M. Keating, V. Vlasov, S. Magnusdóttir, C. Y. Ng, G. Preciat, A. Žagare, S. H. Chan, M. K. Aurich, C. M. Clancy, J. Modamio, J. T. Sauls, A. Noronha, A. Bordbar, B. Cousins, D. C. El Assal, L. V. Valcarcel, I. Apaolaza, S. Ghaderi, M. Ahookhosh, M. Ben Guebila, A. Kostromins, N. Sompairac, H. M. Le, D. Ma, Y. Sun, L. Wang, J. T. Yurkovich, M. A. Oliveira, P. T. Vuong, L. P. El Assal, I. Kuperstein, A. Zinovyev, H. S. Hinton, W. A. Bryant, F. J. Aragón Artacho, F. J. Planes, E. Stalidzans, A. Maass, S. Vempala, M. Hucka, M. A. Saunders, C. D. Maranas, N. E. Lewis, T. Sauter, B. Palsson, I. Thiele, and R. M. Fleming, "Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0," *Nature Protocols*, vol. 14, no. 3, pp. 639–702, 2019.

[60] Z. A. King, A. Dräger, A. Ebrahim, N. Sonnenschein, N. E. Lewis, and B. O. Palsson, "Escher: A Web Application for Building, Sharing, and Embedding Data-Rich Visualizations of Biological Pathways," *PLoS Computational Biology*, vol. 11, no. 8, pp. 1–13, 2015.

[61] E. Fischer and U. Sauer, "Large-scale in vivo flux analysis shows rigidity and suboptimal performance of Bacillus subtilis metabolism," *Nature Genetics*, vol. 37, no. 6, pp. 636–640, 2005.

[62] BD, "BD Bionutrients ™ Technical Manual." 2015.

[63] S. H. Gorissen, J. J. Crombag, J. M. Senden, W. A. Waterval, J. Bierau, L. B. Verdijk, and L. J. van Loon, "Protein content and amino acid composition of commercially available plant-based protein isolates," *Amino Acids*, vol. 50, no. 12, pp. 1685–1695, 2018.

[64] M. Dauner, T. Storni, and U. Sauer, "Bacillus subtilis metabolism and energetics in carbon-limited and excess-carbon chemostat culture," *Journal of Bacteriology*, vol. 183, no. 24, pp. 7308–7317, 2001.

[65] P. D. Karp, R. Billington, R. Caspi, C. A. Fulcher, M. Latendresse, A. Kothari, I. M. Keseler, M. Krummenacker, P. E. Midford, Q. Ong, W. K. Ong, S. M. Paley, and P. Subhraveti, "The BioCyc collection of microbial genomes and

metabolic pathways," *Briefings in Bioinformatics*, vol. 20, no. 4, pp. 1085–1093, 2018.

[66] F. Fisher, "polyparci." MATLAB Central File Exchange, `https://www.mathworks.com/matlabcentral/fileexchange/39126-polyparci`, 2014.

[67] Engineering Toolbox, "Air - Composition and Molecular Weight." `https://www.engineeringtoolbox.com/air-composition-d_212.html`, 2003.

[68] J. Schellenberger, J. O. Park, T. M. Conrad, and B. Ø. Palsson, "BiGG: a Biochemical Genetic and Genomic knowledgebase of large scale metabolic reconstructions," *BMC Bioinformatics*, vol. 11, no. 1, p. 213, 2010.

[69] F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessières, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S.-K. Choi, J.-J. Codani, I. F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Düsterhöft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S.-Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G. Grandi, G. Guiseppi, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Hénaut, H. Hilbert, S. Holsappel, S. Hosono, M.-F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S.-M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauël, C. Médigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. O'Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S.-H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, G. Rapoport, M. Rey, S. Reynolds, M. Rieger, C. Rivolta, E. Rocha, B. Roche, M. Rose, Y. Sadaie, T. Sato, E. Scanlan, S. Schleich, R. Schroeter, F. Scoffone, J. Sekiguchi, A. Sekowska, S. J. Seror, P. Serror, B.-S. Shin, B. Soldo, A. Sorokin, E. Tacconi, T. Takagi, H. Takahashi, K. Takemaru, M. Takeuchi, A. Tamakoshi, T. Tanaka, P. Terpstra, A. Tognoni, V. Tosato, S. Uchiyama, M. Vandenbol, F. Vannier, A. Vassarotti, A. Viari, R. Wambutt, E. Wedler, H. Wedler, T. Weitzenegger, P. Winters, A. Wipat, H. Yamamoto, K. Yamane, K. Yasumoto, K. Yata, K. Yoshida, H.-F. Yoshikawa, E. Zumstein, H. Yoshikawa, and A. Danchin, "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis," *Nature*, vol. 390, pp. 249–256, nov 1997.

[70] S. Freeman, *Biological Science.* Prentice Hall, 2 ed., 2005.

[71] G. P. Rédei, *Encyclopedia of genetics, genomics, proteomics, and informatics.* Springer Science & Business Media, 2008.

[72] Y. Soma, K. Tsuruno, M. Wada, A. Yokota, and T. Hanai, "Metabolic flux redirection from a central metabolic pathway toward a synthetic pathway using a metabolic toggle switch," *Metabolic Engineering*, vol. 23, pp. 175–184, 2014.

[73] K. G. Gadkar, F. J. Doyle, J. S. Edwards, and R. Mahadevan, “Estimating optimal profiles of genetic alterations using constraint-based models,” *Biotechnology and Bioengineering*, vol. 89, no. 2, pp. 243–251, 2005.

[74] M. Kanehisa, “KEGG: Kyoto Encyclopedia of Genes and Genomes,” *Nucleic Acids Research*, vol. 28, pp. 27–30, jan 2000.

[75] C. Stephens, “Bacterial sporulation: A question of commitment?,” *Current Biology*, vol. 8, no. 2, pp. 45–48, 1998.

[76] P. Pharkya, A. P. Burgard, and C. D. Maranas, “OptStrain: A computational framework for redesign of microbial production systems,” *Genome Research*, vol. 14, no. 11, pp. 2367–2376, 2004.

[77] W. Malfait, *Analyse & calculus.* Leuven: Acco, 2016.