Inauguration Dissertation
for
obtaining the doctoral degree
of the
Combined Faculty of Mathematics, Engineering and Natural Sciences
of the
Ruprecht-Karls-University
Heidelberg

Presented by
Seyed Babak Loghmani, M.Sc.
born in: Tehran, Iran
Oral examination: 5th of April 2022

Combining omics data and genome-scale models to understand metabolic adaptations in lactic acid bacteria to changing environments

Referees: Prof. Dr. Ursula Kummer
Prof. Dr. Stefan Wölfl

To my lovely family

Acknowledgements

Writing my doctoral thesis was a hard but also an emotional process for me. Getting close to accomplishing one of the biggest goals of my life has not been without emotional consequences. Therefore, I would like to take this opportunity to express my gratitude to everyone whose intellectual and/or emotional support helped me throughout my journey of doctoral study.

First and for most, I would like to dedicate my special thanks to The supervisor, Prof. Ursula Kummer to support me for achieving one of the prime objectives of my life. Ever since I joined the department of modelling of biological processes, I have enjoyed the scientifically advanced yet friendly environment she provided me with. These all have been a great source of inspiration to me. Your support has been beyond words Ursula.

I would like to thank Prof. Stefan Wölfl, my second supervisor, for his nice and positive attitude towards my research in my TAC meetings.

My special appreciation goes to Dr. Sven Sahle (my mentor), for all the mind blowing scientific discussions, as well as intellectually inspiring talks outside of work. Your open mind, enthusiasm for knowledge, and humble attitude have been an ongoing lesson to me.

My Friend, Jocelyn Faberman, thank you for always bringing a smile to my face and being so caring about me.

My colleague and officemate, Dr. Frank Bergmann, always provided the fastest answers to any programming question I had. Thank you Frank for all your persistent support.

I would like to thank my colleagues, Dr. Nadine Veith (my mentor) and Dr. Ruth Großeholz, for their constructive scientific discussions before and after starting my PhD.

I would like to thank Klaus Kappings and Dr. Nana Duhme for all the nice and funny conversations we had, either in German, or English, or half-half. We are also the reigning champions of Ruperto Carola Cup as members of Borussia Bioquant. Cheers guys!

My special thanks goes to people from HGS MathComp, especially Dr. Michael Winckler and Sarah Steinbach, for organizing one of the most amazingly smooth graduate schools in the world. I would also like to thank HGS MathComp and Graduate Academy at Heidelberg University for financing my study.

As part of my PhD project, I had the great opportunity to spend six weeks at Rostock Medical University to perform experiments. During this time, I enjoyed the great supervision from PD Dr. Tomas Fiedler, and substantial support from Eric Zitzow. I would also like to thank Prof. Bernd Kreikemeyer for receiving me in his department. Without the support of the people I just mentioned, this thesis wouldn't have existed.

I would like to thank The Metabolomics Core Technology Platform at COS - Heidelberg Uni-

versity, especially Prof. Rüdiger Hell, Dr. Gernot Poschet and Dr. Hagen Gegner, for offering me the opportunity to collaborate in their project and also contribute to financing the last period of my doctoral study.

My warm gratitude to the friends from my parallel master study, Tamara Bendig, Anne Newrly and Tamar Nizharadze for being the nicest classmates, and all the nice moments we've had throughout these years.

I would like to thank my dear old friend from the time we lived in Tehran, Elnaz Pegah, who was the best host right after I arrived in Heidelberg, and did not let me feel lonely when I was just new.

My smart and lovely friend, Navid (Emil), thank you for all the intellectual inputs for my brain which always brought my thoughts to the next level. And of course, my special thanks to Aristae, for all the warm sense of friendship, nice discussions and beautiful moments we've shared.

My doctoral study would not have been as easy without a group of friends who supported me whenever I needed. First of all, I would like to thank a friend, who by the way is exceptionally nerd, Dr. Yasamin Dabiri, for showing me how a high level scientist thinks, works, acts, and also makes a great friend. Thank you Yasi for all the good moments. I would like to thank my friend, Dr. Shirin Rahmanian, who is one of the most descent persons I've ever known in Heidelberg. We miss you Shirin here in Heidelberg. My friend and partner in crime during thesis writing and submission, Yalda Abadi, I don't know if I would have been able to submit my thesis without your help. Thank you for being very supportive, especially during the last days. My guys, Dr. Farshad Shobeiry and Dr. Reza Tale, thank you for always being up for having fun moments, not to mention the warm friendship we have. My dear Friend, Noujan Ganjian, I wish we knew each other earlier, but even during this short period, I enjoyed the shared moments with you beyond what I thought. Johanna, you are one of the loveliest crazy persons I've ever known, your energy is always a trigger for life. Thank you.

It was not only the support I received from my surrounding people in Heidelberg, but also from the ones overseas. I would like to thank my friend Saghar Zokaei for being one of the best friends I have. I also would like to dedicate my thanks to Melika Chehrazi for being such a nice and lovely friend.

My good friend, Dr. Sally Liechocki. You have been one of my best friends in this beautiful city. I experienced many beautiful moments with you. You are so special. Thanks for being such a good friend.

My friend, my dude, my always available buddy, Dr. Fereydoun Taheri. You're such a great friend. I'll never forget all the moments you were beside me. I know if I need a friend, you're the one to count on. This is such an exceptional quality. Thank you man!

My dearest friend in the city, Dr. Lilija Wehling. You have been an incredible friend for me throughout my whole doctoral study. Your kind and lovely character has been such an inspiration to me. Without you, these four and a half years would not have been the same. Thank you for all the great moments.

At the end, I would like to dedicate my sincere thanks to my great family. I would not have been here without their support throughout my whole life. I would like to thank my father for his never ending support. His appreciation for education and literacy has been the greatest inspiration in my life. I would like to thank my sister, who has always been happy for my achievements even more than myself. And I want to dedicate my greatest appreciation to the best mother in the world. Mom, you are the best person I've ever known. Your unlimited kindness, your pure emotions, your admiration for the finest human qualities has made me the person I am. For me, these are the greatest things I've learned in my life. I love you mom. You are always in my heart.

List of Publications

Seyed Babak Loghmani, Nadine Veith, Sven Sahle, Frank T Bergmann, Brett G Olivier, and Ursula Kummer. Inspecting the solution space of genome-scale metabolic models. Metabolites, 12(1):43, 2022.

Seyed Babak Loghmani, Eric Zitzow, Gene Ching-Chiek Koh, Andreas Ulmer, Nadine Veith, Ruth Großeholz, Madlen Rossnagel, Maren Loesch, Ruedi Aebersold, Bernd Kreikemeyer, et al. All driven by energy demand? integrative comparison of metabolism of enterococcus faecalis wildtype and a glutamine synthase mutant. bioRxiv, 2021. Accepted in Microbiology Spectrum

Abstract

Lactic acid bacteria are a group of bacteria that share the characteristic of lactate fermentation, and are in particular focus of microbiological research not only because of their involvement in human health, but also due to their role in the food industry. On the one hand, they can be used as probiotics, contributing to healthy micro flora in the human body. On the other hand, they can take part in the production of fermented foods and flavour development. *Enterococcus faecalis* and *Streptococcus pyogenes* are two lactic acid bacteria that cause several infections in the human body. Therefore, they have been in the focus of clinical studies for the past few decades. The rising trend of resistance to multiple antibiotics makes the treatment of the infections caused by theses two pathogens very hard. To overcome this progressive trend of resistance, it is important to find novel drug targets in these pathogens. In the present study, I investigated the metabolic characteristics of these two pathogens using an integrative method, comprising multi-omics data integrated with the respective genome-scale metabolic models under the conditions comparable to different tracts in the human body. First, I investigated the effect of glutamine auxotrophy on the metabolic adjustments of *E. faecalis* (in the case of a Δ*glnA* mutant) in response to a change in environmental pH, using an integrative approach combining metabolic and proteome data with genome-scale modelling. The result suggested that the higher energy demand in the Δ*glnA* mutant of *E. faecalis* is most likely due to the lack of control on glutamine transport system as a result of the absence of glnA in the mutant. In the next part, I developed a method for functional analysis of the solution space of the genome-scale metabolic models. This method employs random perturbation to discover the reliability of flux distribution in the network. Additionally, it allows to find out which type of experimental data is most effective in limiting the solution space when the data are used as constraints. Finally, I generated tract-specific genome-scale metabolic models for *E. faecalis* and *S. pyogenes* in order to find tract-specific drug targets in their metabolic networks. I used multi-omics profiles (metabolic, transcriptome and proteome data) obtained under the conditions comparable to natural physiological condition in the human body, namely root canal, unrinary tract and plasma, and used the data to constraint the respective genome-scale metabolic models. The models were used to find potential drug targets using different levels of threshold for metabolic flux values and growth rate of the bacteria. The results suggested that there exist potential drug targets in different subsystems in the metabolic network, from central carbon metabolism to transport system. The presented profiles of drug targets have to be validated experimentally in order to be used for the development of new treatment approaches.

Zusammenfassung

Milchsäurebakterien sind eine Gruppe von Bakterien, die das Merkmal der Milchsäuregärung teilen. Sie stehen nicht nur wegen ihrer Bedeutung für die menschliche Gesundheit, sondern auch wegen ihrer Rolle in der Lebensmittelindustrie im besonderen Fokus der mikrobiologischen Forschung. Sie können sowohl als Probiotika, die zu einer gesunden Mikroflora im menschlichen Körper beitragen, als auch zur Herstellung von fermentierten Lebensmitteln und Entwicklung von Aromen genutzt werden. *Enterococcus faecalis* und *Streptococcus pyogenes* sind zwei Milchsäurebakterien, die verschiedene Infektionen im menschlichen Körper verursachen. Aus diesem Grund stehen sie seit mehreren Jahrzehnten im Mittelpunkt klinischer Studien. Die Tendenz zur zunehmenden Resistenz gegen multiple Antibiotika macht die Behandlung von Infektionen, die durch diese beiden Erreger verursacht werden, sehr schwierig. Um diesen fortschreitenden Trend der Resistenz zu überwinden ist es wichtig neue Ansatzpunkte für Medikamente gegen diese Krankheitserreger zu finden. In der vorliegenden Studie untersuchte ich die Stoffwechseleigenschaften dieser beiden Erreger mit einer integrativen Methode, bei welcher Multi-Omics-Daten die in jeweiligen Stoffwechselmodellen auf Genomebene unter Bedingungen integriert wurden, die mit verschiedenen Abschnitten des menschlichen Körpers vergleichbar sind. Zu Beginn untersuchte ich die Auswirkungen der Glutamin-Auxotrophie auf die metabolischen Anpassungen von *E. faecalis* (im Falle einer Δ*glnA*-Mutante) als Reaktion auf eine Veränderung des pH-Werts in der Umwelt. Dazu verwendete ich einen integrativen Ansatz, der Stoffwechsel- und Proteomdaten mit einer Modellierung auf Genomebene kombiniert. Das Ergebnis deutet darauf hin, dass der höhere Energiebedarf in der ΔglnA-Mutante von *E. faecalis* höchstwahrscheinlich auf die fehlende Kontrolle des Glutamintransportsystems infolge des Fehlens von glnA in der Mutante zurückzuführen ist. Im folgenden Teil habe ich eine Methode zur funktionellen Analyse des Lösungsraums der genomweiten Stoffwechselmodelle entwickelt. Bei dieser Methode werden zufällige Störungen genutzt, um die Zuverlässigkeit der Flussverteilung im Netz zu ermitteln. Zusätzlich erlaubt sie es herauszufinden, welche Art von experimentellen Daten den Lösungsraum am effektivsten begrenzen, wenn diese als Einschränkungen (not sure how to translate constrains the best) verwendet werden. Schließlich habe ich traktspezifische Stoffwechselmodelle auf Genomebene für E. faecalis9 und *S. pyogenes* erstellt, um traktspezifische Wirkstoffziele in deren Stoffwechselnetzwerken zu ermitteln. I would change this sentence to: Ich verwendete Multi-Omics-Profile (Stoffwechsel-, Transkriptom- und Proteomdaten), die unter Bedingungen gewonnen wurden, die mit den natürlichen physiologischen Bedingungen des Wurzelkanals, Harnwegs und Plasmas im menschlichen Körpers vergleichbar sind und nutzte diese Daten, um die jeweiligen

genomweiten Stoffwechselmodelle einzuschränken. Die Modelle wurden eingesetzt, um potenzielle Wirkstoffziele in Abhänigkeit verschiedener Grenzwerte für den Stoffwechselfluss und die Wachstumsrate der Bakterien zu finden. Die Ergebnisse weisen darauf hin, dass es potenzielle Wirkstoffziele in verschiedenen Teilsystemen des Stoffwechselnetzes gibt, angefangen beim zentralen Kohlenstoffmetabolismus bis hin zu Transportsystemen. Die dargelegten Profile von Wirkstoffzielen müssen experimentell validiert werden, um für die Entwicklung neuer Behandlungsansätze genutzt werden zu können.

Contents

List of Tables

List of Figures

Abbreviations

ΔglnA	Glutamine synthetase mutant
ACALDt	Acetaldehyde reversible transport
acety-ser	Acetyl serine
ALCD2x	Alcohol dehydrogenase (ethanol: NAD)
ASNTAL	Aspartyl-tRNA(Asn):L-glutamine amido-ligase (ADP-forming)
CK	Carbamate kinase
CYSTL	Cystathionine b-lyase
FBA	Flux balance analysis
fpram	2-(Formamido)- N1-(5-phospho- D-ribosyl) acetamidine
fru	Fructose
FVA	Flux variability analysis
gam6p	Glucoseamine 6 phosphate
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GEM	Genome-scale model
GF6PTA	Glutamine-fructose-6-phosphate transaminase
GHMT	Glycine hydroxymethyltransferase
glc	Glucose
glnL	L-glutamine
glnA	L-glutamine synthetase
GlnR	Regulatory protein GlnR
GLUPRT	Glutamine phosphoribosyldiphosphate amidotransferase
gmp	Guanosine monophosphate
GMPS2	GMP synthase
ILEt6	Isoleucine transporter
LAB	Lactic acid bacteria
LDH	Lactate dehydrogenase
LDHL	L-lactaate dehydrogenase
Lib	Library
METabc	Methionine ABC transporter
mt	Mutant
MTHFD	Methylenetetrahydrofolate dehydrogenase (NADP)
OCBT	Ornithine carbamoyltransferase
PCA	Principal component analysis

PDH	Pyruvate dehydrogenase
PFK	Phosphofructokinase
PFL	Pyruvate formate lyase
PGDH	6-phosphogluconate dehydrogenase
pram	5-Phospho-beta-D-ribosylamine
PRFGS	Phosphoribosylformylglycinamidine synthase
PRPPS	Phosphoribosylpyrophos- phate synthetase
PTS	Phosphoenolpyruvate:carbohydrate phosphotransferase system
PYK	Pyruvate kinase
pyr	Pyruvate
serL	L-serine
SERAT	Serine O-acetyltransferase
SERD	Serine dehydrogenase
sertrna	Serine-tRNA
suc	Sucrose
THRD	L-threonine deaminase
TPI	Triose-phosphate isomerase
VALt6	Valine transporter
wt	Wildtype

Chapter 1

INTRODUCTION

1.1 Systems biology

Studying the characteristics of living organisms has been the focal point of many research projects. For the most part, after the development of the modern scientific approaches emerging in the 19th century, living organisms were subjected to investigations that were focused on unravelling the features of their individual components. However, the invention of high-throughput technologies boosted the process of data generation on numerous components at the same time. The development of genomics facilities has resulted in whole genome sequencing of various organisms. On the other hand, the application of other technologies, such as mass-spectrometry, and more recently, next generation sequencing resulted in the analysis of biological phenotypes at multiple scales, both qualitatively and quantitatively. Following these developments, we are currently able to quantify the expression level of a desired set of genes at the level of mRNA or protein, or measure the concentration of certain metabolites or signalling components in a living organism. Taking advantage of these breakthrough developments and large data sets, we have been able to switch from an individualistic to a collective standpoint, which enhances the already existing comprehension of living organisms as systems.

A biological system like any other system comprises a massively organized structure and function which can be decomposed into different levels of complexity and studied at different scales. From this perspective, a biological system can be investigated at the scale of an organ, a tissue, a cell or a pathway, for instance. Each of these scales can then be broken up into its components in order to determine the structure of the system. Moreover, each system can be studied at different levels. A metabolic pathway can be studied either at the metabolic level by measuring the concentration of the involved metabolites, or at the proteome level by quantifying the expression level of enzymes. Depending on the question in place, a certain single or multi-omics profile would help to find the best answer. To make the study of a biological system more efficient, it is important to keep the following principles in mind.

The information flow in biological systems was determined by Francis Crick in 1970, known as central dogma [Crick, 1970]. He explained how the information stored in DNA translates into the functional component, namely protein, through mRNA transcription. The information flow of course is not exclusive to this simplistic view, and occurs in other forms as well. An

interesting aspect of central dogma is that from DNA to protein, the universality of information decreases, while the specificity increases. In a broader spectrum, ranging from the information storage part (DNA) to the functional parts, whether it is a single protein, or a signalling or metabolic pathway, units are more universal when they are dealing with information storage and more specific when they are functional. That is, while a certain open reading frame (ORF) might exist in several organisms, various protein isoforms might be expressed in each organism due to reasons such as post translational modification [Brett et al., 2002] or single nucleotide polymorphism (SNP) [Lander et al., 2001]. Moreover, the expression level of the identical proteins might be different between various organisms due to dissimilar regulatory regimes. Therefore, while the specificity increases along with the functionality of units, the information storing units are quite universal among biological systems.

Taking this scheme into account, the more universal part of the central dogma is a very practical resource to reconstruct the draft of any biological network (bottom-up approach) [Palsson, 2006], since the universality of genetic information allows a rough determination of the genetic products by means of bioinformatics tools such as sequence alignment. However, in order to make the reconstructed network organism-specific, the refinement of the network using the data from the more specific part is essential (top-down approach). In essence, a biological system is characterized by the connection between its components [Palsson, 2006]. While components come and go, a persistent characteristic of any complex system is the interconnection between its different parts [Palsson, 2006]. As the number of components in a system grows, the number of potential connections increases exponentially. Since not all of these connections occur in reality, it is crucial to eliminate the unrealistic ones [Palsson, 2006]. Therefore, the integration of the bottom-up and top-down approaches helps to come up with a well designed structure which can be further used to predict the function [Palsson, 2006].

Defining the structure of any biological system is the first step in simulating/predicting its behaviour. Structure determines what are the inputs of the system, how the input material is processed through the network, and what are the outcomes of it. In a signalling network, the input of the system might be an extracellular signal, triggering a signalling pathway in a cell. The input signal is then processed through a pathway by for example phosphorylation/dephosphorylation of various proteins which eventually alter the expression level of a certain gene [Ardito et al., 2017]. In a metabolic network, the input can be a carbohydrate, serving as an energy resource for the cell. The carbohydrate will be metabolized through glycolysis and produces energy or converts to other metabolites to produce other necessary components of a cell [Alberts et al., 2003]. A comprehensive network structure brings the opportunity to consider all the possibilities in a network and to predict the physiologically relevant functional states. However, defining the structure is far from enough when it comes to determining the dynamics of a system and predicting the correct functions.

A key step in predicting the behaviour of biological systems is to find out the rates at which system's components are interconverted to each other, all the way from the input to the output [Klipp et al., 2016]. This information helps to determine what a real functional state looks

like, and what phenotypes are likely to emerge as a consequence. Acquiring such information requires a proper experimental setup under desired conditions. As a result, the dynamics of, e.g., protein phosphorylation in a signalling pathway, or the uptake rate of a carbohydrate in a metabolic pathway can be discovered [Klipp et al., 2016]. The integration of these data results not only in a better fit of the parameters in a model, but also increases the accuracy of phenotype prediction [Klipp et al., 2016].
By reconstructing well structured networks that can successfully simulate the functional states of a given system, an integrative analysis of biological systems becomes feasible. Integrative analysis has become more popular in the scientific community, as it enables scientists to predict phenotypes at the system level, which is a huge advantage over individualistic approaches. This integrative approach, meaning the combination of preferably but not exclusively large data sets at the system level, with mathematical and computational approaches which enable the simulation of a biological phenomenon, is referred to as systems biology.

1.1.1 Mathematical modelling in biology

Mathematical models are a key part of the systems biology approach. The outcome of a complex system cannot be predicted by intuition, since the resultant phenotype of any system is not equivalent to the sum of its parts [Klipp et al., 2016]. Mathematical models are a tool to make the integrative analysis of a complex system possible. This is something that a purely experimental approach is unable to do. Experiments in general help to generate hypotheses about a phenomenon and the ways it emerges, but they often come up short in providing a global overview of the whole process. Using mathematical models, we are able to achieve a more comprehensive knowledge about the dynamics of the system, the nature of the internal processes and test the experimentally generated hypotheses [Klipp et al., 2016].
A mathematical model is used to answer specific questions [Klipp et al., 2016], [Palsson, 2006]. Therefore, a model should be developed according to the problems to be solved and its level of confidence as well as limitations should be clear. To achieve this goal, the structure of the model should be faithfully taken from the real system and the experimental data integrated into the model should be chosen specifically with respect to the question in place [Klipp et al., 2016], [Palsson, 2006]. The level of complexity in a model is not a criterion for a model to be right or wrong, it just determines the extent to which a model can make predictions [Klipp et al., 2016]. The model should be able to predict the functional state of the system with acceptable precision, which should be further validated experimentally or by extrapolations from comparable systems.
A biological process can be modelled using different approaches, depending on the nature of the process, the problem to be solved and the availability of the data. If the time course of a metabolic or signaling pathway is going to be simulated, kinetic models using ordinary differential equations (ODE) are powerful methods to mimic the time course [Klipp et al., 2016]. If the spatial information of the system is also a point of interest, then spatio-temporal modelling using partial differential equations (PDE) will be the correct approach to use [Resat et al.,

2011]. On the other hand, if the steady-state flux distribution of a metabolic system (at large or small scale) is the aim of study, constrained-based models provide the respective answers to the question [Klipp et al., 2016], [Palsson, 2006]. Furthermore, other modelling formalisms such as agent-based models (ABM) can be used if the interaction of autonomous parts is the point of interest [Klipp et al., 2016]. In the paragraphs ahead, I will elaborate on the background and the applications of ODE models with a particular focus on metabolism. The constraint-based modelling, which is the approach I used for this project, will be explained separately in more detail in the next section.

The kinetic models find their roots in the early years of the twentieth century, when Leonor Michaelis and Maud Menten proposed a mathematical model for enzyme activity [Michaelis et al., 1913]. This model described the rate of an enzymatic reaction by forming a relationship among substrate concentration, the maximum rate at which an enzyme can catalyse a reaction (Vmax), and Km, which is a numerical constant, representing the concentration of the substrate at which the reaction rate is half of the Vmax. Using Michaelis-Menten kinetics as a foundation, a series of various kinetic formula have been proposed and applied to simulate different metabolic pathways in different organisms.

Kinetic models have been widely used to simulate different biological processes. Among all, modelling the central carbon metabolism and glycolysis for different organisms have been the subject of numerous studies. Hess and Boiteux showed that the oscillation in glycolysis in yeast is independent of the current state of the cell and can emerge by the addition of a variety of substrates [Hess and Boiteux, 1968]. Cortassa and Aon applied metabolic control analysis (MCA) to determine the rate-controlling steps in glycolysis, in aerobic chemostat cultures of *Saccharomyces cerevisiae* [Cortassa and Aon, 1994]. In another study on *S. cerevisiae*, kinetic models were used to assess the effect of a glucose pulse on glycolysis in a steady-state continuous culture [Rizzi et al., 1997]. Furthermore, glycolytic pathway in prokaryotes heve also been modelled using kinetic laws. Chassagnole and colleagues developed a kinetic model for central carbon metabolism in *Escherchia Coli* that connected the phosphotransferase system (PTS) to glycolysis and pentose-phosphate shunt, allowing to simulate the transient concentration of intracellular metabolites [Chassagnole et al., 2002]. The model was also used to perform MCA to gain an insight into the functional characteristics of the system [Chassagnole et al., 2002]. In another project, Levering and colleagues used kinetic modelling to study the role of phosphate in central carbon metabolism and its role in glucose uptake in two lactic acid bacteria, *Lactococcus lactis* and *Streptococcus pyogenes* [Levering et al., 2012]. On the other hand, kinetic models can be used to spot targets for metabolic engineering. Andreozzi and colleagues employed large-scale kinetic models of a recombinant *Escherichia coli* to increase the production of 1,4-butanediol. They reported that the computationally detected targets were in good agreement with experimental results [Andreozzi et al., 2016]. Kinetic modelling has been also used for drug target identification in the case of numerous pathogens. To name a few, Haanstra and colleagues employed kinteic modelling to find drug targets in the metabolic network of *Trypanosoma brucei* (a parasite causing sleeping sickness) that do not interfere with host me-

tabolism. They computationally identified and experimentally verified that glyceraldehyde-3-phosphate dehydrogenase and glucose transport are very effective drug targets to combat this parasite [Haanstra et al., 2017]. Kinetic models were also used to determine the metabolite enzyme interactions in micro organisms. Link and colleagues performed metabolite tracing experiments in the example of *E. coli* to identify the allosteric interaction of metabolites and proteins. They fit the data to a kinetic model of *E. coli* comprising glycolytic reactions, suggesting the allosteric interactions involved in the glycolysis/gluconeogenesis switch, one of the being the interaction between pyruvate and fructose-1,6-bisphosphatase [Link et al., 2013]. In another study, Hackett and colleagues found three previously unknown cross pathway regulatory interactions in yeast, using a kinetic model [Hackett et al., 2016]. The model showed that pyruvate kinase (PYK) and pyruvate decarboxylase are inhibited by citrate and phenylpyruvate, respectively. The results were validated experimentally [Hackett et al., 2016]. Of course, the examples of the application of ODE models are not limited to the ones that were mentioned in this paragraph, and the extent of research subjects using this method is very large.
While kinetic models are usually focused on small-scale metabolic pathways, large-scale pathways simulation is often done using another modelling formalism called constraint-based modelling. In the absence of detailed kinetic data, constraint-based models provide a platform to discover the metabolic characteristics when pathways are analysed in ,e.g., genome-scale.

Constraint-based modelling

Constraint-based modelling is an approach for simulating the flux distribution of metabolic networks at steady-state. The objective function of a constraint-based model, which can be the production rate of biomass, or uptake rate of a certain metabolite, can be optimised by fulfilling a set of constraints [Palsson, 2006]. The primary constraint in such a model is the mass and charge conservation [Palsson, 2006]. As a consequence, the exact same amount of mass and charge entering the system must be converted to biomass or exported out of the system. This shows the importance of the exchange reactions in constraint-based models, as they determine the amount of mass and charge that can be imported in and exported from the system. Another type of constraints in a constraint-based model are the thermodynamic constraint, determining the the reversibility of reactions [Palsson, 2006]. In addition to these intrinsic constraints, there are also user-defined constraints which can considerably increase the accuracy of flux distribution in a model, such as reaction flux boundaries. The upper and lower boundaries of reactions can be adjusted to direct the flux towards the intended directions based on experimental data, whether it is a metabolite conversion rate, or gene expression data. The aggregation of all these constraints determines the outcome of a constraint-based model. Constraint-based models can be developed either at the level of a single pathway, or at genome-scale. Since in this thesis I present the results from two constraint-based models at genome-scale, for the rest of this thesis I use the term genome-scale model (GEM).
To solve a GEM mathematically, the system has to be converted into mathematical terms. In the constraint-based approach, metabolic reactions are translated into a stoichiometric matrix in

which rows correspond to metabolites and columns correspond to reactions. The stoichiometric matrix is then used to calculate the flux distribution using the following equation:

$$S.v = 0$$

where S is the stoichiometric matrix, and v is the vector of flux which satisfies the steady-state assumption.This equation, also known as general equation [Llaneras and Picó, 2008], maintains the mass balance constraint for the whole system, one of the most fundamental constraints that significantly determines the outcome of the system.

GEMs have been used for a variety of tasks in the past few decades. They are used for metabolic engineering in cell factories, drug target identification, studying host-pathogen interactions, investigation of novel enzymatic functions, etc. Therefore, a broad spectrum of approaches can be used to investigate the metabolic capabilities of an organism by means of GEMs. For instance, if metabolic engineering is intended, OptKnock, OptGene and RobustKnock can be used to consider single or multiple gene deletion scenarios[Simeonidis and Price, 2015]. The *in silico* generated scenarios can then be used to manipulate the organism of interest experimentally, to increase the production of certain metabolites, gain/loss of function or other engineering tasks. For example, a previously developed *E. coli* GEM [Orth et al., 2011] was used for GEM-directed metabolic engineering of *E. coli* to increase the production of D-phenyllactic for aromatic polymers production[Yang et al., 2018]. In another study, a GEM of *Yarrowia lipolytica* was used to apply strain design strategies to enhance the production of dodecanedioic acid [Mishra et al., 2018]. GEM-directed metabolic engineering was also used in the case of eukaryotic organisms. A GEM of *S. cerevisiae* was used by Bordina and colleagues to simulate the effect of introducing a new synthetic pathway, enabling the de novo biosynthesis of 3-hydroxypropionic acid in this organism [Borodina et al., 2015].

Another important application of GEMs is in drug target identification. Using GEMs, it is possible to spot essential genes and reactions. The suppression of essential genes results in the growth inhibition of pathogens. On the other hand, one can find the essential reactions whose inhibition stops the growth of pathogens. However, a more realistic approach would be to find the reaction whose flux reduction to a certain extent limits the growth of a pathogen, as neither the enzyme nor growth inhibition can be achieved in absolute form. In one of the earliest attempts to investigate drug targets using GEMs, Beste and colleagues probed the gene essentiality of known drug targets in *Mycobacterium tuberculosis* [Beste et al., 2007]. They predicted an additional 220 essential genes using their models, suggesting potential drug targets to overcome the multi-drug resistance of *M. tuberculosis* [Beste et al., 2007], [Espinal, 2003], [Raviglione et al., 2001]. In another study, a set of 26 non condition-specific essential genes and 17 combinations of 21 condition-specific genes were identified in *Pseudomonas aeruginosa*, without having homology with any human gene [Sigurdsson et al., 2012].

The comparison GEM-generated essential genes and experimental data can bring a deep insight into the predictability of GEMs and the confidence level of the predicted drug targets. For this matter, Kim and colleagues performed a gene essentiality analysis on a GEM of *Acinetobacter*

baumannii AYE and compared the results to the experimental observations from single-gene deleted mutants of *A. baylyi ADP1* [Kim et al., 2010]. They found that the predicted profile of essential genes has an overall 72% consistency with the experimental data. Interestingly, the *in silico* profile was always smaller than the experimental profile (meaning that the number of computationally predicted targets was smaller than the experimentally detected ones), both when it was compared at the whole-metabolism level, or at the individual subsystem level. They also performed metabolite essentiality analysis and reported a set of 9 metabolites that do not interfere with human metabolism, and the enzymes for which these metabolites were a substrate were reported as potential drug targets [Kim et al., 2010]. In another investigation, using the same approach to screen essential metabolites, the authors recognized five essential metabolites in the opportunistic human pathogen *Vibrio vulnificus* which were further experimentally validated to be essential for the cell survival [Kim et al., 2011]. GEMs also provide a powerful platform to discover different drug responses, not only between different species in the same genus, but also between different life cycles of a single organism. A GEM-driven drug response discovery on the *Plasmodium* genus determined stage- and species-specific drug targets in these parasites [Abdel-Haleem et al., 2018]. They reported that the stage-specific models showed differences in central carbon metabolism, while species-specific did so in thiamine, choline and pentothenate metabolism.

Linear optimization

To solve the general equation, which was introduced earlier in this chapter, several algorithms have been developed based on linear optimization, allowing the calculation of the flux vectors. The most famous one is called flux balance analysis (FBA) [Orth et al., 2010]. In FBA, a vector of flux distribution is calculated following the optimisation of the objective function, which by no means is unique. In fact, as the system is mathematically underdetermined (considering that there are more reactions (n) than metabolites (m) in the system and therefore more unknown variables than equations), solutions can be found in a convex space, usually referred to as the solution space [Orth et al., 2010], [Llaneras and Picó, 2008], [Famili and Palsson, 2003]. The solution space contains all the feasible flux distribution profiles resulting in the optimal value of the objective function. As there is no unique solution for the general equation, alternative optimal values exist for some of the reactions, which can be found using flux variability analysis (FVA) [Mahadevan and Schilling, 2003]. FVA tries to find the feasible flux interval for each reaction by optimising the objective function in the first place, and using the value as an additional constraint. This results in two additional values for variable reactions, minimum and maximum boundaries. An FVA problem can be formulate as follows:

$$Max f^T v,$$

$$s.t.S.v = 0,$$

$$v_{min} \leq v \leq v_{max}$$

where f is the vector of the objective function and vmax and vmin are the vectors of maximum and minimum allowable flux values, respectively, for each reaction [Mahadevan and Schilling, 2003]. FVA is therefore the most straightforward approach to investigate the solution space at the level of individual reactions.

Solution space of constraint-based models

The solution space of constraint-based models is a consequence of having more unknown variables(reactions) than equations (since there are more reactions than compounds) in the system. In a stoichiometric matrix of a metabolic network with m row and n column (corresponding to metabolites and reactions, respectively), the system has n-m degree of freedom [Orth et al., 2010], resulting in a very large solution space. This solution space is convex [Famili and Palsson, 2003], containing all the flux distribution profiles resulting in the optimal value of the objective function and, therefore, all these profiles are considered as optimal. When the system is solved using a linear algorithm like FBA, one single flux distribution profile is arbitrarily reported as an optimal distribution. However, it is crucial to keep in mind that it is semantically wrong to refer to this profile as 'the optimal', since there are too many other profiles that are equally optimal. Generally speaking, a large solution space emerges from having numerous alternative optima for individual reactions. The combination of these alternative values results in a huge number of solutions. However, the alternative optima does not exist for all the reactions in the network. Depending on the degree to which the system is constrained, the number of variable reactions differs.

The presence of various alternative solutions limits the reliability of the outcome of FBA. Therefore, to increase the predictability of a model, it is important to shrink the solution space by integrating additional constraints, which are ideally obtained from experimental data. Additional constraints can be applied in different ways. A very common method is to integrate the uptake and production rate of metabolites into the boundaries of the respective exchange reactions [Veith et al., 2015], [Levering et al., 2016], [Großeholz et al., 2016]. This is a very efficient approach to increase the predictability of a GEM, as it determines the input and output of mass in the system. Adjusting the constraints which determine the flow of mass into the system has a significant impact on the outcome of optimization, as it deals with the mass-balance constraint, which is one of the most fundamental constraints in GEMs. On the other hand, constraints can be applied at the level of internal reactions. One possibility is to deactivate the reactions for which an enzyme was not observed in the expression profile. With this approach, one can employ a proteome profile, for instance, and deactivate the absent enzymes [Großeholz et al., 2016]. Moreover, the flux boundaries of the internal reactions can also be subjected to adjustment based on the experimental measurements. Therefore, different experimental results such as fluxome, transcriptome or proteome profiles can be used for this purpose. To increase the precision of flux distribution using fluxomics profiles, a method called metabolic flux analysis

(MFA) has been developed which incorporates flux values obtained by labeled tracer to estimate the flux distribution in a network [Crown and Antoniewicz, 2013], [Leighty and Antoniewicz, 2013], [Swarup et al., 2014]. The major problem with MFA was that the precision of flux estimation drops, going from central to peripheral metabolism. However, different methods have been developed in recent years to expand the accuracy of flux distribution to the genome-scale [García Martín et al., 2015], [McCloskey et al., 2016]. In a different approach, the internal flux boundaries can be adjusted also based on the transcriptome profiles [Kim and Lun, 2014]. For example, a method (PROM) has been developed to adjust the maximal flux value of a reaction by a factor of expression probability of its enzyme relative to the activity of its transcription factor [Chandrasekaran and Price, 2010]. In another method (MADE), the flux boundaries of reactions are adjusted based on the differential expression of the respective genes from one condition to another [van Berlo et al., 2009]. This is the assumption that is employed in this study to integrate the transcriptome data into the models. Finally, the internal flux boundaries can be adjusted based on the comparative proteomics, adjusting reaction boundaries relative to the change in enzyme abundances [Großeholz et al., 2016]. This is the approach I used in this thesis to incorporate the proteome data into the models.

Solution spaces have been studied from various perspectives. Using convex analysis, two fundamental types of pathways, namely elementary flux modes [Schuster and Hilgetag, 1994], [Schuster et al., 1999], and extreme pathways [Schilling et al., 2000] have been determined. The identification of the elementary flux modes resulted in the calculation of the minimal set of fluxes resulting in the optimal value of the objective function. This approach was used to redirect fluxes to enhance the production of aromatic amino acids in *E. coli* [Liao et al., 1996]. Extreme pathways, on the other hand, are a unique set of systematically independent biochemical pathways (vectors) that fully characterize the steady-state flux distribution of a network [Schilling et al., 2000]. This concept was recently used to perceive the regulatory structures from metabolic pathways [Xi and Wang, 2019]. In a fundamental analysis, Famili and Palsson showed that the left null space of the stoichiometric matrix contains information about the time invariant metabolite concentrations in combination, which can be further used for classification of the metabolic networks [Famili and Palsson, 2003]. For instance, metabolically relevant pools can be determined using a convex basis for the null space [Famili and Palsson, 2003]. Moreover, they classified metabolic pools into three different categories according to the incorporation of primary metabolites (molecules that undergo conversion, Type A), primary and secondary metabolites (the combination of primary and secondary metabolites, Type B), and only secondary metabolites (cofactors, Type C) [Famili and Palsson, 2003]. Type A pools account for the conservation of biochemical elements and therefore contain only primary metabolites, which are analogous to type I pathways (flux pathways). Type B pools determine the conservation of exchanged biochemical moieties and incorporate both primary and secondary metabolites. Type B pools are analogous to Type II pathways (futile cycles coupled to cofactors with internal input and output). Type C pools account for cofactor conservation and only incorporate secondary metabolites and are analogous to Type III pathways (internal cycles

with no input and output) [Famili and Palsson, 2003]. All these were possible using the analogy of the left null space to the right null space [Famili and Palsson, 2003]. The metabolic pools can then be used to visualize pathways and to find out which reactions take part in the conservation relationships [Famili and Palsson, 2003].

From a new perspective, Kelk and colleagues developed a method to study the solution space which defines a set of modules comprising variable reactions that are linearly independent [Kelk et al., 2012]. This method divides various flux routes into three categories, vertices, rays and linealities, which correspond to paths, reversible and irreversible cycles in a model, respectively. Therefore, reactions in the same modules are the ones whose flux values are correlated across all vertices, regardless of appearing in the same flux route or not. Solution space has also been extensively analysed by methods based on Monte-Carlo sampling. The majority of Monte-Carlo methods focus on analysing the solution space with respect to individual reactions. For instance, these methods have been used to calculate the probability distribution of individual flux values or defining correlated reaction sets [Price et al., 2004]. Correlated reaction/gene sets reportedly enable defining the regulatory networks [Papin et al., 2004]. Moreover, Monte-Carlo sampling has been used to calculate the average and standard deviation of flux values, allowing the comparison between different conditions and to compare the simulation results to the transcriptome profiles [Bordel et al., 2010]. In another study, sampling the solution space was applied to discover the steady-state flux distributions that are consistent with experimental data in human mitochondrion [Thiele et al., 2005]. A major drawback of the Monte-Carlo methods is the need for a very large sample size and, consequently, expensive computations. However, there have been some attempts recently to increase the efficiency of these methods with respect to computation time [Fallahi et al., 2020]. For example, coordinated hit-and-run with rounding (CHRR) is a computationally efficient method that is used to study the properties of the bacterial GEMs [Fallahi et al., 2020], [Pinhal et al., 2019]. Furthermore, sampling the solution space has been used to investigate the impact of changing ATP maintenance (ATPm), biomass precursors and metabolites uptake/production rates following temporal fluctuations in environments, on the outcomes of GEMs [Dinh et al., 2021]. The authors then suggested that the integration of these constraints significantly increases the predictability of a model [Dinh et al., 2021].

Strictly speaking, a large solution space would mean that many of the solutions correspond to biologically unrealistic phenotypes. For example, it has been reported that the occurrence of internal loops in a model can result in the generation of energy resources like ATP, or in the interconnectivity of two electron pools [Maranas and Zomorrodi, 2016]. Therefore, the large size of the solution space of GEMs is considered as one of the major drawbacks of this modelling platform. This signifies the importance of shrinking the solution space by means of case-specific experimental data.

1.2 Lactic acid bacteria

As previously explained by several examples, mathematical modelling and more specifically, constraint-based modelling has been used to simulate and investigate the metabolic characteristics of different microorganisms including bacteria. A group of bacteria that are of significant importance, not only because of their role in human health, but also in the food industry, are lactic acid bacteria. So far, a number of lactic acid bacteria have been studied using different mathematical modelling formalisms [Veith et al., 2015], [Levering et al., 2016], [Großeholz et al., 2016], [Levering et al., 2016], [Teusink et al., 2006]. In this thesis, I employed constraint-based modelling to study the metabolic network of two lactic acid bacteria, namely *Enterococcus faecalis* and *Streptococcus pyogenes*. In the rest of this chapter, I give a general introduction on lactic acid bacteria, and introduce the two aforementioned species in more detail.

Lactic acid bacteria (LAB) are a large group of bacteria sharing the characteristic of lactate fermentation as the main strategy to produce energy [Holzapfel and Wood, 1995]. These bacteria are gram-positive, low G+C content organisms, non-respiring cocci or rods that are classified based on their physiological and metabolic characteristics [Franz et al., 1999], [Dworkin, 2006], [Salminen and Von Wright, 2004], [Stiles and Holzapfel, 1997]. This group comprises a wide variety of species, classified into different genera such as *Enterococcus*, *Streptococcus*, *Lactococcus*, *Lactobacillus*, *Leuconostoc*, *Carnobacterium*, *Enterococcus*, *Wissella*, *Oenococcus* and *Pediococcus* [Stiles and Holzapfel, 1997]. Lactic acid bacteria are in particular focus of microbiological research, not only because of their pathogenic characteristics, but also due to their role in the food industry [Schillinger et al., 1996]. In general, those strains of lactic acid bacteria that are used in the food industry are referred to as GRAS (generally regarded as safe) and can be used as, for example, probiotics [AFRC, 1989] or cheese starter [Dahlberg and Kosikowsky, 1948]. These bacteria can also be used for food preservation as a result of producing lactic acid and other organic acids; flavour formation and also the production and secretion of bacteriocins, which reduce the growth of pathogens and therefore sustain a high quality in microflora [Schillinger et al., 1996], [Stiles and Holzapfel, 1997], [Leroy and De Vuyst, 2004], [Franz et al., 1999]. Probiotic lactic acid bacteria mostly belong to the genera of *Lactobacillus* and *Bifidobacterium* and are prominently used in milk-based products [Lavermicocca et al., 2021]. To name a few species that are widely used in the food industry: *Lactobacillus lactis* is used in the production of various types of cheese [Leroy and De Vuyst, 2004], *Lactobacillus plantarum* is employed in the fermentation of pickles, olives and sauerkraut [Leroy and De Vuyst, 2004], and *Lactobacillus delbrueckii* and *Streptococcus thermophilus* are used in the cheese and yoghurt production [Leroy and De Vuyst, 2004]. The use of probiotic lactic acid bacteria facilitates the growth of beneficial species and restrains the growth of pathogens by modulating the constituents of the host intestinal immune system, which might boost the immune response in the host [Chugh and Kamal-Eldin, 2020]. The probiotic species, however, must be screened to make sure of the absence of antimicrobial resistance genes and transferable virulence factors [Baccouri et al., 2019].

On the other hand, there are numerous species in the group of lactic acid bacteria that are serious pathogens for human [Stiles and Holzapfel, 1997] and cause serious problems in hospital environments. The role of several lactic acid bacteria in human health was reviewed in the 1990's by Aguirre and Collins [Aguirre and Collins, 1993], [Aguirre and Collins, 1993], bringing the attention of the community to the important role of this group in bacterial infections. Pathogenic lactic acid bacteria belong to different genera, such as *Enterococcus* and *Streptococcus*. In the present study, I investigated the metabolic characteristics of *Enterococcus faecalis* and *Streptococcus pyogenes*, using an integrative approach, comprising the integration of multi-omics data into the genome-scale models of these two bacteria. In the paragraphs ahead, I introduce these two species and denote their most relevant characteristics.

1.2.1 *Enterococcus faecalis*

E. faecalis belongs to the group of lactic acid bacteria and is usually referred to as an opportunistic pathogen [Franz et al., 1999], [Moreno et al., 2006], [Murray, 1990]. Depending on the strain, *E. faecalis* can be considered as a pathogen, a probiotic or a food-production supplement. This organism can be used as a starter for cheese production [Dahlberg and Kosikowsky, 1948], or as a probiotic, which can contribute to health [AFRC, 1989]. *E. faecalis* is a commensal organism which is a natural inhabitant of the human gastrointestinal tract, where it can take part in the maintenance of a healthy flora. Conversely, *E. faecalis* can cause serious infections in different human tracts, such as urinary tract infection, bacteremia, endocarditis, root canal and wound infection [Kristich et al., 2014], [Kau et al., 2005]. Many strains of *E. faecalis* have developed resistance against multiple antibiotics, making the treatment of the infections caused by this pathogen very hard [Kau et al., 2005]. *E. faecalis* makes up a large part of the nosocomial infections in hospital environments [Pöntinen et al., 2021], such as post surgical contaminations [Franz et al., 1999].

Comprising one chromosomal DNA and three plasmids, the genome of *E. faecalis* contains 3337 open reading frames (ORF) with a G+C content of 37.5% (in the chromosome) [Paulsen et al., 2003]. More than a quarter of its genome is made of probable mobile or foreign DNA, being characterized by seven probable integrated phage regions, 38 IS (insertion sequence) elements, a putative pathogenicity island, multiple conjugative and composite transposons and integrated plasmid genes [Paulsen et al., 2003]. This very diverse genome composition might be the underlying reason for the very high survival capability of *E. faecalis*, resonating in its ability to colonize various environments and resist to antimicrobial agents. The natural habitats of *E. faecalis* range from the gastrointestinal tract of human and animals, to water, plant material and soil [Moreno et al., 2006]. This organism is intrinsically resistant to several antibiotics, such as vancomycin [Cetinkaya et al., 2000], [Kafil and Asgharzadeh, 2014] and cephalosporins [Snyder et al., 2014]. In general, *E. faecalis* owns three major mechanisms of resistance against antibiotics, namely beta-lactam resistance, vancomycin resistance and aminoglycoside resistance [Cetinkaya et al., 2000]. Moreover, *E. faecalis* acquires more flexibility and stress resistance by producing biofilms and bacteriocins such as enterocines [Snyder et al., 2014], [Jayaraman,

2009]. *E. faecalis* is reportedly the most prominent cause of enterococcal infections and its treatment is particularly hard due to its resistance to numerous antibiotics [Cetinkaya et al., 2000], [Sahm et al., 1989], [Norrby et al., 2005].
The very versatile metabolism of *E. faecalis* enables its growth on various carbon sources, ranging from primary ones such as glucose, sucrose and fructose, to the secondary ones such as lactate, malate, citrate, amino acids and glycerols [De la Maza et al., 2013]. Interestingly, similar to several other bacteria, *E. faecalis* is able to employ the "phpshoenolpyruvate phosphotransferase system" (PTS) system to pair sugar uptake with phosphorylation, resulting in a very effective use of energy resources in comparison to non-PTS sugar uptake [Gilmore et al., 2002]. *E. faecalis* can withstand ethanol, heavy metals, azide [Stuart et al., 2006], and oxidative stress [De la Maza et al., 2013]. In order to combat the infections caused by this highly flexible pathogen, a system-wide analysis of its metabolic characteristics is essential.

1.2.2 *Streptococcus pyogenes*

S. pyogenes, also known as group A streptococcus (GAS), is a non-sporeforming and non-motile, facultative anaerobic bacteria, triggering various infections in human body. This organism triggers infections such as scarlet fever, puerperal sepsis, bacteremia, pharyngitis, impetigo, pneumonia, streptococcal toxic shock (STSS), root canal infection, blood stream infection, endocarditis and necrotizing fasciitis [Walker et al., 2014]. *S. pyogenes* also causes immune-mediated post-infectious disorders like rheumatic heart disease (RHD) and rheumatic fever (ARF) [Walker et al., 2014]. The genome of this bacterium is composed of one genomic DNA, comprising 1699 ORFs [McShan et al., 2008]. 85.3% of its DNA is used as a coding sequence. Similar to other lactic acid bacteria, *S. pyogenes* possesses a low G+C content genome with 38.57%, which in the ORFs is slightly increased to 39.18% [McShan et al., 2008]. Similar to *E. faecalis*, foreign DNA makes up a part of the genome of *S. pyogenes*. There are three different prophage regions in the genome of *S. pyogenes*, introducing 112 ORFs to the organism. Despite one susceptible virulence gene, prophage 1 plays no part in virulence. However, phage 2 and phage 3 contain several virulence genes, such as streptococcal pyrogenic exotoxin H (speH) (prophage 2), streptodornase (spd3), paratox genes and hyaluronidase (prophage 3) [McShan et al., 2008]. In addition to prophage regions, IS elements also exist in its genome which take part in the virulence of *S. pyogenes* as well. Interestingly, there are several CRISPR elements as spacers in the genome of *S. pyogenes*, which are most likely involved in the lytic cycle [Mojica et al., 2005]. Among other genomic regions coding for virulence factors, there are virulence-associated hyper variable regions encoding different virulence factors such as major antiphagocytic M protein (emm), vir regulon, streptococcal pilus or T-antigen: Cpa, FctA and PrtF2 [McShan et al., 2008]. Additionally, there are several unique gene clusters like Nudix hydrolase, which plays a role in host cell invasion, and NudA, nudB and nudC, which most likely take no part in the virulence of this pathogen [McShan et al., 2008].
From a metabolic standpoint, *S. pyogenes* is not as versatile as *E. faecalis*, although it can grow on several energy resources besides glucose, which is the primary energy resource. [Gera and

McIver, 2013], [Ferretti et al., 2016]. *S. pyogenes* owns an intact glycolysis comprising 12 enzymes, some of them such as GAPDH and enolase can be exported to the cell surface too [Ferretti et al., 2016]. Nevertheless, the mechanism of this export is not known [Ferretti et al., 2016]. Surface GAPDH is reported in several studies to take part in the virulence of *S. pyogenes* [Jin et al., 2011], *seidler2013*. *S. pyogenes* is capable of using various kinds of carbohydrates based on the content of the extracellular environment. For instance, during biofilm formation in saliva, it takes up maltodextrin, which is more abundant in saliva compared to glucose. However, when colonizing the throat (pharyngitis) or in the case of deep wound infection, rapidly metabolized mono- and di-saccharides are its favourite choice [Ferretti et al., 2016]. Moreover, several vitamins are required for the growth of *S. pyogenes* such as biotin, riboflavin and thiamine [Koser et al., 1968]. Likewise, amino acids are also very important for *S. pyogenes*. The growth rate of *S. pyogenes* is significantly influenced by the lack of amino acids, except for glutamine, glutamate, cysteine, proline, cystine, asparagine, alanine and aspartate [Levering et al., 2016]. However, the same study reported that *S. pyogenes* can adapt to the lack of glycine and serine in the environment [Levering et al., 2016].
To impose its virulence capability, *S. pyogenes* employs LPXTG-containing proteins to attach to human cells. LPXTG is known as the major cell-anchoring motif in many species of gram-positive bacteria [Terao, 2012]. Among LPXTG-containing proteins, a fibronectin (fn) binding protein has been characterised (fbaA) which is experimentally shown to attach to human fn [Terao, 2012]. On the other hand, it is shown that when spreading to deep human tissues, *S. pyogenes* disrupts the complement immune system in the human body [Terao, 2012]. To do so, it uses an extracellular cysteine protease, SpeB [Wexler and Cleary, 1985], [Chen and Cleary, 1989] to cleave C3b, an important component of the human complement immunity [Terao, 2012]. The same study also showed that SpeB increases the survival rate of *S. pyogenes* when growing on human blood [Terao, 2012]. Other studies also have shown that *S. pyogenes* employs fn binding proteins together with M protein and LTA as a device to attach to the host tissues [Todar, 2006].

1.2.3 multi-omics data to study lactic acid bacteria

multi-omics refers to multiple libraries of large scale data sets which are produced with the aim of studying a particular system as a whole. Following a top-down approach, omics data sets bring knowledge about as many as possible components of a system, on the contrary to the reductionist approach. Therefore, the use of omics libraries would enable researchers to study the overall phenotypic landscape of a biological system under certain circumstances. Multi-omics libraries are composed of different types of data. Most conventionally, genomics, transcriptomics and proteomics libraries are used to characterize the genotype as well as the phenotypes of different organisms at different levels. More recently, metabolomics and epigenomics are also used to answer the particular questions requiring those data [O'Donnell et al., 2020].
The technologies producing omics data have been widely used to study lactic acid bacteria [O'Donnell et al., 2020], [Huang et al., 2017]. The use of metabolomics and comparative

genomics studies in the case of LAB was reported as early as 2004, showing that these technologies had been used to improve our understanding of food microbes [Leroy and De Vuyst, 2004]. In recent years, these technologies have been used, singly or in combination, to study different characteristics of LAB. Lahtvee and colleagues used microarray transcriptomics together with Mass-spectrometry (MS) based proteomics to investigate the growth capabilities and amino acid metabolism in *Lactococcus lactis*, [Lahtvee et al., 2011]. In another study, authors used multi-omics libraries to have a better-designed platform to select probiotics, aiming at the attenuation of infectious diseases such as chytridiomycosis [Rebollar et al., 2016]. Furthermore, multi-omics data sets can be used to study microbial communities [O'Donnell et al., 2020], [Wang et al., 2019],[Sirén et al., 2019]. As an example, DNA and cDNA libraries (generated by NGS) together with Mass-spectrometry were used to identify and assess the microbial community as well as compounds related to the organoleptic properties in milk whey [Sattin et al., 2016]. In another study, 16S RNA profiling, shotgun metagenomics and metabolomics profiling were used to investigate metabolic interactions in the bifidobacterial community and their impact on gut-microbiota [Turroni et al., 2016].

Different types of omics and multi-omics libraries such as metabolic, transcriptome and proteome libraries have been generated to study LAB [O'Donnell et al., 2020]. A metabolic profile can reflect how an organism reacts to different environmental conditions as well as to explain the metabolic interaction in bacterial communities. This method has been widely used to study the metabolic behaviours of numerous species of LAB which are involved in the production of fermented foods, such as *Streptococcus* strains found in fermented soymilk [Mozzi et al., 2013]. On another level, transcritome data have been used to study different aspects of LAB, especially their virulence. Vebo and colleagues used transcriptome profiles of *E. faecalis* to assess the expression level of the genes involved in virulence, as well as other characteristics such as modulation of fatty acid composition in membrane, in a comparative study between the cultures of 2xYT with 10% blood and 100% blood [Vebø et al., 2009]. In another investigation, Solheim and colleagues reported that bovine bile impacts the expression level of drug resistance transporters, using a transcriptome data set from *E. faecalis* [Solheim et al., 2007]. At the proteome level, proteome profiles have been used to study the impact of tract-specific environments on the protein abundances of *E. faecalis*. For instance, the expression level of proteins involved in the secretome of *E. faecalis* when growing in urine was investigated and the effect of the environment on the expression of antigens and other components was observed [Arntzen et al., 2015]. Moreover, the proteome profile of *E. faecalis* cultured in bovine bile was used to study the impact of the environment on the metabolic programming of the bacterium, such as on the fatty acid and phospholipid biosynthesis and energy metabolism [Bøhle et al., 2010].

Omics data sets have been previously used here in the department of biological processes (Bioquant, Heidelberg University) to study *E. faecalis* and *S. pyogenes*. Veith and colleagues used metabolic profile of *E. faecalis* to study amino acid metabolism, in an integrative study using genome-scale modelling [Veith et al., 2015]. In a similar approach Levering and colleagues

investigated amino acid essentiality of *S. pyogenes* [Levering et al., 2016]. Furthermore, a workflow was developed by Großeholz and colleagues to integrate proteome data into genome-scale models, enabling the application of multi-omics data in integrative studies using genome-scale metabolic models [Großeholz et al., 2016]. This is the approach that I also used in the present study, which was previously published in an article too [Loghmani et al., 2021].

1.3 Aim of study

In this project, I studied the metabolic characteristics of two human pathogens, *E. faecalis* and *S. pyogenes*, using an integrative approach comprising experimentally acquired multi-omics data and mathematical modelling, with the aim of finding tract-specific drug targets in their metabolic networks. The tract-specific new drug targets were in particular focus due to the rising trend of multi-resistance to conventional antibiotics in these two species. As mentioned earlier in this chapter, metabolism has been one of the primary candidates for drug target identification in human pathogens. To acquire a comprehensive understanding of metabolic characteristics as well as adjustments, different experiments were designed to acquire omics data at different levels. First, I studied adaptive metabolism of a Δ*glnA* mutant of *E. faecalis* during a pH shift experiment, using metabolic and proteomic data, combined with mathematical modelling. The integrative analysis of adaptive metabolism enables us to understand the behaviour of the organism under different environmental conditions, which in this case was the extracellular pH. Considering the fact that this pathogen colonizes different tracts, it is subjected to a wide range of environmental conditions, many of which might have different pH values. Taking advantage of its highly versatile metabolism, *E. faecalis* is able to survive different types of stress and shock, and further grow and colonize different tissues. Moreover, studying the characteristics of the Δ*glnA* mutant and comparing it to those of the wildtype would provide a deep insight towards the nitrogen metabolism in *E. faecalis*, as glutamine is the main donor of the amino groups in the metabolic network.

Next, in order to discover the sensitivity/robustness of the results generated by genome-scale models, I developed a method based on random perturbation. Genome-scale modelling have been shown to be powerful a approach to simulate metabolic networks and the conclusions drawn based on its results have helped scientists to find answers to their questions of interest. However, a major drawback of this modelling formalism is a very large solution space as a consequence of having a high degree of variability in the system. The flux distribution profiles obtained from this large solution space are, therefore, highly variable, many of them are contradictory to actual biological phenotypes. The most powerful approach to shrink the solution space of genome-scale models is to integrate experimentally measured data into models. In this study, I developed a method to analyse the reliability of the results generated by genome-scale models. Using random perturbation to sample the solution space, the method allows to functionally analyse the solution space and to determine what fraction of the solutions are consistent with the actual biological phenotypes. The method is particularly useful when the flux

distribution of a set of internal reactions is of interest, which is the case when looking for drug targets in the metabolic networks of pathogens. Moreover, the method allows to discover the impact of different types of experimental data on shrinking the solution space and increasing the predictability of a model.
The final part of this study was developing tract-specific genome-scale models in the example of *E. faecalis* and *S. pyogenes* in order to find tract-specific drug targets in their metabolic network. Considering the fact that genome-scale models are often developed based on experimental data generated from the standard lab conditions, there is a relatively big gap between the insight that these models provide us with and the actual physiological phenotypes of a pathogen in human body. The need for tract-specific analysis of pathogens becomes even more significant, given the progressive trend of bacterial resistance to current antimicrobial agents. To bridge this gap, I collected different types of omics data and integrated into the genome-scale metabolic models of these two pathogens to simulate their metabolic behaviour under the conditions comparable to actual physiological conditions in human body. Using this approach, I generated a set of libraries containing potential drug targets under each condition. The computationally found drug targets are still to be validated experimentally. Finding tract-specific drug targets would be the first step in overcoming the drug resistance in these two pathogens and bringing the treatment strategies to the next level.

Chapter 2

Materials and Methods

2.1 Studying the impact of glutamine auxotrophy on the adaptive metabolism of *Enterococcus faecalis*

The method for this part of the thesis is obtained from my previous publication on this part [Loghmani et al., 2021].

2.1.1 Experimental

Bacterial strains and culture conditions

Enterococcus faecalis V583 Δ*glnA* [Veith et al., 2015] was grown in batch cultures at 37 °C in a chemically defined medium for lactic acid bacteria (CDM-LAB [Levering et al., 2016], pH 7.5 and 6.5). The CDM-LAB medium contained the following per liter: 1 g K2HPO4, 5 g KH2PO4, NaHCO3, 0.6 g ammonium citrate, 1 g acetate, 0.25 g tyrosine,0.24 g alanine, 0.5 g arginine, 0.42 g aspartic acid, 0.13 g cysteine, 0.5 g glutamic acid, 0.15 g histidine, 0.21 g isoleucine, 0.475 g leucine, 0.44 g lysine, 0.275 g phenylalanine, 0.675 g proline, 0.34 g serine, 0.225 g threonine, 0.05 g tryptophan, 0.325 g valine, 0.175 g glycine, 0.125 g methionine, 0.1 g asparagine, 0.2 g glutamine, 10 g glucose, 0.5 g L-ascorbic acid, 35 mg adenine sulfate, 27 mg guanine, 22 mg uracil, 50 mg cystine, 50 mg xanthine, 2.5 mg D-biotin, 1 mg vitamin B12, 1 mg riboflavin, 5 mg pyridoxamine-HCl, 10 mg p-aminobenzoicacid, 1 mg pantothenate, 5 mg inosine, 1 mg nicotinic acid, 5 mg orotic acid, 2 mg pyridoxine, 1 mg thiamine, 2.5 mg lipoic acid, 5 mg thymidine, 200 mg MgCl2, 50 mg CaCl2, 16 mg MnCl2, 3 mg FeCl3, 5 mg FeCl2, 5 mg ZnSO4, 2.5 mg CoSO4, 2.5 mg CuSO4, and 2.5 mg (NH4)6Mo7O24.

pH shift experiments in chemostat cultures

The pH shift experiments were carried out as previously described [Großeholz et al., 2016]. In short, *E. faecalis V583 DeltaglnA* was grown in glucose-limited chemostat cultures in Biostat B Plus benchtop bioreactors (Sartorius) in 750 ml CDM-LAB with a dilution rate of 0.15/h at 37 °C and gassing with 0.05 L/min nitrogen and stirring with 250 rpm. The pH was kept at the desired level by titrating with 2 M KOH. Initially, the pH was kept constant at 7.5 until

a steady-state was reached. Steady-state was assumed when no glucose was detectable in the culture supernatant anymore, and dry mass and optical density (600 nm) were constant on two consecutive days. For the pH shift, the pH control was switched off until the desired pH (6.5) value was reached. The cultivation was continued until the steady-state was reached again. Samples were taken at steady state pH 7.5 and at several time points during and after the pH shift as indicated in Figure 1. Per sampling point, samples for determination of dry mass, extracellular metabolites, and proteomic analysis were taken as previously described [Großeholz et al., 2016].

Chemostat cultures for determination of ATP maintenance

For determination of ATPmaintenance (ATPm), *E. faecalis V583 DeltaglnA* was grown in glucose-limited chemostats as described above (except for pH shift) at two different dilution rates, 0.15 h^{-1}, and 0.05 h^{-1}, with three biological replicates per dilution rate. At steady-state samples were taken and processed as described above.

Quantification of extracellular metabolites

For samples from pH shift experiments, quantification of amino acids in media and culture supernatants was done by Frank Gutjahr Chromotgraphie (Balingen, Germany); quantification of lactate, formate, acetate, glucose, acetoin, 2,3-butanediol, ascorbate, citrate, pyruvate, and ethanol were done by Metabolomics Discoveries GmbH (Potsdam, Germany). For quantification of amino acids, glucose, and fermentation products in CDM-LAB and culture supernatants of samples from ATPmaintanance experiments, the following two methods were used:
Method 1: an Agilent 1260 Infinity II HPLC system was used. The system was controlled by OpenLAB CDS Workstation software. For amino acids analysis, sample supernatants were filtered through a 0.22 μm syringe filter into an HPLC sample vial. Amino acids were derivatized, separated on a reversed-phase column (Agilent Poroshell 120 EC-C18 4.6x100mm, 2.7μm), detected with a diode array detector (DAD G7117A), and quantified following manufacturer's guidelines (AdvanceBio Amino Acid Analysis, © Agilent Technologies, Inc. 2018). Standards ranging from 5 μM to 30 mM were used to quantify aspartate, glutamate, asparagine, serine, glutamine, histidine, glycine, threonine, arginine, alanine, tyrosine, valine, methionine, tryptophan, phenylalanine, isoleucine, leucine, lysine, and proline.
For the analysis of organic compounds, samples were prepared as follows: 100 μl 35 % perchloric acid was added to 1 ml sample, mixed, and placed on melting ice for 10 minutes. Subsequently, 55 μl potassium hydroxide solution (7 M) was added, and the sample was centrifuged for 2 min at 20,000 g. The supernatant was filtered through a 0.22 μm syringe filter into an HPLC sample vial. Separation of sugars and fermentation products in the sample was performed by using an Agilent Hi-Plex H column (4.6x250 mm, 8 μm) with a working temperature of 65 °C using 10 mM H2SO4 as a mobile phase with a flow rate of 0.4 ml/min. For detection, a refraction index detector (RID) with a working temperature of 35 °C and a diode array detector

(DAD) with a wavelength of 210nm/4nm with a reference wavelength of 360nm/100nm were used. Standards ranging from 50 μM to 150 mM were used for the quantification of glucose, ethanol, citrate, lactate, pyruvate, formate, and acetate.
Method 2: Sugars and organic acids in the supernatant were measured with an isocratic Agilent 1200 series HPLC system equipped with a Phenomenex guard carbo-H column (4 by 3.0 mm) and a Rezex ROA organic acid H (8%) column (300 by 7.8 mm, 8 μm; Phenomenex) maintained at 50°C. Analytes were separated and detected using 5 mM H2SO4 with a constant flow rate of 0.4 mL min-1. Prior to analysis, samples were pretreated for precipitation of abundant phosphate by adding 4 M NH3 and 1.2 M MgSO4 solution, followed by incubation with 0.1 M H2SO4. Absolute concentrations were obtained by standard-based external calibration and normalization with L-rhamnose as internal standard.

Determination of protein abundances

All proteomics experiments and relevant downstream data analyses were done as part of a previous study as essentially described and published in [Großeholz et al., 2016], where the methods were detailed. The following sections briefly describe the proteome sample preparation and quantification of protein abundances using SWATH-MS.

Proteome sample preparation

Bacterial cell pellets were washed three times with PBS and kept frozen until experimentation began. These non-viable cell pellets were processed in two technical replicates using BarocyclerrNEP2320 (PressureBioSciences, Inc, South Easton, MA). Briefly, samples were lysed in buffer containing 8 M urea, 0.1 M ammonium bicarbonate, 10% trifluoroethanol, and completeTM protease inhibitor under pressure cycling (PCT) program (198 cycles, 20 seconds 45 kpsi, 10 seconds 0 kpsi) at 35 °C. Whole cell lysates were then sonicated for 30 seconds with 1-minute interval on ice for 3 times. Cellular debris was removed by centrifugation and sample protein concentration was determined by BCA assay prior to protein reduction with 10 mM TCEP for 25 min at 35 °C, and alkylation with 40 mM iodoacetamide in the dark for 30 min at room temperature. LysC digestion (1/50, w/w) was performed in 6 M urea under PCT program: 90 cycles, 25 seconds 22 kpsi, 10 seconds 0 kpsi at 35 °C; subsequent trypsin digestion (1/30, w/w) was performed at further diluted urea (1.6 M) under PCT program: 180 cycles, 25 seconds 22 kpsi, 10 seconds 0 kpsi 35 °C. Digestion was stopped by acidification with trifluoroacetic acid (TFA) to a final pH of approximately 2 before C18 column desalting using SEP-PAK C18 cartridges (Waters Corp., Milford, MA, USA).

Data acquisition and quantification of protein abundances using SWATH-MS

We used available, published SWATH MS Spectral and assay library generated by [Großeholz et al., 2016]. For SWATH-MS data acquisition, the same mass spectrometer and LC-MS/MS

setup was operated essentially as described before20, except that 64 windows of variable effective isolation widths were used (with an additional 1 Da overlap on the left side of the window), with a dwell time of 100 ms to cover the mass range from 400 - 1200 m/z in 3.3 s. The collision energy for each window was set using the collision energy of a 2+ ion centered in the middle of the window with a spread of 15 eV. The SWATH targeted data analysis was carried out using (OpenMS 1.12) analysis workflow (OpenSwathWorkflow21, http://www.openswath.org) running on an internal computing cluster and consists of the following steps. First, fragment-ion chromatograms were extracted for each peptide precursor in its appropriate SWATH-MS window based on the target and decoy assays in TraML format, with an extraction width of 0.05 Thomson (OpenSwath ChromatogramExtractor) and a retention time extraction window of $\pm$300 seconds around the expected retention time. Additionally, ion chromatograms for the iRT retention time standard peptides were extracted to facilitate projection of the assays from the normalized iRT retention time space into the retention time space for each individual run (OpenSwath RTNormalizer). Peak groups from the extracted fragment-ion chromatograms were formed and scored according to their elution profiles, similarity to the target assay in terms of retention time and relative fragment-ion intensity, as well as features from the full MS2 SWATH spectrum extracted at the chromatographic peak apex (OpenSwath Analyzer). Finally, the optimal separation between true and false peak groups was achieved using a linear discriminant model training with 60-fold semi-supervised learning iterations; and the score distribution from the shuffled decoy assays was used to estimate the false discovery rate using pyProphet (0.9.2) (https://pypi.python.org/pypi/pyprophet21) based on the mProphet algorithm22 and filtered using 1% FDR at the peptide feature level. Further, peak-groups were aligned among all 48 SWATH runs (24 wt and 24 GlnA mutant samples) using the OpenSwath feature aligner to ensure the consistent quantification of peak groups (peptide features) that could otherwise not be confidently identified above the FDR cut-off from a single run alone. Requantification option was also enabled to provide an upper bound for the intensity of target analyte where no peak-group passed the confidence filter so that the final data matrix did not contain any missing data point. Protein quantification was computed using R package, MSStats.daily 2.3.523. Briefly, we preprocessed the dataset from openSWATH extraction by log2 transformation and quantile normalization and generated the protein quantity matrix from the fragment ion level data using the 'groupComparison' and 'quantification' function of MSstats. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org/) via the PRIDE partner repository with the data set identifier PXD030778. Users can sign in via http://www.ebi.ac.uk/pride/archive/ to access the SWATH data with: Reviewer account details:
Username: reviewer_pxd030778@ebi.ac.uk Password: SfhNDiaz

2.1.2 Computational

Determination of non-growth associated ATPmaintenance

The determination of non-growth associated ATP (ATPm) was performed as described in [Teusink et al., 2006]. Thus, the measured flux value for the carbohydrates, organic acids, and amino acids were integrated in the genome-scale model as constraints. The biomass reaction was fixed at the respective growth rate (dilution rate), and the flux of the ATPm reaction was maximized as the objective function. The obtained values were used to fit a linear function, for which the y-intercept determines the required energy for the organism at a zero growth rate. This value is then applied to the model as the lower bound of the ATPm reaction.

Software, model, and computational methods

PySCeS-CBMPy[Olivier et al., 2005] was used for constrained-based modelling. The genome-scale metabolic model of E. faecalis [Veith et al., 2015] was used for all the computational analyses (BioModels: MODEL1510010000). To optimize the growth rate, ATPm, and obtaining flux distribution profiles, flux balance analysis (FBA)11 was used. This method calculates a vector of flux values while optimizing the user-defined objective function. To calculate the feasible range of each reaction that results in the optimized value of the objective function, flux variability analysis (FVA) [Mahadevan and Schilling, 2003] was applied. FVA uses the optimized value of the objective function as an additional constraint and subsequently calculates the maximum and minimum boundary of the feasible interval for each reaction.

Integration of constraints to the genome-scale model

The integration of constraints to the genome-scale model was done as indicated in [Großeholz et al., 2016]. To integrate the metabolic data, a tolerance level of 40% was applied to the measured flux rates to account for measurement errors. The obtained values were applied to the upper (+ 20%) and lower (-20%) bounds of the respective exchange reactions at both conditions. Regarding the proteome data, reactions with no experimental evidence (after the comparison to the list of essential genes and reactions at the proteome level at pH 7.5 were deactivated. For the model to have a feasible solution, several genes had to be reactivated (EF0387, EF3069, EF0108, EF0929, EF0547, EF2442, EF3015, EF3199, EF3200, EF3117). To represent the significant fold changes of proteins in response to pH shift, the log2 fold-changes of protein abundances were multiplied by 40% (tolerance level) and then applied to the maximum and minimum value of respective reactions, obtained by flux variability analysis (FVA)26 at pH 7.5.

2.2 Inspecting the solution space of genome-scale metabolic models

The method for this part is obtained from my previous publication on 'Inspecting the solution space of genome-scale metabolic models' [Loghmani et al., 2022].

Models, experimental data and constraints integration

The genome-scale metabolic models of *E. faecalis* (wildtype) [Veith et al., 2015] and of a knock-out mutant of glutamine synthase (*DeltaglnA*) of *Enterococcus faecalis* [Veith et al., 2015] were used for the initial analysis. The experimental data was obtained from [Großeholz et al., 2016] for the wildtype and [Loghmani et al., 2021] for the mutant. The findings reported here were validated using the genome-scale models of *Streptococcus pyogenes* [Levering et al., 2016] and *Lactococcus lactis* [Flahaut et al., 2013] together with the respective published experimental data. The models with no constraints (denoted by nc throughout the text) had all the upper bounds set to +1000 (representing +infinity) and lower bounds to -1000 (representing -infinity) except for the biomass reaction that was fixed at the intended objective function value. Constraints were integrated at three different steps using the respective experimental data, medium composition, metabolite uptake/secretion rates (organic acids and amino acids) and proteome data. To integrate the medium composition data, the upper and lower boundaries of the respective reactions were adjusted. To integrate the metabolite uptake and secretion rates, the experimentally measured flux value of metabolites was used as the basis and a tolerance value of 40% was considered to account for measurement error. Consequently, 20% were added to the measured value and used as the upper bound and the subtraction of 20% was used as the lower bound [Großeholz et al., 2016]. To integrate proteome data, the flux value of the reactions whose respective proteins were not detected by the proteomics experiment were set to zero [Großeholz et al., 2016]. In summary, the E. faecalis model contains 709 reactions (in the case of the wildtype, 708 in the case of the mutant), 644 metabolites and 688 genes. The model of S. pyogenes contains 576 reactions, 558 metabolites and 481 genes. The model of L. lactis contains 754 reactions, 650 metabolites and 516 genes.

Perturbation procedure

The perturbation procedure was used to determine the effect of alternative optimal values taken by each reaction on the flux distribution profile of the network. Given the fact that parameters within the permissible flux interval calculated by FVA can result in the optimal value of the objective function, we wanted to find out how each point in the permissible interval affects the quantitative and qualitative flux distribution in the network. The steady-state flux distribution of a constraint-based model can be obtained by:

$$S.v = 0$$

where S represents the stoichiometric matrix and v is the vector of flux distribution. Here, the maximum and minimum values of individual reactions can be obtained using FVA [Mahadevan and Schilling, 2003] by maximizing the objective function and using the respective value as an additional constraint:

$$Max f^T v,$$

$$s.t.S.v = 0,$$

$$v_{min} \leq v \leq v_{max}$$

where f is the objective function vector and v_{max} and v_{min} are the vectors of maximum and minimum allowable flux values, respectively, for each reaction. Using this characteristic of constraint-based models, the perturbation procedure we proposed is based on the idea that a change in the flux value of a reaction would result in a different combination of fluxes in the network, as shown in the Figure, in the example of two flux combinations:

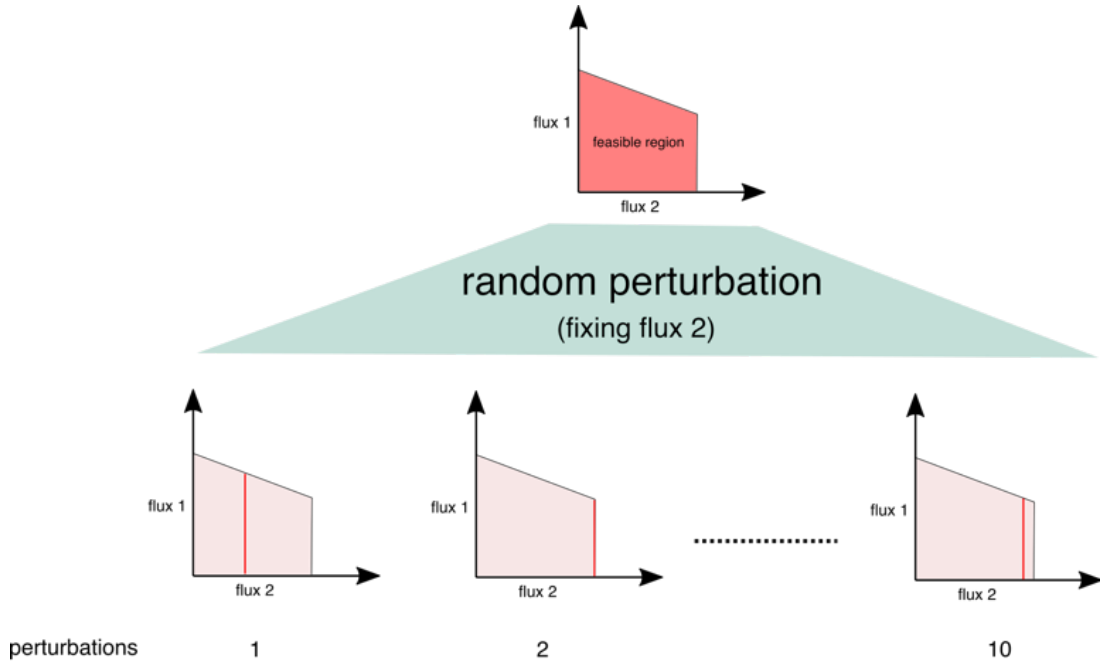

Figure 2.1: Perturbation procedure to determine the robustness of FBA/FVA outcome. The figure shows how fixing one reaction at various random values results in a different range for flux combinations between two fluxes.

Consequently, the robustness of model predictions, whether they are qualitative biological phenotypes, or flux values/range for a several reactions can be examined and the overall predictability of the model can be determined.

All analyses were done using the FBA perturbation toolbox which is available on Github (https://github.com/babakml/FBA$_p$*erturbation.git*), *developed for this project. The toolbox was primarily devel based modelling on MATLAB R2018a* [*MATLAB*, 2018] *on MacOS Mojave version* 10.14.6 (*Apple.com*). *To perfo 6 was applied to consider a reaction as variable. Next, for each variable reaction*, 10 *random values within the deter*

Table 2.1: Bacterial strains used for the tract-specific drug target identification

Species	strain	Reference
Enterococcus faecalis	V583	[Paulsen et al., 2003]
Streptococcus pyogenes M49	NZ131	[McShan et al., 2008]

*CBMPy*0.8.0[*OlivierandGottstein*, 2021]*onPython*2.7[*VanRossumFredLandJr*, 1995]*andthefluxdistributior*
of the original flux value. In the cases where the original flux value was zero, the threshold was set to 10-6.

Analysing the solution space using CoPE-FBA

A standard analysis of the solution space was performed using CoPE-FBA with CPLEX as solver [Kelk et al., 2012]. For this purpose, FVA was performed and variable reactions were determined with the interval size of 10-6 as a threshold. Afterwards, flux modules, which are the sets of variable reactions that are linearly independent, were determined. The modules were then used to analyse the solution space.

2.3 Integrative tract-specific drug target identification in *E. faecalis* and *S. pyogenes*

Bacterial strains

Standard cultivation of bacteria

For standard cultivation, *E. faecalis* and *S. pyogenes* were grown in THY medium or on Columbia agar plates containing 5% sheep erythrocytes at 37 °C in 5% CO2 enriched atmosphere. Cultures on agar plates were kept in the refrigerator and were used to inoculate fresh liquid cultures for maximum 4 weeks.

Batch growth experiments

For batch growth experiments, *E. faecalis* and *S. pyogenes* were grown in overnight cultures in THY medium, with the culture volume of ten percent of the main culture for CDM-LAB, artificial saliva and artificial urine, and 15% of the main culture for the plasma. The overnight cultures were spun down by centrifugation at 3345 x g for 10 min. The supernatants were discarded and the pellets were washed twice with PBS. Afterwards, bacteria were suspended in 1 ml PBS and added to the medium.
For CDM-LAB, artificial saliva, and artificial urine, a culture volume of 500 ml was used. Cultures were incubated at 37 °C. After two hours, a 15 ml sample was collected and bacteria were spun down at 3345 x g for 10 minutes. The supernatant were collected and stored at -20 °C for the metabolic analysis. At the second time point, two hours after inoculation, a 400 ml

sample was taken and spun down at 12000 x g for 5 min. A sample of 15 ml supernatant was taken and stored in -20 °C for analysis of extracellular metabolites. The bacterial pellets were suspended in PBS and transferred to 15 ml tubes for centrifugation at 3345 x g for 10 min. Then, the supernatant was discarded and pellets were suspended in 2 ml PBS and divided into 1.5 ml tubes, each containing 1 ml. The samples were spun down again with a centrifuge at 3099 x g for 10 min. The supernatant was discarded and the pellets were stored at -20 °C for transcriptomic and proteomic analysis.

Plasma cultures were prepared in a total volume of 100 ml. The first sample for metabolic analysis was taken at four hours after the start of the incubation, the OD was measured and supernatant was collected in the same way as in the other media. For the second time point, five hours after the start of the incubation, the OD was measured and the remaining 85 ml of the cultures were spun down using centrifuge at 8000 rpm for five minutes, 15 ml of supernatants were collected for metabolic analysis and stored at -20 °C. Pellets for transcriptomic and proteomic analysis were collected in the same way as in the other media.

Measurement of organic acids and sugars via HPLC Agilent InfinityLab LC Series 1260 Infinity II Quaternary System

The method for this part is provided by Eric Zitzow, who performed the measurement of the organic acids in Rostock medical university.

Detectable substances: glucose, pyruvate, acetate, citrate, (ethanol,) formate, lactate

Mobile phase: 10 mM H2SO4 (555,7 μl conc. sulfuric acid (96%) in 1 l water)

Column: -Hi-Plex H 4.6 x 250 mm, 8 μm (column storage condition: water) - 65 °C (max. temperature for the column: 70°C; pH range: 0-7; separation mode ion exchange (IEX); H-phase)

Flow rate: 0.4 ml/min (max. flow rate: 0.4 ml/min(!!!)) Duration: 25 min Injection volume: 20 μl

DAD: 210 nm/4nm; reference: 360nm/100nm (main detector is RID) RID: 35 °C

Sample preparation: -taking of 1 ml sample (e.g. culture supernatant) -addition of 100 μl 35% PCA (perchloric acid) solution -mix and place on melting ice for 10 minutes -addition of 55 μl 7 M KOH solution -mix shortly and centrifuge at maximum speed for 2 minutes -filter supernatant through a 0.22 μm syringe filter into an HPLC sample vial

Source:

-HPLC Analysis of Sugars and Glycoproteins in Biological Systems

https://www.agilent.com/cs/library/applications/SI-01410.pdf

-Agilent Hi-Plex Columns for Carbohydrates, Alcohols, and Acids

https://www.agilent.com/cs/library/applications/5990–8264EN.pdf

Table 2.2: Gradient program

Time (min)	% B
0	2
0.35	2
13.4	57
13.5	100
15.7	100
15.8	2
18	end

Amino acid measurement via HPLC Agilent InfinityLab LC Series 1260 Infinity II Quaternary System

The method for this part is provided by Eric Zitzow, who performed the measurement of the amino acids in Rostock medical university.
Detectable substances: alanine, arginine, asparagine, aspartate, cysteine, glutamate, glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, serine, threonine, tryptophan, tyrosine, valine, (hydroxyproline, sarcosine, proline,norvaline)
Mobile phase A: 10 mM Na2HPO4, 10 mM Na2B4O7 pH 8.2 To prepare 1 L, weigh out 1.4 g anhydrous Na2HPO4 and 3.8 g Na2B4O7 10H2O in 1 L water. Adjust to approximately pH 8.4 with 1.2 mL concentrated HCl, then add small drops of acid and adjust to a final pH of 8.2. Allow stirring time for complete dissolution of borate crystals before adjusting pH. Filter through 0.45 μm (or 0.22 μm) regenerated cellulose membranes (p/n3150-0576).
Mobile phase B: Acetonitrile:methanol:water (45:45:10, v:v:v) (Mobile phase A is consumed twice as fast as mobile phase B) (A: ca. 1,123 ml/min; B: ca. 0,377 ml/min per run)-¿ Volume: 1/3 A= B
Derivatization reagents: Injection diluent: 10 ml mobile phase A with 40 μl conc. H3PO4 OPA: 100 μl aliquots (vial with insert), potent for 7 to 10 days FMOC: 100 μl aliquots (vial with insert), potent for 7 to 10 days Borate buffer: change every 3 days
Column: Agilent Poroshell 120 EC-C18 4.6x100mm, 2.7μm 40 °C (max. temperature for the column: 60 °C; pH range: 2-8; separation mode: Reversed phase, EC-C18 phase), Column storage condition: 60% ACN, 40% H2O
Derivatization: 1. Draw 2.5 μL from borate vial (p/n5061-3339) 2. Draw 1.0 μL from sample vial 3. Mix 3.5 μL in wash port five times 4. Wait 0.2 minutes 5. Draw 0.5 μL from OPA vial (p/n 5061-3335) 6. Mix 4.0 μL in wash port 10 times default speed 7. Draw 0.4 μL from FMOC vial (p/n 5061-3337) 8. Mix 4.4 μL in wash port 10 times default speed 9. Draw 32 μL from injection diluent vial 10. Mix 20 μL in wash port eight times 11. Inject 12. Wait 0.1 minutes 13. Valve bypass
Flow rate: 1.5 ml/min for columns with 4.6 mm inner diameter (like our column) 0.62 ml/min for columns with 3 mm inner diameter
Injection volume: 1 μl

DAD: Signal A: 338 nm, 10 nm bandwith; reference: 390 nm, 20 nm bandwith Signal B: 262 nm, 16 nm bandwith; reference: 324 nm, 8 nm bandwith
Duration: 20 min Sample preparation: -taking of 1 ml sample (e.g. culture supernatant) -addition of 100 μl 35 %PCA (perchloric acid) solution -mix and place on melting ice for 10 minutes -addition of 55 μl 7 M KOH solution -mix shortly and centrifuge at maximum speed for 2 minutes -filter supernatant through a 0.22 μm syringe filter into an HPLC sample vial

Calculation of the growth rate and flux values

The growth rate of the bacteria were calculated using the following formula for the exponential growth [Seviour and Nielsen, 2010] using the measured OD values:

$$lnX = lnXo + \mu(t - to)$$

where X is a straight line showing the growth rate, Xo is the initial OD, μ is the specific growth rate constant (h^{-1}), t is the time point and t_0 is the initial time point.
The measured flux value of metabolites were calculated according to the exponential growth rate by fitting the data to the following linear equation:

$$S = So - (q/\mu) * (X - Xo)$$

where S is the substrate concentration, S_0 is the substrate value at initial time point. The method for the calculation of the growth rate and substrate uptake/production rate was obtained from [**?**].

The measurement of the protein abundances

The method for this part is provided by Dr. Sergo Kasvandik who performed the measurement of the protein abundances in Proteomics Core Facility, Institute of Technology, University of Tartu, Tartu, Estonia.
Sample preparation:

Cell pellets were suspended in 10 volumes of 6 M guanidine-HCl, 100 mM Tris-HCl pH 8.0, 20 mM DTT buffer. Samples were heated at 95°C for 10 min, followed by sonication in a Bioruptor (Diagenode) sonicator at 4°C. Samples were further homogenized with FastPrep24 (MP Biomedicals) bead beating device two times at 4 m/s for 30 s with cooling between cycles. After removal of beads, the samples were cleared with centrifugation at 17 000 g for 10 min at 4°C. Cell proteins were precipitated at 4oC by adding 1/4 volume of 0.4% sodium-deoxycholate, 20% trichloroacetic acid. Protein concentrations were measured with the Micro-BCA Assay (Thermo Fisher Scientific). Pellets were then suspended in 30 μl of 7 M urea / 2 M thiourea, 100 mM ABC, 20 mM methylamine solution, followed by disulfide reduction and cysteine alkylation with 5 mM DTT and 10 mM chloroacetamide for 30 min each at room

Table 2.3

Fraction	Elution buffer
BRP1	200 mM AF* pH 10
BRP2	200 mM AF pH 10, 10% ACN
BRP3	200 mM AF pH 10, 15% ACN
BRP4	200 mM AF pH 10, 20% ACN
BRP5	200 mM AF pH 10, 25% ACN
BRP6	200 mM AF pH 10, 50%* ACN

temperature. Proteins were predigested with 1:50 (enzyme to protein) Lys-C (Wako Chemicals) for 4 h, diluted 5 times with 100 mM ABC and further digested with trypsin (Sigma Aldrich) overnight at room temperature. *Enterococcus faecalis* digests were desalted using C18 StageTips and resuspended in 0.5% trifluoroacetic acid (TFA) ready for LC/MS/MS injection. *Streptococcus pyogenes* digests were further fractionated into six fractions using basic reversed phase chromatography with in-house made C18 StageTips using following elution buffers:

*AF – ammonium formatted, ACN – acetonitrile

Nano-LC/MS/MS measurement:

Samples were injected to an Ultimate 3000 RSLCnano system (Dionex) using a C18 trap-column (Dionex) and an in-house packed (3 μm C18 particles, Dr Maisch) analytical 50 cm x 75 μm ID emitter-column (New Objective). Peptides were eluted at 250 nl/min with a 8-40% B 240 min gradient (buffer B: 80% acetonitrile + 0.1% formic acid, buffer A: 0.1% formic acid) to a Q Exactive Plus (Thermo Fisher Scientific) mass spectrometer (MS) using a nano-electrospray source (spray voltage of 2.5 kV). The MS was operated with a top-10 data-dependent acquisition strategy. Briefly, one 350-1400 m/z MS scan at a resolution setting of R=70 000 at 200 m/z was followed by higher-energy collisional dissociation fragmentation (normalized collision energy of 26) of 10 most intense ions (z: +2 to +6) at R=17 500. MS and MS/MS ion target values were 3e6 and 5e4 with 50 ms injection times. Dynamic exclusion was limited to 60 s.

Raw Data Processing:

Mass spectrometric raw files were analyzed the with the MaxQuant software package (version 1.6.15.0). Methionine oxidation, glutamine/asparagine deamidation and protein N-terminal acetylation were set as variable modifications, while cysteine carbamidomethylation was defined as a fixed modification. Search was performed against UniProt (www.uniprot.org) Enterococcus faecalis and Streptococcus pyogenes reference proteome database using the tryptic digestion rule (including cleavages after proline). Only identifications with minimally 1 peptide 7 amino acids long were accepted and transfer of identifications between runs was enabled. Label-free quantification normalization with MaxQuant LFQ algorithm was enabled with default settings. Peptide-spectrum match and protein false discovery rate (FDR) was kept below 1% using a target-decoy approach. All other parameters were default.

Calculation of the significant fold changes in protein abundances

To calculate the significant protein abundances between the different conditions, I used the DEP package (Differential Enrichment analysis of Proteomics data) [Zhang et al., 2018] in R [R Core Team, 2021]. For the statistical analysis, I filtered out the proteins with missing values in the replicates of the same condition. Afterwards, I corrected the data for background and normalized using variance stabilizing transformation method [Huber et al., 2002]. The data was then imputed using random draws from a Gaussian distribution centered around a minimal value. The imputed data were used for the differential enrichment analysis using linear models and empirical Bayes statistics, provided by limma [Ritchie et al., 2015]. The p-values were adjusted for multiple testing by Benjamini-Hochberg (BH) multiple testing correction. An adjusted p-value of 0.05% was considered as a threshold for significance. A more elaborate description on data analysis can be found in the results chapter.

Integration of experimental data into genome-scale metabolic models

To integrate the metabolic data, I used the method published in [Großeholz et al., 2016]. Therefore, a 40% measurement error was applied to the measured values to calculate the flux boundaries. A +20% measurement error was applied to the flux values to calculate the upper bound, while a -20% was applied to the flux values to calculate the lower bound. For the integration of the transcriptome data, I used the published in [Großeholz et al., 2016] for the integration of the proteome data.The calculated significant fold changes were applied to the flux intervals obtained by FVA for the control conditions. An error rate of 40%, and when necessary, 60% was applied to the calculated values for the model to have a feasible solution.

Identification of the drug targets

In order to find the drug targets in a model. I used three different strategies. In the first method, I inactivated all the reactions in a model and those whose inhibition reduced the flux value of the bacteria to 30%, 20% and 10% were regarded as potential drug targets. In the second approach, I decreased the flux interval of all the reactions to 30%, 20% and 10% of the original value, and the reactions which a reduction in their flux values resulted in zero growth of the bacteria were considered as potential drug targets. In the third approach, I combine the two previous methods, meaning that I reduced the flux intervals of all reactions to 30%, 20% and 10%, and the reactions which a reduction in their flux reduced the growth rate of the bacteria to 30% or lower were considered as potential drug targets. A more detailed description of the integrative drug target identification can be found in the results chapter.

Chapter 3

RESULTS

In this chapter, I present the result of my thesis. The aim of this study was to discover tract-specific drug targets in the metabolic network of two pathogenic bacteria, *E. faecalis* and *S. pyogenes*, using an integrative approach that combines multi-omics profiles with genome-scale modelling. To do so, I employed numerous experimental as well as computational approaches, not only to uncover the phenotypes of these two pathogens under certain circumstances, but also to develop and improve the required computational tools. First, I studied the role of glutamine auxotrophy in *E. faecalis* by analyzing the metabolic and proteomic profiles of a Δ*glnA* mutant of *E. faecalis*. I also integrated the two profiles into the genome-scale model of the Δ*glnA* mutant to discover the role of glutamine auxotrophy on the metabolic behaviour of this organism in large scale by comparing the results to those of the wildtype, and also to predict certain metabolic adjustments by means of the genome-scale model. Considering the fact that glutamine is the main nitrogen donor in *E. faecalis* metabolism, studying the role glutamine auxotrophy brings a profound knowledge about the characteristics of not only the nitrogen metabolism, but also the carbohydrate metabolism and protein expression profiles. Next, I developed a method that enables us to study the sensitivity of genome-scale models by means of random perturbation. The method allows to find out how reliable the flux distributions in different parts of the network are, and how this reliability can be improved by integrating different types of experimental data. Being the major draw back of genome-scale models, internal flux distributions are often highly variable, containing a very large number of solutions that many of them are contradictory to biological reality. Therefore, a key step to generate highly reliable predictions with genome-scale models is to determine which set of results are relatively more reliable compared to the others. Having said that, when trying to find drug targets in a metabolic network, this method allows to compare the sensitivity of the potential drug targets. As a result, the potential drug targets can be presented with their degree of confidence. Last, I tried to find tract-specific drug targets in the metbolic network of *E. faecalis* and *S. pyogenes* by integrating the multi-omics profiles to the respective genome-scale models. To do so, I performed batch culture experiments in Rostock Medical University and collected samples for metabolic, transcriptome and proteome measurments. These data were then integrated into the genome-scale models and tract-specific drug targets were determined.

3.1 The effect glutamine auxotrophy on metabolic characteristics of *E. faecalis*

As mentioned above, I discovered the role of glutamine auxotrophy on the metabolism of the Δ*glnA* mutant of *E. faecalis* by analysing its metabolic and proteome profiles, integrated into the respective genome-scale metabolic model. As glutamine is the main provider of amino groups in metabolism, it has a comprehensive impact on the growth and phenotype of organisms in general. Therefore, studying the effect of glutamine auxotrophy would provide important information of metabolic adjustments and performance of *E. faecalis*. The results of this section is previously published in [Loghmani et al., 2021].
To find out the effect of glutamine auxotrophy on metabolic characteristics of *E. faecalis*, the gluthamine synthetase Δ*glnA* mutant of *E. faecalis* was subjected to pH shift in the course of 21 hours (Figure 3.1). The time course comprised two steady-states, at pH 7.5 and pH 6.5. Samples were taken at 8 different time points, time point 1 (t1) was the steady-state at pH 7.5, t2 the transition state, and t3, t4, t5, t6 and t7 were at pH 6.5, 80, 100, 120 and 240 minutes after the start of the pH shift. T8 was the steady-state at pH 6.5, 21 hours after the start of the pH shift. Samples were then used to calculate the exchange fluxes of organic acids, carbohydrates and amino acids, and to quantify the protein concentration at each time point.

3.1.1 Effect of pH on the growth rate

To find out about the effect of a pH shift on the growth rate of Δ*glnA E. faecalis*, the biomass was measured at different time points and were compared to those of the wildtype to discover the effect of glutamine auxotrophy. As shown in fig 3.2, the biomass of the Δ*glnA E. faecalis* decreased from 1.54 to 1.15 g/l, showing that the more acidic condition reduces the growth capabilities of the organism. Compared to the wildtype, the biomass production of the mutant is lower at all time points, pointing out to the fact that glutamine auxptrophy negatively affects growth of the bacterium. However, the trend of relative biomass reduction in response to pH shift is similar (approximately 25%) between the wildtype and the mutant.

3.1.2 Effect of the pH shift on protein expression profile of Δ*glnA E. faecalis*

To uncover the effect of the pH shift on the protein expression profile of Δ*glnA E. faecalis*, the significant fold change of all the detected proteins compared to t1 were calculated fig 3.3. The complete list of all the significant fold changes can be found in the supplementary data. Overall, 1681 open reading frames (ORF) were detected by SWATH-MS. While no significant fold change was detected at t2, t3 contained the highest number of significant fold changes with more than 40 proteins being affected by the pH shift. I performed a pathway analysis and the result showed that 11 of these proteins were involved in membrane and cell wall production and

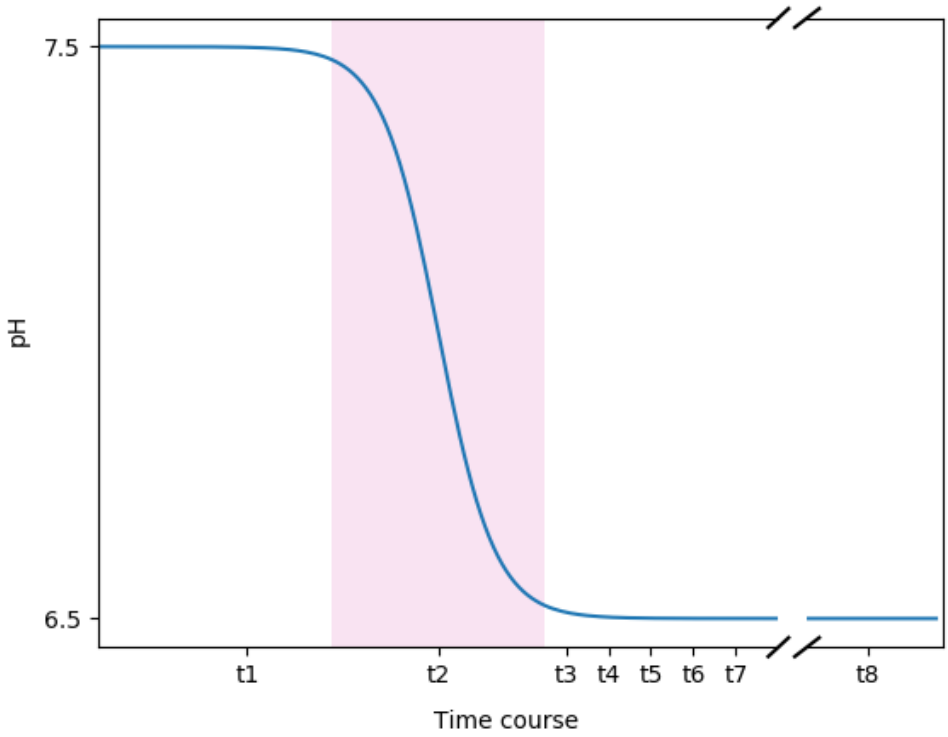

Figure 3.1: The pH shift experiment over the course of 21 hours. The experiment started at steady-state at pH 7.5 (t1), followed by a transition state (t2, the pink area) to pH 6.5. Samples were taken at pH 6.5 at t3-t7, as well as t8, the steady-state at pH 6.5. The figure is adapted from [Loghmani et al., 2021]

two proteins were involved in peptidoglycan biosynthesis. This finding suggests that the pH shift has a prompt effect on the restructuring and reshaping of the cell membrane and envelope. In contrast, at t3, there were only a few proteins being affected by the pH shift from t4 to t7 and all were down-regulated. The number of significant fold changes increased to 40 at t8, 21 hours after the pH shift. A major part of these proteins were down-regulated. Among all, it involved several proteins in the nucleotide metabolism, all down-regulated. Given the fact that de novo biosynthesis of nucleotides is energy consuming, the downregulation of nucleotide metabolism is consistent with the lower growth rate and higher energy demand in a more acidic environment.

I also compared the protein expression profile of the mutant to that of the wildtype in order to investigate the effect of glutamine auxotrophy on protein expression in *E. faecalis* in response to pH shift (Figure 3.3). The number of significant changes at the protein level in the wildtype was higher than in the mutant, except for t8. At t3, a large number of significantly affected proteins in both genotypes were involved in membrane and cell envelope biosynthesis, pointing out to the fact that the restructuring of cell membrane and envelope happens early in response to pH shift [Großeholz et al., 2016]. Unlike the mutant, several glycolytic enzymes were affected by the pH shift in the wildtype. Considering the fact that the growth and survival of *E. faecalis* at

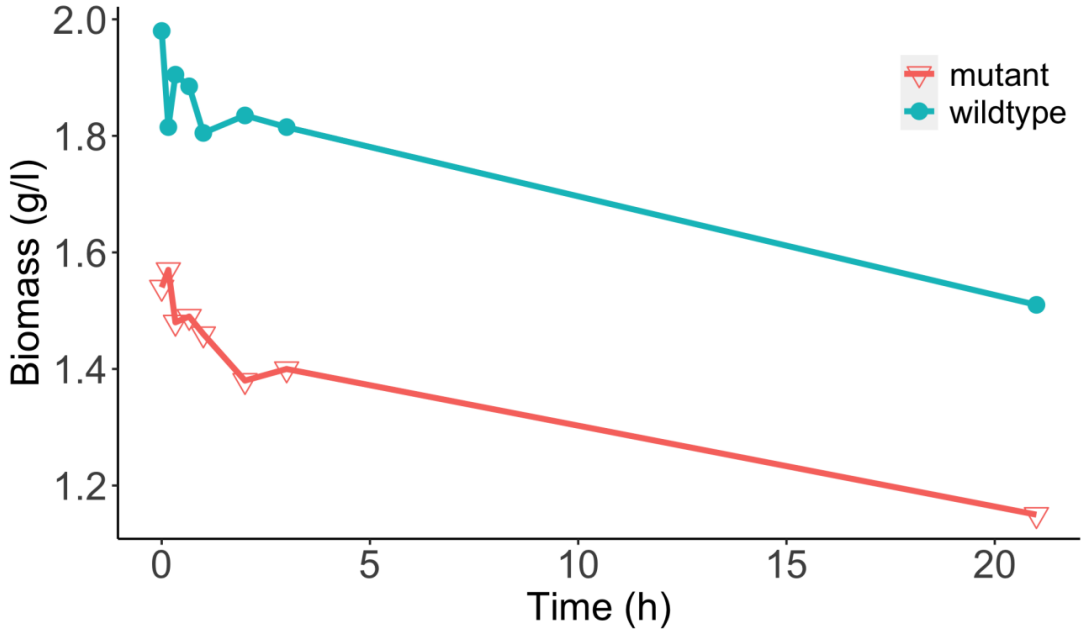

Figure 3.2: The biomass production of the Δ*glnA* mutant and the wildtype *E. faecalis* in response to pH shift over the course of 21 hours. The biomass value of the mutant is lower than of the wildtype at all during the whole time course. The figure is adapted from [Loghmani et al., 2021].

pH 6.5 is more energy demanding (compared to pH 7.5), due to the metabolic adjustments such as maintenance of internal pH homoeostasis; as well as the central role of glycolysis in energy metabolism, the increase in the expression of glycolytic enzymes is predictable. However, the lack of upregulation in glycolytic enzymes in the mutant was surprising, suggesting that this genotype had already been introduced to high energy demand as a consequence of glutamine auxotrophy. On the other hand, both wildtype and the mutant showed a similar pattern of down-regulation of genes involved in nucleotide metabolism.

3.1.3 Integrative computational metabolic analysis of Δ*glnA E. faecalis*

To get a comprehensive understanding of metabolic adjustments in Δ*glnA E. faecalis* in response to pH shift, I used a previously published genome-scale metabolic model of *E. faecalis* [Veith et al., 2015] and integrated the metabolic and proteome data into the model as constraints. The metabolic and proteomic constraints were integrated into the model using a previously developed method that translates flux values and significant fold changes into the flux boundaries of the respective reactions [Großeholz et al., 2016]. I used the integrative analysis to contextualize the quantitative metabolic and proteomic data within the framework of whole-system computational metabolic analysis.

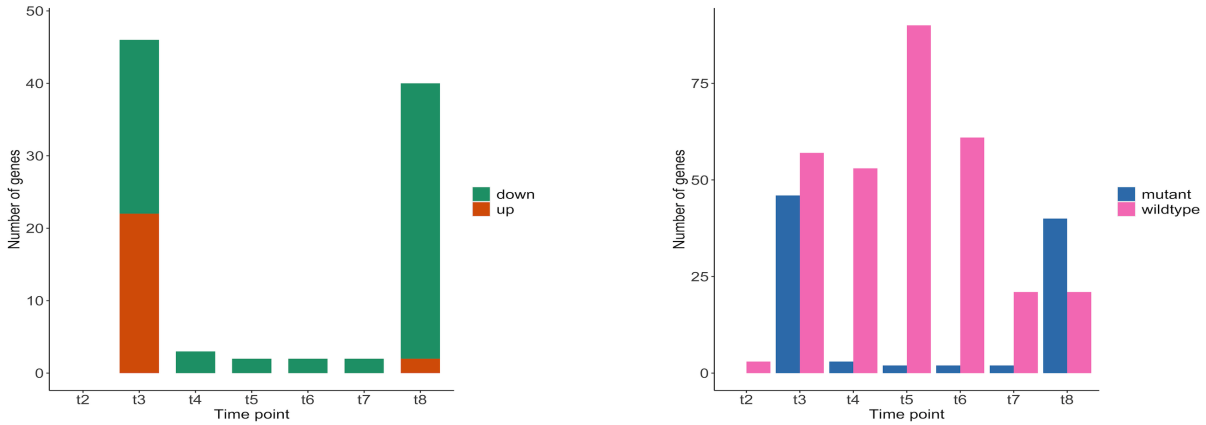

Figure 3.3: Left panel: the number of significant fold changes at the protein level in Δ*glnA E. faecalis* at different time points. Right panel: the number of significant fold changes at the protein level in the wildtype (pink) and the Δ*glnA* mutant of *E. faecalis* at different time points. The figures are adapted from [Loghmani et al., 2021]

Integrative determination of ATP maintenance

To improve the accuracy and predictability of the genome-scale model, I calculated the ATPM value using experimental data integrated with the genome-scale model of Δ*glnA E. faecalis*. ATPM is the amount of energy required for the organism to sustain life at zero growth rate. This is a particularly important parameter in genome-scale modelling and helps to improve the predictability of FBA results. To estimate the ATPM value using an integrative approach, comprising experimental data and genome-scale modelling, I performed chemostat experiments under the supervision of PD Dr. Tomas Fiedler at Rostock medical university. I grew Δ*glnA E. faecalis* in chemostat cultures at two different dilution rates, 0.15 and 0.05 h^{-1} at pH 7.5 and 6.5. I let the cultures continue until reaching the steady-state and took samples for metabolic analysis. I spun down the samples and supernatant and took the supernatant to analyse extracelullar concentration of glucose, organic acids and amino acids. Using the metabolites concentration, I calculated the flux values (uptake/production rate) for each individual metabolites and used the values as the flux boundaries of the exchange reactions for respective metabolites. After integrating the constraints, I fixed the biomass production rate at the growth rate (equal to the dilution rate) and set the objective function to ATPM reaction to obtain the maximum value. I performed this process for both dilution rates and used the ATPM flux values to fit a linear function with respect to growth rate. In this sense, if we plot the linear line as a function of growth rate (x-axis), the *y*-intercept would be the ATPM value at zero growth rate. This value is considered as the minimum amount of energy that is required only for survival and does not contribute to growth. The ATPM value is used as the lower boundary of the ATPM reaction to make sure that the minimum required amount of energy is produced by the system, while maximising the growth rate.

Using the above mentioned procedure, the estimated ATPM value for Δ*glnA E. faecalis* was

5.977 and 6.224 mmol/$g^{-1}_{DW}h^{-1}$ at pH 7.5 and 6.5, respectively. However, the integration of these values into the genome-scale model resulted in an overproduction of biomass. In fact, the excretion rate of fermentation products is at such a high value that the produced ATP would excessively boost the biomass production, while the ATPM value is at this low level. Therefore, the lower bound of the ATPM reaction has to be set on a higher value at both pH to make sure that the biomass production is inline with the experimental data. There are several factors allowing such an adjustment without violating modelling rules. For one, the ATPM value is an estimated value being the outcome of integrating around 30 metabolite's measurements into the genome-scale model. This makes the optimization process very error prone, since the accumulation of all the measurement errors would affect the result. As an example, in the data set used for the estimation of the ATPM value, the glucose uptake rate at pH 7.5 at the dilution rate of 0.15 h^{-1} was 6.32 mmol/$g^{-1}_{DW}h^{-1}$, while in the data set from the pH shift experiment it was 7.17 mmol/$g^{-1}_{DW}h^{-1}$. The 0.85 mmol/$g^{-1}_{DW}h^{-1}$ difference between these two measurements would make a difference of 2 mmol/$g^{-1}_{DW}h^{-1}$ at the estimated value of ATPM. Therefore,the accumulation of measurement errors for about 30 metabolites would obviously have a significant impact on the optimization result. Moreover, ATPM is a parameter to make sure that the minimum required energy of the cell (non-growth associated) is produced while maximizing the biomass production. Thus, the divergence from the ATPM value to a higher level would not be a violation of the ATPM assumption, whereas the deviation to a lower level should be done by having more solid evidences. In the case of this study, the estimated value for both pH conditions seem to be very low and therefore had to be increased. If lactate production is considered as an indicator of the energy production state in the system, the fact that lactate production in the mutant is several times higher than in the wildtype would imply that the ATPM value of the mutant should be considerably higher than in the wildtype (the estimated ATPM value for the wildtype is reportedly 3.9 and 8.4 mmol/$g^{-1}_{DW}h^{-1}$ at pH 7.5 and 6.5, respectively [Großeholz et al., 2016]). As a result, the lower bound of the ATPM reaction (ATPM value) was increased to 9.7 and 10.6 mmol/$g^{-1}_{DW}h^{-1}$ for pH 7.5 and 6.5, respectively. At pH 7.5, this value ensures the correct production rate of biomass. Regarding pH 6.5, the value can possibly go higher, but it does not change the outcome of optimization. These values enable the accurate prediction of the production rate of biomass and fermentation products. It has to be mentioned that the chemostat cultures growing the mutant were not only glucose-limited, but also glutamine limited. Therefore, it can be suggested that glutamine was an additional growth limiting factor in these cultures.

Predicted flux through the energy metabolism in the Δ*glnA* mutant

As described above, I used a previously developed framework for integrating the metabolic and proteome data into genome-scale metabolic models [Großeholz et al., 2016] to perform the integrative analysis of Δ*glnA E. faecalis* metabolism during the pH shift experiment. For this purpose, the update/production rate of carbohydrates, organic acids and amino acids were inte-

grated into the upper/lower boundaries of the respective exchange reactions (measured flux ± 20%). For the integration of the proteome data, the flux of reactions associated with no proteins in the proteome data was set to zero. To account for the protein fold changes after the pH shift, the log fold change value of the respective enzymes were used to adjust the reaction boundaries.

The experimental data showed that the mutant has a higher energy demand compared to the wildtype. This was successfully captured by the model, reproducing a higher ATP production rate and a higher flux through glycolysis as the central part of energy metabolism. The higher energy demand was also reflected in the higher flux through lactate dehydrogenase (LDH) which can be considered as an indicator of the energy requirements of the system.

I also compared the flux distribution profile in glycolysis between the mutant and the wildtype to find out about the differences in flux values in energy metabolism. The flux through glycolysis in the mutant was higher at pH 7.5, consistent with the higher energy demand in the mutant as a result of glutamine auxotrophy. The need for extracellular glutamine in the model is translated to glutamine uptake either by GLNab (glutamine ATP binding cassette) transporter or by glutamine permease. Using GLNabc would require ATP to import glutamine into the system, while glutamine permease would import a proton along with glutamine which further has to be actively pumped to extracellular environment. Either way, glutamine uptake would add additional energy costs to the cell. However, glutamine synthesis from glutamate would also consume one ATP per glutamine, preventing to draw any conclusion on the underlying reason for higher energy demand in the mutant at this point.

Impact of glutamine uptake in the model of Δ*glnA E. faecalis*

Glutamine auxotrophy in the mutant requires the organism to take up glutamine from extracellular environment at a higher level compared to the wildtype. Based on experimental data, the glutamine uptake in the mutant increased from 0.147 to 0.197 mmol/$g^{-1}_{DW}h^{-1}$ from pH 7.5 to pH 6.5. The model originally contained two mechanisms for glutamine uptake, namely GLNabc and glutamine permease, as mentioned previously. Based on the proteome data, the ATP binding site (EF0760) in GLNabc protein complex was up-regulated following the pH shift, suggesting an increase in transporter demand which is expected based on the metabolic data. Since a lot of membrane associated proteins were absent in the proteome data due to technical issues (protein isolation), it is possible that other subunits of GLNabc were also subjected to an upregulation after the pH shift. However, after applying the upregulation of transporter into the model, the original transporter design in the model failed to reproduce the higher glutamine flux by means of GLNabc. The flux distribution obtained from FBA showed a zero flux through GLNabc reaction at both pH levels. Therefore, I changed the design of the glutamine transport system in the model to enable it to reproduce the experimental data. Reportedly, the glutamine permease in *E. faecalis* is not an amino acid specific enzyme, but is rather shared by multiple

amino acids [Holden and Bunch, 1973]. Among all, glutamine has the highest affinity for the permease, while asparagine and threonine have a lower affinity. Therefore, to account for the difference in the level of affinity of different amino acids, I introduced a new permease reaction to the model, carrying two molecules of glutamine together with one asparagine, one threonine and 4 protons. Accordingly, the single amino acid permease reactions of all three amino acids were deactivated. The newly developed transport system correctly mimics the uptake rate of all three amino acids and also brings the glutamine ABC transporter into use at pH 6.5. This system results in a more accurate reproduction of the actual transport system in the organism as it takes the shared permease system into account and also makes use of the glutamine ABC transporter when necessary. The involuntarily uptake of other amino acids besides glutamine either by means of permease or active transport might provide a potential answer on the higher energy demand in the mutant. Furthermore, the absence of glnA potentially prevents GlnR (the regulatory protein) to properly function, as it is reported that GlnR needs glnA to control the glutamine uptake in *Streptococcus pneumonia* [Kloosterman et al., 2006]. As a result, the lack of control over glutamine uptake might cause an unnecessary uptake of glutamine and potentially all the other amino acids that share the same transporter together. Hence, the excessive import of amino acids would force additional energy costs to the organism.

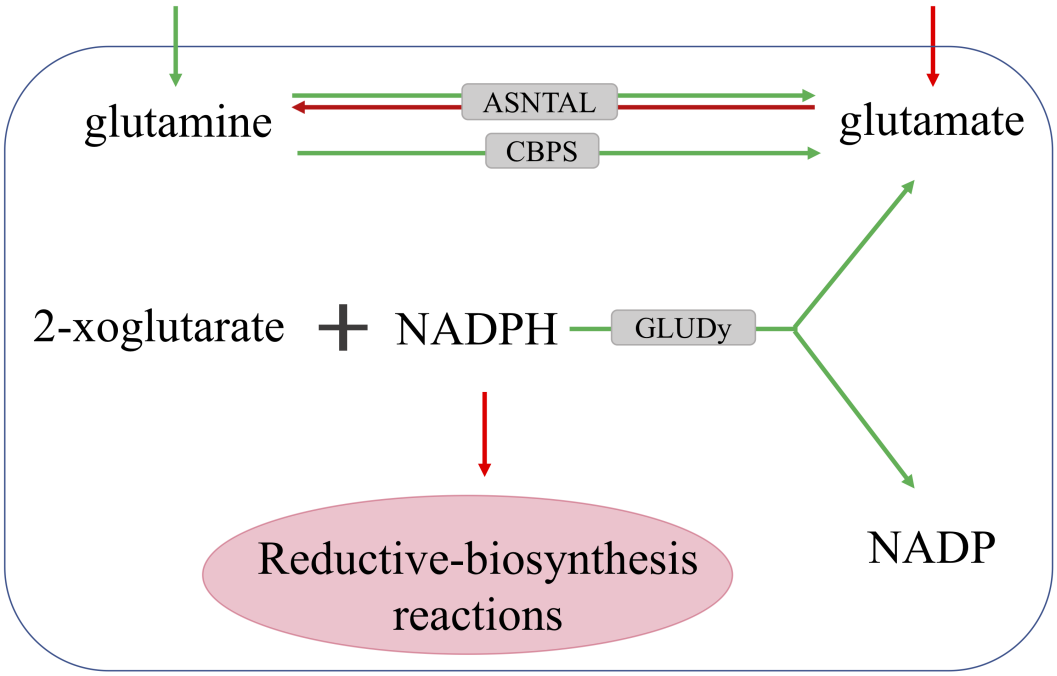

Figure 3.4: The effect of increased glutamine uptake and decreased glutamate uptake on the glutamine/glutamate metabolism. Following an increase in glutamine uptake, more glutamate is produced from glutamine by means of ASNTAL (aspartyl-tRNA(Asn):L- glutamine amido-ligase (ADP-forming)) and CBPS (carbamoyl-phosphate synthase (glutamine-hydrolysing)). Moreover, the increased glutamate production by GLUDy (glutamate dehydrogenase) results in less NADPH being available for reductive-biosynthesis reactions.

Predicted flux through glutamine/glutamate metabolism

I further used the flux distribution profile of the genome-scale model obtained from FBA and FVA to analyse the glutamine/glutamate metabolism in response to pH shift. This pathway was particularly important as the aim of this study was to investigate the effect of glutamine auxotrophy on metabolic characteristics of *E. faecalis*. The increase in glutamine uptake and the decrease in glutamate uptake following the pH shift in the genome-scale model is translated into a higher conversion of glutamine to glutamate by means of a flux switch through two reactions, namely aspartyl-tRNA(Asn):L- glutamine amido-ligase (ADP-forming) (ASNTAL) and carbamoyl-phosphate synthase (glutamine-hydrolysing) (CBPS). Moreover, the model predicted that the glutamine-fructose-6-phosphate (gam6p) transaminase reaction would be shut down. This reaction produces gam6p by catalyzing a transaminase reaction between glutamine and fructose-6-phosphate (f6p). The model rather produced gam6p by assimilating ammonia into f6p. Furthermore, the model produced glutamate from 2-oxoglutarate by increasing the flux towards the reverse direction of glutamate dehydrogenase (GLUDy). The directionality of GLUDy is considered as an indicator of the balance between the carbon and nitrogen metabolism [Hwang and Lee, 2018]. The redirection of flux towards glutamate production would change the NADPH/NADP ratio such that less NADPH would be available for, e.g., amino acid biosynthesis. This is consistent with the lower biomass production at pH 6.5. The decrease in the NADPH level also prevents reductive synthesis reactions to occur (Figure 3.4), which can be also observed in the proteome data by the significant downregulation of proteins involved in nucleotide metabolism. Another effect of the reverse direction of GLUDy reaction is consuming one proton, which is important for the cell in a more acidic environment. Moreover, the production and the degradation of some other amino acids were also affected following the direction change in GDH flux. For example, the conversion rate of glutamate to alanine and aspartate was predicted to be higher, with alanine being produced by the organism after the pH shift, based on the metabolic data.

3.2 Inspecting the solution space of genome-scale metabolic model

Having studied the impact of glutamine auxotrophy on metabolic adjustments, protein expression and flux distribution in the genome-scale model of *E. faecalis*, it became clear that the genome-scale model can mimic the actual physiological condition to a large extent. However, as mentioned before, there are limitations in the extent to which the results of the model can be interpreted, the most prominent being the internal flux distribution in the model. Therefore, it is necessary to determine the degree of robustness of the genome-scale model from a functional stand point. This would enable to determine which parts of the predictions made by the model are more reliable. to do so, I developed a method to discover the degree of sensitivity/robustness in genome-scale models. In this sub chapter, I present the results of the method development,

which is previously published in [Loghmani et al., 2022].
As described previously in the introduction, there is no single solution for genome-scale metabolic networks, but rather and often solutions are contained in a huge solution space. This is due to the fact that they are mathematically underdetermined, meaning that there are more unknown (reactions) than known (metabolites) variables in these systems. Inevitably, a large part of the solutions provided by these systems are contradictory to biological phenotypes. This shortcoming is usually overcome by integrating experimental data as additional constraints into models, attempting to shrink the solution space. However, even after integrating additional constraints the solution space although smaller than before, remains large enough so the probability of acquiring contradictory results is still high. Therefore, whenever a genome-scale model is used to draw a biological conclusion, it is necessary to discover and analyse numerous sets of alternative solutions. In this perspective, taking the flux range of reactions into account would be a more solid basis to draw conclusions rather than just considering a single value obtained by FBA. For this purpose, FVA is a common approach as it provides the feasible range of all the reactions for which an alternative optimal value exists. Nonetheless, FVA does not provide any information on how different combinations of these alternative optimal values shape the flux distribution profiles. To acquire such knowledge, one way to go is to calculate all the possible flux combinations and enumerate the number of solutions. Unfortunately, this approach is not computationally feasible when the solution space is relatively large (which is often the case). A feasible alternative to this process would be to have an estimation of how the overall pattern of solutions looks like by sampling the solution space. A comprehensive overview of the background and perspective of different sampling methods is described in the introduction. Here, I present the result of the new method I developed for sampling the solution space by means of random perturbation. Using this method, one would be able to assess the sensitivity and robustness at the whole-network level as well as investigating various potential scenarios at the branching points in the network, where the qualitative aspect of solutions are decided. As the case study, I used the previously published genome-scale metabolic model of *E. faecalis* (wildtype and the Δ*glnA* mutant) [Veith et al., 2015], genome-scale model of *S. pyogenes* [Levering et al., 2016] and *Lactococcus lactis* [Flahaut et al., 2013]. I also used the previously mentioned framework for the integration of metabolic and proteomic data [Großeholz et al., 2016] to examine the extent to which various constraints affect the solution space.

3.2.1 Effect of different constraints on the flux variability in the network

As the first step to analyse the effect of different constraints on the solution space, I analysed how various sets of constraints affected flux variability in the network. For this purpose, I used the genome-scale metabolic model of *E. faecalis* in the wildtype and the Δ*glnA* mutant from [Veith et al., 2015] and integrated the constraints stepwise. Starting from a completely non-constrained model, I first integrated the medium composition data, followed by the integration of the metabolic data, namely the uptake rate of glucose, production rate of organic acids (fer-

Table 3.1: Number of variable reactions in differently constrained genome-scale models of *E. faecalis* wildtype (wt) and the Δ*glnA* mutant (mt). Here, "nc" indicates model version without any constraints, 181 "med" indicates integration of medium composition, "met" the additional integration of data on 182 metabolite uptake and release and "pro" the additional integration of proteome data. The table is adapted from [Loghmani et al., 2022].

Model Name	**# Variable Reaction Variability $>10^{-6}$**	**# Variable Reaction Variability $>10^{-3}$**	**No of Reactions**
mt + nc	397	397	708
mt + med	362	340	708
mt + med + met	347	315	708
mt + med + met + pro	298	289	708
wt + nc	398	398	709
wt + med	363	341	709
wt + med + met	362	340	709
wt + med + met + pro	307	85	709

mentation products) and amino acid, and finally, the proteome data. To determine the level of variability in the system at each step, the feasible flux interval for each reaction was calculated by FVA (Table 3.1). The addition of constraints at each step resulted in a decrease in the number of variable reactions in the network. This information is important as it's the variability at the level of individual reactions which gives rise to the very large number of feasible solutions. Therefore, the reduction in the number of variable reactions gives an initial hint on how the solution space changes following the integration of constraints.

3.2.2 Inspecting the solution space using random perturbations

As mentioned in the previous paragraph, the information regarding the number of variable reactions in the network provides a good introductory insight on how the solution space is affected by the integration of constraints. However, it is incapable of providing a deep insight on how biological phenotypes are distributed in the solution space. To obtain such information, I developed a method to randomly perturb the results of FBA and assess how robust the flux distribution and the resulting conclusions are. Therefore, I first obtained the feasible flux interval for all the variable reactions. An interval size of 10^{-6} was considered as a threshold for a reaction being "variable". All the other reactions were considered as "stable". Interestingly, the majority of the variable reactions in these two models (wildtype and mutant *E. faecalis*) had an interval size of larger than 10^{-3}. Next, for each variable reaction, I randomly chose 10 different values, fixed the reactions at the random values (by making the upper and lower bounds equal to that value) and performed FBA/FVA to obtain the alternative flux distribution profiles. This procedure will be referred throughout this work as the "perturbation". In order to have an accurate account on the behaviour of reactions during the perturbation process, the variable reactions whose flux value in response to perturbation changed for more than $\pm 5\%$ will be referred to as

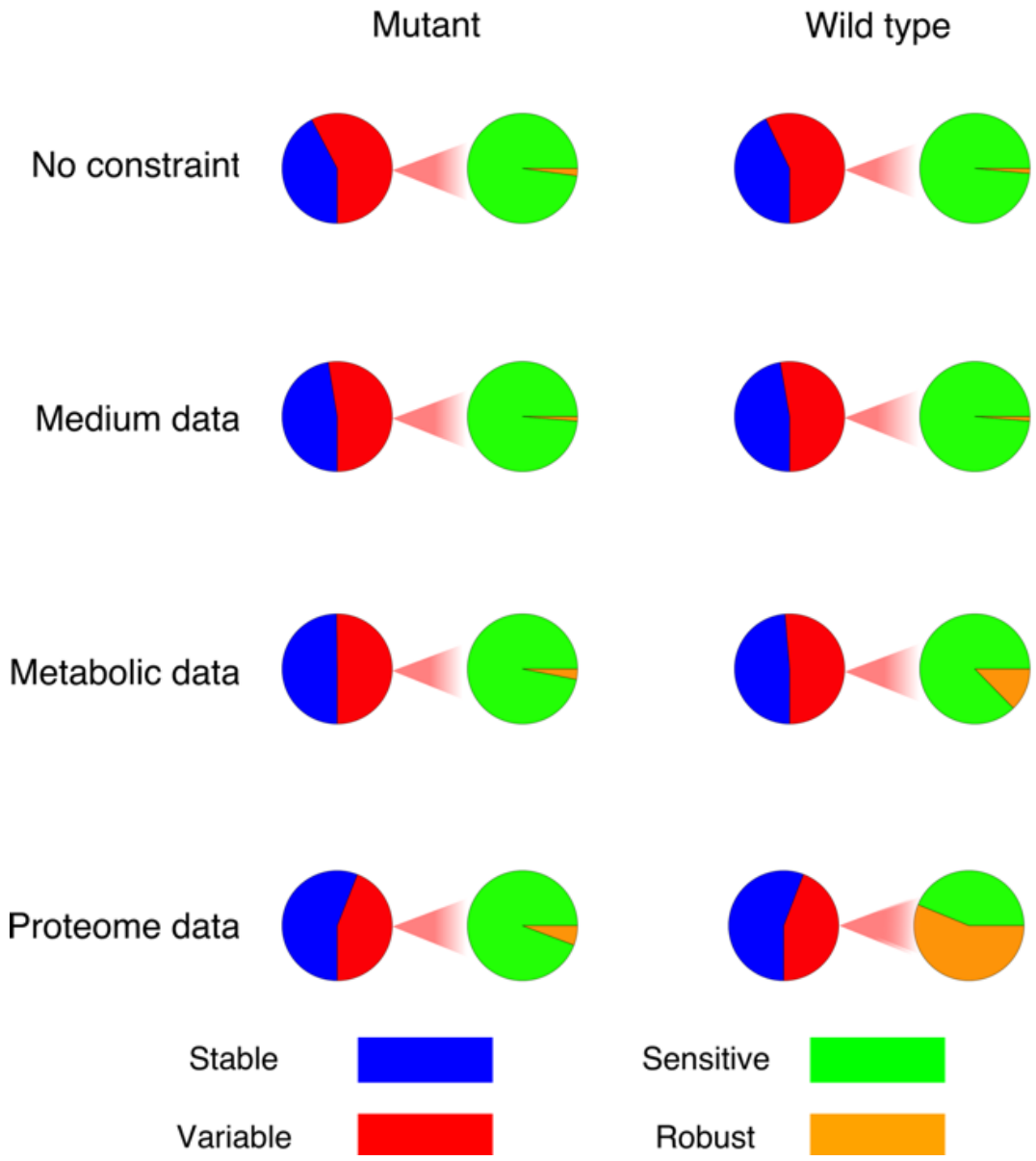

Figure 3.5: The fraction of variable reactions that are sensitive or robust in each model. The term "robust" in this figure refers to those variable reactions whose variability interval (feasible flux interval) is smaller than the $\pm 5\%$ of the original value. The figure is adapted from [Loghmani et al., 2022].

"sensitive", while the rest of variable reactions which take more or less the same flux value all the time, irrespective of any perturbation, will be referred to as "robust". The $\pm 5\%$ threshold provided a good basis for distinction of characteristics between different models in this setup, but it can be changed to other values if necessary. It has to be pointed out that the concepts of "stable" and "robust" here are essentially different, as "stable" refers to reactions for which an alternate optima does not exist according to FVA, while "robust" refers to those of "variable" reactions which did not change their flux in response to perturbation by more than $\pm 5\%$. I also found that the width of the feasible interval has a direct impact on whether a reaction is sensitive or robust according to the provided definition of sensitivity. That means, if the width of the feasible interval in a reaction is larger than $\pm 5\%$ of the original flux value, the reaction is most likely sensitive, and robust otherwise. Therefore, to calculate the number of sensitive reactions, it is not necessary to perform the perturbation procedure. However, perturbation is necessary to investigate additional properties of sensitive reactions.

I applied the above-mentioned strategy on the wildtype and the Δ*glnA* mutant of *E. faecalis* before and after the integration of experimental constraints. First, I calculated the number of sensitive reactions in each model fig 3.5. Expectedly, the number of sensitive reactions decreased following the integration of constraints. However, the trend in which the number of sensitive reactions changed was surprisingly independent of that of the variable reactions. For instance, after the integration of proteome data, the number of variable reactions in the model of wildtype was higher than in the model of mutant, but its number of sensitive reactions was considerably lower. Furthermore, the integration of constraints at each step resulted in a considerable drop of the number of reactions that changed the directionality of their flux in response to perturbation.

Having defined the concepts of sensitivity and robustness, a more in-depth analysis is necessary to grasp a better understanding of these concepts in constraint-based models. One of the key findings in this study was that the overall sensitivity in the model, namely the number of sensitive reactions, has a direct correlation with and therefore can be used as an indicator of the size of the solution space. Later in this chapter, I would discuss how the wildtype model constrained with medium composition, metabolic and proteome data (wt+med+met+pro) has a considerably more limited solution space compared to its mutant counterpart. The comparison between these two cases is particularly informative as the former model has a much lower number of sensitive reactions than the latter.

In order to have a more elaborate analysis of the sensitivity in the genome-scale models, I performed a comprehensive statistical analysis on the results of the perturbation. For one, I calculated the average number of the times that one perturbation in a variable reaction resulted in a flux change in a sensitive reaction. This can be looked at in two different ways: 1. the number of perturbations in a reaction that resulted in a flux change in other reactions, and 2: the number of reactions that changed their flux in response to perturbation in that particular reaction. I refer to the first case as "affecting average perturbation-wise" and the second

case as “affecting average reaction-wise“ (Figure 3.6). The distinction between the two is necessary and will be discussed in more details later in this chapter. The comparison of these two statistical indices among different models showed that the models of wildtype and mutant represent a different pattern after integrating the constraints. In the model of wildtpye, the integration of constraints consistently reduced the “affecting average“ index, both reaction-wise and perturbation-wise. Whereas in the model of mutant, a considerable decrease occurred after the integration of medium composition data, but the index increased in the next two steps, again both reaction-wise and perturbation-wise. The results are shown in the fig 3.7. A more in-depth analysis is necessary to uncover the reason for discrepancy between the results of the two models. However, to the extent of this study, it can be suggested that this index can be hugely affected by the numerical routines and software implementations.

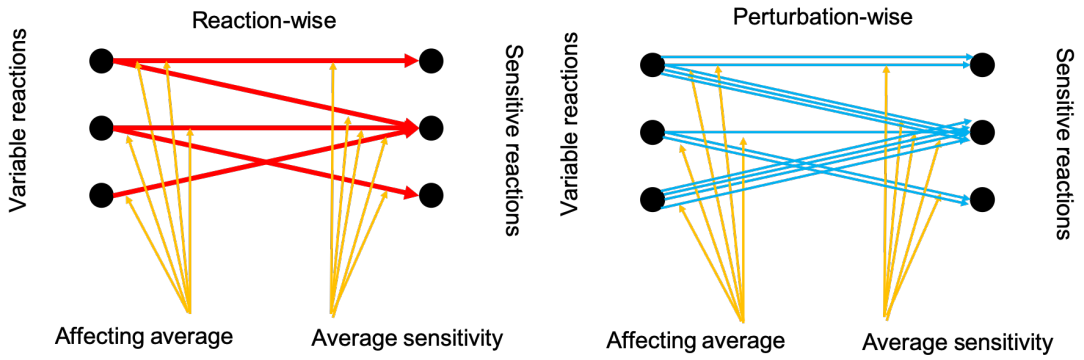

Figure 3.6: A hypothetical representation of affecting average and average sensitivity. Affecting average shows how often on average a perturbation in a reaction causes a flux change in another reaction. Average sensitivity denotes the average number of time a reaction undergoes a flux change as a result of a perturbation in another reaction. The reaction wise analysis reports these two statistical indices with respect to the number of reactions, while perturbation wise analysis does so with respect to the number of perturbations.

To gain a better understanding of how perturbations in some reactions would affect the flux value in some other reactions, the same analysis was performed from a different perspective. This time, I calculated the average number of times that a reaction changes its flux in response to perturbation in another reaction. I refer to this index as “average sensitivity“. Similar to the previous index, I distinguished between the “reaction-wise“ and the “perturbation-wise“ indices. The “reaction-wise“ index indicates the average number of variable reactions that a sensitive reaction changed its flux in response to their perturbations. The “perturbation-wise“ in turn shows the average number of perturbations in variable reactions in response to which a sensitive reaction changed its flux. The trend that “average sensitivity“ developed after the integration of constraints was very similar to that of “affecting average“, again both reaction-wise and perturbation-wise (Figure 3.8). Likewise, the average number of flux changes in response to perturbations in exchange reactions in both models changed similarly to the two previous indices(Figure 3.9).

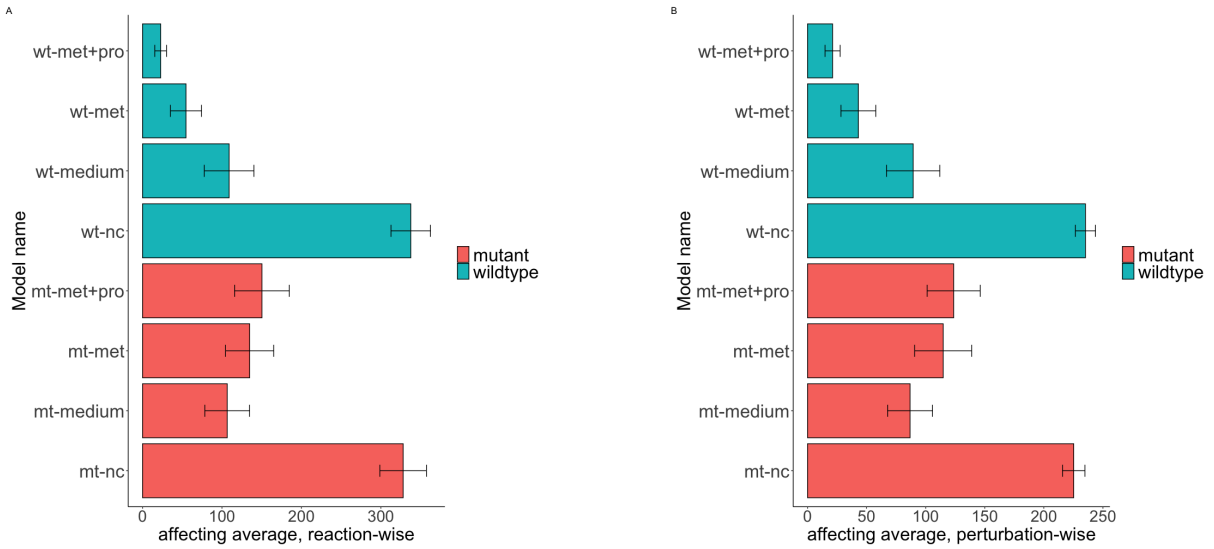

Figure 3.7: The average number of flux changes (reaction-wise and perturbation-wise) in response to a perturbation in a variable reaction

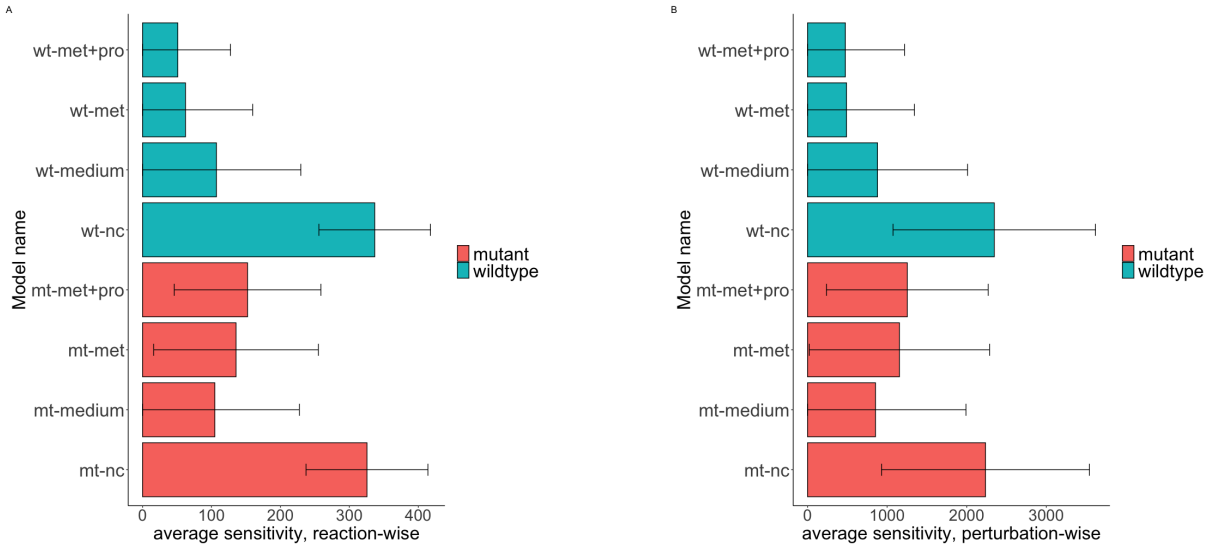

Figure 3.8: The average number of times a sensitive reaction responded to perturbations in variable reactions (reaction-wise and perturbation-wise)

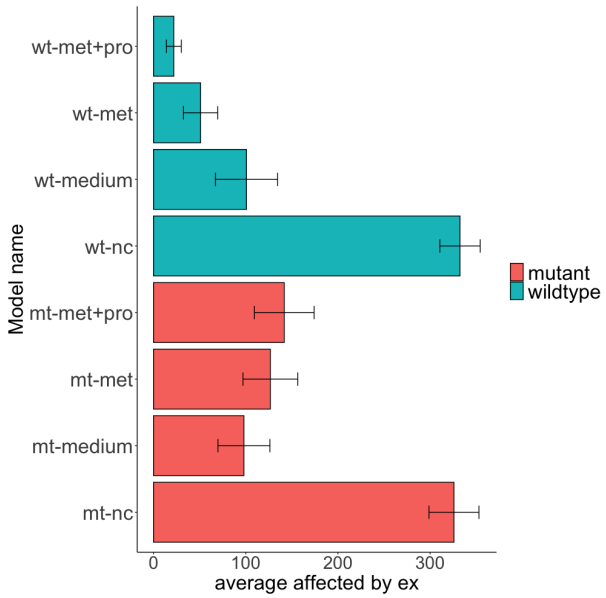

Figure 3.9: The average number of flux changes in response to perturbations in exchange reactions.

As previously mentioned, to find out whether numerical libraries and softwares impose bias and affect the result of the perturbation procedure, I repeated the procedure using other software and numerical solvers. The original study was performed using COBRA toolbox on MATLAB, CPLEX as the numerical solver, and the flux distribution profiles obtained from fastFVA function (a faster implementation of fluxvariabilityanalysis in COBRA). Therefore, to assess the effect of library/function -generated biases, I repeated the analysis in COBRA on MATLAB, obtaining the flux distribution profiles using FBA (optimizeCBmodel function), with CPLEX and glpk as numerical solvers. I also repeated the analysis in python, using PySCeS-CBMPy, CPLEX as the solver and the flux distribution profiles obtained with FVA. Surprisingly, the comparison of the results from the different methods showed a huge disagreement, not only in the numerical values of the different indices, but also in their trend of development after integrating the constraints (Figure 3.10). As shown in the figure, the most pronounced difference appeared when the affecting average and the average sensitivity indices were compared among the different approaches reaction-wise. The affecting-average index increased from mut-med onwards when using FVA in COBRA and CBMPy, while using FBA in COBRA resulted in a decrease from mut-med to mut-med-met and an increase to mut-med-met-pro, in the case of the both solvers (CPLEX and glpk). Here, there is clearly a bias imposed by the method of use (FVA or FBA) to obtain flux distributions. Between the two software, COBRA and CBMPy, however, the numerical values are considerably different, while the trend into which the index developed was similar when using FVA. Furthermore, despite a small difference in numerical value, the use of different numerical solvers, namely CPLEX and glpk did not significantly impact the result of the perturbation procedure. A relatively similar pattern was observed when average sensitivity was compared between the two software. However, the trend to which the average sensitivity developed was different between COBRA and CBMPy. The average sensi-

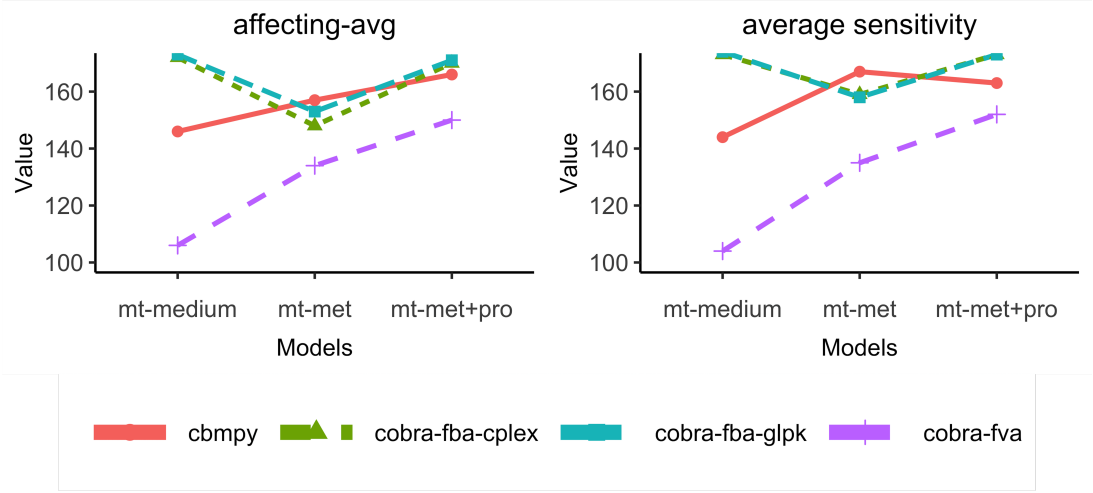

Figure 3.10: The comparison of the affecting average and average sensitivity indices, reaction-wise, among different methods. The figure shows that there are not only numerical differences among the results of different methods, but also qualitative differences in the trend at which the two statistical indices change following the integration of constraints.

tivity generated by COBRA followed a similar pattern to that of the affecting average, while it decreased from mut-med-met to mut-med-met-pro when generated by CBMPy. Here, a clear bias towards the software of use can be observed.

I also compared the affecting average and the average sensitivity indices among the different approaches in perturbation-wise manner (Figure 3.11). The perturbation-wise results showed a higher level of consistency among the different approaches. The affecting average index increased from mut-med onwards in the results from all the software/functions/solvers. However, when CBMPy (FVA) was used, the increase from mut-med-met to mut-med-met-pro was very small. Here, the choice of software did not have a pronounced impact on the trend to which the index developed, but it had a huge impact on the individual numerical values of the index. When comparing the average sensitivity results, however, the trend was not consistent among all the approaches. The use of CBMPy (FVA) resulted in an increase from mut-med to mut-med-met and a further decrease to mut-med-met-pro, while all the methods in COBRA (FBA using CPLEX and glpk, FVA) reported a constant increase at each step. Interestingly, the results from CPLEX and glpk (FBA in COBRA) were almost the same with negligible differences. Similar to the previous cases, the choice of method and software had a big impact on the numerical value of the index. All in all, the comparison of the results among the different methods showed a clear bias towards the software/function/solver of choice. Consequently, I regard the analysis of the affecting average and the average sensitivity as inconclusive, as a consistent reproducibility of the indices is impossible by various method. Moreover, the increasing trend in the model of mutant (from the medium level to proteome level), while the otherwise is expected, cannot be explained at this stage of the study and more detailed analysis is necessary to uncover the underlying reasons.

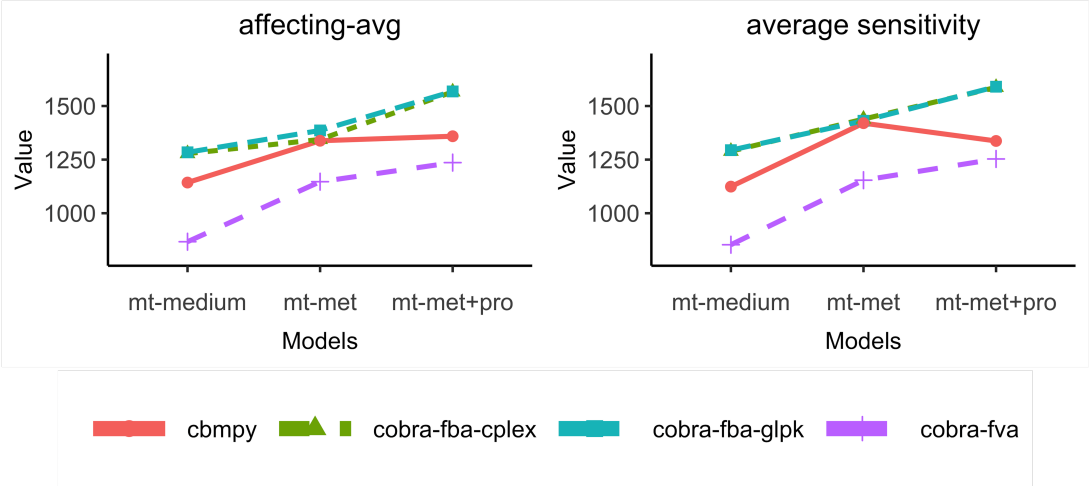

Figure 3.11: The comparison of the affecting average and average sensitivity indices, perturbation-wise, among different methods. Similar to the previous figure, both quantitative and qualitative differences exist in the results obtained by different methods.

In the previous paragraphs I discussed the consistent and reproducible statistical indices (number of variable and sensitive reactions) and irreproducible indices (affecting average and average sensitivity) in the used genome-scale metabolic models. As this study aims to uncover the properties of the solution space in these models, I tried to find the consistent routine in which perturbation affects the sensitive reactions. Considering the fact that it is the sensitive reactions which give rise to a big solution space in genome-scale models, uncovering the trend in which sensitive reactions perform would help us to grasp a better understanding of the solution space. Therefore, I investigated the interval in which the sensitive reactions responded to perturbation, as well as the frequency of individual flux values within those intervals. Performing such analysis in the case of the three models, namely mut-med, mut-med-met, mut-med met-pro, where the number of sensitive reactions decreased consistently and expectedly, and the affecting average and the average sensitivity indices increased unexpectedly would help to find an overall consistent and conclusive pattern on how sensitive reactions respond to perturbation. For that matter, I first divided the individual flux values taken by the sensitive reactions (in response to perturbation) into 20 bins and calculated the number of values in each bin (Figure 3.12 panel A, B, C). Second, I compared the interval in which those values were taken (sensitivity interval) to the interval given by FVA (variability interval) (Figure 3.12 panel D, E, F). The result showed that not surprisingly, the flux values taken by the sensitive reactions in response to perturbations spread in larger intervals in the mut-med model compared to mut-med-met and mut-med-met-pro (the models integrated with more constraints). Following the addition of constraints, more and more fluxes sat on the boundaries of the variability intervals. This holds true both regarding the number of reactions which take flux values at the boundaries of the variability intervals, and the frequency of flux values sitting at the boundaries of the variability intervals. In another word, in the mut-med model, the variability and the sensitivity intervals are relatively distanced from each other in several reactions, but the distance shrinks following the addition of constraints in the mut-med-met and mut-med-met-pro models. In the mut-med-met-pro model, the

two intervals are aligned in almost all reactions (Figure 3.13).

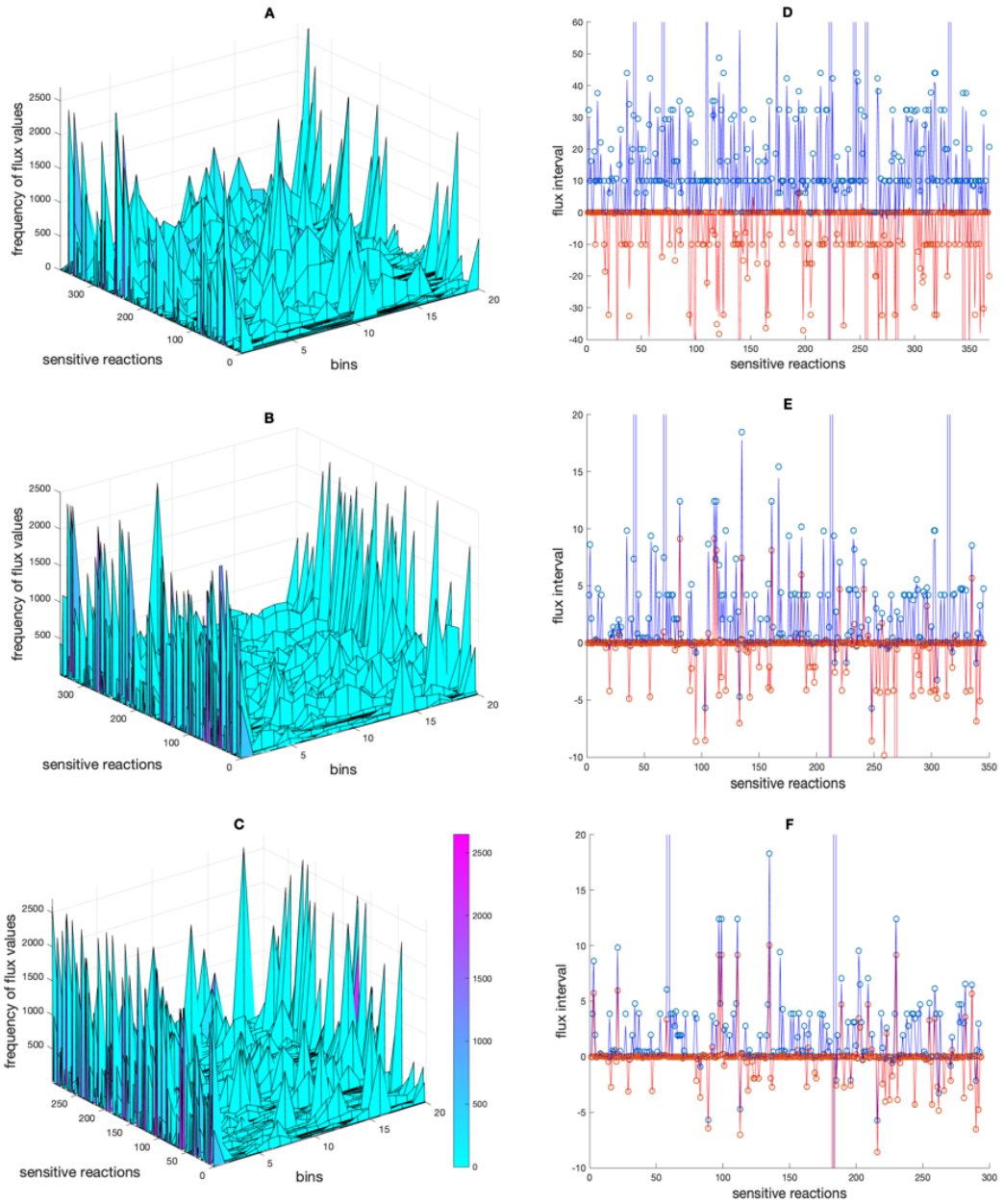

Figure 3.12: The distribution of alternative flux values across the flux intervals. Panel (A–C) show the frequency of flux values of sensitive reactions divided into 20 bins in the mt + med, mt + med + met and mt + med + met + pro models, respectively. Panel (D) to F show the intervasl in the respective models in which sensitive reactions responded to perturbations (red and blue lines, indicating lowest and highest flux values, respectively), and the interval given by FVA, indicated by red dots (lower bounds) and blue dots (upper bounds). For the sake of clarity, a few extreme points in panel (D–F) are excluded. The figure is adapted from [Loghmani et al., 2022].

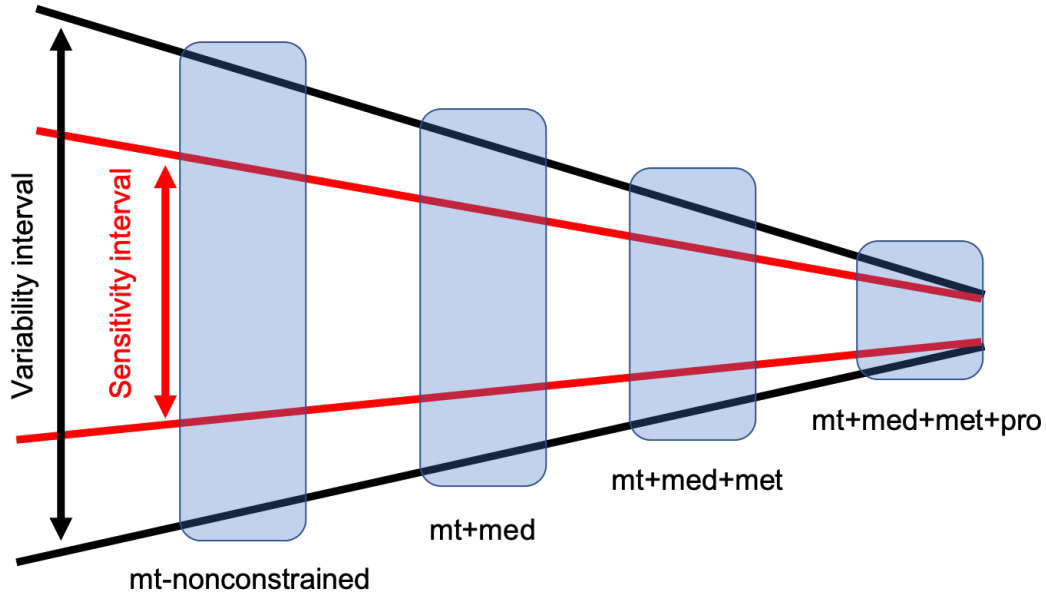

Figure 3.13: Hypothetical depiction of how the integration of constraints shrinks the gap between variability intervals and sensitivity intervals. According to the results of the perturbation procedure, the variability interval and sensitivity interval are not necessarily the same, meaning that the part of the variability interval is sometimes unused when the solution space is sampled. Following the integration of experimental data and shrinking the solution space, the distance between the two intervals decreases.

The result suggests that FVA is a key, initial step in the analysis of the solution space. It provides the mathematically feasible range for individual reactions. However, it does not provide any information on the interval that alternate optima is most likely taken. This study showed that this latter interval, namely the sensitivity interval, is a more informative indication of the size of the solution space, and the perturbation procedure presented in this study is an efficient way to calculate the sensitivity interval.

3.2.3 Investigating biological phenotypes in FBA results

In order to investigate the variation in biological phenotypes in the solution space of FBA, I analysed the flux distribution at three branching points in the two models of *E. faecalis* using the perturbation approach. Branching points provide a good example on how biological phenotypes can significantly vary within a given set of intervals. That said, by taking the results of FVA into account, one would be able to gain knowledge on the feasible range of individual reactions, but no information on different flux combinations that would change the quality of biological phenotypes.

First, I analysed the flux distribution at the pyruvate branching point. There, pyruvate is distributed into two major fermentation routes, lactate fermentation and mixed-acid fermentation comprising ethanol, acetate, and formate production (Figure 3.14). The predominantly lactate production is referred to as homolactic fermentation, while the otherwise is called mixed-acid fermentation. This branching point is of biotechnological importance as it decides for the out

coming fermentation products. Furthermore, it is often regarded as a deciding factor and therefore an indicator of the state of energy production in the system. Normally, lactic acid bacteria use homolactic fermentation under high energy demand and mixed-acid fermentation under the otherwise. To analyse the flux distribution through this branch, I calculated the ratio between the homolactic and mixed-acid fermentation, generated by various FVA runs during the perturbation process in the models with different sets of constraints (Figure 3.15).

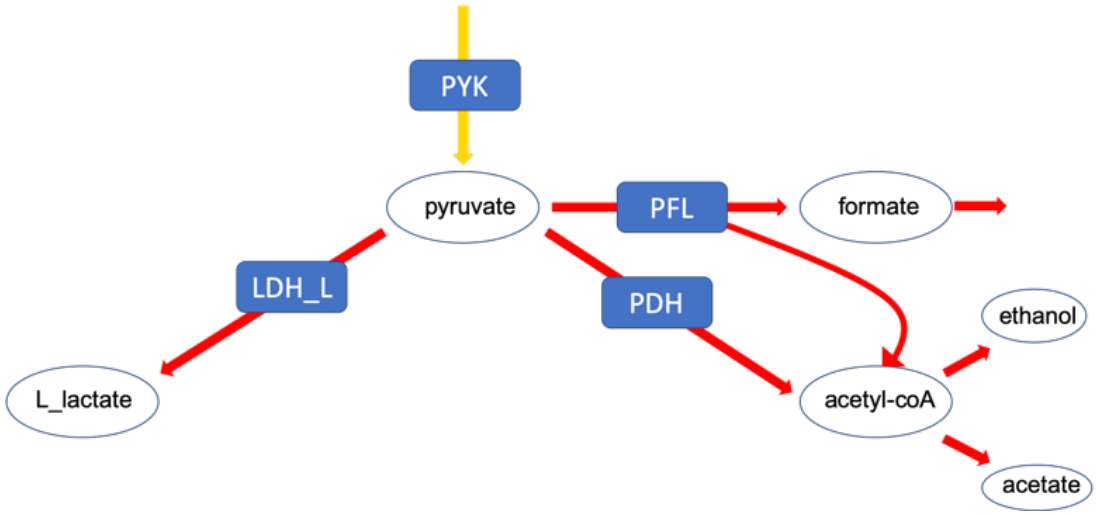

Figure 3.14: The pyruvate branching point. This branch decides for the type of fermentation in the system. The figure is adapted from [Loghmani et al., 2022].

As shown in fig 3.15, the integration of constraints at each step resulted in a reduction in the variability of relative flux distribution at the branching point. The non-constrained model, both in the case of the wildtype and the mutant, showed a very highly variable flux distribution, ranging from entirely homolactic fermentation to a lactate-free fermentation, expectedly so as there is no constraint in place. The integration of medium-composition constraints resulted in the removal of the lactate-free fermentation, but still an entirely homolactic fermentation along with mixed-acid fermentation exist in the solution space. This is a noticeable finding as FBA very often is performed using only medium-composition data, which is clearly shown that is not sufficient. The turning point in the stepwise integration of constraints into the models, both in the wildtype and the mutant, was the integration of metabolic data (uptake and production rates) which had an enormous impact on limiting the flux distribution. However, the extent of this impact was very different between the wildtype and the mutant. In the wildtype, in which the mixed-acid fermentation was dominant, the variability was almost vanished. Whereas in the mutant model which produced a higher percentage of lactate (based on experimental data), the extent of variability was much larger, ranging from homolactic to mixed-acid (less than 50% lactate) fermentation. The integration of proteome data helped to shrink the range of variability, but did not eliminate it.

I showed that the integration of constraints in the model of wildtype resulted in elimination of physiologically irrelevant results from the solution space. Hence, it can be suggested that

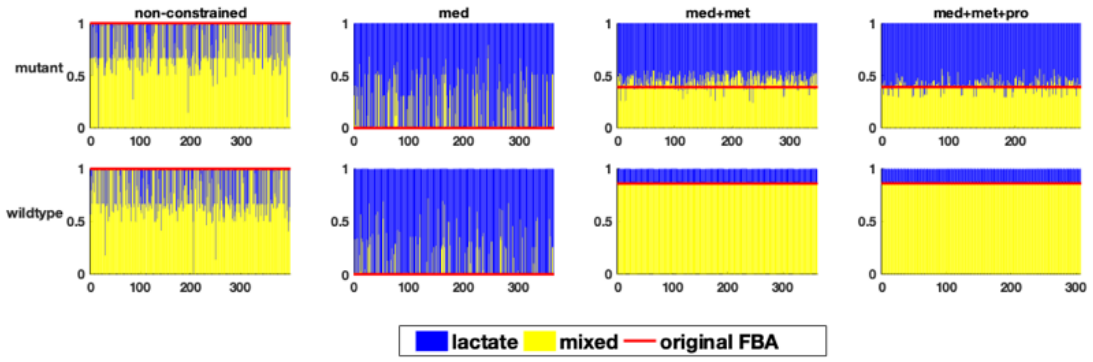

Figure 3.15: The relative flux distribution in the branching point of carbohydrate fermentation (y-axis) in response to one perturbation in each of the variable reactions (x-axis) in the two studied genome- scale models of *E. faecalis*, resulting in homolactic or mixed acid fermentation in the two genome- scale models of *E. faecalis*. The figure is adapted from [Loghmani et al., 2022].

the model of wildtype has a more limited and smaller solution space compared to the model of mutant. This is consistent with the results from the previous section on sensitivity analysis, showing that there were fewer sensitive reactions in the wildtype model and therefore suggesting a smaller solution space compared to the mutant model. It has to be emphasized that all the flux distribution profiles were obtained while satisfying the optimum value of the objective function (biomass production).

In the second example, I analysed the flux distribution at a branching point in serine metabolism (Figure 3.16). The flux through this branching point would spread to amino acid metabolism, tRNA loading and central metabolism by producing pyruvate.

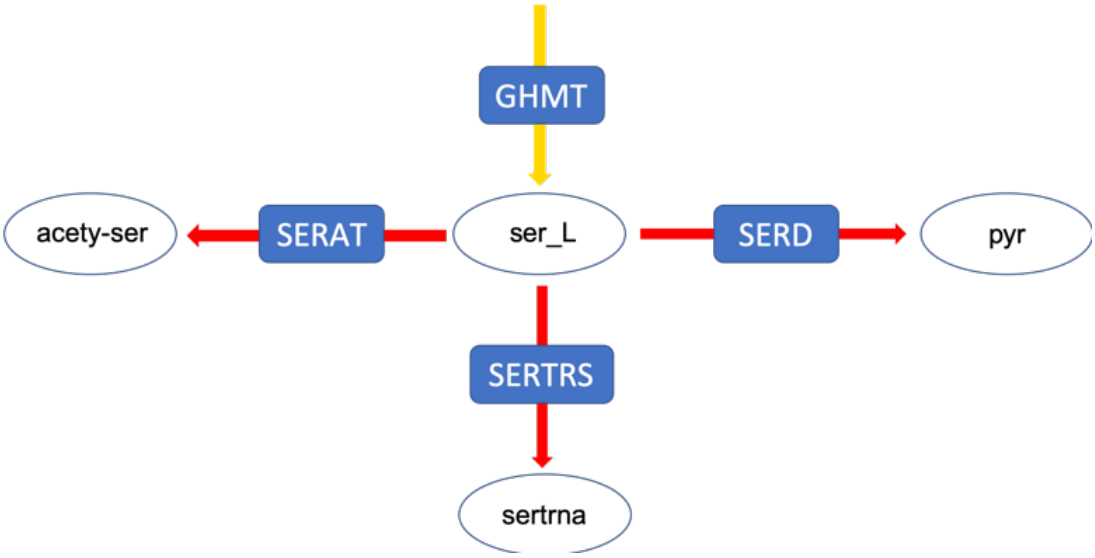

Figure 3.16: The serine branching point by which serine is distributed. Pyr: pyruvate; sertrna: L-seryl- tRNA; acetyl-ser: Acetyl-serine. The figure is adapted from [Loghmani et al., 2022].

The integration of medium-composition data in both models resulted in a very large fraction of flux being redirected into serine O-acetyltransferase reaction. The integration of metabolic data,

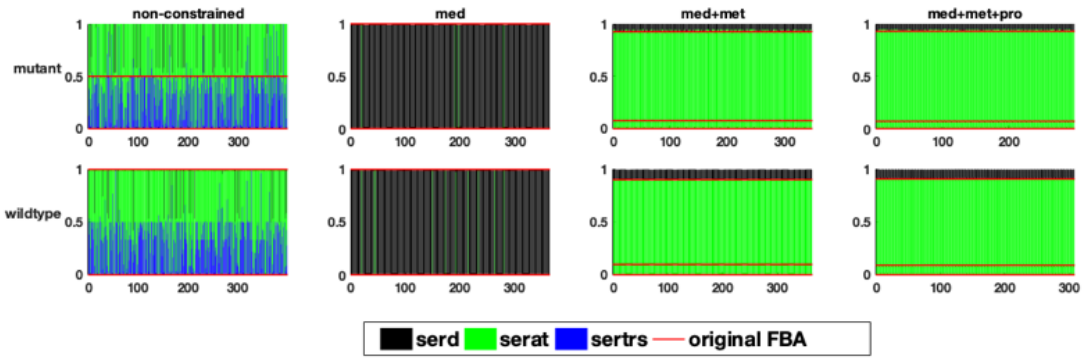

Figure 3.17: The relative flux distribution in the branching point of serine metabolism (y-axis) in response to one perturbation in each of the variable reactions (x-axis) in the two studied genome- scale models of *E. faecalis*, resulting in the production of acetyl serine, or seryl-tRNA or serine secretion. The figure is adapted from [Loghmani et al., 2022].

however, shifted the flux distribution to the other two reactions, namely serine dehydrogenase and seryl-tRNA synthetase. I have to mention that although I did not explicitly include the tRNA loading data in this study, but recently there have been efforts to include these data into genome-scale models, facilitating the integration of expression data [Gu et al., 2019] [Oftadeh et al., 2021] [Salvy and Hatzimanikatis, 2020] [Lloyd et al., 2018] [Garcia et al., 2020]. In the mutant model constrained with all the data, around 95% of the flux went through serine dehydrogenase reaction, which could vary by 1% in the solution space (Figure 3.17). The variation here was considerably smaller compared to the variation at the pyruvate branching point. In the case of the wildtype model, there is no variation at this branching point which again supports the idea that the solution space in the wildtype model is more limited and smaller than in the mutant model.

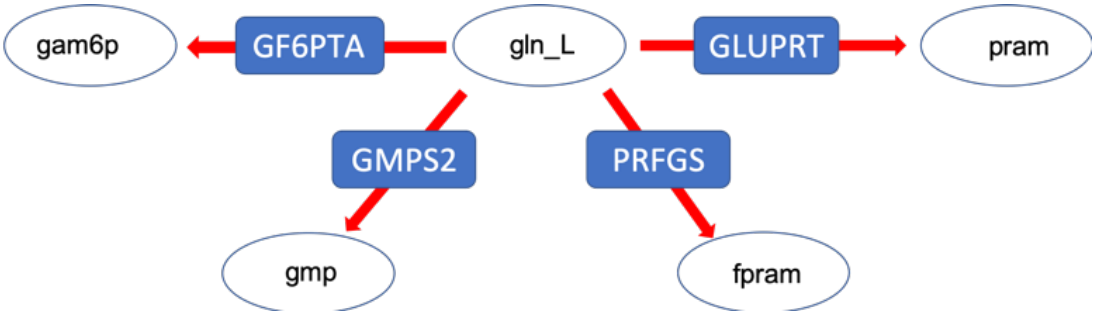

Figure 3.18: The glutamine branching point distributing amino-groups via the amino acid L-glutamine. Gln: glutamine, pram: 5-Phospho-beta-D-ribosylamine, fpram: 2-(Formamido)-N1-(5-phospho- D-ribosyl) acetamidine, gmp: guanosine monophosphate, gam6p: glucoseamine 6 phosphate. The figure is adapted from [Loghmani et al., 2022].

In the third example, I studied a branching point in glutamine metabolism which distributes nitrogen groups via the amino acid L-glutamine to different parts of metabolism (fig, 3.18). Similar to the previous two cases, the two models with no constraint and the medium-composition data showed a huge variability, whereas the integration of metabolic and further proteome data

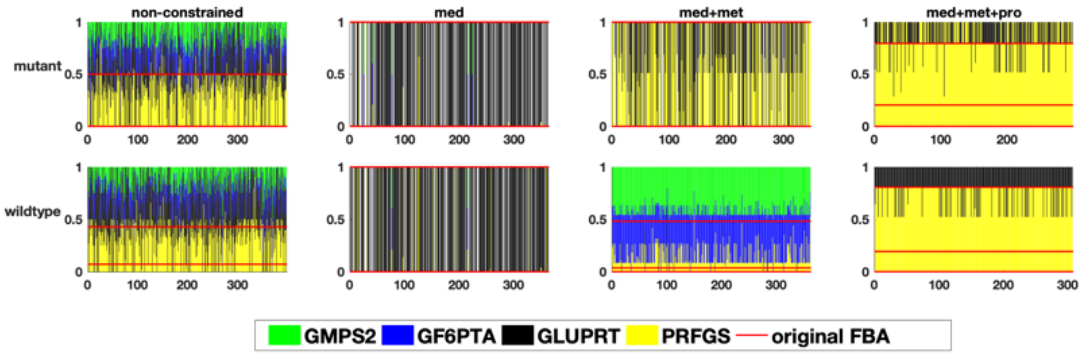

Figure 3.19: The relative flux distribution in the glutamine branching point(y-axis) in response to one perturbation in each of the variable reactions (x-axis)) in the two studied genome-scale models of *E. faecalis*, resulting in the distribution of glutamine in different pathways, namely amino acid, purine and pyrimidine metabolism.The figure is adapted from [Loghmani et al., 2022].

resulted in a more limited and precise flux distribution pattern. As a result, the integration of proteome data in both models resulted in the large part of the fluxs channel towards pyrimidine and purine metabolism, by means of PRFGS and GLUPRT reactions (Figure 3.19). In the wild-type model, the integration of the metabolic data (without the integration of proteome data) led to a very diverse flux distribution profile, which signifies the importance of the proteome data.

3.2.4 The influence of specific quantitative constraints on the solution space

As discussed above, the results from the perturbation procedure revealed that the solution space in the model of wildtype is considerably more limited than that of the model of mutant. The number of sensitive reactions in the wildtype model was nearly as half as of the mutant model, when constrained with all the experimental data. This can be considered as an informative indicator of the size of the solution space. Furthermore, the analysis of the flux distribution at branching points showed little to no variation in the model of wildtype, Whereas in the model of mutant, especially in the case of the pyruvate branching point, the solution space contained a highly variable results, many of them contradicted the experimental data.
In order to validate that the wildtype model is more accurate in simulating the experimental data as a result of the proclaimed smaller solution space, I compared the production rate of organic acids, obtained from the perturbation approach, to the experimental data in both models. Based on the results, in the model of mutant, only 2.8% of the ethanol production values, 7.9% of the formate and 50% of the acetate production values were within $\pm$ 10% interval of the measured experimental values. Whereas, in the model of wildtype, none of the production values in the flux distribution profiles obtained from the perturbation process were beyond $\pm$ 10% interval of the experimental data. This was no surprise, as the flux distribution analysis at the branching points previously showed that there is no variation at the pyruvate branching point in the wild-type model.

To find out whether it is the constraints on the fermentation products which results in the smaller solution space in the wildtype model, the constraints of the four fermentation products from the mutant model, namely lactate, ethanol, acetate and formate were integrated into the wildtype model. The newly constrained model was used to perform the perturbation procedure and the results revealed that the sensitivity of the model increased to a comparable level to that of the mutant model. The results propose that the more limited and smaller solution space in the model of wildtype is primarily the consequence of the specific constraints on the exchange reactions of the four fermentation products. This is, however, no surprise as these reactions are involved in energy metabolism and the flux ratio at this branching point determines the state of energy production in the system. Considering the fact that maximising the growth associated energy is the primary objective of FBA, the flux distribution at this branching point has an expectedly enormous impact on the solution space of FBA. Overall, the results showed that the functional analysis of the solution space is very informative and necessary in the assessment of the biologically relevant phenotypes. Moreover, it helps to investigate the functional consequences of larger or smaller solution spaces.

3.2.5 Analysing the solution space using CoPE-FBA

The results from the perturbation procedure showed that in a relatively computationally cheap fashion, we would be able to analyse the overall sensitivity in the network, compare the level to which solution space of different models is limited, and also functionally analyse the solution space. To find out whether the results from this approach can be investigated using other techniques, I used CoPE-FBA to analyse the solution space of the aforementioned models. CoPE-FBA is an open source software, trying to estimate the size of the solution space by determining a set of modules of variable reactions which share certain characteristics together. I wanted to see whether the modules created by CoPE-FBA can reflect the difference in the size of the solution space in different models, and whether they can determine the biological reproducibility of each model.
For that matter, I investigated the quantitative impact of different constraints on the modules generated by CoPE-FBA. The modules generated by this software consist of variable reactions whose fluxes are linearly independent of each other, and the combinations of their fluxes give rise to big solution spaces in genome-scale models.
I again used the wildtype and the mutant models of *E. faecalis*, for which various experimental data, namely medium-composition data, metabolic and proteome data was available. Similar to the perturbation approach, I investigated the size of the solution space of the models in a stepwise fashion by generating modules for each model. Here, the higher number of modules with the lower number of reactions in each would imply a smaller solution space. The result is shown in Table 3.2. As the result shows, the integration of proteome data had the most enormous impact on the solution space in the case of both models, while the integration of metabolic

data did so to a much smaller degree.

Table 3.2: The number of reactions in the existing modules in each model when their solution space was analysed with CoPE-FBA. The table is adapted from [Loghmani et al., 2022].

Model name	Number of reactions in each module
mt+nc	399
mt+med	360, 4
mt+med+met	345, 4
mt+med+met+pro	286, 5, 4, 4
wt+nc	400
wt+med	361, 4
wt+med+met	360, 4
wt+med+met+pro	295, 5, 4, 4

When compared to the results from the perturbation procedure, CoPE-FBA could successfully determine the quality of shrinking the solution space by integrating more constraints. However, it could not reflect the extent to which the solution space in each model was limited. The perturbation procedure showed that the integration of metabolic data had a significant impact on the sensitivity of both models both in the general term and the functional term (flux distribution at the branching points). Moreover, the perturbation procedure reported that the solution space of the fully constrained wildtype model is considerably smaller than that of the mutant model, which is also not reflected in the CoPE-FBA results. However, it has to be mentioned that CoPE-FBA could not be performed to its full extent, as the enumeration of individual solutions in the case of these models was computationally infeasible due to the enormously large solution spaces. Therefore, it can be suggested that the perturbation procedure is not only a more efficient platform, but also provides more aspects to the analysis of solution space, such as functional analysis.

3.2.6 The influence of ATP maintenance on the solution space

The maximal growth rate is very often the objective function in use, especially in the case of bacterial genome-scale models, which is formulated as the biomass reaction. The stoichiometry of this reaction accounts for the growth associated ATP. Therefore, while maximizing the growth rate, the model maximizes the growth associated ATP and minimizes the non growth associated ATP in turn. The non-growth associated ATP is often formulated in a generic reaction (ATPM), accounting for all the processes that use ATP, but do not contribute to the growth of the organism. As a result, the boundaries of both reactions are important parameters in the model and contribute to shaping the solution space, as the main criteria in the optimization is maximizing the growth associated ATP (when biomass is the objective function). It has been

shown that ATPM is an important parameter in FBA and has a direct impact on the consistency of the biomass production rate [Dinh et al., 2021].

There have been several studies in which the ATPM value is estimated in an integrative way (using genome-scale models integrated with experimental data) and used to constrain the model in a biologically consistent way [Veith et al., 2015], [Teusink et al., 2006], [Loghmani et al., 2021]. A constrained model with the estimated ATPM value would account for the minimum amount of energy required for non-growth purposes, when maximizing the growth associated ATP. To find out whether the value of ATPM has a direct impact on the solution space, the wildtype model was subjected to a flux scan over the whole feasible range of ATPM reaction and the solution space was calculated using CoPE-FBA accordingly (Figure 3.20).

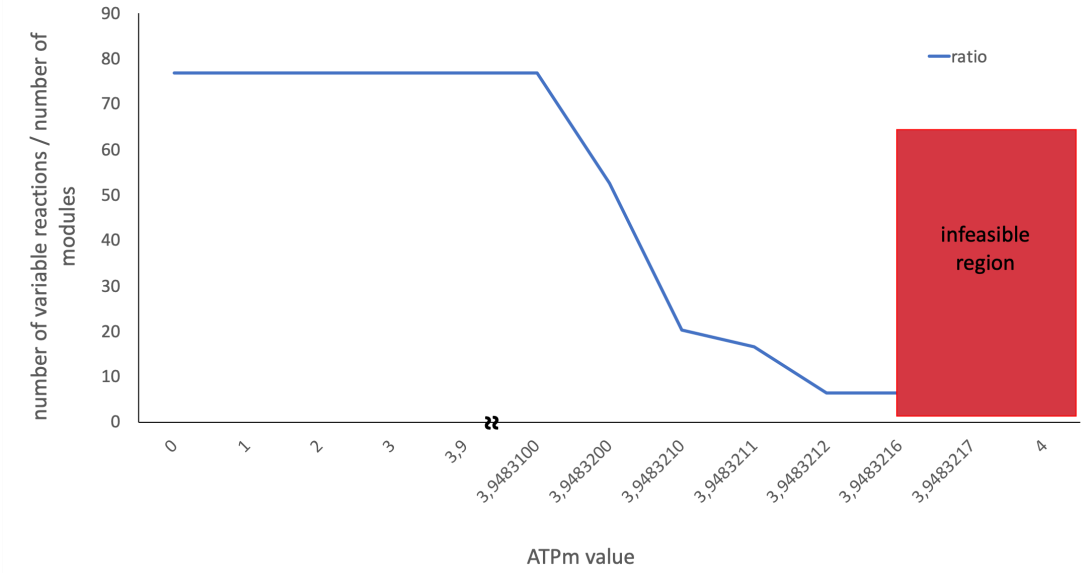

Figure 3.20: Flux scan of the ATPm value over the feasible region of the wildtype model. The blue line shows the ratio between the number of variable reactions and the number of modules generated by CoPE-FBA. The figure is adapted from [Loghmani et al., 2022].

The results indicated that the ATPM value did not have a significant impact on the solution space, up to the boundary of the feasible region. The module analysis using CoPE-FBA at this point reported a much smaller solution space (in an implicit way) compared to previous points. However, when calculating the modules, FVA failed to determine the feasible interval for several reactions at this point. This problem is supposedly a consequence of numerical instability at this point. I performed the same flux scan process on the mutant model in order to make sure that this is not a model-specific problem and it resulted in the same kind of artefact. The result proposes that when a parameter (a flux boundary in this case) is adjusted at the point which is in line with the boundary of the feasible region (possibly several digits after the decimal), FVA fails to calculate the feasible interval for several reactions. In the case of this study, this failure and the resulting artefact led to the calculation of a false smaller solution space.

To find out whether this artefact exclusively exists in the case of ATP/energetically-involved reactions, the same flux scan was performed on two other reactions: 1. lactate dehydrogenase, as an example of an energetically important reaction which does not include ATP in its stoichiometry, and 2. ribulose 5-phosphate 3-epimerase as an example of the reactions with non related energetic importance. The results showed that the same phenomenon happened in the case of both reactions, although two different extents. All in all, the results suggest that as a universal characteristic of genome-scale models, applying constraints up to to the edge of the feasible region in a reaction would result in numerical instabilities, which further leads to some artefacts in downstream analyses. Especial considerations should be taken in the case of the constraints whose values are obtained from the optimization process, such as ATPM, as these values usually contain several digits after the decimal.

It has to be mentioned that the problem of having numerical artefacts was first found in the originally published genome-scale metabolic model of the wildtype *E. faecalis* when constrained with the experimental data and the optimality tolerance of FVA was set to 100% [Großeholz et al., 2016]. The ATPM value in this case was optimized in a way that it was right at the edge of the feasible region. Therefore, I investigated two different ways of avoiding the problem of having numerical artefacts. The first way to avoid this problem is the one that was also used in the original study [Großeholz et al., 2016], which is decreasing the optimality tolerance of FVA from 100% to 99.9%. This approach results in a higher number of variable reactions and potentially a larger feasible interval for several reactions. Another approach would be to change the flux boundary of one or few reactions to a small extent, so the model would not be at the edge of feasibility any more. Therefore, in this case, I decreased the lower boundary of the formate exchange reaction by 0.05 unit, allowing the constraint on the ATPM reaction to remain at the optimized value. This is the approach that I took throughout this study. In order to find out how these different approaches of avoiding artefacts (as a result of the model being at the edge of feasibility) would affect the analysis of the solution space as well as the biologically relevant results, I compared the results from these two different approaches using perturbation procedure and CoPE-FBA. The results are shown in Table 3.2. I found that when the optimality tolerance is reduced to 99.9%, a higher number of variable reactions was reported by FVA for both models (the model with original parametrisation and the adjusted model). The extent of difference however is dissimilar between the two and is more dramatic in the case of the model at the edge of feasibility. Of course this is no surprise as I previously mentioned that FVA is not able to find the feasible interval for several reactions due to numerical instabilities. The comparison of the number of sensitive reactions in each model also showed the same trend. However, the difference in this case was less dramatic. Regarding the CoPE-FBA results, as discussed above, the solution space of the model is considerably smaller when the model is at the edge of feasibility, which is clearly an artefact. When the optimality tolerance is reduced to 99.9% or the model is adjusted for not being at the edge, the calculated solution space by CoPE-FBA is similar between the two models and is relatively large. Moreover, I analysed the impact of reducing the optimality tolerance of FVA on the flux distribution at the branching

points in metabolism and I found no significant impact on the biologically relevant results in this case. The results can be found in the supplement

Table 3.3: Number of variable reactions according to FVA, number of sensitive reactions according to the solution space inspection 530 procedure (perturbation analysis), and the number of reactions in existing modules in each model (CoPE-FBA). All three methods 531 were used with two optimality tolerance value (100% and 99.9%) and the respective results are compared. The table is adapted from [Loghmani et al., 2022].

	Wt+med+met+pro-edge		Wt+med+met+pro	
optimality tolerance	100	99.9	100	99.9
FVA	209	387	307	387
#reactions, sensitive to perturbation	87	137	133	147
#modules according to CoPE-FBA	4, 13, 7, 5, 4, 4, 12, 3	295,5,4,4	295,5,4,4	295,5,4,4

3.2.7 Validating the results using models of other species

In order to find out whether the results of the above-mentioned analyses, namely the functional analysis of the solution space and its following conclusions are general characteristics of all genome-scale models or rather limited to the genome-scale model of *E. faecalis*, I performed parts of the analyses on the example of two other genome-scale models, *S. pyogenes* [Levering et al., 2016] and *L. lactis* [Flahaut et al., 2013]. In the case of *S. pyogenes*, the previously published genome-scale model together with the experimental data including medium composition and metabolites uptake/production rates were used. In the absence of proteome data, two sets of artificially generated data were used to constrain the model and study the resulting impacts on the solution space. Similar to the case of *E. faeclis*, the data were integrated into the genome-scale metabolic model in a stepwise manner. The two artificially generated sets of proteome data were integrated into the model separately. Each set comprised 11 reactions, with one set containing 5 variable and 6 stable reactions, and the other containing 1 variable and 10 stable reactions. The definition of variable and stable was done based on the original model before the integration of proteome data. Similar to *E. faecalis* and expectedly, the integration of data at each step resulted in shrinking the solution space, reflected in the lower number of variable and sensitive reactions. Regarding the artificially generated proteome data, both data sets resulted in reducing the size of the solution space, although the extent of impact was different. The integration of the data set containing a higher number of variable reactions had a more pronounced impact on shrinking the solution space compared to the integration of the other data set. It has to be pointed out again that the integration of proteome data here means the deactivation of the respective reactions in the model.

In order to ensure that the inactivation of variable reactions, resulting in the increase in the proportion of stable reactions in the model, has a reproducible effect on the decreasing the number of sensitive reactions (and therefore shrinking the solution space), again, two sets of

artificially generated proteome data were used to constrain the genome-scale metabolic model of *L. lactis* [Flahaut et al., 2013]. Here, one data set comprised only 15 variable reactions and the other contained only 15 stable reactions. The integration of the first data set decreased the number of sensitive reactions by 44, whereas the integration of the second data set did so by only 3 reactions. Therefore, it can be proposed that the inactivation of reactions such that the proportion of variable reactions in the model decreases (deactivating more variable than stable reactions), results in a more robust behaviour and consequently a smaller solution space. The deactivation of stable reactions does not necessarily have the similar effect.

3.3 Integrative tract-specific drug target identification in *E. faecalis* and *S. pyogenes*

In the previous chapters, I studied different aspects of metabolic behaviour in *E. faecalis*, through studying the impact of glutamine auxotrophy and comparing it to those of the wildtype. It became clear that *E. faecalis* has a very adaptable metabolism, even under suboptimal conditions such as uncontrolled amino acid uptake, lower pH level, etc. Afterwards, I presented the results from the method I developed to investigate the robustness/sensitivity of genome-scale models. The results showed that genome-scale models are powerful tools to capture actual physiological conditions, of course with many limitations. However, I showed that by integrating different types of experimental data, one can significantly increase the predictability of genome-scale models. Therefore, to reach the ultimate goal of this project, which is finding tract-specific drug targets in the metbolic network of *E. faecalis* and *S. pyogenes*, a comprehensive experimental set up was designed to acquire different types of omics data to constrain the respective genome-scale models. In the rest of the results chapter I introduce the results from the experimental and computational parts of this project. It is noteworthy to mention that I performed the bacterial cell cultures in Rostock Medical University under the supervision of PD Dr. Tomas Fiedler. The metabolic profiles were analysed in the same group by Eric Zitsow, and the proteomics experiments were performed in University of Tartu, Tartu, Estonia.

In order to find tract-specific drug targets in the metabolic network of *E. faecalis* and *S. pyogenes*, a comprehensive study was designed to acquire multi-omics data from the physiological conditions comparable to the natural conditions in human body. The acquired data were then integrated into the genome-scale metabolic model of the respective organism to develop tract-specific models and finally suggest drug targets. Therefore, two tract-specific conditions were used for each of the bacteria and the data were compared to those of the standard lab conditions (CDM-LAB). For *E. faecalis*, artificial saliva and artificial urine were used to mimic the physiological condition during root canal infection and urinary tract infection. For artificial saliva, a second condition containing 11 g/l of glucose (similar to the glucose content in CDM-LAB) was used as well, representing a high sugar condition in the mouth. For artificial urine, two different sugar contents were used: 1. 1 g/l of glucose, representing diabetic condition, and 2. 0.7 g/l

of fructose + 0.3 g/l sucrose, representing healthy condition. The standard lab condition for all of the above treatments was CDM-LAB at pH 6.5. For *S. pyogenes*, the exact same conditions for artificial saliva as for *E. faecalis* were used. Moreover, natural human plasma was used to mimic the physiological condition during blood stream infection caused by *S. pyogenes*. For artificial saliva, CDM-LAB at pH 6.5 was used as the control condition, whereas for plasma, CDM-LAB at pH 7.4 was the standard lab condition.
To acquire multi-omics data, samples were used for quantitative metabolic, transcriptomic and proteomic analysis. The transcritome data analysis was published in my master thesis [Loghmani, 2020], but the data was used for the tract-specific modelling in this thesis. The metabolic and proteomic data were analysed and used for the modelling and will be presented in this thesis. In the following, I will present the results of the experiments, metabolic and proteomic data analysis, and finally, the integrative drug target identification using the two genome-scale metabolic models.

3.3.1 Cell culture and growth rates

First, the growth capabilities of the *E. faecalis* and *S. pyogenes* wild type strains in media resembling environments under infection conditions in comparison to CDM-LAB should be analysed. For that purpose, I cultured *E. faecalis* in batch in 50 ml CDM-LAB, artificial saliva, and artificial urine, and *S. pyogenes* in CDM-LAB, artificial saliva and natural human plasma. I grew the bacteria in the media for six hours, and took samples at each hour for the measurement of the ODs. Figure 3.21 shows the growth curves of both species in different media. Both bacteria had the highest growth rate in CDM-LAB, with the exponential growth between the first and the fourth hours after the start of the experiment. Although to a lesser extent, both bacteria also grew exponentially in artificial saliva, *E. faecalis* from time point zero until two hours later and *S. pyogenes* from time point zero until three hours later. Surprisingly, *E. faecalis* did not grow in artificial urine, most likely due to the lack of a proper carbon source.
To assess whether an additional carbon source increases the growth rate of the two bacterial species in the artificial media, I repeated the growth experiments using artificial saliva and artificial urine supplemented with glucose at the same concentration as in CDM-LAB (11 g/l). The results showed that the addition of glucose to artificial urine enables *E. faecalis* to reach an OD level comparable to that in CDM-LAB (Figure 3.22). As the recipe of artificial urine contains none of the primary carbon sources, the result suggests that *E. faecalis* fails to use alternative carbon sources in this medium and does not grow on the original artificial urine as a result. Interestingly, while the addition of glucose to artificial saliva increased the growth rate of *E. faecalis* compared to the original condition (artificial saliva with no additional glucose), it did not have a considerable impact on the growth rate of *S. pyogenes*. The additional glucose enabled *E. faecalis* to growth at a higher rate compared to the original condition. The OD of the *E. faecalis* culture three hours after the start of the incubation with additional glucose was 0.52, which is approximately twice as high as the OD in the original medium that was 0.25. On the other hand, the growth of *S. pyogenes* did not remarkably change by the addition of glucose

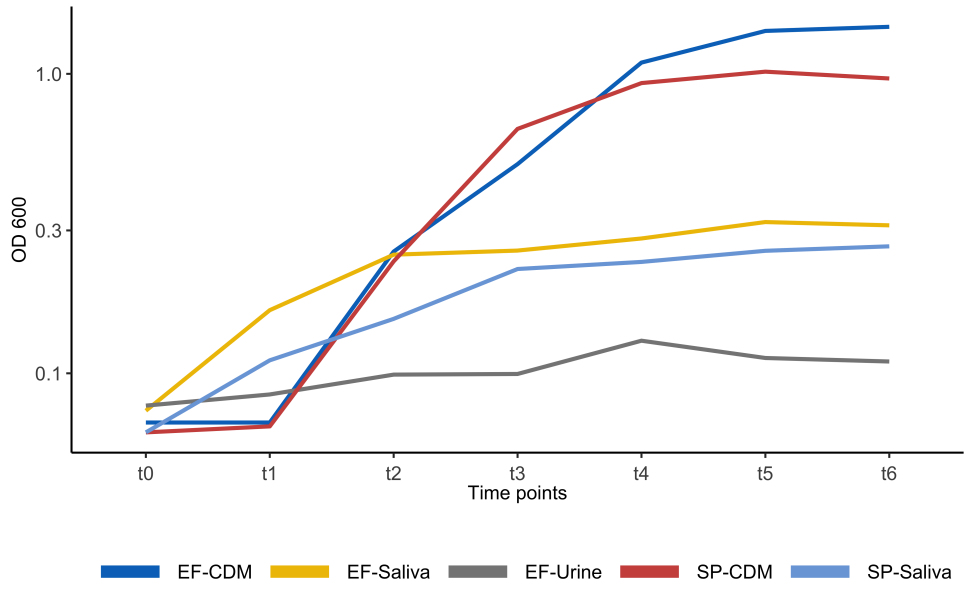

Figure 3.21: The growth curve of *E. faecalis* and *S. pyogenes* growing on CDM-LAB, artificial saliva and artificial urine (only *E. faecalis*). Despite the faster growth on artificial saliva at early time points, both bacteria reached a higher OD level after on CDM-LAB at the end of the time course. The growth of *E. faecalis* on artificial urine was very poor. EF-CDM: *E. faecalis* growing in CDM-LAB; EF-Saliva: *E. faecalis* growing in artificial saliva; EF-Urine: *E. faecalis* growing in artificial Urine; SP-CDM: *S. pyogenes* growing in CDM-LAB; SP-Saliva: *S. pyogenes* growing in artificial saliva.

to the medium. Comparing the OD values of the two conditions, *S. pyogenes* started to grow exponentially one hour later in the presence of additional glucose, however, the OD values of the two conditions were comparable during the exponential phase. At the stationary phase, the OD value of the original medium was 0.26, while it was 0.31 in the other medium. The growth curves in the glucose supplemented media are shown in the figure 3.22.

In order to investigate the obtainable amount of dry mass of the bacteria growing in the different media, I performed an experiment to measure the dry mass at different OD values. Both bacteria were grown in CDM-LAB and artificial saliva, and *E. faecalis* was grown in artificial urine as well. Both artificial media were supplemented with additional glucose. Initially, each treatment was performed with two technical replicates in a culture volume of 100 ml, with the starting dilution of 1.5% and samples were taken at three hours, followed by four and a half and six hours after the start of the incubation. At each time point, samples were taken in 20 ml. After measuring the OD, pellets were dried, and the respective weight was measured. However, the cell amount was too low to reliably measure the dry weight. To increase the reliability of the measurements, the culture volume was increased to 400 ml and samples of 200 ml were taken after one and four hours for all the treatments. The samples were spun down after measuring the OD and the respective dry mass values were calculated in gram per 200 ml, and extrapolated for gram per 1 Litre. The measurement of the dry mass values suggested that an acceptable measurement of biomass for all the treatments seems to be possible.

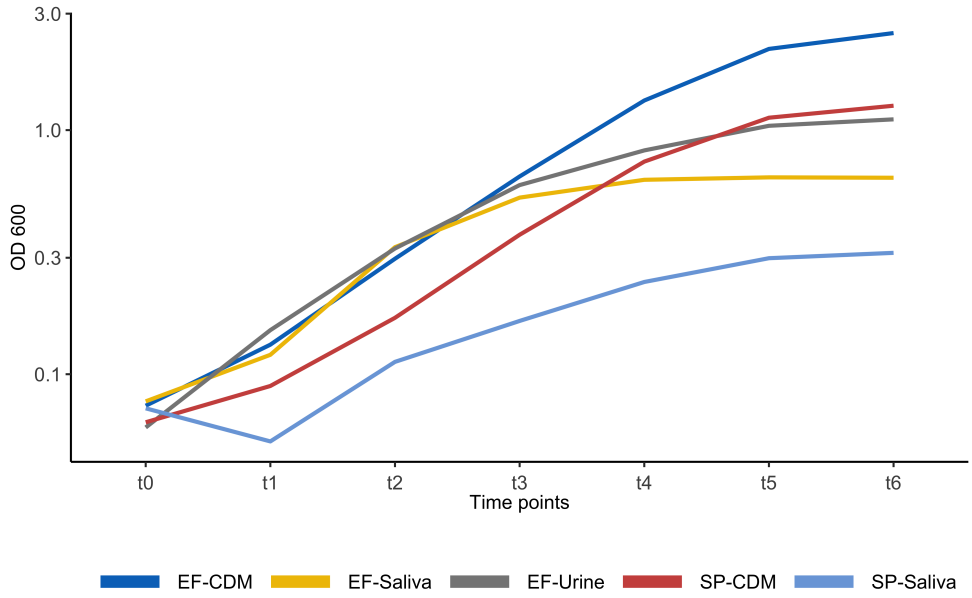

Figure 3.22: The growth curve of *E. faecalis* and *S. pyogenes* growing on CDM-LAB, artificial saliva with additional glucose, and artificial urine with additional glucose (only *E. faecalis*). The addition of glucose resulted in a considerable in increase in the growth of the bacteria on artificial media. EF-CDM: *E. faecalis* growing in CDM-LAB; EF-Saliva: *E. faecalis* growing in artificial saliva; EF-Urine: *E. faecalis* growing in artificial Urine; SP-CDM: *S. pyogenes* growing in CDM-LAB; SP-Saliva: *S. pyogenes* growing in artificial saliva.

3.3.2 Correlation of OD and dry mass

Next, in order to see whether the dry mass to OD ratio is independent of the medium, I grew *E. faecalis* and *S. pyogenes* in CDM-LAB and artificial saliva. If the dry mass to OD ratio is independent of the medium, the dry mass can be calculated from the OD value using a calibration curve as obtained from this experiment. This was an essential step in the project since neither of the bacteria grows well enough in the artificial media to acquire enough palettes for reliable dry mass measurement as well as omics experiments (transcriptomics and proteomics) from the same batch. To overcome this limit, I tried to create calibration curves to estimate the dry mass based on the OD values for each bacteria and each condition. Using a calibration curve, there was no need to collect bacterial pellets for the dry mass measurements, but they could rather be used for transcriptomics and proteomics. The experiment was designed for two biological replicates, each having one technical replicate. I grew both species in 500 ml of each medium until a high OD was reached. For that, I measured the OD once an hour to make sure that a proper OD has been reached. Afterwards, a 200 ml sample was taken, its OD was measured, and the sample was kept for dry mass measurement. Another 200 ml sample was taken and diluted with 200 ml of fresh medium, out of that, a 200 ml sample was taken for OD and dry mass measurement. The rest of the diluted sample was further diluted with 200 ml fresh medium, and 200 ml of that was taken for OD and dry mass measurement. Diluting the sample twice with the fresh medium supposedly decreases the OD to its half at each step. If the dry mass value

of each diluted sample decreases proportionally to its OD in both media, one could conclude that the dry mass to OD ratio is independent of the medium. Figure 3.23 shows the result of the calibration curve for *E. faecalis* and *S. pyogenes* growing in CDM-LAB and artificial saliva. Based on the figure, the dry mass to OD is nicely proportioned for *E. faecalis* in both media. Despite the small differences between the values in the two media, the overall slope of the curve is approximately the same, fulfilling the initial expectation. In the case of *S. pyogenes*, the values are not as nicely proportioned as for *E. faecalis*. However, both numerical discrepancies in CDM-LAB and artificial saliva have occurred of lower OD values, which might be the result of measurement errors, either in the case of OD or dry mass. Hence, it was assumed that the dry mass value of *S. pyogenes* can be estimated from the OD value independent of the medium similar to the case of *E. faecalis*, based on the similarity of the overall shape of the curves in the two media and the respective slopes.

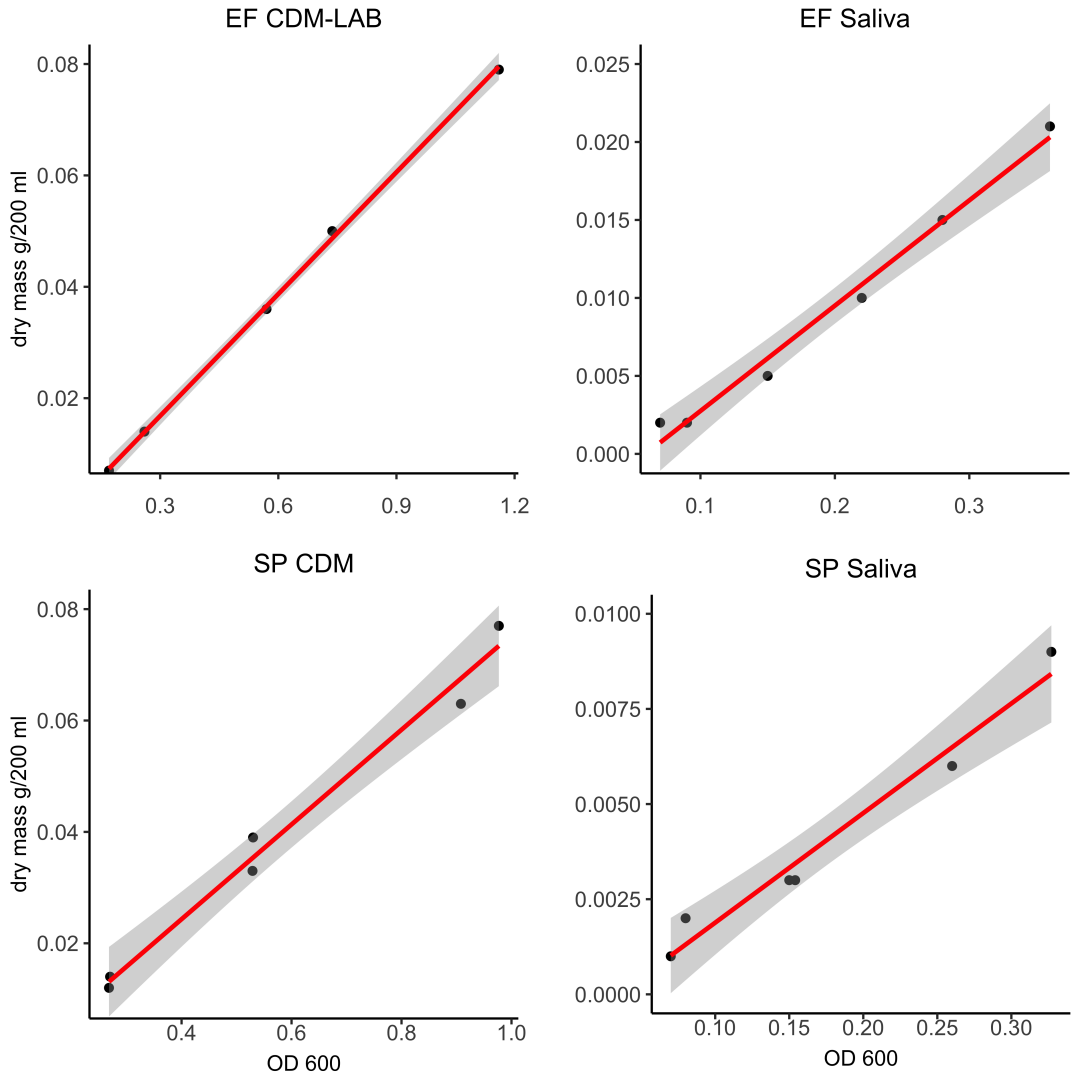

Figure 3.23: Th calibration curve derived from the results of the experiments to estimate the dry mass based on OD for *E. faecalis* and *S. pyogenes* on CDM-LAB and artificial saliva. EF CDM-LAB: *E. faecalis* growing in CDM-LAB; EF Saliva: *E. faecalis* growing in artificial saliva; SP CDM: *S. pyogenes* growing in CDM-LAB; SP Saliva: *S. pyogenes* growing in artificial saliva.

In order to analyse the growth capabilities of *E. faecalis* and *S. pyogenes* in human plasma, I grew both species in two technical replicates in citrated human plasma (pooled from three donors), and the respective OD was measured once an hour. Figure 3.24 shows the time course data. While *S. pyogenes* grew well until six hours after the start of the incubation, *E. faecalis* barely grew at all. All cultures were inoculated with an initial OD value of 0.3. After increasing to 0.4 after one hour, *E. faecalis* barely reached an OD of 0.5 even after six hours, suggesting that the medium is not suitable for its growth. On the other hand, *S. pyogenes* continued to grow until six hours after the start of the incubation and reached a final OD of approximately 1.4. Based on the preliminary experiment's results, we decided to use only *S. pyogenes* for further experiments in human plasma. Due to the shortage of available human plasma, and the very promising growth of *S. pyogenes* in it, a medium size of 100 ml was chosen for the main experiment, with samples being taken four and five hours after the start of the incubation.

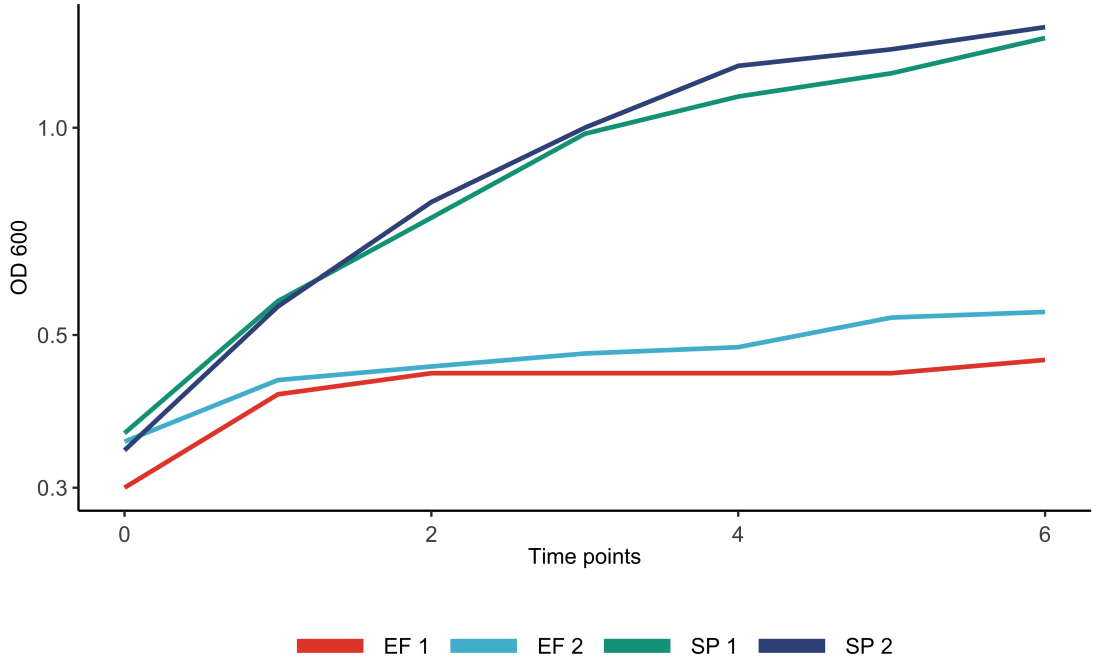

Figure 3.24: The growth curve of *E. faecalis* and *S. pyogenes* when growing on natural human plasma. While *S. pyogenes* grew very fast, *E. faecalis* hardly showed any growth after the first hour. EF 1: *E. faecalis* culture replicate 1; EF 2: *E. faecalis* culture replicate 2; SP 1: *S. pyogenes* culture replicate 1; SP 2: *S. pyogenes* culture replicate 2.

3.3.3 Main experiments and sample collection

To measure the growth rate of the two bacteria in different media and collect samples for the omics studies, I performed batch culture experiments in three biological replicates per medium, each having one technical replicate. For the growth of *E. faecalis* in CDM-LAB, artificial saliva and artificial urine, as well as *S. pyogenes* in CDM-LAB and artificial saliva, 500 ml cultures

were prepared, and samples were taken at one hour after inoculation for metabolic analysis and at two hours after inoculation for metabolic, transcriptomic and proteomic analysis. For the growth of *S. pyogenes* in natural human plasma, cultures of 100 ml were used, and samples were taken after four hours of growth for metabolic analysis, and after five hours of growth for metabolic, transcriptomic, and proteomic analysis. *E. faecalis* was used to grow on CDM-LAB at pH 6.5, artificial saliva, artificial saliva with additional glucose (saliva+glc), artificial urine with additional glucose (urine+glc) and artificial urine with additional sucrose and fructose (urine+fru+suc). *S. pyogenes* was used to grow on CDM-LAB at pH 6.5 and 7.4, artificial saliva and artificial saliva with additional glucose, and natural human plasma. Interestingly, *E. faecalis* grew faster in two hours in artificial saliva than in CDM-LAB, which had been observed also in the test experiments. Although *E. faecalis* in a more extended period (e.g., 6 hours) reaches a higher OD in CDM-LAB, it does not grow as fast as in artificial saliva at initial time points. Starting at an OD of approximately 0.14, artificial saliva facilitates the growth of *E. faecalis* to the OD of 0.24 after two hours, while CDM-LAB enables the growth until the OD of 0.17 - 0.19. The addition of glucose to artificial saliva did not change the growth capability of *E. faecalis* remarkably, as it grows from 0.15 to 0.25. The addition of glucose to artificial urine had a more pronounced impact on the growth of *E. faecalis*. According to the experiments, it grew on artificial urine with additional glucose very fast, and the OD increased from 0.13 to 0.26 within 2 hours. It grew even faster on urine with additional sucrose and fructose, as the OD increased from 0.13 to 0.33 within 2 hours. In general, while the growth capability of *E. faecalis* on CDM-LAB is higher in a more prolonged period, artificial media facilitate a faster growth at early time points.

Table 3.4: The OD value of the *E. faecalis* from the three biological replicates when growing on CDM-LAB at pH 6.5, artificial saliva, artificial saliva + glucose, artificial urine + fructose and sucrose, artificial urine + glucose.

Medium	Replicate	Initial OD	1st h OD	2nd h OD
CDM-LAB pH 6.5	1	0,14	0,17	0,19
	2	0,13	0,148	0,17
	3	0,13	0,15	0,176
Artificial Saliva	1	0,14	0,18	0,24
	2	0,12	0,17	0,24
	3	0,17	0,2	0,24
Artificial Saliva + glc	1	0,15	0,18	0,25
	2	0,18	0,2	0,29
	3	0,15	0,19	0,25
Artificial Urine + glc	1	0,16	0,24	0,37
	2	0,12	0,14	0,2
	3	0,13	0,16	0,26
Artificial Urine + fru + suc	1	0,16	0,21	0,36
	2	0,11	0,16	0,27
	3	0,13	0,19	0,33

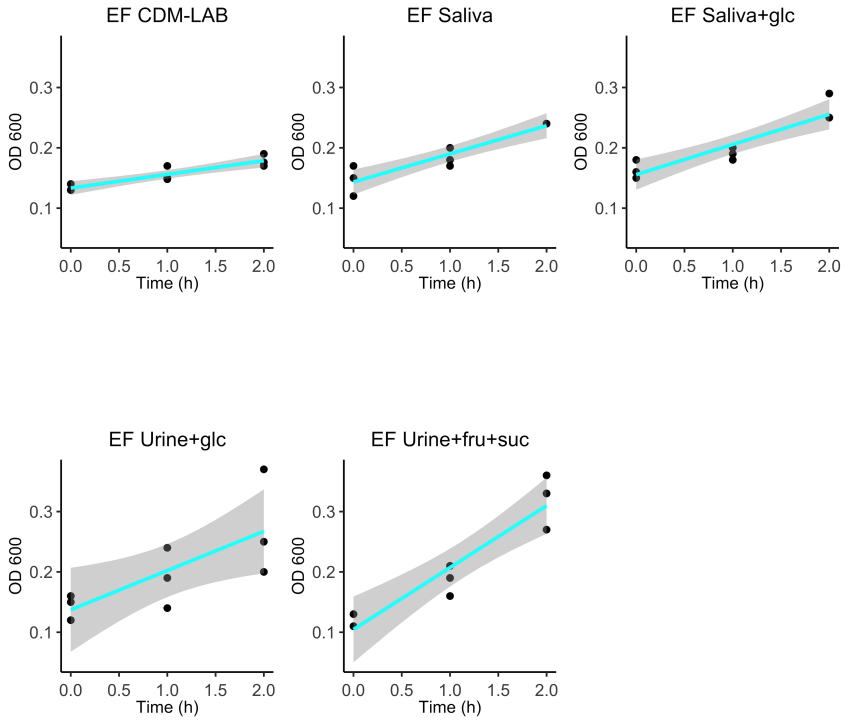

Figure 3.25: The growth curve of *E. faecalis* when grew on different media. Each experiment contained three biological replicates. EF CDM-LAB: *E. faecalis* growing in CDM-LAB; EF Saliva: *E. faecalis* growing in artificial saliva; EF Saliva+glc: *E. faecalis* growing in artificial saliva+glc; EF Urine+glc: *E. faecalis* growing in artificial urine+glc; EF Urine+fru+suc: *E. faecalis* growing in artificial urine+fru+suc.

I grew *S. pyogenes* in CDM-LAB at pH 6.5 and 7.4, artificial saliva, artificial saliva with additional glucose and natural human plasma. Based on the experiments, CDM-LAB at pH 7.4 facilitates a faster growth compared to pH 6.5. While at pH 6.5 the OD of the culture increased from 0.11 to 0.14, at pH 7.4 the OD had a more noticeable increase, starting from 0.1 going to 0.15-0.16 within 2 hours. This is no surprise as it is known that lactic acid bacteria have their fastest growth at pH 7 to 7.5. The growth rate of *S. pyogenes* at artificial saliva was slightly higher than the growth in CDM-LAB at pH 6.5, but the difference between the two was not as large as in the case of *E. faecalis*. However, the difference between the growth rate on original artificial saliva and artificial saliva with additional glucose was more pronounced than in *E. faecalis*. The addition of glucose to artificial saliva enabled *S. pyogenes* to grow faster, and the culture OD increased from 0.12 to 0.17. Similar to the test experiments, the most consistent growth condition for *S. pyogenes* was in natural human plasma. Due to the small size of the culture, the bacteria were incubated for five hours to obtain more bacteria for omics experiments. Starting at the OD value of 0.17, *S. pyogenes* reached the OD of 1.11 after five hours. The growth curve of all cultures of *E. faecalis* and *S. pyogenes* are shown in figure 3.25 and 3.26, respectively. All samples from different cultures were used to analyse the concentration

Table 3.5: The OD value of the *S. pyogenes* from the three biological replicates when growing on CDM-LAB at pH 6.5, CDM-LAB at pH 7.4,artificial saliva, artificial saliva + glucose, natural human plasma.

Medium	Replicate	Initial OD	1st h OD	2nd h OD	4th h OD	5th h OD
CDM-LAB pH 6.5	1	0,11	0,13	0,14		
	2	0,11	0,12	0,136		
	3	0,12	0,13	0,15		
CDM-LAB 7.4	1	0,1	0,12	0,135		
	2	0,1	0,128	0,16		
	3	0,14	0,17	0,22		
Artificial Saliva	1	0,11	0,14	0,15		
	2	0,11	0,13	0,16		
	3	0,09	0,13	0,16		
Artificial Saliva + glc	1	0,15	0,16	0,2		
	2	0,13	0,14	0,16		
	3	0,11	0,14	0,18		
Human Plasma	1	0,16			0,74	1,12
	2	0,17			0,7	1,08
	3	0,17			0,83	1,11

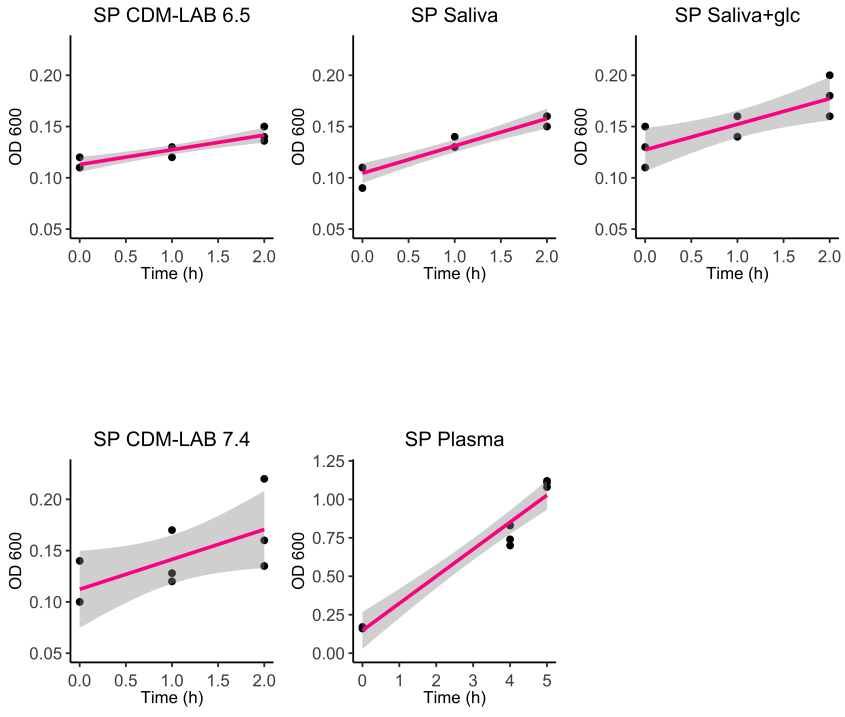

Figure 3.26: The growth curve of *S. pyogenes* when grew on different media. Each experiment contained three biological replicates. SP CDM-LAB 6.5: *S. pyogenes* growing in CDM-LAB at pH 6.5; SP CDM-LAB 7.4: *S. pyogenes* growing in CDM-LAB at pH 7.4; SP Saliva: *S. pyogenes* growing in artificial saliva; SP Saliva+glc: *S. pyogenes* growing in artificial saliva+glc; SP Plasma: *S. pyogenes* growing in natural human plasma.

of extracellular metabolites, namely carbohydrates, organic acids and amino acids. Moreover, transcriptome and proteome analysis were performed for all the samples.

3.3.4 Comparative proteome data analysis of *E. faecalis* and *S. pyogenes*

To investigate the effect of different physiological conditions on the proteome profile of *E. faecalis* and *S. pyogenes*, bacterial pellets obtained from the cell cultures were used for the proteomics study. For *E. faecalis*, pellets from CDM-LAB as well as saliva+glc and urine+glc were used for proteomics. For the *S. pyogenes*, pellets from CDM-LAB at pH 7.4 and plasma were used for proteome analysis. In the following paragraphs I present the results from the comparative proteome data analysis and highlight the notable findings.

To calculate the significant fold changes between the conditions, I prepared the data in a way that suits best for statistical analysis. Therefore, I first transformed all the signal intensities into log2-transformed and then filtered the proteins with missing values across the samples. In the case of the *E. faecalis*, samples from all conditions contained an almost similar number of proteins, whereas in the case of *S. pyogenes*, samples from plasma contained a considerably higher number of proteins compared to the ones from CDM-LAB (Figure 3.27). According to this figure, there were around 500 proteins available in the plasma samples and absent in the ones from CDM-LAB. Considering the fact that plasma medium was natural human plasma, it contained a lot of human proteins as well. Therefore, to be able to analyse the bacterial proteins without having the effect of human proteins on the statistical analysis, I excluded the human proteins and reperformed the analysis using only bacterial proteins. As a result, the number of proteins between the two conditions, CDM-LAB and plasma became almost similar. Additionally, I performed principal component a analysis (PCA) to find out how different samples are grouped together. While *E. faecalis* samples were grouped in a reasonable way, the first CDM sample of *S. pyogenes* was very distant from the other two across the first component. Therefore, I decided to exclude the first sample and performed all the downstream analyses using two samples for CDM-LAB. I also calculated the number of overlaps between the samples for both bacteria and the results showed that a lot of the proteins were identified in all samples.
Next, the data were normalized by variance stabilizing transformation and then checked to see whether missing values are biased to specific samples.The results showed that in both bacteria, there is a bias towards samples and certain proteins are missing in there(figure 3.28). As can be seen in the figure, there are large groups of proteins existing in certain conditions and not in the others. This is even more eye catching in the case of *S. pyogenes*, where there are so many proteins that are available in the CDM-LAB library, but not in the plasma library. Therefore, to see whether this systematic loss of proteins is due to the fact that the signal intensity of those missing proteins are close to the detection limit, the signal intensity as well as the cumulative fraction values were calculated against log2 transformed values. The result showed that the signal intensity of the missing proteins were generally lower than the proteins without the missing values, pointing out to the fact that the missing values are possibly missed as a result of lower

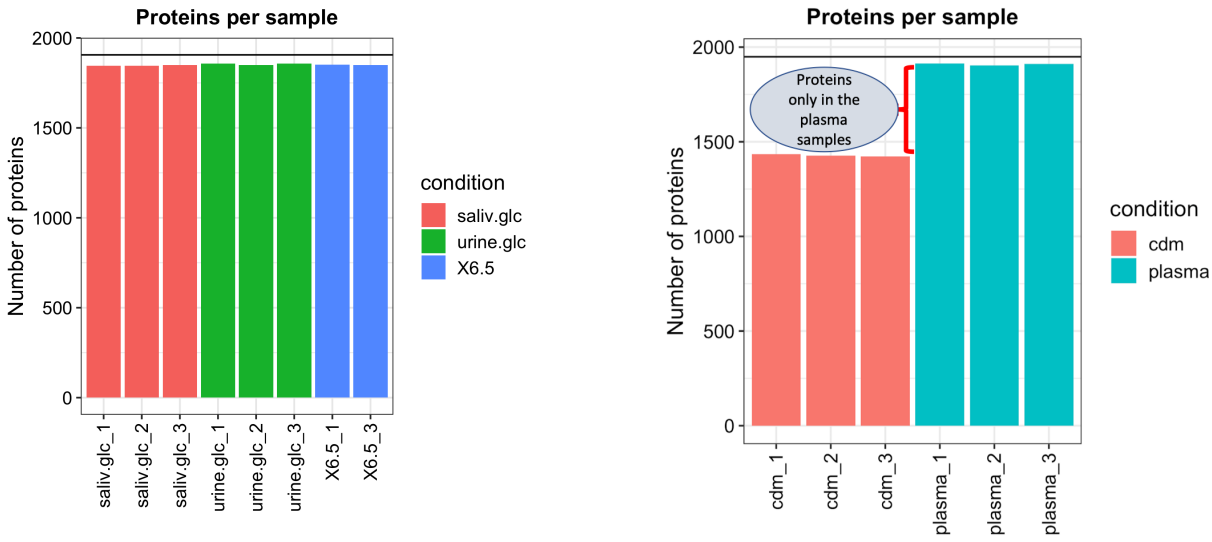

Figure 3.27: The protein content of the each sample; left: *E. faecalis*, right : *S. pyogenes*. While samples of *E. faecalis* contain almost a similar number of proteins, *S. pyogenes* samples in plasma contain almost 500 more proteins compared to CDM-LAB, due to the presence of human proteins in the cultures.

intensities and therefore they had to be imputed.

Finally, to find out about the significant fold changes between the conditions, I used protein-pairwise model together with empirical Bayes statistics. The p-values were adjusted for multiple testing by Benjamini-Hochberg (BH) multiple testing correction. In the case of *E. faecalis*, the significant fold changes between saliva+glc and CDM-LAB as well as urine-glc and CDM-LAB were investigated. For *S. pyogenes*, samples from plasma were compared to those from CDM-LAB at pH 7.4. Principal component analysis (PCA) was used to gain a high level overview over the data. Using PCA to analyse the *E. faecalis samples*, it became cleart that while saliva+glc samples, and to a lower extent, CDM-LAB samples were well grouped together, samples from urine+glc were quiet distant from each other. In the case of *S. pyogenes*, plasma samples were almost perfectly grouped together, whereas the CDM-LAB samples showed some variations. The results from PCA were further validated by the results from the pearson correlation analysis (Figure 3.29). In the case of *E. faecalis*, the pearson correlation analysis shown that the saliva+glc samples are more distanced from the control condition compared to the urine+glc samples. In the case of *S. pyogenes*, plasma samples are perfectly distanced from the control condition. The results from the significant fold change analysis were plotted in volcano plots for all the intended comparisons (figure 3.30 and 3.31).

To find out about the significant changes in protein abundances, I performed pairwise enrichment analysis. For *E. faecalis*, saliva+glc and urine+glc samples were compared to those of CDM-LAB. In saliva+glc, there were 69 proteins that were differentially expressed compared to CDM-LAB. More specifically, seven proteins were involved in quorum sensing (all down-regulated), five in two-component system (all up-regulated) and four in beta-lactam resistance

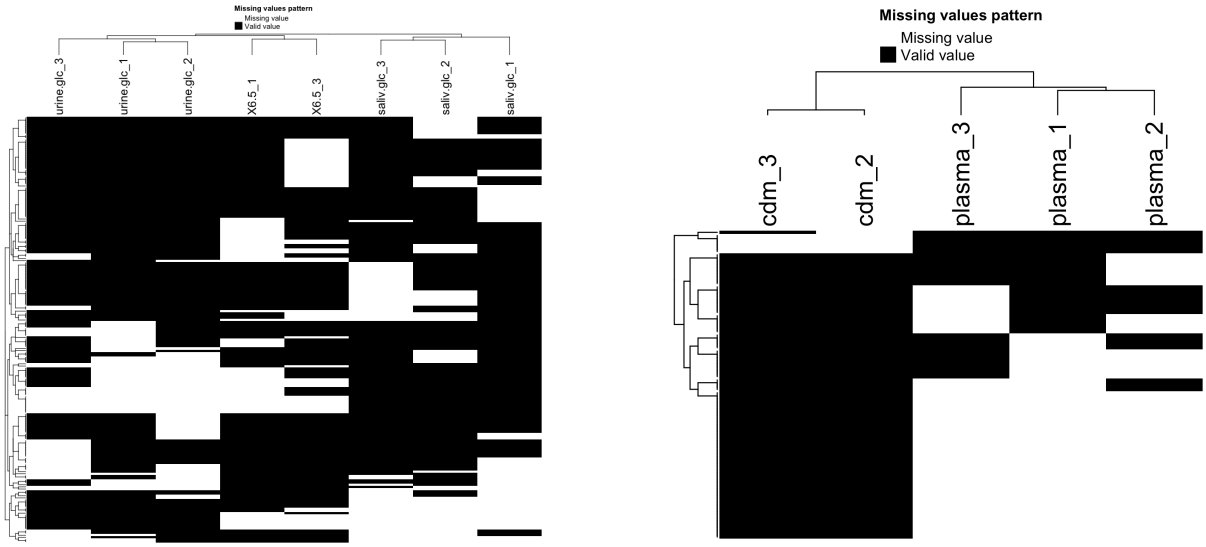

Figure 3.28: The log2-transformed intensity of the normalised and fitted data; left: *E. faecalis*, right : *S. pyogenes*. The figure shows that there is a bias in the missing proteins with respect to samples.

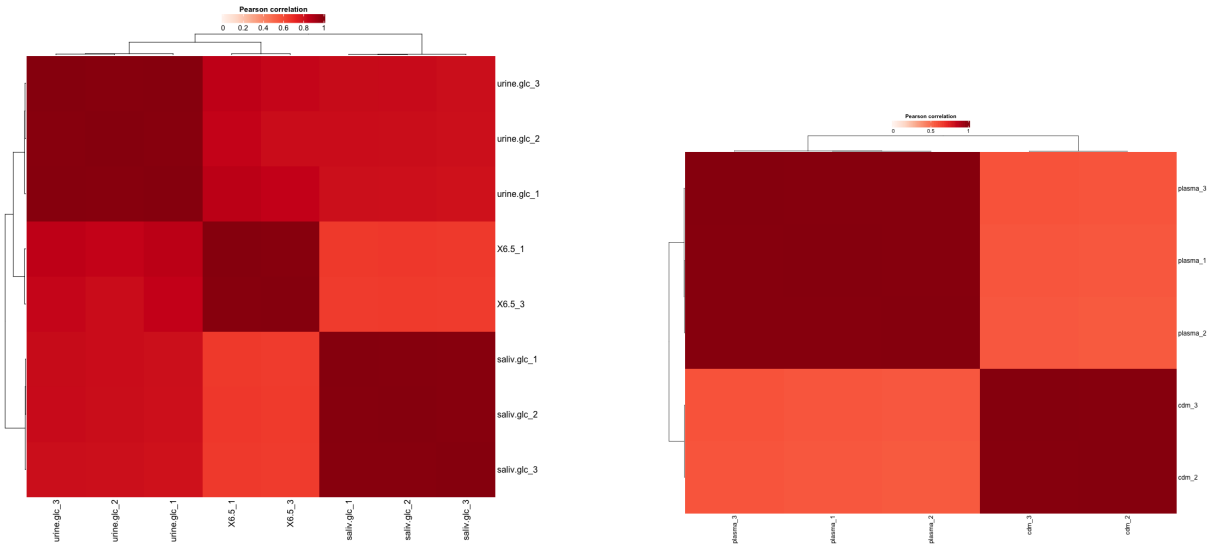

Figure 3.29: Pearson correlation analysis; left: *E. faecalis*, right : *S. pyogenes*. The figure shows that the samples are relatively distant from each other, especially in *S. pyogenes*.

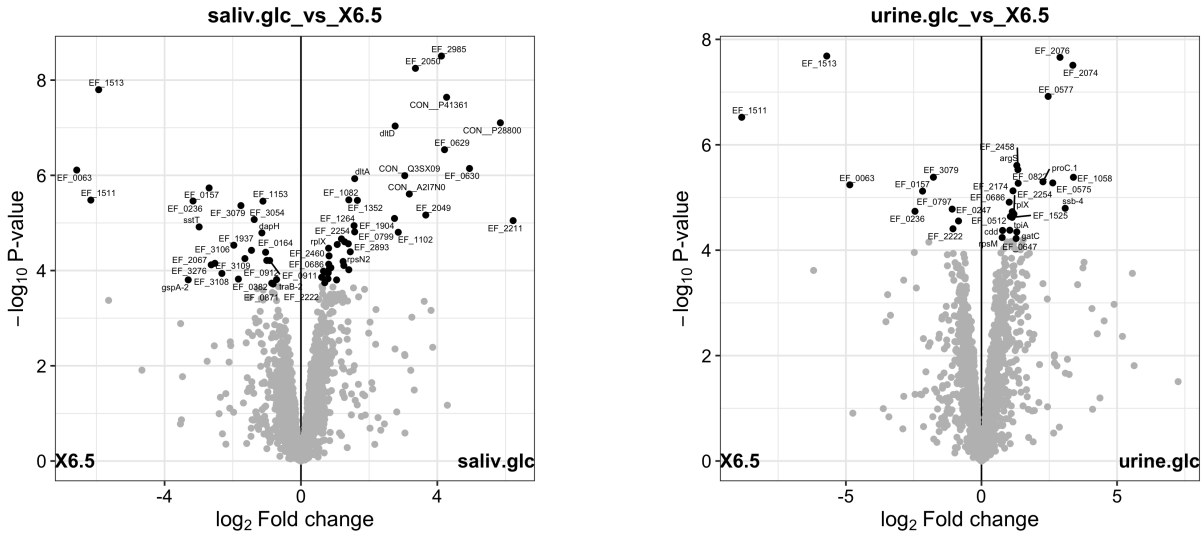

Figure 3.30: Volcano plot showing significantly changed proteins; left: *E. faecalis* on artificial saliva+glc, right : *E. faecalis* on artificial urine+glc.

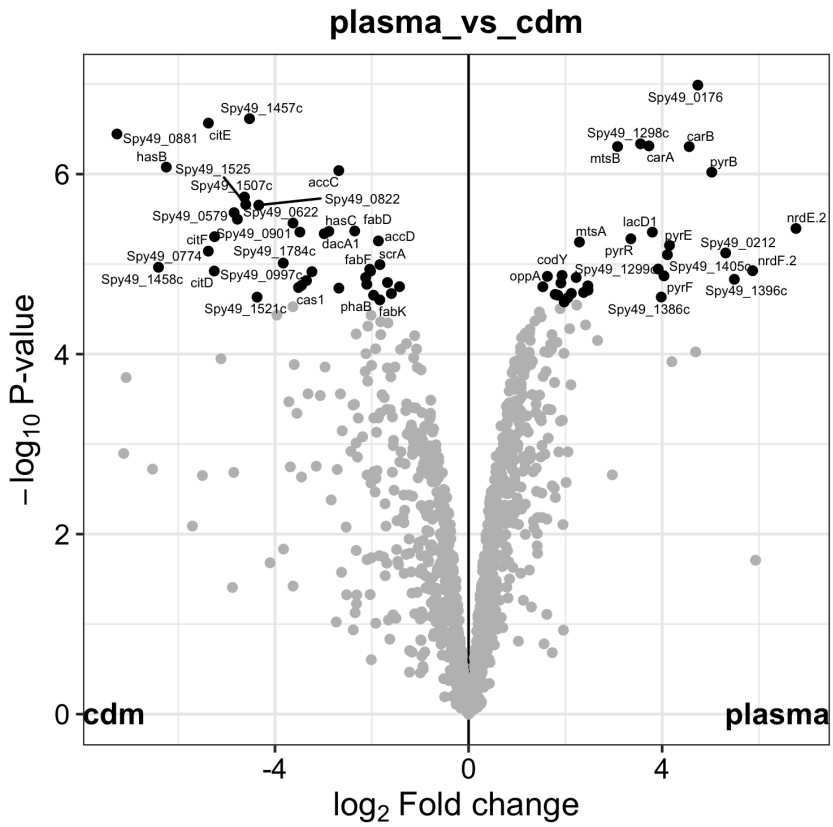

Figure 3.31: Volcano plot showing the significantly changed proteins on natural human plasma in *S. pyogenes*.

(all down-regulated), highlighting the impact of a natural physiological condition on the proteins involved in pathogenic characteristics of *E. faecalis*. Moreover, five of the affected proteins belonged to the family of ABC Transporters, expectedly so, given the difference in the composition of the two media. In urine+glc, there were a total of 30 proteins being differentially expressed compared to CDM-LAB. Eight of the 30 proteins belonged to ABC Transporters, with four up-regulated and four down-regulated. Additionally, two proteins were involved in beta-Lactam resistance (both down-regulated), and two in quorum sensing (both down-regulated). I also performed pairwise significant fold change analysis for *S. pyogenes*. In total, there were 65 proteins being significantly changed between plasma and CDM-LAB. Interestingly, none of the major metabolic pathways were significantly affected by the medium in both species, as there were multiple pathways having one protein being significantly changed between the two media. The complete list of all the significantly expressed proteins can be found in the supplement.

3.3.5 Integrative investigation of tract-specific drug targets

As described previously in the introduction, tract-specific drug targets provide a more accurate targeted therapy which potentially restrain the trend of multi resistance. In order to investigate potential tract-specific drug targets in the metabolic networks of *E. faecalis* and *S. pyogenes*, I used the respective genome-scale metabolic models constrained by metabolic and transcriptome data obtained from batch culture experiments.

The metabolic profiles comprised the uptake and production rate of carbon sources (sugars), organic and amino acids. The extracellular concentration of each metabolite was quantified using HPLC (at Rostock Medical University) at each time point as well as in the fresh media. The values were then used to calculate the uptake/secretion rates with respect to the exponential growth rate of the bacteria under each condition. The calculated flux values were integrated into the respective genome scale models as metabolic constraints. A 40% measurement error was considered and applied to the values when integrated into the models. However, due to the poor quality of the measurements, a lot of the flux boundaries had to be adjusted to values far beyond the error intervals. Unfortunately, this decreased the expected impact from the metabolic constraints, which earlier in this thesis were shown to be quite useful to limit the solution space. Nevertheless, redoing the experiments was beyond the time frame of this project and the data had to be used with a lot of adjustments. Therefore, a conclusive biological discussion of the metabolic profiles of the two bacteria under different circumstances was not possible. I rather used the data only to constrain the models, but having in mind that these are only very rough estimates and therefore only very lose contraints. However, establishing the technology and work flow will later on allow a very fast integration of repeated metabolic profiles. The measured concentration of the metabolites under each condition and the respective calculated uptake/production rate can be found in the supplement.

In order to find out tract-specific drug targets, I first developed tract-specific genome-scale metabolic models using different types of constraints. To do so, I first integrated the metabolic data for CDM-LAB (at pH 6.5 for *E. faecalis* and *S. pyogenes* and 7.4 for *S. pyogenes*) into

the respective genome-scale metabolic models. Afterwards, I calculated the flux intervals for variable reactions using FVA. Moreover, I determined the essential reactions for each bacteria growing on CDM-LAB at the respective pH values. To develop tract-specific models compared to CDM-LAB, I first integrated the respective metabolic data into models and calculated the flux intervals for variable reactions using FVA. Afterwards, I used the transcriptome data and applied the respective fold changes to the flux intervals obtained with FVA from the CDM-LAB models. Whenever necessary, the boundaries were adjusted to the point allowing a feasible solution. Finally, the essential reactions as well as potential tract-specific drug targets were investigated using the approaches that will be described in the next paragraph.

The main idea behind finding tract-specific drug targets in this study is that the integration of tract-specific constraints enables us to simulate the metabolic behaviour of the pathogen while growing in a tract environment. Therefore, the enzymes whose inhibition result in a lower growth rate would be considered as drug targets. To be more specific, I distinguished between enzymes whose inhibition fully inhibits the growth (essential enzymes) and those which only restrain the growth of the bacteria. Although essential genes for all the different conditions were investigated, I did not regard them as drug targets. The primary reason for that was the fact that there is hardly any drug which results in a 100% inhibition of an enzyme. There is always a certain level of an inhibited reaction that takes place, even if the inhibitor is very strong. Hence, the assumption that an essential reaction (whose inhibition results in zero growth rate) fully kills the bacteria is not realistic. To have a more biologically feasible assumption, I used two different approaches. First, I investigated the enzymes whose inhibition would reduce the growth rate to 30%, 20% and 10% of the original rate. This way, the enzymes whose inhibition had more impact on the growth rate would be considered potentially more effective drug targets. In a different approach, I reduced the flux boundaries of all the reactions in the network to 30%, 20% and 10% of the original flux boundaries. The enzymes whose reduced flux interval resulted in zero growth rate would be considered as potential drug targets. With this method, a more effective drug target would be the one that inhibits the growth with lower reduction to its flux boundaries. I also used the two methods in combination, meaning that reducing the flux boundaries to 30%, 20% and 10% of their original values, and those which reduced the growth rate to 30% or lower of the original growth rates would be considered as potential drug targets. Of course, the reactions with lower reduction in their flux interval would be considered as more effective drug targets. An overview of the strategy to find the drug targets and the points at which libraries of drug targets were generated is shown in figure 3.32. Each library contains a set of single reactions whose inhibition (to different extent) affects the growth rate in the model.

I have to point out that despite introducing the calibration curve to estimate the dry mass based on the OD value earlier in this chapter, I decided to use the growth rate as OD/hour. The reason for that was the very unrealistic values for metabolite uptake and production rates, when the estimated gram/hour was used for the calculation of flux rates. Therefore, in the rest of this chapter, all the introduced results were based on the models using OD/hour as the growth rate,

and metabolite uptake and production rates as millimolar/OD/h.

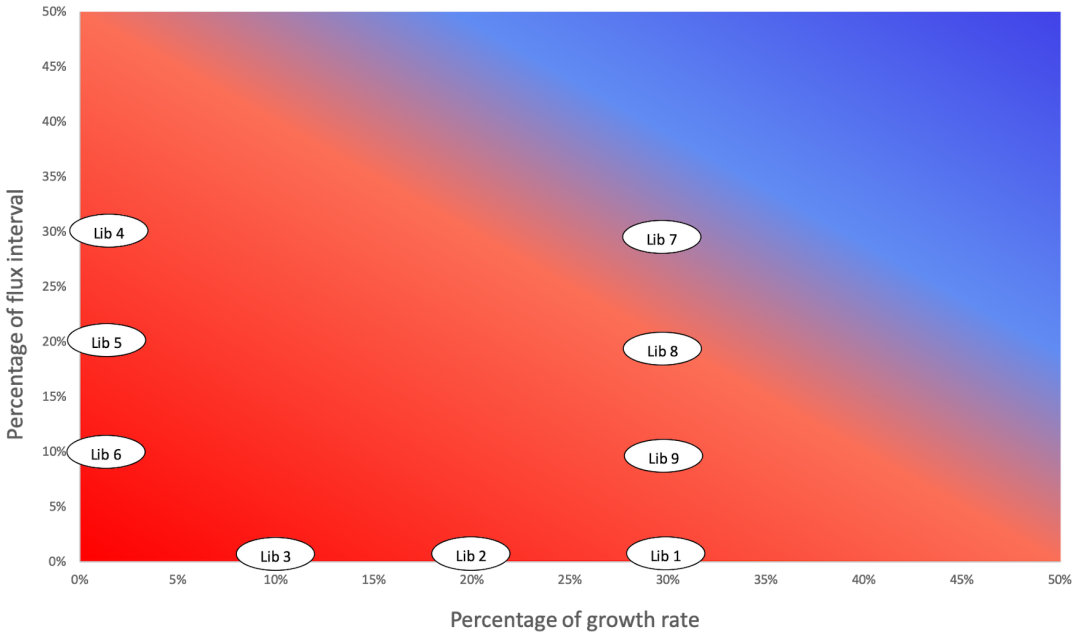

Figure 3.32: Spread of libraries of potential drug targets in the landscape of reaction flux intervals and growth rate. Each library contains a set of reactions whose inhibitions (at the level of single reaction) or flux reduction (at the level of single reaction) reduces the growth rate to different extent. According the assumption used in this study, Lib 7 supposedly suggests a more realistic set of drug targets compared to the others.

For *E. faecalis*, I first investigated the drug targets for the CDM-LAB environment. As mentioned above, the model was developed based on the data from the CDM-LAB. At CDM-LAB at pH 6.5, *E. faecalis* grew at the rate of 0.146 OD/h further adjusted for the model to have a feasible solution. The FVA result of the CDM-LAB model can be found in the supplementary information. A total of 243 essential reactions were found in the CDM-LAB model. The list of essential reactions can be found in the supplementary information. I also calculated the potential drug targets in CDM-LAB using the three above-mentioned approaches. The number of potential drug targets in each library from the different approaches is shown in Table 3.6. Among all the libraries, I consider the Lib 7, which actually contains the drug targets whose flux interval reduction to 30% reduces the growth rate of the bacteria to 30% or lower of its original value, as the most realistic and therefore accurate one.

Table 3.6: The number of drug targets in *E. faecalis* according to each libraries.

	Lib 1	Lib 2	Lib 3	Lib 4	Lib 5	Lib 6	Lib 7	Lib 8	Lib 9
CDM-LAB 6.5	251	249	244	70	72	73	71	238	245
Saliva	289	289	289	61	68	257	63	248	257
Saliva+glc	308	308	308	108	108	287	108	289	293
Urine+glc	282	282	282	71	76	79	72	247	256
Urine+fru+suc	234	232	232	53	60	64	53	221	224

Table 3.7: The number of drug targets in *S. pyogenes* according to each libraries.

	Lib 1	Lib 2	Lib 3	Lib 4	Lib 5	Lib 6	Lib 7	Lib 8	Lib 9
CDM-LAB 6.5	217	217	217	27	29	29	27	196	205
Saliva	220	220	220	39	204	210	40	206	211
Saliva+glc	249	249	249	55	225	231	57	225	235
CDM-LAB 7.4	237	236	236	48	48	51	48	221	226
Human plasma	226	224	224	41	43	216	43	206	218

In the next step, I performed pathway analysis to put the findings in a metabolic context. The result of the pathway analysis for Lib 7 from CDM-LAB at pH 6.5 is shown in Table 3.8. According to the table, 3 of the 71 potential drug targets involved to glycolysis / gluconeogenesis, including lactate dehydrogenase (LDH) and glyceraldehyde-3-phosphate dehydrogenase (GAPDH or GAPD). This is no surprise, as *E. faecalis* significantly rely on glycolysis for energy production and growth. More specifically, LDH in general play a significant role in energy production in *E. faecalis* and other lactic acid bacteria. On the other hand, the role of GAPDH as a target for antimicrobial agents used against gram-positive bacteria has been previously reported [Gómez et al., 2019]; pointing out to the fact that GAPDH is potentially an effective drug target in *E. faecalis*. Morovere, three enzymes were involved in amino acid biosynthesis, including serine dehydrogenase (SERD) and aspartyl-tRNA(Asn):L-glutamine amido-ligase (ADP-forming) (ASNTAL). The third enzyme, GAPDH belongs also to glycolysis. The pathway analysis showed to the important role of central metabolism in the growth of the bacteria, as expected.

Table 3.8: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *E. faecalis* from CDM-LAB.

Pathway	No. Reactions
Metabolic pathways	7
Biosynthesis of secondary metabolites	5
Microbial metabolism in diverse environments	4
Biosynthesis of amino acids	3
Carbon metabolism	3
Cysteine and methionine metabolism	3
Glycolysis / Gluconeogenesis	3
Pyruvate metabolism	2

Searching for the tract-specific drug targets, I investigated the targets in the *E. faecalis* metabolic network when growing on artificial saliva. The growth rate of *E. faecalis* under this condition was equal to 0.263 OD/h. First, I calculated the essential reactions and compared it to the ones from the CDM-LAB condition. I would like to remind that the model for saliva condition is

constrained with metabolic and transcriptome data. There were 296 essential reactions in the model for the saliva condition, which denotes an increase of 53 reactions compared to CDM-LAB. This is translated to 162 essential genes that are necessary for the model to have a feasible solution. The Lib 7 library of the potential drug targets in artificial saliva showed that there are three transporters, ILEt6 (for isoleucine transport), VALt6 (for valine transport) and METabc (for methionine transport), whose flux reduction reduced the growth. Methionine uptake rate was one of the cases in which the flux boundary of the exchange reaction had to be adjusted after the integration of the metabolic data. A more accurate measurement of the methionine uptake rate might bring more insight in this regard. Interestingly, none of the glycolytic enzymes existed in this library, which might be due to the fact that none of the primary carbon sources exist in artificial saliva. However, other enzymes from central metabolism were identified as drug targets such as enzymes involved in amino acid biosynthesis, including SERD (similar to the library from CDM-LAB), L-threonine deaminase (THRD) and cystathionine b-lyase (CYSTL). The results from the pathway analysis is shown in Table 3.9.

Table 3.9: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *E. faecalis* from artificial saliva.

Pathway	No. Reactions
ABC transporters	3
Biosynthesis of amino acids	3
Biosynthesis of secondary metabolites	3
Cysteine and methionine metabolism	3
Metabolic pathways	3
Carbon metabolism	2
Glycine, serine and threonine metabolism	2
Two-component system	2

I also investigated the drug targets in the metabolic network of *E. faecalis* when growing on saliva+glc. *E. faecalis* grew at the rate of 0.249 OD/h under this condition .I found 310 essential reactions and 175 essential genes for the model to have a feasible solution under this condition. Similar to the drug target profile of the CDM-LAB condition, GAPDH, LDH and PGDH from glycolysis and pentose phosphate pathway existed in the Lib 7 of the saliva+glc condition, which is most likely due to the addition of glucose into the medium. This might suggest that under the conditions which are rich in terms of primary carbon sources, enzymes in central carbon metabolism might be a good target for inhibiting the growth of the bacteria. The pathway analysis also showed that there are 7 enzymes involved in the biosynthesis of secondary metabolites, most of them were involved also in central metabolic pathways such as glycolysis or amino acid biosynthesis, including, LDH, GAPDH, PGDH, CYSTL and cystathionine g-lyase (substrate homoserine) (CYSTH). The results from the pathway analysis is shown in Table 3.10.

Table 3.10: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *E. faecalis* from artificial saliva+glc.

Pathway	No. Reactions
Metabolic pathways	10
Biosynthesis of secondary metabolites	7
Microbial metabolism in diverse environments	6
Biosynthesis of amino acids	5
ABC transporters	4
Carbon metabolism	3
Cysteine and methionine metabolism	3
Glycolysis / Gluconeogenesis	3

In Urine-glc, *E. faecalis* grew at the rate of 0.34 OD/h. The model integrated with the data from urine+glc comprised 282 and 170 essential reactions and genes, respectively. The addition of glucose to artificial urine did not have the same effect as it had on artificial saliva, in the sense that none of the glycolytic enzymes were identified as a potential drug target in the Lib 7 library of urine+glc condition; only one enzyme from pentose phosphate pathway, PGDH existed in the library. On the other hand, there were several enzymes such as SERD, CYSTL and ornithine carbamoyltransferase (OCBT) belonging to the biosynthesis of secondary metabolites and amino acids. The latter case, together with carbamate kinase (CK), are involved in arginine biosynthesis, which might suggest the important role of arginine metabolism under this condition. The results from the pathway analysis is shown in Table 3.11.

Table 3.11: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *E. faecalis* from artificial urine+glc.

Pathway	No. Reactions
Metabolic pathways	7
Biosynthesis of secondary metabolites	4
Biosynthesis of amino acids	3
Carbon metabolism	3

Interestingly, when fructose and sucrose were used as carbon sources, *E. faecalis* grew at a higher rate, 0.44 OD/h, compared to the previous condition. Under this condition, the model comprised 235 essential reactions and 126 essential genes. In the Lib 7 library of potential drug targets, LDH was one of the eye identified targets. Considering the fact that LDH was absent from the Lib 7 of the urine+glc condition, it might be suggested that lactate fermentation under this condition, with fructose and sucrose as primary energy resources, is more of a limiting factor of the growth rate, compared to the condition with glucose as the energy resource. Additionally, the transporters of isoleucine and leucine were also among the identified targets. Other

targets involved in other metabolic pathways such as biosynthesis of secondary metabolites as well as amino acids, including THRD-L and CYSTL. The results from the pathway analysis is shown in Table 3.12.

Table 3.12: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *E. faecalis* from artificial urine+fru+suc.

Pathway	No. Reactions
Biosynthesis of secondary metabolites	4
Metabolic pathways	4
Biosynthesis of amino acids	3
Cysteine and methionine metabolism	3

The growth rate of *S. pyogenes* in CDM-LAB at pH 6.5 was lower than that of *E. faecalis*, being 0.113 OD/h. First, I looked for the potential drug targets. There were 209 essential reactions and 197 essential genes for the model to have a feasible solution. The Lib 7 library of the potential drug targets showed 11 reactions involved in glycolysis, including triose-phosphate isomerase (TPI), phosphofructokinase (PFK), GAPDH, LDH and several others. Again, having glucose as the primary energy sugar in CDM-LAB, there is no surprise that glycolysis has a great impact on the growth of the organism. A lot of these enzymes also play a role in other metabolic pathways such as biosynthesis of secondary metabolites and amino acids, fructose and mannose metabolism and microbial metabolism in diverse environments. Moreover, glutamine ABC transporte (GLNabc) as well as two enzymes involved in fatty acid biosynthesis, namely alcohol dehydrogenase (ethanol: NAD) (ALCD2x) and acetaldehyde reversible transport (ACALDt) were also identified as potential drug targets. The results from the pathway analysis is shown in Table 3.13.

Table 3.13: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *S. pyogenes* from CDM-LAB at pH 6.5.

Pathway	No. Reactions
Biosynthesis of secondary metabolites	12
Metabolic pathways	12
Glycolysis / Gluconeogenesis	11
Microbial metabolism in diverse environments	11
Biosynthesis of amino acids	8
Carbon metabolism	8
Methane metabolism	4
Fructose and mannose metabolism	3
Pentose phosphate pathway	3
Pyruvate metabolism	3

The increase at the pH level (from 6.5 to 7.4) resulted in a higher growth rate of *S. pyogenes* when growing on CDM-LAB. *S. pyogenes* grew at 0.203 OD/h under this condition. For the *S. pyogenes* model constrained by the data obtained from CDM-LAB at pH 7.4, I found 226 essential reactions and 208 essential genes. Similar to the previous conditions, several glycolytic enzymes were identified as potential drug targets in the Lib 7 library, a lot of them being also involved in the biosynthesis of amino acids. In addition to enzymes involved in central metabolism, others which are involved in other pathways such as biosynthesis of cofactors (e.g., methylenetetrahydrofolate dehydrogenase (NADP) (MTHFD)) were also identified as potential targets. The results from the pathway analysis is shown in Table 3.14.

Table 3.14: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *S. pyogenes* from CDM-LAB at pH 7.4.

Pathway	No. Reactions
Metabolic pathways	23
Biosynthesis of secondary metabolites	17
Microbial metabolism in diverse environments	14
Carbon metabolism	13
Glycolysis / Gluconeogenesis	12
Biosynthesis of amino acids	11
Biosynthesis of cofactors	6
Methane metabolism	5
Alanine, aspartate and glutamate metabolism	4
Glycine, serine and threonine metabolism	4
One carbon pool by folate	4
Pyruvate metabolism	4
Fructose and mannose metabolism	3
Pentose phosphate pathway	3

Similar to *E. faecalis*, *S. pyogenes* was grown on artificial saliva with and without additional glucose. The growth rate of *S. pyogenes* growing on artificial saliva (without additional glucose) was 0.21 OD/h. I found 220 essential reactions and 211 essential genes for the model to have a feasible solution. Unlike, in *E. faecalis*, when growing on saliva, the Lib 7 library showed 8 glycolytic reactions as potential drug targets. Not surprisingly, drug targets were also identified in other central metabolic pathways such as biosynthesis of amino acids and central carbon metabolism. Interestingly, the addition of glucose to the medium did not affect the number of glycolytic enzymes which were identified as drug targets. Likewise, a number of central and secondary metabolic pathways contained identified drug targets in the Lib7 library of potential drug targets. The growth rate of *S. pyogenes* under this condition was equal to 0.16 OD/h. I found 249 essential reactions and 219 essential genes. The results from the pathway analysis are shown in Table 3.15 and 3.16 for saliva saliva+glc conditions, respectively.

Table 3.15: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *S. pyogenes* from artificial saliva.

Pathway	No. Reactions
Metabolic pathways	20
Biosynthesis of secondary metabolites	15
Biosynthesis of amino acids	14
Microbial metabolism in diverse environments	12
Carbon metabolism	11
Glycolysis / Gluconeogenesis	8
Methane metabolism	6
Cysteine and methionine metabolism	5
Biosynthesis of cofactors	3
Fructose and mannose metabolism	3
Pentose phosphate pathway	3
Propanoate metabolism	3
Purine metabolism	3
Pyruvate metabolism	3

Table 3.16: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *S. pyogenes* from artificial saliva+glc.

Pathway	No. Reactions
Metabolic pathways	27
Biosynthesis of secondary metabolites	18
Biosynthesis of amino acids	16
Carbon metabolism	16
Microbial metabolism in diverse environments	14
Glycolysis / Gluconeogenesis	8
Methane metabolism	7
Biosynthesis of cofactors	5
Cysteine and methionine metabolism	5
One carbon pool by folate	4
Pentose phosphate pathway	4
Alanine, aspartate and glutamate metabolism	3
Fructose and mannose metabolism	3
Glycine, serine and threonine metabolism	3
Purine metabolism	3

Finally, I sought for the drug targets in the metabolic network of *S. pyogenes* when growing

Table 3.17: The result of the pathway analysis on drug targets from the Lib 7 of the cultures of *S. pyogenes* from natural human plasma

Pathway	No. Reactions
Metabolic pathways	23
Biosynthesis of secondary metabolites	14
Biosynthesis of amino acids	13
Carbon metabolism	12
Microbial metabolism in diverse environments	11
Biosynthesis of cofactors	7
Glycolysis / Gluconeogenesis	7
Methane metabolism	5
One carbon pool by folate	4
Alanine, aspartate and glutamate metabolism	3
Cysteine and methionine metabolism	3
Fructose and mannose metabolism	3
Glycine, serine and threonine metabolism	3
Pentose phosphate pathway	3
Purine metabolism	3

on natural human plasma. The growth rate of *S. pyogenes* under this condition was equal to 0.94 OD/h. I found 224 essential reactions and 205 essential genes. Similar to the previous conditions, a number of drug targets were identified in central metabolism, such as glycolytic enzymes (TPI, PFK, GAPDH), biosynthesis of amino acids (SERD, phosphoribosylpyrophosphate synthetase (PRPPS)). The results from the pathway analysis is shown in Table 3.17.
The complete list of all the potential drug targets can be found in the appendix. The list of exclusive tract-specific drug targets for each conditions based on the combo-approach at 30% flux boundaries can be seen in the figures 3.33 and 3.34.
Considering the fact that proteome data in the case of both organisms contained non of the available enzymes in the genome-scale models, the integration of proteome data had no effect on constraining the models and tract-specific drug target identification. This shortcoming can be overcome by repeating the experiments using a set up that results in higher yields of bacteria, which in turn brings a higher protein content for quantification.
The results from the integrative tract-specific drug target identification suggested that the mathematical modelling of pathogens under tract-specific conditions is essential, as the phenotypes of pathogens vary considerably according to environmental conditions. Therefore, integrative tract-specific analysis of would bring a deeper insight towards the adaptive metabolism of these micro organisms and help to design new treatment strategies.

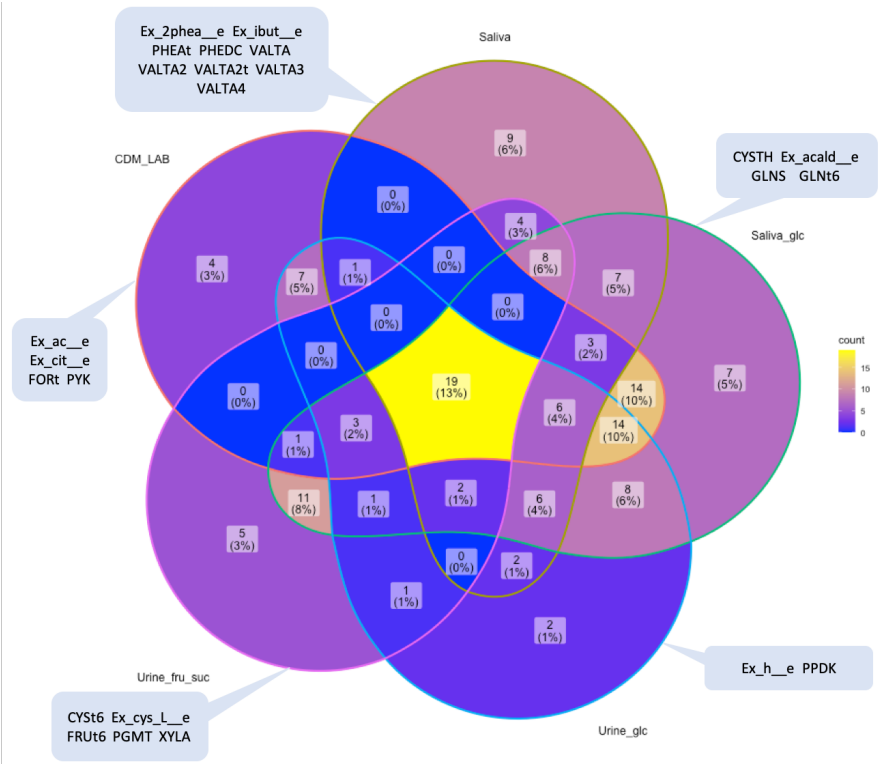

Figure 3.33: The number of common drug targets between different conditions, as well as the exclusive targets under each condition in *E. faecalis*.

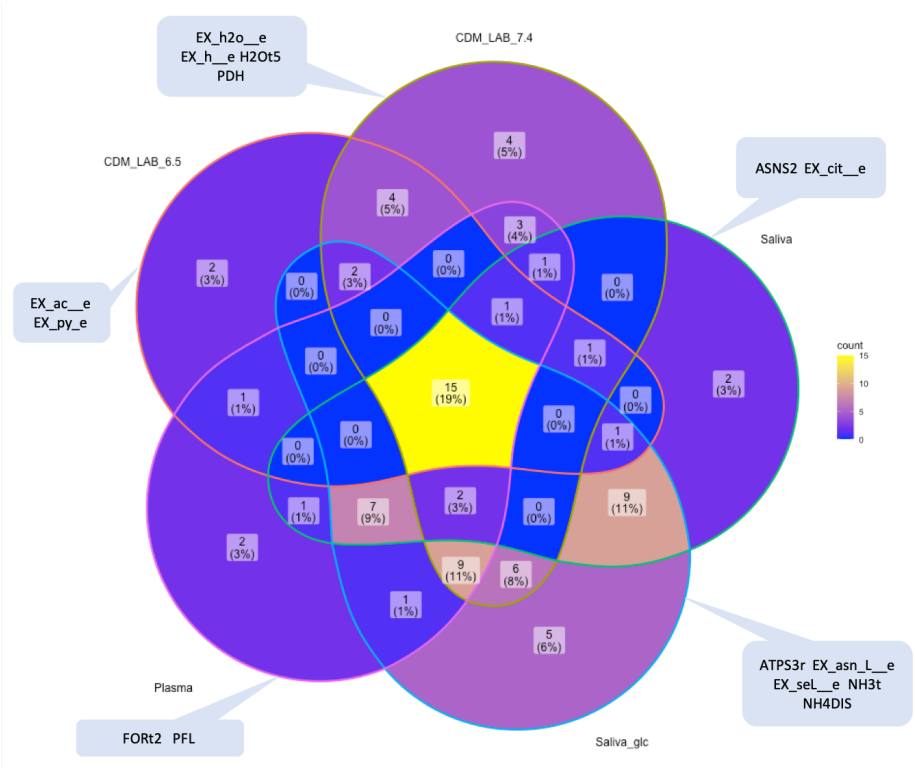

Figure 3.34: The number of common drug targets between different conditions, as well as the exclusive targets under each condition in *S. pyogenes*.

Chapter 4

Discussion

Studying the metabolic characteristics of pathogens is a common approach to uncover their strategies to grow, survive, colonize and resist to antimicrobial agents in the human body. Some of the pathogenic bacteria possess a versatile metabolism that allows them to adapt to various environments, whether its different parts of the human body for its goods (working as probiotics) and bads (causing infections), or plant and dairy products that usually results in developing fermentation products and flavours. In this study, I investigated the metabolic attributes of two pathogenic bacteria, *E. faecalis* and *S. pyogenes*, using their multi-omics profiles (metabolic, transcriptome and proteome) together with their respective genome-scale models, aiming at finding tract-specific drug targets. Prior to this main goal, I performed a more in-depth analysis of *E. faecalis* metabolism by studying the properties of a (Δ*glnA*) mutant, which helped to better understand the metabolic and proteomic adjustments of this species under different environmental conditions.

4.1 The effect of glutamine auxotrophy on metabolic characteristics of *E. faecalis*

E. faecalis is a growing concern in hospital environments as more of its strains are progressively developing resistance against conventional antibiotics. On the other hand, *E. faecalis* plays an important role in the food industry, specifically in the dairy sector. As a result, studying its adaptive behaviour is a key step in understanding the survival strategies of this distinctly flexible microorganism. The integrative analysis of metabolism comprising metabolic, proteomic and mathematical modelling presented in this study provides a subtle ground to characterize different aspects of *E. faecalis* adaptive metabolism. Here, I studied the impact of dropping environmental pH from 7.5 to 6.5 over the course of 21 hours on a glutamine synthetase (Δ*glnA*) mutant of *E. faecalis*. Considering the fact that *E. faecalis* colonizes different tracts in the human body, it is often exposed to the conditions with highly variable parameters, one of them being the environmental pH. Therefore, studying the adaptation strategies of this species from one pH level to another would help us to understand the underlying mechanisms, which can be useful when designing treatment plans. To uncover the role of glutamine auxotrophy on the

pH adaptation of the mutant, I compared the results to those of the wildtype [Großeholz et al., 2016]. A focal point in adapting to the new and more acidic environmental pH shared by the mutant and the wildtype was maintaining the internal pH homoeostasis by pumping protons to extra cellular environment, which of course increases the energy demand. To enable such a mechanism, proteome and metabolic profiles had to be adjusted accordingly. Therefore, there were many similarities between the wildtype and the mutant following the decline in the environmental pH, while there were of course some differences. Although the two genotypes shared the quality of increasing energy demand in response to pH shift, the quantity of energy demand reflected in biomass production was dissimilar.

The most prominent consequence of the increasing energy demand was the decline in biomass production that occurred both in the wildtype and the mutant. Moreover, the fermentation pattern shifted from mixed acid fermentation to homolactic fermentation in response to pH shift. The change in fermentation pattern was observed in the case of both genotypes, although the degree of change was considerably higher in the mutant, with the larger part of the fermentation profile being made of lactate. This verifies the fact that under a more energy demanding condition, being the Δ*glnA* mutant or a more acidic environment, *E. faecalis* changes its fermentation pattern from mixed acid to homolactic fermentation. Based on the stoichiometry of the fermentation pathway, mixed acid fermentation produces one more ATP compared to homolactic fermentation. However, it is reported that lactic acid bacteria like *L. plantarum* and *L. lactis* use homolactic fermentation under conditions such as high substrate availability, rapid growth rate or high glycolytic flux [Teusink et al., 2011]. As a result of a higher energy demand, more glucose is taken up from the medium, which further increases the flux through glycolysis as well as the NDAH/NAD ratio. The higher ratio of NADH/NAD is reported to up-regulate the activity of lactate dehydrogenase (LDH) in *L. lactis* [Garrigues et al., 1997]. Although the up-regulation of lactate production under energetically demanding conditions is widely reported in the literature, a conceivable biological explanation of why an energetically less efficient fermentation pattern is used by lactic acid bacteria is still to be discovered.

The overwhelming impact of increased energy demand in a more acidic environment was again reflected in the amino acid uptake/production profile. Except for serine, arginine and glutamine, the uptake rate of amino acids decreased in response to pH shift. Decreasing amino acid uptake might be due to regulation as a result of decreasing biomass production, in addition to the fact that it saves energy. Generally, amino acid uptake is coupled to either ATP hydrolysis or proton symport, which in both cases increases the energy demand of the cell. However, as arginine and serine can be used for energy production, the increased uptake rate of these two amino acids in the more acidic environment can be easily explained. As reported in the literature, arginine is widely used as a valuable energy resource in lactic acid bacteria, especially in the ones that are used in the food industry [Hwang and Lee, 2018] [Tonon and Lonvaud-Funel, 2000]. When used for energy production, arginine is converted into ornithin instead of citruline, whether arginnine deaminase is used or not [Hwang and Lee, 2018] [Tonon and Lonvaud-Funel, 2000]. Likewise, serine is reportedly deaminated to ammonia and pyruvate

(which further promotes the production of acetate and formate) and therefore contributes to energy production in the growth cells of *L. plantarum* isolated from cheese [Liu et al., 2003]. The flux distribution profile obtained from the genome-scale model in this study successfully replicated this behaviour, predicting an increase in arginine uptake in response to pH shift that results in an increase in ornithine production. The model also correctly predicted the conversion of citruline to ornithine by ornithine carboyl transferase, a reaction that is previously reported to take place in *E. faecalis* [Knivett, 1954]. The produced ornithine is further used to import arginine by arginine-ornithine antiporter.

The metabolic profiles obtained from the chemostat experiments showed a very big difference in glutamine/glutamate metabolism between the wildtype and the mutant. Following the pH shift, the glutamate uptake rate considerably decreased in both genotypes. However, the glutamine uptake rate increased only marginally in the wildtype, while it increased substantially in the mutant. As discussed earlier in the results chapter, there are multiple mechanisms that could potentially lead to this observation. For one, as it is reported on *S. pneumoniae* [Kloosterman et al., 2006], it can be proposed that the regulatory effects of GlnR (a transcription factor), including the control over glutamine transport depend on the presence of an intact gene for GlnA. As the gene for GlnA is absent in the Δ*glnA* mutant, GlnR might lose its functionality and it results in an unregulated glutamine uptake. Moreover, the change in the directionality of the GDH and its following regulatory effects is only possible if more 2-oxoglutarate is available. For this to happen, glutamine is required. Also, as reflected in the results from the genome-scale model, the lack of specificity of the amino acid transporters plays an important role and has to be taken into account. The uptake rate of the co-transported amino acids with glutamate such as aspartate, which are not very easy to catabolize (and their uptake rate decreased in response to pH shift) underwent a decrease, while the ones that are co-transported with glutamine such as arginine can be used for energy production and their uptake rate increased following the pH shift.

Ammonium assimilation takes place mainly by the glutamine synthetase reaction (catalyzed by the enzyme GlnA) [Forchhammer, 2007]. This reaction of course does not occur in the Δ*glnA* mutant. In this study, glutamine uptake happens at such a high rate that part of the imported glutamine is further converted into glutamate and ammonium by a deaminase reaction. This might explain the slight amount of glutamate export at pH 6.5 and is also successfully replicated by the model. Additionally, the higher uptake rate of arginine and serine in the mutant compared to the wildtype and their respective deamination would increase the amount of intracellular ammonium even further. The high intracellular concentration of ammonium is reported to decrease the growth rate in bacteria [Müller et al., 2006]. However, it is suggested that it is rather the general osmotic or ioinic effects of ammonium, than its toxicity, which affects the growth rate [Müller et al., 2006]. Although not being completely clear, the ammonium export occurs either using a proton antiporter or active transport by using ATP, which in either case adds up to the energy expense of the mutant.

The proteome data used in this study uncovered some additional aspects of the metabolic ad-

justments in the mutant in response to pH shift. These findings paralleled those of the wildtype in some ways, while being different in other ways. The down regulation of genes in nucleotide metabolism after pH shift occurred in both genotypes, as the bacteria grew at a lower rate, less nucleotides were required for cells to grow. In addition, de novo synthesis of nucleotides is not energetically favourable, so the bacteria could take up the required nucleotides from the medium when energy was more of a limiting factor. Moreover, the up-regulation of enzymes involved in the restructuring of the cell membrane and cell wall, which are necessary to control the proton leak in a more acidic environment, was shared between the wildetype and the mutant. On the other hand, the up-regulation of glycolytic enzymes at the beginning of the pH shift which was previously observed in the wildtype, was not found in the mutant. Considering the fact that the glycolytic flux in the mutant was already at a higher level at the start of the experiment in the mutant, it can be suggested that the expression of the glycolytic enzymes was already at a higher level in the first place.

The genome-scale model successfully predicted a large fraction of the experimental data. However, the model had to be adjusted with respect to the stoichiometry of amino acid transporters in order to reflect their real unspecific character, so it would be able to mimic the experimental data. As far as this study is concerned, such adjustment has no precedence in bacterial genome-scale models.

The prominent outcome of the integrative analysis of the metabolic network in the Δ*glnA* mutant was a higher energy demand compared to the wildtype. The results suggested that this is perhaps a consequence of losing control on the glutamine transport system. To the extent of my knowledge, glutamine is the main nitrogen donor in bacteria. Glutamine has to be provided in surplus in the cell to ensure the sufficient supply of amino-groups [Reitzer, 2014]. Glutamine can either be synthesised by glutamine synthetase, or taken up from the extracellular environment. While both can happen in the wildtype, the mutant is only capable of doing the latter. The stochiometric analysis of both processes suggests that one ATP has to be consumed to provide one molecule of glutamine, either way. So the energetic cost would be similar. However, if glutamine transport is unregulated due to the lack of GlnA, an excessive amount of glutamine would be imported into the cell, which is costly. Furthermore, the uncontrolled glutamine transporter would import other amino acids in surplus as a consequence of being unspecific. These two factors together with the need to export the resultant high level of ammonia would add additional costs to the cell.

The metabolic and proteome data used in this part of the study could reflect the biological phenotypes of the Δ*glnA* mutant to an acceptable extent. The metabolic data clearly showed the fermentation pattern shift, both in the wildtype and the mutant, which is reportedly a shared characteristic of LABs under more energetically demanding conditions [Teusink et al., 2011]. However, there were possible measurement flaws as well. For what I expect, the glutamate should still be taken up from the medium after the pH shift, while the data is suggesting otherwise. Overall, the metabolic data presented a good and acceptable account of the metabolic behaviour of the Δ*glnA* mutant under both pH conditions. The proteome profile contained the

relative abundances of around 1600 gene products, presenting a very large fraction of the whole proteome of *E. faecalis*. As a result, most of the metabolic genes were detected and subjected to the relative significance analysis. However, the absence of membrane integrated proteins in the data was noticeable and limited the room to explain the adjustments in response to pH shift, especially when amino acids uptake and production were the most probable reason for the major outcome of the study, which was the higher energy demand in the mutant.

From the modelling point of view, genome-scale model of *E. faecalis* was a powerful tool to analyse the metabolic behaviour of this species and predict the phenotypes that were not reflected in the data, which of yourse have to be experimentally verified. Despite the overall success of the genome-scale model in simulating the experimental data, the model possess a very large solution space containing highly variable flux distribution profiles. This imposes limitations to the conclusions drawn from modelling. The approach used in this study to shrink the solution space was to integrate omics profiles obtained from experimental data, which supposedly increases the predictability of models. In order to find out how efficient the data reduced the solution space of the genome-scale models of *E. faecalis*, and in general any other genome-scale model, and to find out what portion of the results in the solution space are biologically feasible, I developed a method to inspect the solution space of genome-scale models.

4.2 Inspecting the solution space of genome-scale metabolic models

Genome-scale metabolic models comprise almost all the existing reactions in an organism and therefore are large systems. As a result of the large number of reactions and parameters, these models are mathematically underdetermined, meaning that they contain more unknown variables than equations. Hence, there is no single solution available for these systems, and solutions have to be rather found in a large solution space. The solution space contains all the flux distribution profiles that satisfy the optimum value of the objective function. These profiles can be highly variable as a result of nonlinearities and branching points in the metabolic network on the one hand, and the lack of enough constraints on the other hand. Therefore, two flux distribution profiles can be enormously different while fulfilling the same value for the objective function. In the present thesis, in order to find out about the extent of variability, as well as the effect of different types of constraints on the results of FBA, a method was developed and the results were introduced in the results chapter. Briefly speaking, the method subjects the results of FBA to random perturbation and collects the resultant flux distribution profiles for downstream statistical analysis. This allows for computationally cheap yet statistically extensive sampling of the solution space. In the case of the wildetype and the mutant models of *E. faecalis*, the method showed that there are not only quantitative differences between different flux distribution profiles that have to be the point of attention, but also qualitative differences which in many cases might result in biologically inconsistent interpretations of FBA results.

From the quantitative standpoint, this method allows us to asses the sensitivity in genome-scale models by defining the number of sensitive reactions and the spread of the flux values within the variability interval. For that, I defined the reactions changing their flux value in response to perturbation for more than $\pm 5\%$ as sensitive, and robust if otherwise. The distinction between sensitive and robust reactions enables us to study the effect of different constraints on the sensitivity and predictability of genome-scale models. I therefore proposed that the number of sensitive reactions in a model is a very good indicator of the size of the solution space. The downstream analysis showed that the models with a lower number of sensitive reactions fit the experimental data better and are potentially more accurate when it comes to predictions.

The high variability among flux distribution profiles acquired by FBA emphasizes the point that an optimal flux distribution profile generated by FBA is just one arbitrary selection of too many fitting profiles and is by no means unique. Therefore, when discussing the results from only one run of FBA, it is important to refer to the results as 'an optimal' rather than 'the optimal' flux distribution.

Moreover, three different aspects of the results have to be discussed further:

1. The integration of biological constraints obtained from experimental data is key to decrease the number of biologically irrelevant results. In this study, the integration of proteome data was the most effective way to decrease the solution space. Nonetheless, metabolic data, namely the uptake and production rate of metabolites, integrated into the boundaries of exchange reactions can also considerably shrink the solution space. The effect of metabolic data perhaps is best reflected on lowering the flux variability at the branching points in metabolism. However, the degree of reduction is actually case-specific, signifying the importance of the second point [Razmilic et al., 2018].

2. It is necessary to analyse the solution space when using FBA. The results from this study showed that the frequently used constraints such as metabolic data can effectively shrink the solution space. As a result, drawing biological conclusions, e.g., the type of fermentation pattern in this case, can be done with higher accuracy. Of course, the ability to acquire such knowledge and the extent of accuracy depends significantly on the type of biological question and available data-sets. The analysis of the solution space can be done in different ways. The method presented in this thesis provides an easy method to study various optimal flux distributions by means of random perturbation. As demonstrated by the results, functional analysis of the solution space using the perturbation method provides a deep insight towards the robustness and reliability of the outcomes obtained from a genome-scale model. The use of this method would also enable us to determine which type of experimental data was most effective in shrinking the solution space and support a better experimental design.

There are also other methods available for the analysis of the solution space that were not used in this study. The most frequently used technique in sampling the solution space is Monte-Carlo

sampling. There have been several methods developed based on the Monte-Carlo sampling, most of them are used to calculate the probability distribution of individual flux values or to determine the correlated reaction sets used for experimental design [Price et al., 2004] [Thiele et al., 2005]. The perturbation method on the other hand, is focused on revealing the existing uncertainty in the interplay of different fluxes. As previously mentioned, correlated reaction sets obtained by Monte-Carlo methods enable the determination of candidate reactions for flux measurements, which can be further used to estimate the flux value of the correlated ones. However, the perturbation method showed that the integration of the metabolic data on the fermentation products for which the internal reactions were reported to be correlated (e.g., LDH, PFK) does not essentially eliminate the biologically irrelevant results (in the case of pyruvate branching point in the mutant). Hence, the functional analysis of the solution space provided by the perturbation method helps to obtain a comprehensive knowledge and overview regarding the behaviour of genome-scale models. Moreover, the perturbation process requires far less data points compared to most of the Monte-Carlo methods, which normally need a very large sample size (250,000 data points in [Price et al., 2004]) in order to deliver a reliable outcome. The perturbation process, as it is defined in this study, needs ~ 10 times the number of variable reactions in the model, which in the examples of the models used as the case study here would be $2900 \sim 3980$ data points. However, as a result of having fewer data points, our method does sample the solution space in a less extensive way and some alternative solutions might be dismissed. Furthermore, despite the fact that the perturbation method has a different focus, it is important to mention that it is computationally more efficient and faster compared to Monte-Carlo methods. As it is reported by Fallahi and colleagues [Fallahi et al., 2020], sampling the solution space in the case of the models of comparable size to the model of *E. faecalis* takes approximately 7.64 to 10.67 minutes, using the CHRR method (as the fastest available Monte-Carlo method) on a processor of intel Core i7 at 2.5 GHz. It is important to mention that in this study, a reduced version of the genome-scale models, in the sense that the reactions carrying no flux were discarded, were used. As a result, the number of reactions of the four models, iLJ478, iSB619, iHN637 and iJN746 were reduced from 652, 743, 785 and 1054 to 380, 450, 522 and 652, respectively [Fallahi et al., 2020]. On the other hand, the perturbation process in this study for the models of *E. faecalis* took between 122 to 175 seconds for different models (wildtype or mutant), and of course depending on how constrained a model was, on a processor of Intel Core i5 2.3 GHz, 16 MB memory and HDD hard drive. The flux distributions in this case were acquired by FBA on MATLAB. The perturbation method also allows to obtain the flux distribution profiles by FVA, which in this case the run time increased to 31 to 51 minutes for the same models using the same hardware. The comparison of the run time together with the number of reactions in each model for CHRR as well as the perturbation method is shown in detail in the supplement. It has to be pointed out that additional statistical analyses take more time.

3. Caution has to be taken when some of the parameter values sit at the edge of the feasible region, as it can significantly affect the result of optimization and therefore, produce artefacts.

Our result showed that particularly the methods using FVA can fail to generate reliable outcomes, as FVA usually fails to fully operate under such circumstances. In the case of this study, the failure of FVA resulted in the calculation of an artificially smaller solution space.

Finally, it has to be mentioned that at this point, we cannot distinguish between the absolute solution space and the ones that are reported by various methods, software implementations and numerical methods. The perturbation method presented in this thesis reported no qualitative difference when analysing the solution space by means of different numerical libraries or software implementations. For example, the decreasing trend of the number of sensitive reactions following the integration of constraints were obtained by all the applied software implementations/numerical libraries. However, there were minor quantitative differences between the outcome of these methods, e.g., the number of sensitive reactions. There were also some major quantitative differences such as the average number of reactions reacting to one perturbation. Hence, it can be suggested that the most reliable indicator of the size of the solution space using the perturbation method is the number of sensitive reactions, which shows only small differences between different implementations/libraries. The qualitative aspect, namely the trend of change in the number of sensitive reactions is also an informative indicator to analyse how different types of experimental data affect the solution space.

So far, I showed that genome-scale models are powerful and useful instruments to analyse metabolism in large scale and capture biological phenotypes. I also explained how experimentally measured omics data can effectively shrink the solution space, increasing the predictability of genome-scale models. As a result, I used the genome-scale model of *E. faecalis* and *S. pyogenes* to find tract-specific drug targets in their metabolic network. I used different types of omics data to make the models tract-specific and shrink the solution space, so the predictions would be more trustworthy.

E. faecalis and *S. pyogenes* are two important bacterial pathogens causing severe problems in hospital environments. They both cause various infections in different tracts. *E. faecalis* for example causes urinary tract infection, root canal infection, bacteremia, endocarditis and wound infection [Fiore et al., 2019]. *S. pyogenes* on the other hand, provokes scarlet fever, puerperal sepsis, pharyngitis, impetigo, pneumonia, root canal infection and blood stream infection. Additionally, *S. pyogenes* can cause immune-mediated post-infectious disorders such as rheumatic heart disease (RHD) and rheumatic fever (ARF) [Walker et al., 2014]. The ongoing trend of multi resistance, especially in the case of *E. faecalis* has resulted in serious problems in the hospital environments and made the treatment of resistant isolates extremely hard.

A wide range of antibiotics have been used to treat the infections caused by these two pathogens. Ampicillin is widely used to treat the infections caused by *E. faecalis* [Kristich et al., 2014]. Penicillin, vancomycin and daptomycin are also used against *E. faecalis*, although to a smaller extent [Kristich et al., 2014]. *E. faecalis*, however, employed various strategies to resist the aforementioned antibiotics, giving rise to the current trend of multi-resistant in this organism [Miller et al., 2014]. On the other hand, *S. pyogenes* is still susceptible to beta-Lactam antibiotics like penicillin, amoxicillin and cephalosporins [Camara et al., 2013]. While the treatment

of *S. pyogenes* using conventional antibiotics is still possible, combating *E. faecalis* strains with multi-resistance character is extraordinarily hard. This ever-growing trend of resistance amongst pathogenic bacteria points out the necessity of developing new strategies to fight these micro organisms. Of course, a big part of such a plan is the development of new antibiotics. However, finding new drug targets, e.g., in the metabolic network of pathogens, which do not allow for the rapid development of resistance, is equally important and recently in the focus of international research. To do so, metabolism of the pathogenic bacteria has to be studied in more detail.

Comprising all the functional biochemical reactions, metabolism determines how a living organism performs under a given condition. Through the extensive exchange of metabolites with extracellular environment, it allocates resources to serve in the best interest of the cell. Therefore, environmental conditions play a significant role on the metabolic characteristics of a cell. There are various aspects of extracellular environment that affect the phenotypes of a living cell. To name a few, extracellular pH can impact the growth rate as well the fermentation pattern of *E. faecalis* (as also mentioned previously in this thesis) [Großeholz et al., 2016], [Loghmani et al., 2021]. Likewise, the glucose content of the extracellular environment and also the use of different carbohydrates are reported to take part in the virulence of *S. pyogenes* [Shelburne et al., 2008], [Thurlow et al., 2009]. Therefore, it is crucial to take the characteristics of the extracellular environment into account when analysing metabolism. This also holds true when metabolism is studied with the aim to find novel drug targets. That being said, the environmental condition in different parts of the human body can be considerably different, forcing pathogens to adapt to each and adjust their metabolic characteristics accordingly. Hence, it is essential to distinguish between the characteristics of different environmental conditions when trying to find efficient drug targets. Such an approach does not only enable us to combat a pathogen very effectively (as a result of identifying the drug targets by taking the tract-specific conditions and the respective metabolic adjustments into account), but also prevents the pathogen from a rapid development of resistance. The latter would be due to a smaller evolutionary pressure on the pathogen as a result of employing different antibiotics to pick out different targets when colonizing different tracts.

4.3 Integrative tract-specific drug target identification in *E. faecalis* and *S. pyogenes*

With such perspective, I used an integrative approach to find new drug targets in the metabolic network of *E. faecalis* and *S. pyogenes*. The experimental design comprised the growth of the bacteria in the media comparable to the actual physiological condition in different tracts in human body, namely root canal (for *E. faecalis* and *S. pyogenes*), urinary tract (for *E. faecalis*) and human plasma (for *S. pyogenes*). In order to account for the variation in sugar content in these environments, an additional treatment with additional glucose (for saliva and urine) and

additional fructose and sucrose (only for urine) were used. To develop the tract-specific omics profiles, metabolic data (uptake and production rate of carbohydrates, organic and amino acids) were quantified in each environment. The transcriptome and proteome profiles, however, were determined relative to the profiles from the standard lab condition (CDM-LAB). The aim of such design was to determine the tract-specific multi-omics profile of the two pathogens and use them for mathematical modelling.

For the bacterial cell cultures, I used artificial saliva and artificial urine because of the consistent composition of the artificial media. As the composition of biological media vary significantly, not only from person to person, but also from time to time, and depending on the specific diet before sampling, the reproducibility of cultures using those media is very low. This is even a bigger problem when quantitative data are needed for mathematical modelling, when even small variability in some parameters enormously affects the outcome. For saliva, the content can be mainly affected by food intake or during sleep, when low oxygen availability in the mouth can promote lactate fermentation. In urine, it has been shown that the composition is not only dependent on the diet, but also on other factors such as age, gender, etc. [Mack et al., 2018]. Therefore, the choice of the artificial media helped to have a higher reproducibility in the experiments, which, however, was achieved at the expense of lower growth rate compared to natural media such as plasma. Regarding plasma, a preliminary test experiment resulted in zero growth of the bacteria, therefore, the only choice was to use natural human plasma.

During the test experiments it became clear that the lack of primary sugar content prevents a high growth rate in the artificial media. It has been previously reported that *E. faecalis* has an overall poor growth rate in artificial saliva [Kampfer et al., 2007]. It is also shown that the growth rate of *E. faecalis* on artificial urine supplemented with additional metabolites such as folic acid can be considerably increased compared to pure artificial urine [Lara et al., 2021]. The same phenomenon has been observed in the case of *S. pyogenes*. More specifically, the addition of artificial saliva to BHI (Brain heart infusion) supplemented with 5% glucose has been shown to reduce the already poor growth rate of *S. pyogenes* [Riani, 2009]. Interestingly, *S. pyogenes M49*, the strain that is used in this thesis, showed one of the lowest growth rates in artificial saliva compared to the other dental colonizing species including *E. faecalis* [Riani, 2009]. The very high growth rate of *S. pyogenes* on natural human plasma, however, very clearly showed the effect of a rich biotic environment on the growth capability of this pathogen. Despite the poor growth rate of *E. faecalis* and *S. pyogenes*, I was able to obtain a sufficient amount of supernatant and bacterial pellets for multi-omics experiments.

As mentioned above, the poor growth rate was a limiting factor in the experimental design. The already low amount of bacterial pellets had to be saved for transcriptomics and proteomics studies. Therefore, I could not afford to sacrifice a large part of the pellets to calculate the dry mass for each treatment. On the other hand, the dry mass calculation was necessary to calculate the uptake/production rate of metabolites. To overcome this issue, I ran an experiment to derive calibration curves to allow for the estimation of the dry mass based on OD values. This is a necessary step, as the estimation of dry mass based on OD needs to be condition-specific. As

the chemical composition of each medium is different, the metabolic phenotype of organisms would vary, which in turn results in the excretion of different metabolites that would affect the OD measurement. As a result, the dry mass of bacterial cultures cannot be simply estimated by the same calibration cure when growing on different media. As I mentioned earlier, the calibration curve was developed with the aim to calculate the metabolite uptake/production rates to be integrated into the genome-scale models. However, when the modelling started, the calculated metabolic fluxes based on the estimated dry mass resulted in very unrealistic values. A very probable reason for such inaccuracy could be the very low quality of the metabolic measurements. Repeating the metabolomics experiments with a better measurement set up would potentially solve this shortcoming. Therefore, I calculated the flux values based on the OD measurements, which resulted in acceptable and realistic values.

To acquire multi-omics profiles, the metabolic, transcriptome and proteome profiles of the two bacteria were determined. As mentioned in the previous paragraph, the quality of measurements in metabolomics experiments were very low. Hence, an informative and reliable biological conclusion from the metabolic profiles is not possible at this point. The transcriptome data has been previously published in my master thesis [Loghmani, 2020].

The proteome profile of the two pathogens was analysed and presented earlier in the resuls chapter To determine the proteome content, shotgun DDA mass-spectrometry was used. The plasma samples were fractionated into 6 fractions before quantification. The samples from other conditions were subjected to proteomics without fractionation. Whole cell proteome experiments have been previously used for tract-specific studies in the case of *E. faecalis*. Arntzen and colleague performed proteomics on *E. faecalis* growing on urine relative to 2xYT [Arntzen et al., 2015]. They reported that six proteins that were exclusively expressed in urine, including endocarditis specific antigen and its homologue, adhesion lipoprotein [Arntzen et al., 2015]. To a lower extent, proteomics was applied to study the membrane associated proteins in *E. faecalis*, although in a non tract-specific study [Yan et al., 2018]. The lack of extensive tract-specific proteomics holds true in the case of *S. pyogenes* too. In one of the few attempts, Edwards and colleagues performed proteome analysis on the *S. pyogenes* isolates obtained from infected patients fluids (empyema, septic arthritis, necrotising fasciitis) in an attempt to find new diagnostic targets [Edwards et al., 2018]. Similar to this, they also faced the problem of identifying many streptococcal proteins in the samples. In another study, the proteome interaction map of *S. pyogenes* and human plasma were investigated [Sjöholm et al., 2014]. To the best of my knowledge, this theses presents the first study trying to determine the multi-omics profile of *E. faecalis* and *S. pyogenes* in tract-specific conditions.

Unlike the transcriptome data, only the cultures with additional glucose as well as the ones from CDM-LAB were used for the proteomics study for *E. faecalis*. In the case of *S. pyogenes*, only the cultures from natural human plasma and CDM-LAB at pH 7.4 were used for the proteomics. The reason for this selection, in the case of *E. faecalis*, was the very minimal difference in the transcriptome profile [Loghmani, 2020] between the different sugar contents of the same media. Therefore, it was assumed that most of the conclusions drawn from the proteome profile of the

samples from one sugar content can be extrapolated to the other. Regarding *S. pyogenes*, there were so few significant fold changes in artificial saliva cultures even at the transcriptome level, that there was no realistic chance of observing significant changes at the proteome level. Hence, the saliva cultures were completely excluded from the proteomics for *S. pyogenes*. It should be mentioned that the following discussion on the up- and down-regulated proteins is primarily done to put the data in context and to mention similar findings in other studies. For a conclusive biological discussion on the findings of the proteomics in this study, the results have to undergo a validation process, as the quality of the data in the present study cannot be guaranteed to be very high.

Surprisingly, the most up-regulated protein in *E. faecalis* in saliva with additional glucose (compared to CDM-LAB) was an uncharacterised protein Q832L5 (EF2211), showing the necessity of more detailed and accurate proteomics study under tract-specific conditions. Among other highly up-regulate proteins, there was the glyoxalase family protein Q838D9 (EF0630), which has been previously shown that its disruption (at the gene level) attenuates the killing ability of *E. faecalis* in the nematode model *Caenorhabditis elegans* [Maadani et al., 2007]. This protein is prominently involved in damage control and repair. I also found a putative permease Q82ZR3 (EF2985) and oxidoreductase, aldo/keto reductase family Q838E0 (EF0629) among the highly up-regulated proteins, for which a connection to virulence has not been reported yet. Among others, there were three transporte proteins that were highly up-regulated, ABC transporter, ATP-binding protein Q82ZR2 (EF2986), ABC transporter, permease protein, putative Q833B5 (EF2049) and ABC transporter, ATP-binding protein Q833B4 (EF2050), showing the important role of transporters in adjusting to different environmental conditions. I also observed an up-regulation in the expression level of a lipoteichoic acid biosynthesis Q830N3 (dltD). Lipoteichoic acid has been reported to take part in the inflammatory response caused by *E. faecalis* [Baik et al., 2008].

On the other hand, there were multiple proteins that were highly down-regulated in *E. faecalis* in the saliva + glucose compared to CDM-LAB. Among the two most down-regulated proteins, there were two proteins involved in pheromone signalling, namely Pheromone binding protein, putative Q839T9 (EF0063) and Pheromone binding protein Q834W4 (EF1513). The pheromone system is a way of communication in enterococci coupled with the conjugation system, which is mostly used to transfer the genes responsible for antimicrobial agents and virulence [Dunny et al., 1995], [Sterling et al., 2020]. The down regulation of proteins involved in pheromone signalling in the saliva environment is a novel finding and can be used for further investigations in clinical and systems biology studies. However, it should be pointed out that the results of the proteomics presented in this thesis have to be validated by additional experiments to prove the validity of observations. The results also showed the down regulation of Peptidase, M20/M25/M40 family Q839D6 (EF0236), which has been previously shown to be down-regulated in a Δ*rnjB* mutant of *E. faecalis* [Gao et al., 2017]. The rnjB gene encodes an active RNase J2 and its deletion has been reported to attenuate the virulence capability of *E. faecalis* [Gao et al., 2010]. This protein has been also found in a proteomics study on the en-

dodonic infection, in which bacterial cells were isolated from patients with root canal infection [Nandakumar et al., 2009]. Seemingly correlated, the role of this protein in root canal infection cannot be conclusively discussed at this point.

The comparative proteome analysis also showed a number of significant fold changes in the urine with additional glucose cultures compared to the ones from CDM-LAB in *E. faecalis*. The most highly up-regulated protein was Universal stress protein family Q836Q0 (EF1058). Universal stress proteins in general help bacteria to cope with the stresses imposed by various factors [Kvint et al., 2003]. ABC transporter, ATP-binding protein Q832Z4 (EF2074) and Cationic ABC transporter, ATP-binding protein H7C6W2 (EF0575) were also up-regulated in artificial urine (supplemented with additional glucose) compared to CDM-LAB. Although the significant fold change of transporter proteins between different media is no surprise, they would significantly impact the outcome of mathematical models in case they are part of the model. Earlier in this chapter I discussed how an up-regulation of the ATP-binding subunit of glutamine ABC transporter directed us to come up with a new transporter design for the genome-scale model of Δ*glnA E. faecalis*. The up-regulation of Endocarditis specific antigen Q832Z2 (EF2076) was also observed in the transcriptome profile [Loghmani, 2020]. While it is reported that this protein is a major factor in root canal infection [Preethee and Deivanayagam Kandaswamy, 2012], its involvement in urinary tract infection has not been reported yet.

For *S. pyogenes*, the comparative proteomics was performed only between plasma and CDM-LAB 7.4 samples. Two of the three most highly up-regulated proteins in plasma compared to CDM-LAB were Ribonucleoside-diphosphate reductase A0A0H3BZA0 (nrdE.2) and Ribonucleoside-diphosphate reductase A0A0H3BYW0 (nrdF.2). Ribunucleotide reductase enzymes have been shown to be essential for the survival of bacteria [Torrents, 2014], but an indication of their involvement in virulence has not been reported so for. Another highly up-regulated protein was ABC-type cobalt transport system, permease component CbiQ A0A0H3C149 (Spy491396c). The role of cobalt transport system has not been studied in *S. pyogenes*, however, a cobalt transporter (Opp1 ABC transporter) has been reported to affect the urease activity and help with colonization and virulence of *Staphylococcus aureus* in urinary tract infection models [Remy et al., 2013]. Sugar ABC transporter, sugar-binding protein A0A0H3BVP3 (Spy490212) is another highly up-regulated protein in plasma. ABC transporters are not only responsible for the active transport of necessary metabolites, but they are also involved in the export of capsular polysacharides (in gram-positive bacteria) [Silver et al., 2001], export of antibiotics in antibiotic producing bacteria [**?**], export of antibiotics resistant agents [van Veen et al., 2001] and also export of cellular toxins [Garmory and Titball, 2004]. Quinolinate phosphoribosyltransferase [decarboxylating] A0A0H3BVM1 (Spy490176) is also highly up-regulated in plasma. Quinolinate- salvage pathway (QSP) pathway has been shown to be important in the virulence of *S. pyogenes* [Sorci et al., 2013]. Among the highly up-regulated proteins in plasma, this has been the only protein that reportedly plays a role in the virulence of this species.

The most highly down-regulate protein in *S. pyogenes* in plasma compared to CDM-LAB was the Putative short chain dehydrogenase/reductase A0A0H3BYU6 (Spy490881). The expression

of Citrate (pro-3S)-lyase subunit beta A0A0H3C0N9 (citE) and Citrate lyase acyl carrier protein B5XLM6 (citD), two subunits of citrate lyase which is an important enzyme in the fermentation of citrate [Subramanian and Sivaraman, 1984] was down-regulated too. It seems that as a result of excessive substrate availability in the plasma environment, *S. pyogenes* down-regulates the citrate lyase that perhaps brings no advantage to the bacteria. Other highly down-regulated proteins were Uncharacterized phage-associated protein A0A0H3C1X0 (Spy491458c) and Putative major head protein (Phage associated) A0A0H3BYL2 (Spy490774).

Unlike in the transcriptome profile, not so many virulence related proteins were shown by the proteome profile to be significantly changed in the tract-specific environment. One possible reason could be the low protein content of the pelletes (as a result of poor growth rate) which does not result in a comprehensive detection of all the significantly changed proteins.

As mentioned earlier in this chapter, I used the metabolic, transcriptome and proteome data to constrain the genome-scale model of *E. faecalis* and *S. pyogenes*. This workflow was designed to find tract-specific drug targets in the respective metabolic networks. Therefore, the metabolic data were integrated to each model independently to determine the state of metabolite uptake and production rate under each condition. Unfortunately, as a result of the poor quality of the data, a lot of adjustments to the flux bounds of the respective exchange reactions had to be made for the models to have a feasible solution. Next, I tried to integrate the transcriptome and proteome data individually to the models to observe the effect of integrated constraints on the models. This process made it clear that only the integration of the transcriptome data had a meaningful outcome. The reason is that there was hardly any protein responsible for the existing reactions in the metabolic networks that were present in the list of significant fold changes in the proteome profiles. Therefore, there was actually no constraint to be integrated. As also discussed earlier, the most possible explanation for that would be the low protein content of the samples, so many of the proteins were below the detection limit and therefore are absent from the statistical analysis. On the other hand, the profiles of the significant fold changes from the transcriptomics contained a large number of metabolic genes. Hence, integration of those sets of constraints affected the genome-scale models considerably.

During the process of drug target identification, I distinguished between essential reactions and drug targets. The use of essential reactions as drug targets would mean that a drug absolutely inhibits an enzyme, which in turn would absolutely kill the bacteria. This is not a very realistic assumption. There is always a limit in the level at which drugs inhibit a biological process. Moreover, the likelihood of a drug killing a micro organism in an absolute way is also very low. Therefore, a realistic assumption would be to consider a margin of dysfunction, not only at the level of enzyme inhibition, but also at the level of killing the pathogen. Having said that, I decided to consider three levels at which an enzyme can be inhibited by a drug, which would be reducing the enzyme activity to 30%, 20% 10% of its original activity. In the first scenario, I assumed that this inhibition absolutely kills the bacteria. Of course this is not realistic, but it brings more input to the context. In the second scenario, with the same aim of having more input, I assumed that a drug would entirely inhibit an enzyme and this reduction would reduce

the growth rate of the pathogen to 30%, 20% 10% of its original value. In the third attempt, which I referred to as the combo approach, I assumed that a drug would reduce the enzyme activity to its 30%, 20% 10% of its original value, and the attenuation of enzyme would reduce the growth rate of the bacteria to 30%, 20% 10% of its original value. From this perspective, a drug target (enzyme) with the need for a lower reduction of its activity, which in turn reduces the growth rate of bacteria more dramatically, would be a more efficient drug target.

I performed the explained processes for all the intended conditions. Not surprisingly, the size of the profiles of potential drug targets, with few exceptions, grew along with a more restricted inhibition, both at the level of enzyme inhibition and the growth inhibition. However, the margin was larger in the case of enzyme inhibition. In general, the libraries were in size of 50 to less than 300 reactions. The potential drug targets belong to different subsystems such as glycolysis, amino acid metabolism, etc. However, the identified drug targets which were previously reported to serve as drug targets, such as GAPDH, are more promising candidates [Gómez et al., 2019]. For each condition, there were a lot of reactions shared between the different libraries. The reason was that a lot of the enzymatic inhibitions resulted in zero growth rate in the system. Also, a lot of the reactions existed in the list of essential reactions too. The small margin of difference in the libraries when distinguishing between the growth rate reduction to 20% or 10%, points out to the fact that genome-scale models are not very flexible with respect to the reduction in enzyme activity. As the models in this study are not designed to simulate a community of bacterial cells, therefore, the system assumes that every single metabolite has to be produced internally (of course other than the ones that are defined by exchange reactions) and the metabolic communication under a normal physiological condition is not taken into account. However, these results give us a rough idea on which enzymes are the most likely successful candidates for inhibition.

Genome-scale metabolic models are widely used to find novel drug targets in various organisms. A common problem in this field is that these studies normally use the unrealistic assumption of full inhibition and full mortality. For instance, Paul and colleagues investigated potential drug targets for cancer by screening the gene knockouts [Paul et al., 2021]. Likewise, Larsson and colleagues investigated drug targets in the case of glioblastoma, assuming full inhibition of 5 gene products [Larsson et al., 2020]. In another attempt, Jerbyn and Ruppin investigated potential drug targets, again in cancer cell lines, by generating a library of synthetic lethal genes whose knockouts would kill the cancer cells [Jerby and Ruppin, 2012]. However, the relative inhibition has also been previously considered by the community. In a study aimed to provide more insight for personalized medicine, a human genome-scale model was used, in which a therapeutic window of 80% or lower rate of enzyme activity which reduces the growth rate to 45% to 75% of its original value is considered to be effective [Raškevičius et al., 2018]. Of course the margin used there is much larger than the one used in this study, which allows for the detection of more potential drug targets.

Genome-scale metabolic models can be used as powerful tools for identification of drug targets. Metabolic drug targets can be more efficient than the ones from the signalling pathways, which

are normally used in cancer therapies, due to the more conserved nature of metabolism [Locasale and Cantley, 2011]. However, the investigation of drug targets should be done by having realistic assumptions, sophesticated experimental design and experimental validation.

Chapter 5

CONCLUSION

In this thesis, I employed the genome-scale metabolic model of two human pathogens, *Enterococcus faecalis* and *Streptococcus pyogenes* to find tract-specific drug targets in their metabolic networks. The idea was promoted given the rising trend of multi-resistance to antibiotics in these two bacteria. First, In order to acquire a better understanding of the adaptive metabolism of *E. faecalis*, I investigated the effect of glutamine auxotrophy on the adaptation of this species to a declining environmental pH (from 7.5 to 6.5) at the metabolic and proteomic level. To get an overall understanding of its metabolic adjustments, I integrated the experimental data into the genome scale metabolic model of a Δ*glnA* mutant of *E. faecalis*. The genome-scale model was successful in reproducing a large part of the experimental data, which was proven by the comparison to the results of the wildtype. The integrative study resulted in a new design for the transport system in the genome-scale model, which accounts for the different level of affinity among the amino acids for the same transporter. This way, the model could successfully reflect the metabolic and proteome data into the flux distribution of the network. I also suggested that the absence of glnA from the genome of the mutant potentially results in the lost control of the glutamine uptake system, which further leads to a higher energy demand in the mutant compared to the wildtype. In the next step, I developed a method for functional analysis of the solution space of genome-scale metabolic models. The large solution space of genome-scale models is considered as a major drawback of this modelling formalism, with part of the flux distribution profiles being contradictory to the actual biological phenotypes. The method presented in this thesis enables one to identify the number of sensitive reactions in a metabolic network, which I showed to be an informative indicator of the size of the solution space. The method also allows for studying the flux distribution profiles at the branching points in the metabolic network, and also to determine how effectively each set of experimental data shrinks the solution space. In the last part, I used several libraries of multi-omics data (metabolic, transcriptome and proteome) to develop tract-specific genome-scale models of *E. faecalis* and *S. pyogenes*. The tract-specific models were then used to find tract-specific drug targets. As the quality of the metabolic data was not very high, an improvement in that part might increase the reliability of the results. The transcriptome data contained information regarding a large number of enzymes under all conditions and therefore had a meaningful effect in constraining the models, although with many adjustments being necessary. The proteome data in the case of the both bacteria did

not contain any information about metabolic enzymes, therefore, was not effective in drug target identification. For each of the bacteria, several libraries of potential drug targets were generated. The drug targets have to be experimentally validated to prove their effectiveness. Overall, the reliability of the suggested libraries of drug targets can be increased by improving the quality of the experimental data which were used as constraints for the genome-scale models.

Appendix A

Supplementary materials

Table A.1: Time course data showing the flux measurement of glucose, organic acids and amino acids in the course of 21 hours (t1-t8) in chemostat cultures of Δ*glnA* mutant of *E. faecalis*. The table is adapted from [Loghmani et al., 2021]

Exchange reaction	t1	t2	t3	t4	t5	t6	t7	t8
R_Ex_glc__e	-7.170	-7.036	-7.455	NA	-7.560	NA	-7.881	-9.593
R_Ex_etoh__e	2.157	2.035	2.134	NA	1.899	NA	1.689	1.490
R_Ex_ac__e	4.037	4.385	4.161	NA	2.931	NA	3.314	2.568
R_Ex_lac_L__e	5.878	6.467	7.671	NA	7.339	NA	10.720	15.026
R_Ex_for__e	7.109	7.522	7.262	NA	5.506	NA	5.570	4.226
R_Ex_asp_L__e	-0.127	-0.117	-0.121	-0.105	-0.102	-0.108	-0.092	-0.095
R_Ex_ser_L__e	-0.289	-0.283	-0.296	-0.292	-0.295	-0.317	-0.312	-0.387
R_Ex_glu_L__e	-0.057	-0.045	-0.043	-0.020	-0.014	-0.014	0.010	0.014
R_Ex_gly__e	-0.085	-0.080	-0.087	-0.075	-0.073	-0.077	-0.062	-0.081
R_Ex_his_L__e	-0.016	-0.012	-0.012	-0.006	-0.004	-0.005	0.002	-0.004
R_Ex_nh4__e	0.253	0.274	0.290	0.338	0.363	0.387	0.409	0.530
R_Ex_arg_L__e	-0.284	-0.279	-0.295	-0.294	-0.300	-0.318	-0.312	-0.380
R_Ex_thr_L__e	-0.054	-0.050	-0.054	-0.045	-0.043	-0.045	-0.034	-0.051
R_Ex_ala_L__e	0.060	0.065	0.059	0.080	0.088	0.093	0.117	0.124
R_Ex_pro_L__e	-0.064	-0.057	-0.074	-0.045	-0.039	-0.047	-0.014	-0.043
R_Ex_cys_L__e	0.035	0.034	0.038	0.045	0.045	0.050	0.054	0.068
R_Ex_tyr_L__e	-0.032	-0.022	-0.031	-0.034	-0.013	-0.055	-0.046	-0.045
R_Ex_val_L__e	-0.084	-0.078	-0.085	-0.072	-0.069	-0.074	-0.056	-0.074
R_Ex_met_L__e	-0.021	-0.019	-0.020	-0.017	-0.016	-0.017	-0.013	-0.019
R_Ex_orn_L__e	0.197	0.198	0.207	0.223	0.233	0.246	0.261	0.316
R_Ex_lys_L__e	-0.075	-0.068	-0.076	-0.058	-0.052	-0.057	-0.035	-0.062
R_Ex_ile_L__e	-0.076	-0.073	-0.076	-0.069	-0.067	-0.070	-0.060	-0.073
R_Ex_leu_L__e	-0.134	-0.125	-0.131	-0.114	-0.108	-0.112	-0.087	-0.106
R_Ex_phe_L__e	-0.046	-0.042	-0.046	-0.038	-0.034	-0.040	-0.029	-0.041
R_Ex_asn_L__e	-0.051	-0.049	-0.051	-0.046	-0.046	-0.047	-0.041	-0.052
R_Ex_gln_L__e	-0.147	-0.144	-0.153	-0.152	-0.155	-0.164	-0.162	-0.197
R_Ex_trp_L__e	-0.006	-0.005	-0.006	-0.006	-0.005	-0.006	-0.004	-0.006
R_Ex_citr_L__e	0.022	0.022	0.021	0.022	0.023	0.022	0.023	0.019
R_Ex_cys_L__e	-0.062	-0.060	-0.061	-0.057	-0.055	-0.055	-0.050	-0.053

Table A.2: Metabolic constraints in the form of lower and upper bounds at pH 7.5. in the genome-scale metabolic model of Δ*glnA E. faecalis*. The table is adapted from [Loghmani et al., 2021]

Exchange reaction	experimental value	calculated lower bound	calculated upper bound	adjusted lower bound	adjusted upper bound
R_Ex_glc__e	-7.170	-8.604	-5.736.000.000.000.000		
R_Ex_etoh__e	2.157	17.256	25.884		
R_Ex_ac__e	4.037	32.296	4.844.399.999.999.990		
R_Ex_lac_L__e	5.878	47.024	70.536		
R_Ex_for__e	7.109	5.687.200.000.000.000	85.308		
R_Ex_asp_L__e	-0.127	-0.1524	-0.10160000000000001		
R_Ex_ser_L__e	-0.289	-0.34679999999999994	-0.2312		
R_Ex_glu_L__e	-0.057	-0.0684	-0.0456		-0.04
R_Ex_gly__e	-0.085	-0.10200000000000001	-0.068		
R_Ex_his_L__e	-0.016	-0.0192	-0.0128	-0.013	0.0
R_Ex_nh4__e	0.253	0.20240000000000002	0.3036		
R_Ex_arg_L__e	-0.284	-0.34079999999999994	-0.22719999999999999		-0.2372
R_Ex_thr_L__e	-0.054	-0.0648	-0.0432		
R_Ex_ala_L__e	0.060	0.048	0.072		0.06
R_Ex_pro_L__e	-0.064	-0.0768	-0.0512	-0.052	0.0
R_Ex_cys_L__e	0.035	0.028000000000000004	0.042		
R_Ex_tyr_L__e	-0.032	-0.0384	-0.0256		
R_Ex_val_L__e	-0.084	-0.1008	-0.06720000000000001		
R_Ex_met_L__e	-0.021	-0.0252	-0.016800000000000002		
R_Ex_orn_L__e	0.197	0.15760000000000002	0.2364		0.2464
R_Ex_lys_L__e	-0.075	-0.09	-0.06		
R_Ex_ile_L__e	-0.076	-0.09119999999999999	-0.0608		
R_Ex_leu_L__e	-0.134	-0.1608	-0.10720000000000002		
R_Ex_phe_L__e	-0.046	-0.0552	-0.0368		
R_Ex_asn_L__e	-0.051	-0.06119999999999999	-0.0408		
R_Ex_gln_L__e	-0.147	-0.17639999999999997	-0.1176		
R_Ex_trp_L__e	-0.006	-0.0072	-0.0048000000000000004		
R_Ex_citr_L__e	0.022	0.0176	0.026399999999999996		0.044
R_Ex_cys_L__e	-0.062	-0.0744	-0.049600000000000005		0.032

Table A.3: Metabolic constraints in the form of lower and upper bounds at pH 7.5. in the genome-scale metabolic model of Δ*glnA E. faecalis*. The table is adapted from [Loghmani et al., 2021]

Exchange reaction	Experimental value	Calculated lower bound	Calculated upper bound	Adjusted lower bound	Adjusted upper bound
R_Ex_glc__e	-9.593	-115.116	-76.744		
R_Ex_etoh__e	1.490	1.192	1.788		
R_Ex_ac__e	2.568	20.544.000.000.000.000	30.816		
R_Ex_lac_L__e	15.026	12.020.800.000.000.000	180.312		
R_Ex_for__e	4.226	33.808.000.000.000.000	50.712		
R_Ex_asp_L__e	-0.095	-0.11399999999999999	-0.07600000000000001		
R_Ex_ser_L__e	-0.387	-0.4644	-0.30960000000000004		
R_Ex_glu_L__e	0.014	0.011200000000000002	0.0168		
R_Ex_gly__e	-0.081	-0.0972	-0.06480000000000001		
R_Ex_his_L__e	-0.004	-0.0048	-0.0032	-0.0062	
R_Ex_nh4__e	0.530	0.42400000000000004	0.636		
R_Ex_arg_L__e	-0.380	-0.45599999999999996	-0.30400000000000005		
R_Ex_thr_L__e	-0.051	-0.06119999999999999	-0.0408		
R_Ex_ala_L__e	0.124	0.09920000000000001	0.1488		
R_Ex_pro_L__e	-0.043	-0.05159999999999999	-0.0344	-1	3
R_Ex_cys_L__e	0.068	0.054400000000000004	0.0816		
R_Ex_tyr_L__e	-0.045	-0.054	-0.036		
R_Ex_val_L__e	-0.074	-0.08879999999999999	-0.0592		
R_Ex_met_L__e	-0.019	-0.022799999999999997	-0.0152		
R_Ex_orn_L__e	0.316	0.2528	0.3792		
R_Ex_lys_L__e	-0.062	-0.0744	-0.049600000000000005		
R_Ex_ile_L__e	-0.073	-0.0876	-0.0584		
R_Ex_leu_L__e	-0.106	-0.12719999999999998	-0.0848		
R_Ex_phe_L__e	-0.041	-0.0492	-0.0328		
R_Ex_asn_L__e	-0.052	-0.0624	-0.0416		
R_Ex_gln_L__e	-0.197	-0.2364	-0.15760000000000002		
R_Ex_trp_L__e	-0.006	-0.0072	-0.0048000000000000004		
R_Ex_citr_L__e	0.019	0.0152	0.022799999999999997		
R_Ex_cys_L__e	-0.053	-0.06359999999999999	-0.0424		0

Table A.4: List of inactivated genes following the integration of proteome data at pH 7.5 from the pH shift experiment. The table is adapted from [Loghmani et al., 2021]

Reactions	**Genes**	**Reactions**	**Genes**
R_TAL	EF3304	R_FRDx	EF2566
R_LDH_D	EF2295	R_ARABte	EF2773
R_MALt	EF1920	R_PNTOt2	EF2657
R_SERt6	EF3015	R_THMASE	EF2767
R_DALAt	EF1103	R_RMK	EF0433
R_TMDPK	EF3117	R_RMI	EF0434
R_RIBFLVt2	EF1541	R_GARFT	EF1779
R_HXANt2	EF2935	R_VANX	EF2293
R_MENE	EF0446	R_GLNS	EF2159
R_ARGORNt	EF0108	R_NH4t	EF0547
R_LYSt6	EF0929	R_UNAGAMAMT	EF1173
R_XYLB	EF0557	R_RAFDH	EF1603
R_XYLA	EF0556	R_CYSTGL	EF3284
R_XYLI	EF0557	R_DSERt	EF1103
R_FUMt	EF1920	R_PIht6	EF2442
R_PTB	EF1663	R_RIBt2	EF2959

Table A.5: List of proteomic constraints in the form of reactions upper and lower bounds at pH 6.5 from the pH shift experiment. The table is adapted from [Loghmani et al., 2021]

Reactions	Adjusted lower bound	Adjusted upper bound
R_PRFGS	0	0
R_SPMDabc	0	0
R_PDE1	0.0	0.0
R_PDE2	0.0	0.0
R_PDE3	0.0	0.0
R_PDE4	0.0	0.0
R_PDE5	0.0	0.0
R_NDPK9	0.0	0.0
R_NDPK8	0.0	501130
R_NDPK7	0.0	0.0010203525938033196
R_NDPK6	0.0	0.0013224859790384868
R_NDPK5	0.0	0.0010203525938033196
R_GLNabc	0.0	0.06837444461554447
R_GMPR	0.0	0.000768957299464954
R_XANt2	0.002012843715119238	0.005257190576108279
R_IMPC	0.0	0.0
R_ME1x	0.0	0.0
R_AIRC	0.0	0.0
R_FMETDF	0.0	0.0
R_TREpts	0.0	0.0
R_ADSS	0.0	0.0
R_GLUPRT	0.0	0.0
R_URAt2	0.0003167241962024313	0.011771168317517122
R_AICART	0.0	0.0
R_XPPT	0.002720007446909017	0.007104176766568783
R_PMANM	0.0	0.0
R_PTRCabc	0.0	0.0
R_CDD	0.0	0.0
R_GLYK	0.0	0.0024090417384447812
R_PRAGS	0.0	0.0
R_MALPP	0.0	0.0
R_OAADC	0.0	0.011374005166922309
R_PRAIS	0.0	0.0

Table A.6: List of missing essential genes from the experimentally detected protein library in the pH shift experiment. The table is adapted from [Loghmani et al., 2021]

Genes	**Reactions**
EF0168	R_PNTK
EF0739	R_NACUP, R_NCAMRNSt
EF0904	R_MEVK
EF0992	R_PAPPT1_A, R_PAPPT1_L
EF1391	R_PZS
EF1392	R_PZS
EF1393	R_PZS
EF1396	R_PZS, R_Moabc
EF2183	R_LTAS2
EF2294	R_VANB
EF2411	R_DAGK_LPL
EF2439	R_UDCPK, R_UDCPDP
EF2494	R_DASYN_LPL
EF2661	R_DAGK_LPL
EF2746	R_DARTAL_EFA, R_DALTAL_EFA
EF2748	R_DARTAL_EFA, R_DALTAL_EFA
EF2973	R_ALKP_Efa
EF3072	R_BTNt2i

Table A.7: List of essential reactions (reactions that are necessary for the Δ*glnA E. faecalis* model to have a feasible solution). The table is adapted from [Loghmani et al., 2021]

R_ACACT1	R_UAGPT1_A	R_DPMVD	R_HBUHL1	R_TMDS
R_ACCOAC	R_UAGPT1_L	R_DTMPK	R_HBUR1	R_TRDR
R_AGAT_EFA	R_UAMAGS	R_EPA_PS_EFA	R_HCO3E	R_TRPTRS
R_ALAR	R_UAMAS	R_Ex_arg_L__e	R_HDDHL5	R_TYRTRS
R_ALATA_Lr	R_UAPGR	R_Ex_btn__e	R_HDDR5	R_HISTRS
R_ALATRS	R_UDCPDP	R_Ex_glyclt__e	R_HDEHL4	R_HMGCOAR
R_ALKP_Efa	R_UDCPDPS	R_Ex_his_L__e	R_HDER4	R_HMGCOAS
R_ARGTRS	R_UDPG4E	R_Ex_ile_L__e	R_HDMAT7	R_HOCHL3
R_ASPTRS	R_UGLDDS1_A	R_Ex_leu_L__e	R_HEMAT2	R_HOCR3
R_BIOMASS	R_UGLDDS1_L	R_Ex_nac__e	R_HHDHL7	R_HODHL8
R_BPPA1	R_URIDK1	R_Ex_pydam__e	R_HHDR7	R_HODR8
R_BPPA1_L	R_VALt6	R_Ex_thm__e	R_HHYHL2	R_HPPK
R_BPPA2	R_VALTRS	R_Ex_trp_L__e	R_HHYR2	R_HTDHL6
R_BPPA2_L	R_VANB	R_Ex_val_L__e	R_HISt6	R_HTDR6
R_BTMAT1	R_WTASI	R_FABM1	R_MCMAT6	R_ILEt6
R_BTNt2i	R_WTASII	R_FABM2	R_MCMAT7	R_ILETRS
R_CLPNS_LPL	R_DARTAL_EFA	R_G1PACT	R_MCMAT8	R_IPDDI
R_CPS_EFA_SYNTH	R_DASYN_LPL	R_G1PTMT	R_PYDAMt	R_kaasIII
R_CPS_PS_EFA	R_DDL	R_GALU	R_PYDXK	R_LEUt6
R_CRCT	R_DDMAT5	R_GAT1_EFA	R_PZS	R_LEUTRS
R_CYSTRS	R_DEMAT4	R_GCALDD	R_RBT5PDHy	R_LPGS_EFA
R_CYTK1	R_DEX_PS_EFA	R_GF6PTA	R_RNAS_LPL	R_LTAS1
R_DAGGT_LPL	R_DHFR	R_GLNTAL	R_SERTRS	R_LTAS2
R_DAGK_LPL	R_DHFS	R_GLUR	R_TAPGL4_EFA	R_LYSTRS
R_DALTAL_EFA	R_DHNPA	R_GLUTRS	R_TDMAT6	R_MACPMT
R_UAAGLS1	R_DHPS3	R_GLYCLTt2r	R_TDPDRE	R_MCMAT2
R_UACGE	R_DMATT	R_GLYTRS	R_TDPGDH	R_MCMAT3
R_UAGCVT	R_DNAS_LPL	R_GRTT	R_THMabc	R_MCMAT4
R_UAGDP	R_DPCOAK	R_GTPCI	R_THRTRS	R_MCMAT5
R_LEUt6	R_LEUTRS	R_LPGS_EFA	R_LTAS1	R_LTAS2
R_LTAS2	R_LYSTRS	R_MACPMT	R_MCMAT2	R_MCMAT3
R_MCMAT4	R_MCMAT5			

Table A.8: List of essential genes and their respective enzymes in the Δ*glnA E. faecalis* model. The table is adapted from [Loghmani et al., 2021]

Genes	**Enzyme**
EF0043	Glutamyl-tRNA synthetase
EF0045	Cysteinyl-tRNA synthetase (Cysteine–tRNA ligase) (CysRS)
EF0059	glucosamine-1-phosphate-N-acetyltransferase
EF0090	Diacylglycerol kinase (LPL specific)
EF0168	pantothenate kinase
EF0268	Lysyl-tRNA synthetase
EF0680	peptidoglycan glycosyltransferase
EF0724	glutamyl-tRNA(Gln):L-glutamine amido-ligase (ADP-forming)
EF0725	glutamyl-tRNA(Gln):L-glutamine amido-ligase (ADP-forming)
EF0726	glutamyl-tRNA(Gln):L-glutamine amido-ligase (ADP-forming)
EF0727	Diacylglycerol kinase (LPL specific)
EF0739	Nicotinic acid uptake
EF0770	Phosphatidylglycerol phosphate phosphatase (LPL specific)
EF0801	Leucyl-tRNA synthetase
EF0843	D-alanine—D-alanine ligase (reversible) (D-Ala-D-Ala ligase) (D-alanylalanine synthetase)
EF0845	UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-L-lysyl-D-alanyl-D-alanine synthetase (alpha-glutamate)
EF0849	alanine racemase
EF0880	dephospho-CoA kinase
EF0901	isopentenyl-diphosphate D-isomerase
EF0902	phosphomevalonate kinase
EF0903	diphosphomevalonate decarboxylase
EF0904	mevalonate kinase
EF0930	Methionyl-tRNA synthetase

EF0981	geranyltranstransferase
EF0991	peptidoglycan glycosyltransferase
EF0992	phospho-N-acetylmuramoyl-pentapeptide-transferase (alpha-glutamate) (D-ala)
EF0993	UDP-N-acetylmuramoyl-L-alanyl-D-glutamate synthetase
EF0994	UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide)pyrophosphoryl-undecaprenol N-acetylglucosamine transferase
EF1003	Isoleucyl-tRNA synthetase
EF1036	nucleoside-diphosphate kinase (ATP:GDP)
EF1115	Phenylalanyl-tRNA synthetase (Phenylalanyl-tRNA synthetase alpha subunit)
EF1116	Phenylalanyl-tRNA synthetase (Phenylalanyl-tRNA synthetase beta subunit)
EF1121	glutamate racemase
EF1148	peptidoglycan glycosyltransferase
EF1364	Hydroxymethylglutaryl CoA reductase
EF1379	Alanyl-tRNA synthetase (AlaRS)
EF1391	precursor Z synthase
EF1392	precursor Z synthase
EF1393	precursor Z synthase
EF1396	precursor Z synthase
EF1547	cytidylate kinase (dCMP)
EF1576	thymidylate synthase
EF1577	dihydrofolate reductase
EF1711	carbonate dehydratase (HCO3 equilibration reaction)
EF1746	UTP-glucose-1-phosphate uridylyltransferase
EF1908	UDP-N-acetylmuramoyl-L-alanine synthetase
EF1970	Aspartyl-tRNA synthetase (AspRS)
EF1971	Histidyl-tRNA synthetase
EF2150	UDP-N-acetylglucosamine 4-epimerase

EF2162	dimethylallyltranstransferase
EF2182	Lipoteichoic acid synthase (LPL specific)
EF2183	Lipoteichoic acid modification (koijbiose)
EF2192	dTDPglucose 4
EF2193	dTDP-4-dehydromannose 3
EF2194	glucose-1-phosphate thymidylyltransferase
EF2294	D-alanine—(R)-lactate ligase
EF2379	Prolyl-tRNA synthetase
EF2396	uridylate kinase (UMP)
EF2406	Glycyl-tRNA synthetase
EF2407	Glycyl-tRNA synthetase
EF2411	Diacylglycerol kinase (LPL specific)
EF2439	Undecaprenyl-diphosphatase
EF2451	pantetheine-phosphate adenylyltransferase
EF2471	Arginyl-tRNA synthetase (ArgRS)
EF2476	peptidoglycan glycosyltransferase
EF2494	CDP-Diacylglycerol synthetase (Phosphatidate cytidylyltransferase) (LPL specific)
EF2495	Undecaprenyl diphosphate synthase
EF2644	Diacylglycerol kinase (LPL specific)
EF2655	phosphopantothenate-cysteine ligase
EF2656	phosphopantothenoylcysteine decarboxylase
EF2658	UDP-N-acetylmuramoylpentapeptide-lysine-N6-alanyltransferase
EF2661	Diacylglycerol kinase (LPL specific)
EF2691	1-Acyl-glycerol-3-phosphate acyltransferase (Efa specific)
EF2746	D-Alanine lipoteichoic acid ligase
EF2747	D-Alanine lipoteichoic acid ligase

EF2748	D-Alanine lipoteichoic acid ligase
EF2749	D-Alanine lipoteichoic acid ligase
EF2764	Thymidylate kinase (dTMP kinase)
EF2858	Threonyl-tRNA synthetase
EF2871	nicotinate-nucleotide adenylyltransferase (nicotinate)
EF2875	acetyl-CoA carboxylase
EF2876	acetyl-CoA carboxylase
EF2877	acetyl-CoA carboxylase
EF2879	acetyl-CoA carboxylase
EF2882	Malonyl-CoA:[acyl-carrier-protein] S-malonyltransferase
EF2883	Tetradecanoyl-[acyl-carrier protein]:malonyl-CoA C-acyltransferase
EF2885	beta-ketoacyl-ACP synthase III
EF2890	1
EF2891	1
EF2908	peptidoglycan glycosyltransferase
EF2928	dihydrofolate synthase
EF2931	Valyl-tRNA synthetase
EF2973	alkaline phosphatase
EF3072	Biotin uptake
EF3112	glycerol 3-phosphate acyltransferase (Efa specific)
EF3129	peptidoglycan glycosyltransferase
EF3148	lysylphosphatidyl-glycerol synthetase
EF3265	dihydropteroate synthase
EF3267	GTP cyclohydrolase I
EF3268	2-amino-4-hydroxy-6-hydroxymethyldihydropteridine diphosphokinase
EF3269	dihydroneopterin aldolase

Table A.9: Significant fold changes at the proteome level between t3 and t1. The table is adapted from [Loghmani et al., 2021]

Reaction abbreviation	Change	p-value
R_CTPS2	1.63890193589	0.04828668
R_PMEVK	1.97630931519	0.049963774
R_TAGA	1.71091285581	0.048401574
R_FMNAT	1.53450840082	0.018418951
R_NNAMr	0.579906671149	0.043743765
R_RNDR2	0.113556516552	0.014524055
R_ASPCT	0.540883578382	0.014436519
R_DEX_PS_EFA	0.677245103099	0.034169764
R_G3PO	2.93224651611	0.049505943
R_OCBT	0.540883578382	0.014436519
R_GMPR	0.580492532748	0.044735034
R_SBTpts	0.208542436348	0.010564345
R_DHNAS	1.54800580009	0.042118513
R_PGL	0.620953374912	0.016626571
R_MALTabc	2.71542379428	0.014436519
R_ARGD	1.57686937955	0.038619556
R_TYRTA	1.83546731126	0.030302482
R_CDPMEK	2.9629291135	0.042488016
R_RNDR4	0.113556516552	0.014524055
R_RNDR3	0.113556516552	0.014524055
R_PGPP_LPL	0.486269996416	0.014436519
R_TAPGL4_EFA	0.540883578382	0.014436519
R_DAPDC	1.60416754956	0.031295061
R_DHFOR2	0.444448732256	0.042674702
R_TDPDRE	0.526923889232	0.040338473
R_UACGE	0.677245103099	0.034169764
R_DHFR	0.444448732256	0.042674702
R_VALTRS	2.33769302453	0.032222324
R_GTHS	2.02787326816	0.031295061
R_CDD	0.486664770062	0.00041297
R_RNDR1	0.113556516552	0.014524055
R_CELBpts	0.532183444906	0.004964827
R_CELBpts	0.532183444906	0.004964827
R_BPPA2_L	0.677245103099	0.034169764
R_RPI	2.73601994462	0.043937668
R_GLUCYSL	2.02787326816	0.031295061
R_EPA_PS_EFA	0.677245103099	0.034169764
R_ALATA_L	2.46626089459	0.04926589
R_ACTPASE	0.543284194457	0.014436519
R_BPPA2	0.677245103099	0.034169764
R_PANB	1.54123791035	0.031265045
R_G1PACT	2.26209396876	0.029276069
R_UAGDP	2.26209396876	0.029276069
R_RBFK	1.53450840082	0.018418951
R_MTHPTGHM	2.01200124101	0.045724593
R_GHMT	2.31513631222	0.031295061

Table A.10: Significant fold changes at the proteome level between t4 and t1. The table is adapted from [Loghmani et al., 2021]

Reaction abbreviation	Change	p-value
R_PGL	0.5738456585	0.019662016
R_CDD	0.58433750945	0.008558951
R_ACTPASE	0.543462288312	0.031954297

Table A.11: Significant fold changes at the proteome level between t5 and t1. The table is adapted from [Loghmani et al., 2021]

Reaction abbreviation	Change	P-value
R_CDD	0.584205462912	0.011112227
R_ACTPASE	0.523644783934	0.025410243

Table A.12: Significant fold changes at the proteome level between t6 and t1. The table is adapted from [Loghmani et al., 2021]

Reaction abbreviation	Change	P-value
R_CDD	0.521029799606	0.001808714
R_ACTPASE	0.477016117438	0.009193028

Table A.13: Significant fold changes at the proteome level between t7 and t1. The table is adapted from [Loghmani et al., 2021]

Reaction abbreviation	Change	P-value
R_CDD	0.526602135869	0.004403881
R_ACTPASE	0.544365914365	0.032780414

Table A.14: Significant fold changes at the proteome level between t8 and t1. The table is adapted from [Loghmani et al., 2021]

Reaction abbreviation	Changes	p-value
R_PRPPS	0.702185299835	0.01759182
R_PRFGS	0.294343721566	0.000171526
R_SPMDabc	0.128708671432	0.000891361
R_PDE1	0.579410399883	0.011248306
R_PDE2	0.579410399883	0.011248306
R_PDE3	0.579410399883	0.011248306
R_PDE4	0.579410399883	0.011248306
R_PDE5	0.579410399883	0.011248306
R_NDPK9	0.357950631741	0.042018929
R_NDPK8	0.357950631741	0.042018929
R_NDPK7	0.357950631741	0.042018929
R_NDPK6	0.357950631741	0.042018929
R_NDPK5	0.357950631741	0.042018929
R_NDPK4	0.357950631741	0.042018929
R_NDPK3	0.357950631741	0.042018929
R_NDPK2	0.357950631741	0.042018929
R_NDPK1	0.357950631741	0.042018929
R_GLNabc	1.55907771659	0.003943372
R_GMPR	0.457885582652	0.006045753
R_XANt2	0.379306679373	0.011248306
R_IMPC	0.116277843614	1.52e-05
R_ME1x	0.615708970929	0.018231213
R_AIRC	0.228685137252	0.000929637
R_FMETDF	1.6307598154	0.023588934
R_PGAMT	0.355650632158	0.009695126
R_TREpts	0.280736842019	5.44e-05
R_ADSS	0.406572914915	2.98e-05
R_GLUPRT	0.337390147352	0.023338551
R_URAt2	0.379306679373	0.011248306
R_AICART	0.116277843614	1.52e-05
R_XPPT	0.512566866275	0.0159729
R_PMANM	0.355650632158	0.009695126
R_PTRCabc	0.128708671432	0.000891361
R_PGMT	0.355650632158	0.009695126
R_CDD	0.395281255682	8.09e-06
R_GLYK	0.591527293003	0.026602872
R_PRAGS	0.366141109905	0.024864028
R_MALPP	0.564735962708	0.023338551
R_OAADC	0.615708970929	0.018231213
R_PRAIS	0.331714599531	0.010924227

Table A.15: Flux measurement of glucose, organic acids and amino acids at two dilatation rates, 0.15 and 0.05 h-1 at pH 7.5 and 6.5. The data is used for the estimation of ATP maintenance. The table is adapted from [Loghmani et al., 2021]

Reactions	**Experimental value**			
	7,5, 0,15	6,5, 0,15	7,5 0,05	6,5, 0,05
R_Ex_glc__e	-6,3230357	-8,74125	-4,1436652	-5,2328571
R_Ex_etoh__e	3,00853211	1,31258619	1,954819	1,43534594
R_Ex_ac__e	3,24047722	1,6646415	2,34504525	1,81889951
R_Ex_lac_L__e	7,54292064	15,9580439	2,60823222	5,56738437
R_Ex_for__e	5,57835409	2,79945752	4,41502043	3,30368615
Succinate	0	0	0	0
R_Ex_pyr__e	0,10182884	0,10656909	0,09474662	0,12763664
Acetoin	0	0	0	0
2.3-butanediol	0	0	0	0
CO2	0	0	0	0
O2	0	0	0	0
R_Ex_asp_L__e	-0,1308159	-0,137878	-0,1313828	-0,1505571
R_Ex_ser_L__e	-0,3023771	-0,3221543	-0,1854702	-0,2508137
R_Ex_glu_L__e	-0,0824365	-0,0705838	-0,0275192	-0,0733358
R_Ex_gly__e	-0,0953275	-0,0597635	-0,0581162	-0,0977894
R_Ex_his_L__e	-0,0333229	-0,0289961	-0,0186223	-0,0300862
R_Ex_nh4__e	-0,5438415	-0,751831	-0,3563948	-0,4500757
R_Ex_arg_L__e	-0,2994718	-0,4119674	-0,121859	-0,2481603
R_Ex_thr_L__e	-0,0737675	-0,0530418	-0,0425567	-0,0646863
R_Ex_ala_L__e	0,00692002	0,15424036	-0,000855	0,00952816
R_Ex_pro_L__e	0,0554295	0,16986676	-0,0973449	-0,1101082
R_Ex_cys_L__e	-0,0859643	-0,1188409	-0,0563348	-0,0711429
R_Ex_tyr_L__e	-0,0770671	-0,1081494	-0,0298488	-0,0621639
R_Ex_val_L__e	-0,1156056	-0,1069205	-0,0658251	-0,1054554
R_Ex_met_L__e	-0,0339244	-0,038035	-0,0570136	-0,0515313
Ornithine	0	0	0	0
R_Ex_lys_L__e	-0,043371	-0,0116138	-0,0139551	-0,0388254
R_Ex_ile_L__e	-0,0906547	-0,0875255	-0,0409405	-0,0770522
R_Ex_leu_L__e	-0,1538447	-0,1398334	-0,0807965	-0,1387774
R_Ex_phe_L__e	-0,0653184	-0,078708	-0,0400491	-0,0649809
R_Ex_trp_L__e	-0,0071603	-0,0063578	-0,0165611	-0,007572
R_Ex_asn_L__e	-0,0527041	-0,0388719	-0,0513122	-0,0335545
R_Ex_gln_L__e	-0,1416857	-0,1958727	-0,0928507	-0,1172571

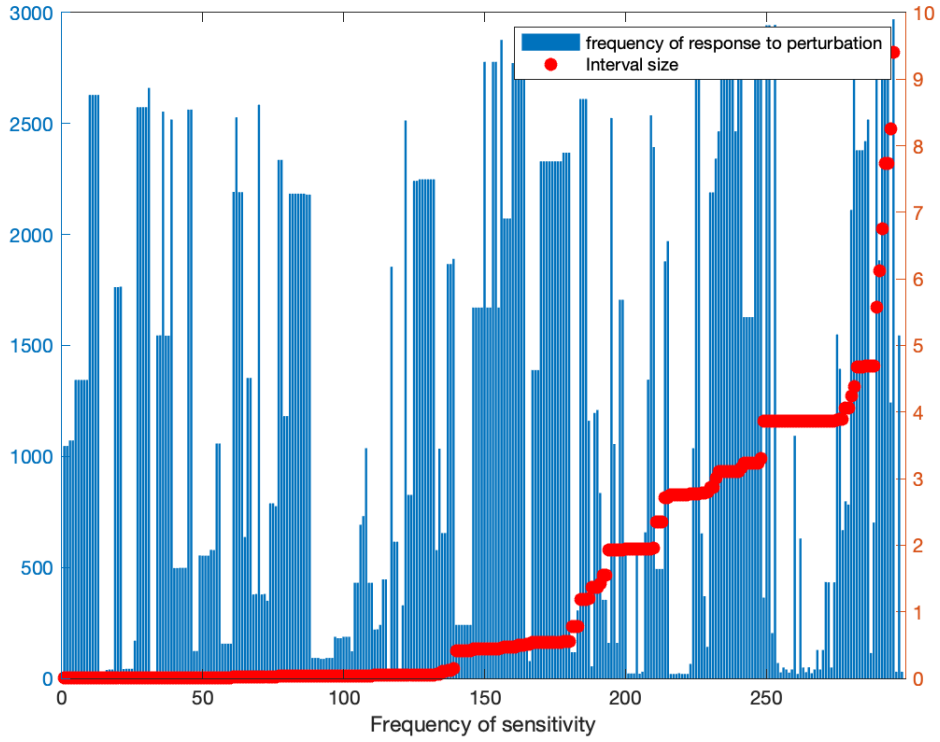

Figure A.1: Comparison between the frequency of response to perturbation (sensitivity frequency) and the FVA interval size in the mutant model integrated with medium composition, metabolic and proteome data. The graph shows that the frequency of sensitivity is not correlated with the FVA interval size. The figure is adapted from [Loghmani et al., 2022]

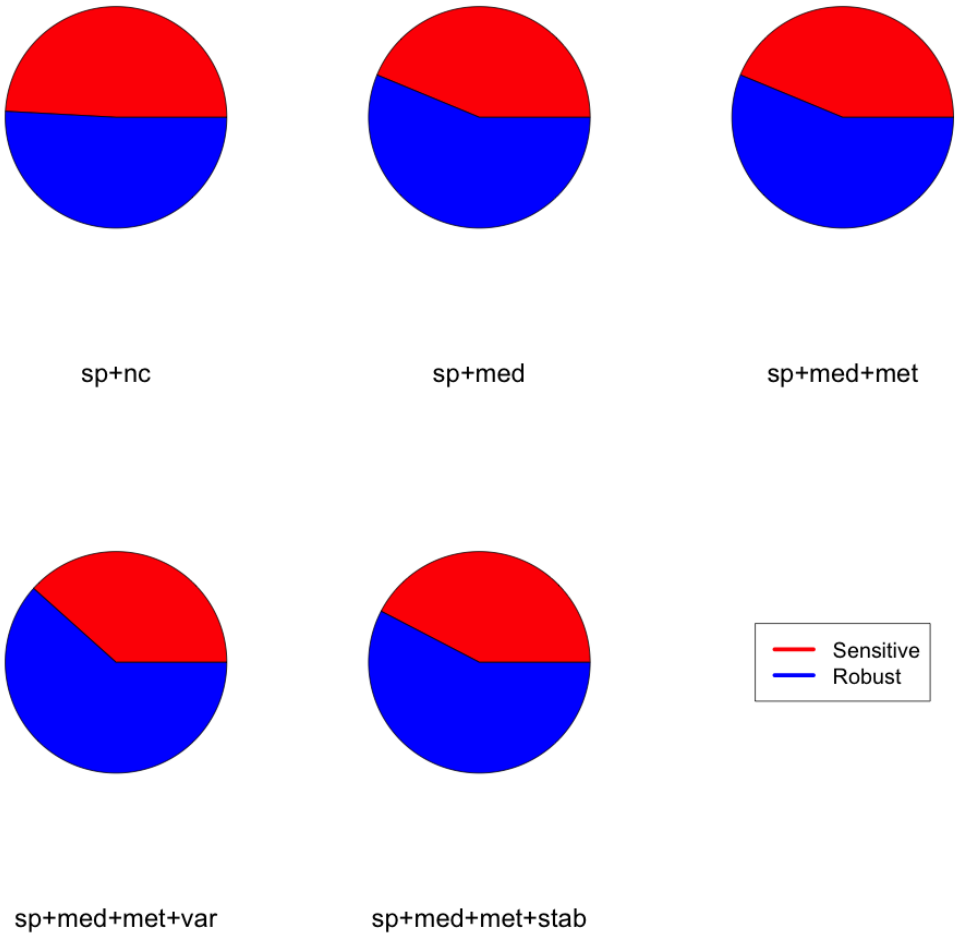

Figure A.2: The proportion of sensitive reactions with respect to perturbations in other reactions in the genome-scale models of *E. faecalis* wildtype (wt) and ΔglnA mutant (mt), when the perturbation procedure was performed with opt-percentage of 99.9 in FVA. The integration of constraints (none, medium composition, metabolic and proteome data, from left to right) into the model results in reducing the number of sensitive reactions (red) and increasing the number of robust reactions (blue). The figure is adapted from [Loghmani et al., 2022]

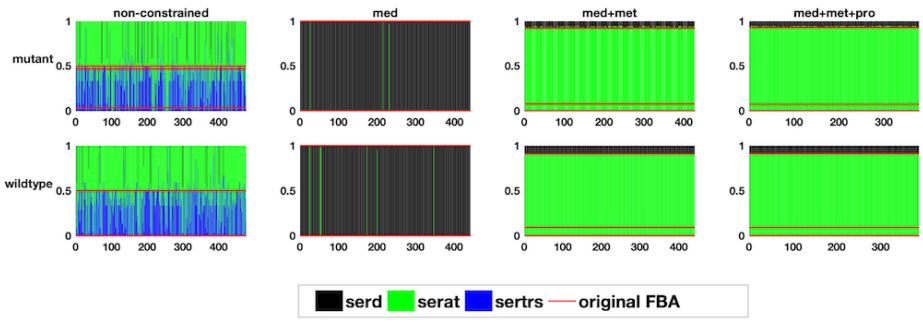

Figure A.4: The relative flux distribution at the branching point in serine metabolism in the two studied genome-scale models of *E. faecalis* with an opt-percentage of 99.9 in FVA, resulting in the production of acetyl serine, seryl-tRNA or serine secretion. The figure is adapted from [Loghmani et al., 2022]

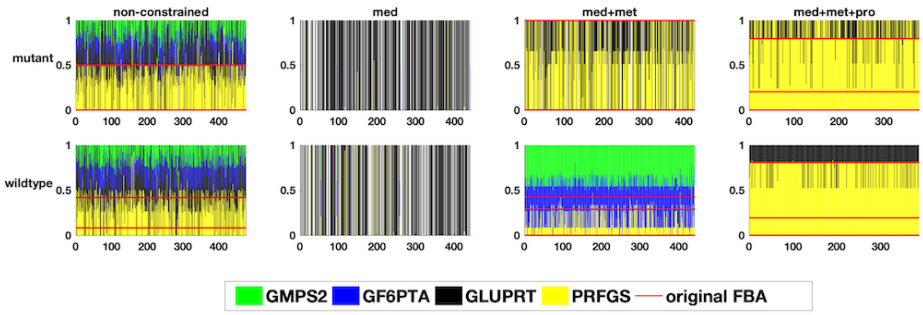

Figure A.5: The relative flux distribution at the branching point in glutamine metabolism in the two studied genome-scale models of *E. faecalis* with opt-percentage of 99.9 in FVA, resulting in the distribution of glutamine in different pathways, namely amino acid, purine and pyrimidine metabolism. The figure is adapted from [Loghmani et al., 2022]

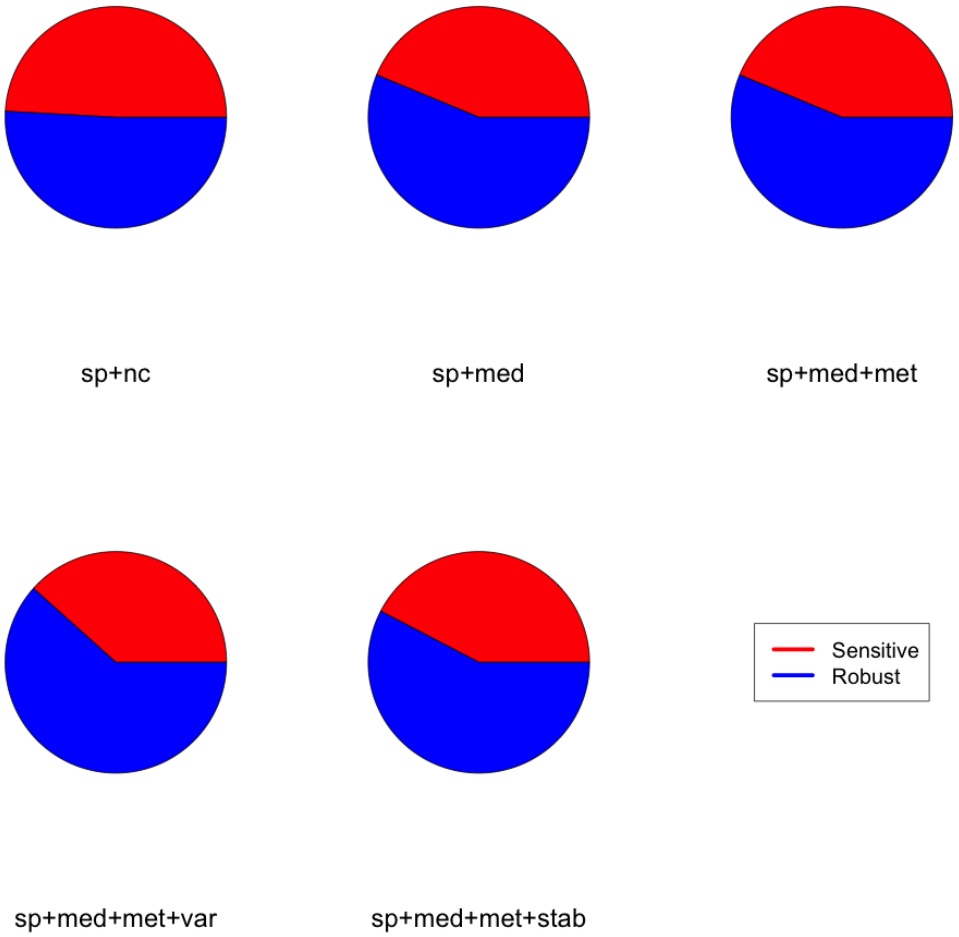

Figure A.6: The proportion of sensitive reactions with respect to perturbations in other reactions in the genome-scale models of *S. pyogenes*. In the two charts on the bottom, "var" refers to the case when more variable reactions were deactivated than stable reactions, and "stab" refers to the case when more stable reactions were deactivated than variable reactions. The integration of constraints (none, medium composition, metabolic and proteome data, from left to right) into the model results in reducing the number of sensitive reactions (red) and increasing the number of robust reactions (blue). The deactivation of a higher number of variable reactions had a slightly more impact on the decrease of the number of sensitive reactions. The figure is adapted from [Loghmani et al., 2022]

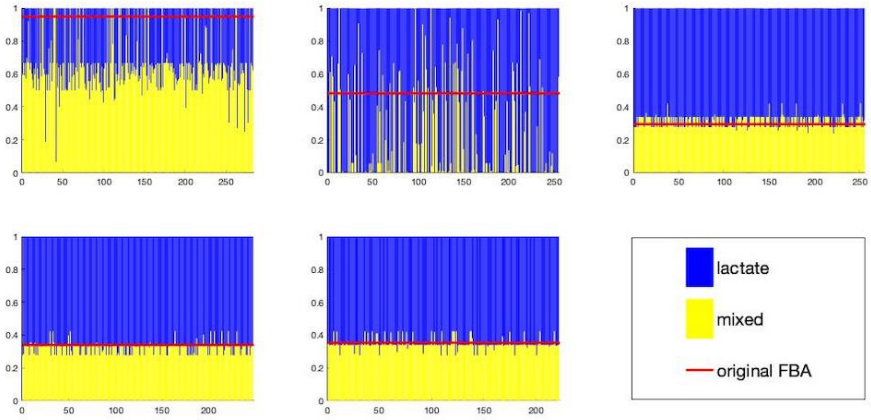

Figure A.7: The relative flux distribution through the carbohydrate branchpoint, resulting in homolactic or mixed acid fermentation in the genome-scale models of *S. pyogenes*. The figure is adapted from [Loghmani et al., 2022]

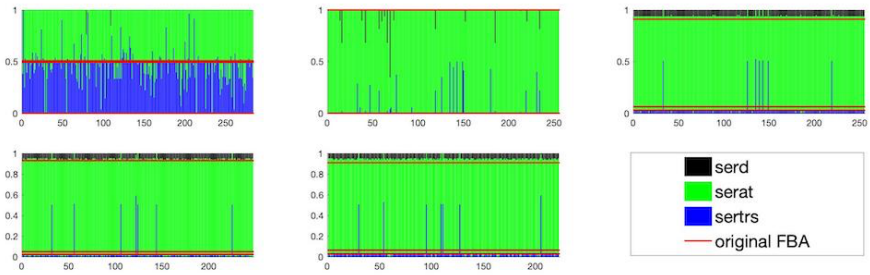

Figure A.8: The relative flux distribution through the serine metabolism in the genome-scale model of *S. pyogenes*, resulting in the production of acetyl serine, or seryl-tRNA or serine secretion. The figure is adapted from [Loghmani et al., 2022]

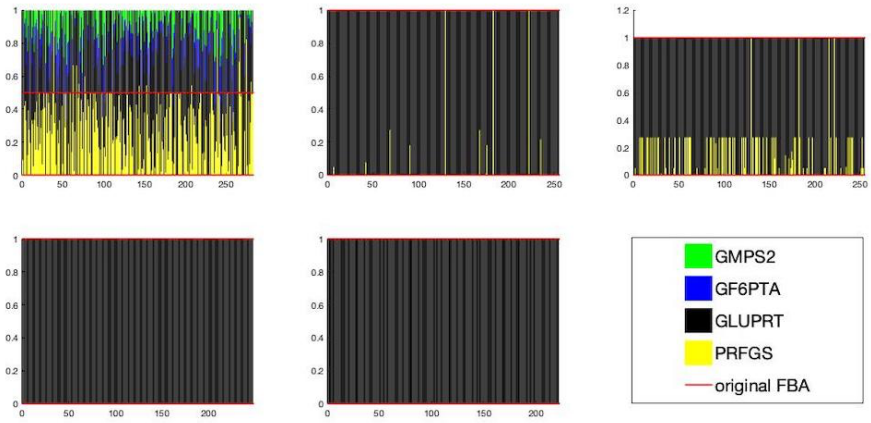

Figure A.9: The relative flux distribution through a branch point in glutamine metabolism in the genome-scale model of *S. pyogenes*, resulting in the distribution of glutamine in different pathways, namely amino acid, purine and pyrimidine metabolism. The figure is adapted from [Loghmani et al., 2022]

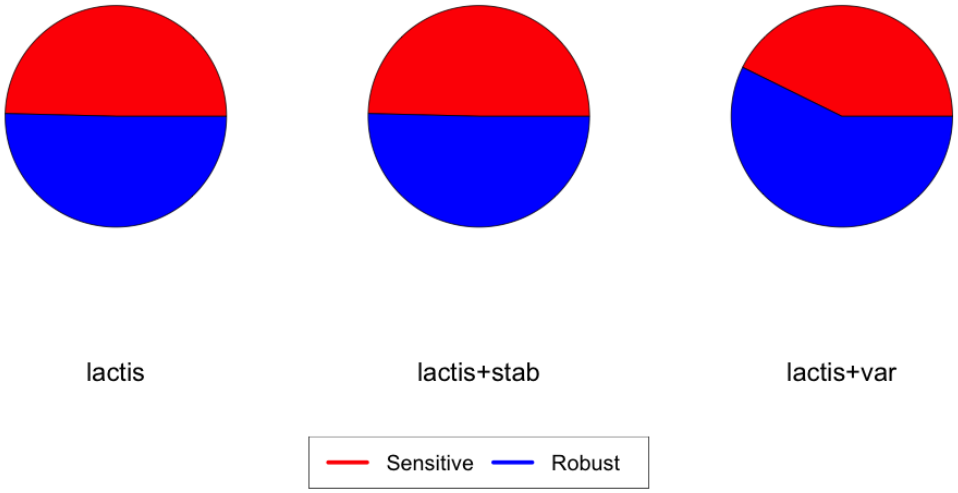

Figure A.10: The proportion of sensitive reactions with respect to perturbations in other reactions in the genome-scale models of *L. lactis* wildtype. "stab" refers to the case when only stable reactions were deactivated, while "var" refers to the case when only variable reactions were deactivated. Not surprisingly, the deactivation of variable reactions had more impact on the number of sensitive reactions. The figure is adapted from [Loghmani et al., 2022]

Table A.16: Calculated flux values for carbohydrates, organic and amino acids under different conditions in the batch culture experiment for identification of tract-specific drug targets.

Exchange reactions	EF CDM-LAB	EF Saliva	EF Saliva+glc	EF Urine+glc	EF Urine+fru+suc	SP CDM 6.5	SP CDM 7.4	SP Saliva	SP Saliva+glc	SP Plasma
R_Ex_cit__e	-7.57	-0.0925	-0.147	-0.14	-0.305	-8.019	-1.097	0.02	-0.141	-0.341
R_Ex_pyr__e	1.389	-2.45	-2.611	0.618	0.586946058		1.033	-0.82	-2.147	0.626
R_Ex_lac_L__e	20.74	1.359	-4.784	2.048	-0.461			3.013	-6.132	7.99
R_Ex_for__e				1.471	-14.53	-5.21				9.115
R_Ex_ac__e	-24.91	-5.02	-4.613	0.235	-1.075	-45.94	-0.852	-1.768	-3.056	1.841
R_Ex_glc__e	-74.105	-1.227	-3.232	-1.227		-110.8	-29.38	0.084	-30.07	-2.246
R_Ex_asp_L__e	1.188	-9.566	-0.062	4.571	-4.584	-5.21	-2.754	2.097	1.068	-3.746
R_Ex_glu_L__e	2.465	-1.465	-1.3064	9.10	-7.87	-4.78	-27.499	5.449	-1.559	3.03
R_Ex_asn_L__e	0.659	-2.02	-0.1136	1.212	-1.237	-6	-3.52	0.438	-0.99	-0.96
R_Ex_ser_L__e	-0.058	-6.35	1.0192	2.85	-4.081	-5.14	-3.033	-1.69	-4.816	-0.667
R_Ex_gln_L__e	-0.1326	-0.7	0.197	0.107	-0.55	-1.794	-1.61	-0.31	-0.46	-3.83
R_Ex_his_L__e	-0.46	-1.22	0.779	0.139	-0.673	-1.52	-0.808	-0.81	-0.202	-1.477
R_Ex_gly__e	-1.68	-2.205	0.477	1.066	-1.747	-3.62	-2.25	-1.174	-1.949	-1.157
R_Ex_thr_L__e	2.349	-1.724	1.734	0.484	-1.403	-2.43	-2.62	-1.78	-2.12	-1.315
R_Ex_arg_L__e	-4.82	-0.8738	0.902	-0.1015	-0.479	-4.95	-1.324	3.779	4.547	-0.038
R_Ex_ala_L__e	-1.47	-8.240	4.108	1.688	-5.58	-5.18	-3.235	-3.566	-1.95	-2.349
R_Ex_tyr_L__e	0.474	-1.421	1.082	7.48	-9.263	-2.023	-1.162	-1.917	-7.713	-1.389
R_Ex_val_L__e	-3.559	-0.117	-0.042	0.251	-0.039	-4.237	-2.628	0.02	-0.508	0.45
R_Ex_met_L__e	21.567	-4.70	-0.939	3.615	-2.38	9.771	May 38	3.873	1.128	-0.04
R_Ex_trp_L__e	0.251	-2.46	0.256	1.01	-1.09	-0.237	-0.183	0.431	0.117	-2.825
R_Ex_phe_L__e	-2.046	-1.776	-1.501	2.113	-1.02	-3.904	-2.015	2.489	3.03	-3.605
R_Ex_ile_L__e	2.109	-8.606	-0.885	5.858	-4.364	-0.97	-1.035	2.372	-4.95	-3.015
R_Ex_leu_L__e	3.975	-1.792	-0.0857	7.78	-9.505	-5.184	-3.061	-2.5044	-4.44	1.078
R_Ex_lys_L__e	0.941	-1.303	7.027	0.7728	-10.66	0.43	-0.316	6.454	-2.7	-1.86
R_Ex_cys_L__e	-0.192	-0.502	-0.339	0.116	-0.148	-0.161	-0.2	50.38	18.53	
R_Ex_fru__e					-3.227					
R_Ex_sucr__e					0.0676					

Table A.17: Metabolic constraints integrated into the genome-scale metabolic model of *E. faecalis* to specify the CDM-LAB condition.

Reaction	Lower bound	Upper bound
R_Ex_cit__e	-9.0858053952	-6.0572035968
R_Ex_pyr__e	1.1119774888	1.6679662332
R_Ex_lac_L__e	16.0	20.0
R_Ex_ac__e	0.0	8.8
R_Ex_glc__e	-80.0	-15.0
R_Ex_asp_L__e	0.950435772	1.425653658
R_Ex_glu_L__e	1.9722732968	2.9584099452
R_Ex_asn_L__e	0.5276403992	0.7914605988
R_Ex_ser_L__e	-0.0705287604	-0.0470191736
R_Ex_gln_L__e	-10.0	10.0
R_Ex_his_L__e	-1000.0	1000.0
R_Ex_gly__e	-2.0268700116	-1.3512466744
R_Ex_thr_L__e	1.8797085064	2.8195627596
R_Ex_arg_L__e	-1000.0	1000.0
R_Ex_ala_L__e	-1.7645513088	-1.1763675392
R_Ex_tyr_L__e	0.37976486	0.56964729
R_Ex_val_L__e	-1000.0	1000.0
R_Ex_met_L__e	0.0	25.880516952
R_Ex_trp_L__e	-10.0	0.30185178
R_Ex_phe_L__e	-2.4554629944	-1.6369753296
R_Ex_ile_L__e	-10.0	2.5319808948
R_Ex_leu_L__e	-10.0	4.771032564
R_Ex_lys_L__e	0.753295696	1.129943544
R_Ex_cys_L__e	-0.2307378216	-0.1538252144

Table A.18: Metabolic constraints integrated into the genome-scale metabolic model of *E. faecalis* to specify the artificial saliva condition.

Reaction	Lower bound	Upper bound
R_Ex_cit__e	-0.1110395292	-0.0740263528
R_Ex_pyr__e	-2.9452427004	-1.9634951336
R_Ex_lac_L__e	0.0	10.0
R_Ex_ac__e	-6.0251192472	-4.0167461648
R_Ex_glc__e	-1.4729048976	-0.9819365984
R_Ex_asp_L__e	-11.4799505136	-0.2
R_Ex_glu_L__e	-17.588414748	10.0
R_Ex_asn_L__e	-2.43410115	-1.6227341
R_Ex_ser_L__e	-7.6286039364	-5.0857359576
R_Ex_gln_L__e	-0.8403180408	-0.5602120272
R_Ex_his_L__e	-1000.0	1000.0
R_Ex_gly__e	-2.646900366	-0.05
R_Ex_thr_L__e	-2.0695203324	-1.3796802216
R_Ex_arg_L__e	-1.0485705036	-0.6990470024
R_Ex_ala_L__e	-9.8885349024	-0.05
R_Ex_tyr_L__e	-17.063932644	-0.05
R_Ex_val_L__e	-0.1412160492	-0.0941440328
R_Ex_met_L__e	-5.647497468	-0.05
R_Ex_trp_L__e	-2.959803234	-1.973202156
R_Ex_phe_L__e	-2.132379954	-1.421586636
R_Ex_ile_L__e	-10.3274020116	-0.05
R_Ex_leu_L__e	-21.514502784	-0.05
R_Ex_lys_L__e	-15.640768056	-0.05
R_Ex_cys_L__e	-0.6024232356	-0.4016154904

Table A.19: Transcriptomic constraints integrated into the genome-scale metabolic model of *E. faecalis* to specify the artificial saliva condition.

Reaction	Lower bound	Upper bound	Reaction	Lower bound	Upper bound
R_CTPS2	0.0	0.045358235004	R_TYRTA	-0.177904448613	-0.072969058704
R_ADCS	0.0	7.79918e-07	R_ASAD	-3.908536943211	-0.864389656599
R_ALATRS	0.0	0.462093088076	R_LTAS1	0.0	0.006244352138
R_ASP2DC	0.0	0.739189880667	R_LPGS_EFA	0.000607625859	0.001419212883
R_GLYCK	0.0	2.068254277975	R_CITt6	0.0	840.139581492607
R_ACACT1	0.0	0.00100484336	R_ASNTRS	0.0	0.096616511483
R_ALAR	0.016234043834	0.037917352965	R_DTMPK	0.000681284165	0.001591254306
R_DEX_PS_EFA	0.0	0.00019377711	R_SHK3D	0.0	0.002779501922
R_TMDS	0.0	0.008355982644	R_TRE6PH	0.0	1.634006329198
R_ASPTA1	-5.358540653769	-1.407600285225	R_DAPDC	0.207199905183	0.490331329466
R_GLCt	0.0	6.587879344828	R_ARGORNt	0.0	86.822676956651
R_UGLDDS1_A	0.0	0.059462642828	R_LEUTRS	0.0	0.098255803102
R_UGLDDS1_L	0.0	0.059462642828	R_XYLA	0.0	324.208853487699
R_GF6PTA	0.0	10.296700850497	R_UACGE	0.0	0.178025157692
R_PROTRS	0.0	0.044477277847	R_TDPGDH	0.0	0.083299451623
R_ARGTRS	0.0	0.048671204997	R_GLNS	0.0	0.126605638935
R_TPI	0.0	90.08586003439	R_ASCt	0.0	4.650374090233
R_NDPK8	0.0	5717662.18629238	R_METTRS	0.0	0.026769162748
R_NDPK7	0.0	0.011331297123	R_PGSA_LPL	0.013653056642	0.031889021185
R_NDPK6	0.0	0.200047367356	R_DRPA	-0.036928574945	0.0
R_NDPK5	0.0	0.011331297123	R_MTHFD	-0.686680217151	-0.287741879381
R_NDPK4	0.0	0.18899703336	R_MTHFC	-0.686680217147	-0.287741879381
R_NDPK3	0.0	0.614683932407	R_RBK	0.0	0.096090242606
R_NDPK2	0.0	1.892404991959	R_TREPP	-2.392325298199	0.0
R_NDPK1	0.0	4.883249598052	R_LEUt6	0.083400827666	0.358793399742
R_ENO	5.0	183.64011056531	R_5RNTP1	-0.000109388957	0.065535188864
R_CELBP	-0.839972245292	0.0	R_BPPA2_L	0.0	0.027555182955
R_GMPR	0.0	0.077783634179	R_DHDPS	0.202935792795	0.480240456621
R_GLUSYN1	-8.718683174643	0.0	R_EPA_PS_EFA	0.0	0.005231981978
R_ILEt6	0.045977379355	0.271472018604	R_CYSTRS	0.0	0.014345239124
R_PGL	0.0	52.061242514703	R_BPPA2	0.0	0.027555182955
R_PGI	-39.43265314777	31.316352762482	R_ADPT	0.0	0.076586858641
R_DURIPP	-0.008340676585	0.0	R_FBA	-0.115696214635	61.715598372292
R_5RNTP2	-0.000109388918	0.0	R_GLYK	0.0	1582.03221867543
R_ASPTRS	0.006604631971	0.015426234167	R_PPM2	-0.016756694571	0.0
R_SERt6	0.002255190308	10.0	R_HISTRS	0.001782683291	0.004163758103
R_ARGD	0.0	218.474210820336	R_ASPK	0.27927410487	5.0

Table A.20: Metabolic constraints integrated into the genome-scale metabolic model of *E. faecalis* to specify the artificial saliva+glc condition.

Reaction	Lower bound	Upper bound
R_Ex_cit__e	-0.1764291744	-0.1176194496
R_Ex_pyr__e	-3.1336668072	-2.0891112048
R_Ex_lac_L__e	0.0	16.0
R_Ex_ac__e	-5.5356831228	-3.6904554152
R_Ex_glc__e	-38.793431628	-25.862287752
R_Ex_asp_L__e	-0.0745011444	-0.0496674296
R_Ex_glu_L__e	-1.567765728	-1.045177152
R_Ex_asn_L__e	-0.1364143104	-0.0909428736
R_Ex_ser_L__e	0.815419712	1.223129568
R_Ex_gln_L__e	0.1579788344	0.2369682516
R_Ex_his_L__e	-1000.0	1000.0
R_Ex_gly__e	-1000.0	0.5726552088
R_Ex_thr_L__e	1.3875473776	2.0813210664
R_Ex_arg_L__e	-1000.0	1.0825183428
R_Ex_ala_L__e	3.286534108	4.929801162
R_Ex_tyr_L__e	0.8657114496	1.2985671744
R_Ex_val_L__e	-0.050440662	-0.033627108
R_Ex_met_L__e	-1.127462148	-0.751641432
R_Ex_trp_L__e	-1000.0	0.3082542132
R_Ex_phe_L__e	-1.8013793616	-1.2009195744
R_Ex_ile_L__e	-1.0631915016	-0.7087943344
R_Ex_leu_L__e	-0.1029384372	-0.0686256248
R_Ex_lys_L__e	5.6221317272	8.4331975908
R_Ex_cys_L__e	-0.4073369328	-0.2715579552

Table A.21: Transcriptomic constraints integrated into the genome-scale metabolic model of *E. faecalis* to specify the artificial saliva+glc condition.

Reaction	Lower bound	Upper bound
R_BUK	0.0000	0.0000
R_FOLt	0.0000	0.0000
R_ASP2DC	0.0000	1.6404
R_MNLpts	0.0000	0.0000
R_MENF	0.0000	0.0000
R_MEND	0.0000	0.0000
R_MENH	0.0000	0.0000
R_ACACT1	0.0000	0.0011
R_PDE1	0.0000	0.0000
R_PDE2	0.0000	0.0000
R_PDE3	0.0000	0.0000
R_PDE4	0.0000	0.0000
R_PDE5	0.0000	0.0000
R_FRUK	0.0000	56.1331
R_ARGTRS	0.0000	0.0504
R_TPI	0.0000	76.8103
R_M1PD	0.0000	0.0000
R_TYRt6	-3.3086	-0.7907
R_ENO	5.0000	157.6671
R_PGK	5.0000	187.4988
R_ASPTRS	0.0041	0.0382
R_SERt6	-10.0000	0.0313
R_DALAt	0.0000	0.0000
R_ADPRDP	0.0000	0.0000
R_UAMAS	0.0000	0.1591
R_ADSS	0.0000	0.1573
R_SHK3D	0.0000	0.0034
R_ARGORNt	1.4204	8.1652
R_XYLB	0.0000	0.0000
R_XYLI	0.0000	0.0000
R_PTB	0.0000	0.0000
R_RIBt2	0.0000	0.0000
R_GLNS	0.0000	0.3778
R_ASCt	0.0000	10.9841
R_RBK	0.0000	0.1277
R_TREPP	-0.1516	0.0000
R_GLYC3Pabc	0.0000	0.0000
R_DSERt	0.0000	0.0000
R_ARABte	0.0000	0.0000
R_MTHPTGHM	0.0000	0.0000
R_ASPK	0.2962	100.0000

Table A.22: Metabolic constraints integrated into the genome-scale metabolic model of *E. faecalis* to specify the artificial urine+glc condition.

Reaction	Lower bound	Upper bound
R_Ex_cit__e	-0.1680858588	-0.1120572392
R_Ex_pyr__e	0.4944608472	0.7416912708
R_Ex_lac_L__e	0.0	16.0
R_Ex_for__e	0.0	5.0
R_Ex_ac__e	0.188004336	0.282006504
R_Ex_glc__e	-1.4732685504	-0.9821790336
R_Ex_asp_L__e	3.657555964	5.486333946
R_Ex_glu_L__e	7.2807040024	10.9210560036
R_Ex_asn_L__e	0.970110976	1.455166464
R_Ex_ser_L__e	2.287170024	3.430755036
R_Ex_gln_L__e	-10.0	0.1294142196
R_Ex_his_L__e	-1000.0	1000.0
R_Ex_gly__e	-1000.0	1.2799458408
R_Ex_thr_L__e	0.3877419584	0.5816129376
R_Ex_arg_L__e	-1000.0	-0.0812764728
R_Ex_ala_L__e	1.3508984912	2.0263477368
R_Ex_tyr_L__e	3.084489	8.9767328352
R_Ex_val_L__e	-10.0	10.0
R_Ex_met_L__e	-10.0	4.33919415
R_Ex_trp_L__e	-10.0	1.2122225328
R_Ex_phe_L__e	-10.0	2.5365151608
R_Ex_ile_L__e	-10.0	7.0297080324
R_Ex_leu_L__e	-10.0	9.3360988308
R_Ex_lys_L__e	0.6182596768	0.9273895152
R_Ex_cys_L__e	0.0934520136	0.1401780204

Table A.23: Transcriptomic constraints integrated into the genome-scale metabolic model of *E. faecalis* to specify the artificial urine+glc condition.

Reaction	Lower bound	Upper bound	Reaction	Lower bound	Upper bound
R_CTPS2	0.0	0.087626334013	R_TRE6PH	0.0	0.186650796687
R_GLCpts	0.0	46.199781800126	R_GLUPRT	0.0	1.415251689591
R_ALATRS	0.0	0.428169938893	R_GALU	0.0	0.467926255596
R_ASP2DC	0.0	0.461376654738	R_ARGORNt	0.0	100.0
R_PPND	0.067848754978	10.0	R_AICART	0.0	0.96664539192
R_ACACT1	0.0	0.000924644451	R_BTNt2i	0.0	0.000273242914
R_SHKK	0.0	0.003642241016	R_GLNS	0.0	0.132900051431
R_FRUK	0.0	75.624389112006	R_DRPA	-0.042126778376	0.0
R_UGLDDS1_A	0.0	0.060600870086	R_RBK	0.0	0.158278121315
R_UGLDDS1_L	0.0	0.060600870086	R_TREPP	-0.194577065149	0.0
R_DHDPRy	0.565184	3.23065	R_LEUt6	0.05169632412	0.222399470174
R_GF6PTA	0.0	7.180629147768	R_5RNTP1	-0.000178938906	0.10720273176
R_ARGTRS	0.0	0.052164505905	R_OMPDC	0.0	2.363574788171
R_TPI	0.0	202.703722364358	R_URIDK2	0.0	0.234622546977
R_G6PDA	-0.381337207678	4.132117566719	R_URIDK1	0.0	0.637314551939
R_ENO	0.0	342.684563453802	R_ADK1	-1.862741217194	0.0
R_GMPR	0.0	1.261911251968	R_PAPPT1_A	0.004539	0.058175
R_ILEt6	0.028499255604	0.168272975891	R_PAPPT1_L	0.00453	0.049775
R_PGK	0.0	274.51428129779	R_CYSTRS	0.0	0.012062860155
R_PGI	-42.262871201178	33.564035834253	R_ADPT	0.0	0.069504041262
R_DASYN_LPL	0.0101	0.13367	R_FBA	-0.233001890895	124.289728612101
R_5RNTP2	-0.000178938842	0.0	R_GLYK	0.0	77.577332570816
R_SERt6	-10.0	0.025250003512	R_DMATT	0.0	0.000111119513
R_ARGD	0.386206922623	100.0	R_PRAGS	0.0	1.803824084126
R_IMPC	-0.966645385676	0.0	R_PRASCS	0.0	1.190080066819
R_ORPT	-1.906558125183	0.0	R_PPM2	-0.016640947627	0.0
R_PYK	0.0	97.359266882558	R_G1PACT	0.0	0.448066349523
R_HCO3E	0.271139379878	10.0	R_UAGDP	0.0	0.448066349523
R_CHORM	0.0	0.00387653107	R_PNTOt2	0.0	1.9881067e-05
R_PSCVT	0.0	0.003823321117	R_ADD	0.0	5.145248211735
R_ADSS	0.0	1.001250151411	R_ACLDC	0.0	0.01358740313
R_SHK3D	0.0	0.001209848482	R_GARFT	0.0	2.550992466621
R_TRE6PH	0.0	0.186650796687	R_PRAIS	0.0	1.465160424421
			R_ASPK	0.659641336758	2.982720251091

Table A.24: Metabolic constraints integrated into the genome-scale metabolic model of *E. faecalis* to specify the artificial urine+fru+suc condition.

Reaction	Lower bound	Upper bound
R_Ex_cit__e	-0.366158136	-0.244105424
R_Ex_pyr__e	0.4695568464	0.7043352696
R_Ex_lac_L__e	-0.553859808	-0.369239872
R_Ex_for__e	0.0	16.0
R_Ex_ac__e	-1.2906393756	-0.8604262504
R_Ex_fru__e	-3.87308541	-1.582057
R_Ex_sucr__e	-1.3944038856	0.0
R_Ex_asp_L__e	-5.5010967552	-3.6673978368
R_Ex_glu_L__e	-9.4550768844	-6.3033845896
R_Ex_asn_L__e	-1.4848437384	-0.9898958256
R_Ex_ser_L__e	-4.8983897988	-3.2655931992
R_Ex_gln_L__e	-0.661173516	-0.440782344
R_Ex_his_L__e	-1000.0	1000.0
R_Ex_gly__e	-2.0970642996	-1.3980428664
R_Ex_thr_L__e	-1.6845762804	-1.1230508536
R_Ex_arg_L__e	-0.5750998476	-0.3833998984
R_Ex_ala_L__e	-6.7073995824	0.0
R_Ex_tyr_L__e	-11.11577145	0.0
R_Ex_val_L__e	-1000.0	-0.0319046248
R_Ex_met_L__e	-2.860459644	2.0
R_Ex_trp_L__e	-1.3189221864	-0.8792814576
R_Ex_phe_L__e	-1000.0	0.0
R_Ex_ile_L__e	-5.2378809396	-3.4919206264
R_Ex_leu_L__e	-11.406732642	-7.604488428
R_Ex_lys_L__e	-12.796433472	0.0
R_Ex_cys_L__e	-0.1786157952	10.0

Table A.25: Transcriptomic constraints integrated into the genome-scale metabolic model of *E. faecalis* to specify the artificial urine+fru+suc condition.

Reaction	Lower bound	Upper bound	Reaction	Lower bound	Upper bound
R_CTPS2	0.0	0.091982824738	R_UACGE	0.0	0.159336979391
R_GLCpts	0.0	54.941109228283	R_AICART	0.0	0.642184009753
R_ALATRS	0.0	0.434146968502	R_PPNCL	6.362031e-06	0.00015
R_ASP2DC	0.0	0.575950572659	R_BTNt2i	0.0	0.000851604802
R_ACACT1	0.0	0.001203278526	R_GLNS	0.0	0.193233022331
R_DEX_PS_EFA	0.0	0.00015175595	R_DRPA	-0.018852261402	0.0
R_TMDS	0.0	0.007890760862	R_MTHFD	-0.767218974591	-0.321490300917
R_UGLDDS1_A	0.0	0.070182	R_MTHFC	-0.767218974587	-0.321490300917
R_UGLDDS1_L	0.0	0.070182	R_RBK	0.0	0.138747506704
R_DHDPRy	1.327852821455	3.142317264689	R_LEUt6	0.039043955016	10.0
R_GF6PTA	0.0	13.399527789717	R_5RNTP1	-0.000129187555	0.077396577114
R_ARGTRS	0.0	0.058282731512	R_OMPDC	0.0	2.971045120413
R_TPI	0.0	104.201249838725	R_URIDK2	0.0	0.23789775568
R_ENO	0.0	202.353977312061	R_URIDK1	0.0	0.646211131546
R_GMPR	0.0	1.409917211213	R_BPPA2_L	0.0	0.024662579579
R_ILEt6	0.021524231611	10.0	R_EPA_PS_EFA	0.0	0.004682755041
R_PGK	0.0	172.534061544285	R_UDCPDPS	7.887059e-06	0.00018
R_PGI	-42.262871201178	33.564035834253	R_BPPA2	0.0	0.024662579579
R_DASYN_LPL	0.009721318991	0.162706	R_ADPT	0.0	0.10246766851
R_5RNTP2	-0.000129187509	0.0	R_FBA	-0.260330015843	138.867315172787
R_SERt6	0.005074441981	10.0	R_GLYK	0.0	73.903114793851
R_ARGD	0.409634300892	1.81337098052	R_PRAGS	0.0	1.293301505965
R_IMPC	-0.642184005605	0.0	R_PRASCS	0.0	0.847366875965
R_ORPT	-2.100846151855	0.0	R_G1PACT	0.0	0.493726602792
R_HCO3E	0.351615133168	0.959942711579	R_UAGDP	0.0	0.493726602792
R_ADSS	0.0	0.85963877323	R_ADD	0.0	10.009102257617
R_SHK3D	0.0	0.001305702644	R_ACLDC	0.0	0.013033908298
R_GLUPRT	0.0	1.133715497739	R_GARFT	0.0	1.730343709752
R_GALU	0.0	0.580091782066	R_PRAIS	0.0	1.087532592403
R_ARGORNt	0.134556132554	0.593514393258	R_ASPK	0.721841359905	3.263971983392

Table A.26: Metabolic constraints integrated into the genome-scale metabolic model of *S. pyogenes* to specify the CDM-LAB condition at pH 6.5.

Reaction	Lower bound	Upper bound
R_EX_cit__e	-9.6235255896	-6.4156837264
R_EX_ac__e	-55.129986552	5.0
R_EX_glc__e	-133.02224724	-10.0
R_EX_asp_L__e	-6.2520499224	0.0
R_EX_glu_L__e	-5.7362575416	-3.8241716944
R_EX_asn_L__e	-7.2009017724	10.0
R_EX_ser_L__e	-6.1695196536	-4.1130131024
R_EX_gln_L__e	-2.1533685192	-1.4355790128
R_EX_his_L__e	-1000.0	1000.0
R_EX_gly__e	-4.3555498884	10.0
R_EX_thr_L__e	-2.9160152508	10.0
R_EX_arg_L__e	-5.9403631632	10.0
R_EX_ala_L__e	-6.220312086	10.0
R_EX_tyr_L__e	-2.42864223	10.0
R_EX_val_L__e	-1000.0	1000.0
R_EX_met_L__e	-10.0	11.7261942768
R_EX_trp_L__e	-0.2844561204	10.0
R_EX_phe_L__e	-4.6855814232	10.0
R_EX_ile_L__e	-1.1667264108	10.0
R_EX_leu_L__e	-6.2215703112	10.0
R_EX_lys_L__e	-10.0	0.5168450544
R_EX_cys_L__e	-0.1937968236	10.0

Table A.27: Metabolic constraints integrated into the genome-scale metabolic model of *S. pyogenes* to specify the CDM-LAB condition at pH 7.4.

Reaction	Lower bound	Upper bound
R_EX_cit__e	-1.3164671244	-0.8776447496
R_EX_pyr__e	0.8269069024	1.2403603536
R_EX_ac__e	-1.0235968464	1.0
R_EX_glc__e	-35.256143556	-23.504095704
R_EX_asp_L__e	-3.3049111296	-2.2032740864
R_EX_glu_L__e	-3.2999236704	-2.1999491136
R_EX_asn_L__e	-4.2329161932	-2.8219441288
R_EX_ser_L__e	-3.640351326	-2.426900884
R_EX_gln_L__e	-10.0	-1.2882045352
R_EX_his_L__e	-1000.0	1000.0
R_EX_gly__e	-2.7074387664	10.0
R_EX_thr_L__e	-3.1555247232	0.0
R_EX_arg_L__e	-1.58946093	0.0
R_EX_ala_L__e	-3.882668694	10.0
R_EX_tyr_L__e	-1.39532211	0.0
R_EX_val_L__e	-1000.0	1000.0
R_EX_met_L__e	-10.0	6.4621847856
R_EX_trp_L__e	-0.2205567996	0.0
R_EX_phe_L__e	-2.4181307064	0.0
R_EX_ile_L__e	-1.24260867	0.0
R_EX_leu_L__e	-3.6736642524	0.0
R_EX_lys_L__e	-0.3800688624	0.0
R_EX_cys_L__e	-0.24030171	0.0

Table A.28: Metabolic constraints integrated into the genome-scale metabolic model of *S. pyogenes* to specify the artificial saliva condition.

Reaction	Lower bound	Upper bound
R_EX_cit__e	-10.0	0.0287347944
R_EX_pyr__e	-0.9915097476	-0.6610064984
R_EX_lac_L__e	2.41	21.2
R_EX_ac__e	-2.122382616	-1.414921744
R_EX_glc__e	0.0678873536	0.1018310304
R_EX_asp_L__e	1.678316024	2.517474036
R_EX_glu_L__e	4.359376276	6.539064414
R_EX_asn_L__e	0.3508170312	0.5262255468
R_EX_ser_L__e	-2.0283197352	-1.3522131568
R_EX_gln_L__e	-10.0	-0.2480597256
R_EX_his_L__e	-1000.0	1000.0
R_EX_gly__e	-1.409	6.0
R_EX_thr_L__e	-2.1453065064	10.0
R_EX_arg_L__e	-10.0	4.534827294
R_EX_ala_L__e	-4.280043126	10.0
R_EX_tyr_L__e	-2.3005155708	0.0
R_EX_val_L__e	-1000.0	1000.0
R_EX_met_L__e	-10.0	4.6478316792
R_EX_trp_L__e	-10.0	0.517699368
R_EX_phe_L__e	-10.0	2.9871315672
R_EX_ile_L__e	-10.0	2.847198168
R_EX_leu_L__e	-3.005394084	10.0
R_EX_lys_L__e	-10.0	7.745355072
R_EX_cys_L__e	10.0	60.467352144

Table A.29: Transcriptomic constraints integrated into the genome-scale metabolic model of *S. pyogenes* to specify the artificial saliva condition.

Reaction	Lower bound	Upper bound
R_HDER4	0.005221	0.026195
R_HTDR6	0.005221	0.026195
R_HHYR2	0.005221	0.026195
R_RHC	0.004003066683	0.040753447996
R_PIt6	-100.0	0.109993561852
R_HOCR3	0.005221	0.026195
R_HDDR5	0.005221	0.026195
R_UAMAGS	0.0036	0.026425
R_ASPCT	0.02428	6.76029
R_VALTRS	0.013975	0.065998
R_HHDR7	0.00472	0.024036
R_GLUR	-0.026092	-0.00389
R_GSHPO	0.0	25.587280606423
R_HODR8	0.003028	0.016073
R_METTRS	0.003527907361	0.018239
R_HBUR1	0.005221	0.026195

Table A.30: Metabolic constraints integrated into the genome-scale metabolic model of *S. pyogenes* to specify the artificial saliva+glc condition.

Reaction	Lower bound	Upper bound
R_EX_cit__e	-0.1693921176	-0.1129280784
R_EX_pyr__e	-2.5767719232	-1.7178479488
R_EX_lac_L__e	-7.3592431392	-4.9061620928
R_EX_ac__e	-3.667778598	-2.445185732
R_EX_glc__e	-36.089528184	-10.0
R_EX_asp_L__e	0.8544850832	1.2817276248
R_EX_glu_L__e	-1.8712230192	-1.2474820128
R_EX_asn_L__e	-1.1914247172	-0.7942831448
R_EX_ser_L__e	-5.779830444	10.0
R_EX_gln_L__e	-10.0	-0.3684283016
R_EX_his_L__e	-1000.0	1000.0
R_EX_gly__e	-2.339162622	10.0
R_EX_thr_L__e	-2.5504304676	0.0
R_EX_arg_L__e	-10.0	5.4569316288
R_EX_ala_L__e	-2.3433739944	10.0
R_EX_tyr_L__e	-9.25615866	0.0
R_EX_val_L__e	-1000.0	1000.0
R_EX_met_L__e	-10.0	1.354339656
R_EX_trp_L__e	-10.0	0.1415635632
R_EX_phe_L__e	-10.0	3.6364343196
R_EX_ile_L__e	-5.9458412172	0.0
R_EX_leu_L__e	-5.3395903032	0.0
R_EX_lys_L__e	-3.3482730072	0.0
R_EX_cys_L__e	14.828757456	22.243136184

Table A.31: Transcriptomic constraints integrated into the genome-scale metabolic model of *S. pyogenes* to specify the artificial saliva+glc condition.

Reaction	Lower bound	Upper bound
R_RHC	0.003866707537	0.039365235949
R_UAMAGS	0.003329	0.019776
R_GRTT	9.848762e-06	3.3e-05
R_GLUR	-0.01996	-0.002166
R_DMATT	9.848762e-06	3.3e-05
R_METTRS	0.002975	0.01395
R_UAGPT1	0.005066	0.019829
R_DASYN_LLA	0.01138	0.045589

Table A.32: Metabolic constraints integrated into the genome-scale metabolic model of *S. pyogenes* to specify the natural human plasma condition.

Reaction	Lower bound	Upper bound
R_EX_cit__e	-0.4103004	-0.2735336
R_EX_pyr__e	0.5014021176	0.7521031764
R_EX_lac_L__e	6.398	22.597
R_EX_for__e	7.2926827712	10.9390241568
R_EX_ac__e	1.473	8.2095
R_EX_glc__e	-2.695	-1.7969
R_EX_asp_L__e	-4.495	-2.9972
R_EX_glu_L__e	2.42471	3.637
R_EX_asn_L__e	-1.1607528696	-0.7738352464
R_EX_ser_L__e	-0.8010762036	-0.5340508024
R_EX_gln_L__e	-4.599068196	-3.066045464
R_EX_his_L__e	-1000.0	1000.0
R_EX_gly__e	-1.388821302	-0.925880868
R_EX_thr_L__e	-1.5783836796	0.0
R_EX_arg_L__e	-0.165972	-0.03064
R_EX_ala_L__e	-2.8190148984	0.0
R_EX_tyr_L__e	-1.6670289696	0.0
R_EX_val_L__e	-1000.0	1000.0
R_EX_met_L__e	-0.098379	0.0
R_EX_trp_L__e	-3.3910862388	0.0
R_EX_phe_L__e	-4.3261799868	0.0
R_EX_ile_L__e	-3.6185486256	10.0
R_EX_leu_L__e	-10.0	1.2940224024
R_EX_lys_L__e	-2.2327705056	0.0

Table A.33: Transcriptomic constraints integrated into the genome-scale metabolic model of *S. pyogenes* to specify the natural human plasma condition.

Reaction	Lower bound	Upper bound	Reaction	Lower bound	Upper bound
R_CITt6	0.0	1.925063216232	R_GMPR	0.0	0.006014577206
R_G6PI	-1.706659545171	0.0	R_AGAT_LLA	0.00227713096	2.0
R_DTMPK	0.004255	0.03294	R_RPE	-0.596412106775	-0.021236008114
R_HDER4	0.00596	0.123895	R_LTAS_LLA	0.0021	0.03144
R_CTPS2	0.0	0.089487595026	R_PGK	31.97499165858	128.093450694278
R_HTDR6	0.00596	0.123895	R_HHDHL7	0.0159	0.103763
R_PAP_LLA	0.0	0.011619397254	R_ALAALA	0.0	0.197417780428
R_HHYR2	0.00596	0.123895	R_CTPS1_1	0.0	0.089487595026
R_RHC	0.065742904643	0.274205016606	R_LDH_L	-289.188534671265	-20.0
R_HCYSMT	0.0	10.0	R_LYSTRS	0.0437	0.232089
R_PIt6	0.0	1.304724689533	R_PGI	15.243928832298	64.876767742903
R_FFSD	0.0	0.753236747085	R_URIDK2	0.0	0.013072501467
R_HOCR3	0.00596	0.123895	R_HODR8	0.00346	0.073859
R_HDDR5	0.00596	0.123895	R_METTRS	0.0121	0.088602
R_UAMAGS	0.0	0.329722407032	R_SUCH	0.0	0.012539981607
R_TDPGDH	0.0	0.007729	R_MACPMT	0.039704196119	2.0
R_G5SD	0.0	0.130095371752	R_CYSTRS	0.0	0.301441238719
R_VALTRS	0.0	0.292254	R_G1PTMT	0.0	0.007309
R_HODHL8	0.010205	0.070865	R_GAPN	0.0	0.00495355176
R_HOCHL3	0.017596	0.120456	R_HBUR1	0.00596	0.123895
R_RPE_L	-1.640584263525	0.0	R_HHYHL2	0.0175	0.120456
R_PROTRS	0.008373	0.173528	R_HTDHL6	0.0175	0.120456
R_HHDR7	0.0054	0.111625	R_HDDHL5	0.01759	0.120456
R_HDEHL4	0.01759	0.120456	R_TDPDRE	0.0	0.007199
R_HBUHL1	0.01759	0.120456	R_CYNt	0.0	0.04949231049
R_FORt2	-10.0	-1.309221213781	R_ACTPASE	0.0	0.030877056598
R_PFK	7.46111	28.6435	R_UAGPT1	0.0	0.253371092041
R_GLUR	-0.175473349266	0.0	R_DASYN_LLA	0.0355	0.262378
R_NADN	0.0	0.000922423278	R_GHMT	-10.0	8.885434429244
R_SERAT	0.001006092738	0.025691847404			

Table A.34: Significant fold changes at the protein level in *E. faecalis* between Saliva+glc and CDM-LAB.

Genes	log fold change	adjusted p-value	Genes	log fold change	adjusted p-value
CON__A2I7N0	3.19	4.09e-05	**EF_1573**	1.28	0.000544
CON__P28800	5.87	4.48e-14	**EF_1753**	836	0.0111
CON__P41361	4.14	4.48e-14	**EF_1904**	2.7	1.2e-09
CON__Q3SX09	2.75	5.53e-12	**EF_1937**	-1.45	0.00156
dapH	-1.15	0.000139	**EF_2049**	3.66	5.19e-06
dltA	1.58	2.41e-10	**EF_2050**	3.36	4.48e-14
dltD	2.76	4.48e-14	**EF_2067**	-2.58	0.00138
EF_0063	-6.69	4.48e-14	**EF_2174**	814	0.0072
EF_0157	-2.82	4.48e-14	**EF_2211**	6.22	1.57e-05
EF_0164	-1.04	0.00196	**EF_2222**	-815	0.0378
EF_0236	-3.17	1.94e-07	**EF_2254**	1.19	0.000361
EF_0382	-1.84	26	**EF_2445**	661	0.0139
EF_0512	793	0.0257	**EF_2458**	724	0.0276
EF_0629	4.21	4.48e-14	**EF_2460**	811	0.00126
EF_0630	4.94	4.41e-12	**EF_2893**	1.39	0.000704
EF_0686	824	0.00304	**EF_2919**	-656	0.0457
EF_0799	1.58	0.000117	**EF_2985**	4.51	4.48e-14
EF_0822	877	0.0101	**EF_3054**	-1.37	1.32e-05
EF_0829	1.24	0.00538	**EF_3079**	-1.76	6.08e-07
EF_0871	-858	0.0356	**EF_3106**	-1.97	0.000839
EF_0911	-918	0.00495	**EF_3108**	-2.04	0.0186
EF_0912	-1	0.0049	**EF_3109**	-1.64	4
EF_1022	1.4	12	**EF_3168**	607	0.0232
EF_1082	1.4	1.59e-07	**EF_3206**	1.28	0.00157
EF_1102	3	7.59e-05	**EF_3276**	-2.66	0.00394
EF_1153	-1.11	2.14e-07	**glnA**	1.05	0.0277
EF_1199	0.95	0.0482	**gspA-2**	-3.15	29
EF_1264	1.56	4.06e-05	**rplX**	1.06	0.000782
EF_1340	698	0.0345	**rpsF**	1.26	0.00806
EF_1352	1.66	1.73e-07	**rpsJ**	0.8	0.0158
EF_1511	-6.38	4.91e-08	**rpsM**	618	0.0441
EF_1513	-5.86	4.48e-14	**rpsN2**	1.45	0.00184
EF_1533	937	0.0465	**rpsS**	719	0.0154
sstT	-2.91	2.43e-06	**tag-2**	-1.45	0.00727
traB-2	-715	0.0273			

Table A.35: Significant fold changes at the protein level in *E. faecalis* between Urine+glc and CDM-LAB.

Genes	log fold change	adjusted p-value
argS	1.3	1.3e-05
cdd	778	0.0289
EF_0063	-4.86	3.1e-05
EF_0157	-2.17	3.14e-10
EF_0236	-2.45	0.00566
EF_0247	-847	0.0137
EF_0512	1.07	0.00901
EF_0575	2.63	0.000155
EF_0577	2.46	1.34e-13
EF_0647	1.28	0.0487
EF_0686	1.02	0.00224
EF_0797	-1.08	0.0047
EF_0822	1.35	0.000164
EF_1057	3.74	0.0299
EF_1058	3.39	7.37e-05
EF_1511	-8.84	2.49e-11
EF_1513	-5.51	1.58e-11
EF_1525	1.13	0.0096
EF_2074	3.37	1.34e-13
EF_2076	2.9	1.34e-13
EF_2174	1.16	0.000499
EF_2222	-1.05	0.0254
EF_2254	1.2	0.00734
EF_2458	1.34	2.52e-05
EF_3079	-1.77	7.43e-05
gatC	1.3	0.0315
proC.1	2.27	0.000136
rplX	1.14	0.00595
rpsM	762	0.0466
ssb-4	2.9	0.0392
tpiA	1.04	0.0282

Table A.36: Significant fold changes at the protein level in *S. pyogenes* between Plasma and CDM-LAB at pH 7.4.

Genes	**log fold change**	**adjusted p-value**	Genes	log fold change	adjusted p-value
accA	-2.13	0.0123	pyrF	4.04	0.0114
accC	-2.68	8.38e-08	pyrR	3.35	0.00089
accD	-1.86	0.00102	recX	-2.1	0.0171
adhE	1.84	0.0271	sgaT	-3.63	0.0463
carA	3.73	4.46e-10	ska	2.24	0.0431
carB	4.56	4.64e-10	Spy49_0176	4.74	2.72e-13
citD	-5.33	4.12e-10	Spy49_0212	5.31	0.00272
citE	-5.59	0.0452	Spy49_0441	-2.68	0.0201
codY	1.93	0.0111	Spy49_0579	-4.82	0.00413
dacA1	-2.98	0.000608	Spy49_0622	-3.81	2.06e-06
fabD	-2.35	0.000495	Spy49_0715	2.22	0.0123
fabF	-2.04	0.00785	Spy49_0756	-2.15	0.00603
fabG	-1.68	0.0157	Spy49_0774	-5.4	0.000129
fabH	-2.02	0.00876	Spy49_0794	-1.6	0.0254
fabK	-1.83	0.0339	Spy49_0822	-3.94	1.74e-08
fruK	1.53	0.0189	Spy49_0881	-7.63	0.00115
hasA	-2.86	0.00615	Spy49_0901	-4.78	0.000159
hasC	-2.89	0.000511	Spy49_0997c	-3.83	0.00544
lacD1	3.8	0.000544	Spy49_1092c	2.05	0.0307
mtsA	2.29	0.0011	Spy49_1224c	1.89	0.0499
mtsB	3.08	4.59e-10	Spy49_1297c	2.46	0.0183
nrdE.2	6.76	0.000407	Spy49_1298c	3.55	3.89e-10
nrdF.2	5.74	0.00569	Spy49_1299c	3.92	0.00782
oppA	1.63	0.0115	Spy49_1320c	-3.35	0.0144
pepC	2.12	0.0253	Spy49_1386c	3.98	0.0297
pepD	2.47	0.0221	Spy49_1396c	5.82	0.0153
phaB	-1.96	0.0272	Spy49_1405c	4.11	0.00311
plr	1.91	0.0161	Spy49_1458c	-5.53	0.000707
ptsH	2.38	0.0244	Spy49_1510c	-3.59	0.0195
purD	1.79	0.0263	Spy49_1525	-4.69	0.00442
pyrB	5.03	1.12e-07	Spy49_1609c	-1.43	0.0188
pyrE	4.15	0.00145	Spy49_1686	1.98	0.0373
Spy49_1784c	-3.86	0.000891			

Table A.37: Essential reactions in the model of *E. faecalis* constrained with the metabolic data from the batch culture experiment in CDM-LAB at pH 6.5.

PRPPS	NDPK3	UAMAS	Ex_ch4s__e	METGL_1	AGAT_EFA
THMabc	NDPK2	MCMAT5	GLYTRS	URIDK1	PROTRS
LYSTRS	NDPK1	MCMAT4	ETOHt1	HISt6	WTASI
DARTAL_EFA	HDDHL5	MCMAT2	UACGE	RPE	ACCOAC
HTDR6	HMGCOAS	OCMAT3	Ex_lac_L__e	BPPA2_L	HHDR7
DAGK_LPL	HMGCOAR	MCMAT8	MCMAT6	PPNCL	GLYt6
DHPS3	TYRt6	BIOMASS	HOCR3	RPI	GRTT
PMEVK	TYRTA	HODR8	BTMAT1	DHDPS	TPI
Ex_ala_L__e	ASPTRS_1	UAMAGS	DHFS	EPA_PS_EFA	MEVK
ADPTA	ASNTAL	MACPMT	DHFR	ALATA_L	NDPK4
HPPK	PYDXK	THRt6	TDPGDH	Ex_leu_L__e	G6PDHy
HHYR2	PZS	PKL	FTHFL	CPS_EFA_SYNTH	CPS_PS_EFA
SERTRS	PTAr	LTAS1	ALCD2x	Ex_ac__e	TDMAT6
WTASII	IPDDI	LTAS2	VALTRS	PAPPT1_A	ENO
SERD_L	CO2t	G1PTMT	METS	UDCPDPS	ASPTRS
DHNPA	ASPt6	ASNt6	PTPAT	PAPPT1_L	LDH_L
HDDR5	CH4St	DAPE	HOCHL3	CYSTRS	SERt6
TRPTRS	ILEt6	HHYHL2	BTNt2i	BPPA2	TAL
ACACT1	CRCT	HTDHL6	HDMAT7	DNAS_LPL	ASAD
Ex_trp_L__e	ALA_Lt6	CYSTGL	DALTAL_EFA	DMATT	GLUR
GAT1_EFA	FORt	ADPDS	PFL	ACt6	ILETRS
ALAR	DEMAT4	PGPP_LPL	DDL	ALKP_Efa	CYTK1
Ex_arg_L__e	FABM1	PYK	APAT	Ex_thm__e	LEUTRS
PGDH	FABM2	ALATA_Lr	UAPGR	TYRTRS	Ex_his_L__e
UDPG4E	Ex_co2__e	PGAMT	METTRS	VANB	GALU
CLPNS_LPL	BPPA1_L	HCO3E	PPND	PROTS_LPL_v6_0	LYSt6
GLUTRS	ASPK	LPGS_EFA	RNAS_LPL	HSDy	Ex_btn__e
HODHL8	PGM	PPCDC	Ex_ile_L__e	G1PACT	TDPDRE
Ex_gln_L__e	PGL	PHEt6	PGSA_LPL	HEMAT2	DAPDC
DEX_PS_EFA	PGK	CITt6	UAGCVT	Ex_cys_L__e	PGSYNTH
TMDS	Ex_glyclt__e	ALATRS	UDCPDP	CITL	
DPMVD	HHDHL7	GLYCLTt2r	DDMAT5	UAGDP	Ex_etoh__e
GLNTAL	DASYN_LPL	DTMPK	HDEHL4	Ex_val_L__e	PYRt2
ASPTA1	H2Ot5	HDER4	GCALDD	ACALDt	RBT5PDHy
PNTK	VALt6	OCDMAT8	Ex_glc__e	HBUR1	Ex_cit__e
UGLDDS1_A	DAGGT_LPL	GLUt6	MTHFD	HISTRS	GAPD
UAGPT1_A	kaasIII	ARGTRS	MTHFC	THRTRS	TKT1
UAGPT1_L	TRDR	UAAGLS1	HBUHL1	GTPCI	TKT2
UGLDDS1_L	ASNPTH	TAPGL4_EFA	DPCOAK	BPPA1	LEUt6
DHDPRy	METFR	MCMAT3	Ex_h2o__e	GHMT	CYSTL
PHETRS	MCMAT7	ACALD			

Table A.38: Potential drug targets in *E. faecalis* under CDM-LAB condition according to the Lib 7.

ADPTA	PYK	PGM	Ex_etoh__e	CO2t	APAT
SERD_L	PHEt6	PGL	PYRt2	ASPt6	PPND
PGDH	CITt6	PGK	Ex_cit__e	CH4St	Ex_glc__e
Ex_gln_L__e	GLUt6	H2Ot5	GAPD	ALA_Lt6	MTHFD
ASPTA1	LYSt6	ASNPTH	TKT1	FORt	MTHFC
DHDPRy	DAPDC	METFR	TKT2	Ex_co2__e	Ex_h2o__e
GLYt6	ETOHt1	G6PDHy	CYSTL	ASPK	THRt6
TPI	Ex_lac_L__e	ENO	METGL_1	HSDy	ASNt6
TYRt6	FTHFL	LDH_L	RPI	CITL	DAPE
TYRTA	ALCD2x	SERt6	DHDPS	ACALDt	CYSTGL
ASPTRS_1	METS	TAL	Ex_ac__e	GHMT	ADPDS
ASNTAL	PFL	ASAD	ACt6	ACALD	

Table A.39: Essential reactions in the model of *E. faecalis* constrained with the metabolic and transcriptome data from the batch culture experiment in Saliva.

PRPPS	GLU5K	OCBT	ASAD	HDER4	DHFS	RBT5PDHy	ALKP_Efa
ADCS	Ex_phe_L__e	IPDDI	IMPC	SHK3D	DHFR	GAPD	Ex_thm__e
ASNN	TMDS	CO2t	GLUR	DHQD	TDPGDH	AIRC	TYRTRS
THMabc	DPMVD	ASPt6	GARFT	OCDMAT8	FTHFL	Ex_ival__e	NACUP
PSCVT	GLNTAL	CH4St	UAMAS	DHQS	ALCD2x	TKT1	VANB
DARTAL_EFA	ASPTA1	ILEt6	MCMAT4	ILETA2	PAt	TKT2	LYSTRS
Ex_2mbut__e	PNTK	DPR	MCMAT2	ARGTRS	VALTRS	LEUt6	P5CRr
ADCL	UGLDDS1_A	CRCT	ILETA4	ILETA2t	METS	CYSTL	PRAGS
HTDR6	UAGPT1_A	ALA_Lt6	LEUTA	UAAGLS1	PTPAT	OMPDC	PROTS_LPL_v6_0
DAGK_LPL	UAGPT1_L	DEMAT4	OCMAT3	TAPGL4_EFA	HOCHL3	METGL_1	HSDy
DHPS3	UGLDDS1_L	VALTA	MCMAT8	MCMAT3	BTNt2i	URIDK1	FRDx
PMEVK	DHDPRy	FABM1	BIOMASS	PRFGS	PDUL	PYDAMK	PRASCS
ADPTA	AGAT_EFA	FABM2	HODR8	ILETRS	GLNS	HISt6	G1PACT
HPPK	PROTRS	Ex_co2__e	UAMAGS	CYTK1	LEUTA2t	Ex_acald__e	HEMAT2
HHYR2	WTASI	BPPA1_L	IMPD	GLUPRT	HDMAT7	HYPOE	CITL
SERTRS	ACCOAC	ASPK	MACPMT	LEUTRS	DALTAL_EFA	RPE	UAGDP
DAHPS	HHDR7	ILETA	THRt6	Ex_his_L__e	PDUW	ORPT	Ex_val_L__e
WTASII	GLYt6	PGM	PKL	GALU	DDL	BPPA2_L	ACALDt
DHNPA	LEUTA2	PGL	LTAS1	ARGORNt	APAT	PPNCL	Ex_pa__e
HDDR5	LEUTA3	PGK	LTAS2	LYSt6	UAPGR	RPI	HBUR1
TRPTRS	LEUTA4	Ex_glyclt__e	G1PTMT	SUCCt6	METTRS	DHDPS	CYSTH
ACACT1	GRTT	HHDHL7	DAPE	Ex_btn__e	PPND	EPA_PS_EFA	HISTRS
Ex_trp_L__e	TPI	DASYN_LPL	HHYHL2	TDPDRE	RNAS_LPL	CK	THRTRS
Ex_gly__e	MEVK	Ex_orn_L__e	HTDHL6	DAPDC	Ex_ile_L__e	ALATA_L	GTPCI
GAT1_EFA	NDPK4	H2Ot5	CYSTGL	PGSYNTH	PGSA_LPL	Ex_leu_L__e	BPPA1
ALAR	NDPK3	VALt6	ADPDS	Ex_ch4s__e	UAGCVT	CPS_EFA_SYNTH	GHMT
GLUDy	NDPK2	DAGGT_LPL	PGPP_LPL	METabc	UDCPDP	PAPPT1_A	ACALD
Ex_arg_L__e	NDPK1	kaasIII	PYK	GLYTRS	DDMAT5	UDCPDPS	MCMAT7
GMPS2	DUTPDP	TRDR	CHORS	ADSL2	HDEHL4	PAPPT1_L	PHETRS
NAPRT	HDDHL5	METFR	ALATA_Lr	ADSL1	GCALDD	ACKr	PRAIS
G5SADs	Ex_succ__e	G6PDHy	PGAMT	ETOHt1	Ex_glc__e	CYSTRS	
PGDH	HMGCOAS	CPS_PS_EFA	HCO3E	UACGE	MTHFD	BPPA2	
ASP1DC	HMGCOAR	TDMAT6	LPGS_EFA	AICART	MTHFC	DNAS_LPL	
UDPG4E	PC	ENO	PPCDC	Ex_lac_L__e	HBUHL1	GLNt6	
CLPNS_LPL	TYRt6	ASPTRS	PHEt6	ILETA3	DPCOAK	DMATT	
SHKK	PRDH	LDH_L	CITt6	HOCR3	Ex_h2o__e	ACt6	
GLUTRS	TYRTA	SERt6	ALATRS	BTMAT1	PGMT	PANB	
G5SD	PYDXK	ARGD	ADSS	MCMAT6	Ex_nac__e	PANC	
HODHL8	PZS	PHETA	GLYCLTt2r	PFLpa	Ex_etoh__e	PYDXPP	
DEX_PS_EFA	PTAr	TAL	DTMPK	MCMAT5	PYRt2	Ex_glu_L__e	

Table A.40: Potential drug targets in *E. faecalis* under artificial saliva condition according to the Lib 7.

ASNN	PC	ILETA4	ILETA2t	NH4t	DHDPS
Ex_nh4__e	TYRt6	THRt6	TRPDC	APAT	ACKr
Ex_2mbut__e	PRDH	DAPE	DAPDC	PPND	ACt6
VALTA3	CH4St	CYSTGL	METabc	MTHFD	PYDXPP
VALTA2	ILEt6	ADPDS	ILETA3	MTHFC	CITL
ADPTA	VALTA	ALATA_Lr	PFLpa	THRD_L	Ex_ibut__e
SERD_L	TRPAt	HCO3E	FTHFL	Ex_h2o__e	PHEDC
PHEAt	ILETA	PHEt6	PAt	PYRt2	METFR
VALTA4	H2Ot5	CITt6	METS	CYSTL	SERt6
DHDPRy	VALt6	ILETA2	PDUL	METGL_1	ARGD
PDUW	PHETA	VALTA2t			

Table A.41: Essential reactions in the model of *E. faecalis* constrained with the metabolic and transcriptome data from the batch culture experiment in Saliva+glc.

PRPPS	GLU5K	OCBT	ASAD	HDER4	DHFS	RBT5PDHy	ALKP_Efa
ADCS	Ex_phe_L__e	IPDDI	IMPC	SHK3D	DHFR	GAPD	Ex_thm__e
ASNN	TMDS	CO2t	GLUR	DHQD	TDPGDH	AIRC	TYRTRS
THMabc	DPMVD	ASPt6	GARFT	OCDMAT8	FTHFL	Ex_ival__e	NACUP
PSCVT	GLNTAL	CH4St	UAMAS	DHQS	ALCD2x	TKT1	VANB
DARTAL_EFA	ASPTA1	ILEt6	MCMAT4	ILETA2	PAt	TKT2	LYSTRS
Ex_2mbut__e	PNTK	DPR	MCMAT2	ARGTRS	VALTRS	LEUt6	P5CRr
ADCL	UGLDDS1_A	CRCT	ILETA4	ILETA2t	METS	CYSTL	PRAGS
HTDR6	UAGPT1_A	ALA_Lt6	LEUTA	UAAGLS1	PTPAT	OMPDC	PROTS_LPL_v6_0
DAGK_LPL	UAGPT1_L	DEMAT4	OCMAT3	TAPGL4_EFA	HOCHL3	METGL_1	HSDy
DHPS3	UGLDDS1_L	VALTA	MCMAT8	MCMAT3	BTNt2i	URIDK1	FRDx
PMEVK	DHDPRy	FABM1	BIOMASS	PRFGS	PDUL	PYDAMK	PRASCS
ADPTA	AGAT_EFA	FABM2	HODR8	ILETRS	GLNS	HISt6	G1PACT
HPPK	PROTRS	Ex_co2__e	UAMAGS	CYTK1	LEUTA2t	Ex_acald__e	HEMAT2
HHYR2	WTASI	BPPA1_L	IMPD	GLUPRT	HDMAT7	HYPOE	CITL
SERTRS	ACCOAC	ASPK	MACPMT	LEUTRS	DALTAL_EFA	RPE	UAGDP
DAHPS	HHDR7	ILETA	THRt6	Ex_his_L__e	PDUW	ORPT	Ex_val_L__e
WTASII	GLYt6	PGM	PKL	GALU	DDL	BPPA2_L	ACALDt
DHNPA	LEUTA2	PGL	LTAS1	ARGORNt	APAT	PPNCL	Ex_pa__e
HDDR5	LEUTA3	PGK	LTAS2	LYSt6	UAPGR	RPI	HBUR1
TRPTRS	LEUTA4	Ex_glyclt__e	G1PTMT	SUCCt6	METTRS	DHDPS	CYSTH
ACACT1	GRTT	HHDHL7	DAPE	Ex_btn__e	PPND	EPA_PS_EFA	HISTRS
Ex_trp_L__e	TPI	DASYN_LPL	HHYHL2	TDPDRE	RNAS_LPL	CK	THRTRS
Ex_gly__e	MEVK	Ex_orn_L__e	HTDHL6	DAPDC	Ex_ile_L__e	ALATA_L	GTPCI
GAT1_EFA	NDPK4	H2Ot5	CYSTGL	PGSYNTH	PGSA_LPL	Ex_leu_L__e	BPPA1
ALAR	NDPK3	VALt6	ADPDS	Ex_ch4s__e	UAGCVT	CPS_EFA_SYNTH	GHMT
GLUDy	NDPK2	DAGGT_LPL	PGPP_LPL	METabc	UDCPDP	PAPPT1_A	ACALD
Ex_arg_L__e	NDPK1	kaasIII	PYK	GLYTRS	DDMAT5	UDCPDPS	MCMAT7
GMPS2	DUTPDP	TRDR	CHORS	ADSL2	HDEHL4	PAPPT1_L	PHETRS
NAPRT	HDDHL5	METFR	ALATA_Lr	ADSL1	GCALDD	ACKr	PRAIS
G5SADs	Ex_succ__e	G6PDHy	PGAMT	ETOHt1	Ex_glc__e	CYSTRS	
PGDH	HMGCOAS	CPS_PS_EFA	HCO3E	UACGE	MTHFD	BPPA2	
ASP1DC	HMGCOAR	TDMAT6	LPGS_EFA	AICART	MTHFC	DNAS_LPL	
UDPG4E	PC	ENO	PPCDC	Ex_lac_L__e	HBUHL1	GLNt6	
CLPNS_LPL	TYRt6	ASPTRS	PHEt6	ILETA3	DPCOAK	DMATT	
SHKK	PRDH	LDH_L	CITt6	HOCR3	Ex_h2o__e	ACt6	
GLUTRS	TYRTA	SERt6	ALATRS	BTMAT1	PGMT	PANB	
G5SD	PYDXK	ARGD	ADSS	MCMAT6	Ex_nac__e	PANC	
HODHL8	PZS	PHETA	GLYCLTt2r	PFLpa	Ex_etoh__e	PYDXPP	
DEX_PS_EFA	PTAr	TAL	DTMPK	MCMAT5	PYRt2	Ex_glu_L__e	

Table A.42: Potential drug targets in *E. faecalis* under artificial saliva+glc condition according to the Lib 7.

ASNN	ARGORNt	PGK	TKT1	PTAr	MTHFC
Ex_2mbut__e	LYSt6	Ex_orn_L__e	TKT2	OCBT	Ex_h2o__e
ADPTA	DAPDC	H2Ot5	LEUt6	CO2t	Ex_etoh__e
Ex_gly__e	METabc	VALt6	CYSTL	ASPt6	PYRt2
GLUDy	ETOHt1	METFR	METGL_1	CH4St	GAPD
Ex_arg_L__e	Ex_lac_L__e	G6PDHy	Ex_acald__e	ILEt6	Ex_ival__e
G5SADs	ILETA3	ENO	RPE	ALA_Lt6	ADPDS
PGDH	PFLpa	LDH_L	RPI	Ex_co2__e	ALATA_Lr
G5SD	FTHFL	SERt6	DHDPS	ASPK	HCO3E
GLU5K	ALCD2x	ARGD	CK	ILETA	PHEt6
Ex_phe_L__e	PAt	PHETA	ALATA_L	PGM	CITt6
ASPTA1	METS	TAL	ACKr	PGL	ILETA2
DHDPRy	PDUL	ASAD	GLNt6	ACALDt	ILETA2t
GLYt6	GLNS	ILETA4	ACt6	CYSTH	PC
LEUTA2	LEUTA2t	LEUTA	PYDXPP	GHMT	TYRt6
LEUTA3	PDUW	THRt6	P5CRr	ACALD	PRDH
LEUTA4	APAT	PKL	HSDy	Ex_glc__e	TYRTA
TPI	PPND	DAPE	CITL	MTHFD	CYSTGL

Table A.43: Essential reactions in the model of *E. faecalis* constrained with the metabolic and transcriptome data from the batch culture experiment in Urine+fru+suc.

PRPPS	UAGPT1_A	VALt6	PPCDC	BTNt2i	CPS_EFA_SYNTH
ASNN	UAGPT1_L	DAGGT_LPL	PHEt6	LEUTA2t	FRUt6
THMabc	UGLDDS1_L	kaasIII	CITt6	HDMAT7	PAPPT1_A
DARTAL_EFA	AGAT_EFA	TRDR	ALATRS	DALTAL_EFA	UDCPDPS
Ex_2mbut__e	PROTRS	METFR	GLYCLTt2r	DDL	PAPPT1_L
HTDR6	WTASI	CPS_PS_EFA	DTMPK	APAT	ACKr
DAGK_LPL	ACCOAC	TDMAT6	HDER4	UAPGR	CYSTRS
DHPS3	HHDR7	ASPTRS	OCDMAT8	METTRS	BPPA2
PMEVK	GLYt6	LDH_L	ILETA2	RNAS_LPL	DNAS_LPL
ADPTA	LEUTA2	SERt6	ARGTRS	Ex_ile_L__e	DMATT
HPPK	LEUTA3	ARGD	ILETA2t	PGSA_LPL	ACt6
CYSt6	LEUTA4	TAL	UAAGLS1	UAGCVT	ALKP_Efa
HHYR2	GRTT	ASAD	TAPGL4_EFA	UDCPDP	Ex_thm__e
SERTRS	MEVK	GLUR	MCMAT3	DDMAT5	TYRTRS
WTASII	NDPK4	UAMAS	ILETRS	HDEHL4	NACUP
DHNPA	NDPK3	MCMAT4	CYTK1	GCALDD	VANB
HDDR5	NDPK2	MCMAT2	TRPDC	MTHFD	LYSTRS
TRPTRS	NDPK1	ILETA4	LEUTRS	MTHFC	P5CRr
ACACT1	HDDHL5	LEUTA	Ex_his_L__e	THRD_L	PROTS_LPL_v6_0
Ex_trp_L__e	HMGCOAS	OCMAT3	GALU	HBUHL1	G1PACT
GAT1_EFA	HMGCOAR	MCMAT8	Ex_btn__e	DPCOAK	HEMAT2
ALAR	PYDXK	BIOMASS	TDPDRE	Ex_h2o__e	Ex_cys_L__e
GLUDy	PZS	HODR8	DAPDC	PGMT	CITL
Ex_arg_L__e	IPDDI	UAMAGS	PGSYNTH	Ex_nac__e	UAGDP
G5SADs	CO2t	MACPMT	GLYTRS	PYRt2	Ex_val_L__e
UDPG4E	ASPt6	THRt6	XYLA	RBT5PDHy	Ex_trpa__e
CLPNS_LPL	ILEt6	LTAS1	UACGE	Ex_seL__e	HBUR1
GLUTRS	CRCT	LTAS2	ILETA3	Ex_ival__e	HISTRS
G5SD	DEMAT4	G1PTMT	HOCR3	TKT1	PGI
HODHL8	FABM1	DAPE	BTMAT1	TKT2	THRTRS
DEX_PS_EFA	FABM2	HHYHL2	MCMAT6	LEUt6	GTPCI
P15DAt	Ex_co2__e	HTDHL6	MCMAT5	CYSTL	BPPA1
GLU5K	TRPAt	CYSTGL	DHFS	URIDK1	GHMT
Ex_phe_L__e	BPPA1_L	ADPDS	DHFR	HISt6	MCMAT7
LYSDC	ILETA	PGPP_LPL	TDPGDH	BPPA2_L	PHETRS
TMDS	Ex_glyclt__e	ALATA_Lr	FTHFL	PPNCL	
DPMVD	HHDHL7	PGAMT	VALTRS	RPI	
GLNTAL	DASYN_LPL	HCO3E	METS	DHDPS	
PNTK	Ex_orn_L__e	Ex_p15da__e	PTPAT	EPA_PS_EFA	
UGLDDS1_A	H2Ot5	LPGS_EFA	HOCHL3	Ex_leu_L__e	

Table A.44: Potential drug targets in *E. faecalis* under artificial urine+fru+suc condition according to the Lib 7.

ASNN	ILETA2	LDH_L	CYSTL	MTHFD	LEUTA3
Ex_2mbut__e	ILETA2t	SERt6	DHDPS	MTHFC	LEUTA4
ADPTA	TRPDC	ARGD	ACKr	THRD_L	ASPt6
CYSt6	DAPDC	ILETA4	ACt6	PGMT	ILEt6
Ex_arg_L__e	XYLA	LEUTA	P5CRr	PYRt2	TRPAt
G5SADs	ILETA3	THRt6	Ex_cys_L__e	Ex_ival__e	ILETA
G5SD	FTHFL	DAPE	CITL	LEUt6	METFR
GLU5K	METS	CYSTGL	PGI	CITt6	APAT
GLYt6	LEUTA2t	ADPDS	GHMT	LEUTA2	

Table A.45: Essential reactions in the model of *E. faecalis* constrained with the metabolic and transcriptome data from the batch culture experiment in Urine+glc.

PRPPS	Ex_phe_L__e	ILEt6	OCMAT3	TAPGL4_EFA	APAT	ALATA_L
ADCS	TMDS	DPR	MCMAT8	MCMAT3	UAPGR	Ex_leu_L__e
THMabc	DPMVD	CRCT	BIOMASS	PRFGS	METTRS	CPS_EFA_SYNTH
Ex_nh4__e	GLNTAL	ALA_Lt6	HODR8	ILETRS	PPND	PAPPT1_A
PSCVT	ASPTA1	DEMAT4	UAMAGS	CYTK1	RNAS_LPL	UDCPDPS
DARTAL_EFA	PNTK	VALTA	IMPD	GLUPRT	Ex_ile_L__e	PAPPT1_L
ADCL	UGLDDS1_A	FABM1	MACPMT	LEUTRS	PGSA_LPL	ACKr
HTDR6	UAGPT1_A	FABM2	THRt6	Ex_his_L__e	UAGCVT	CYSTRS
DAGK_LPL	UAGPT1_L	Ex_co2__e	PKL	GALU	UDCPDP	BPPA2
DHPS3	UGLDDS1_L	BPPA1_L	LTAS1	ARGORNt	DDMAT5	DNAS_LPL
PMEVK	DHDPRy	PGM	LTAS2	LYSt6	HDEHL4	DMATT
ADPTA	AGAT_EFA	PGL	Ex_h__e	SUCCt6	GCALDD	ACt6
HPPK	PROTRS	PGK	G1PTMT	Ex_btn__e	MTHFD	PANB
CYSt6	WTASI	Ex_glyclt__e	ASNt6	TDPDRE	MTHFC	PANC
HHYR2	ACCOAC	HHDHL7	DAPE	DAPDC	HBUHL1	PYDXPP
SERTRS	HHDR7	DASYN_LPL	HHYHL2	PGSYNTH	DPCOAK	ALKP_Efa
DAHPS	GLYt6	Ex_orn_L__e	HTDHL6	METabc	Ex_h2o__e	Ex_thm__e
WTASII	GRTT	H2Ot5	CYSTGL	GLYTRS	PGMT	TYRTRS
SERD_L	MEVK	VALt6	ADPDS	ADSL2	Ex_nac__e	NACUP
DHNPA	NDPK4	DAGGT_LPL	PGPP_LPL	ADSL1	PYRt2	VANB
HDDR5	NDPK3	kaasIII	CHORS	UACGE	RBT5PDHy	LYSTRS
TRPTRS	NDPK2	TRDR	ALATA_Lr	AICART	GAPD	PRAGS
ACACT1	NDPK1	ASNPTH	PGAMT	HOCR3	AIRC	PROTS_LPL_v6_0
Ex_trp_L__e	DUTPDP	METFR	HCO3E	BTMAT1	TKT1	FRDx
Ex_gly__e	HDDHL5	G6PDHy	LPGS_EFA	MCMAT6	TKT2	PRASCS
GAT1_EFA	Ex_succ__e	CPS_PS_EFA	PPCDC	MCMAT5	LEUt6	G1PACT
Ex_met_L__e	HMGCOAS	TDMAT6	PHEt6	DHFS	CYSTL	HEMAT2
ALAR	HMGCOAR	ENO	CITt6	DHFR	OMPDC	CITL
Ex_arg_L__e	PC	ASPTRS	ALATRS	TDPGDH	URIDK1	UAGDP
GMPS2	TYRt6	SERt6	ADSS	FTHFL	PYDAMK	Ex_val_L__e
NAPRT	PRDH	ARGD	GLYCLTt2r	VALTRS	HISt6	HBUR1
PGDH	TYRTA	PHETA	DTMPK	METS	HYPOE	HISTRS
ASP1DC	ASPTRS_1	TAL	HDER4	PTPAT	RPE	PGI
UDPG4E	ASNTAL	ASAD	SHK3D	HOCHL3	ORPT	THRTRS
CLPNS_LPL	PYDXK	IMPC	DHQD	BTNt2i	BPPA2_L	GTPCI
SHKK	PZS	GLUR	OCDMAT8	HDMAT7	PPNCL	BPPA1
GLUTRS	OCBT	GARFT	GLUt6	DALTAL_EFA	RPI	GHMT
HODHL8	IPDDI	UAMAS	DHQS	PFL	DHDPS	ACALD
Ex_gln_L__e	CO2t	MCMAT4	ARGTRS	NH4t	EPA_PS_EFA	MCMAT7
DEX_PS_EFA	ASPt6	MCMAT2	UAAGLS1	DDL	CK	PPDK
PHETRS	PRAIS					

Table A.46: Potential drug targets in *E. faecalis* under artificial urine+glc condition according to the Lib 7.

Ex_nh4__e	DAPE	ASPTRS_1	APAT	PHETA	ACt6
ADPTA	CYSTGL	ASNTAL	PPND	TAL	PYDXPP
SERD_L	ADPDS	OCBT	MTHFD	ASAD	CITL
Ex_gly__e	ALATA_Lr	CO2t	MTHFC	THRt6	PGI
Ex_arg_L__e	HCO3E	ASPt6	Ex_h2o__e	Ex_h__e	GHMT
PGDH	PHEt6	ALA_Lt6	PYRt2	ASNt6	PPDK
Ex_gln_L__e	CITt6	Ex_co2__e	TKT1	TYRt6	METS
Ex_phe_L__e	GLUt6	PGL	TKT2	PRDH	PFL
ASPTA1	ARGORNt	Ex_orn_L__e	CYSTL	TYRTA	NH4t
DHDPRy	LYSt6	H2Ot5	RPE	G6PDHy	CK
GLYt6	DAPDC	ASNPTH	RPI	SERt6	ALATA_L
PC	FTHFL	METFR	DHDPS	ARGD	ACKr

Table A.47: Essential reactions in the model of *S. pyogenes* constrained with the metabolic data from the batch culture experiment in CDM-LAB at pH 6.5.

CITt6	EX_lys_L__e	HDEHL4	ADK1	LYSTRS	EX_tyL__e
ALATRS	RBK	GRTT	METACH	TRPTRS	ACALD
PRPPS	ALAR	XPPT	FABM1	FABM	ENO
ASNTRS	TDMAT6	MEVK	FABM2	SERt6	UAGPT1
DTMPK	DAGK_LLA	NDPK3	RPE	PGI	PHETRS
HDER4	PGGT2	NDPK2	LTAS_LLA	LPGS_LLA	PPCDC
CPSS_LLA	NAPRT	NDPK1	NNAT	EX_lac_L__e	L_LACt2
THMabc	ORPT	DUTPDP	DPCOAK	akg_demand	PHEt6
DALTAL_LLA	HOCR3	DNAS_LLA	RPI	HODR8	
EX_met_L__e	BTMAT1	G1PTMT	PGM	PTPAT	
EX_nac__e	PPNCL	HBUHL1	RNAS_LLA	UAGCVT	
EX_py_e	DHFR	GLNabc	PGK	PROTS_LLA	
HTDR6	TDPGDH	CYSM	HHDHL7	TMDPK	
MCMAT7	ALCD2x	HMGCOAS	EX_glu_L__e	G1PACT	
ARGTRS	GLUTRS	HMGCOAR	EX_xan__e	UAMAS	
EX_arg_L__e	VALTRS	AHCYSNS	H2Ot5	MCMAT6	
MCMAT5	EX_pnto__e	TYRt6	EX_phe_L__e	MCMAT4	
PMEVK	HODHL8	PGMT	CFAS180	MCMAT2	
HEMAT2	HOCHL3	ALAALA	UDPG4E	OCMAT3	
EX_etoh__e	OCDMAT8	PFK	UDCPDPS	MCMAT8	
PAPPT1	THRTRS	PYRt2	ACACT1	UAGDP	
DEMAT4	HDMAT7	NDPK4	HSAT	UAMAGS	
ILETRS	TMDS	MCMAT3	ACKr	MACPMT	
CYTK1	GLNTAL	UAAGLS1	PPA	DDMAT5	
HHYR2	ASPTA1	GLUR	GMPS2	CYSTRS	
SERTRS	DMATT	GAT1_LLA	GALTAL	EX_h2o__e	
HISt6	UAPGR	EX_val_L__e	FBA	PGPP_LLA	
THRS	NADS1	PGSA_LLA	GLYK	EX_trp_L__e	
EX_gln_L__e	ACt6	PTAr	PNTK	HBUR1	
THRt6	DPMVD	biomass	EX_his_L__e	GAPD	
CITL	UGT1_LLA	UGLDDS1	TRDR	PNTOt2	
RHC	GF6PTA	SERAT	METTRS	HTDHL6	
IPDDI	PROTRS	HHYHL2	DASYN_LLA	HISTRS	
GALU	HSK	LEUt6	TYRTRS	METAT	
TDPDRR	EX_leu_L__e	AGAT_LLA	VALt6	XANt2	
LYSt6	ACCOAC	CLPNS_LLA	EX_thL__e	TRPt6	
HDDR5	HHDR7	OMPDC	ASPTRS	kaasIII	
LEUTRS	UDCPDP	EX_thm__e	NACUP	PGAMT	
GLYTRS	EX_ac__e	ILEt6	LDH_L	HDDHL5	
ETOHt1	TPI	EX_ile_L__e		TDPDRE	

Table A.48: Potential drug targets in *S. pyogenes* under CDM-LAB condition at pH 6.5 according to the Lib 7.

R_CITt6	R_PYRt2
R_EX_pyr__e	R_PGM
R_EX_etoh__e	R_PGK
R_EX_gln_L__e	R_EX_glu_L__e
R_CITL	R_FBA
R_ETOHt1	R_LDH_L
R_ALCD2x	R_SERt6
R_ASPTA1	R_PGI
R_ACt6	R_EX_lac_L__e
R_EX_ac__e	R_akg_demand
R_TPI	R_GAPD
R_GLNabc	R_ACALD
R_PFK	R_ENO
R_L_LACt2	

Table A.49: Essential reactions in the model of *S. pyogenes* constrained with the metabolic data from the batch culture experiment in CDM-LAB at pH 7.4.

CITt6	GLYTRS	GLYt6	CO2t	PNTK	GFT1
ALATRS	ETOHt1	TPI	LEUt6	EX_his_L__e	PGPP_LLA
PRPPS	EX_lys_L__e	HDEHL4	EX_ala_L__e	EX_h__e	EX_trp_L__e
ASNTRS	EX_co2__e	GRTT	AGAT_LLA	TRDR	HBUR1
DTMPK	RBK	XPPT	CLPNS_LLA	METTRS	GAPD
HDER4	ALAR	MEVK	OMPDC	DASYN_LLA	PNTOt2
CPSS_LLA	TDMAT6	MTHFD	EX_thm__e	TYRTRS	HTDHL6
ASNN	DAGK_LLA	NDPK3	ILEt6	VALt6	HISTRS
THMabc	PGGT2	NDPK2	ALA_Lt6	EX_orot__e	METAT
DALTAL_LLA	ORPT	NDPK1	EX_ile_L__e	EX_thL__e	XANt2
EX_met_L__e	HOCR3	DUTPDP	URIDK1	ASPTRS	PYK
HTDR6	BTMAT1	DNAS_LLA	ADK1	LDH_L	TRPt6
MCMAT7	PPNCL	ASPt6	METACH	LYSTRS	kaasIII
ARGTRS	DHFR	G1PTMT	FABM1	TRPTRS	PGAMT
EX_arg_L__e	OROt6	HBUHL1	FABM2	FABM	HDDHL5
MCMAT5	TDPGDH	GLNabc	FGF	SERt6	TDPDRE
PMEVK	ALCD2x	CYSM	RPE	PGI	EX_tyL__e
HEMAT2	GLUTRS	HMGCOAS	LTAS_LLA	LPGS_LLA	ACALD
EX_etoh__e	ASPCT	HMGCOAR	DPCOAK	EX_lac_L__e	PDH
PAPPT1	VALTRS	AHCYSNS	RPI	akg_demand	ENO
DEMAT4	EX_pnto__e	TYRt6	GHMT	HODR8	UAGPT1
ILETRS	HODHL8	PGMT	PGM	PTPAT	PHETRS
CYTK1	O2t	ALAALA	RNAS_LLA	UAGCVT	PPCDC
HHYR2	HOCHL3	PFK	PGK	PROTS_LLA	L_LACt2
SERTRS	OCDMAT8	PYRt2	HHDHL7	TMDPK	PHEt6
HISt6	THRTRS	NDPK4	EX_xan__e	G1PACT	
THRS	HDMAT7	MCMAT3	H2Ot5	UAMAS	
EX_gln_L__e	TMDS	UAAGLS1	ALATA_L	MCMAT6	
THRt6	GLNTAL	GLUR	EX_phe_L__e	MCMAT4	
CITL	ASPTA1	GAT1_LLA	CFAS180	MCMAT2	
SERD_L	DMATT	EX_gly__e	UDPG4E	OCMAT3	
RHC	UAPGR	EX_val_L__e	UDCPDPS	MCMAT8	
IPDDI	DPMVD	ADEt2	FITHFC	UAGDP	
GALU	UGT1_LLA	EX_o2__e	ACACT1	EX_ade__e	
TDPDRR	GF6PTA	PGSA_LLA	HSAT	UAMAGS	
FAH	PROTRS	biomass	PPA	MACPMT	
LYSt6	HSK	UGLDDS1	GMPS2	DDMAT5	
HDDR5	EX_leu_L__e	SERAT	GALTAL	CYSTRS	
DHORTS	ACCOAC	HHYHL2	FBA	EX_h2o__e	
LEUTRS	HHDR7		GLYK	ASNt6	

Table A.50: Potential drug targets in *S. pyogenes* under CDM-LAB condition at pH 7.4 according to the Lib 7.

CITt6	ALA_Lt6
ASNN	FGF
EX_etoh__e	GHMT
EX_gln_L__e	PGM
CITL	PGK
SERD_L	H2Ot5
FAH	ALATA_L
DHORTS	FITHFC
ETOHt1	FBA
OROt6	EX_h__e
ALCD2x	EX_orot__e
ASPCT	LDH_L
O2t	SERt6
ASPTA1	PGI
GLYt6	EX_lac_L__e
TPI	akg_demand
MTHFD	EX_h2o__e
ASPt6	ASNt6
GLNabc	GFT1
PFK	GAPD
PYRt2	ACALD
EX_gly__e	PDH
EX_o2__e	ENO
EX_ala_L__e	L_LACt2

Table A.51: Essential reactions in the model of *S. pyogenes* constrained with the metabolic and transcriptome data from the batch culture experiment in Saliva.

CITt6	GALU	HSK	CYSM	TYRTRS	PNTOt2
ALATRS	TDPDRR	EX_leu_L__e	EX_thm__e	VALt6	HTDHL6
PRPPS	LYSt6	ACCOAC	ILEt6	EX_thL__e	HISTRS
ASNTRS	HDDR5	HHDR7	SERAT	ASPTRS	METAT
DTMPK	DHORTS	UDCPDP	EX_ile_L__e	NACUP	XANt2
FOLt	LEUTRS	TPI	URIDK1	LDH_L	THRTRS
HDER4	EX_cyn__e	HDEHL4	ADK1	LYSTRS	TRPt6
CPSS_LLA	GLYTRS	GRTT	ORPT	TRPTRS	kaasIII
HCYSMT	EX_lys_L__e	XPPT	METACH	FABM	PGAMT
PFL	RBK	GAT1_LLA	FABM1	SERt6	HDDHL5
EX_fol__e	ALAR	NDPK4	FABM2	LPGS_LLA	TDPDRE
THMabc	TDMAT6	NDPK3	EX_cit__e	EX_lac_L__e	EX_tyL__e
DALTAL_LLA	DAGK_LLA	NDPK2	RPE	HODR8	CYNt
EX_met_L__e	PGGT2	NDPK1	LTAS_LLA	PGSA_LLA	ENO
EX_nac__e	NAPRT	DUTPDP	NNAT	EX_pi__e	UAGPT1
GLUt6	HOCR3	DNAS_LLA	DPCOAK	PTPAT	PHETRS
HTDR6	BTMAT1	ASPt6	RPI	ADPT	PPCDC
MCMAT7	PPNCL	G1PTMT	PGM	UDCPDPS	L_LACt2
ARGTRS	DHFR	HBUHL1	RNAS_LLA	UAGCVT	PHEt6
EX_arg_L__e	TDPGDH	GLNabc	PGK	PROTS_LLA	
MCMAT5	GLUTRS	HMGCOAS	ASNS2	TMDPK	
PMEVK	VALTRS	HMGCOAR	HHDHL7	G1PACT	
DASYN_LLA	EX_pnto__e	AHCYSNS	EX_xan__e	UAMAS	
ATPS3r	HODHL8	TYRt6	CYSTGL1	MCMAT6	
PAPPT1	HOCHL3	PGMT	EX_phe_L__e	MCMAT4	
DEMAT4	OCDMAT8	MEVK	CFAS180	HEMAT2	
ILETRS	HDMAT7	PYRt2	UDPG4E	MCMAT2	
CYTK1	TMDS	MCMAT3	ACACT1	OCMAT3	
CYSt6	PFK	UAAGLS1	HSAT	MCMAT8	
HHYR2	GLNTAL	GLUR	ACKr	UAGDP	
SERTRS	ASPTA1	EX_val_L__e	EX_his_L__e	EX_ade__e	
HISt6	DMATT	ADEt2	PPA	UAMAGS	
THRS	UAPGR	PTAr	GMPS2	MACPMT	
EX_gln_L__e	NADS1	biomass	GALTAL	DDMAT5	
THRt6	ACt6	UGLDDS1	FBA	CYSTRS	
CITL	DPMVD	HHYHL2	GLYK	ASNt6	
SERD_L	UGT1_LLA	LEUt6	PNTK	PGPP_LLA	
RHC	GF6PTA	AGAT_LLA	ALAALA	EX_trp_L__e	
DHFOR2	PROTRS	CLPNS_LLA	TRDR	HBUR1	
IPDDI	TCYS	OMPDC	METTRS	GAPD	

Table A.53: Essential reactions in the model of *S. pyogenes* constrained with the metabolic and transcriptome data from the batch culture experiment in Saliva+glc.

CITt6	GALU	ACt6	MCMAT3	HHDHL7	ADPT
ALATRS	TDPDRR	DPMVD	UAAGLS1	EX_xan__e	UDCPDPS
PRPPS	FAH	UGT1_LLA	NH4DIS	H2Ot5	UAGCVT
ASNTRS	LYSt6	GF6PTA	GLUR	CYSTGL1	PROTS_LLA
DTMPK	HDDR5	EX_actn__e	EX_gly__e	ALATA_L	TMDPK
FOLt	LEUTRS	PROTRS	EX_val_L__e	EX_phe_L__e	G1PACT
HDER4	EX_cyn__e	TCYS	ADEt2	CFAS180	UAMAS
CPSS_LLA	GLYTRS	HSK	EX_o2__e	UDPG4E	MCMAT6
ASNN	ETOHt1	EX_leu_L__e	PTAr	FITHFC	MCMAT4
HCYSMT	EX_lys_L__e	ACCOAC	biomass	ACACT1	HEMAT2
EX_fol__e	RBK	HHDR7	UGLDDS1	HSAT	MCMAT2
THMabc	ALAR	UDCPDP	ALA_Lt6	ACKr	OCMAT3
DALTAL_LLA	TDMAT6	GLYt6	HHYHL2	EX_his_L__e	MCMAT8
EX_met_L__e	NH3t	TPI	LEUt6	PPA	UAGDP
EX_nac__e	DAGK_LLA	HDEHL4	EX_ala_L__e	GMPS2	EX_ade__e
HTDR6	PGGT2	GRTT	AGAT_LLA	EX_btd_R_e	UAMAGS
MCMAT7	NAPRT	XPPT	CLPNS_LLA	GALTAL	MACPMT
ARGTRS	HOCR3	ACTNdiff	OMPDC	FBA	DDMAT5
EX_arg_L__e	BTMAT1	GAT1_LLA	CYSM	GLYK	CYSTRS
MCMAT5	PPNCL	NDPK4	EX_thm__e	EX_seL__e	EX_h2o__e
PMEVK	DHFR	NDPK3	ILEt6	PNTK	ASNt6
DASYN_LLA	TDPGDH	NDPK2	SERAT	ALAALA	PGPP_LLA
EX_etoh__e	ALCD2x	NDPK1	EX_ile_L__e	TRDR	EX_trp_L__e
ATPS3r	GLUTRS	DUTPDP	URIDK1	METTRS	HBUR1
PAPPT1	VALTRS	DNAS_LLA	ADK1	TYRTRS	GAPD
DEMAT4	EX_pnto__e	ASPt6	ORPT	VALt6	PNTOt2
ILETRS	HODHL8	G1PTMT	METACH	EX_thL__e	HTDHL6
CYTK1	O2t	BTDD_RR	FABM1	ASPTRS	HISTRS
CYSt6	HOCHL3	HBUHL1	FABM2	NACUP	METAT
HHYR2	MTHFD	GLNabc	FGF	LYSTRS	XANt2
SERTRS	OCDMAT8	FORt2	RPE	TRPTRS	THRTRS
HISt6	HDMAT7	HMGCOAS	LTAS_LLA	FABM	TRPt6
THRS	TMDS	HMGCOAR	NNAT	SERt6	kaasIII
EX_gln_L__e	EX_fo_e	AHCYSNS	DPCOAK	PGI	PGAMT
THRt6	PFK	TYRt6	RPI	LPGS_LLA	HDDHL5
CITL	GLNTAL	PGMT	EX_asn_L__e	akg_demand	TDPDRE
SERD_L	ASPTA1	BTDt1_RR	GHMT	HODR8	EX_tyL__e
RHC	DMATT	G6PDA	PGM	PGSA_LLA	ACALD
DHFOR2	UAPGR	MEVK	RNAS_LLA	EX_pi__e	CYNt
IPDDI	NADS1	PYRt2	PGK	PTPAT	ENO
UAGPT1	GFT1	PHETRS	PPCDC	L_LACt2	PHEt6

Table A.54: Potential drug targets in *S. pyogenes* under artificial saliva+glc condition according to the Lib 7.

CITt6	EX_gly__e	PFK	FITHFC	RBK	PGM
PRPPS	EX_o2__e	ASPTA1	ACKr	NH3t	PGK
ASNN	PTAr	ACt6	PPA	O2t	CYSTGL1
HCYSMT	ALA_Lt6	TCYS	FBA	MTHFD	ALATA_L
ATPS3r	EX_ala_L__e	GLYt6	EX_ser_L__e	NH4DIS	GAPD
CYSt6	CYSM	TPI	SERt6	ENO	METAT
EX_gln_L__e	SERAT	ASPt6	PGI	GFT1	CYNt
CITL	ADK1	GLNabc	akg_demand	L_LACt2	GHMT
SERD_L	FGF	AHCYSNS	ADPT	FAH	EX_asn_L__e
PYRt2	ASNt6	EX_cyn__e			

Table A.55: Essential reactions in the model of *S. pyogenes* constrained with the metabolic and transcriptome data from the batch culture experiment in Plasma.

CITt6	GALU	UGT1_LLA	LEUt6	TRDR	ASNt6
ALATRS	TDPDRR	GF6PTA	AGAT_LLA	METTRS	PGPP_LLA
PRPPS	FAH	PROTRS	CLPNS_LLA	TYRTRS	EX_trp_L__e
ASNTRS	LYSt6	HSK	CYSM	VALt6	HBUR1
DTMPK	HDDR5	EX_leu_L__e	EX_thm__e	EX_orot__e	GAPD
FOLt	DHORTS	ACCOAC	ILEt6	EX_thL__e	PNTOt2
HDER4	LEUTRS	HHDR7	SERAT	ASPTRS	HTDHL6
CPSS_LLA	GLYTRS	UDCPDP	EX_ile_L__e	NACUP	HISTRS
ASNN	EX_lys_L__e	GLYt6	URIDK1	LDH_L	METAT
HCYSMT	RBK	TPI	ADK1	LYSTRS	THRTRS
PFL	ALAR	HDEHL4	METACH	TRPTRS	TRPt6
EX_fol__e	TDMAT6	GRTT	FABM1	FABM	kaasIII
THMabc	DAGK_LLA	GAT1_LLA	FABM2	SERt6	PGAMT
DALTAL_LLA	PGGT2	NDPK4	FGF	PGI	HDDHL5
EX_met_L__e	NAPRT	NDPK3	RPE	LPGS_LLA	TDPDRE
EX_nac__e	HOCR3	NDPK2	LTAS_LLA	EX_lac_L__e	EX_tyL__e
GLUt6	BTMAT1	NDPK1	NNAT	HODR8	ENO
HTDR6	PPNCL	DUTPDP	DPCOAK	PGSA_LLA	UAGPT1
MCMAT7	DHFR	DNAS_LLA	RPI	EX_pi__e	GFT1
ARGTRS	OROt6	ASPt6	GHMT	PTPAT	PHETRS
EX_arg_L__e	TDPGDH	G1PTMT	PGM	ADPT	PPCDC
MCMAT5	GLUTRS	HBUHL1	RNAS_LLA	UDCPDPS	L_LACt2
PMEVK	ASPCT	GLNabc	PGK	UAGCVT	PHEt6
DASYN_LLA	VALTRS	FORt2	HHDHL7	PROTS_LLA	
PAPPT1	EX_pnto__e	HMGCOAS	EX_glu_L__e	TMDPK	
DEMAT4	HODHL8	HMGCOAR	EX_phe_L__e	G1PACT	
ILETRS	HOCHL3	AHCYSNS	CFAS180	UAMAS	
CYTK1	MTHFD	TYRt6	UDPG4E	G3PD1	
HHYR2	OCDMAT8	PGMT	FITHFC	MCMAT6	
SERTRS	HDMAT7	MEVK	ACACT1	MCMAT4	
HISt6	TMDS	PYRt2	HSAT	HEMAT2	
THRS	EX_fo_e	MCMAT3	ACKr	MCMAT2	
G3PD2	PFK	UAAGLS1	EX_his_L__e	OCMAT3	
EX_gln_L__e	GLNTAL	GLUR	PPA	MCMAT8	
THRt6	ASPTA1	EX_val_L__e	GALTAL	UAGDP	
CITL	DMATT	ADEt2	FBA	EX_ade__e	
SERD_L	UAPGR	PTAr	GLYK	UAMAGS	
RHC	NADS1	biomass	PNTK	MACPMT	
DHFOR2	ACt6	UGLDDS1	ALAALA	DDMAT5	
IPDDI	DPMVD	HHYHL2	EX_h__e	CYSTRS	

Table A.56: Potential drug targets in *S. pyogenes* under natural human plasma condition according to the Lib 7.

CITt6	CITL	MTHFD	FORt2	PGK	EX_lac_L__e
PRPPS	SERD_L	PFK	AHCYSNS	EX_glu_L__e	ADPT
ASNN	FAH	ASPTA1	PYRt2	FITHFC	ASNt6
HCYSMT	DHORTS	GLYt6	ADK1	PPA	GAPD
PFL	RBK	TPI	FGF	FBA	METAT
GLUt6	OROt6	ASPt6	GHMT	EX_orot__e	ENO
EX_gln_L__e	ASPCT	GLNabc	PGM	SERt6	GFT1
L_LACt2					

Appendix B

Scripts

The following scripts were developed to produce the results of this thesis. Theses scripts are were developed for inspecting the solution space of genome-scale metabolic models, using Matlab. The scripts are obtained from my previous publication [Loghmani et al., 2022], and can also be accessed from my github repository : https://github.com/babakml/FBA$_{p}erturbation$.

```
1
2  function [fbasol_fin, randval_fin] = perturb(model_n, solver_n, method)
3
4  %pertub function discovers the effect of random perturbation in
5  %FBA/FVA results. It selects 10 random values in the feasible interval
       of
6  %each variable reaction (with an interval larger than 0.000001), fixes
7  %reactions in those values and collect the flux distribution profile of
       the
8  %whole metabolic network
9
10 %USAGE
11
12 %    [fbasol_fin, randval_fin] = perturb(model_n, solver_n, method)
13
14
15 %INPUT
16
17 % model_n: metabolic model in SBML format
18 % solver_n: solver name
19 % method: 'fba' or 'fva', to determine the method of use to obtain flux
20 % distributions
21
22 % OUTPUT
23
24 % fbasol_fin: flux distribution profiles following 10 perturbation in
       each
25 % variable reaction
26
27 % randval_fin: random values at which variable reactions were fixed
28
```

```
%Example:

% [fbasol_fin, randval_fin] = perturb('mut-chem.xml','ibm_cplex','fba')

% Authors:

% Seyed Babak Loghmani

% Last updated: August 2021




initCobraToolbox;

%defining the solver
changeCobraSolver(solver_n);
model = readCbModel(model_n);

[minFluxF1, maxFluxF1, optsol, ret, fbasol, fvamin, fvamax,
   statussolmin, statussolmax] = fastFVA(model);




da = maxFluxF1;
db = minFluxF1;
fbasol_fin = [];
randval_fin =[];
dif = (da - db);

rxn_n = numel(model.rxns);

%runing the model while fixing each reaction in 10 random values
for i =1:rxn_n
    fbasol_m = [];
            randval_m = [];
            dif = (da(i) - db(i));
    if dif > 0.000001
    if dif ~= 0
        for j = 1:10

        %finding a random value
        randval = random('unif',db(i),da(i));
        %saving the original lower and upper bounds
        pu = model.ub(i);
        pl = model.lb(i);
        %fixing the reaction at the randomly selected value
```

```
        model.ub(i) = randval;
        model.lb(i) = randval;
        i
        da(i)
        db(i)


        da(i) - db(i)
        %collecting flux distributions using FBA
        if method == 'fba'
            sol_dist = optimizeCbModel(model);
            fbasol_m = [fbasol_m, sol_dist.v];

        %collecting flux distributions using FVA
        elseif  method == 'fva'
            [minFluxF1, maxFluxF1, optsol, ret, fbasol, fvamin, fvamax,
               statussolmin, statussolmax] = fastFVA(model);
            fbasol_m = [fbasol_m, fbasol];
        end
        %saving the random value in randval_m
        randval_m = [randval_m, randval];


        %bringing the original uppre and lower bound back
        model.ub(i) = pu;
        model.lb(i) = pl;

        end

        %saving the flux distribution profiles & random values for 10
           perturbation in one
        %reaction in fbasol_fin & randval_fin, respectively
        fbasol_fin = [fbasol_fin; fbasol_m];
        randval_fin = [randval_fin; randval_m];




    end
    end
end

save fbasol_fin.dat fbasol_fin -ascii -double
save randval_fin.dat randval_fin -ascii -double
```

```
function [per_result] = calculation(model_n, fbasol_fin, randval_fin,
   method)

% calculation finds the optimal values in flux distribution profiles
```

```
    following perturbation that
% are significantly different (>0.05 Or >0.000001) from the original
    FBA/FVA flux distribution
% profiles.

%USAGE:
%     [per_result] = calculation(model, fbasol_fin, randval_fin)

%INPUT

%model: metabolic model in SBML format

%fbasol_fin: flux distribution profiles following perturbation

%randval_fin: random values at which variable reactions were fixed

%method: 'fba' or 'fva', to determine the method of use to obtain flux

%OUTPUT:

%per_result: significantly different flux values in a sorted format

%Example:

% [per_result] =
    calculation('mut-chem.xml','fbasol_fin.dat','randval_fin.dat','fba')

% Authors:

% Seyed Babak Loghmani

% Last updated: August 2021

initCobraToolbox;

model = readCbModel(model_n);
load(fbasol_fin)
load(randval_fin)

%FVA
[minFluxF1, maxFluxF1, optsol, ret, fbasol, fvamin, fvamax,
    statussolmin, statussolmax] = fastFVA(model);

if method == 'fva'
    opt_sol = fbasol;
elseif method == 'fba'
    sol_dist = optimizeCbModel(model);
    opt_sol = sol_dist.v;
```

```
48 end
49
50 %finding reactions with fva intervals larger that 0.000001
51 fva_n = maxFluxF1 - minFluxF1;
52 r=fva_n > 0.000001;
53 fva_n_f=find(r);
54
55 n = numel(model.rxns);
56 opt_som = [];
57 sols = [];
58 per_result = [];
59 %opt_sol = fbasol(1:n);
60 opt_sol_abs = abs(opt_sol);
61
62 % defining the 5% tolerance
63 tol = 5*opt_sol_abs/100;
64 s = size(randval_fin);
65 f = 0;
66 rxn_n = numel(model.rxns);
67
68
69 %finding and sorting significant flux changes, assigning their indices
       and random values in which they
70 %got fixed
71
72 num_f_p = numel(fva_n_f)/50;
73 num_f = ceil(num_f_p);
74 rg = 1:num_f;
75 c=1;
76 file_names = [];
77
78 for i = 1:numel(fva_n_f)
79     k = i-1;
80
81     l = (k*n) + 1;%lowe3r end of the data set
82     h = (i*n);%higher end of the data set
83
84
85 if i == rg(c)*50%counter to save the data regarding ecery 50 reactions
       at once
86     f = 1
87     c = c+1;
88 elseif c == max(rg)
89
90     if i == numel(fva_n_f)
91        f = 1
92     end
93 end
```

```
94
95
96  if i ==1
97
98      sols = fbasol_fin(1:rxn_n,1:10);%extracting the data for the first
            reaction
99  else
100     sols = fbasol_fin(l:h,1:10);%extracting the data for the other
            reactions
101 end
102
103 for j = 1:10
104
105     dif = opt_sol - sols(:,j);
106     dif_abs = abs(dif);
107     for k = 1:rxn_n
108
109         %finding changes in flux values larger than 0.000001 for
110         %reactions whose original value was 0
111
112         if tol(k)==0
113            if dif_abs(k) > 0.000001
114               opt_som = [i,j, k, sols(k,j), opt_sol(k),tol(k),
                     randval_fin(i,j)];
115               per_result = [per_result; opt_som];
116
117            end
118
119         else
120
121            %finding changes in flux values larger than 5% tolernace
122
123            if dif_abs(k) > tol(k)
124
125                %saving significantly different flux values in per_result
126
127                opt_som = [i,j, k, sols(k,j), opt_sol(k),tol(k),
                       randval_fin(i,j)];
128                per_result = [per_result; opt_som];
129            end
130
131         end
132
133     end
134 end
135
136 %saving sorted flux values in multiple files for the sake of memory
       usage
```

```

if f ==1
filename = sprintf('%s%d','per_result_new',i, '.dat');
save(filename,'per_result','-ascii','-double');

f_str = string(filename)
file_names = [file_names, f_str];

f=0;
clear per_result

per_result =[];

    end
end


%merging all the files into one comprehensive file
for i = 1:numel(file_names);

    file = load(file_names(i));
    per_result = [per_result;file];
end
save('per_result','per_result','-ascii','-double');
```

```

function table = stat(model_n, per_result, print)

% stat function calculates various statiscal measures on the flux values
% following perturbation in metabolic network

%USAGE: table = stat(model_name, per_result, print)


%INPUT

%model_n: metabolic model in SBML format

%per_result: significantly different flux values in a sorted format

%print: printing the statistics in an excel file (default: false)

%OUTPUT:

%table: statistical categories obtained from analysis of flux
    distribution
%profiles

%Example:
```

```
24
25 %table = stat('mut-chem.xml', 'per_result') using the perturbation
       results
26 %from MATLAB version
27
28 %table = stat('mut-chem.xml', 'final.csv') using the perturbation result
29 %from the python version
30
31
32
33 % Authors:
34
35 % Seyed Babak Loghmani
36
37 % Last updated: August 2021
38
39
40 if (nargin < 3)
41     print = false;
42 end
43
44 % loading model and sorted results from perturbation
45 model = readCbModel(model_n);
46 final = load(per_result);
47
48 % FVA
49 [minFluxF1, maxFluxF1, optsol, ret, fbasol, fvamin, fvamax,
       statussolmin, statussolmax] = fastFVA(model);
50
51 % finding the number of variable and stable reactions
52 fva_n = maxFluxF1 - minFluxF1;
53 fva_ind = find(fva_n);
54 stable_ind = find(~fva_n);
55 fva=numel(find(fva_n)); %number of variable reactions
56 stable = numel(stable_ind); %number of stable reactions
57 % finding variable reactions with the interval size larger than 10e-6,
       used
58 % for perturbation
59 r=fva_n > 0.000001;
60 fva_n_f=find(r);
61
62 rxn_n = numel(model.rxns);
63 rxn_n2 = rxn_n +1;
64
65
66 reac =[];
67 reac_no=[];
68 rand_no1=[];
```

```
rand = [];
rand_no = [];
prt_rec = final(:,1);
sz = size(final);
mx = final(sz(1));
robust_var= [];

%finding how many reactions were affected by each reaction

for i = 1:mx
    %picking the respective data for each perturbed reaction in the
       sorted
    %results
    r_a = find(ismember(prt_rec, i));
    r_a_min = min(r_a);
    r_a_max = max(r_a);
    r_a_n = numel(r_a);
    r_af = final(r_a_min:r_a_max,2);

    for j = 1:10
        r_af_i = find(ismember(r_af, j));
        d1 = numel(r_af_i);
        rand = [i,j,d1]; %finding the number of times each perturbation
           in each reaction affected other reactions
        rand_no = [rand_no;rand]; % saving the results for all 10
           perturbations
    end

    r_a_al = final(r_a_min:r_a_max,3);%finding affected reactions by
       each reaction
    r_a_al_u = unique(r_a_al);
    r_a_num = numel(r_a_al_u); %finding the index of affected reaction
    r_a_n2 = r_a_n/10;
    reac = [i,r_a_n,r_a_n2,r_a_num]; % the index of the perturb
       reaction, the number of resultant flux changes, the number of
       resultant flux changes per 10 perturbations, the number of
       affected reactions
    reac_no = [reac_no;reac]; %size equal to fva result

end
rec = reac_no(:,3);
re_r = reac_no(:,4);
mean_re = mean(rec); %affecting avg(perturbation wise)
mean_re2 = mean(re_r); %affecting avg(reaction wise)
std_dev_meanre= std( rec ); %standard deviation
std_dev_meanre2= std( re_r ); %standard deviation
min_re = min(rec); %affecting min
max_re = max(rec); %affecting max
```

```
110
111 %finding sensitivity of each reaction
112 %column= affected, row = affecting
113
114 reac_sen =[];
115 ou = [];
116 robust=[];
117 pul = 0;
118
119 %creating a matrix containing the information of affected and affecting
    reactions
120 mat = cell(rxn_n2:rxn_n2)
121 mat(2:rxn_n2) = model.rxns;
122 mat(1,2:rxn_n2) = model.rxns;
123 for i = 1:rxn_n
124     rc = find(ismember(final(:,3), i));
125     t = numel(rc); %the number of significant flux changes in the
        respective sensitive reaction
126     ou = [];
127
128     if isempty(rc) == 1
129         robust=[robust; i];%robust reaactions
130     else
131
132     for j =1:numel(rc)
133         ind = rc(j);
134         ou =[ou; final(ind, 1)];%saving the index of reactions that
            affected each sensitive reaction
135     end
136
137     ou_u = unique(ou);
138     for s = 1:numel(ou_u)
139         s_id = find(ismember(ou, ou_u(s)));
140         s_id_u = numel(s_id);
141         nb = ou_u(s);
142         i2 = fva_n_f(nb)+1; %the index of affecting reaction in the
            'mat' matrix
143         i3 = i+1; %the index of affecrted reaction in the 'mat' matrix
144
145         mat{i2, i3} = s_id_u; %implementing the information regarding
            the number of times the reaction i2 affected the reaction i3
146     end
147     gu = numel(ou);
148     n_ou = numel(ou_u);%the number of reactions that affected the
        respective sensitive reaction
149     reac_sen = [reac_sen; i, model.rxns(i), n_ou, t]; %sensitive to
        perturbation
150
```

```
151     end
152 end
153 
154     for i = 1:rxn_n2
155         mat{i,i}=0;%clearing the diagonal information to avoid counting
                the self affecting reactions
156     end
157 
158 sz_sen = size(reac_sen);
159 szs = sz_sen(1);
160 
161 
162 sensitivity=[];
163 for i = 1:szs
164     u=reac_sen{i,1};
165     sensitivity=[sensitivity;u];%sensitive reactions
166 end
167 
168 num_sen = numel(sensitivity);%number of sensitive reactions
169 num_robust = numel(robust);%number of robust reactions
170 
171 
172 
173     %average and maximum and minimum number of sensitivities
174 
175     sen1 = cell2mat(reac_sen(:,3));
176     sen2 = cell2mat(reac_sen(:,4));
177     msen = mean(sen1); %average sensitivity
178     msen2 = mean(sen2); %average sensitivity(perturbation wise)
179     std_dev_sen= std( sen1 ); %standard deviation
180     std_dev_sen2= std( sen2 ); %standard deviation 2
181     maxsen = max(sen1); %max sensitivity
182     minsen = min(sen1); %min sensitivity
183 
184 
185     %robusts with and without flux
186 
187     rec_f = find(fbasol);
188     r_w_f = find(ismember(rec_f,robust));
189     ind_rwf = rec_f(r_w_f);
190     robust_wf = model.rxns(ind_rwf); %robust with flux
191 
192     %finding robust reactions that are variable
193 
194     robust_var = [];
195     for i = 1:numel(robust)
196         ek = find(ismember(fva_ind, robust(i)));
197         if numel(ek) == 1
```

```
198             robust_var = [robust_var; robust(i)];
199         end
200     end
201
202     num_robust_var = numel(robust_var);%number of robust-variable
        reactions
203     num_robust_wf = numel(robust_wf);%number of robust reactions with
        flux in original FBA/FVA
204     num_robust_wof = numel(model.rxns) - num_robust_wf -
        num_sen;%number of robust reactions without flux in original
        FBA/FVA
205
206
207     %variability caused by exchange reactions
208     num_imp = [];
209     for i = 2:rxn_n2
210         c_count = 0;%counter
211         ex_ind = findstr(mat{i,1}, 'Ex');
212         if numel(ex_ind) ==1
213             mat_row = mat(i,:);
214             for j = 2:rxn_n2
215                 if j~=i
216                     mat_cell = mat_row{j};
217                     if numel(mat_cell) == 1
218                         c_count = c_count + 1;
219                     end
220                 end
221             end
222         end
223         num_imp = [num_imp; i-1,c_count];%saving the number of
            significant flux changes caused by perturbation in exchange
            reactions
224     end
225
226     %maximum and average number that reactions got affected by exchange
227     %reactions
228 nimp = num_imp(:,2);
229 f_nimp = find(nimp);
230 nimp_e=nimp(f_nimp);
231 avg_ex = mean(nimp_e); %avg affected by Ex
232 std_dev_ex= std( nimp_e ); %standard deviation
233 max_ex = max(nimp_e); %max affected by Ex
234
235 out={'model␣
    name','variable','stable','sensitive','robust','affecting-avg(reaction-wise)','st
    with␣flux','robust␣without␣
    flux','robust-w-variability','avg-affected␣by␣
    ex','std-ex','max-affected␣by␣ex','max-affected␣by␣
```

```
    all','min-affected␣by␣
    all';model.description,fva,stable,num_sen,num_robust,mean_re2,std_dev_meanre2,mea
table=cell2table(out);
if (print)
    writetable(table,'statistics.xls','WriteVariableNames',0)
end
```

The following scripts were developed based on python to find potential drug targets in tract-specific genome-scale mmetabolic models.

```
#Inserting the model

modelDir = os.getcwd()
cmod=cbm.CBRead.readSBML3FBC('saliva-fc-ful-3.xml', modelDir)

#FVA

f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)


#finding drug targets that reduce the growth rate by 30%


cmod.setReactionBounds('R_BIOMASS', 0,1)
f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)

react_list=cmod.getReactionIds()
drug_30_react = []
for h in react_list:
    r=cmod.getReactionBounds(h)

    cmod.setReactionBounds(h, 0,0)
    t= cbm.analyzeModel(cmod)
    if t <= 0.0789:

        g_o = cmod.getGPRforReaction(h)
        if bool(g_o)==True:
            g = g_o.getGeneIds()
            drug_30_react.append([h,g,t])


    elif np.isnan(t)==True:
        g_o = cmod.getGPRforReaction(h)
        if bool(g_o)==True:
            g = g_o.getGeneIds()
            drug_30_react.append([h,g,t])

     #elif t==0:
```

```
      #    drug_30_react.append(h)
     #elif t==0.0:
      #    drug_30_react.append(h)

    cmod.setReactionBounds(h, r[1],r[2])


textfile = open("drug_30_percent_growth.txt", "w")
for element in drug_30_react:
    textfile.write(str(element[0]) + '\t' + str(element[1])+ '\t' +
       str(element[2])+ "\n")
textfile.close()



#finding drug targets that reduce the growth rate by 20%

cmod=cbm.CBRead.readSBML3FBC('saliva-fc-ful-3.xml', modelDir)
f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)
cmod.setReactionBounds('R_BIOMASS', 0,1)


react_list=cmod.getReactionIds()
drug_20_react = []
for h in react_list:
    r=cmod.getReactionBounds(h)

    cmod.setReactionBounds(h, 0,0)
    t= cbm.analyzeModel(cmod)
    if t <= 0.0526:
        g_o = cmod.getGPRforReaction(h)
        if bool(g_o)==True:
            g = g_o.getGeneIds()
            drug_20_react.append([h,g,t])

    elif np.isnan(t)==True:

        g_o = cmod.getGPRforReaction(h)
        if bool(g_o)==True:
            g = g_o.getGeneIds()
            drug_20_react.append([h,g,t])

     #elif t==0:
      #    drug_30_react.append(h)
     #elif t==0.0:
      #    drug_30_react.append(h)

    cmod.setReactionBounds(h, r[1],r[2])
```

```
87
88
89  textfile = open("drug_20_percent_growth.txt", "w")
90  for element in drug_20_react:
91      textfile.write(str(element[0]) + '\t' + str(element[1])+ '\t' +
            str(element[2])+ "\n")
92  textfile.close()
93
94
95  #finding drug targets that reduce the growth rate by 10%
96
97  cmod=cbm.CBRead.readSBML3FBC('saliva-fc-ful-3.xml', modelDir)
98  f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)
99  cmod.setReactionBounds('R_BIOMASS', 0,1)
100
101
102 react_list=cmod.getReactionIds()
103 drug_10_react = []
104 for h in react_list:
105     r=cmod.getReactionBounds(h)
106
107     cmod.setReactionBounds(h, 0,0)
108     t= cbm.analyzeModel(cmod)
109     if t <= 0.0263:
110
111         g_o = cmod.getGPRforReaction(h)
112         g = g_o.getGeneIds()
113         if bool(g_o)==True:
114             g = g_o.getGeneIds()
115             drug_10_react.append([h,g,t])
116
117     elif np.isnan(t)==True:
118
119         g_o = cmod.getGPRforReaction(h)
120         if bool(g_o)==True:
121             g = g_o.getGeneIds()
122             drug_10_react.append([h,g,t])
123
124      #elif t==0:
125       #    drug_30_react.append(h)
126      #elif t==0.0:
127       #    drug_30_react.append(h)
128
129     cmod.setReactionBounds(h, r[1],r[2])
130
131
132 textfile = open("drug_10_percent_growth.txt", "w")
133 for element in drug_10_react:
```

```
134     textfile.write(str(element[0]) + '\t' + str(element[1])+ '\t' +
            str(element[2])+ "\n")
135 textfile.close()
136
137
138
139 #finding drug targets that 30% reduction in theri flux kills the
        bacteria
140
141 cmod=cbm.CBRead.readSBML3FBC('saliva-fc-ful-3.xml', modelDir)
142 f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)
143 cmod.setReactionBounds('R_BIOMASS', 0.,1)
144
145 drug_30_flux = []
146 for h in react_list:
147     ind = n.index(h)
148     r=cmod.getReactionBounds(h)
149
150     n_l_b = 0.3*f[ind][2]
151     n_u_b = 0.3*f[ind][3]
152
153     cmod.setReactionBounds(h, n_l_b,n_u_b)
154     t= cbm.analyzeModel(cmod)
155     if np.isnan(t)==True:
156         g_o = cmod.getGPRforReaction(h)
157         if bool(g_o)==True:
158             g = g_o.getGeneIds()
159             drug_30_flux.append([h,g,t])
160
161     elif t==0.0:
162
163         g_o = cmod.getGPRforReaction(h)
164         if bool(g_o)==True:
165             g = g_o.getGeneIds()
166             drug_30_flux.append([h,g,t])
167
168     cmod.setReactionBounds(h, r[1],r[2])
169
170
171 textfile = open("drug_30_percent_flux.txt", "w")
172 for element in drug_30_flux:
173     textfile.write(str(element[0]) + '\t' + str(element[1])+ '\t' +
            str(element[2])+ "\n")
174 textfile.close()
175
176
177
178
```

```
179 #finding drug targets that 20% reduction in theri flux kills the
        bacteria
180
181 cmod=cbm.CBRead.readSBML3FBC('saliva-fc-ful-3.xml', modelDir)
182 f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)
183 cmod.setReactionBounds('R_BIOMASS', 0.,1)
184
185 drug_20_flux = []
186 for h in react_list:
187     ind = n.index(h)
188     r=cmod.getReactionBounds(h)
189
190     n_l_b = 0.2*f[ind][2]
191     n_u_b = 0.2*f[ind][3]
192
193     cmod.setReactionBounds(h, n_l_b,n_u_b)
194     t= cbm.analyzeModel(cmod)
195     if np.isnan(t)==True:
196         g_o = cmod.getGPRforReaction(h)
197         if bool(g_o)==True:
198             g = g_o.getGeneIds()
199             drug_20_flux.append([h,g,t])
200     #if t<= 0.1:
201      #    drug.append(h)
202     elif t==0.0:
203
204         g_o = cmod.getGPRforReaction(h)
205         if bool(g_o)==True:
206             g = g_o.getGeneIds()
207             drug_20_flux.append([h,g,t])
208
209     cmod.setReactionBounds(h, r[1],r[2])
210
211
212 textfile = open("drug_20_percent_flux.txt", "w")
213 for element in drug_20_flux:
214     textfile.write(str(element[0]) + '\t' + str(element[1])+ '\t' +
        str(element[2])+ "\n")
215 textfile.close()
216
217
218 #finding drug targets that 10% reduction in theri flux kills the
        bacteria
219
220 cmod=cbm.CBRead.readSBML3FBC('saliva-fc-ful-3.xml', modelDir)
221 f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)
222 cmod.setReactionBounds('R_BIOMASS', 0.,1)
223
```

```
drug_10_flux = []
for h in react_list:
    ind = n.index(h)
    r=cmod.getReactionBounds(h)

    n_l_b = 0.1*f[ind][2]
    n_u_b = 0.1*f[ind][3]

    cmod.setReactionBounds(h, n_l_b,n_u_b)
    t= cbm.analyzeModel(cmod)
    if np.isnan(t)==True:
        g_o = cmod.getGPRforReaction(h)
        if bool(g_o)==True:
            g = g_o.getGeneIds()
            drug_10_flux.append([h,g,t])
    #if t<= 0.1:
     #   drug.append(h)
    elif t==0.0:

        g_o = cmod.getGPRforReaction(h)
        if bool(g_o)==True:
            g = g_o.getGeneIds()
            drug_10_flux.append([h,g,t])

    cmod.setReactionBounds(h, r[1],r[2])


textfile = open("drug_10_percent_flux.txt", "w")
for element in drug_10_flux:
    textfile.write(str(element[0]) + '\t' + str(element[1])+ '\t' +
       str(element[2])+ "\n")
textfile.close()


#finding drug targets that 30% reduction in their flux resulting in 30%
   reduction in growth rate or kill

cmod=cbm.CBRead.readSBML3FBC('saliva-fc-ful-3.xml', modelDir)
f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)
cmod.setReactionBounds('R_BIOMASS', 0.,1)

drug_30_combo = []
for h in react_list:
    ind = n.index(h)
    r=cmod.getReactionBounds(h)

    n_l_b = 0.3*f[ind][2]
    n_u_b = 0.3*f[ind][3]
```

```
270
271     cmod.setReactionBounds(h, n_l_b,n_u_b)
272     t= cbm.analyzeModel(cmod)
273
274     if t <= 0.0789:
275
276         g_o = cmod.getGPRforReaction(h)
277         if bool(g_o)==True:
278             g = g_o.getGeneIds()
279             drug_30_combo.append([h,g,t])
280
281
282     elif np.isnan(t)==True:
283
284         g_o = cmod.getGPRforReaction(h)
285         if bool(g_o)==True:
286             g = g_o.getGeneIds()
287             drug_30_combo.append([h,g,t])
288     #if t<= 0.1:
289      #    drug.append(h)
290     elif t==0.0:
291
292         g_o = cmod.getGPRforReaction(h)
293         if bool(g_o)==True:
294             g = g_o.getGeneIds()
295             drug_30_combo.append([h,g,t])
296
297     cmod.setReactionBounds(h, r[1],r[2])
298
299
300 textfile = open("drug_30_percent_combo.txt", "w")
301 for element in drug_30_combo:
302     textfile.write(str(element[0]) + '\t' + str(element[1])+ '\t' +
        str(element[2])+ "\n")
303 textfile.close()
304
305
306 #finding drug targets that 20% reduction in their flux resulting in 20%
    reduction in growth rate or kill
307
308 cmod=cbm.CBRead.readSBML3FBC('saliva-fc-ful-3.xml', modelDir)
309 f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)
310 cmod.setReactionBounds('R_BIOMASS', 0.,1)
311
312 drug_20_combo = []
313 for h in react_list:
314     ind = n.index(h)
315     r=cmod.getReactionBounds(h)
```

```
316
317     n_l_b = 0.2*f[ind][2]
318     n_u_b = 0.2*f[ind][3]
319
320     cmod.setReactionBounds(h, n_l_b,n_u_b)
321     t= cbm.analyzeModel(cmod)
322
323     if t <= 0.0789:
324
325         g_o = cmod.getGPRforReaction(h)
326         if bool(g_o)==True:
327             g = g_o.getGeneIds()
328             drug_20_combo.append([h,g,t])
329
330
331     elif np.isnan(t)==True:
332
333         g_o = cmod.getGPRforReaction(h)
334         if bool(g_o)==True:
335             g = g_o.getGeneIds()
336             drug_20_combo.append([h,g,t])
337     #if t<= 0.1:
338      #    drug.append(h)
339     elif t==0.0:
340
341         g_o = cmod.getGPRforReaction(h)
342         if bool(g_o)==True:
343             g = g_o.getGeneIds()
344             drug_20_combo.append([h,g,t])
345
346     cmod.setReactionBounds(h, r[1],r[2])
347
348
349 textfile = open("drug_20_percent_combo.txt", "w")
350 for element in drug_20_combo:
351     textfile.write(str(element[0]) + '\t' + str(element[1])+ '\t' +
        str(element[2])+ "\n")
352 textfile.close()
353
354
355 #finding drug targets that 10% reduction in their flux resulting in 10%
    reduction in growth rate or kill
356
357 cmod=cbm.CBRead.readSBML3FBC('saliva-fc-ful-3.xml', modelDir)
358 f, n = cbm.FluxVariabilityAnalysis(cmod, optPercentage=99.9)
359 cmod.setReactionBounds('R_BIOMASS', 0.,1)
360
361 drug_10_combo = []
```

```
for h in react_list:
    ind = n.index(h)
    r=cmod.getReactionBounds(h)

    n_l_b = 0.1*f[ind][2]
    n_u_b = 0.1*f[ind][3]

    cmod.setReactionBounds(h, n_l_b,n_u_b)
    t= cbm.analyzeModel(cmod)

    if t <= 0.0789:

        g_o = cmod.getGPRforReaction(h)
        if bool(g_o)==True:
            g = g_o.getGeneIds()
            drug_10_combo.append([h,g,t])


    elif np.isnan(t)==True:

        g_o = cmod.getGPRforReaction(h)
        if bool(g_o)==True:
            g = g_o.getGeneIds()
            drug_10_combo.append([h,g,t])
    #if t<= 0.1:
     #   drug.append(h)
    elif t==0.0:

        g_o = cmod.getGPRforReaction(h)
        if bool(g_o)==True:
            g = g_o.getGeneIds()
            drug_10_combo.append([h,g,t])

    cmod.setReactionBounds(h, r[1],r[2])


textfile = open("drug_10_percent_combo.txt", "w")
for element in drug_10_combo:
    textfile.write(str(element[0]) + '\t' + str(element[1])+ '\t' +
       str(element[2])+ "\n")
textfile.close()
```

Bibliography

Alyaa M Abdel-Haleem, Hooman Hefzi, Katsuhiko Mineta, Xin Gao, Takashi Gojobori, Bernhard O Palsson, Nathan E Lewis, and Neema Jamshidi. Functional interrogation of plasmodium genus metabolism identifies species-and stage-specific differences in nutrient essentiality and drug targeting. *PLoS computational biology*, 14(1):e1005895, 2018.

R FULLER AFRC. Probiotics in man and animals. *Journal of applied bacteriology*, 66(5): 365–378, 1989.

M Aguirre and MD Collins. Lactic acid bacteria and human clinical infection. *Journal of Applied Bacteriology*, 75(2):95–107, 1993.

Bruce Alberts, Alexander Johnson, Julian Lewis, Martin Raff, Keith Roberts, Peter Walter, et al. Molecular biology of the cell. *SCANDINAVIAN JOURNAL OF RHEUMATOLOGY*, 32(2): 125–125, 2003.

Stefano Andreozzi, Anirikh Chakrabarti, Keng Cher Soh, Anthony Burgard, Tae Hoon Yang, Stephen Van Dien, Ljubisa Miskovic, and Vassily Hatzimanikatis. Identification of metabolic engineering targets for the enhancement of 1, 4-butanediol production in recombinant e. coli using large-scale kinetic models. *Metabolic engineering*, 35:148–159, 2016.

Fatima Ardito, Michele Giuliani, Donatella Perrone, Giuseppe Troiano, and Lorenzo Lo Muzio. The crucial role of protein phosphorylation in cell signaling and its use as targeted therapy. *International journal of molecular medicine*, 40(2):271–280, 2017.

Magnus Øverlie Arntzen, Ingrid Lea Karlskås, Morten Skaugen, Vincent GH Eijsink, and Geir Mathiesen. Proteomic investigation of the response of enterococcus faecalis v583 when cultivated in urine. *PloS one*, 10(4):e0126694, 2015.

Olfa Baccouri, Amine Mohamed Boukerb, Leila Ben Farhat, Arthur Zébré, Kurt Zimmermann, Eugen Domann, Mélyssa Cambronel, Magalie Barreau, Olivier Maillot, Isabelle Rincé, et al. Probiotic potential and safety evaluation of enterococcus faecalis ob14 and ob15, isolated from traditional tunisian testouri cheese and rigouta, using physiological and genomic analysis. *Frontiers in microbiology*, 10:881, 2019.

Jung Eun Baik, Young Hee Ryu, Ji Young Han, Jintaek Im, Kee-Yeon Kum, Cheol-Heui Yun, Kangseok Lee, and Seung Hyun Han. Lipoteichoic acid partially contributes to the inflammatory responses to enterococcus faecalis. *Journal of endodontics*, 34(8):975–982, 2008.

Dany JV Beste, Tracy Hooper, Graham Stewart, Bhushan Bonde, Claudio Avignone-Rossa, Michael E Bushell, Paul Wheeler, Steffen Klamt, Andrzej M Kierzek, and Johnjoe McFadden. Gsmn-tb: a web-based genome-scale network model of mycobacterium tuberculosis metabolism. *Genome biology*, 8(5):1–18, 2007.

Liv Anette Bøhle, Ellen M Færgestad, Eva Veiseth-Kent, Hilde Steinmoen, Ingolf F Nes, Vincent GH Eijsink, and Geir Mathiesen. Identification of proteins related to the stress response in enterococcus faecalis v583 caused by bovine bile. *Proteome science*, 8(1):1–12, 2010.

Sergio Bordel, Rasmus Agren, and Jens Nielsen. Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS computational biology*, 6(7):e1000859, 2010.

Irina Borodina, Kanchana R Kildegaard, Niels B Jensen, Thomas H Blicher, Jérôme Maury, Svetlana Sherstyk, Konstantin Schneider, Pedro Lamosa, Markus J Herrgård, Inger Rosenstand, et al. Establishing a synthetic pathway for high-level production of 3-hydroxypropionic acid in saccharomyces cerevisiae via β-alanine. *Metabolic engineering*, 27:57–64, 2015.

David Brett, Heike Pospisil, Juan Valcárcel, Jens Reich, and Peer Bork. Alternative splicing and genome complexity. *Nature genetics*, 30(1):29–30, 2002.

Makhtar Camara, Assane Dieng, and Cheikh Saad Bouh Boye. Antibiotic susceptibility of streptococcus pyogenes isolated from respiratory tract infections in dakar, senegal. *Microbiology insights*, 6:MBI–S12996, 2013.

Yesim Cetinkaya, Pamela Falk, and C Glen Mayhall. Vancomycin-resistant enterococci. *Clinical microbiology reviews*, 13(4):686–707, 2000.

Sriram Chandrasekaran and Nathan D Price. Probabilistic integrative modeling of genome-scale metabolic and regulatory networks in escherichia coli and mycobacterium tuberculosis. *Proceedings of the National Academy of Sciences*, 107(41):17845–17850, 2010.

Christophe Chassagnole, Naruemol Noisommit-Rizzi, Joachim W Schmid, Klaus Mauch, and Matthias Reuss. Dynamic modeling of the central carbon metabolism of escherichia coli. *Biotechnology and bioengineering*, 79(1):53–73, 2002.

CECIL C Chen and P PATRICK Cleary. Cloning and expression of the streptococcal c5a peptidase gene in escherichia coli: linkage to the type 12 m protein gene. *Infection and immunity*, 57(6):1740–1745, 1989.

Bhawna Chugh and Afaf Kamal-Eldin. Bioactive compounds produced by probiotics in food products. *Current Opinion in Food Science*, 32:76–82, 2020.

Sonia Cortassa and Miguel A Aon. Metabolic control analysis of glycolysis and branching to ethanol production in chemostat cultures of saccharomyces cerevisiae under carbon, nitrogen, or phosphate limitations. *Enzyme and microbial technology*, 16(9):761–770, 1994.

Francis Crick. Central dogma of molecular biology. *Nature*, 227(5258):561–563, 1970.

Scott B Crown and Maciek R Antoniewicz. Parallel labeling experiments and metabolic flux analysis: Past, present and future methodologies. *Metabolic engineering*, 16:21–32, 2013.

AC Dahlberg and FV Kosikowsky. The development of flavor in american cheddar cheese made from pasteurized milk with streptococcus faecalis starter. *Journal of Dairy Science*, 31(4): 275–284, 1948.

Luis M De la Maza, Marie T Pezzlo, Janet T Shigei, Grace L Tan, Ellena M Peterson, et al. *Color atlas of medical bacteriology.* Number Ed. 2. ASM press, 2013.

Hoang V Dinh, Debolina Sarkar, and Costas D Maranas. Quantifying the propagation of parametric uncertainty on flux balance analysis. *Metabolic Engineering*, 2021.

Gary M Dunny, BA Leonard, and Peter J Hedberg. Pheromone-inducible conjugation in enterococcus faecalis: interbacterial and host-parasite chemical communication. *Journal of bacteriology*, 177(4):871–876, 1995.

Martin Dworkin. *The prokaryotes: vol. 4: bacteria: firmicutes, cyanobacteria*. Springer Science & Business Media, 2006.

Robert J Edwards, Marta Pyzio, Magdalena Gierula, Claire E Turner, Vahitha B Abdul-Salam, and Shiranee Sriskandan. Proteomic analysis at the sites of clinical infection with invasive streptococcus pyogenes. *Scientific reports*, 8(1):1–9, 2018.

Marcos A Espinal. The global situation of mdr-tb. *Tuberculosis*, 83(1-3):44–51, 2003.

Shirin Fallahi, Hans J Skaug, and Guttorm Alendal. A comparison of monte carlo sampling methods for metabolic network models. *Plos one*, 15(7):e0235393, 2020.

Iman Famili and Bernhard O Palsson. The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools. *Biophysical journal*, 85(1):16–26, 2003.

Joseph J Ferretti, Dennis L Stevens, and Vincent A Fischetti. Streptococcus pyogenes: basic biology to clinical manifestations [internet]. 2016.

Elizabeth Fiore, Daria Van Tyne, and Michael S Gilmore. Pathogenicity of enterococci. *Microbiology spectrum*, 7(4):7–4, 2019.

Nicolas AL Flahaut, Anne Wiersma, Bert van de Bunt, Dirk E Martens, Peter J Schaap, Lolke Sijtsma, Vitor A Martins Dos Santos, and Willem M de Vos. Genome-scale metabolic model for lactococcus lactis mg1363 and its application to the analysis of flavor formation. *Applied microbiology and biotechnology*, 97(19):8729–8739, 2013.

Karl Forchhammer. Glutamine signalling in bacteria. *Front Biosci*, 12(2069):10–2741, 2007.

Charles MAP Franz, Wilhelm H Holzapfel, and Michael E Stiles. Enterococci at the crossroads of food safety? *International journal of food microbiology*, 47(1-2):1–24, 1999.

Peng Gao, Kenneth L Pinkston, Sreedhar R Nallapareddy, Ambro van Hoof, Barbara E Murray, and Barrett R Harvey. Enterococcus faecalis rnjb is required for pilin gene expression and biofilm formation. *Journal of bacteriology*, 192(20):5489–5498, 2010.

Peng Gao, Kenneth L Pinkston, Agathe Bourgogne, Barbara E Murray, Ambro van Hoof, and Barrett R Harvey. Functional studies of e. faecalis rnase j2 and its role in virulence and fitness. *PloS one*, 12(4):e0175212, 2017.

Sergio Garcia, R Adam Thompson, Richard J Giannone, Satyakam Dash, Costas D Maranas, and Cong T Trinh. Development of a genome-scale metabolic model of clostridium thermocellum and its applications for integration of multi-omics datasets and computational strain design. *Frontiers in bioengineering and biotechnology*, 8:772, 2020.

Héctor García Martín, Vinay Satish Kumar, Daniel Weaver, Amit Ghosh, Victor Chubukov, Aindrila Mukhopadhyay, Adam Arkin, and Jay D Keasling. A method to constrain genome-scale models with 13c labeling data. *PLoS computational biology*, 11(9):e1004363, 2015.

Helen S Garmory and Richard W Titball. Atp-binding cassette transporters are targets for the development of antibacterial vaccines and therapies. *Infection and immunity*, 72(12):6757–6763, 2004.

Christel Garrigues, Pascal Loubiere, Nic D Lindley, and Muriel Cocaign-Bousquet. Control of the shift from homolactic acid to mixed-acid fermentation in lactococcus lactis: predominant role of the nadh/nad+ ratio. *Journal of Bacteriology*, 179(17):5282–5287, 1997.

Kanika Gera and Kevin S McIver. Laboratory growth and maintenance of streptococcus pyogenes (the group a streptococcus, gas). *Current protocols in microbiology*, 30(1):9D–2, 2013.

Michael S Gilmore, Don B Clewell, Patrice Courvalin, Gary M Dunny, Barbara E Murray, Louis B Rice, et al. *The enterococci: pathogenesis, molecular biology, and antibiotic resistance*, volume 10. ASM press Washington, DC, 2002.

Sara Gómez, Javier Querol-García, Gara Sánchez-Barrón, Marta Subias, Àlex González-Alsina, Virginia Franco-Hidalgo, Sebastián Albertí, Santiago Rodriguez de Cordoba, Francisco J Fernández, and M Cristina Vega. The antimicrobials anacardic acid and curcumin are non-competitive inhibitors of gram-positive bacterial pathogenic glyceraldehyde-3-phosphate dehydrogenase by a mechanism unrelated to human c5a anaphylatoxin binding. *Frontiers in microbiology*, 10:326, 2019.

Ruth Großeholz, Ching-Chiek Koh, Nadine Veith, Tomas Fiedler, Madlen Strauss, Brett Olivier, Ben C Collins, Olga T Schubert, Frank Bergmann, Bernd Kreikemeyer, et al. Integrating highly quantitative proteomics and genome-scale metabolic modeling to study ph adaptation

in the human pathogen enterococcus faecalis. *NPJ systems biology and applications*, 2(1): 1–9, 2016.

Changdai Gu, Gi Bae Kim, Won Jun Kim, Hyun Uk Kim, and Sang Yup Lee. Current status and applications of genome-scale metabolic models. *Genome biology*, 20(1):1–18, 2019.

Jurgen R Haanstra, Albert Gerding, Amalia M Dolga, Freek JH Sorgdrager, Manon Buist-Homan, Francois Du Toit, Klaas Nico Faber, Hermann-Georg Holzhütter, Balázs Szöör, Keith R Matthews, et al. Targeting pathogen metabolism without collateral damage to the host. *Scientific reports*, 7(1):1–15, 2017.

Sean R Hackett, Vito RT Zanotelli, Wenxin Xu, Jonathan Goya, Junyoung O Park, David H Perlman, Patrick A Gibney, David Botstein, John D Storey, and Joshua D Rabinowitz. Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science*, 354(6311), 2016.

Laurent Heirendt, Sylvain Arreckx, Thomas Pfau, Sebastián N Mendoza, Anne Richelle, Almut Heinken, Hulda S Haraldsdóttir, Jacek Wachowiak, Sarah M Keating, Vanja Vlasov, Stefania Magnusdóttir, Chiam Yu Ng, German Preciat, Alise Žagare, Siu H J Chan, Maike K Aurich, Catherine M Clancy, Jennifer Modamio, John T Sauls, Alberto Noronha, Aarash Bordbar, Benjamin Cousins, Diana C El Assal, Luis V Valcarcel, Iñigo Apaolaza, Susan Ghaderi, Masoud Ahookhosh, Marouen Ben Guebila, Andrejs Kostromins, Nicolas Sompairac, Hoai M Le, Ding Ma, Yuekai Sun, Lin Wang, James T Yurkovich, Miguel A P Oliveira, Phan T Vuong, Lemmer P El Assal, Inna Kuperstein, Andrei Zinovyev, H Scott Hinton, William A Bryant, Francisco J Aragón Artacho, Francisco J Planes, Egils Stalidzans, Alejandro Maass, Santosh Vempala, Michael Hucka, Michael A Saunders, Costas D Maranas, Nathan E Lewis, Thomas Sauter, Bernhard Ø Palsson, Ines Thiele, and Ronan M T Fleming. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*, 14(3):639–702, 2019. ISSN 1750-2799. doi: 10.1038/s41596-018-0098-2. URL `https://doi.org/10.1038/s41596-018-0098-2`.

Benno Hess and Arnold Boiteux. Mechanism of glycolytic oscillation in yeast, i. aerobic and anaerobic growth conditions for obtaining glycolytic oscillation. 1968.

Joseph T Holden and James M Bunch. Asparagine transport in lactobacillus plantarum and streptococcus faecalis. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 307(3):640–655, 1973.

WH Holzapfel and BJB Wood. Lactic acid bacteria in contemporary perspective. In *The genera of lactic acid bacteria*, pages 1–6. Springer, 1995.

Sijia Huang, Kumardeep Chaudhary, and Lana X Garmire. More is better: recent progress in multi-omics data integration methods. *Frontiers in genetics*, 8:84, 2017.

Wolfgang Huber, Anja von Heydebreck, Holger Sueltmann, Annemarie Poustka, and Martin Vingron. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics*, 18 Suppl. 1:S96–S104, 2002.

Hyelyeon Hwang and Jong-Hee Lee. Characterization of arginine catabolism by lactic acid bacteria isolated from kimchi. *Molecules*, 23(11):3049, 2018.

IBM ILOG. CPLEX user's manual. page 596, 2017. URL `https://www.ibm.com/docs/en/icos/12.8.0.0?topic=cplex-users-manual`.

R Jayaraman. Antibiotic resistance: an overview of mechanisms and a paradigm shift. *Current science*, pages 1475–1484, 2009.

Livnat Jerby and Eytan Ruppin. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clinical Cancer Research*, 18(20):5572–5584, 2012.

Hong Jin, Shivangi Agarwal, Shivani Agarwal, and Vijay Pancholi. Surface export of gapdh/sdh, a glycolytic enzyme, is essential for streptococcus pyogenes virulence. *MBio*, 2(3):e00068–11, 2011.

Hossein Samadi Kafil and Mohammad Asgharzadeh. Vancomycin-resistant enteroccus faecium and enterococcus faecalis isolated from education hospital of iran. *Maedica*, 9(4):323, 2014.

J Kampfer, TN Göhring, T Attin, and M Zehnder. Leakage of food-borne enterococcus faecalis through temporary fillings in a simulated oral environment. *International endodontic journal*, 40(6):471–477, 2007.

Andrew L Kau, Steven M Martin, William Lyon, Ericka Hayes, Michael G Caparon, and Scott J Hultgren. Enterococcus faecalis tropism for the kidneys in the urinary tract of c57bl/6j mice. *Infection and immunity*, 73(4):2461–2468, 2005.

Steven M Kelk, Brett G Olivier, Leen Stougie, and Frank J Bruggeman. Optimal flux spaces of genome-scale stoichiometric models are determined by a few subnetworks. *Scientific reports*, 2(1):1–7, 2012.

Hyun Uk Kim, Tae Yong Kim, and Sang Yup Lee. Genome-scale metabolic network analysis and drug targeting of multi-drug resistant pathogen acinetobacter baumannii aye. *Molecular BioSystems*, 6(2):339–348, 2010.

Hyun Uk Kim, Soo Young Kim, Haeyoung Jeong, Tae Yong Kim, Jae Jong Kim, Hyon E Choy, Kyu Yang Yi, Joon Haeng Rhee, and Sang Yup Lee. Integrative genome-scale metabolic analysis of vibrio vulnificus for drug targeting and discovery. *Molecular systems biology*, 7 (1):460, 2011.

Min Kyung Kim and Desmond S Lun. Methods for integration of transcriptomic data in genome-scale metabolic models. *Computational and structural biotechnology journal*, 11 (18):59–65, 2014.

Edda Klipp, Wolfram Liebermeister, Christoph Wierling, and Axel Kowald. *Systems biology: a textbook*. John Wiley & Sons, 2016.

Tomas G Kloosterman, Wouter T Hendriksen, Jetta JE Bijlsma, Hester J Bootsma, Sacha AFT van Hijum, Jan Kok, Peter WM Hermans, and Oscar P Kuipers. Regulation of glutamine and glutamate metabolism by glnr and glna in streptococcus pneumoniae. *Journal of Biological Chemistry*, 281(35):25097–25109, 2006.

VA Knivett. The anaerobic interconversion of ornithine and citrulline by streptococcus faecalis. *Biochemical Journal*, 58(3):480, 1954.

Stewart Arment Koser et al. Vitamin requirements of bacteria and yeasts. *Vitamin requirements of bacteria and yeasts.*, 1968.

Christopher J Kristich, Louis B Rice, and Cesar A Arias. Enterococcal infection—treatment and antibiotic resistance. *Enterococci: from commensals to leading causes of drug resistant infection [Internet]*, 2014.

Kristian Kvint, Laurence Nachin, Alfredo Diez, and Thomas Nyström. The bacterial universal stress protein: function and regulation. *Current opinion in microbiology*, 6(2):140–145, 2003.

Petri-Jaan Lahtvee, Kaarel Adamberg, Liisa Arike, Ranno Nahku, Kadri Aller, and Raivo Vilu. Multi-omics approach to study the growth efficiency and amino acid metabolism in lactococcus lactis at various specific growth rates. *Microbial Cell Factories*, 10(1):1–12, 2011.

Eric S Lander, Lauren M Linton, Bruce Birren, Chad Nusbaum, Michael C Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, William FitzHugh, et al. Initial sequencing and analysis of the human genome. 2001.

Elena G Lara, Isabelle van der Windt, Douwe Molenaar, Marjon GJ de Vos, and Chrats Melkonian. Using functional annotations to study pairwise interactions in urinary tract infection communities. *Genes*, 12(8):1221, 2021.

Ida Larsson, Mathias Uhlén, Cheng Zhang, and Adil Mardinoglu. Genome-scale metabolic modeling of glioblastoma reveals promising targets for drug development. *Frontiers in genetics*, 11:381, 2020.

Paola Lavermicocca, Cristina Reguant, and Joaquin Bautista-Gallego. lactic acid bacteria within the food industry: what is new on their technological and functional role. *Frontiers in Microbiology*, 12, 2021.

Robert W Leighty and Maciek R Antoniewicz. Complete-mfa: complementary parallel labeling experiments technique for metabolic flux analysis. *Metabolic engineering*, 20:49–55, 2013.

Frédéric Leroy and Luc De Vuyst. Lactic acid bacteria as functional starter cultures for the food fermentation industry. *Trends in Food Science & Technology*, 15(2):67–78, 2004.

Jennifer Levering, Mark WJM Musters, Martijn Bekker, Domenico Bellomo, Tomas Fiedler, Willem M de Vos, Jeroen Hugenholtz, Bernd Kreikemeyer, Ursula Kummer, and Bas Teusink. Role of phosphate in the central metabolism of two lactic acid bacteria–a comparative systems biology approach. *The FEBS journal*, 279(7):1274–1290, 2012.

Jennifer Levering, Tomas Fiedler, Antje Sieg, Koen WA van Grinsven, Silvio Hering, Nadine Veith, Brett G Olivier, Lara Klett, Jeroen Hugenholtz, Bas Teusink, et al. Genome-scale reconstruction of the streptococcus pyogenes m49 metabolic network reveals growth requirements and indicates potential drug targets. *Journal of biotechnology*, 232:25–37, 2016.

James C Liao, Shao-Yi Hou, and Yun-Peng Chao. Pathway analysis, engineering, and physiological considerations for redirecting central metabolism. *Biotechnology and bioengineering*, 52(1):129–140, 1996.

Hannes Link, Karl Kochanowski, and Uwe Sauer. Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. *Nature biotechnology*, 31(4): 357–361, 2013.

S-Q Liu, R Holland, P McJarrow, and VL Crow. Serine metabolism in lactobacillus plantarum. *International journal of food microbiology*, 89(2-3):265–273, 2003.

Francisco Llaneras and Jesús Picó. Stoichiometric modelling of cell metabolism. *Journal of bioscience and bioengineering*, 105(1):1–11, 2008.

Colton J Lloyd, Ali Ebrahim, Laurence Yang, Zachary A King, Edward Catoiu, Edward J O'Brien, Joanne K Liu, and Bernhard O Palsson. Cobrame: A computational framework for genome-scale models of metabolism and gene expression. *PLoS computational biology*, 14(7):e1006302, 2018.

Jason W Locasale and Lewis C Cantley. Metabolic flux and the regulation of mammalian cell growth. *Cell metabolism*, 14(4):443–451, 2011.

S.B Loghmani. Comparative transcriptome data analysis of the human pathogens, enterococcus faecalis and streptococcus pyogenes. 2020.

Seyed Babak Loghmani, Eric Zitzow, Gene Ching-Chiek Koh, Andreas Ulmer, Nadine Veith, Ruth Großeholz, Madlen Rossnagel, Maren Loesch, Ruedi Aebersold, Bernd Kreikemeyer, et al. All driven by energy demand? integrative comparison of metabolism of enterococcus faecalis wildtype and a glutamine synthase mutant. *bioRxiv*, 2021.

Seyed Babak Loghmani, Nadine Veith, Sven Sahle, Frank T Bergmann, Brett G Olivier, and Ursula Kummer. Inspecting the solution space of genome-scale metabolic models. *Metabolites*, 12(1):43, 2022.

Arash Maadani, Kristina A Fox, Elftherios Mylonakis, and Danielle A Garsin. Enterococcus faecalis mutations affecting virulence in the caenorhabditis elegans model host. *Infection and immunity*, 75(5):2634–2637, 2007.

Carina I Mack, Christoph H Weinert, Björn Egert, Paola G Ferrario, Achim Bub, Ingrid Hoffmann, Bernhard Watzl, Hannelore Daniel, and Sabine E Kulling. The complex human urinary sugar profile: determinants revealed in the cross-sectional karmen study. *The American journal of clinical nutrition*, 108(3):502–516, 2018.

R Mahadevan and CH Schilling. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metabolic engineering*, 5(4):264–276, 2003.

Andrew Makhorin. GLPK (GNU linear programming kit). *http://www. gnu. org/s/glpk/glpk. html*, 2008.

Costas D Maranas and Ali R Zomorrodi. *Optimization methods in metabolic networks*. John Wiley & Sons, 2016.

MATLAB. MATLAB version 9.4.0.813654 (R2018a). *The MathWorks Inc. Natick, Massachusetts*, 2018.

Douglas McCloskey, Jamey D Young, Sibei Xu, Bernhard O Palsson, and Adam M Feist. Modeling method for increased precision and scope of directly measurable fluxes at a genome-scale. *Analytical chemistry*, 88(7):3844–3852, 2016.

W Michael McShan, Joseph J Ferretti, Tadahiro Karasawa, Alexander N Suvorov, Shaoping Lin, Biafang Qin, Honggui Jia, Steve Kenton, Fares Najar, Hongmin Wu, et al. Genome sequence of a nephritogenic and highly transformable m49 strain of streptococcus pyogenes. *Journal of bacteriology*, 190(23):7773–7785, 2008.

Leonor Michaelis, Maud L Menten, et al. Die kinetik der invertinwirkung. *Biochem. z*, 49 (333-369):352, 1913.

William R Miller, Jose M Munita, and Cesar A Arias. Mechanisms of antibiotic resistance in enterococci. *Expert review of anti-infective therapy*, 12(10):1221–1236, 2014.

Pranjul Mishra, Na-Rae Lee, Meiyappan Lakshmanan, Minsuk Kim, Byung-Gee Kim, and Dong-Yup Lee. Genome-scale model-driven strain design for dicarboxylic acid production in yarrowia lipolytica. *BMC systems biology*, 12(2):9–20, 2018.

Francisco JM Mojica, Chc Díez-Villaseñor, Jesús García-Martínez, Elena Soria, et al. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, 60(2):174–182, 2005.

MR Foulquié Moreno, P Sarantinopoulos, E Tsakalidou, and L De Vuyst. The role and application of enterococci in food and health. *International journal of food microbiology*, 106(1): 1–24, 2006.

Fernanda Mozzi, Maria Eugenia Ortiz, Juliana Bleckwedel, Luc De Vuyst, and Micaela Pescuma. Metabolomics as a tool for the comprehensive understanding of fermented and functional foods with lactic acid bacteria. *Food Research International*, 54(1):1152–1161, 2013.

Tim Müller, Britta Walter, Astrid Wirtz, and Andreas Burkovski. Ammonium toxicity in bacteria. *Current microbiology*, 52(5):400–406, 2006.

Barbara E Murray. The life and times of the enterococcus. *Clinical microbiology reviews*, 3(1): 46–65, 1990.

R Nandakumar, Nandakumar Madayiputhiya, and AF Fouad. Proteomic analysis of endodontic infections by liquid chromatography–tandem mass spectrometry. *Oral microbiology and immunology*, 24(4):347–352, 2009.

S Ragnar Norrby, Carl Erik Nord, Roger Finch, et al. Lack of development of new antimicrobial drugs: a potential serious threat to public health. *The Lancet infectious diseases*, 5(2):115–119, 2005.

Omid Oftadeh, Pierre Salvy, Maria Masid, Maxime Curvat, Ljubisa Miskovic, and Vassily Hatzimanikatis. A genome-scale metabolic model of saccharomyces cerevisiae that integrates expression constraints and reaction thermodynamics. *BioRxiv*, 2021.

Brett Olivier and Willi Gottstein. CBMPy release 0.8.2. *Zenodo*, 2021. doi: 10.5281/ZENODO.5546608.

Brett G Olivier, Johann M Rohwer, and Jan-Hendrik S Hofmeyr. Modelling cellular systems with pysces. *Bioinformatics*, 21(4):560–561, 2005.

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

Jeffrey D Orth, Tom M Conrad, Jessica Na, Joshua A Lerman, Hojung Nam, Adam M Feist, and Bernhard Ø Palsson. A comprehensive genome-scale reconstruction of escherichia coli metabolism—2011. *Molecular systems biology*, 7(1):535, 2011.

Shane Thomas O'Donnell, R Paul Ross, and Catherine Stanton. The progress of multi-omics technologies: determining function in lactic acid bacteria using a systems level approach. *Frontiers in microbiology*, 10:3084, 2020.

Bernhard Ø Palsson. *Systems biology: properties of reconstructed networks*. Cambridge university press, 2006.

Jason A Papin, Jennifer L Reed, and Bernhard O Palsson. Hierarchical thinking in network biology: the unbiased modularization of biochemical networks. *Trends in biochemical sciences*, 29(12):641–647, 2004.

Abhijit Paul, Rajat Anand, Sonali Porey Karmakar, Surender Rawat, Nandadulal Bairagi, and Samrat Chatterjee. Exploring gene knockout strategies to identify potential drug targets using genome-scale metabolic models. *Scientific reports*, 11(1):1–13, 2021.

Ian T Paulsen, L Banerjei, GSA Myers, KE Nelson, Rekha Seshadri, Timothy D Read, Derrick E Fouts, Jonathan A Eisen, Steven R Gill, JF Heidelberg, et al. Role of mobile dna in the evolution of vancomycin-resistant enterococcus faecalis. *Science*, 299(5615):2071–2074, 2003.

Stéphane Pinhal, Delphine Ropers, Johannes Geiselmann, and Hidde de Jong. Acetate metabolism and the inhibition of bacterial growth by acetate. *Journal of bacteriology*, 201(13): e00147–19, 2019.

Anna K Pöntinen, Janetta Top, Sergio Arredondo-Alonso, Gerry Tonkin-Hill, Ana R Freitas, Carla Novais, Rebecca A Gladstone, Maiju Pesonen, Rodrigo Meneses, Henri Pesonen, et al. Apparent nosocomial adaptation of enterococcus faecalis predates the modern hospital era. *Nature communications*, 12(1):1–13, 2021.

Thomas Preethee and Rosaline Hannah Deivanayagam Kandaswamy. Molecular identification of an enterococcus faecalis endocarditis antigen efaa in root canals of therapy-resistant endodontic infections. *Journal of conservative dentistry: JCD*, 15(4):319, 2012.

Nathan D Price, Jan Schellenberger, and Bernhard O Palsson. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophysical journal*, 87(4):2172–2186, 2004.

R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2021. URL `https://www.R-project.org/`.

Vytautas Raškevičius, Valeryia Mikalayeva, Ieva Antanavičiūtė, Ieva Ceslevičienė, Vytenis Arvydas Skeberdis, Visvaldas Kairys, and Sergio Bordel. Genome scale metabolic models as tools for drug design and personalized medicine. *PloS one*, 13(1):e0190636, 2018.

MARIO C Raviglione, RAJESH Gupta, CHRISTOPHER M Dye, and MARCOS A Espinal. The burden of drug-resistant tuberculosis and mechanisms for its control. *ANNALS-NEW YORK ACADEMY OF SCIENCES*, 953:88–97, 2001.

Valeria Razmilic, Jean F Castro, Barbara Andrews, and Juan A Asenjo. Analysis of metabolic networks of streptomyces leeuwenhoekii c34 by means of a genome scale model: Prediction of modifications that enhance the production of specialized metabolites. *Biotechnology and bioengineering*, 115(7):1815–1828, 2018.

Eria A Rebollar, Rachael E Antwis, Matthew H Becker, Lisa K Belden, Molly C Bletz, Robert M Brucker, Xavier A Harrison, Myra C Hughey, Jordan G Kueneman, Andrew H Loudon, et al. Using "omics" and integrated multi-omics approaches to guide probiotic selection to mitigate chytridiomycosis and other emerging infectious diseases. *Frontiers in microbiology*, 7:68, 2016.

L Reitzer. Amino acid synthesis. 2014.

Laetitia Remy, Marie Carrière, Aurélie Derré-Bobillot, Cécilia Martini, Maurizio Sanguinetti, and Elise Borezée-Durant. The s taphylococcus aureus opp1 abc transporter imports nickel and cobalt in zinc-depleted conditions and contributes to virulence. *Molecular microbiology*, 87(4):730–743, 2013.

Haluk Resat, Michelle N Costa, and Harish Shankaran. Spatial aspects in biological system simulations. *Methods in enzymology*, 487:485–511, 2011.

Catur Riani. *Effects of oral streptococci and selected probiotic bacteria on the pathogen Streptococcus pyogenes: viability, biofilms, molecular functions, and virulence traits*. PhD thesis, Universiät Rostock. Mathematisch-Naturwissenschaftliche Fakultät, 2009.

Matthew E Ritchie, Belinda Phipson, Di Wu, Yifang Hu, Charity W Law, Wei Shi, and Gordon K Smyth. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47, 2015. doi: 10.1093/nar/gkv007.

Manfred Rizzi, Michael Baltes, Uwe Theobald, and Matthias Reuss. In vivo analysis of metabolic dynamics in saccharomyces cerevisiae: Ii. mathematical model. *Biotechnology and bioengineering*, 55(4):592–608, 1997.

Daniel F Sahm, Jessica Kissinger, Michael S Gilmore, Patrick R Murray, Ross Mulder, Joanne Solliday, and Barbara Clarke. In vitro susceptibility studies of vancomycin-resistant enterococcus faecalis. *Antimicrobial agents and chemotherapy*, 33(9):1588–1591, 1989.

Seppo Salminen and Atte Von Wright. *Lactic acid bacteria: microbiological and functional aspects*, volume 139. CRC Press, 2004.

Pierre Salvy and Vassily Hatzimanikatis. The etfl formulation allows multi-omics integration in thermodynamics-compliant metabolism and expression models. *Nature communications*, 11(1):1–17, 2020.

Eleonora Sattin, Nadia A Andreani, Lisa Carraro, Rosaria Lucchini, Luca Fasolato, Andrea Telatin, Stefania Balzan, Enrico Novelli, Barbara Simionati, and Barbara Cardazzo. A multi-omics approach to evaluate the quality of milk whey used in ricotta cheese production. *Frontiers in microbiology*, 7:1272, 2016.

Christophe H Schilling, David Letscher, and Bernhard Ø Palsson. Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective. *Journal of theoretical biology*, 203(3):229–248, 2000.

U Schillinger, R Geisen, and WH Holzapfel. Potential of antagonistic microorganisms and bacteriocins for the biological preservation of foods. *Trends in Food Science & Technology*, 7(5):158–164, 1996.

Stefan Schuster and Claus Hilgetag. On elementary flux modes in biochemical reaction systems at steady state. *Journal of Biological Systems*, 2(02):165–182, 1994.

Stefan Schuster, Thomas Dandekar, and David A Fell. Detection of elementary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. *Trends in biotechnology*, 17(2):53–60, 1999.

Robert Seviour and Per Halkjær Nielsen. *Microbial ecology of activated sludge*. IWA publishing, 2010.

Samuel A Shelburne, David Keith, Nicola Horstmann, Paul Sumby, Michael T Davenport, Edward A Graviss, Richard G Brennan, and James M Musser. A direct link between carbohydrate utilization and virulence in the major human pathogen group a streptococcus. *Proceedings of the National Academy of Sciences*, 105(5):1698–1703, 2008.

Gunnar Sigurdsson, Ronan MT Fleming, Almut Heinken, and Ines Thiele. A systems biology approach to drug targets in pseudomonas aeruginosa biofilm. *PLoS One*, 7(4):e34337, 2012.

Richard P Silver, Kelli Prior, Christiane Nsahlai, and Lori F Wright. Abc transporters and the export of capsular polysaccharides from gram-negative bacteria. *Research in microbiology*, 152(3-4):357–364, 2001.

Evangelos Simeonidis and Nathan D Price. Genome-scale modeling for metabolic engineering. *Journal of Industrial Microbiology and Biotechnology*, 42(3):327–338, 2015.

Kimmo Sirén, Sarah Siu Tze Mak, Ulrich Fischer, Lars Hestbjerg Hansen, and M Thomas P Gilbert. Multi-omics and potential applications in wine production. *Current opinion in biotechnology*, 56:172–178, 2019.

Kristoffer Sjöholm, Christofer Karlsson, Adam Linder, and Johan Malmström. A comprehensive analysis of the streptococcus pyogenes and human plasma protein interaction network. *Molecular BioSystems*, 10(7):1698–1708, 2014.

Holly Snyder, Stephanie L Kellogg, Laura M Skarda, Jaime L Little, and Christopher J Kristich. Nutritional control of antibiotic resistance via an interface between the phosphotransferase system and a two-component signaling system. *Antimicrobial agents and chemotherapy*, 58 (2):957–965, 2014.

Margrete Solheim, Agot Aakra, Heidi Vebø, Lars Snipen, and Ingolf F Nes. Transcriptional responses of enterococcus faecalis v583 to bovine bile and sodium dodecyl sulfate. *Applied and Environmental Microbiology*, 73(18):5767–5774, 2007.

Leonardo Sorci, Ian K Blaby, Irina A Rodionova, Jessica De Ingeniis, Sergey Tkachenko, Valérie de Crécy-Lagard, and Andrei L Osterman. Quinolinate salvage and insights for targeting nad biosynthesis in group a streptococci. *Journal of bacteriology*, 195(4):726–732, 2013.

Amy J Sterling, William J Snelling, Patrick J Naughton, Nigel G Ternan, and James SG Dooley. Competent but complex communication: The phenomena of pheromone-responsive plasmids. *PLoS pathogens*, 16(4):e1008310, 2020.

Michael E Stiles and Wilhelm H Holzapfel. Lactic acid bacteria of foods and their current taxonomy. *International journal of food microbiology*, 36(1):1–29, 1997.

Charles H Stuart, Scott A Schwartz, Thomas J Beeson, and Christopher B Owatz. Enterococcus faecalis: its role in root canal treatment failure and current concepts in retreatment. *Journal of endodontics*, 32(2):93–98, 2006.

Subhalakshmi Subramanian and C Sivaraman. Bacterial citrate lyase. *Journal of Biosciences*, 6(4):379–401, 1984.

Aditi Swarup, Jing Lu, Kathleen C DeWoody, and Maciek R Antoniewicz. Metabolic network reconstruction, growth characterization and 13c-metabolic flux analysis of the extremophile thermus thermophilus hb8. *Metabolic engineering*, 24:173–180, 2014.

Yutaka Terao. The virulence factors and pathogenic mechanisms of streptococcus pyogenes. *Journal of Oral Biosciences*, 54(2):96–100, 2012.

Bas Teusink, Anne Wiersma, Douwe Molenaar, Christof Francke, Willem M De Vos, Roland J Siezen, and Eddy J Smid. Analysis of growth of lactobacillus plantarum wcfs1 on a complex medium using a genome-scale metabolic model. *Journal of Biological Chemistry*, 281(52): 40041–40048, 2006.

Bas Teusink, Herwig Bachmann, and Douwe Molenaar. Systems biology of lactic acid bacteria: a critical review. *Microbial Cell Factories*, 10(1):1–17, 2011.

Ines Thiele, Nathan D Price, Thuy D Vo, and Bernhard Ø Palsson. Candidate metabolic network states in human mitochondria: Impact of diabetes, ischemia, and diet. *Journal of Biological Chemistry*, 280(12):11683–11695, 2005.

Lance R Thurlow, Vinai Chittezham Thomas, and Lynn E Hancock. Capsular polysaccharide production in enterococcus faecalis and contribution of cpsf to capsule serospecificity. *Journal of bacteriology*, 191(20):6203–6210, 2009.

Kenneth Todar. Todar's online textbook of bacteriology, 2006.

T Tonon and A Lonvaud-Funel. Metabolism of arginine and its positive effect on growth and revival of oenococcus oeni. *Journal of Applied Microbiology*, 89(3):526–531, 2000.

Eduard Torrents. Ribonucleotide reductases: essential enzymes for bacterial life. *Frontiers in cellular and infection microbiology*, 4:52, 2014.

Francesca Turroni, Christian Milani, Sabrina Duranti, Leonardo Mancabelli, Marta Mangifesta, Alice Viappiani, Gabriele Andrea Lugli, Chiara Ferrario, Laura Gioiosa, Alberto Ferrarini, et al. Deciphering bifidobacterial-mediated metabolic interactions and their impact on gut microbiota by a multi-omics approach. *The ISME journal*, 10(7):1656–1668, 2016.

Rogier JP van Berlo, Dick de Ridder, Jean-Marc Daran, Pascale AS Daran-Lapujade, Bas Teusink, and Marcel JT Reinders. Predicting metabolic fluxes using gene expression differences as constraints. *IEEE/ACM transactions on computational biology and bioinformatics*, 8(1):206–216, 2009.

Guido Van Rossum Fred L and Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.

Hendrik W van Veen, Christopher F Higgins, and Wil N Konings. Multidrug transport by atp binding cassette transporters: a proposed two-cylinder engine mechanism. *Research in microbiology*, 152(3-4):365–374, 2001.

Heidi C Vebø, Lars Snipen, Ingolf F Nes, and Dag A Brede. The transcriptome of the nosocomial pathogen enterococcus faecalis v583 reveals adaptive responses to growth in blood. *PloS one*, 4(11):e7660, 2009.

Nadine Veith, Margrete Solheim, Koen WA van Grinsven, Brett G Olivier, Jennifer Levering, Ruth Grosseholz, Jeroen Hugenholtz, Helge Holo, Ingolf Nes, Bas Teusink, et al. Using a genome-scale metabolic model of enterococcus faecalis v583 to assess amino acid uptake and its impact on central metabolism. *Applied and environmental microbiology*, 81(5):1622–1633, 2015.

Nikos Vlassis, Maria Pires Pacheco, and Thomas Sauter. Fast reconstruction of compact context-specific metabolic network models. *PLoS computational biology*, 10(1):e1003424, jan 2014. ISSN 1553-7358 (Electronic). doi: 10.1371/journal.pcbi.1003424.

Mark J Walker, Timothy C Barnett, Jason D McArthur, Jason N Cole, Christine M Gillen, Anna Henningham, KS Sriprakash, Martina L Sanderson-Smith, and Victor Nizet. Disease

manifestations and pathogenic mechanisms of group a streptococcus. *Clinical microbiology reviews*, 27(2):264–301, 2014.

Qing Wang, Kaicen Wang, Wenrui Wu, Eleni Giannoulatou, Joshua WK Ho, and Lanjuan Li. Host and microbiome multi-omics integration: applications and methodologies. *Biophysical reviews*, 11(1):55–65, 2019.

DANIEL E Wexler and P PATRICK Cleary. Purification and characteristics of the streptococcal chemotactic factor inactivator. *Infection and immunity*, 50(3):757–764, 1985.

Yanping Xi and Fei Wang. Extreme pathway analysis reveals the organizing rules of metabolic regulation. *PloS one*, 14(2):e0210539, 2019.

Jia Yan, Yun Xia, Mi Yang, Jiaqi Zou, Yingzhu Chen, Dawei Zhang, and Liang Ma. Quantitative proteomics analysis of membrane proteins in enterococcus faecalis with low-level linezolid-resistance. *Frontiers in microbiology*, 9:1698, 2018.

Jung Eun Yang, Si Jae Park, Won Jun Kim, Hyeong Jun Kim, Bumjoon J Kim, Hyuk Lee, Jihoon Shin, and Sang Yup Lee. One-step fermentative production of aromatic polyesters from glucose by metabolically engineered escherichia coli strains. *Nature communications*, 9(1):1–10, 2018.

Xiaofei Zhang, Arne H Smits, Gabrielle BA van Tilburg, Huib Ovaa, Wolfgang Huber, and Michiel Vermeulen. Proteome-wide identification of ubiquitin interactions using ubia-ms. *Nature protocols*, 13(3):530–550, 2018.