Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

Sofie Tande-Petersen

Analysis of *Bacillus subtilis* Biomass Composition to Determine the Effects of Condition-Specific Biomass Objective Functions in *i*Bsu1147

Master's thesis in MBIOT5
Supervisor: Eivind Almaas
Co-supervisor: Vetle Simensen and Christian Schulz
June 2023

NTNU
Norwegian University of
Science and Technology

Sofie Tande-Petersen

Analysis of *Bacillus subtilis* Biomass Composition to Determine the Effects of Condition-Specific Biomass Objective Functions in *i*Bsu1147

Master's thesis in MBIOT5
Supervisor: Eivind Almaas
Co-supervisor: Vetle Simensen and Christian Schulz
June 2023

Norwegian University of Science and Technology
Faculty of Natural Sciences
Department of Biotechnology and Food Science

NTNU
Norwegian University of
Science and Technology

Abstract

Genome-scale metabolic models (GEMs) are mathematical models of an organisms cellular metabolism under the assumption of steady-state. They are used for computation of metabolic capabilities and phenotype of a given organism [1, 2]. The biomass objective function (BOF) is a central function in a GEM, and consists of the precursor-components for synthesis of cellular biomass, for example DNA, RNA, proteins and lipids [2, 3]. The composition of these precursors have been found to vary with organism, strain, and environmental conditions [2, 4], and predictions by GEMs have proven to be sensitive to changes in the BOF [3]. Still, it is common practice that the coefficients for the BOF are derived from experiments performed with closely related organisms, rather than the one being modeled; or from experiments with the organism in other growth conditions. Because the predictions by GEMs are sensitive to changes in biomass composition, this utilization of data from closely related strains or conditions could impact the validity of the predictions [2, 4].

The aim of this project is therefore to highlight the significance of implementing experimentally derived coefficients in the BOF from the strain and condition being modeled. In order to achieve this, *Bacillus subtilis* was cultivated by batch-fermentation in minimal media with different carbon sources. The cells were harvested during exponential growth followed by analysis of the biomass, demonstrating that the biomass composition of *B. subtilis* can be measured experimentally, and that it changes in response to media composition. The resulting biomass compositions were implemented in the GEM *i*Bsu1147 as new, condition-specific BOFs to explore the effects on growth predictions. The results following pFBA and FVA analyses with the new BOFs demonstrated that the growth rate predictions are quite robust to changes in the biomass composition, but that the flux distribution is more sensitive.

Sammendrag

Genom-skala metabolske modeller (GEM) er matematiske modeller av en organismes cellulære metabolisme under forutsetningen av "steady-state". De brukes til å beregne metabolsk kapasitet og fenotype til en gitt organisme [1, 2]. Biomasseobjektivfunksjonen (BOF) er en sentral funksjon i GEMs, og består av forløperkomponenter for syntese av cellulær biomasse, for eksempel DNA, RNA, proteiner og lipider [2, 3]. Sammensetningen av disse forløperne har vist seg å variere med organisme, stamme og miljøforhold [2, 4], og prediksjoner fra GEMs har vist seg å være sensitive for endringer i BOF [3]. Likevel er det vanlig praksis at koeffisientene i BOF er avledet fra eksperimenter utført med nært beslektede organismer i stedet for den som skal modelleres; eller fra eksperimenter med organismen under andre vekstforhold. Ettersom prediksjonene fra GEMs er følsomme for endringer i biomassekomposisjonen, kan denne bruken av data fra nært beslektede stammer eller andre forhold påvirke gyldigheten til prediksjonene [2, 4].

Målet med dette prosjektet er derfor å synliggjøre betydningen av å implementere eksperimentelt utledede koeffisienter i BOF fra stammen og tilstanden som modelleres. For å oppnå dette ble *Bacillus subtilis* dyrket ved batch-fermentering i minimalt medie fem ganger, tilsatt ulik karbon-kilde. Cellene ble høstet under eksponentiell vekst etterfulgt av analyse av biomassen. Med dette ble det demonstrert at biomassekomposisjonen kan kvantifiseres eksperimentelt, og at biomasse komposisjonen endrer seg i respons på tilgjengelig karbonkilde. De resulterende biomassesammensetningene ble implementert i GEM *i*Bsu1147 som nye BOF-er spesifikke til karbonkilden som var tilgjengelig i mediet, for å undersøke effekten av dette på prediksjonene av modellen. Resultatene etter pFBA- og FVA- analyser med de nye BOF-ene viste at veksthastighet prediksjoner er robuste for endringer i biomassesammensetningen, men at fluksfordelingen er mer følsom.

Acknowledgments

At first, I want to express my heartfelt appreciation to the entire group at Almaas Lab. Thank you all for being so inclusive and contributing to what has been the best year of my studies. A special thank you to my main supervisor Eivind Almaas. Thank you for the motivational speeches during stressful times, the continuous wit, and the spontaneous mortar-and-pestle-technique seminar in the fall of 2022. More importantly, I want to express my deepest gratitude for allowing me to be a part of such a remarkable team, and for your support and guidance.

Another special thank you to my fantastic team of co-supervisors, Vetle Simensen and Christian Schulz. Throughout the course of this thesis, I have crossed the threshold of your office more times than I can count, carrying with me a range of mental states: moments of being unmotivated, motivated, stressed, eager, or nervous. However, every time I've left the office I have felt a sense of joy and motivation. Thank you both for the patience, for everything you have taught me, and for the jokes and laughter. The guidance and support has been invaluable.

Kåre Kristiansen, thank you for your expertise, help, and support in the lab. I would also like to thank Laura García Calvo and Divyata Vilas for everything you have taught me. My gratitude to Siri Stavrum for the assistance in the lab! Thank you Emil Karlsen and Jacob Pettersen for being so supportive and for all the interesting conversations about everything and anything.

Linn Sandvik, thank you for being there through a year of incredible highs and incredible lows. Thank you for the jokes and laughter when the days were long. Thank you for being such a great lab-partner, and more importantly for being such a good friend. Du er hel ved!

Last but not least, I want to express appreciation to my friends, partner, and family for unwavering support throughout this journey!

List of abbreviations

B. subtilis	*Bacillus subtilis*
BOF	Biomass objective function
cDW	Cellular dry weight
DO	Dissolved oxygen
FBA	Flux balance analysis
FVA	Flux variability analysis
GEM	Genome-scale metabolic model
HPLC	High-performance liquid chromatography
LB	Lennox broth
MOMA	Minimization of metabolic adjustment
MS	Mass spectrometry
MQ-water	Milli-Q water
nBOF	New biomass objective function
NMR	Nuclear magnetic resonance
oBOF	Original BOF
OD	Optical density
PES	Polyethersulfone
pFBA	Parsimonious flux balance analysis
PPP	Pentose phosphate pathway
RCF	Relative centrifucal force
rpm	Rotations per minute
RQ	Respiratory quotient
TCA	The citric acid cycle
TMS	Trace mineral solution

List of Figures

List of Tables

Contents

1. Introduction

Biological systems such as cells, can be described as systems within systems. Within a cell, there is a network of multiple interacting components [5] such as proteins, genes, metabolites [6] connected by associations, reactions and pathways. Each reaction and pathway have characteristics and distinct properties. Together they combine and make up a cell with it's own emergent properties - properties that are not inherent to the components of the system themselves, but exhibited by the system as a whole. Understanding how a system is assembled is important, however it is also important to understand why systems are assembled the way that they are [5]. For instance, being familiar with the mechanics of gene expression such as translation and transcription, does not in itself explain or illustrate how this process relies on coordinated action of several gene products at the same time [7, 8]. Thus, a biological system can be described as hierarchical structures of subsystems and these subsystems are linked together to make up complex networks [5] that describe how the components are related and how they function together [9].

Since the full genome sequence of *Haemophilus influenza* was published in the mid 1990s, it became possible to map all gene products involved in biological processes in an organism [7]. Further, increased availability of genomic and metabolomic data (Big Data) were enabled in the 21st century as technological advancements enabled more cost-efficient generation of data. There was an exponential increase in available biological data, but processing it all became a challenge. Systems biology and machine learning approaches helped integrate and structure this information in these datasets to understand the interactions between the components [6].

One method of structuring the available data is with Genome scale metabolic models (GEMs). GEMs are mathematical models of an organisms cellular metabolism [1]. The models contain information about the genome, enzymes and gene-protein associations, reactions and metabolites that are linked by their interactions [6]. The information is structured by reaction and metabolite stoichiometry under the assumption of steady-state. GEMs are used for computation of the metabolic capabilities of an organism and its phenotype [1], and can be analysed quantitatively by the addition of constraints as mathematical functions, followed by reevaluation of feasible phenotypic states [8]. One such analysis is Flux Balance analysis (FBA) where the flow of metabolites (flux) through the model is simulated during maximal growth.

Central in the GEM and for analysis of the model with FBA, FVA and MOMA is the biomass objective function (BOF). The BOF is implemented in the GEM as a reaction consisting of macromolecules that are necessary for cell growth and proliferation. It contains the building-blocks of the cells of the organism the GEM is modeling [10], and so the formulation of an accurate BOF is essential reaction for quality model predictions [11]. To construct the

BOF, information about the biomass composition of the cell and its energy requirements are necessary [12]. As the proportions of an organisms biomass is not genetically specified, the composition should be experimentally measured for the organism in study [2, 4]. The compositions of biomass is known to change with environmental conditions and growth rate, as well as between strains and organisms [2, 13]. However, the BOFs are often constructed automatically, based on data from published research which might not correspond to the organism the BOF is intended for, or the conditions that the organism is modeled in. Therefore, the resulting predictions made by such models might not represent the organism [4, 2].

A precise estimation of the BOF is essential for accurate model predictions, but it is often constructed based on published research where strain, nutrients, or overall conditions are different from the target organism [4, 2]. For instance, *i*Bsu1147 is a GEM of the bacterium *Bacillus subtilis* that was constructed based on two older models, *i*Bsu1103 and *i*YO844. Interestingly the BOF has not changed much across these reconstructions, with *i*Bsu1147 using the same research papers for their BOF coefficients as the two aforementioned models [14]. Using previously published research is in itself is not a problem, but the coefficients are from experiments that are performed in different conditions, various strains and methods of cultivation, drawing the validity of the *i*Bsu1147 BOF into question.

As stated, it has been shown that changes in BOF coefficients affect the predictions of the model, and that the biomass composition changes in response to environmental conditions [2, 11, 13]. The aim of this study was to perform biomass analyses of *B. subtilis* cultivated in various media compositions by batch fermentation to asses whether the conditions affect the biomass composition. Further, the measured biomass compositions were used to formulate new, condition-specific BOFs. These were implemented in *i*Bsu1147 to assess how this affects the model predictions. The results from these analyses were examined to determine the necessity of condition specific BOFs.

This project can be divided into four parts - Cultivation of *B. subtilis*; biomass analyses of the harvested cells of the bacterium; implementation of the experimentally derived data in a genome scale metabolic model (*i*Bsu1147); and various analyses of the model with the experimental data. *B. subtilis* was cultivated by batch fermentation in bioreactors containing minimal media supplemented with different carbon-sources. The experiments were performed together with Linn Sandvik, another masters student. During the fermentation, various analyses were performed allowing for calculation of growth-rates as well as uptake- and secretion rates for various compounds. The cells were harvested in the exponential growth phase and the biomass composition was quantified following the pipeline established by Simensen et al. (2021) [2]. The measured biomass composition was used to construct a new condition-specific BOFs for each cultivation. These were implemented in *i*Bsu1147, and the model was analysed by FBA, FVA and MOMA to see how the new BOFs affect the model predictions.

2. Background

This chapter presents the theoretical foundation for understanding the project. It is initiated with information about the organism of study, *B. subtilis*, and its metabolism. This is important for being able to analyze how the bacterium adapts to the carbon-sources used in the cultivation-experiments. Then, metabolic modeling is presented along with the analyses that were performed - flux balance analysis (FBA), flux variability analysis (FVA) and minimization of metabolic adjustment (MOMA). The genome scale metabolic model *i*Bsu1147, is presented with special attention to the biomass objective function, as this is the reaction that is updated in this project. The background chapter ends with the theory behind the methods used to experimentally quantify the amount of DNA, RNA, proteins and lipids — the biomass — of the bacterium.

2.1 *Bacillus subtilis*

The use of microorganisms in fermentation, bio-production, and -transformation continues to develop. As *B. subtilis* is generally recognized as safe (GRAS), is able to metabolize a wide range of substrates, and has well-characterized and efficient protein secretion pathways, it can be used as a powerful cell factory [15]. It is already widely used in industrial production of various products such as enzymes and vitamins [16]. The bacterium lacks an outer cell membrane[15], which classifies it as Gram-positive [16]. This is a beneficial trait in the industry, as it allows more efficient relocation of proteins into the extracellular space (ECM) [15].

B. subtilis has been thoroughly studied over more than 40 years. Its biochemistry, physiology and genetics are therefore well characterized [17]. Since the last annotation of its genome, more than eight thousand new references have been uploaded to PubMed that are related to the bacterium. Many of the publications are about the function of identified genes, or about the potential it has in metabolic engineering. Although it is well characterised, a better understanding of its metabolism is essential for better use of *B. subtilis* in industrial applications, for example metabolite production [18].

2.1.1 *B. subtilis* metabolism

B. subtilis metabolizes glucose as its preferred source of energy and carbon [19, 20]. Glucose enters glycolysis through phosphorylation and is further metabolized in the pentose phosphate pathway (PPP), and the citric acid cycle (TCA). The sugar is oxidized to carbon dioxide

and in the process ATP is generated [19]. However, the bacterium can metabolize a wide range sugars and organic compounds [20]. Other sugars and sugar alcohols can be used by *B.subtilis* by being phosphorylated and entering glycolysis similarly to glucose. Organic acids are utilized by conversion to intermediates in the TCA, gluconeogenesis or PPP [19].

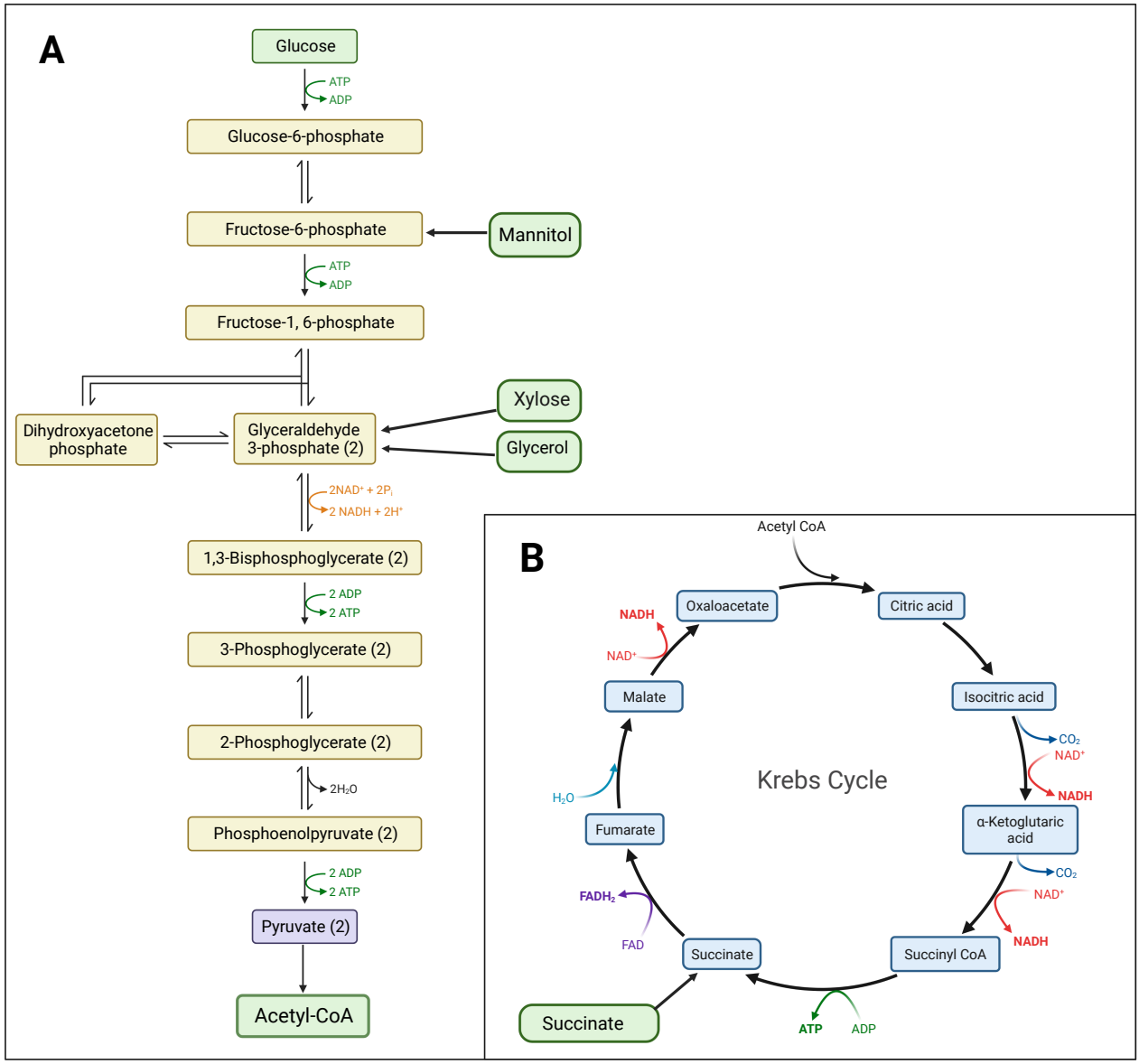


FIGURE 2.1. Illustration of where various carbon sources enter the glycolysis or TCA in *Bacillus subtilis*. **(A)** shows where in the glycolytic pathway xylose, mannitol and glycerol enter glycolysis. **(B)** shows where succinate enters the citric acid cycle. Figure adapted from [21] using Biorender.com.

Figure 2.1 illustrates various carbon sources that *B. subtilis* can utilize and how they enter the metabolic pathways. Xylose is taken up via the arabinose transporter, converted to xylulose-5-phosphate via the isomerase pathway, and metabolised through the PPP before entering glycolysis as glyceraldehyde-3-phosphate [22]. Mannitol, a sugar alcohol [23], can also be used as a sole source of carbon and energy for *B. subtilis*. It is taken up by a phosphotransferase system specific to mannitol as mannitol-1-phosphate. This is then either converted to fructose-6-phosphate and enters the glycolytic pathway, or it is dephosphorylated and stored inside the cell [24]. Glycerol is metabolised by glycerol kinase and GP dehydrogenase to form dihydroxyacetone phosphate which is an intermediate in glycolysis ([25]). Succinate enters the metabolism as an intermediate in the citric acid cycle [21] as illustrated in Figure 2.1. However, growth on organic acids such as succinate require the cells to reverse the flux of the glycolytic pathway to gluconeogenesis in order to produce hexose and pentose-phosphates which are important biomass building blocks [26].

2.2 Bioreactors

The technology for cultivation of bacteria has in recent years developed rapidly [27]. Bioreactors are one example of cultivation equipment for fermentations, developed and designed to accommodate cellular growth in controlled environments. They are vessels containing necessary nutrients, substrates, and microbial cells where the end goal usually is the production of a desired product. The main objective of a bioreactor is to maintain specific conditions needed for the organisms to grow, and generally perform a biochemical transformation. Though the bioreactor has existed in various forms throughout history, there has recently been significant advancements in bioreactor technology due to progression in design and an increased demand for biochemical products [28].

As the field of biochemical engineering has progressed, the cultivation process could be performed with an increasing amount of control over the conditions. Solutions for measuring and monitoring the physical and chemical environment developed, allowing adjustments to be made during fermentation to control conditions of growth. Physical variables that can be controlled include temperature and agitation, as well as gas and liquid flow rates. The chemical environment, such as the pH and partial gas pressure can be registered using electrodes submerged in the media. Additionally, gas concentrations can be measured with a mass spectrometer. The ability to measure these parameters in the environment is fundamental to controlling the cultivation [28].

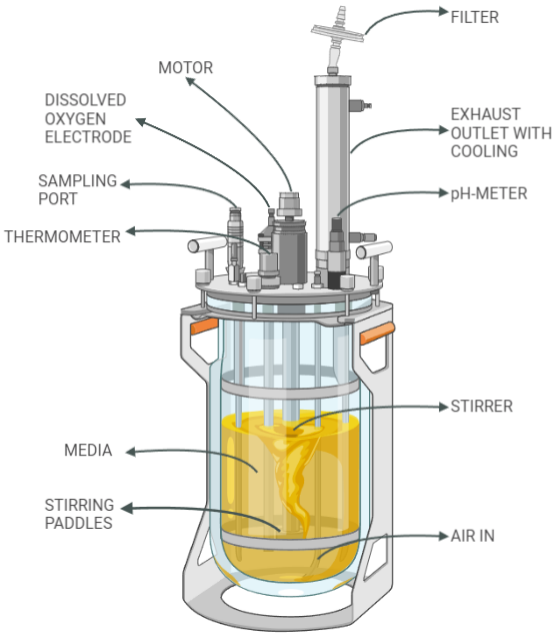


FIGURE 2.2. A bioreactor with labels on important components for controlling growth conditions when fermenting bacteria or other microorganisms. The media is loaded into the wessel and is stirred by a motor-powered stirrer. Air comes in at the bottom of the vessels directly into the media, and goes out through the exhaust outlet. Media, OD and dryweight samples are taken from the sampling port. The dissolved oxygen (DO) and pH-meter measure the oxygen levels in the media and pH, respectively. The illustration was created using BioRender.com.

The organism, its genetics and strain, and the conditions in the bioreactor all affect the rate of growth, as well as the number of microorganisms and the availability of nutrients [28]. This brings into attention another important aspect of bioreactor cultivation. Besides providing the ability for microbial growth in optimal conditions, the substrate concentration is also controlled. The concentration may vary and change over time, depending on factors such as the number of microorganisms and whether the growth process is a batch-, fed batch-operation [29] or continuous. The difference between these modes is whether or not the system is closed. In a fed-batch fermentation, growth limiting nutrients are added, and in a continuous culturing (chemostat) there is constant removal of old media and addition of fresh media [30].

2.2.1 Batch-mode fermentation

Fermentation in batch-mode entails that the system is partially closed. There is no addition or withdrawal from the culture. Growth media and most materials required are aseptically added to the bioreactor before inoculation of the particular microorganism [31]. Since the system is partially closed, all substrate is added to the media before the start and then the process runs its course without addition of more substrate [30]. There is withdrawal of media

and biomass only for sampling or harvesting at the end of the fermentation. Air, anti-foam, and pH-controlling agents are added during the fermentation as needed. Other than that, the system is closed until the fermentation is completed. During the batch fermentation, there is a program that monitors and controls the conditions. This ensures the environment is stable, and the only changes occurring is what is being taken up from and released into the media, and the microorganisms growing [31].

Compared to fed-batch or continuous mode of fermentation where substrate or media is added during the fermentation, the stages of microbial growth in batch operations are not controllable. This means that the bacteria can not be kept in a specific state of growth [30]. The process is started when a small amount of a cell culture is added to the media in the reactor. Unless the cells are harvested at some point, the growth runs its course through the growth phases described below [31].

Typically, growth in a batch fermentor can be divided into four phases, as Figure 2.3 illustrates. The cell culture is added to the fermentor from a starter-culture [31], usually followed by what is called a lag-phase [31, 32]. In the lag phase, the cells reorganize their metabolism to the new environment. This is, for example done by accelerating the synthesis of enzymes that are needed depending on the nutrients available in the medium [33]. After the cells have adjusted they enter a growth phase, usually characterized by an exponential increase in the number of cells. Unless the fermentation process is stopped, the cells will continue to grow into the next phase. The growth slows down in this stage because most of the nutrients, or a growth limiting nutrient, have already been consumed during exponential growth. This is often referred to as the second growth phase, where growth slows down. The next phase is the stationary phase, where essential nutrients for cell growth have been depleted, followed by the death phase. This is the final stage. Here, the concentration of viable cells rapidly decreases [31, 32].

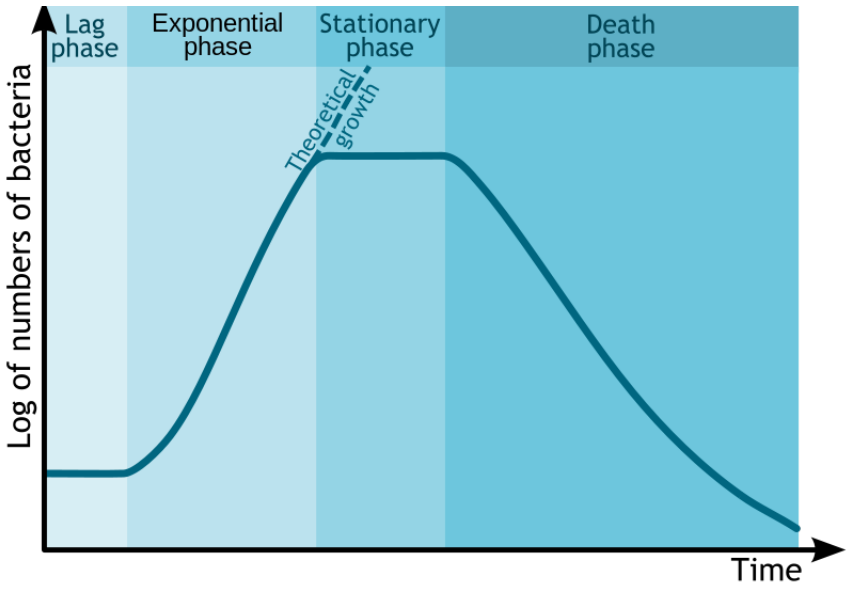


FIGURE 2.3. Illustration of the four growth phases of bacteria in batch fermentation. After introduction to a new media, there is a lag-phase where the bacteria reorganize their metabolism to adapt to the environment. When the bacteria has adjusted to the environment, they enter a phase of exponential growth. The dashed line representing "theoretical growth" illustrates the expected growth if the old media was replaced with fresh media, as in a chemostat set up, allowing the bacteria to continue exponential growth [34]. Because the figure illustrates the growth in batch fermentation however, the bacteria continue to stationary phase [31]. Figure retrieved from Komorniczak (2009) [35].

2.3 Metabolic modelling

Biological systems are networks of mutually dependent, interacting, components that integrate to form a unified entity with its own emergent properties - properties only possessed by the whole system, not by its individual components in isolation. There are several layers of hierarchy in biological systems. A metabolic pathway is one example of a system at lower level, where substrates are converted to products by enzymes. A cell is a biological system at a higher hierarchical level, where pathways are mutually dependent and interact, resulting in emergent properties of the cell. The various levels of hierarchy are interlinked. Mutations in the a gene can affect the emergent properties of the cell, its phenotype [5]. A phenotype is a characteristic of cells that can be observed. In metabolic modeling the focus is often on growth phenotypes, which refers to characteristics of growth such as if it grows and how quick the organism proliferates [36].

Understanding the complex relationship between genes and phenotypes has long been a challenge in molecular biology, as most phenotypic traits are a result of coordinated action between several gene-products [7]. Additionally, gene-regulatory networks and pathways interact with each other. Thus, the complexity of biological systems can not fully be understood by studying genes, proteins and pathways in isolation. To achieve a comprehensive under-

standing, the dynamic interactions of the components and structure of the cellular system have to be examined [37].

Advancements in the field of molecular biology enabled the collection of extensive sets of data through genome sequencing and high-throughput measurements [37]. Genome sequencing has enabled the identification of gene products involved in complex pathways in an organism. The increase in available metabolic data enabled the reconstructions of metabolic networks of given organisms on genome-scale, called genome-scale metabolic models (GEMs) [7]. These are reconstructions of an organisms metabolism, reconstructions, systematized as a mathematical representation [6, 38]. One of the objectives in metabolic modeling is to use GEMs to predict the behavior of cells, for example their growth phenotype [36].

2.3.1 Genome scale metabolic models

GEMs enable *in silico* exploration and analyses of microorganisms, their metabolism and genomic data [39]. These mathematical models contain biochemical, genetic and genomic data such as information about enzymes, gene-protein associations, reactions and metabolites [6], and allow for computation of phenotypic traits [7, 39]. They are mathematical representations of an organisms metabolism [1], and have become powerful tools for studying the genotype-phenotype relationship [40].

GEMs are used for computation of the metabolic capabilities of an organism, its phenotype. Because growth is directly related to gene-protein relationships, it is important that the model is constructed based on data that provides insight about how the cells adapt to various conditions [36]. It is also important that the models are structured in such a way that they account for stoichiometry and reversibility of the reactions, assuming the metabolic system is in a steady-state [1, 6], and and how the the underlying metabolic fluxes — the flow of metabolites — are distributed across the metabolic model [41].

A GEM is constructed by first listing all available data on reaction equations, compounds and compartments of the organism. The reaction equations are then converted to a matrix with the stoichiometric coefficients from the reactions, as shown in Figure 2.4. The relationship between reactions and compounds of the network is defined by this matrix [39]. The stoichiometric matrix ($\mathbf{S}$) has a size of $m \times n$, where m are the number of metabolites in the model and n are the number of reactions [42]. As illustrated in Figure 2.4B, each column represents a reaction, and contains the stoichiometric coefficients of the metabolites associated with the given reaction. Each row is associated with a given metabolite. If an entry in $\mathbf{S}$ has a negative value, this indicates the substrate was consumed. A zero means it is not a part of the reaction and a positive value indicates the production of the compound. This matrix is the first step in defining the framework of the organisms metabolism, and thus represents the first level of constraints on the model [39, 42]. Each coefficient in the matrix constrains the model by imposing limitations on the flow of metabolites, or flux, through the network, thereby ensuring mass balance in the system. The total mass produced is equal to the amount

consumed [42]. This way, the stoichiometric matrix defines the space of achievable metabolic phenotypes [39].

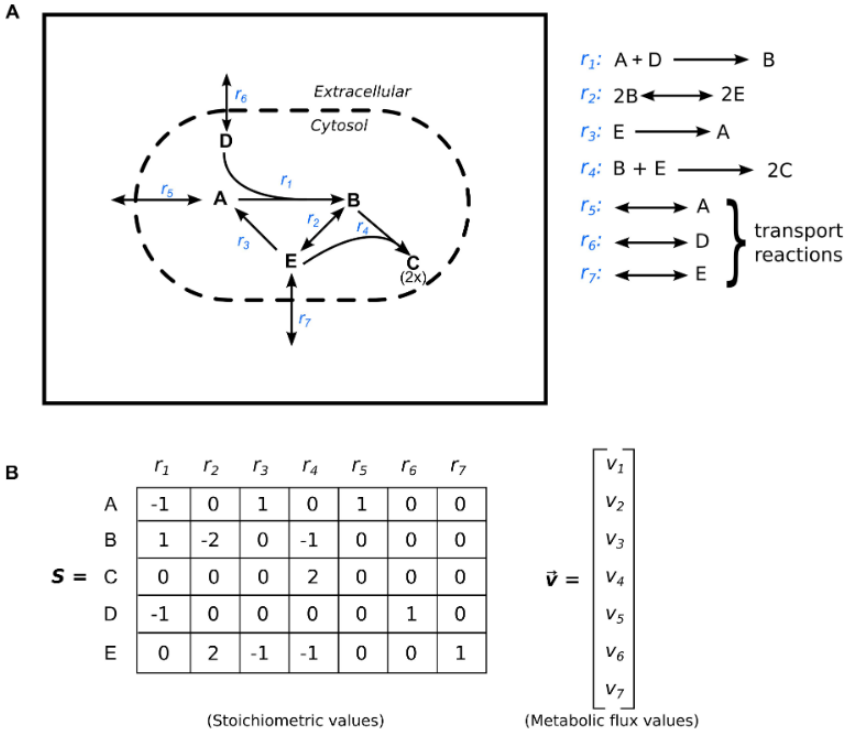


FIGURE 2.4. **(A)** shows an example of a bacterial model and how the reactions for it can be listed. The reactions are labeled in blue text. In figure **(B)** the stoichiometric matrix for that bacterial model has been made, illustrating how one uses -1 or 1 to indicate consumption or production of the metabolites in the reaction. Figure retrieved from Cuevas et al. (2016).

The compartment of the substrates or compounds in a reaction is denoted in the stoichiometric matrix. This is another constraint in the system as it enables the differentiation of compounds inside the cell from those outside the cell. It also enables the model to include reactions of transportation across the outer membrane or cell wall of the organism. Other higher-level constraints on the model are media composition, as well as uptake- and secretion from the media, defined by exchange reactions in the stoichiometric matrix. However, the model also includes boundaries for each reaction. These boundaries limit the flux through the reaction. Otherwise, one or several reactions could achieve unrealistically high or low fluxes. Flux boundaries are also used to define whether or not a reaction is reversible [39].

The next step for construction of a GEM is setting the objective [42]. This means defining the reaction that represents the main objective of the cell [39]. The objective is normally growth, represented by biomass production, in which case it is referred to as the biomass objective function (BOF). This function defines the the conversion of metabolic compounds to biomass constituents, such as lipids, proteins, DNA, RNA etc [39, 42]. The BOF is implemented in the model as an additional reaction in a column of stoichiometric coefficients, where precursor

metabolites are consumed in a manner that stimulates production of biomass [42]. The BOF is scaled so that the flux through this reaction equals the growth rate [42].

The BOF is constructed based on data on the organisms biomass composition. This data should be based on experimentally derived measurements of biomass composition for the organism being modeled [42]. This is important for accuracy in model predictions because biomass composition has been shown to vary with organism, strain and environmental conditions. However, data on biomass composition is not available for most organisms [2]. Therefore, the biomass composition is often derived from experiments performed with closely related organisms, another organism entirely, or from experiments performed in different conditions than what is being modeled. This can affect the validity of the predictions made by the model [2, 4]. This will be explained in further detail in section 2.5.

In order to construction a GEM that well represents the genotype-phenotype relationship, the generation of the model must be based on a few fundamental features. A few of which have been introduced so far; all cellular functions can be described by chemical equations, genome sequencing and high-throughput measurements enable constructions of metabolic models [37, 43], cells are subjected to constraints, cells adapt to their environment, and the mass is conserved. When these features are combined, they represent the concept of what is called Constraint-Cased Reconstruction and Analysis (COBRA)[43]. When the stoichiometric matrix is made, the reconstructed metabolic model can be analysed following the constraint based approach. Constraints-based approaches entail identification of feasible metabolic states, or growth phenotypes of the biological network that satisfies the constraints that have been applied [44] .

2.3.2 Flux balance analysis

After the construction of the stoichiometric matrix is completed, and the media has been defined, boundaries for compounds and reactions are set and the objective function has been defined, the GEM is ready for analysis [39]. A common approach to study the flow of metabolites (the flux) through the network is flux balance analysis (FBA) [42]. The purpose of FBA is to find a feasible set of steady state fluxes that optimizes a specified objective of the cell, for example maximizing biomass production within the space that the constraints of the network define [45]. Essentially an FBA harnesses the information in the GEMs to make predictions on growth and corresponding flux distribution of the model [42]. The biological basis of the analysis is the knowledge that an organisms over time will adapt to their growth environment by altering their metabolism to achieve optimal growth given the conditions. In FBA, the growth rate and corresponding fluxes are based on the assumption that the organism has evolved to an optimal state [46].

The system is solved using linear programming in FBA [39, 42]. "Solving" the system refers to maximizing or minimizing the objective-function, depending on what is being analyzed [42]. In biological systems, there are usually more reactions than metabolites. This means that

the stoichiometric matrix (**S**) has more columns than rows, resulting in plurality of possible distributions of metabolites through the network (flux distributions). This means there are several feasible solutions in the model, but under a particular set of constraints, for example definition of the media, only a set of feasible phenotypes can be expressed [8]. While the stoichiometric coefficients, flux values and objective of the system define a range of possible solutions for solving the system, an FBA will identify single points within this range, as illustrated in Figure 2.5 [42].

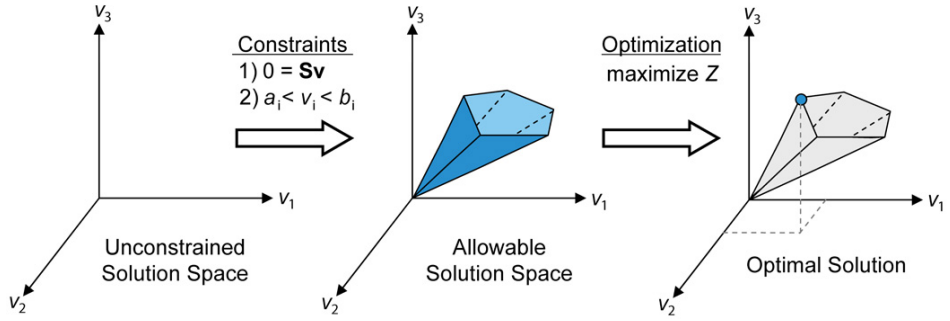


FIGURE 2.5. An FBA analysis can be performed on a Genome Scale Metabolic network (GEM), in order to identify the optimal flux distribution. The solution is on the edge of the solution-space allowed by the stoichiometric matrix, reaction fluxes, media composition and the objective of the model. Figure retrieved from Orth et al. (2010) [42]

.

With the objective being biomass production, FBA seeks to maximize the flux through the BOF as much as possible given the restraints on the system [42]. When the network has been constructed and the stoichiometric matrix has been defined, the mass balance of the system can be defined as

$$\frac{dx}{dt} = \mathbf{S} \cdot \mathbf{v}. \tag{2.1}$$

S is the stoichiometric matrix and **v** is the vector of fluxes that correspond to each reaction in the network. dx/dt denotes the rate of change of a vector x, a vector with the concentrations of all the metabolites in the system, with respect to time. Overall the equation represents how the rate of change of metabolite concentrations is determined by the stoichiometry of the reactions in the model and the flux of the reactions [45].

There is no change in the amount of compound x over time, t when the system is in steady-state. In this case, the reaction defined in equation 2.1 can be written as

$$\mathbf{S} \cdot \mathbf{v} = \mathbf{0}. \tag{2.2}$$

This represents the steady-state constraint of FBA. The optimized solution is calculated assuming the system is in steady-state [45], and is found by maximizing or minimizing the objective function which is defined by

$$Z = \mathbf{c}^T \mathbf{v}. \tag{2.3}$$

In Equation 2.3, $\mathbf{c}$ is a vector that contains the weights of how much each reaction contributes to the BOF. $\mathbf{c}$ defines the objective as a linear combination of fluxes in the model, which commonly is a vector with all zeros except a single non-zero element for the BOF. In FBA the equation 2.2 below is solved given a set of upper- and lower bounds for the reactions in $\mathbf{v}$, and a linear combination of fluxes as an objective function, as shown below with equations 2.3, 2.2 and 2.4. The output of this analysis is a flux distribution $\mathbf{v}$ that maximizes the BOF [42].

$$\max \quad Z = \mathbf{c}^T \mathbf{v} \tag{2.3}$$

$$\text{s. t.} \quad \mathbf{S} \cdot \mathbf{v} = \mathbf{0} \tag{2.2}$$

$$\mathbf{v_l} \quad \leq \quad \mathbf{v} \quad \leq \quad \mathbf{v_u}. \tag{2.4}$$

The system of equations above formulate the FBA-problem as a linear programming problem, where equation 2.3 is solved such that (s. t.) the upper (v_u) and lower bounds (v_l) for the reactions in $\mathbf{v}$ are upheld [42].

2.3.3 Flux variability analysis

A flux variability analysis (FVA) is used to calculate the maximum and minimum flux for each reaction in a GEM while satisfying the objective, for example biomass production [47]. Because a FBA solution normally is not unique, FVA can be used to find the range of possible fluxes that reside in the solution of the FBA-problem [48]. The maximum and minimum value is found for each reaction given the highest and lowest achievable flux possible through each reaction, given the objective. This analysis is optimal for studying flux distributions and the robustness or flexibility of the GEM. $\mathbf{c}^T$ represents the objective, for example biomass production. The FVA is performed by solving two optimizations for each flux (v_i) [47],

$$\begin{aligned} &\max/\min\ v_i \\ &\text{s. t. } \mathbf{Sv} = \mathbf{0} \end{aligned} \tag{2.5}$$

$$\begin{aligned} \mathbf{c}^T \cdot \mathbf{v} &\geq \gamma Z_0 \\ \mathbf{v}_l \leq \mathbf{v} &\leq \mathbf{v}_u. \end{aligned}$$

γ is a parameter that defines whether the analysis is performed in optimal ($\gamma = 1$) or sub-optimal ($0 \leq \gamma < 1$) conditions. $\mathbf{v}_u$ and $\mathbf{v}_l$ represent the upper and lower bounds possible for a given reaction. The optimal solution is where $Z = \mathbf{c^T v}$. During FVA, each reaction is iterated through and two LP optimization problems are solved [47].

2.3.4 Minimization of Metabolic Adjustment analysis

As mentioned, FBA assumes that the organism is growing optimally. However, this is not always the case, as for example with unevolved mutants or during response to metabolic limitations. With MOMA, one can predict the behavior of strains in such conditions where they exhibit sub-optimal growth [49]. This analysis is performed to find a solution-vector that is closest to a given reference-vector of fluxes ($\mathbf{w}$). The output of this analysis is a vector $\mathbf{x}$ that contains a list of fluxes that are feasible and closest to the fluxes in vector $\mathbf{w}$. The aim is to minimize the Euclidean distance from the reference fluxes to fluxes that are actually possible for the network to function with, using quadratic programming. It predicts a metabolic phenotype by minimizing the distance in flux space. The analysis can be used with experimental data in vector $\mathbf{w}$. The objective in MOMA is to minimize the Euclidean distance, D [46], which is defined as

$$D(\mathbf{w}, \mathbf{x}) = \sqrt{\sum_{i=1}^{N} (\mathbf{w_i} - \mathbf{x_i})^2}. \tag{2.6}$$

where distance between vectors $\mathbf{w}$ and $\mathbf{x}$ is calculated by finding the sum of squared differences between the two, and then calculating the square root of the sum. In MOMA, the original distance function has been transformed [46]. The objective is to identify a feasible flux state $\mathbf{v}$ that is closest to the reference flux distribution $\mathbf{w}$(Ref. equation 2.7), while still maintaining the steady-state-(Ref. equation 2.8), and the capacity constraints in the model(Ref. equation 2.9) [43],

$$\min||\mathbf{v} - \mathbf{w}||^2, \tag{2.7}$$

$$\mathbf{S} \cdot \mathbf{v} = 0, \tag{2.8}$$

$$v_{j,min} \leq v_j \leq v_{j,max}. \tag{2.9}$$

2.3.5 Constructing a biomass objective function

An accurate representation of the composition of the cell that is needed for proliferation is a requirement for GEM predictions that well represents the organism in study. The BOF is the function in the model that defines this composition, by specifying the components needed for growth and their respective amounts required [2]. It is central part of analyses such as FBA, FVA and MOMA. Therefore, an objective function that is based on specific physiological measurements is key in successful FBA analyses [45, 2]. In order to formulate this function, the composition of the cell must be known as well as the energy requirements for biomass generation [12].

The BOF, as mentioned, is represented as a reaction in the stoichiometric matrix. The stoichiometric coefficients of this reaction represent the requirements for biomass generation at steady-state [11]. Depending on the available data, the BOF of a given organism falls into one of three categories. A basic BOF only includes the macromolecular composition of the cell, meaning the amount and synthesis of macromolecules such as proteins, RNA, DNA and lipids. A BOF at intermediate level also includes maintenance costs of synthesis of macromolecules and for the cell in general, meaning the energy costs of driving biosynthetic processes and for synthesizing building blocks. Advanced BOFs have metabolites specific to the organism, such as coenzymes, cell wall components etc. [12]. Eventually an advanced BOF will have a structure similar to the example equation 2.10, where c_i represent the coefficients for each component in the BOF.

$$\begin{gathered} c_1 \text{ protein} + c_2 \text{ DNA} + c_3 \text{ RNA} + c_4 \text{ lipid} + c_5 \text{ cell wall} \\ c_6 \text{ cofactors and ions} + c_7 \text{ ATP} + c_7 \text{ water} \\ \rightarrow \\ c_7 \text{ ADP} + c_7 \text{ phosphate} \end{gathered} \tag{2.10}$$

To construct a BOF, the coefficients (c_i) have to be calculated for the macromolecular components that are necessary for biomass production [11] - DNA, RNA, proteins, lipids, carbohydrates, etc. The coefficients represent the relative amount of a each component in the BOF measured in gram per gram dry weight (g/gDW) of the biomass. These coefficients are scaled so that all the components in the BOF together adds up and represents 100% of the cellular dry-weight, or biomass [16]. Scaling the coefficients is a critical step of assuring

valid predictions based on the BOF. It enables quantitative comparisons to be made between biomass production and growth yields, such as growth rate. The scaling of the BOF is important for the model to account for stoichiometry in substrate conversion to biomass. It also ensures accuracy as it enables the comparison of simulated predictions to experimental data. The biomass produced must have a molecular weight of 1 g/mmol to ensure the steady-state assumption is also applied to the the biomass function [50].

Synthesis reactions for the components in the BOF have to be defined as well, in order for the components of the BOF to be produced by the organism. Otherwise, the constituents of the BOF are not produced and there will be no growth. For example, DNA is a component in the BOF. The synthesis reaction for DNA is the reaction in which dAMP, dGMP, dCMP and dTMP make DNA [16];

$$x_1 \text{ dAMP} + x_2 \text{ dGMP} + x_3 \text{ dCMP} + x_4 \text{ dTMP} \rightarrow \text{DNA}. \tag{2.11}$$

The coefficients for the components of the synthesis reactions for the macromolecular components in the BOF are also scaled. For example for the reaction in equation 2.11, the coefficients for the nucleotides (x_i) are scaled so that the g/gDW (%) of the four substrates adds up to represent 100% of the g/gDW of DNA. By doing this, the coefficients in the BOF can be implemented directly as g/gDW DNA found experimentally of the total biomass of the modeled organism [16].

2.4 A genome scale metabolic model of *B. subtilis* - *i*Bsu1147

The genome scale metabolic model *i*Bsu1147 is one of the newest models available of *B. subtilis*. It is based on the two older models, *i*YO844 and *i*Bsu1103 [14]. Because the BOF is central to the analyses described (FBA, FVA and MOMA), the focus is on how it was developed.

The oldest of the two models, *i*YO844, was one of the first highly detailed genome scale metabolic networks for *B. subtilis* and was constructed by Oh et al. (2007). It contains 844 genes, 1020 reactions and 988 metabolites. The objective function for this model was developed to account for the metabolites that go into biomass. It consists of six macromolecular components; DNA, RNA, protein, lipid, lipoteichoic acid and cell wall components, and ions and metabolites as well [16].

The coefficients for the components in the Oh et al. (2007) BOF was gathered from previously published experimental data by Dauner and Sauer (2001) [16].They performed experiments with glucose-limited continuous cultures with *B. subtilis*. In these experiments they measured the content of RNA and proteins at various dilution rates, and found there

was a linear relationship between RNA content and growth rate, and proteins and growth rate. This means that an increase in growth rate leads to an increase in RNA- and protein-concentrations. Their measured concentrations were used to find the coefficients for RNA- and protein-content in the *i*YO844 BOF. The coefficients other macromolecules in the BOF are derived from this same paper, however they were not experimentally measured by Dauner and Sauer. They were found in other published research that Dauner and Sauer refer to in their paper [13]; the DNA and lipid content is based on a paper from Bishop et al. (1997); metabolites and ions are based on data from Umbarger (1996); Lipoteichoic acid and cell wall composition are both based on data from Foster et al. (2002) and Dauner and Sauer themselves [16, 13] The composition of amino acids and nucleotides in the model are adopted from two papers; Dauner and Sauer (2001) [13], and Sauer et al. (1996) [51]. The lipid components are adopted from Matsumoto et al. (1998) [52].

The *i*Bsu1103 model was constructed after the model by Oh et al. (2007). It contains 1436 reactions that are associated with 1103 genes. This model also includes Gibbs free energy change to improve reaction reversibility prediction. Further, *i*Bsu1103 was updated with annotation data generated by the SEED project. The SEED project provides information that is continuously updated, yielding annotations that are highly accurate [53].

The BOF in *i*Bsu1103 was reconstructed based on the same BOF used in *i*YO844. However, the 61 components that make up the Oh et al. (2007) BOF were not associated with the genes involved in their synthesis. In *i*Bsu1103 the macromolecular components were categorized into seven synthesis reactions; DNA synthesis; RNA synthesis; protein synthesis; lipid content; lipoteichoic acid; cell wall synthesis and biomass synthesis. The macromolecule synthesis reactions (like the synthesis reaction described in section 2.3.5) are consumed as reactants along with 22 co-factors and ions in the biomass objective function. In doing so the BOF complexity is reduced. This also allows the synthesis of the BOF components to be associated with their respective genes [53].

The stoichiometric coefficients in *i*Bsu1103 BOF are derived from the same papers as the *i*YO844 model. The coefficients proteins, RNA, DNA, and lipids in *i*Bsu1103 are derived from the article by Dauner and Sauer (2001) [13]. To improve gene essentiality predictions, two new biomass precursors were added to the *i*Bsu1103 BOF; Coenzyme A (CoA) and Acyl-Carrier-protein (ACP). These are important carrier compounds in *i*Bsu1147. Except for the changed structure of the function by lumping the reactions of the BOF together, and the addition og CoA and ACP, the BOF remained unchanged. It was not updated with newer data on DNA-, RNA-, protein-, lipid-, cell-wall- or lipoteichoic acid composition [53].

*i*Bsu1147, the model used in this project, is based on *i*Bsu1103 but has been upgraded with additional data from KEGG and Uniprot. 99% of the reactions in *i*Bsu1103 are included in this newer model. The reactions that were excluded, were removed because they were either replaced by a new, more specifically defined reaction or because they were lumped reactions that could be represented as un-lumped in *i*Bsu1147. The biomass objective function remained unchanged and was adopted in this model as it was defined in *i*Bsu1103 [14].

2.5 Methods for analysing biomass composition and its importance in growth predictions

It is known that the environmental- and growth conditions affect the composition of biomass in bacteria [13]. In computational analyses of metabolism, cellular growth or biomass is normally an important output. In order to have successful growth predictions, it is critical to have accurate measurements of the biomass components of the organism [2, 4]. Differences in biomass composition have been shown to significantly affect model predictions, as the weight of the components in the BOF (the coefficients) influence the requirement of the precursors in the synthesis-reactions, and so on. For example, the coefficient for DNA in the BOF will affect how much is needed of the necessary nucleotides for generation of DNA. The objective function sets the proportions needed of biomass components, and this further dictates the flux distribution through other reactions in the model [4].

While the technology for generating metabolic models for predictions of growth phenotype automatically is rapidly increasing, it is still essential that the BOFs in these models are based on measured mass fractions. Rather than quantifying the macromolecular composition, coefficients for the components of the BOF are rather found in published literature [4], as with *i*Bsu1147 (Section 2.4). This in it self is not a problem, but often the adopted coefficients are based on experiments performed in other environments, or with another strain or organism, than what is being modeled. This potentially harms the accuracy of the model predictions, as the values adopted from literature might not be representative for other organisms [2, 4]. Additionally, biomass has been shown to adjust in response to environmental conditions. Utilizing data gathered from other experimental conditions than what is being modeled, is therefore not an optimal approach [2]. Accurate quantification of biomass is critical for predicting the potential production of the products under various conditions [4, 2]. Ratios between the different groups of macromolecular constituents, for example RNA, DNA and proteins, might vary and be correlated to specific properties, for example growth rate [4].

There are several methods available for quantification of macromolecular composition. In most protocols available, DNA and RNA composition is measured by spectroscopic methods, while the protein content is measured by high-performance liquid chromatography following acid hydrolysis [4]. Lipids are usually quantified gravimetrically following a methanol/chloroform extraction [54, 55].

2.5.1 Analysis DNA and RNA by spectroscopic methods

DNA can be analyzed using spectroscopic methods. However, the DNA has to be extracted from the rest of the cell tissue before it can be analyzed. The cytoplasmic and nuclear membranes have to be disrupted which is commonly achieved with liquid-liquid extraction, for example with phenol and chloroform. The cells are first treated with a lysis buffer to

disrupt the membranes, as mentioned. Then a phenol/chloroform mix is added. The lysis buffer sufficiently denaturates proteins. The organic chloroform/phenol phase in the mixture will contain all organic solvents and hydrophobic cellular components. The DNA can thereby be separated by extracting the aqueous phase. Any potential RNA in the sample can be removed by incubating the sample with RNAase [56].

Before RNA can be analyzed by spectroscopic methods, it has to be extracted [56]. For RNA, extraction can be carried out by perchloric acid extraction, as perchloric acid had been proven to extract RNA from tissue without extracting DNA. Instead the DNA is denaturated. The resulting perchloric acid solution containing the RNA can therefore by analysed by spectroscopic methods [57] .

2.5.2 Protein extraction by acid hydrolysis and analysis with HPLC

To determine the amino acid composition in biomass and protein content, the sample containing the proteins and biomass, has to be hydrolysed. This is normally done with acid. After the sample has been hydrolyzed, the amino acids can be analyzed by high performance liquid chromatography (HPLC) [58].

Separation, identification and purification of compounds in a mixture containing molecualr components of various sizes, molecular weights and affinities can be done with chromatography. It is a method of physical separation of molecules [59], based on their ability to interact with a stationary- and mobile phase [60]. The principle that chromatography is based on, is that molecules in a mixture that are applied on a surface or a fluid stationary phase, can be separated as they move from one side of the stationary phase to the other, aided by a mobile phase [61, 59]. Affinity and molecular weight are some of the characteristics that define the speed at which the molecules will move with the mobile phase. Based on the number of interactions with the stationary phase, some molecules will move quick into the mobile phase while others remain in the stationary phase longer. This separates the molecules of a mixture. Such methods are efficient for identification and separation of small molecules, for example amino acids [61, 60].

HPLC is a column chromatography method commonly used to identify and quantify compounds of a mixture. The mobile phase in HPLC flows through columns that hold the stationary phase. It has a pump that moves the mobile phase under pressure of 10-400 atm through the columns [62, 61]. The sample that is being analyzed is added to the stream of mobile phase, and is delayed because it interacts physically and chemically with the stationary phase. The time it takes for the compound being analyzed to come through the column, is called the retention time.A detector measures the retention time. The retention time varies depending on the interaction between the stationary phase, the molecules analyzed and the solvent used [62]. HPLC is a good method for separation and identification of for example amino acids [61].

2.5.3 Lipid quantification by liquid-liquid extraction

Lipid content can be quantified by combining liquid-liquid extraction with methanol, chloroform and water followed by evaporation of chloroform. Most lipids are insoluble in water, and must be extracted with organic solvents. As many lipids are bound to proteins, a mix of organic solvents should be used. For example, methanol will separate lipids from proteins, facilitating their extraction by enabling them to dissolve in chloroform. Chloroform will dissolve lipids [63]. By homogenizing the cells in water, methanol and chloroform, a biphasic system is created where the lower phase of mostly chloroform contains the lipids from the sample. The organic phase containing the lipids can be extracted [54, 55, 63], and by letting the chloroform evaporate, the amount extracted can be quantified gravimetricallys [2].

2.6 NMR analysis

Nuclear magnetic resonance (NMR) spectroscopy is a tool for analysing the structure and dynamics of molecules [64], and for identification and quantification of compounds in complex mixtures.NMR is based on the principle that energy transitions related to nuclear orientation take place in the presence of a strong magnet [65]. Most of the elements have at least one isotopes that is magnetic [64] which comes from the magnetic moment associated with nuclear spin - an angular momentum of the nuclei [66]. There are two possible orientations for the nucleus to adapt when a magnetic field surrounds the isotope, both of which are associated with a different energy state [64]. One energy-state is associated with the orientation of the nuclei being parallel to the magnetic field, while the other is anti-parallel to the magnetic field [66]. The energy state depends on the strength of interactions between the isotope nucleus and the magnetic field. If a radiofrequency (RF) pulse is applied to the magnetic field at various frequencies, the energy state of the nuclei can be measured. The RF pulse causes the nuclei to flip from a low energy orientation to a high energy orientation. The elements, or isotopes, can be distinguished from one another by changing the strength of the magnetic field, because each nuclei has its own characteristic magnetic moment. Additionally, the chemical environment of the nucleus, meaning the molecule it is in, affects the the resonance frequency of the nuclei. This means that for example, if an NMR analysis is performed on ethanol (CH_3CH_2OH), the H atoms in CH_3, CH_2 and OH will yield three separate peaks for H-atoms as they have different resonance frequencies [66, 64]. Based on these principles, NMR can be used to analyse complex mixtures as it is able to distinguish between organic molecules and metabolites [65].

NMR can be used to identify changes in the flux of metabolites being secreted in to the media or being taken up by the organism. This information can be used to identify the flux rates for metabolites, or to see if any metabolites are limiting growth. The analysis is used to quantify the metabolites in the media. If this is done at various times through the cultivation, the change in metabolite concentration over the course of growth is determined [67].

3. Methods

This project can be divided into four parts, two parts of which were performed in the laboratory; Growth experiments, biomass analyses, and two that were performed *in silico*; implementation of the experimentally derived coefficients in the model, and analysis of how the new coefficients affect model predictions. In this chapter the methods for all four parts are presented, starting with the experiments performed in the lab. hyperlinks are provided in blue to relevant to supplementary data in GitHub, such as excel sheets used for calculations and scripts. The link to the repository in its entirety is here: Supplementary data. In case of potential issues with the hyperlinks, a full URL is available in Appendix C.

The first part of this chapter presents the protocols for for the fermentations. Then the protocols for biomass analysis are presented, followed by how the experimental data was implemented in the model, and a description of how the MOMA-, FBA-, and FVA-analyses were performed. The terms nBOF and oBOF refer to "new biomass objective function" and "original biomass objective function", respectively. The carbon-source from the fermentation will be used to identify that fermentation run, as this is the only variance between runs. For example, nBOF glucose refers to the new biomass coefficient in minimal media containing glucose.

3.1 Preparation of glycerol stock solutions

In order to ensure the cells were preserved and to retain their characteristics [68], new glycerol stock cultures were prepared from a preexisting stock of *Bacillus subtilis* 168 prepared in Lennox broth (LB) medium. These were stored in -80 °C after preparation until needed. A protocol by [69] was followed for the preparation of the glycerol stocks.

This protocol was completed over three days. The first day the shake flasks and LB medium were prepared, with the components and their concentrations as described in Table 3.1. The flasks and media were sterilized by autoclavation at 121 °C for 20 minutes.

TABLE 3.1. Compounds and their concentration in LB-media for growth of *Bacillus subtilis*.

Chemical:	Concentration (g/L):	Supplier:	CAS Number:
Tryptone	10	Sigma-Aldrich	91079-40-2
Yeast Extract	5	Sgima-Aldrich	8013-01-2
Sodium Chloride	10	VWR chemicals	7647-14-5

The next day, 25 µL of *Bacillus subtilis* 168 from the preexisting glycerol stocks were inocu-

lated in 50 mL LB medium in 250mL baffled flasks. The cells were incubated overnight, for 16 hours, before the culture was harvested into a sterile 50mL Falcon tube, and centrifuged (5 min, 4000 RCF). Some of the supernatant was discarded, leaving a total volume of 16 mL together with the pellet. The pellet was resuspended by gentle pipetting. 4 mL of 18% sterile glycerol was added to reach a final concentration of 16 % in the glycerol stock, and the sample was mixed by inverting the tube gently. The glycerol-culture solution was transferred into sterile cryotubes with 1 mL in each. These were snap-frozen in liquid nitrogen before being stored at -80°C.

3.2 Preparation of stock solutions for the media

To improve preparation-time for both over-night cultures and for the fermentations, stock solutions were pre-made and stored ready to use. All stocks except the stocks of various carbon sources, the $MgSO_4$ stock, and the trace mineral solution (TMS) were sterilized by autoclavation at 121°C for 20 minutes. The carbon source stocks and $MgSO_4$ were sterilized by filtering through a 0.22 µm polyethersulfone (PES) filter. The stock solutions in Table 3.2 were prepared by dissolving the components in MQ-water.

TABLE 3.2. Three of the stock-solutions pre-made for use in defined media for *Bacillus subtilis* 168 and the concentrations of the stocks.These three stock solutions were made by dissolving the compounds in MQ-water

compound:	**chemical formula:**	**stock concentration (g/L):**	**supplier:**	**cas number:**
sodium chloride	NaCl	50	VWR	7647-14-5
ammonium chloride	Nh_4Cl	60	Merck	12125-02-9
magnesiumsulfate	$MgSO_4 \cdot H_2O$	246,5	Sigma-Aldrich	10034-99-8

3.2.1 Trace mineral solution

The trace mineral solution (TMS) was mixed by dissolving the components as listed in Table 3.3 in 500 mL of 5M Hydrochloric acid (HCl), and subsequently stored at 4 °C wrapped in aluminium foil. As the TMS was made with HCl, the solution was not further sterilized.

TABLE 3.3. The concentrations of the compounds in the trace mineral solution for minimal media, dissolved in HCl (500mL, 5M) and their supplier and CAS-number. The stock was prepared and kept ready-to use, stored at 4°C.

Chemical formula:	Concentration (g/L):	Supplier:	CAS-number:
$FeCl_2 \cdot 4\,H_2O$	7.2		13478-10-9
$ZnCl_2$	0.5		7646-85-7
$CaCL_2 \cdot 2\,H_2O$	0.5	Sigma-Aldrich	10035-04-8
$CuCL_2 \cdot 2\,H_2O$	1.0	Sigma-Aldrich	10125-13-0
$MnCL_2 \cdot 4\,H_2O$	0.2	Sigma-Aldrich	13446-34-9
$CoCL_2 \cdot 6\,H_2O$	0.05	Sigma-Aldrich	7791-13-1
$Na_2MoO_4 \cdot 2\,H_2O$	0.01	Sigma-Aldrich	10102-40-6

3.2.2 Phosphate buffer

The phosphate buffer consisted of sodium phosphate dibasic heptahydrate ($Na_2HPO_4 \cdot 7\,H_2O$, Sigma-Aldrich, 7782-77-0) and potassium phosphate monobasic (KH_2PO_4, Sigma-Aldrich, 7778-77-0). The final concentrations of the components were 112 g/L and 30 g/L respectively. After mixing the two components in MQ-water, the pH was adjusted to 7.2 using 4M sodium hydroxide (NaOH, VWR, 1310-73-2).

3.2.3 Carbon source stocks

Stocks were made of each carbon source. Depending on the solubility of the compound, the stocks had different concentrations. This was corrected for when making the defined media in order to achieve the same concentration of carbon source in each medium. The amount of MQ-water added to the media was adjusted thereafter. The various stocks and their concentrations are listed in Table 3.4. They were mixed by adding the carbon source to MQ-water while the water was stirred. When the compounds had completely dissolved, the stocks were sterilized by filtering through a 0.22 µm polyethersulfone (PES) filter.

Because the succinate stock was made using disodium succinate, corrections had to be made in order to get the same weight of carbon atoms as in the other stocks. This was done by first calculating what grams were needed for a stock with a concentration of 130 g/L. Then the molar mass of succinate was used to find the mass-percentage of succinate in disodium succinate. This was found to be 71.62%. This percentage was used to scale or up-regulate the amount of disodium succinate added to reach a concentration of 130 g/L of succinate in the stock.

TABLE 3.4. The stock solutions made of the different carbon sources and their concentrations for use in minimal media for cultivation of *B. subtilis*. They were made by dissolving the components in MQ-water under constant stirring and subsequently sterilized by filtering through a 0.22 µm polyethersulfone (PES) filter.

Carbon-source	**Solubility (g/L):**	**Stock concentration (g/L):**	**Supplier:**	**CAS:**
D-glucose	909	400	VWR	50-99-7
Glycerol	1260	400	VWR	56-81-5
Xylose	1170	300	Sigma-Aldrich	58-86-6
Mannitol	182	80	Sigma-Aldrich	69-65-8
Succinate*	350	130	Sigma-Aldrich	150-90-3

*The succinate stock was made using disodium succinate.

3.3 Preparation for Batch fermentation

100 µL from the glycerol stocks with *Bacillus subtilis* 168 were first inoculated in LB-medium directly from the glycerol stocks made as described in subsection 3.1. The LB medium was made with the components as described in Table 3.1. After inoculation the flasks were incubated at 37°C in an Infors HT Ecotron incubator shaker at 200 rpm for 8 hours. After incubation, cells from the LB medium were transferred to flasks containing defined medium for further incubation over night. The defined media consisted of the stock-solutions as described in Table 3.5, MQ-water was added to reach a final volume of 100 mL.

TABLE 3.5. Media composition for over-night cultures with volumes of 100 mL in 500 mL shake-flasks. Depending on the carbon source stock used, the amound of MQ water added to give the media its final volume varied.

Stock:	**Stock concentration (g/L):**	**Concentration in media (g/L):**	**Stock addded to media (mL):**
NaCl	50	0.5	1.0
Nh_4Cl	60	3	5.0
$MgSO_4 \cdot H_2O$	246.5	0.493	0.2
Phosphate buffer	see Table 3.2.2	-	10.0
TMS	see Table 3.3	-	0.13
C-source stock:	see Table 3.4	10	-

The over-night cultures in the defined media were inoculated in 500 mL baffled shake-flasks, with 100 mL of media. The NaCl-, Nh_4Cl-stock, and MQ-water were mixed in the shake-flask and sterilized by autoclavation at 121°C for 20 minutes. 5.0 mg L-tryptophane (Sigma-Aldrich, 73-22-3) and 200mg L-glutamic acid (Sigma-Aldrich, 56-86-0) was also added to the flasks before autoclavation reaching a concentration of 50.0 mg/L and 2.0 g/L respectively, in the media. The TMS, $MgSO_4$-, and the carbon source stock were added aseptically to the

media in a Thermo Scientific Safe 2020 sterile bench. The cells were inoculated in the defined media overnight for 12-15 hours. Because of the time required to get the fermentation started the next day could vary for a number of reasons, two flasks were used per carbon source for over-night cultures. For each carbon source there was one flask inoculated with 75 µL from the LB-flask, and one with 150 µL. Cells for the fermentation would be transferred from the flask where cells had grown enough to be in exponential phase, and not yet reached the stationary phase. The cells from the over-night culture with 75 µL of media transferred from the LB-culture was used to inoculate the bioreactors for all carbon sources.

3.4 Batch fermentation protocol and procedure

The fermentation of *Bacillus subtilis* for this project was performed using New Brunswick BioFlo®/CelliGen® 115 Benchtop 3L bioreactors. The final volume for each fermentation was 1.5 L of defined media, following the same recipe as for the overnight cultures but adjusted to the larger volume (ref. Table 3.6). The media was mixed directly in the bioreactor, except the carbon source-, TMS-, and $MgSO_4$-stock. 75 mg L-tryptophane and 3g L-glutamic acid was also added to the bioreactor before autoclavation reaching a concentration of 50.0 mg/L and 2.0 g/L respectively, in the media. The bioreactor with the media was autoclaved for 20 minutes at 121°C. The day of the fermentation, the rest of the stocks for the media were added aseptically to the bioreactor inside the sterile bench.

TABLE 3.6. Media composition of minimal media used in bioreactors for batch fermentations with *B. subtilis*. The amount of carbon source stock (C-source stock) added depended on the concentration of the stock solution. MQ-water was added to reach a final volume of 1.5L.

Stock:	**Stock concentration (g/L):**	**Concentration in media (g/L):**	**Stock addded to media (mL):**
NaCl	50	0.5	15
Nh_4Cl	60	3	75
$MgSO_4 \cdot H_2O$	246.5	0.493	3
Phosphate buffer	see Table 3.2.2	-	150
TMS	see Table 3.3	-	2
C-source stock:	see Table 3.4	10	-

The bioreactor was then connected to the Brunswick BioFlo®/CelliGen® 115 Benchtop system, and properly set up. The temperature was set to 37°C. The pH and dissolved oxygen (DO) electrode were calibrated for each fermentation. The pH electrode was first calibrated for pH 7 and then pH 4 using buffer solutions before the bioreactor was autoclaved. The DO electrode was calibrated for 0% using nitrogen gas before autoclavation, and then for 100% after the bioreactors had been set up. To do this, the agitation was set to 500, and after the temperature had reached 37°C in the media, the DO electrode was calibrated for 100%. After the DO electrode had been calibrated, a cascade was set connecting agitation

to amount of DO in the media. The cascade was set from 200 to 1000rpm. This allowed for the stirring to increase as the amount of oxygen consumed in the media increased. The increase in stirring was set to start at a threshold value of 40% Dissolved oxygen.

4M sodium chloride (NaOH, VWR, 1310-73-2) was added in a cylinder that was attached with a hose on the bottom to the pump-system of the reactor. NaOH was thereby added automatically when the system registered a pH below 7. For the fermentation with succinate as the carbon source, 4M Hydrochloric acid (HCl, Sigma-Aldrich, 7647-01-0) was used in stead of NaOH. The amount of base or acid added to the system was logged manually.

During the fermentation, the New BrunswickTM BioCommand ® Software continously logged pH-, DO-, agitation- and temperature-values. Gas analysed from the system was measured with a Thermo ScientificTM Prima BT Benchtop Process Mass Spectrometer. The software "Gasworks" tracked the concentrations of various gasses such as oxygen and carbon dioxide in the air leaving the bioreactor through the exhaust outlet, ref Figure 2.2. In addition, samples were extracted manually from the bioreactor via the sampling port for other analyses, such as dry weight measurements, NMR, and OD measurements. The time points for extraction of media for these analyses were logged manually by the hour and minute.

For the OD measurements, 1 mL of media was extracted from the system using a syringe that attached to the sampling port of the bioreactor. The OD was measured using the VWR V-1200 Spectrophotometer and was used to track the growth of the bacteria in the bioreactors. The cells would ideally be harvested from the bioreactor around OD 3. The growth curves and data that was logged during the fermentations can be found in the GitHub repository for this project, in the Fermentation data and growth rate calculations Excel document.

3.4.1 Protocol for collection of dry weight- and media samples

The dry weight measurements and media samples were gathered during the fermentation by extracting 3 mL of sample and transferring them to VWR 15 mL centrifuge tubes. The samples were centrifuged for 4 minutes at 4500 RCF. The supernatant was transferred into a syringe with a 0.22 µm polyethersulfone (PES) filter and filtered into a 3 mL Eppendorf tube. This tube was immediately stored in -20°C and was later used for media analysis by NMR. The cell pellet in the 15mL centrifuge tube was resuspended in 3 mL of MQ-water. 1 mL of this cell solution was transferred with a pipette onto a previously weighed and dried aluminium pan. For each time-point, three parallels were made. The aluminium pans were placed in a VWR DRY-Line® Prime drying oven for drying at 110°C for 3 days. These were weighed again and the weight of the aluminium pan was subtracted to get the actual weight of the cells in the sample. These measurements were used to calculate growth rates, while the media samples stored in the freezer were used to calculate the uptake- and secretion rates various compounds in the media.

The protocol for the medium analysis by NMR was based on Søgaard et al. (2018) [70]. The frozen samples were thawed out and shortly vortexed. 800 µL of sample was transferred into a 1,5mL Eppendorf tube. 80 µL 0.75% deuterium oxide (D_2O-TSP, Sigma-Aldrich, 7789-20-0) was added to the sample before it was shortly vortexed again. 600 µL was transferred from this new Eppendorf tube into a 5mm Wilmad® NMR tube. For each time point three replicates were made. After the samples were prepared they were analyzed in a 400MHz (14.6 T) Bruker NMR spectrometer by Christian Schulz as described in the paper by Simensen et al. (2021) [2].

3.4.2 Cell harvesting and washing and washing the cells

Around OD 3, the cells were harvested. 60 mL syringes were used to transfer all media in the bioreactor into several 50mL VWR centrifuge tubes. The tubes were centrifuged at 4°C for 5 minutes at 4500 RCF. The supernatant was discarded and and the cell pellet was washed by re-suspending the cells in approximately 15 mL of NaCl-solution (0.9%). The cells were resuspended by vortexing. This washing-step with NaCl was repeated, centrifuging for 4°C for 5 minutes before the second wash. After the NaCl supernatant was discarded after the second wash, the final wash was performed using approximately 20 mL of MQ-water per tube. The cells were resuspended in the MQ-water by vortexing before they were centrifuged for 10 minutes at 4°C at 4500 RCF. As the supernatant was discarded between washings and the cell pellet was re-suspended in the NaCl or MQ-water, the samples could be combined. The final biomass harvested was stored in 2-4 50mL tubes per fermentation. The tubes with the cells were stored at -80°C until further treatment.

3.4.3 Freeze drying of the cells

After all the fermentation experiments were completed the harvested cells were stored at -80°C. The next step was to freeze dry the cells using the Christ Alpha 3-4 LSCbasic freeze dryer. The 50 mL VWR tubes were covered with parafilm, and small holes were made so moisture could escape the tubes as the sample dried. The samples were then freeze dried for 3 days at -140°C in near vacuum before the now dry cells were stored again in -80°C.

3.5 The protocols for biomass composition analyses

3.5.1 DNA

The DNA was extracted following a protocol described by Wright et al. (2017) [71] using individual enzymes and a phenol/chloroform separation method. For each carbon source,

three parallels with 10 mg of cells was weighed in 2 mL Eppendorf tubes. The method involved using a lysis buffer to lyse the cells. This buffer was prepared by mixing the chemicals and a TE buffer as listed in table 3.7. The TE buffer was made with mixing 0.5 mL Tris EDTA buffer solution (pH 8.0), that contains 10mM Tris-Cl and 1mM EDTA, with MQ-water (49.5 mL). 600 µL of the lysis buffer was added to the cells and the samples were incubated in a VWR COOL THERMAL SHAKER TOUCH 230V for 30 minutes at 55°C. After the incubation and the samples had cooled to room temperature, 600 µL phenol/chloroform mix (1:1) (Sigma-Aldrich, 136112-00-0) was added. The sample was mixed by inverting the tubes and after the phases had mixed completely it was centrifuged at 12000 RCF for 5 minutes. The upper aqueous layer was transferred to a new tube. This phenol/chloroform washing step was repeated twice, transferring the upper aqueous phase to a new 2mL Eppendorf tube each time. Following the last round of washing, 600 µL chloroform (Sigma-Aldrich, 67-66-3) was added to the sample, in order to remove any phenol left. The tubes were mixed by inverting them, and then centrifuged at 12000 RCF for 5 minutes. Then, 40 µL of 3M sodium acetate (Sigma-Aldrich, 126-96-5) and 1 mL 99% ice-cold ethanol (VWR, 64-17-5) was added and the samples were incubated overnight at -20°C.

TABLE 3.7. The components in the lysisbuffer used for analysis lysis of *B. subtilis* as an initial step in measuring the DNA-content of its biomass.

Stock:	**mL in lysis buffer**	**Supplier**	**CAS:**
TE buffer	9.34	Sigma-Aldrich	*SLCG3955
10% SDS	0.6	-	-
Proteinase K (20mg/mL)	0.06	MERCK	39450-01-6

*CAS number not available. Lot number used instead.

After incubation, the cells were centrifuged for 15 minutes at 12000 RCF at 4°C. The supernatant was discarded and the DNA pellet was washed with 1 mL 70% EtOH. After centrifuging the sample again for 2 minutes at 12 000 RCF, the supernatant was again discarded and the DNA pellet was dried using a VWR COOL THERMAL SHAKER TOUCH 230V at 37°C. After the samples had dried, the DNA pellet was resuspended in 50 µL TE-buffer. 1 µL of RNase A (Thermo Scientific, RNase A, Lotnr:01001341) was added and the samples were incubated again in the Thermoshaker for 15 minutes at 37°C. After this incubation the samples were analyzed using Nanodrop with the setting "double stranded DNA", measuring the DNA concentration. The experimental data and calculations for the DNA content in the biomass composition can be viewed in the "DNA" sheet in the biomass analyses Excel document.

3.5.2 RNA

The method for RNA quantification followed the protocol described by Benthin, Nielsen and Villadsen (1991) [72]. Three parallels of cells from each growth-condition of around 10 mg each was prepared in 15 mL VWR polypropylene centrifuge tubes. In order to degrade

the cell walls, the cells were washed three times with 3 mL 0.7M perchloric acid ($HClO_4$, Sigma-Aldrich, 7601-90-3). In between washes, the tubes were centrifuged at 4500 RCF for 10 minutes at 4°C, and the supernatant was discarded in between washes. After the third wash, the pellet was re-suspended in 3 mL 0.3M potassium hydroxide (KOH, Sigma-Aldrich, 1310-58-3). The cells were then incubated for 1 hour at 800 rpm at 37°C in the VWR COOL THERMAL SHAKER TOUCH 230V. After the incubation and the samples had cooled down, 1 mL of 3M $HClO_4$ was added. The samples were centrifuged at 4500 RCF for 10 minutes at 4°C, and the supernatant was transferred into a new 50 mL VWR polypropylene centrifuge tube. The pellet was washed two more times, with 4 mL of 0.5M $HClO_4$, and after each round in the centrifuge the supernatant was added to the 50 mL tube. 3 mL 0.5M $HClO_4$ (3 mL, 0.5M) was added to this 50mL tube with extracts in order to get a final volume of 15 mL and was subsequently centrifuged again 4500 RCF for 10 minutes at 4°C once more to remove any precipitates of KOH. The samples were analyzed using NanoDrop using the setting for "single-stranded RNA". NanoDrop gives the concentration of RNA in the sample. The raw data and calculations for RNA content of the biomass can be viewed in the "RNA" sheet in the biomass analyses Excel document.

3.5.3 Proteins

The quantification of the content of proteins in the cells was measured by amino acid derivatization through HPLC after acid hydrolysation, following the protocol as described by Noble et al. [58]. About 10 mg of cells from each growth condition were weighed out in Schott flasks. 500 µL of 6M HCl was added and the samples were weighed before they were put in the oven to boil at 105°C for 24 hours. After the boiling, and the samples had cooled down they were again weighed, before they were neutralized with 500 µL of 6M NaOH. Then three dilutions were made of each sample - 1:50, 1:100 and 1:250 - in HPLC glass vials and stored in -20°C until further analysis.

These dilutions were analyzed by HPLC by Siri Stavrum. The method followed a protocol as described by Lindroth and Mopper in 1979 [73]. The pre-column derivatization was performed using OPA (o-Phthaldialdehyde Reagent Solution, Sigma P0532). The column used was a Waters Nova-Pac C18 (4µm), with a dionex RF2000 flourescence detector. The software used for sample analysis was Chromeleon^{TM} Chromatography Data system. Based on the resulting concentrations, the dilution that gave a concentration of an amino acid closest to 2.5 µmol/L would be used for further analysis. This is important to ensure the concentrations measured are within the upper boundary for linearity from the HPLC analysis.

Due to partial or complete degradation during hydrolysis, not all twenty amino acids could be measured following this protocol, for example methionine and cysteine. The concentrations of these amino acids were predicted by linear regression, as described by [2]. An example of a the linear regression for the amino acid distribution in the media containing glucose is illustrated in Figure 3.1. The linear regression was preformed using the concentrations of the amino acids that could be measured, and their prevalence in protein-coding genes. The

amino acid prevalence and distribution were determined using the "Amino acid distribution" Python script by written by Vetle Simensen.

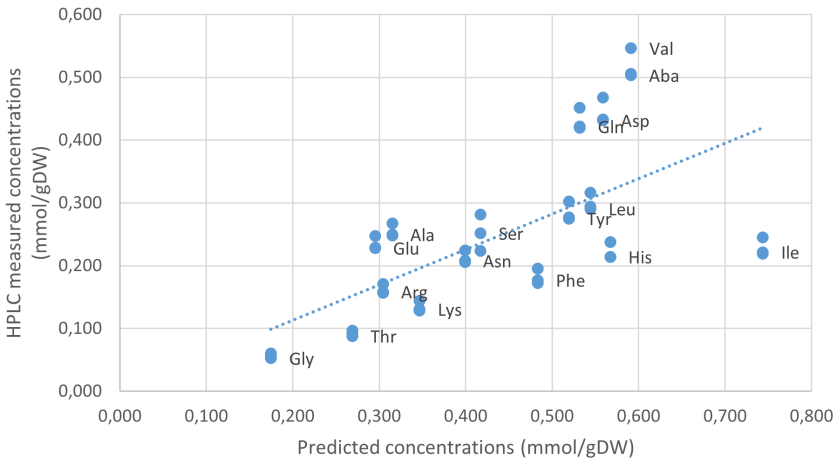


FIGURE 3.1. To correct for partial or complete degradation of Arg, Met, Cys, Pro, and Trp during amino acid derivatization protocol, a linear regression was performed on the measured concentrations of the other amino acids against their relevant prevalence calculated from protein coding genes in *B. subtilis*. The equation of the linear regression was used to predict the concentrations of the remaining amino acids by their prevalence.

Another correction had to made for amino acids with overlapping retention times. The concentrations of these amino acids were predicted as well, using the same protocol. Glutamine is deaminated to glutamate, and asparagine to aspartate. To get the correct concentrations, these were predicted as well [2]. The Excel sheets containing the calculations and raw data from the protein analyses for each carbon source can be viewed in the HPLC calculations folder in the GitHub repository.

Due to technical problems, the HPLC results were quite delayed. As a safety measure to ensure data on protein analysis, a Bradford protein assay was performed to get the quantity of proteins. As the HPLC results eventually were received in time, this data and the description of the Bradford protein assay is described in appendix A.

3.5.4 Lipids

A chloroform/methanol extraction protocol was followed to gravimetrically quantify the amount of lipids [55, 54]. Three parallels of 40 mg of cells were weighed out for each growth condition in 2 mL Eppendorf tubes, as well as 6 blank tubes per round performed of the protocol. The cells were rehydrated by adding 0.15 mL MQ-water and vortexing them shortly at low rpm. After re-hydration, 1.4 mm zirconium beads (0,5 g, Precellys Bulk bead for 500 preps, lotnr: 210622-830) and 0.4 mL methanol (VWR chemicals, 67-56-1) was added to the

samples. The samples were then homogenized at 5500 rpm at 20 second intervals for 2 cycles. The samples stored on ice between the cycles. After this step, 0.8 mL of chloroform (67-66-3) was added, and the samples were vortexed for 20 minutes. Then 100 µL of MQ-water was added to the samples and they were vortexed again for 10 minutes before being centrifuged for 4 minutes using a table centrifuge at 12000 RCF. After the centrifuge, the organic phase was collected from the samples and transferred to a dark MS vial using a syringe. These vials and their lids had been weighed before the organic phase was collected in them. This washing step of the samples was repeated, but now with 0.6 mL chloroform before they were mixed by vortex for 10 minutes, and then centrifuged again for 4 minutes at 12000 RCF. The organic phase was extracted using a syringe and collected in the same MS vial. After collecting the organic phase twice, the dark MS vials were weighed, and then put back in the fume hood without their lids in order for the chloroform to evaporate. After 3 and 4 days the samples were weighed. To get the amount of lipids in the sample, the average weight of the blank samples were subtracted from each of the weights of the samples containing the cells. The experimental data and calculations for lipid content in the biomass can be viewed in the "Lipid" sheet in the Biomass analyses Excel document.

3.6 Determination of experimentally measured growth rates

The growth rates were calculated based on the dry weight measurements. The natural logarithm of the dry weight measurements were plotted against time. By preforming a linear regression in Excel, the slope of the line equals the growth rate [32]. The experimental data and calculations of growth rates for each carbon source are available in the Fermentation data and growth rate calculations Excel document. The standard error for the growth rates was found by performing a linear regression in RStudio. That script (RStudio - Linear regression growth rates) is available in the GitHub repository. The equations from the linear regressions of the dry-weight measurements was later used to predict dry-weights for other calculations, such as the media analyses and gas uptake- and secretion rates.

3.7 Uptake- and secretion rates

3.7.1 Uptake- and secretion from the media

The previously prepared samples for media analysis with NMR as described in section 3.4.1 were analyzed and the resulting concentrations were used to calculate uptake- and secretion-rates of various compounds in the media. The measured concentrations for each compound were log transformed and plotted against time. A linear regression was performed for the

measurements taken during exponential phase. The equation for the regression line was used to predict concentrations of the compounds. These predicted concentrations were used, together with predicted dry-weight measurements in the same time-points to calculate specific uptake- and secretion rates using formula 3.1, which is based on the protocol used by Bartosova et al. (2021) [74]. An average of the last three measurements before harvest gave the final uptake- and secretion rates. The calculations and experimental data is in the NMR raw data and calculations Excel document.

$$\frac{\frac{mmol/L_{t_0} - mmol/L_{t_1}}{t_1 - t_0}}{\frac{\Sigma_{T_0}^{T_1} gCDW/L}{2}} = \frac{mmol}{g \times h} \tag{3.1}$$

3.7.2 O_2-uptake and CO_2-secretion

During the fermentation, the off-gas was measured by Thermo ScientificTM Prima BT Benchtop Process Mass Spectrometer. The mass spectrometer analyzed the constituents of the gas leaving the bioreactor and gave the values of percentage out of the air. The percent-values were used to calculate mmol/h of gas for O_2 and CO_2, which was used to calculate the specific uptake- and secretion rates of the two gasses. The calculations and experimental data are in the Gas - Uptake and secretion Excel document.

By predicting the dry weight of cells at the same time-points as there are gas measurements, this could be used to calculate the specific uptake-rate per time point using the formulas below 3.2, 3.3, 3.4. For each time point an average of the five measurements leading up to that time point was used for further calculations. Because the amount of gas measured is provided as a percentage, the flow-rate (mL/min) of gas entering the system was used to find the liters per minute of O_2 and CO_2, as shown in equation 3.2. This was finally divided by the predicted total content of cellular bacteria in the bioreactor at that time. In order to find the content of CO_2 and O_2 in the air entering the reactor, the average was taken of 15 measurements of gas flowing through the system before inoculation (sparge air). The average uptake- and secretion rates were then calculated by taking the average of the final 10 measurements before harvesting.

$$\frac{recorded(\%) - sparge(\%)}{100} \times Airflow(ml/min) = \frac{ml}{min} \tag{3.2}$$

$$\frac{ml}{min} \times N = \frac{mmol}{min} \tag{3.3}$$

The N was found using the equation below, 3.4. This was calculated for O_2 and CO_2 and was used to find the mmol/min of gas, the sparge air, flowing into the system.

$$\frac{mmol}{L} = N \tag{3.4}$$

3.8 Construction of the new condition-specific biomass objective functions

The construction of the new BOFs (nBOFs) started with sorting the substrates of the original BOF (oBOF) reaction in seven groups; Lipteichoic acid content, cell wall composition, cofactors and ions, protein, DNA, RNA, lipids. This was done to ease the ability to compare the nBOFs to each other, and to ease the process of running the model with the nBOFs. All the twenty-seven substrates in the original *i*Bsu1147 BOF were sorted in one of these seven groups, except for ATP, water, orthophosphate and Apo-acyl-carrier-protein.

A critical feature of Genome Scale metabolic models to correctly predict growth yields is that the biomass produced has a molecular weight of 1g/mmol [50]. The first part of constructing the BOFs for the various media compositions was first to find the preceding reaction of the macromolecular components that were measured (DNA, RNA, proteins, lipids) and scale these reactions so that the g/gDW (%) measured of each component in these experiments could be used as coefficients in the nBOF. The substrates that were lumped into the co-factor group were scaled as well.

To scale the synthesis reaction, the chemical formulas for each metabolite as defined in the model were used to calculate the molecular weight for each substrate. The molecular weight for each compound was used with the coefficient from the reaction to find the gram per gram dry weight (g/gDW). The sum of the g/gDW of all substrates in the synthesis reaction was calculated. For most reactions the sum was somewhere around 0.9. The sum of the g/gDW was used to scale the coefficient of the substrate in the reaction (mmol/gDW). To check if the substrates were scaled correctly, the molecular weight was again used to calculate the g/gDW from the scaled coefficients, and the sum of of the scaled g/gDW was again calculated to check that it would be 1. For the cofactor function, the new coefficient for that lumped reaction was 0.044, which was the sum of g/gDW before the coefficients were scaled. The Excel document, BOF and synthesis reactions, contains all calculations of these data.

The new coefficients for the synthesis reactions were loaded into the model using Python. The model was loaded and the reactions were added to the model. The old synthesis reactions were deleted. Each carbon source has its individual script where this is performed. All Python scripts are available in the Functions and reactions GitHub folder. The script for the model in glucose medium contains explanations in the code. The functions to create reactions from the experimental data were modified from the functions script by Linn Sandvik. The new synthesis reactions were all labeled with "new" behind the reaction ID. Additionally, the model was updated with the media composition for each of the fermentation analyses. Excel

documents with the coefficients for the nBOFs, BOF-new, is available in the Gihub repository. The Excel document with the new coefficients for the synthesis reactions, Macromolecule-synthesis, is available there as well.

3.8.1 Network analyses with the newly constructed BOFs

Three different types of analyses were performed on the network comparing the new BOFs to the original *i*Bsu1147 BOF in the same media compositions. In total 10 models were analysed, of which 5 had the original BOF (oBOF), but the measured uptake- and secretion rates that were found experimentally or by MOMA. The last 5 models each had a newly constructed, experimentally derived BOF (nBOFs), with the measured uptake-rates. The analyses were performed with nBOF and oBOF in minimal media containing their respective carbon source.

To update the model with the media composition, a list of exchange reactions that correlated to components in the media was imported. The model was then updated with new flux rates based on this list. If an exchange ID was in the media, the model was updated with flux range -1000 as the lower bound, and 1000 as the upper bound. This means that the compound corresponding to that exchange reaction can be taken up by the bacterium, and secreted into the media. If and exchange ID from the model was not in the list of media components, the model was updated with flux range 0 as the lower bound and 1000 as the upper bound. This means the bacterium could not take up that compound, but could secrete it. An overview of the exchange reactions and their corresponding media component per carbon source is available in the Medium Excel document.

3.8.2 Minimization of metabolic adjustment analysis for adjusting measured fluxes

Because the model was not able to run with both the experimental growth rate and uptake- and secretion rates for glucose and mannitol, a MOMA-analysis was performed to find flux-values that were compatible in the model with the measured growth rates.

The analysis was performed in Python, using the ReFramed library created by Daniel Machado [75]. The growth-rates for glucose and mannitol were set, and the nBOFs were implemented with their synthesis reactions. By adjusting the upper boundary to the measured growth-rate plus 2 $\times$ the standard deviation, and the lower boundary to the measured growth-rate minus 2 $\times$ the standard deviation. The model was then reloaded as a cb model and the MOMA analysis was performed. The experimentally measured fluxes were used as reference-values. The feasible resulting solution from the analysis will thereby be the closest possible to the experimentally measured values while maintaining the growth-rate within the locked upper and lower bounds. The scripts for the MOMA-analysis is in the "Functions and

reactions" files for glucose and mannitol in the supplementary data repository in GitHub.

MOMA results in a list of feasible fluxes for the reactions. These fluxes were used for the reactions that would have been updated with experimental values; O_2, carbon source, glutamate, tryptophane uptake and CO_2 and acetate secretion. Glucose and mannitol nBOFs were still used as derived experimentally, only the exchange reactions and growth rate was updated with the fluxes suggested from MOMA.

3.8.3 Parsimonious Flux balance analyses with experimental data and original data

The FBA was performed using the parsimonious FBA function from the COBRA toolbox in Python. This was performed on the model with the nBOF and oBOF for each of the experimental conditions. The models were updated with the fluxes that had been experimentally measured, or that were updated from the MOMA analysis. The script for this analysis, pFBA-script takes in the models, performs the pFBA analysis and returns a plot of the results.

The reason a pFBA was performed and not FBA is because the results were to be compared between nBOFs and oBOFs in their respective media composition. In pFBA, the solution is found based on optimizing flux through the objective function, then minimize the fluxes through the other reactions of the model [76]. This means that the presented solution from pFBA is one of the solutions from FBA, but between the nBOF and oBOF, and between conditions, the solution from the analysis is made on the same basis, and therefore results can be compared.

The resulting flux distributions were visualized in scatter plots. One plot was made for each media composition, with the flux values from the nBOF and oBOF. The plots then visualize a the change in flux distribution between the model with the oBOF and the nBOF in the same media composition to illustrate how the nBOF affects model predictions. Further, the reaction IDs with large variations in flux between the nBOF and oBOF were identified using KEGG REACTION [77, 78, 79]. This helped understand how the new BOF coefficients altered the flux distribution through the model.

3.8.4 Flux variance analysis with experimental data and original data

Following the FBA, a FVA was performed on the models for nBOF and oBOF in the different media compositions. The FVA function from the COBRA toolbox was used on the model. The script used, FVA-script, takes in the nBOF and oBOF and performs the FVA analysis, and returns a plot with the resulting range and mid-values. This analysis returns the reaction

IDs and their corresponding feasible minimum flux and maximum flux. The optimality constraint in this analysis was set to 95%.

Following this analysis the range and mid-value were calculated for each reaction. These values were then illustrated in scatter plots to visualize the change between the model with the nBOF and the oBOF. The calculations were performed in Python, and are available in the FVA-script.

4. Results and analyses

As previously mentioned in the introduction to chapter two, the project can be divided into four parts; Growth experiments, biomass analyses, implementation of the experimentally derived coefficients in the model, and analysis of how the new coefficients affect model predictions. This chapter is initiated with the results from a preliminary-, exploration phase of the project. The second part is the results from the batch fermentations that were performed with *B. subtilis* in minimal media. Each fermentation is referred to by the carbon source available in the media, as all other components in the media are the same. Then the results from the biomass analyses are presented and discussed, followed by the measured uptake- and secretion-rates. The chapter then moves on discussing the new biomass objective functions (nBOFs). The nBOFs from all five carbon sources are compared to the original BOF (oBOF) in the same media composition with the experimentally measured uptake- and secretionrates. The final section of this chapter presents the results from FBA, FVA and MOMA analyses that were performed to explore the effects of the nBOFs on model predictions.

The fermentation experiments and biomass analyses were performed in cooperation with another masters student Linn Sandvik. Both individual projects included 4 to 5 batch fermentation, and we decided to perform them together. A significant amount of work is associated with the preparation for, and performance of batch fermentations. Both projects also involved biomass analyses. By cooperating, there was the opportunity to exchange ideas and deliberate while still independently planning and carrying out the individual projects.

4.1 Preliminary research

4.1.1 Shake flask experiments to confirm growth with various carbon sources

Initial shake-flask experiments were performed to evaluate potential carbon-sources for batch-cultivations of *B. subtilis*. The objective was to identify carbon-sources that *B. subtilis* is able to metabolize effectively. The following carbon sources were tested in the shake-flask experiments: Glucose, glycerol, mannitol, xylose, succinate, acetate, and fructose. One flask was prepared with every media-component without addition of a carbon source to confirm that there would not be any growth without it. The media prepared followed the same recipe as described in Table 3.5, and the flasks were inoculated with 100µL of *B. subtilis* from the glycerol stocks and incubated at 200 rpm at 37° C.

However, there was little to no growth in the flasks. This experiment could not be used to draw any conclusion on what carbon sources to use, because the cells did not grow in any of the media composition. One of the trouble-shooting steps performed was to measure the pH of the media in the flasks, and it was measured to be 2.4 after a few hours of cultivation. The pH measured for pure media was 3.3. One of the causes for the low pH uncovered was the adjustment that had been made to properly dissolve all the components of the TMS. The media composition described in Table 3.5 has been modified from the original recipe. In the original recipe, 2mL of TMS is used in 100mL media. However, in that case, the TMS was made using MQ-water. The TMS used for these experiments was made dissolving the components in 5M HCl, after discovering that the components did not properly dissolve in water. Adding 2mL of the TMS dissolved in 5M HCl significantly affected the pH, which was supposed to be around 7.

To find a solution to the pH problems, a new shakeflask experiment was performed with minimal media containing glucose. This time the media was made with different volumes of TMS added to find a concentration that would not affect the pH, and still add enough minerals for the bacteria to grow. The amounts of TMS added were 130µL and 200µL. There was growth in both flasks, but more growth in the flask containing 130µL. The final volume of TMS to add in a 100mL shakeflask was therefore adjusted in the recipe to be 130µL.

After resolving the TMS- and pH-issues, the original shake flask experiment with growth on the various carbon-sources was performed again, with the media composition as described in Table 3.5. The carbon sources that were selected for further experiments were those that exhibited interesting variations in growth; glucose, glycerol, mannitol, xylose and succinate.

4.1.2 Minimizing the duration of the lag-phase in batch fermentation

A few practice rounds of fermentation were performed before the actual experiments, in order to get familiar with the protocol and the bacterium. In the first few rounds of fermentation the lag-phase was about 4 hours long for *B. subtilis* cultivated in defined media containing glucose. To prepare for this fermentation, the bacterium was inoculated directly from the glycerol stocks in defined media the day before fermentation. The overnight culture was incubated in 100mL media in a 500 mL shake flask for about 15-17 hours.

The same overnight culture procedure was performed one more time. But this flask was used to try to identify why the lag-phase had been so long in the bioreactor. The OD was measured and showed that the cells reached stationary phase in the over-night culture. They would still grow when transferred to the fermentor, but it would take the cells some time to start growing again. In order to cut the lag phase, the over-night culture protocol as described in the methods section 3.3 was attempted for *B. subtilis* 168, and a new round of fermentation was performed. This time, the lag-phase was around 2 hours, a significant reduction compared to the approximate 4 hours for the first fermentation (Fig. 4.1).

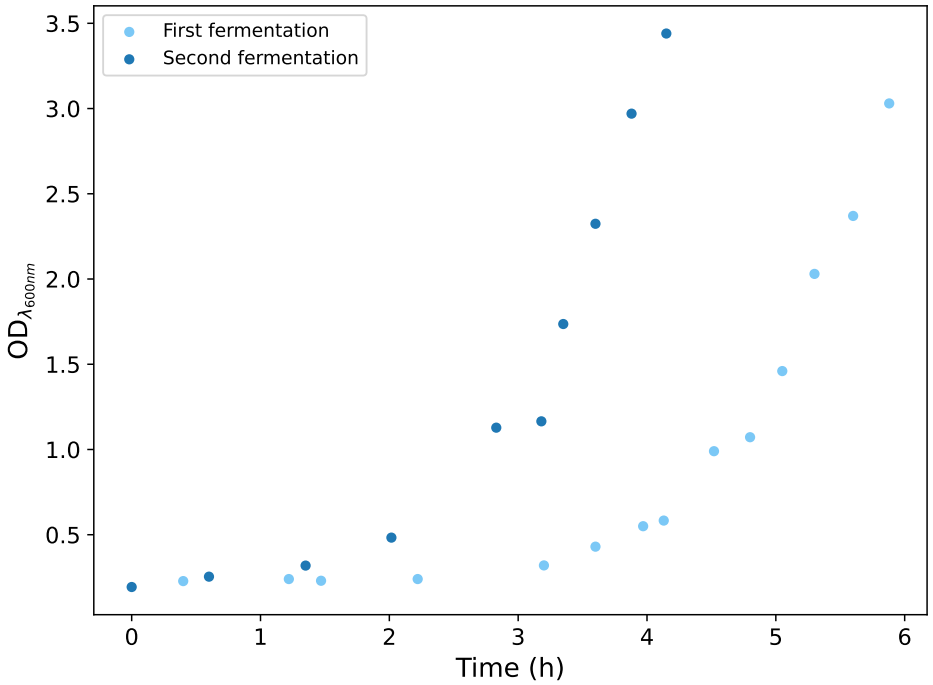


FIGURE 4.1. Improvements were made to the protocols of making starter cultures for the batch fermentation experiments with *B. subtilis* in minimal media. By changing the protocol, the duration of the lag-phase was reduced. This plot shows the reduction of time in the lag-phase with OD-measurements from both fermentation experiments.

Because this preliminary research only was performed on *B. subtilis* in glucose, another measure to reduce the risk of the bacteria reaching stationary in the flasks with the other carbon-sources was implemented. Two over-night cultures in defined media were made for each fermentation. In one 75µL of culture was transferred from the LB-media, in the other 150 µL was transferred. That way, the OD could be measured in each flask the morning of the fermentation, and the flask with an OD closest to exponential growth would be used to inoculate the fermentors.

By implementation of these steps described here; inoculation in LB medium and then minimal medium over night, the lag-phase was reduced. The lag phase is the phase where the metabolism is reorganized to adjust to a new environment [31, 33]. One of the factors that affect the lag phase duration is how much the environment changes from the media they were in compared to the media they are transferred to [33]. By initiating the cell recovery phase from glycerol stock to shake-flask in LB-medium, the cells were allowed to re-initiate growth in the same type of environment as was used for cultivating the glycerol stocks. This requires little adjustment, and reduces the lag-phase of re-initiating growth.

After growth in LB-medium, the cells were transferred to minimal medium, a more nutrient limited environment. This transfer requires the cells to adapt their metabolism to the nutrients available. By using two inoculum sizes, the risk of transferring cells in stationary phase to the fermentor was reduced. The OD could be measured in the flasks, and the flasks with cells still in exponential phase could be used. Stationary phase is characterized by nutrient depletion and a decrease in growth rate [31]. In this phase the cells re-adjust their metabolism in preparation for starvation. When starved cells are transferred to a new environment where resources are available, it has been shown that they do not immediately resume growth. Instead, they show a delay in re-initiating growth at population level [80]. Therefore, it would reduce the lag-time if the cells had not reached stationary which is why two inoculum sizes were used for the flasks. A higher cell count will metabolize the available substrates faster, reaching stationary phase quicker than a media inoculated with fewer cells. Further, by allowing the re-adjustment of metabolism to happen over night and transferring already rapidly growing cells into to the same media composition in the bioreactor, the lag-phase is reduced.

4.2 Fermentation analyses

4.2.1 ODs at harvest of *B. subtilis* from bioreactors

The cells would ideally be harvested during exponential growth, when the cells are still exponentially increasing, just before the growth starts to decrease. This is important because the cells in exponential growth are said to be growing in a "steady-state", meaning they are harvested before they are restricted by nutrient depletion [81]. The steady-state is achieved in the exponential phase, because the total biomass in the bioreactor increases exponentially with time. A characteristic of this growth phase is that the concentration of macromolecules in the bacteria is constant and increases proportionally with time [82]. Because there is no addition of fresh media in batch fermentation, the exponential growth phase is the only phase where the bacteria can achieve the highest possible growth rate, given the nutrients available [83].

As previously mentioned in subsection 2.3.2, predictions in GEMs (for example FBA) are performed with the assumption that the system exhibits steady-state properties [84, 42]. As the results from the biomass analyses of the harvested bacteria will be implemented in the GEM *i*Bsu1147, it is important that the cells are harvested at "steady-state" as this is a central assumption in the model.

In order to get a good estimate of an approximate OD where the cells are growing exponentially, a fermentation was performed letting the cells reached stationary phase. From looking at the growth curve in appendix B (Figure 3) from this fermentations, one can see that OD 3 is the point where the growth transitions from exponentially increasing to growth decreasing.

The actual ODs of harvest for each carbon source is listed in Table 4.1. Some of the ODs are a bit higher than this ideal value. This might have affected the state of the bacteria at harvest. If they had started decreasing in growth this might for example affect the growth rate and biomass composition. The effects of the growth phase on the biomass composition is further discussed in chapter 5.

TABLE 4.1. The ODs of which *Bacillus subtilis* was harvested from the bioreactor from each round of batch-fermentation in defined media containing different carbon sources.

Carbon-source	**Harvest OD:**
Glucose	3.4
Mannitol	3.7
Glycerol	3.4
Xylose	3.0
Succinate	3.5

4.2.2 The results and analysis of the experimentally derived growth rates

Growth rates were calculated for *B. subtilis* cultivated in the five different carbon sources. In order to find the growth rates, the three parallels of measurements from each time-point were log-transformed. A linear regression analysis was preformed, and the growth rate was then determined by calculating the slope of the line from the linear regression. A Table with the growth rates is in Appendix C (ref. Table 1), and they are illustrated in Figure 4.2. As illustrated, there are differences in the growth rates achieved by *B. subtilis* with different carbon sources available. An ANOVA analysis of the results gave a P-value < 0.05 meaning there is significant variance. By performing a post hoc analysis, it was revealed that it is only between glycerol and glucose, mannitol and glucose, and glycerol and succinate that there is no significant variance. The experimental data and the scripts for the statistical analyses available in appendix C.

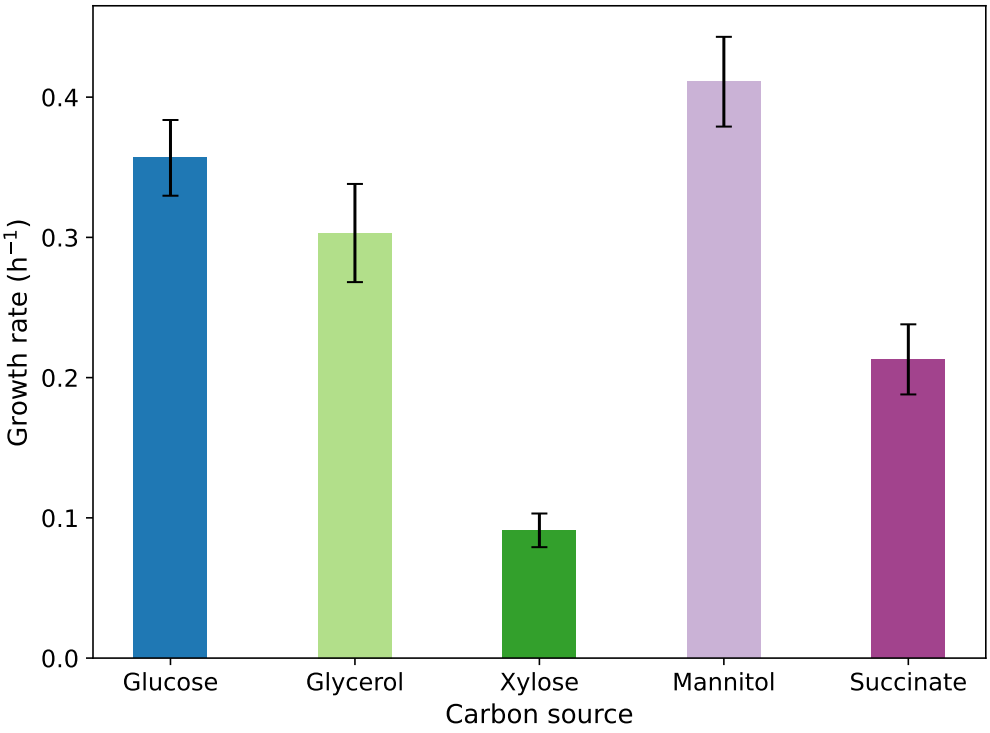


FIGURE 4.2. Estimated specific growth rates for *Bacillus subtilis* cultivated by batch fermentation in minimal media containing one available carbon source per round of fermentation. Error bars denote one standard deviation of uncertainty.

The ability for a cell to grow fast on a given substrate is connected to its ability to adapt its carbon- and energy metabolism to the nutrient available [26]. Utilization of different sources of carbon is associated with expression of different sets of genes to synthesise the specific transporters, regulators and enzymes needed to metabolize the given compound [85]. Additionally, different carbon sources have different energy content, meaning that some contribute more energy to the cell than others [86]. The resulting growth rates found for *B. subtilis* in the various carbon sources indicate that the bacteria is able to better utilize glucose, glycerol and mannitol, compared to xylose and succinate.

B. subtilis can reach high growth rates on glycerol, as it can be rapidly imported via a selective facilitator [87]. Glucose and mannitol are taken up by efficient, substrate-specific transporters and enter the metabolism early in glycolysis [19, 24]. The efficient uptake is part of the reason growth on these substrates is faster. The growth rate for the media containing xylose is much lower than what was calculated for *B. subtilis* in the media containing the other carbon sources. This can be explained by the fact that *B. subtilis* does not have a xylose-specific transporter, and therefore cannot efficiently transport xylose rapidly into the

cell to metabolize it [27].

The growth rate for *B. subtilis* in succinate is modest as well. Succinate is metabolized in the TCA. Utilization of succinate requires the cells to reverse the flow of metabolites in their metabolism, the flux, from glycolysis to gluconeogenesis in order to synthesise precursors needed for the metabolism [26, 19]. Waschina et al. (2016) [86] investigated the cost of amino acid production on different carbon sources and characterized three groups of amino acids depending on the type of carbon substrate their production is associated with. They found that there is a cost trade-off between these groups. The amino acids that are metabolically cheaper to produce when glycolytic substrates are utilized, are produced at a higher metabolic cost when carbon sources utilized are gluconeogenetic (such as succinate), and *vice versa*. Additionally, there are several reactions in gluconeogenesis that cause additional metabolic costs [86]. The growth rate measured for growth on succinate, indicate that the metabolic cost of succinate utilization might be higher compared to the other carbon sources used.

When taking dry-weight samples, 3 ml of media was extracted and centrifuged, and the pellet was resuspended in 3 mL of MQ-water. 1 mL was transferred to each aluminium-pan for dry-weight samples. This protocol had some room for error. In some cases it was clear that the pellet had not been properly resuspended and there was a high standard deviation between replicates from the same time-point. A better protocol would likely have been to do three replicates with 1 mL tubes. Then, rather than a pellet from 3mL of sample being resuspended in 3mL of MQ-water, three pellets would be resuspended in 1 mL of water, avoiding the issue of improper re-suspension or uneven distribution of the cells on the aluminium pans. However, the standard deviation between measurements were not a problem for all fermentation rounds, indicating that error of execution might have been the biggest challenge, not necessarily the protocol itself.

4.3 Experimental determination of biomass composition in different media compositions

It is known that the environmental- and growth conditions affect the composition of biomass in bacteria [13]. Various analyses were executed to measure the biomass of *B. subtilis* grown in environments containing different carbon sources to see how the composition might vary. The overall resulting g/gDW with corresponding standard deviations is illustrated in Figure 4.3. The experimental values for growth rate, DNA, RNA, proteins and lipids are listed in tables in Appendix C as well, along the results from statistical analyses performed. All ANOVA and post hoc analyses were performed in Python, and can be in the statistical analyses script.

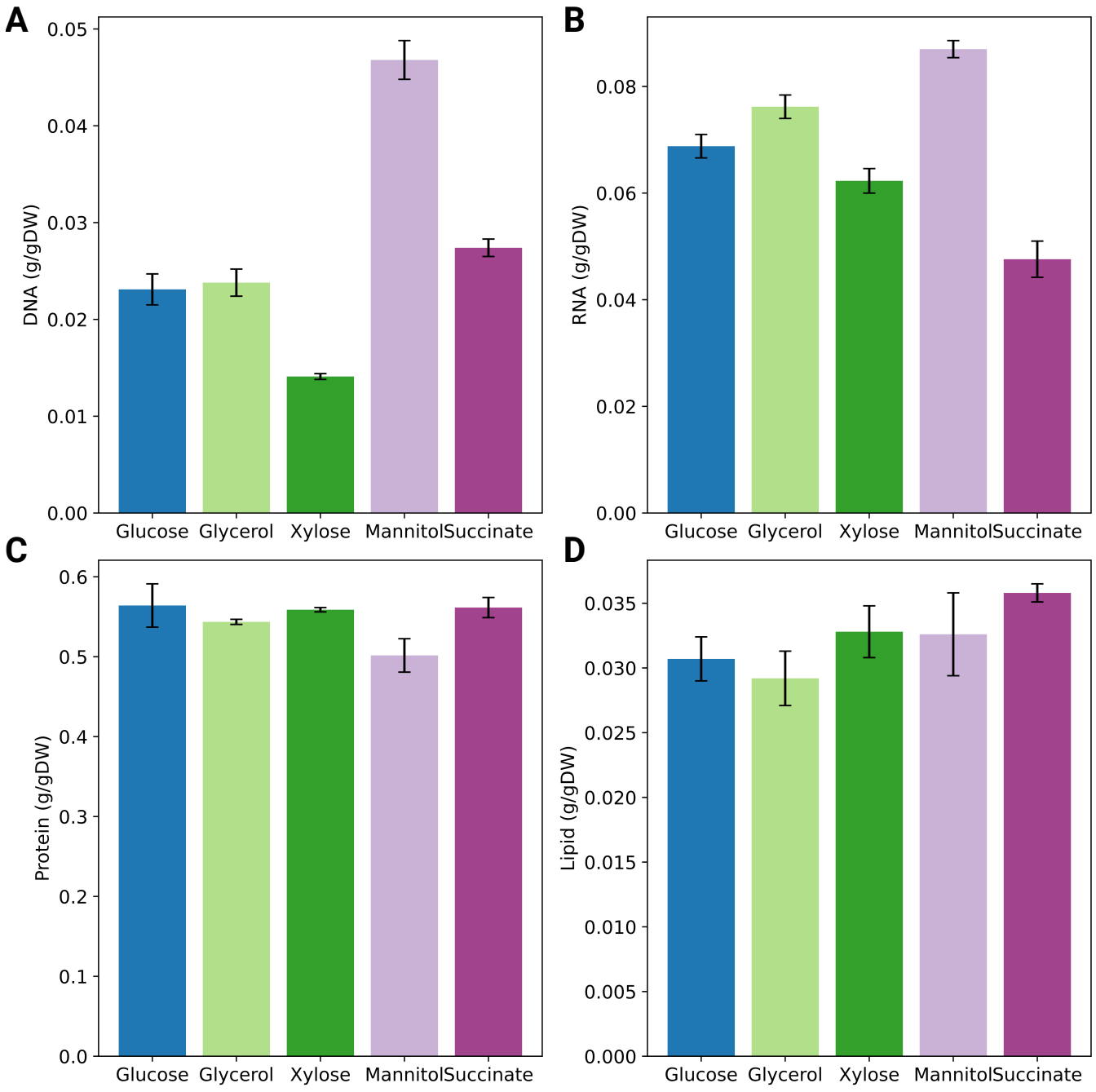


FIGURE 4.3. The amount of A) DNA, B) RNA, C) proteins and D) lipids in g/gDW measured in *Bacillus subtilis* in the exponential growth phase of batch fermentations performed in minimal media containing different carbon sources.

4.3.1 DNA content

The DNA content was measured using NanoDrop following a phenol/chloroform liquid-liquid separation protocol. The contents measured are presented in the Figure 4.3A. As can be seen in the Figure, the amount of DNA found in *B. subtilis* growing on mannitol as the sole source of carbon is the highest. An ANOVA analysis gave a P-value < 0.001, and post hoc analysis showed that it was only DNA content measured from the glucose and glycerol fermentations that were not significantly different from each other. The ANOVA analyses indicate that there the amount of DNA varies with the carbon-source available.

When xylose was the only available source of carbon, the DNA content was measured to be 1,41 g/gDW(%), compared to 4,68 g/gDW(%) in mannitol. The growth rates in media with the same carbon sources were found to be 0.911 h^{-1} and 0.411 h^{-1}, respectively. These results are in accordance with research showing there is a positive correlation between growth rate and DNA-content, where faster growing cells contain more DNA [88]. Looking at the results in Figure 4.3A, and comparing the amount of DNA to the growth-rates, it looks like this might be the case. *B. subtilis* in xylose, which as mentioned had low amounts of DNA, also had a low growth rate. The opposite is true for the bacterium in mannitol, where the growth rate and amount of DNA is significantly higher than for the rest. Research on mero-olioploidity in *B. subtilis* might be a part of the explanation for this. Experiments in exponential growth phase with *B. subtilis* by Böttinger et al. (2018) found that the average number of origins was around 5.9, and the number of termini was 1.2. This indicates that the bacteria is mero-oligoploid, which means it has multiple copies of its genome during exponential growth. Their results showed that most of the cells had between 4 and 8 origins of replication, which aligns with previous research on *B. subtilis* by Sharpe et al. (1998), Webb et al. (1998), Kodoya et al. (2002) and Moriya et al. (2009) [89, 90, 91, 92] that showed the DNA content increased with increasing growth rate. The number or origins decreases when the cells reach stationary phase, indicating that the copy number is positively correlated to growth rate [88]. The results from these fermentations are in accordance with the findings by Böttinger et al. (2018), as the amount of DNA measured is higher in carbon-sources where a high growth rate was also observed.'

A Pearson correlation analysis gave a correlation coefficient of 0.76, indicating a positive linear correlation between DNA amount and growth rates in this experiment as well. However, the P-value for this correlation-analysis is 0.14 which is too significant to draw a conclusion on correlation between DNA and growth rates on this data alone. There are some deviations from the aforementioned research in the results from these fermentation experiments. The growth rate for *B. subtilis* cultivated in media containing succinate was the second lowest compared to growth on the other carbon sources. However, the amount of DNA found in *B. subtilis* in the succinate media is only surpassed by the amount found in mannitol. The results could indicate that the amount of DNA is not only correlated to the growth rate, but is also affected by the media composition. It might be that a positive correlation between growth rate and DNA-content would be seen at different growth rates in the same carbon source.

4.3.2 RNA content

The amount of RNA (g/gDW) was found to vary with the carbon-source. An ANOVA-test proved that the variations were statistically significant. The mass fractions (g/gDW) and the corresponding standard deviations are illustrated in Figure 4.3B. As can be observed, there is quite some variation in the amount of RNA found in cultivation of *B. subtilis* with different carbon sources available, and an ANOVA analysis gave a P-value < 0.001. There is especially a high variation in amount of RNA between mannitol compared to xylose and succinate. The growth rates for succinate and xylose were the lowest out of the five media compositions that were used for these fermentations.

The amount of RNA in *B. subtilis* has been shown by Dauner and Sauer (2001) to increase linearly with increasing growth rate [13]. The correlation between RNA and growth rate is associated with the increasing need for ribosomes to meet protein synthesis demands, meaning that the rRNA (ribosomal RNA) levels increase [93]. The expectation was therefore to observe that the cells that had faster growth rates would also have a higher amount of RNA. RNA content (g/gDW) was plotted against growth rate to illustrate the results. From observing Figure 4.4 there seem to be a correlation and linear relationship between the two. A pearson correlation analysis revealed a correlation coefficient of 0.69, however the P-value for the analysis was too significant to make a conclusion on this based on the available results from these experiments.

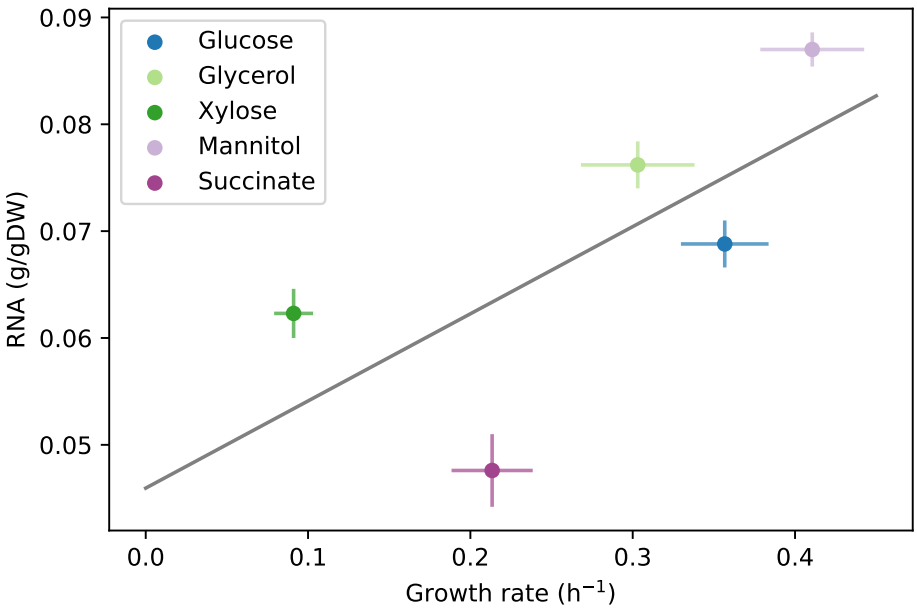


FIGURE 4.4. The calculated RNA-amount (g/gDW) of *Bacillus subtilis* from batch fermentations performed in minimal media with different carbon sources available, plotted against the corresponding growth rate (h^{-1}) found from biomass analysis.

As can be observed in Figure 4.4, the levels of RNA in xylose is higher than in succinate, but succinate has a higher growth rate. Same can be observed for glucose and glycerol. This is in disagreement with the research by Dauner and Sauer (2001) indicating a positive, linear relationship between RNA content and growth rate. However, the results in these experiments are from cultivations with different carbon sources, and they could indicate that the amount of RNA varies with the carbon source available.

Comparing these results with those of Dauner and Sauer (2001) in a direct manner is not entirely appropriate. The experiments they conducted were performed using a chemostat set-up in minimal media containing glucose, harvesting *B. subtilis* for biomass analysis at various growth rates. The potential explanation for why the same correlation is not observed in the results from these experiments could be attributed to the different carbon-sources used for each measurement.

4.3.3 Protein content

The protein analysis was performed following an amino acid derivatization with HPLC after an acid hydrolyzation protocol. The resulting concentrations of each amino acid was used to find the total content of proteins (g/gDW) which is illustrated with corresponding standard deviations in Figure 4.3C. As can be observed in Figure 4.3C, there is not much variance in total amount of protein between the different media compositions. An ANOVA analysis gave a P-value < 0.01, showing some significant variance between the groups. A post hoc analysis showed there is only significant variance between the bacteria cultivated in mannitol compared to glucose, xylose and succinate. These results indicate that there is not much variance in protein content as a result of variation in available carbon source.

Dauner and Sauer (2001) found there to be a positive correlation between protein content and growth rate in *B. subtilis* [13]. A Pearson correlation analysis performed on this data-set resulted is a correlation coefficient of -0.61, indicating the opposite of their discoveries. However, the P-value for the correlation coefficient was 0.28, and so no conclusion can be drawn on whether the amount of protein increases with growth-rate based on this data. The negative correlation and high P-value could be due to a low number of samples, or as a result of the effect of carbon source. However, as mentioned with the RNA results, the experiments from Dauner and Sauer (2001) were performed using a chemostat-setup and glucose as the carbon source at various dilution rates. A comparison of correlation between proteins and growth rate from these batch fermentations with different carbon-sources to their results is therefore not optimal.

As described in Section 3.5.3 the HPLC analysis gave the concentration of each amino acid in the samples. A one-sample T-test was performed comparing the experimentally measured amino acid triplicates to the molar-fraction in the model for each amino acid. Though a few amino acids were not significantly different in concentration from the model coefficient, most were. Because there was significant variance, the concentrations of amino acids were used

in the model. The synthesis reactions for protein were updated with the scaled amino acid concentrations.

The relative distribution of amino acid composition in all carbon sources is illustrated in a radial plot (ref. Figure 4.5). There is not much variance in the relative distribution, except for Leu and Arg.

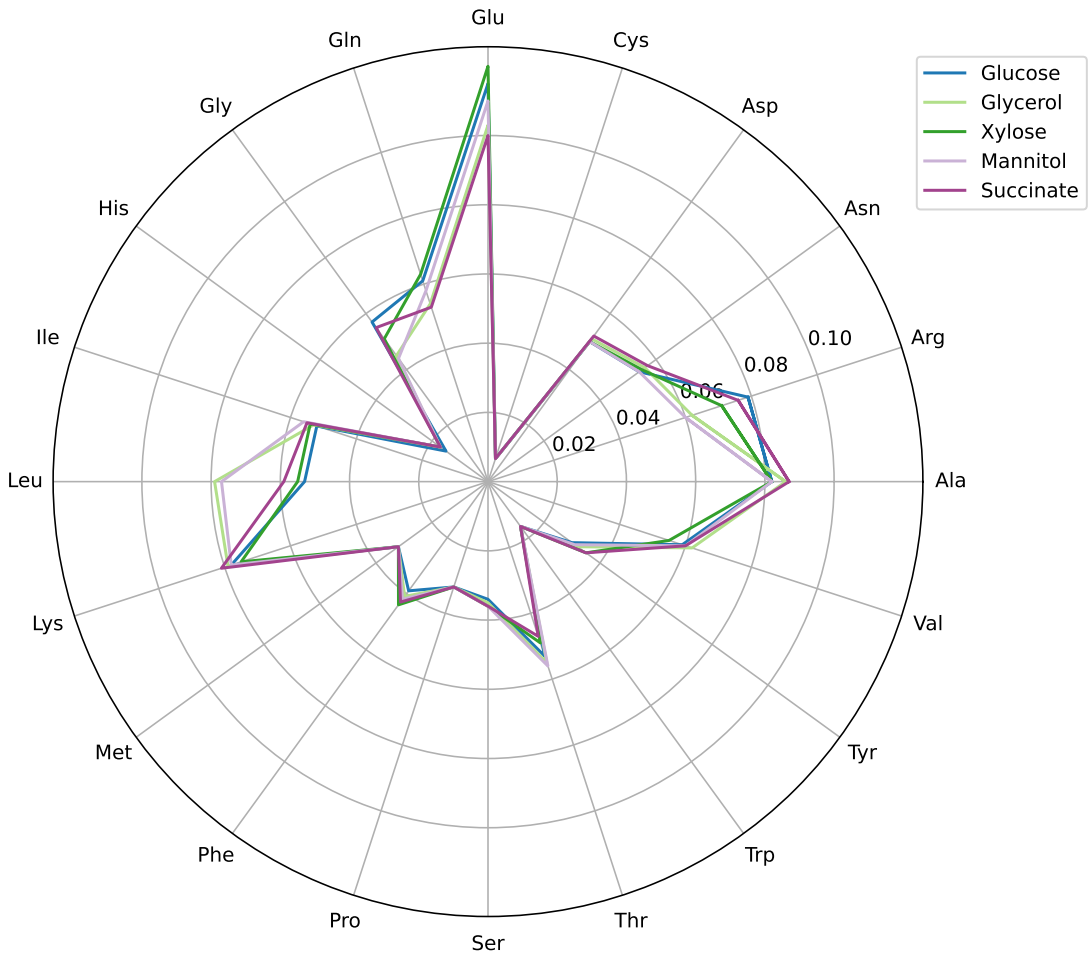


FIGURE 4.5. Radial plot of the relative distribution of amino acids in g/gDW fractions in protein samples from *B. subtilis* dry weight samples harvested from five batch fermentations performed in minimal media with different carbon sources.

Amino acids are enzyme-building blocks. The sequence of amino acids determine the fold of proteins and thereby the functional structure of the enzyme. The composition of amino acids has been shown to vary between species, additionally Simensen et al. (2022) [94] found that the amino acid distribution of the yeast *Saccharomyces cerevisiae* varied in response to the external environment. However, there is variance in the degree to which amount of individual amino acids tend to vary with the environment [94].

The variance in amount of amino acids reflect the demand of amino acids by the cell [94]. The metabolism has to be adapted in response to the available carbon source. This adaption is associated with expression of different genes to produce the needed transporters and enzymes [85], and amino acids are as mentioned the building blocks of proteins and enzymes [94]. Variation in amino acid distribution could be an indication of such adaptions to a new environment where different proteins are needed.

4.3.4 Lipid content

Lipids were extracted from the freeze-dried cells by chloroform/methanol extraction as described in section 3.5.4. The total lipid content extracted from cells and the corresponding standard deviations are shown in Figure 4.3D. An ANOVA analysis on the data gave a P-value of 0.289. The following post hoc analysis showed that there was only a statistically significant variance between the bacteria grown in succinate and glycerol. The results indicate that the amount of lipids is not affected by different carbon sources.

It was expected to see some variation in lipid content based on growth rate. It is known that cell size is positively correlated to growth rate in environments characterized by limited nutrient availability. Further, the growth rate is related to metabolic rates and cell size to phospholipid synthesis, meaning that an increase in growth rate with a subsequent increase in metabolic rates, leads to a larger cell size and higher production of phospholipids for the cell membrane [95]. In *B. subtilis* this relationship between cell size and growth rate is linear [96]. However, though differences in growth rates are observed in these experiments, all measured growth rates are between 0.1 h^{-1} and 0.5 h^{-1}. In this part of the linear regression of cell size and growth rate, the cell size does not increase by a large amount [96].

4.4 The experimental uptake- and secretion rates

4.4.1 Carbon, glutamate, tryptophane, and acetate

During the rounds of fermentation, samples were extracted from the media to perform analyses of what is taken up from and released into the media by *B. subtilis*. These samples were analyzed by NMR. The results from the analyses were used to calculate uptake- and secretion rates for glutamate, L-tryptophane, Acetate, Pyruvate and the respective carbon-source present in the media. The calculated uptake rates for *B. subtilis* cultivated in different carbon sources are shown in Table 4.2.

TABLE 4.2. The calculated rates of uptake- and secretion of various compounds measured from five batch fermentations of *B. subtilis* in minimal media containing different carbon sources. The rate is measured at mmol $gCDW^{-1}$ h^{-1}. The carbon-source, glutamate and tryptophane are taken up from the media, while acetate is secreted in to the media.

Fermentation:	**Carbon source:**	**Acetate**	**Glutamate**	**Tryptophane**
Glucose	3.52 ± 1.0	3.60 ± 0.12	1.99 ± 0.65	0.021 ± 0.006
Glycerol	14.2 ± 2.8	3.04 ± 0.17	2.30 ± 0.55	0.031 ± 0.007
Xylose	0.58 ± 0.04	0.07 ± 0.00	1.59 ± 0.29	0.008 ± 0.001
Mannitol	4.79 ± 1.33	3.01 ± 0.01	1.84 ± 0.65	0.013 ± 0.003
Succinate	5.76 ± 1.97	0.23 ± 0.04	2.84 ± 1.55	0.017 ± 0.006

The results from an ANOVA analysis showed there is statistically significant variance in uptake-rates between all compounds except for Glutamate. When it comes to the uptake of carbon sources, it was only the variance between glucose uptake with xylose, mannitol and succinate that was not statistically significant. As can be seen in Table 4.2 the difference in uptake is quite large. *B. subtilis* in glycerol had the highest uptake-rate of carbon source, while the uptake of xylose was quite low. As has been discussed already, the uptake of xylose is likely reduced by the fact that *B. subtilis* does not have a xylose-specific transporter [22]. In contrast, the bacterium is able to import glycerol quite efficiently via a glycerol facilitator [87].

Dauner and Sauer (2001) did analyse the uptake of glucose and secretion of acetate in correlation with growth rate. Based on their results, the uptake-rate of glucose calculated here was lower than expected. With a growth rate of 0.3567 their research suggests the uptake-rate should be around 6 mmol g^{-1} h^{-1}. The calculated rate here was 3.52 mmol g^{-1} h^{-1}, which is lower than their measured value. For acetate secretion the measured value from glucose-media in this experiment is close to the values measured at similar growth rates for glucose found by Dauner and Sauer(2001) [13].

There is also a large variance in secretion of acetate, especially between succinate and xylose compared to *B. subtilis* in the other media compositions. The secretion of Acetate is associated with "overflow" metabolism, meaning a switch in metabolism happens if the cell has reached its respiratory capacity or during fermentation at high glucose concentrations [97]. *B. subtilis* growing in glucose, glycerol and mannitol were the fastest growing. The results indicate that the cells in media containing xylose and succinate did not enter the "overflow" stage, perhaps because growth on these substrates proved inefficient.

4.4.2 Calculated O_2-uptake and CO_2-secretion

During the fermentation, analyses were automatically performed of the gas flowing through the system with a Thermo ScientificTM Prima BT Benchtop Process Mass Spectrometer. The off-gas was sampled continuously throughout the fermentation process, and the resulting concentrations of O_2 and CO_2 were used to calculate the uptake and secretion rates. Uptake-

and secretion rates were calculated based on the specific rates from the 10 last measurements before the cells were harvested and are listed in Table 4.3.

TABLE 4.3. The calculated rates of uptake of oxygen and secretion of carbon dioxide measured from five batch fermentations of *B. subtilis* in minimal media containing different carbon sources. The rate is measured at mmol $gCDW^{-1}$ h^{-1}. O_2 is taken up by *B. subtilis*, and CO_2 is secreted.

	Oxygen		**Carbon dioxide**		
	Rate:	Std. Dev:	Rate:	Std. Dev:	RQ:
Glucose	5.22	0.30	5.49	0.81	1.05
Glycerol	9.85	0.06	7.23	0.12	0.73
Xylose	6.06	0.06	6.68	0.03	1.10
Mannitol	5.58	0.25	4.96	0.26	0.88
Succinate	14.2	0.9	16.7	1.0	1.17

An ANOVA analyses on the O_2 uptake-rates gave a P-value < 0.001. A Tukey's post hoc test was performed and showed that there is statistically significant variance between most rates calculated for the various carbon sources, except between mannitol with glucose and xylose. The ANOVA analysis for CO2 also showed statistical significance with a P-value below P-value < 0.001. Glucose and mannitol, and glycerol and xylose were two only combinations where the variance was not found to be significant for CO_2 secretion.

The respiratory quotient was calculated to determine whether the uptake of oxygen and release of carbon dioxide is balanced for growth in the various substrates. The RQ indicates how the carbon-source is utilized or to what degree it is reduced, and is therefore an important part of understanding the metabolic activity [98]. The RQ value will be 1 for glucose under aerobic conditions if all reactions in the respiratory chain are active, and no alternative electron receptors are in use [99]. The RQ can vary with the carbon-source, and it is higher than 1 when more CO_2 is produced than oxygen consumed, and vice versa [98]. As observed in these experiments, the RQ varies with the carbon-source. The RQ for *B. subtilis* in glucose is almost 1, indicating that the oxygen consumption and carbon dioxide secretion is happening at similar rates. *B. subtilis* in xylose and succinate have a higher production of CO_2 than O_2. They were also not able to grow efficiently on these substrates.

4.5 New experimentally derived, condition specific biomass objective functions

The first step in creating the new biomass objective functions was to lump the cofactors and ions together in one new synthesis reaction where the product was a co-factor metabolite that was created. This reaction was called the cofactor synthesis reactions and was added to the model as well as the '"cofactor" metabolite. The new BOF was structured as shown in equation 4.1.

$$
\begin{aligned}
&c_1 \text{ protein} + c_2 \text{ DNA} + c_3 \text{ RNA} + c_4 \text{ lipid} + 0.0304 \text{ lipoteichoic acid} \\
&+ 0.224 \text{ cell wall} + 0.0446 \text{ cofactor} + 105.00 \text{ ATP} + 105.00 \text{ } H_2O \\
&+ 0.00027 \text{ Acyl- carrier protein} \qquad (4.1) \\
&\rightleftharpoons \\
&104.99 \text{ ADP} + 104 \text{ Orthophosphate} + 0.00027 \text{ Acyl-carrier protein}
\end{aligned}
$$

The only coefficients updated in this experiment are the DNA, RNA, protein and lipids coefficients. All substrates in the updated BOFs are the same as in the *i*Bsu1147 model, only the metabolites and ions are lumped in the cofactor synthesis reaction. After having structured the BOF, the synthesis reactions for DNA, RNA, proteins and lipids were found. Those reactions were scaled so that the total g/gDW of all substrates and products in each reaction would have a sum of 1 not counting the macromolecule. That way, the measured g/gDW of the macromolecular constituents could be used directly as coefficients in the BOF. The new, scaled BOF-coefficients in g/gDW are listed in Table 4.4.

TABLE 4.4. The experimentally derived biomass objective function coefficients (g/gDW) found in batch fermentation analyses performed with *B. subtilis* in minimal medium with a different carbon-source (C-source) available. These are the scaled g/gDW coefficients from the experimentally measured values. They are scaled so that the sum of coefficients for all components in the BOF together add up to 1 g/gDW. The original BOF coefficients from the model are listed at the bottom of the table as a comparison to the measured values obtained by these experiments.

	DNA	**RNA**	**Protein**	**Lipid**	**Sum (%)**
Glucose	0.0234 ± 0.0015	0.0697 ± 0.0022	0.572 ± 0.027	0.0311 ± 0.0017	69.6
Glycerol	0.0245 ± 0.0014	0.0784 ± 0.0022	0.559 ± 0.003	0.0301 ± 0.0021	69.2
Xylose	0.0145 ± 0.0003	0.0642 ± 0.0023	0.576 ± 0.003	0.0369 ± 0.0019	69.2
Mannitol	0.0484 ± 0.0019	0.0899 ± 0.0015	0.519 ± 0.021	0.0336 ± 0.0032	69.1
Succinate	0.0282 ± 0.0009	0.0490 ± 0.0034	0.578 ± 0.013	0.0369 ± 0.0007	69.2
*i*Bsu1147	0.026	0.0655	0.528	0.076	69.55

There is some variation observed between the content measured from these experiments and the coefficients found in the model. However, the coefficients in the oBOF are not measured from *B. subtilis* in the same media-composition, in the same phase of growth, or are necessarily derived from experiments of *B.subtilis* [16, 13]. A one sample T-tests were performed comparing experimentally measured g/gDW to the expected value for that coefficient in the oBOF. The expected value used in the T-test was the corresponding coefficient value in the oBOF. The experimental values for the nBOF coefficients were scaled using the same scaling-factor as for the final coefficients used in the BOFs. Otherwise raw data would be compared to a scaled value from the model. The one sample T-test showed that all nBOF coefficients for glycerol and succinate were significantly different from the original coefficients. For mannitol and glucose all but the protein coefficients were significantly different from each other.

For xylose it was only the RNA coefficient that did not statistically vary from the value in the original BOF. The results from the T-test are in appendix D.

The sum (%) in Table 4.4 refers to the total percentage of the biomass composition that is represented by these new coefficients. The row on the far right contains the corresponding coefficients from the *i*Bsu1147 model. The BOF in the model was created using stoichiometric coefficients from previous lab-experiments from Dauner and Sauer (2001) [13]. There was a significant variance in the lipid coefficient derived from these experimental measurements, compared to the coefficient in the model. However, this was not surprising as the lipid coefficient in the oBOF is based on data by Bishop (1967) [13], but the bacteria in the experiments were cultivated until they reached stationary phase [100]. As mentioned in section 4.3.4 cell size is known to vary with growth rate and the rate metabolic activity [95]. Because the cells were in stationary, they likely exhibited different growth characteristics than the bacteria in these experiments, harvested in exponential phase. Some variation in lipid-content was therefore expected. However, this brings to question why the amount of lipids measured here is lower.

The protein coefficients measured experimentally in the nBOFs exceed the corresponding coefficient in the oBOF, with the exception of *B. subtilis* in mannitol. The coefficient in the model is calculated from the equation of the linear regression Dauner and Sauer (2001) performed with protein content against growth rate. The coefficient is calculated with a growth rate of 0.1 [16, 13], which is a lower growth rate than *B. subtilis* exhibited when the protein content was calculated. This might explain why the experimentally measured protein-coefficients for the nBOFs surpass the oBOF coefficient. As the variance between the coefficients for the nBOFs compared to the coefficients in the oBOF was significant, the new nBOFs were implemented in the model to explore what effect they have on model predictions.

4.6 Implementation of the condition specific BOFs in *i*Bsu1147

After having found the nBOF coefficients (Table 4.4), these were implemented in the *i*Bsu1147 GEM. The next step of the project was to see how the new condition-specific BOFs affect model predictions. For mannitol and glucose, the measured growth rates implemented together with the measured uptake- and secretion rates led to the model not being able to find a feasible solution. For those two carbon sources new flux rates were found using MOMA, and the resulting uptake- and secretion rates from the MOMA analysis was further used in FBA. All models were analysed with FBA and FVA to see how the flux distribution and growth rate is affected by the nBOFs.

4.6.1 Identifying feasible solutions using MOMA

With the new fluxes measured in glucose and mannitol, the model was not able to find feasible solutions. As described in section 2.3.4, a MOMA analysis can be performed to find a solution vector of fluxes that is as close as possible to the fluxes provided in the reference vector [46]. Therefore, the experimentally measured fluxes were used as a reference vector in the MOMA analysis to find the closest, feasible uptake- and secretion rates given the measured growth rate. The upper- and lower bound for the growth-rates was set to the measured rate ± 2× the standard deviation, to allow for some flexibility. The MOMA-analysis was performed with the nBOF and oBOF from both media compositions. The resulting list of fluxes, Table 4.5, was used for further analysis with FBA and FVA.

TABLE 4.5. A MOMA analysis was performed to find a feasible vector of fluxes given the measured growth rates. The experimentally measured fluxes were used as a reference, to find the closest possible, feasible fluxes to the experimentally measured uptake- and secretion rates. The growth rate for the analysis was locked at experimentally measured value ± 2 × std. dev. The feasible fluxes from the analyses are listed in this table, and these flux values were used as constraints in the model for the nBOF and oBOF for these media compositions for pFBA and FVA.

	Glucose		Mannitol	
Flux	nBOF	oBOF	nBOF	oBOF
Growth rate	0.303	0.303	0.379	0.379
C-source	4.86	4.95	6.79	6.84
O2	6.74	6.83	9.12	9.12
CO2	6.00	6.03	6.01	6.03
Acetate	3.79	3.52	3.01	3.01
Glutamate	2.06	2.07	2.01	2.09
Tryptophane	0.01	0.02	0.02	0.02

The suggested fluxes from the MOMA analysis are quite similar between the oBOF and the nBOF for glucose. The coefficients in the nBOF shifts the MOMA-results a slightly towards the rates that were measured experimentally, compared to the oBOF. The same applies for the MOMA results for mannitol. Although the flux values suggested for the nBOF and oBOF by MOMA analysis vary slightly from each other, the growth rates suggested for the model with the nBOF and the oBOF are the same. These observations were consistent for both glucose and mannitol. Indicating that the growth rate predictions are quite robust against changes in the BOF coefficients.

The MOMA analyses for both conditions suggested higher uptake- and secretion rates than what had been measured, and a slightly lowered growth rate. It has been observed in cases with other models that the growth rate predictions of the model can be slower than what is measured experimentally. Potential explanations for this is that the efficiency in the model for generating biomass is too low compared to what is observed *in vivo* [76]. The highest growth rates in these experiments were measured for *B. subtilis* in glucose and mannitol. It

might be that the measured uptake-rates were not significant enough to support this growth in the model. Or it might be that the model needs improvement when it comes to generation of biomass as mentioned in the article by Lewis et al. (2010) [76].

4.6.2 Analysis of flux patterns associated with the new BOFs using FBA

The FBA analyses were performed comparing the nBOF with the oBOF in the each of the different media compositions, for example nBOF and oBOF from glucose media with the measured uptake- and secretion-rates derived experimentally for that media composition. The type of FBA analysis performed was a Parsimonious FBA (pFBA). This analysis assumes that there is a selection of the fastest growing strain that requires the least amount of flux through the network during exponential growth. For the bacteria this means that under growth selective pressure, the cells will over time adapt to a higher growth rate. In this process there will be a down-regulation of expression of genes and proteins in non-functional reactions. This down-regulation saves resources which then can be allocated towards further increasing the growth rate. This assumption of down-regulation separates pFBA from FBA [76].

As mentioned, pFBA finds the solution in the space of feasible solutions which yields a high growth rate combined with the least amount of flux through the model [76]. This means the pFBA can be used to provide one feasible solution that is made on the same basis for all the carbon sources in these experiments, and ensures that the results are comparable.

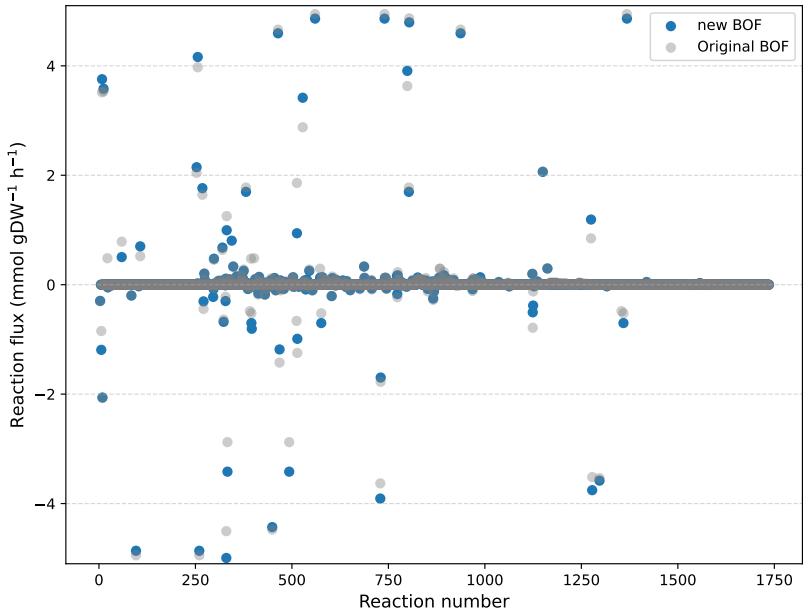


FIGURE 4.6. A scatter-plot of the resulting fluxes following a parsimonious flux balance analysis (pFBA) on the GEM *i*Bsu1147 with the new experimentally derived BOF (nBOF) compared to the flux distribution following a FBA in the same model with the original BOF (oBOF), both in minimal media containing glucose. The y-axis was cut at 5 and -5 so the variety in flux-distribution could more easily be studied. The synthesis reactions for the macromolecular constituents of the BOF are not included in the plot, due to differences in the nBOF and oBOF formulation.

The resulting flux-distribution from the pFBA of the model with the nBOF and oBOF in minimal media containing glucose show some variance in flux distribution, as illustrated in Figure 4.6. KEGG REACTION database [77, 78, 79] was used to see what pathways the reactions with biggest variance between nBOF and oBOF reactions belong to. For glucose the biggest flux variances were observed between oBOF and nBOF in reactions that are a part of carbohydrate, lipid, and amino acid synthesis. This was expected to see. The amino acid composition in the protein synthesis reaction was updated with experimental values, as well as the coefficients for DNA, RNA, proteins and lipids. Changing these coefficients mean that their weight of significance has been changed. This will impact the other reactions in the model - the flux level is adjusted by the new biomass requirements [12]. Between the nBOF and oBOF in the other media compositions the same can be observed, as seen in Figure 4.7. There are adjustments in the flux distribution as a result of the updated amino acid composition in the model as well as the coefficients in the nBOFs.

There is variance in the flux patterns between the media compositions, as expected. The bacterium has to adapt the metabolism to the available carbon source. This adaption to the carbon source is observed in reaction numbers 1 to 250, which represent the exchange reactions, and one can see that the flux pattern for these reactions is different for each carbon source. *B. subtilis* reached a higher growth rate in glycerol and mannitol than in xylose and succinate. In the flux-plots for glycerol 4.7A and mannitol 4.7C the distribution is more

wide-spread and there are more reactions of higher flux, and the comparison between nBOFs and oBOFs show some variance. In xylose 4.7B and succinate 4.7D, the bacterium's growth rate was significantly lower. In these plots, the flux distributions are less dispersed and the variance in flux between nBOF and oBOF is less obvious.

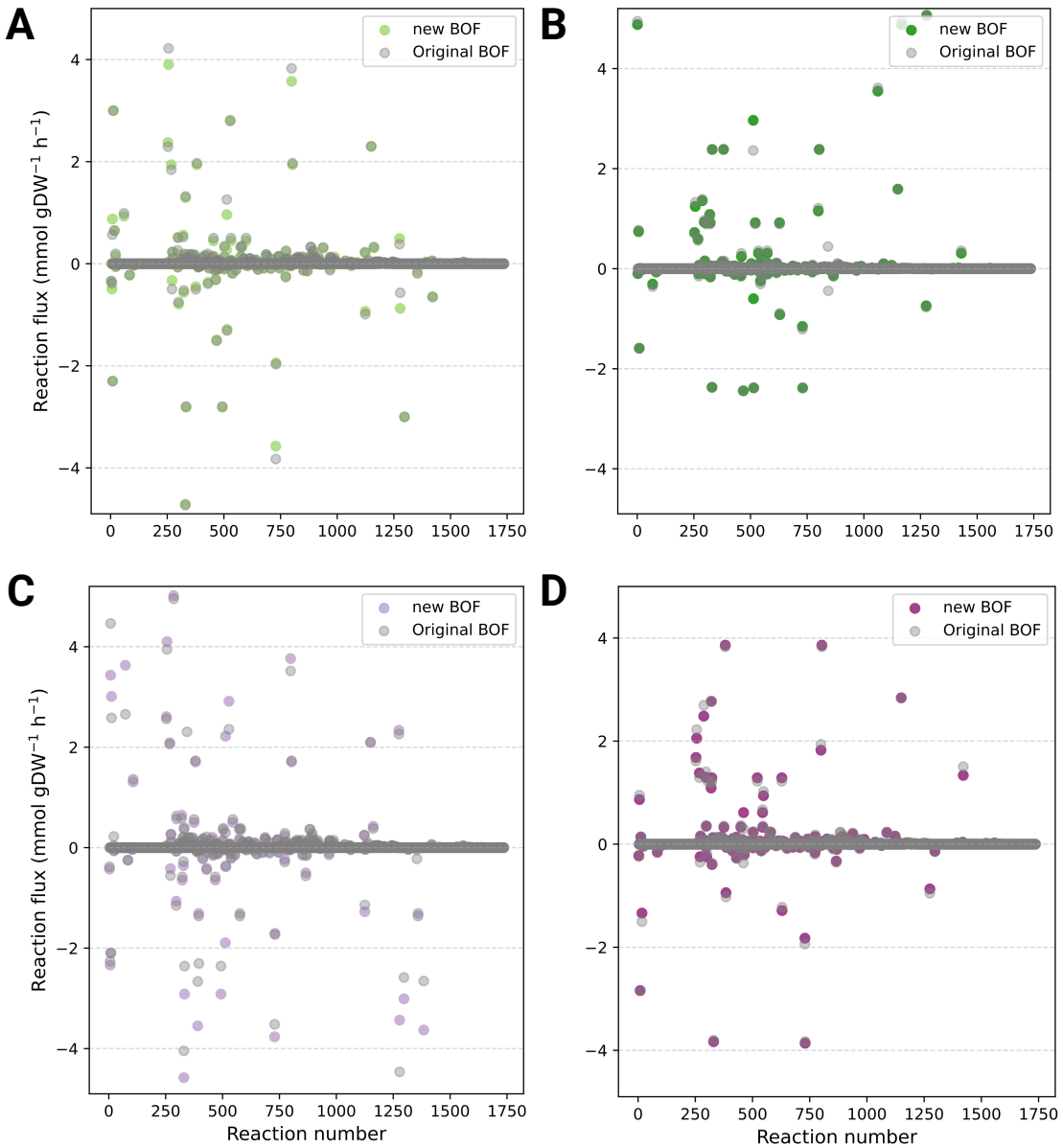


FIGURE 4.7. Plots of the flux ranges per reaction in *i*Bsu1147 found by performing pFBA on the model with the original BOF (oBOF) and the new BOF (nBOF). **A** shows the flux distribution in minimal media containing glycerol as sole source of carbon. **B** shows the flux distribution in minimal media containing xylose. **C** in mannitol, and **D** succinate. As all four flux distributions had some flux values of much greater value than the rest, the plots y-axes were cut of at 5 mmol gDW^{-1} h^{-1}, in order to better visualize the flux distribution for the majority of the reactions.

The growth rates predicted by pFBA with the nBOFs and oBOF in the same media compositions show consistent results, indicating that the model is resilient when it comes to growth rate predictions. A group of researchers studying diatoms, Lavoie et al. (2020), found that that the GEM for the organism in study was robust to changes in biochemical composition, except for carbon uptake rate [101]. In this case, the pFBA analyses were only conducted between nBOF and oBOF with the same uptake-rates for carbon, so the same observation can not be derived for *i*Bsu1147 on these data alone. However, the data shows that the growth rate predictions are robust to changes in biomass composition.

4.6.3 Assessing the impact of the new BOFs on optimal steady state using FVA

A FVA analysis was performed with the new flux- and growth rates for all five carbon sources. For comparison, the FVA was performed on the model with the oBOF for all 5 flux-environments. This way, the resulting flux ranges for the model with the nBOFs could be compared to the results from the same analysis with the oBOF.

The reactions from 1 to 252 are the exchange reactions in the model. As the uptake- and secretion rates for glucose and mannitol are from the MOMA analysis and therefore not exactly the same between nBOF and oBOF. There are small variances, as seen in Table 4.5. Overall, that area of the curve for mid-value, as seen in Figure 4.8A and B, is more or less the same between mBOF and oBOF. In these experiments, only the uptake- and secretion-rates for O_2, CO_2, carbon-source, acetate, tryptophane and glutamate are measured and updated. This means only 6 of the flux-variables are updated out of 252 exchange reactions. So it was not expected to see a major change in this area of the plot.

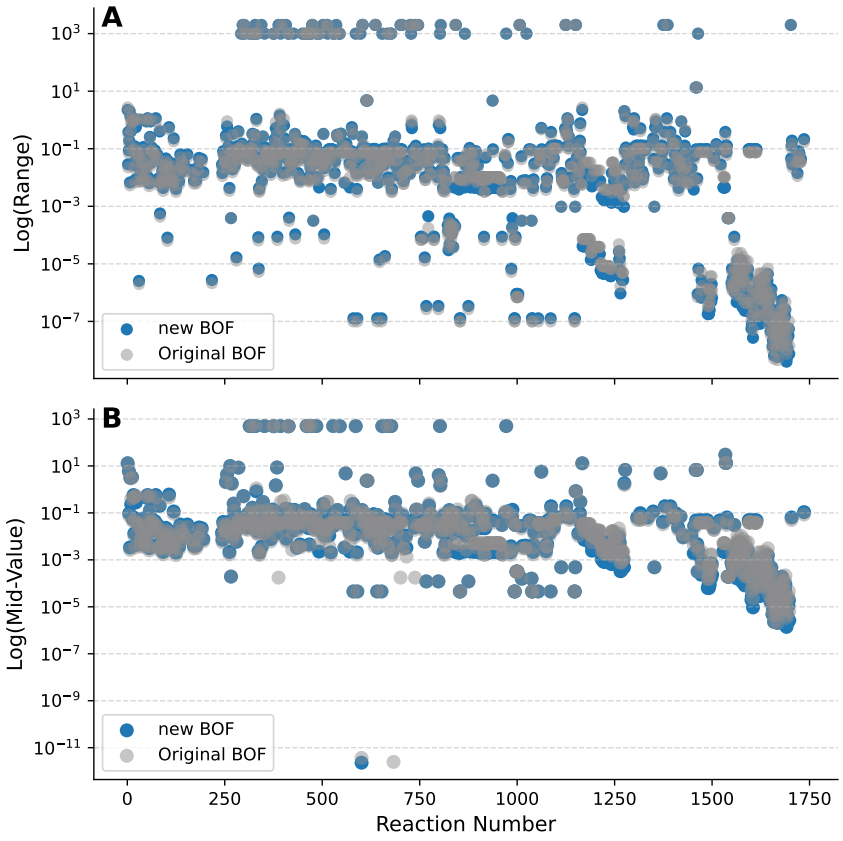


FIGURE 4.8. A FVA analysis was performed on the GEM *i*Bsu1147 with an experimentally updated BOF (nBOF) and for the same network with the original BOF (oBOF). Both analyses were performed for the network in minimal media with glucose as the available carbon source. Plot **A** shows the resulting log of the range values, and plot **B** shows the log of the mid-values from the analysis.

The reactions from around reaction number 850 to 950 are related to synthesis of lipids. All nBOFs have coefficients for lipids almost half the weight of the oBOF. This is visualized in both plot A and B. There is a line of reactions around those reaction numbers where the range and mid-values for nBOF are slightly lower on the y-axis than for the oBOF. Overall Figure 4.8 shows there is a slight reduction in the span of max and min flux values for the model, indicating that the solution space has shifted.

For the other media compositions, Figure 4.9, the same is observed. For the lipid-reactions, there has been a shift to a lower range and lower mid-value for the nBOFs. However, because the only the content was measured for DNA, RNA and lipids, and not the composition, it was not expected to see great changes in flux distribution. What was expected to be observed was a shift, depending on whether the measured content was higher or lower than the coefficient in the model. As for lipids, the measured content resulted in a coefficient for the BOF that was significantly lower than in the oBOF. Therefore, the maximum and minimum possible flux values have shifted, but the composition remains the same and that is why the pattern of flux value is mostly unchanged.

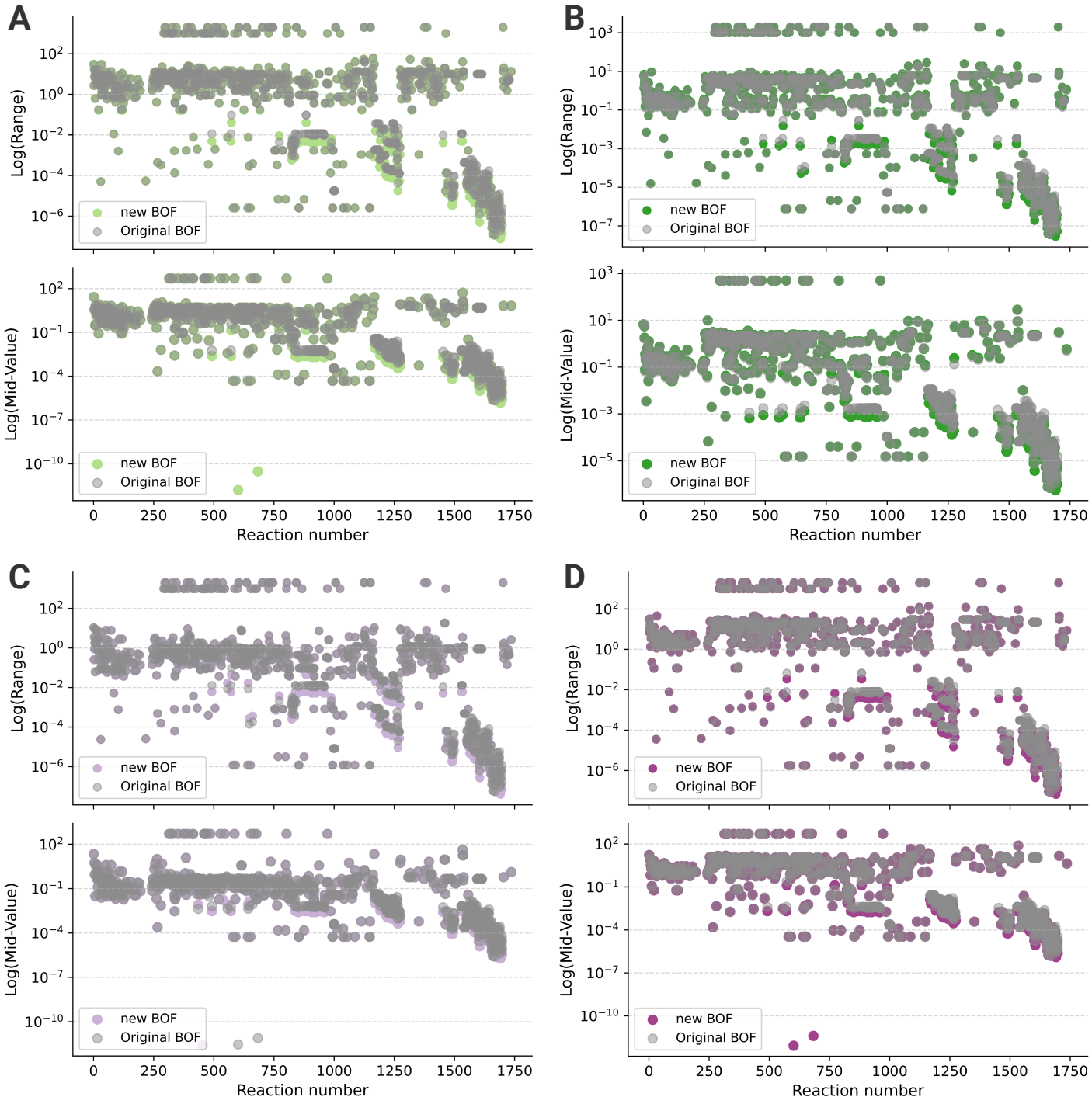


FIGURE 4.9. Plots of the resulting distribution of range and mid-values calculated after FVA analysis of the network with the oBOF and nBOF in the model *i*Bsu1147, in minimal media containing different carbon source. **A** visualizes the resulting range and mid-values for *B. subtilis* in media containing glycerol as sole source of carbon, **B** is in xylose, **C** is in mannitol, and **D** is in succinate.

5. Discussion

Initial discussions around the results were presented in chapter 4. In this chapter the scope is broadened and the discussion will be centered around the aim of this project, building on the discussions in chapter 4. To begin with, the experiments performed in the lab are discussed, before moving on to a discussion on the analyses of how the experimental BOFs affect the model predictions. This will lead to the section on whether updating the BOFs with experimental data affects model predictions. As presented in the Introduction, the aim of the study is to perform experiments on *B. subtilis* to map its biomass composition and update *i*Bsu1147 BOF with the measured coefficients. The results from these analysis should help answer the question of whether BOF-coefficients should be condition-specific.

5.1 Evaluation of the protocol for measuring biomass composition

5.1.1 Challenges and considerations with batch-fermentation

There are two experimental parts to the quantification of biomass; cultivation of the bacteria and then the actual protocols for measuring its biomass. There are several available protocols, and each comes with its own set of considerations [4]. Because several carbon sources were tested in this project, and the main objective was to examine if changes in available carbon source affects biomass composition, the chosen cultivation method was batch fermentation.

There are multiple benefits to batch-fermentations. Because the organism and media are loaded in to the reactor aseptically at the beginning, there is little risk of contamination during the cultivation process [102, 31]. Further, the growth-conditions can be controlled in many aspects, for example temperature, agitation, aeration and pH [31]. However, because there is no addition of fresh media, the cells added at the beginning will continue through all phases until the death phase, unless they are harvested before nutrient depletion. They cannot be maintained in exponential phase as can be achieved with a chemostat setup [31, 32]. This can affect how well experimentally measured values align with the underlying assumptions of an analysis, such as the steady-state constraint of FBA.

It takes generations for bacteria to properly adapt its metabolism and then acquire an optimal growth phenotype. For example, Lewis et al. (2010) found that it took hundreds of generations of adaptive evolution before the growth of *Escherichia coli* was in accordance with pFBA results. As described previously, the assumption behind pFBA is that the cells

will evolve to achieve higher growth rate, while minimizing the flux through the reactions in its metabolism. Because an excess of enzymes require large amounts of energy, the idea is that the organism will adapt to not over-produce in order to allocate as much energy as possible into growth and proliferation [76]. In batch-fermentation the cells are not maintained in exponential phase for a prolonged amount of time. The cells go through metabolic transition in the lag phase where their metabolism is adjusted to the new environment. After the adjustment the cells enter the exponential growth phase [31]. They have adjusted their metabolism, but they might not have achieved this optimal state prior to harvest. If this is the case, the experimental conditions deviate from the underlying assumption of pFBA. However, the aim of this project was to examine changes in biomass with environmental condition. For this purpose it should be sufficient that the biomass is measured after metabolic adjustments have been made.

The alternative to batch fermentation would have been to perform the experiments using a chemostat setup. In this process, the cells can be kept in exponential growth longer by addition of fresh media. This is combined with an overflow device so that fresh medium replaces old, and continuous growth in exponential phase can be maintained. If the addition of the media is kept at such a rate that the organism maintains a stable growth rate, eventually a steady state is achieved. The total amount of mass leaving the system is equal to the mass entering the system. Introducing fresh media facilitates cell division, and the old media with the deceased cells are removed [32]. In batch-fermentation, the growth can not be stabilized like this over time. However, as previously stated, one can assume that the culture has steady state properties during exponential growth [81]. This is why it was essential that the cells were harvested during exponential growth in these experiments. The analysis of the biomass composition and its incorporation into the model could be completed ensuring that the experimental additions to the model were in alignment with its underlying steady-state assumptions.

Another consideration in terms of cultivating the cells through batch-fermentation is that there is great variability from one run to another [34]. While there are technical replicates for the biomass analyses, and triplicates for growth rate determination, only one batch fermentation analysis was performed for each carbon source. Usually, the biological variability is higher than the technical variability [103]. Having biological replicates would strengthened the observations related to biomass composition and growth rates. With multiple replicates the analyses would have been more robust, and this would have enhanced the reliability of the results.

5.1.2 Addressing the protocol for biomass quantification

In these experiments the DNA, RNA, protein, amino acid distribution and lipid content was measured following the pipeline by Simensen et al. (2021) [2] which is based on the workflow by Beck et al. (2018) [4]. Simensen et al. (2021) characterized 91.6% of the bacterial mass of *Escherichia coli*. This percentage is significantly higher than achieved in these experiments.

However, they also performed an analysis of carbohydrate content, and so it was not expected to see the same coverage from these analyses. Further, because *B. subtilis* as a gram-positive bacteria has a higher content of cell-wall material, a lower percentage of the cellular dry weight is RNA [13]. The coverage following the same pipeline using *B. subtilis* was therefore expected to be slightly lower for the macromolecular constituents of the BOF that were quantified in these experiments.

The total coverage is an important factor of evaluation of the protocols for biomass quantification because when constructing the BOF, the coefficients in the reactions are scaled such that the coefficients add up to represent 100% of the cellular dry weight or biomass, and the biomass produced must make add up to 1g/mmol to relate biomass to other characteristics such as growth rate [2, 50]. If the total coverage is high, it provides a more accurate understanding of the proportions of the components in the BOF, and the need for loss-adjustment is reduced. Thereby, a high total coverage provides a BOF that more accurately represents the organism [2].

Not all components of the BOF were found experimentally in this project, and there is not a lot of available data on *B. subtilis* biomass composition. Therefore, it is challenging to draw a final conclusion on the total coverage achieved in these experiments. A carbohydrate analysis would have been interesting in addition, as this is an important macromolecular constituent in the BOF. However, compared to the coefficients in the model, the total coverage of the measured coefficients for DNA, RNA, protein and lipids is in close range to the total representation by the model coefficients 4.4.

The amount of proteins is normally the largest among the macromolecules in the BOF. The quantification based on amino acid analysis is one of the methods for protein quantification with less bias compared to UV absorbance and dye-methods. It is also more accurate than these methods. Further, using different dilutions of the samples ensures that amino acids of various abundance in the cell can be detected. One challenge with this method is that cysteine and tryptophane are degraded during hydrolysis. Additionally, methionine is only present in small amounts and usually the HPLC measurements have a high standard deviation [4]. However, it is possible to correct for these issues, as was done in these experiments. In this case the measured values of the other amino acids were plotted against the predicted concentrations based on the protein-coding genes in the *B. subtilis* 168 genome. A linear regression was performed and the concentrations of cysteine, methionine and tryptophane were found by their predicted concentrations. The same challenge is amino acids with overlapping retention times. Corrections can be made for this scenario as well, on the same assumptions as just described using the relative distribution [2]. Even though adjustments have to be made for some amino acids, this protocol has proven to be highly accurate for amino acid distribution and protein content [4].

The amount of RNA is a reliable indicator of the metabolic activity of the cell, as the RNA content is correlated with the amount of protein being synthesized [72]. An accurate measurement of RNA is therefore key for understanding the rate of expression and production. There are many methods for quantification of RNA, one of which is by OD-measurements.

Though this is precise, it requires a pure sample, and can not be performed if it is a mix of RNA and DNA [4]. The cells were therefore analyzed with NanoDrop, which allows for direct quantitative analysis of nucleotides by UV-visible spectroscopy [2]. Because NanoDrop does not require large volumes for analysis compared to OD-measurements for example, it reduces the amount of sample needed, and increases the the range of concentrations of nucleic acids that can be examined, making it an efficient method for quantification [104]. The NanoDrop was also used for the analysis of DNA, ensuring efficient measurements of DNA-content as well.

Lipid quantification can efficiently be performed gravimetrically. A challenge surrounding this protocol however, is loss of sample during the protocol. There are several steps of extraction of the organic phase using a syringe. It is important not to get any of the aqueous solution in this extraction process, because then the water-soluble components of the cell will be weighed as well. However, by rather being a bit conservative with the extraction, any loss of organic phase can be corrected for by using the density of chloroform to find the weight of organic phase that would have been extracted if the entire phase was transferred [2].

All protocols are associated with some form of correction to be made or concerns to account for. However, the main take-away from this is that it is possible to quantidy the components of the biomass of *B. subtilis*. In this project only a few of the BOF constituents were accounted for. Carbohydrates, cell-wall components or lipoteichoic acid are other macromoleculuar components it would have been interesting to quantify. Additionally several other compounds and biopolymers are also important factors in biomass productions that were not measured in these experiments. There are several cofactors, ions and metabolites that have not been quantified. They do not contribute as much to the actual biomass, but they are essential for biomass production and therefore an important part of the BOF [2]. An examination of their effect on the predictions by the model would therefore have been interesting.

5.2 Uncovering inconsistencies in *i*Bsu1147 BOF construction

The total mass quantified in these experiments after scaling the coefficients are around 69.1-69.6% of the cellular dry-weight. This is close to the mass represented by the same coefficients in the model (69.5%). At first glance this might give the impression that the oBOF represents what is seen experimentally well. However, there are some important differences to note behind these sums, and all of these will be discussed in this section. First, the distribution of the weights of the coefficients in the BOF are different. Additionally, the experiments behind the coefficients in the oBOF have not all been performed under the same conditions. Some of the experiments behind the coefficients are also performed using a recombinant strain of *B. subtilis*. In addition, the assumption behind the DNA coefficient diverge from observations in other relevant studies. The total content of DNA was assumed to be constant when con-

structing the BOF for *i*Bsu1147 [16, 13], however more recent research show that the amount of DNA increases with increasing growth rates [88]. Dauner and Sauer (2001) [13] point out in the same paper as the coefficients are adopted from that the macromolecular composition of biomass changes in response to environmental conditions, and in response to growth rate. Yet the BOF coefficients in *i*Bsu1147 are based on data from various environments and growth rates.

In these experiments the protein content was measured to be higher than the coefficient is in the model, and the reverse is true for the content of lipids. One potential reason the experimental coefficients are so different from the ones in the model, might be that the coefficients from the model are derived from experiments performed in various conditions - as was shortly discussed in chapter 4.4. The lipid components in the oBOF are derived from experiments performed with cells in stationary phase, cultivated in minimal medium. Bishop et al. (1967) found the lipid content to be 5.2% of the dry-weight of the cells, which by Dauner and Sauer was used in their model. It is from this model the coefficient for the BOF in *i*Bsu1147 comes from. The protein-coefficient was derived from experiments by Dauner and Sauer (2001). However, these experiments were performed in chemostat culture. They harvested cells at various growth rates in the exponential phase. This means that in the same BOF, two of the coefficients are harvested from cells at different phases of growth. This is not ideal, because the biomass composition composition has been shown to vary between different stages of growth. The amount of DNA, RNA, proteins, and lipids have been shown to increase as a result of increased growth rate in the exponential phase [88, 13, 95]. Stationary phase is associated with several molecular changes such as synthesis of proteins related to survival, reduction in cell size, and an increase in ratio of DNA compared to proteins [105]. In other words, the composition of the biomass of the same organism is not expected to be the same in different growth phases. The oBOF for *i*Bsu was constructed using biomass composition data derived from cell in conditions that are expected to yield distinct biomass compositions.

Not only are the conditions of the growth experiments behind the coefficients for the model not the same. The strain that has been used in the experiments also differs. In the experiments performed by Dauner and Sauer (2001), they used a recombinant, riboflavin producing strain of *B. subtilis* [13]. As biomass composition has also proven to be organism and strain specific [2, 4] this is another aspect of the BOF used in *i*Bsu1147 that draws its validity into question. The BOF consists of data from different strains, cultivated in different conditions, and harvested in different phases of growth. While the significance of the BOF for accurate model predictions is clear, it is still not necessarily formulated from biomass composition data on the organism in study [106], as the BOF in *i*Bsu1147 exemplifies.

DNA-content in the composition of *B. subtilis* in *i*Bsu1147 is considered constant at 2.6%, based on research by Bishop et al. (1967) [13]. However, these cells were harvested from stationary phase [100]. In this phase, characterized by the depletion of a necessary nutrient in the media [31, 32], the cells adapt to new stresses and divert energy away from growth. One of the consequences of this is nucleoid condensation to protect the DNA, and to fit the

chromosome in the now-smaller sized cells. The only production of proteins is limited to those only necessary for cell viability [105, 107]. In the exponential phase however, under optimal conditions, it has been observed that the cells of *B. subtilis* have 4 to 8 genomes per cell. This is referred to as mero-oligoploidity, where the number of origins is up-regulated. This up-regulation was also found to vary with growth rate. In stationary phase the number of origins was down-regulated again. This would lead to there being less DNA in the cell in stationary phase [88]. Based on these new findings, and the difference in characteristics of *B. subtilis* across different growth-phases, the validity of maintaining this BOF coefficient as is raises some concerns. The DNA content measured in these experiments revealed significant variation between growth conditions and/or growth rate, further undermining the underlying assumption that DNA-content remains constant.

5.3 The impact of carbon source on biomass composition

The quantification of the biomass of *B. subtilis* showed, as discussed in chapter 4.3, that there is significant variance in biomass depending on the growth condition. It was expected to observe variations, as this is something that has been reported previously [2, 3, 83]. Although the content of lipids remained consistent across the different conditions, the DNA-, RNA-, and protein content varied significantly. The amino acid composition showed some variance as well. Adapting to different carbon sources requires a metabolic reorganization in the cell [31, 32], which often involve making alterations in gene expression. For example, succinate will require a shift in metabolism such that the gluconeogenesis and pentose phosphate pathway are active in order to produce metabolic precursors as well as energy [19]. In other words, metabolic redistribution happens in adaption to the media, consequently influencing the composition of the cell.

The calculated growth rates that *B. subtilis* was able to achieve in these batch-fermentations varied with the carbon source available as well. It is difficult to conclude whether the carbon source available or the growth rate achieved has the impact on biomass composition without conducting additional experiments. However, these experiments show that the growth rate was significantly different when *B. subtilis* was cultivated with different carbon sources. This is likely result of the various transportation methods that can be used to import the carbon source [87, 19, 24, 15, 22, 26], as discussed in section 4.2.2, or due to different energy requirements associated with maintaining metabolism on different sources of carbon [86]. Considering that biomass composition, as indicated in multiple research articles [95, 88, 13] is known to vary with growth rate, the observed differences in growth rates measured in these experiments with different sources of carbon further support the concept of condition-specific BOFs. For instance, if *B. subtilis* can not achieve a high growth rate on xylose but can do so in glucose, and this leads to different biomass compositions, then the argument for condition specific BOFs remain valid.

5.4 The effect of updated BOF on model predictions and a call for condition specific BOFs

As observed in the results from the FBA and FVA, there is a change in flux distribution for certain groups of reactions - especially those related to lipids and exchange reactions - in adaption to the nBOFs. Because the BOF sets the objective of the cell, and defines the weight of which its components must be synthesised for biomass production, the coefficients in the BOF alter the down-stream distribution of flux through the model [4]. The internal fluxes have been shown to be sensitive to changes in the biomass composition implemented by the BOF. Experimentation performed on a GEM for *E. coli* found that the internal fluxes are especially sensitive to changes in protein- and lipid composition when it came to macromolecular changes. The experiments also showed that the flux distribution is sensitive to changes is amino acid composition [3]. The protein and lipid compositions found experimentally were significantly different from the coefficients in the model, and so was the amino acid distribution. Therefore, it was expected to see variation in the flux distributions between oBOFs and nBOFs.

However, the BOF contains many components that were not measured in these experiments. The BOF was only updated in the DNA, RNA, protein and lipid coefficient. Further, only the composition of amino acids was examined. The composition of lipid-groups was not. As mentioned this affects the distribution of fluxes in the model. When the composition remains the same but the amount is changed, there is a shift in the fluxes, but the flux is not redistributed between the lipid-groups. The carbohydrate content was not measured, and neither were other metabolites that contribute to biomass production. While alterations have been made to the oBOF to make it specific to the conditions *B. subtilis* was cultivated in, several other components that also might affect model predictions were not measured.

Though the oBOF coefficients have been derived from various conditions and two different strains, the growth rate predictions seem robust. pFBA was performed on the model in the different media compositions, comparing the flux distribution with the nBOF to the oBOF. Though the flux distribution through the model is changed when the nBOF is implemented, the growth rate predictions were always the equal between the two BOFs for each condition. As shortly discussed in section 4.6.2, Lavoie et al. (2020) found that growth rate predictions in GEMs were robust to changes in biomass composition, which is what is implemented with the BOF. Changes in the carbon-uptake rate however, affected the growth rate predictions [101]. In this case, the pFBA analyses were only conducted between nBOF and oBOF with the same media composition and the same carbon uptake rate.

Changes in environment forces the cells to adapt on a transcriptional level, by adjusting the expression of genes the productions of compounds is altered to be able to utilize the available resources for vitality and proliferation [106, 19]. The argument can be made that measuring the composition for every relevant environmental condition is unrealistic [106]. However, the significant variance in biomass composition and growth rates in different media

composition, and the subsequent effect on flux distributions highlight the significance of constructing condition specific BOFs.

6. Conclusion and outlook

The biomass composition of bacteria is known to vary between organism, strain and conditions. However, the coefficients for this central equation to phenotype predictions in GEMs are often derived from research on other closely related strains or from other conditions [2, 4]. In the case of the GEM *i*Bsu1147, the BOF coefficients have been derived from multiple experiments, and so the coefficients are based on two strains, bacteria cultivated by different methods and harvested in different phases of growth [14, 16, 13].

Following the pipeline by Simensen et al.(2021), the biomass composition of *Bacillus subtilis* was quantified after cultivation by batch-fermentation in minimal media containing different sources of carbon - glucose, glycerol, mannitol, xylose and succinate. The cells were harvested in the exponential growth phase and subsequently analyzed. The macromolecular compounds that were analysed were DNA, RNA, protein and lipids, which are all components in the BOF. The analysis of biomass composition of *B. subtilis* cultivated with different carbon sources show significant variation, further demonstrating the dynamic nature of biomass composition. The resulting g/gDW found of DNA, RNA, protein and lipids were implemented in the model by creating five new condition-specific BOFs.

The total mass quantified in these experiments after scaling the coefficients are around 69.1-69.6% of the cellular dry-weight. This is close to the mass represented by the same coefficients in *i*Bsu1147 (69.5%). However, the distribution of weights in the original BOF is significantly different from the weights in the new BOFs. The levels of lipids were found to be lower experimentally, and the measured protein content was higher. The new conditions-specific BOFs were implemented in the model, and pFBA and FVA was performed to see if and how they affected model predictions. The growth rate predictions proved to be quite robust, however the flux rates and distribution throughout the model were more sensitive to changes in biomass composition, implemented in the model with the nBOFs.

Further analyses of *i*Bsu1147 with the new BOFs would be necessary to conclude any further on the specifics of how these variances in flux distributions affect the phenotype predictions. However, the variances in biomass between growth conditions is significant, and it is possible to measure it. It might be time consuming, however these results illustrate how the experimentally measured, conditions-specific BOFs affect the flux distribution in the model predictions. They are also all significantly different from the original BOF. Implementing BOFs based on experimentally measured biomass compositions, and condition-specific BOFs, would help increase the specificity of the BOF and thereby improve the validity of GEM phenotype predictions.

To improve the strength of these results, it would have been beneficial with biological replicates from the different carbon sources. There is variation associated with batch-

fermentations [34], and biological variation is usually greater than technical variation [103]. To substantiate the reliability of the results from this project of the significant difference in biomass composition as a result of the conditions of cultivation, biological replicates are key.

In addition to DNA, RNA, proteins and lipids, the BOF consists of numerous other components. By measuring the content of these additional components the BOF would be further specified and provide an even better representation of the cells biomass composition. Carbohydrates represent another macromolecular constituent [2, 4]. While the carbohydrate layer for *E. coli* was found to be around 7.2% of the biomass, it is expected to be higher for the gram-positive *B. subtilis* because of the thicker peptidoglycan layer of the cell wall [108, 13]. It was planned to perform an analysis following the protocol by Rühman et al. (2016) [109], as part of the pipeline by Simensen et al. (2021) [2], however the pipeline for mass-spectrometry analysis was not ready in time for the protocol to be performed in time.

All of the energy in a cell cannot be allocated towards growth. Other processes require energy as well, such as motility, cellular osmolarity maintenance, correction of errors [43] by proofreading followed by repairs of for example proteins or RNA [110]. The energy required for these processes is represented in GEMs by the maintenance coefficient. The maintenance energy required impacts the biomass yield and overall resource allocation in the cell [43]. The flux through the model is therefore distributed to satisfy both the objective, which in this case is growth, but also the cellular maintenance [111]. The maintenance energy in *i*Bsu1147 was set according to the energy coefficient of *B. subtilis* 168, and is a static component in the model [14]. However, it has been observed in *E. coli* that maintenance energy is lower at higher growth rates and higher in unfavorable environments with lower growth rates [112]. It would have been interesting to explore the potential effects of the maintenance energy on the model predictions, and whether presuming a uniform maintenance requirement is a valid assumption.

Bibliography

[1] Blas Blázquez et al. "New Insights on Metabolic Features of Bacillus subtilis Based on Multistrain Genome-Scale Metabolic Modeling". In: *International Journal of Molecular Sciences* 24.8 (Apr. 2023). ISSN: 1422-0067. DOI: 10.3390/ijms24087091.

[2] Vetle Simensen et al. "Quantification of macromolecular biomass composition for constraint-based metabolic modeling". In: *bioRxiv* (Jan. 2021), p. 2021.08.20.457062. DOI: 10.1101/2021.08.20.457062. URL: http://biorxiv.org/content/early/2021/08/20/2021.08.20.457062.abstract.

[3] Yoon-Mi Choi et al. "Mitigating biomass composition uncertainties in flux balance analysis using ensemble representations". In: *bioRxiv* (Mar. 2022).

[4] Ashley Beck, Kristopher Hunt, and Ross Carlson. "Measuring Cellular Biomass Composition for Computational Biology Applications". In: *Processes* 6.5 (Apr. 2018), pp. 38–65. ISSN: 2227-9717. DOI: 10.3390/pr6050038.

[5] Anthony Trewavas. "A Brief History of Systems Biology". In: *The Plant Cell* 18.10 (Oct. 2006), pp. 2420–2430. ISSN: 1040-4651. DOI: 10.1105/tpc.106.042267.

[6] Anurag Passi et al. "Genome-Scale Metabolic Modeling Enables In-Depth Understanding of Big Data". In: *Metabolites* 12.14 (Dec. 2021). ISSN: 2218-1989. DOI: 10.3390/metabo12010014.

[7] Bernhard Palsson. "Metabolic systems biology". In: *FEBS Letters* 583.24 (Dec. 2009), pp. 3900–3904. ISSN: 00145793. DOI: 10.1016/j.febslet.2009.09.031.

[8] Bernhard Ø. Palsson. *Systems Biology: Properties of Reconstructed Networks*. 1st ed. 2006.

[9] Mariano Bizzarri, Alessandro Palombo, and Alessandra Cucina. "Theoretical aspects of Systems Biology". In: *Progress in Biophysics and Molecular Biology* 112.1-2 (May 2013), pp. 33–43. ISSN: 00796107. DOI: 10.1016/j.pbiomolbio.2013.03.019.

[10] Adam M. Feist et al. "Reconstruction of biochemical networks in microorganisms". In: *Nature Reviews Microbiology* 7 (Feb. 2009), pp. 129–143. ISSN: 1740-1526. DOI: 10.1038/nrmicro1949.

[11] Jean-Christophe Lachance et al. "BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data". In: *PLOS Computational Biology* 15.4 (Apr. 2019), e1006971. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006971.

[12] Adam M Feist and Bernhard O Palsson. "The biomass objective function". In: *Current Opinion in Microbiology* 13.3 (June 2010), pp. 344–349. ISSN: 13695274. DOI: 10.1016/j.mib.2010.03.003.

[13] Michael Dauner and Uwe Sauer. "Stoichiometric growth model for riboflavin-producingBacillus subtilis". In: *Biotechnology and Bioengineering* 76.2 (Sept. 2001), pp. 132–143. ISSN: 0006-3592. DOI: 10.1002/bit.1153.

[14] Tong Hao et al. "In silico metabolic engineering of Bacillus subtilis for improved production of riboflavin, Egl-237, (R,R)-2,3-butanediol and isobutanol". In: *Molecular BioSystems* 9.8 (2013), p. 2034. ISSN: 1742-206X. DOI: 10.1039/c3mb25568a.

[15] Seo A. Park et al. "Bacillus subtilis as a robust host for biochemical production utilizing biomass". In: *Critical Reviews in Biotechnology* 41.6 (Aug. 2021), pp. 827–848. ISSN: 0738-8551. DOI: 10.1080/07388551.2021.1888069.

[16] You-Kwan Oh et al. "Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data". In: *Journal of Biological Chemistry* 282.39 (Sept. 2007), pp. 28791–28799. ISSN: 00219258. DOI: 10.1074/jbc.M703759200.

[17] F. Kunst et al. "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis". In: *Nature* 390 (Nov. 1997), pp. 249–256. ISSN: 0028-0836. DOI: 10.1038/36786.

[18] Erhard Bremer et al. "A model industrial workhorse: ¡i¿Bacillus subtilis¡/i¿ strain 168 and its genome after a quarter of a century". In: *Microbial Biotechnology* (Apr. 2023), pp. 1–29. ISSN: 1751-7915. DOI: 10.1111/1751-7915.14257.

[19] Oliver Schilling et al. "Transcriptional and Metabolic Responses of ¡i¿Bacillus subtilis¡/i¿ to the Availability of Organic Acids: Transcription Regulation Is Important but Not Sufficient To Account for Metabolic Adaptation". In: *Applied and Environmental Microbiology* 73.2 (Jan. 2007), pp. 499–507. ISSN: 0099-2240. DOI: 10.1128/AEM.02084-06.

[20] Frederik M. Meyer et al. "Malate-Mediated Carbon Catabolite Repression in Bacillus subtilis Involves the HPrK/CcpA Pathway". In: *Journal of Bacteriology* 193.24 (Dec. 2011), pp. 6939–6949. ISSN: 0021-9193. DOI: 10.1128/JB.06197-11.

[21] Xin Wang et al. "Growth strategy of microbes on mixed carbon sources". In: *Nature Communications* 10 (Mar. 2019). ISSN: 2041-1723. DOI: 10.1038/s41467-019-09261-3.

[22] Birthe Halmschlag et al. "Comparison of Isomerase and Weimberg Pathway for γ-PGA Production From Xylose by Engineered Bacillus subtilis". In: *Frontiers in Bioengineering and Biotechnology* 7 (Jan. 2020). ISSN: 2296-4185. DOI: 10.3389/fbioe.2019.00476.

[23] H.W Wisselink et al. "Mannitol production by lactic acid bacteria: a review". In: *International Dairy Journal* 12.2-3 (Jan. 2002), pp. 151–161. ISSN: 09586946. DOI: 10.1016/S0958-6946(01)00153-4.

[24] Thanh Nguyen et al. “Targeting Mannitol Metabolism as an Alternative Antimicrobial Strategy Based on the Structure-Function Study of Mannitol-1-Phosphate Dehydrogenase in Staphylococcus aureus”. In: *mBio* 10.4 (Aug. 2019). ISSN: 2161-2129. DOI: 10.1128/mBio.02660-18.

[25] Viveka Lindgren and Lars Rutberg. “Glycerol Metabolism in ¡i¿Bacillus subtilis¡/i¿ : Gene-Enzyme Relationships”. In: *Journal of Bacteriology* 119.2 (Aug. 1974), pp. 431–442. ISSN: 0021-9193. DOI: 10.1128/jb.119.2.431-442.1974.

[26] Marieke F. Buffing et al. “Capacity for instantaneous catabolism of preferred and non-preferred carbon sources in Escherichia coli and Bacillus subtilis”. In: *Scientific Reports* 8 (Aug. 2018). ISSN: 2045-2322. DOI: 10.1038/s41598-018-30266-3.

[27] Yong-Cheol Park, Soo Young Jun, and Jin-Ho Seo. “Construction and characterization of recombinant Bacillus subtilis JY123 able to transport xylose efficiently”. In: *Journal of Biotechnology* 161.4 (Nov. 2012), pp. 402–406. ISSN: 01681656. DOI: 10.1016/j.jbiotec.2012.07.192.

[28] L.E. Erickson. “Bioreactors”. In: *Encyclopedia of Microbiology*. Ed. by Moselio Schaechter. 3rd ed. Elsevier, 2009, pp. 206–211. DOI: 10.1016/B978-012373944-5.00136-X.

[29] G. Lidén. “Understanding the bioreactor”. In: *Bioprocess and Biosystems Engineering* 24 (Sept. 2001), pp. 273–279. ISSN: 1615-7591. DOI: 10.1007/s004490100263.

[30] Shijie Liu. *Bioprocess Engineering - Kinetics, Sustainability, and Reactor Design*. 2nd ed. Elsevier, 2017, pp. 1007–1058. ISBN: 9780444637833. DOI: 10.1016/C2015-0-04891-2.

[31] Ali Cinar, Satish J. Parulekar, and Cenk Ûndey. *Batch fermentation: Modeling, monitoring and control*. New York: Marcel Dekker, 2003.

[32] Peter F. Stanbury, Allan Whitaker, and Stephen J. Hall. *Principles of fermentation technology*. 3rd ed. Elsevier, 2017, pp. 1–19.

[33] Matthew D. Rolfe et al. “Lag Phase Is a Distinct Growth Phase That Prepares Bacteria for Exponential Growth and Involves Transient Metal Accumulation”. In: *Journal of Bacteriology* 194.3 (Feb. 2012), pp. 686–701. ISSN: 0021-9193. DOI: 10.1128/JB.06112-11.

[34] Shijie Liu. “Cell Cultivation”. In: *Bioprocess Engineering*. Elsevier, 2017, pp. 699–782. DOI: 10.1016/B978-0-444-63783-3.00012-5.

[35] Michal Komorniczak. *Bacterial growth*. 2009. URL: https://commons.wikimedia.org/wiki/File:Bacterial_growth_en.svg.

[36] Barry R. Bochner. “Global phenotypic characterization of bacteria”. In: *FEMS Microbiology Reviews* 33.1 (Jan. 2009), pp. 191–205. ISSN: 1574-6976. DOI: 10.1111/j.1574-6976.2008.00149.x.

[37] Hiroaki Kitano. “Systems Biology: A Brief Overview”. In: *Science* 295.5560 (Mar. 2002), pp. 1662–1664. ISSN: 0036-8075. DOI: 10.1126/science.1069492.

[38] Aarash Bordbar et al. “Constraint-based models predict metabolic and associated cellular functions”. In: *Nature Reviews Genetics* 15 (Feb. 2014), pp. 107–120. ISSN: 1471-0056. DOI: 10.1038/nrg3643.

[39] Daniel A. Cuevas et al. “From DNA to FBA: How to Build Your Own Genome-Scale Metabolic Model”. In: *Frontiers in Microbiology* 7 (June 2016). ISSN: 1664-302X. DOI: 10.3389/fmicb.2016.00907.

[40] Ilyas Kabimoldayev et al. “Basics of genome-scale metabolic modeling and applications on C1-utilization”. In: *FEMS Microbiology Letters* 365.20 (Oct. 2018). ISSN: 1574-6968. DOI: 10.1093/femsle/fny241.

[41] K. Raman and N. Chandra. “Flux balance analysis of biological systems: applications and challenges”. In: *Briefings in Bioinformatics* 10.4 (July 2009), pp. 435–449. ISSN: 1467-5463. DOI: 10.1093/bib/bbp011.

[42] Jeffrey D Orth, Ines Thiele, and Bernhard Ø Palsson. “What is flux balance analysis?” In: *Nature Biotechnology* 28 (Mar. 2010), pp. 245–248. ISSN: 1087-0156. DOI: 10.1038/nbt.1614.

[43] Bernhard O. Palsson. *Systems biology: Constraint-Based Reconstruction and Analysis*. 1st ed. Cambridge University Press, 2015.

[44] Ali Ebrahim et al. “COBRApy: COnstraints-Based Reconstruction and Analysis for Python”. In: *BMC Systems Biology* 7 (Dec. 2013). ISSN: 1752-0509. DOI: 10.1186/1752-0509-7-74.

[45] J. M. Lee, E. P Gianchandani, and J. A. Papin. “Flux balance analysis in the era of metabolomics”. In: *Briefings in Bioinformatics* 7.2 (Apr. 2006), pp. 140–150. ISSN: 1467-5463. DOI: 10.1093/bib/bbl007.

[46] Daniel Segrè, Dennis Vitkup, and George M. Church. “Analysis of optimality in natural and perturbed metabolic networks”. In: *Proceedings of the National Academy of Sciences* 99.23 (Nov. 2002), pp. 15112–15117. ISSN: 0027-8424. DOI: 10.1073/pnas.232349399.

[47] Steinn Gudmundsson and Ines Thiele. “Computationally efficient flux variability analysis”. In: *BMC Bioinformatics* 11 (Sept. 2010), p. 489. ISSN: 1471-2105. DOI: 10.1186/1471-2105-11-489.

[48] Dustin Kenefake et al. “An improved algorithm for flux variability analysis”. In: *BMC Bioinformatics* 23 (Dec. 2022). ISSN: 1471-2105. DOI: 10.1186/s12859-022-05089-9.

[49] Joonhoon Kim and Jennifer L Reed. “RELATCH: relative optimality in metabolic networks explains robust metabolic and regulatory responses to perturbations”. In: *Genome Biology* 13 (2012). ISSN: 1465-6906. DOI: 10.1186/gb-2012-13-9-r78.

[50] Siu H J Chan et al. “Standardizing biomass reactions and ensuring complete mass balance in genome-scale metabolic models”. In: *Bioinformatics* 33.22 (Nov. 2017), pp. 3603–3609. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btx453.

[51] U Sauer et al. “Physiology and metabolic fluxes of wild-type and riboflavin-producing Bacillus subtilis”. In: *Applied and Environmental Microbiology* 62.10 (Oct. 1996), pp. 3687–3696. ISSN: 0099-2240. DOI: 10.1128/aem.62.10.3687-3696.1996.

[52] Kouji Matsumoto et al. “Cloning, Sequencing, and Disruption of the ¡i¿Bacillus subtilis psd¡/i¿ Gene Coding for Phosphatidylserine Decarboxylase”. In: *Journal of Bacteriology* 180.1 (Jan. 1998), pp. 100–106. ISSN: 0021-9193. DOI: 10.1128/JB.180.1.100-106.1998.

[53] Christopher S Henry et al. “iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations”. In: *Genome Biology* 10 (2009). ISSN: 1465-6906. DOI: 10.1186/gb-2009-10-6-r69.

[54] Fredrik S. Skedsmo et al. “Demyelinating polyneuropathy in goats lacking prion protein”. In: *The FASEB Journal* 34.2 (Feb. 2020), pp. 2359–2375. ISSN: 0892-6638. DOI: 10.1096/fj.201902588R.

[55] E. G. Bligh and W. J. Dyer. “A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION”. In: *Canadian Journal of Biochemistry and Physiology* 37.8 (Aug. 1959), pp. 911–917. ISSN: 0576-5544. DOI: 10.1139/o59-099.

[56] Preetha J Shetty and Mariyam Dairawan. “The Evolution of DNA Extraction Methods”. In: *American Journal of Biomedical Science & Research* 8.1 (Mar. 2020), pp. 39–45. ISSN: 26421747. DOI: 10.34297/AJBSR.2020.08.001234.

[57] WILLIAM G. ALDRIDGE and MICHAEL L. WATSON. “PERCHLORIC ACID EXTRACTION AS A HISTOCHEMICAL TECHNIQUE”. In: *Journal of Histochemistry & Cytochemistry* 11.6 (Nov. 1963), pp. 773–781. ISSN: 0022-1554. DOI: 10.1177/11.6.773.

[58] J. E. Noble et al. “A Comparison of Protein Quantitation Assays for Biopharmaceutical Applications”. In: *Molecular Biotechnology* 37.2 (Sept. 2007), pp. 99–111. ISSN: 1073-6085. DOI: 10.1007/s12033-007-0038-9.

[59] Muhammad Sajid Hamid Akash and Kanwal Rehman. “Introduction to Chromatographic Techniques”. In: *Essentials of Pharmaceutical Analysis*. Singapore: Springer Nature Singapore, 2020, pp. 147–156. DOI: 10.1007/978-981-15-1547-7{_}11.

[60] Yasser M. Moustafa and Rania E. Morsi. *Column Chromatography*. Ed. by Dean Martin and Barbara B. Martin. InTech, Apr. 2013, pp. 1–31. ISBN: 978-953-51-1074-3. DOI: 10.5772/47823.

[61] Ozlem Coskun. “Separation Tecniques: CHROMATOGRAPHY”. In: *Northern Clinics of Istanbul* 3.2 (2016), pp. 156–160. ISSN: 21484902. DOI: 10.14744/nci.2016.32757.

[62] Rishabha Malviya et al. “High performance liquid chromaatography: A short review”. In: *Journal of Global Pharma Technology* 2.5 (June 2010), pp. 22–26.

[63] Toshiaki Okuno and Takehiko Yokomizo. “Basic Techniques for Lipid Extraction from Tissues and Cells”. In: *Bioactive Lipid Mediators*. Tokyo: Springer Japan, 2015, pp. 331–336. DOI: 10.1007/978-4-431-55669-5{_}23.

[64] P. J. Hore. *Nuclear Magnetic Resonance*. 1st ed. New york: Oxford University Press, 1995. ISBN: 0198556829.

[65] Sofia Moco. "Studying Metabolism by NMR-Based Metabolomics". In: *Frontiers in Molecular Biosciences* 9 (Apr. 2022). ISSN: 2296-889X. DOI: 10.3389/fmolb.2022.882487.

[66] Vladimír Mlynárik. "Introduction to nuclear magnetic resonance". In: *Analytical Biochemistry* 529 (July 2017), pp. 4–9. ISSN: 00032697. DOI: 10.1016/j.ab.2016.05.006.

[67] Jane L. Wagstaff et al. "1H NMR Spectroscopy Profiling of Metabolic Reprogramming of Chinese Hamster Ovary Cells upon a Temperature Shift during Culture". In: *PLoS ONE* 8.10 (Oct. 2013). ISSN: 1932-6203. DOI: 10.1371/journal.pone.0077195.

[68] Homer F. Swift. "PRESERVATION OF STOCK CULTURES OF BACTERIA BY FREEZING AND DRYING". In: *Journal of Experimental Medicine* 33.1 (Jan. 1921), pp. 69–75. ISSN: 1540-9538. DOI: 10.1084/jem.33.1.69.

[69] Michael R. Green and Joseph Sambrook. "Caring for ¡i¿Escherichia coli¡/i¿". In: *Cold Spring Harbor Protocols* 2018.2 (Feb. 2018). ISSN: 1940-3402. DOI: 10.1101/pdb.prot101337.

[70] Caroline Krogh Søgaard et al. ""Two hits - one stone"; increased efficacy of cisplatin-based therapies by targeting PCNA's role in both DNA repair and cellular signaling". In: *Oncotarget* 9.65 (Aug. 2018), pp. 32448–32465. ISSN: 1949-2553. DOI: 10.18632/oncotarget.25963.

[71] Mitchell Henry Wright, Joseph Adelskov, and Anthony Carlson Greene. "Bacterial DNA Extraction Using Individual Enzymes and Phenol/Chloroform Separation". In: *Journal of Microbiology & Biology Education* 18.2 (Sept. 2017). ISSN: 1935-7877. DOI: 10.1128/jmbe.v18i2.1348.

[72] S. Benthin, J. Nielsen, and J. Villadsen. "A simple and reliable method for the determination of cellular RNA content". In: *Biotechnology Techniques* 5.1 (1991), pp. 39–42. ISSN: 0951-208X. DOI: 10.1007/BF00152753.

[73] Peter. Lindroth and Kenneth. Mopper. "High performance liquid chromatographic determination of subpicomole amounts of amino acids by precolumn fluorescence derivatization with o-phthaldialdehyde". In: *Analytical Chemistry* 51.11 (Sept. 1979), pp. 1667–1674. ISSN: 0003-2700. DOI: 10.1021/ac50047a019.

[74] Zdenka Bartosova et al. "Combined Metabolome and Lipidome Analyses for In-Depth Characterization of Lipid Accumulation in the DHA Producing Aurantiochytrium sp. T66". In: *Metabolites* 11.3 (Feb. 2021). ISSN: 2218-1989. DOI: 10.3390/metabo11030135.

[75] Daniel Machado, Nicolas Ochsner, and Rodrigo Amarante Colpo. *reframed 1.4.0*. 2023. DOI: https://doi.org/10.5281/zenodo.7955995.

[76] Nathan E Lewis et al. "Omic data from evolved ¡i¿E. coli¡/i¿ are consistent with computed optimal growth from genome-scale models". In: *Molecular Systems Biology* 6.1 (Jan. 2010). ISSN: 1744-4292. DOI: 10.1038/msb.2010.47.

[77] Minoru Kanehisa et al. “KEGG for taxonomy-based analysis of pathways and genomes”. In: *Nucleic Acids Research* 51.D1 (Jan. 2023), pp. D587–D592. ISSN: 0305-1048. DOI: 10.1093/nar/gkac963.

[78] M. Kanehisa and S. Goto. “KEGG: Kyoto Encyclopedia of Genes and Genomes”. In: *Nucleic Acids Research* 28.1 (Jan. 2000), pp. 27–30. ISSN: 13624962. DOI: 10.1093/nar/28.1.27.

[79] Minoru Kanehisa. “Toward understanding the origin and evolution of cellular organisms”. In: *Protein Science* 28.11 (Nov. 2019), pp. 1947–1951. ISSN: 0961-8368. DOI: 10.1002/pro.3715.

[80] Stefany Moreno-Gámez et al. “Wide lag time distributions break a trade-off between reproduction and survival in bacteria”. In: *Proceedings of the National Academy of Sciences* 117.31 (Aug. 2020), pp. 18729–18736. ISSN: 0027-8424. DOI: 10.1073/pnas.2003331117.

[81] Thomas Egli. “Microbial growth and physiology: a call for better craftsmanship”. In: *Frontiers in Microbiology* 6 (Apr. 2015). ISSN: 1664-302X. DOI: 10.3389/fmicb.2015.00287.

[82] John T. Sauls et al. “Control of Bacillus subtilis Replication Initiation during Physiological Transitions and Perturbations”. In: *mBio* 10.6 (Dec. 2019). ISSN: 2161-2129. DOI: 10.1128/mBio.02205-19.

[83] M. Schaechter, O. MaalOe, and N. O. Kjeldgaard. “Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium”. In: *Journal of General Microbiology* 19.3 (Dec. 1958), pp. 592–606. ISSN: 0022-1287. DOI: 10.1099/00221287-19-3-592.

[84] Wei-Hsiang Lin et al. “Origin of exponential growth in nonlinear reaction networks”. In: *Proceedings of the National Academy of Sciences* 117.45 (Nov. 2020), pp. 27795–27804. ISSN: 0027-8424. DOI: 10.1073/pnas.2013061117.

[85] Yane Luo, Tao Zhang, and Hui Wu. “The transport and mediation mechanisms of the common sugars in Escherichia coli”. In: *Biotechnology Advances* 32.5 (Sept. 2014), pp. 905–919. ISSN: 07349750. DOI: 10.1016/j.biotechadv.2014.04.009.

[86] Silvio Waschina et al. “Metabolic network architecture and carbon source determine metabolite production costs”. In: *The FEBS Journal* 283 (June 2016), pp. 2149–2163. ISSN: 1742464X. DOI: 10.1111/febs.13727.

[87] Aida Kalantari et al. “Conversion of Glycerol to 3-Hydroxypropanoic Acid by Genetically Engineered Bacillus subtilis”. In: *Frontiers in Microbiology* 8 (Apr. 2017). ISSN: 1664-302X. DOI: 10.3389/fmicb.2017.00638.

[88] Benjamin Böttinger et al. “Regulated ploidy of Bacillus subtilis and three new isolates of Bacillus and Paenibacillus”. In: *FEMS Microbiology Letters* 365.4 (Feb. 2018). ISSN: 1574-6968. DOI: 10.1093/femsle/fnx282.

[89] Michaela E. Sharpe et al. "¡i¿Bacillus subtilis¡/i¿ Cell Cycle as Studied by Fluorescence Microscopy: Constancy of Cell Length at Initiation of DNA Replication and Evidence for Active Nucleoid Partitioning". In: *Journal of Bacteriology* 180.3 (Feb. 1998), pp. 547–555. ISSN: 0021-9193. DOI: 10.1128/JB.180.3.547-555.1998.

[90] Chris D. Webb et al. "Use of time-lapse microscopy to visualize rapid movement of the replication origin region of the chromosome during the cell cycle in Bacillus subtilis". In: *Molecular Microbiology* 28.5 (June 1998), pp. 883–892. ISSN: 0950-382X. DOI: 10.1046/j.1365-2958.1998.00808.x.

[91] Ryosuke Kadoya et al. "Two separate DNA sequences within oriC participate in accurate chromosome segregation in Bacillus subtilis". In: *Molecular Microbiology* 45.1 (July 2002), pp. 73–87. ISSN: 0950-382X. DOI: 10.1046/j.1365-2958.2002.03016.x.

[92] Shigeki Moriya et al. "Effects of oriC relocation on control of replication initiation in Bacillus subtilis". In: *Microbiology* 155 (Sept. 2009), pp. 3070–3082. ISSN: 1350-0872. DOI: 10.1099/mic.0.030080-0.

[93] Anne-Gaëlle Planson et al. "Bacterial growth physiology and RNA metabolism". In: *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms* 1863.5 (May 2020). ISSN: 18749399. DOI: 10.1016/j.bbagrm.2020.194502.

[94] Vetle Simensen, Yara Seif, and Eivind Almaas. "Phenotypic response of yeast metabolic network to availability of proteinogenic amino acids". In: *Frontiers in Molecular Biosciences* 9 (Aug. 2022). ISSN: 2296-889X. DOI: 10.3389/fmolb.2022.963548.

[95] Zhichao Zhang et al. "Quantitative Connection between Cell Size and Growth Rate by Phospholipid Metabolism". In: *Cells* 9.2 (Feb. 2020), p. 391. ISSN: 2073-4409. DOI: 10.3390/cells9020391.

[96] John T. Sauls et al. "Gram-positive and Gram-negative Bacteria Share Common Principles to Coordinate Growth and the Cell Cycle at the Single-cell Level". In: *bioRxiv* (Aug. 2019). DOI: https://doi.org/10.1101/726596.

[97] Fabrizio Carteni et al. "A General Process-Based Model for Describing the Metabolic Shift in Microbial Cell Cultures". In: *Frontiers in Microbiology* 11 (Sept. 2020). ISSN: 1664-302X. DOI: 10.3389/fmicb.2020.521368.

[98] Benedikt Heyman et al. "Online monitoring of the respiratory quotient reveals metabolic phases during microaerobic 2,3-butanediol production with ¡i¿Bacillus licheniformis¡/i¿". In: *Engineering in Life Sciences* 20.3-4 (Mar. 2020), pp. 133–144. ISSN: 1618-0240. DOI: 10.1002/elsc.201900121.

[99] Oliver Dilly. "Regulation of the respiratory quotient of soil microbiota by availability of nutrients". In: *FEMS Microbiology Ecology* 43.3 (Apr. 2003), pp. 375–381. ISSN: 01686496. DOI: 10.1111/j.1574-6941.2003.tb01078.x.

[100] D. G. Bishop, L. Rutberg, and B. Samuelsson. "The Chemical Composition of the Cytoplasmic Membrane of Bacillus subtilis". In: *European Journal of Biochemistry* 2.4 (Nov. 1967), pp. 448–453. ISSN: 0014-2956. DOI: 10.1111/j.1432-1033.1967.tb00158.x.

[101] Michel Lavoie et al. "Genome-Scale Metabolic Reconstruction and in Silico Perturbation Analysis of the Polar Diatom Fragilariopsis cylindrus Predicts High Metabolic Robustness". In: *Biology* 9.2 (2020). ISSN: 2079-7737. DOI: 10.3390/biology9020030.

[102] Haruki Ishizaki and Keiji Hasumi. "Ethanol Production from Biomass". In: *Research Approaches to Sustainable Biomass Systems*. Elsevier, 2014, pp. 243–258. DOI: 10.1016/B978-0-12-404609-2.00010-6.

[103] Paul Blainey, Martin Krzywinski, and Naomi Altman. "Replication". In: *Nature Methods* 11 (Aug. 2014), pp. 879–880. ISSN: 1548-7091. DOI: 10.1038/nmeth.3091.

[104] Philippe Desjardins and Deborah Conklin. "NanoDrop Microvolume Quantitation of Nucleic Acids". In: *Journal of Visualized Experiments* (Nov. 2010). ISSN: 1940-087X. DOI: 10.3791/2565.

[105] Jananee Jaishankar and Preeti Srivastava. "Molecular Basis of Stationary Phase Survival and Applications". In: *Frontiers in Microbiology* 8 (Oct. 2017). ISSN: 1664-302X. DOI: 10.3389/fmicb.2017.02000.

[106] Christian Schulz et al. "Genome-scale metabolic modelling when changes in environmental conditions affect biomass composition". In: *PLOS Computational Biology* 17.5 (May 2021). ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1008528.

[107] Joanna Hołówka and Jolanta Zakrzewska-Czerwińska. "Nucleoid Associated Proteins: The Small Organizers That Help to Cope With Stress". In: *Frontiers in Microbiology* 11 (Apr. 2020). ISSN: 1664-302X. DOI: 10.3389/fmicb.2020.00590.

[108] Thomas. J. Silhavy, Daniel. Kahne, and Suzanne. Walker. "The Bacterial Cell Envelope". In: *Cold Spring Harbor Perspectives in Biology* 2.5 (May 2010). ISSN: 1943-0264. DOI: 10.1101/cshperspect.a000414.

[109] Broder Rühmann, Jochen Schmid, and Volker Sieber. "Automated Modular High Throughput Exopolysaccharide Screening Platform Coupled with Highly Sensitive Carbohydrate Fingerprint Analysis". In: *Journal of Visualized Experiments* (Apr. 2016). ISSN: 1940-087X. DOI: 10.3791/53249.

[110] Christopher P. Kempes et al. "Drivers of Bacterial Maintenance and Minimal Energy Requirements". In: *Frontiers in Microbiology* 8 (Jan. 2017). ISSN: 1664-302X. DOI: 10.3389/fmicb.2017.00031.

[111] Edward J. O'Brien, Jonathan M. Monk, and Bernhard O. Palsson. "Using Genome-scale Models to Predict Biological Capabilities". In: *Cell* 161.5 (May 2015), pp. 971–987. ISSN: 00928674. DOI: 10.1016/j.cell.2015.05.019.

[112] Nabil Nancib, Ridha Mosrati, and Joseph Boudrant. "Modelling of batch fermentation of a recombinant Escherichia coli producing glyceraldehyde-3-phosphate dehydrogenase on a complex selective medium". In: *The Chemical Engineering Journal* 52.2 (Aug. 1993), B35–B48. ISSN: 03009467. DOI: 10.1016/0300-9467(93)80056-T.

Appendix A - Bradford protein assay

The first step was to make the standard curve. This was done by using a BSA protein standard (100µg/mL). The BSA stock was used to make a series of dilutions with MQ-water, as listed in Figure 1 below. There were three replicates from each concentration. The average of all three replicates per concentration were used to make the standard curve. The equation for the regression line between these points was used to calculate the concentration of proteins in the bacterial samples later. When measuring the OD for the replicates, the OD is set to zero only with the first blank.

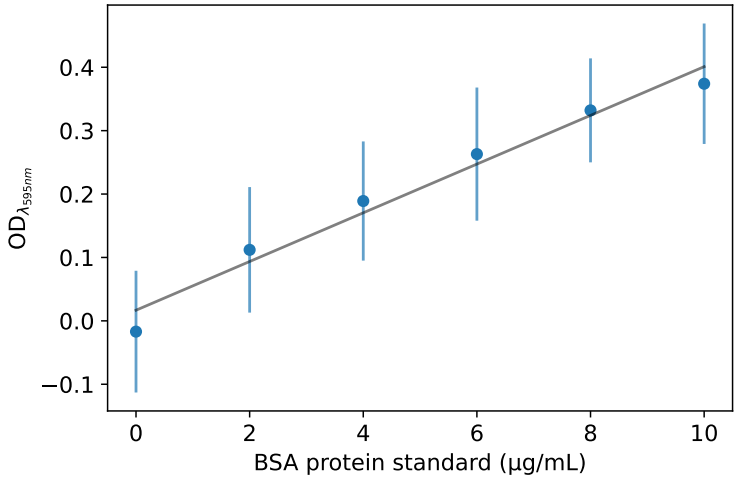


FIGURE 1. The standard curve for analysis of protein content, made with samples of BSA protein standard, to be used for protein content analysis in experiments of *B. subtilis* cultivated in minimal media containing different carbon sources.

The equation for the regression line was found with Excel, and was $y = 0,0384x + 0,017$, with an R^2 of 0.97. In order to find the amount of proteins, this was solved to find the x for each OD measured in the protein samples.

After making the regression curve, the cells from the fermentations were analyzed. 10 mg of cells were weighed out, and MQ-water (30 μL) was added to re-hydrate the cells. To lyse the cells 1 mL of Bacterial Protein Extraction Reagent (B-PER) was mixed with lysozyme (2µL, 50µg/mL). Because B-PER is best suited for gram-negative bacteria, the lysozyme was added for increased yield. The samples were incubated at room temperature in a thermocycler for some slow stirring (500 RPM) for 3 hours. After incubation the samples were centrifuged at 7200 RCF for 10 minutes. 150 µL of the the supernatant was extracted and diluted with

MQ-water to a total volume of 10mL. The samples were further diluted 1:4000, and 2.4 mL of this dilution was mixed with 0.6mL Bio-Rad protein assay. The samples were mixed with by vortex, before a 10 minute incubation. Then the samples were analyzed by OD-measurements at 595nm. The resulting amount of proteins found is illustrated in Figure 2.

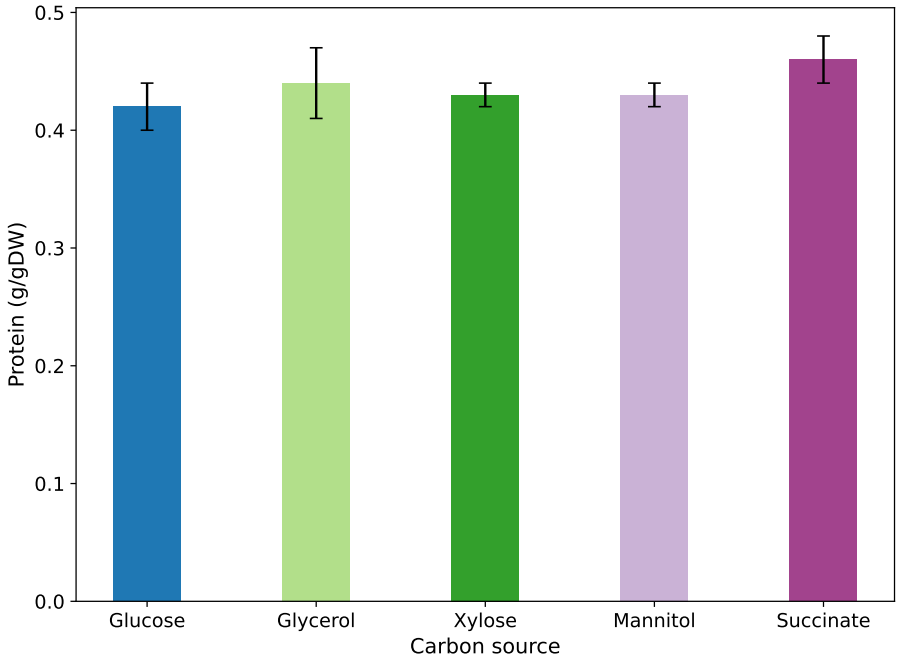


FIGURE 2. The resulting amount of proteins (g/gDW) found by bradford protein assay analysis performed on cells of *B. subtilis* cultivated in minimal media containing different carbon source

The analysis was performed with triplicates from each carbon source. The calculations and raw data can be viewed in the "Proteins - Bradford" sheet in the Biomass analyses excel document.

Appendix B - Full growth curve for *B. subtilis* in minimal media

This full growth curve with *B. subtilis* illustrates how the growth shifts around OD 3 to decreasing. This plot was used to find an appropriate OD for harvesting the cells in the later batch-fermentations, to ensure that the cells used in the biomass analyses were in exponential growth at the time of harvest.

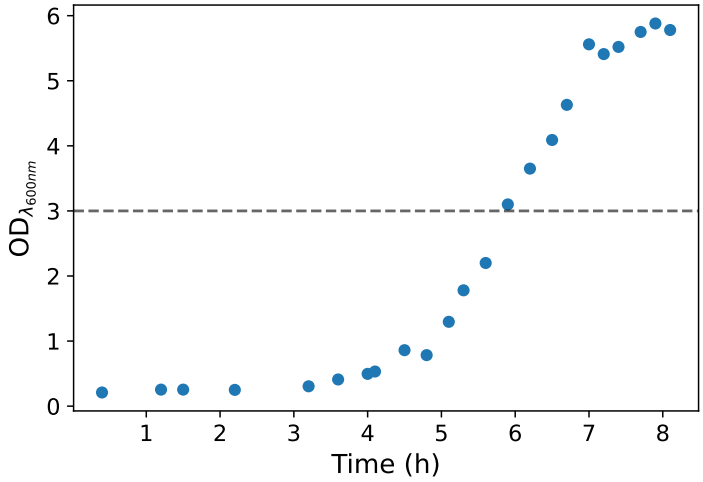


FIGURE 3. A full growth curve of *B. subtilis* cultivarted by batch-fermentation in minimal media containing glucose to explore the duration of the exponential phase. This was used to determine an approximate OD appropriate for harvesting the cells in later batch-fermentations for biomass analyses.

Appendix C - Experimental data

In case there are any issues with the hyperlinks in this thesis, the full URL to the supplementary data in GitHub is here: https://github.com/Sofieta95/Supplementary-data---Master-s-project.git

Experimental growth rate values

This section contains the growth rates with corresponding standard deviations and R^2, listed in Table 1. The standard deviations were found using RStudio, with the Linear regression Growth rates-script. Between Excel and RStudio, the growth rate predictions were the same, but the R^2-values varied between the two programs, and so the values from both are included.

TABLE 1. Experimentally determined growth rates for *B. subtilis* in minimal media containing different carbon sources. The growth rates were found by linear regression in Excel of log-transformed dry weight measurements against time. The slope of the equation for the regression line corresponded to the growth rate. The standard deviation (std. dev) was found using RStudio regression analysis.

	Growth rate	Std. Dev	R^2 excel	R^2 Rstudio
Glucose	0.357	0.027	0.97	0.91
Glycerol	0.303	0.036	0.83	0.79
Xylose	0.091	0.012	0.86	0.76
Mannitol	0.411	0.032	0.96	0.93
Succinate	0.213	0.025	0.95	0.79

The script for the ANOVA and post hoc analyses performed is available in the Statistical analyses Python script.

Experimentally determined DNA content

The experimentally measured values are listen in Table 2 with corresponding standard deviations.

The ANOVA analysis was performed using Python. The script is in the "Statistical analyses" file in the GitHub repository. The results from the one-way ANOVA analysis performed with the measured DNA content between the carbon sources was 2.82×10^{-38}. The results from the post hoc- and the Pearson correlation analysis between DNA content and growth rate can be viewed in the script Statistical analyses.

TABLE 2. The experimentally measured values for DNA content in *B. subtilis* cultivated by batch-fermentation, and harvested in exponential growth phase for biomass composition analysis.

	DNA (g/gDW)	Std. Dev
Glucose	0.0231	0.0016
Glycerol	0.0238	0.0014
Xylose	0.0141	0.0003
Mannitol	0.0468	0.0020
Succinate	0.0274	0.0009

Experimentally determined RNA content

The experimentally measured values for RNA are listen in Table 3 with corresponding standard deviations.

TABLE 3. The experimentally measured values for RNA content in *B. subtilis* cultivated by batch-fermentation, and harvested in exponential growth phase for biomass composition analysis.

	RNA (g/gDW)	Std. Dev
Glucose	0.0688	0.0022
Glycerol	0.0762	0.0022
Xylose	0.0623	0.0023
Mannitol	0.0870	0.0016
Succinate	0.0476	0.0034

The ANOVA analysis was performed using Python. The script is in the "Statistical analyses" file in the GitHub repository. The results from the one-way ANOVA analysis performed with the measured RNA content between the carbon sources was $1.9\times\ 10^{-46}$. The results from the post hoc- and the Pearson correlation analysis between RNA content and growth rate can be viewed in the script Statistical analyses.

Experimentally determined protein content

The experimentally measured protein content (g/gDW) is listed in Table 4.

TABLE 4. The experimentally measured values for protein content in *B. subtilis* cultivated by batch-fermentation, and harvested in exponential growth phase for biomass composition analysis.

	Protein (g/gDW)	Std. Dev
Glucose	0.564	0.027
Glycerol	0.544	0.003
Xylose	0.559	0.003
Mannitol	0.502	0.021
Succinate	0.562	0.013

The ANOVA analysis was performed using Python. The script is in the "Statistical analyses" file in the GitHub repository. The results from the one-way ANOVA analysis performed with the measured RNA content between the carbon sources was 0.004. The results from the post hoc- and the Pearson correlation analysis between protein content and growth rate can be viewed in the script Statistical analyses.

Experimentally determined lipid content

The experimentally measured lipid content (g/gDW) is listed in Table 5.

TABLE 5. The experimentally measured values for lipid content in *B. subtilis* cultivated by batch-fermentation, and harvested in exponential growth phase for biomass composition analysis.

	Lipid (g/gDW)	Std. Dev
Glucose	0.0307	0.0017
Glycerol	0.0292	0.0021
Xylose	0.0328	0.0020
Mannitol	0.0326	0.0032
Succinate	0.0358	0.0007

The ANOVA analysis was performed using Python. The script is in the "Statistical analyses" file in the GitHub repository. The results from the one-way ANOVA analysis performed with the measured RNA content between the carbon sources was 0.028. The results from the post hoc- can be viewed in the script Statistical analyses.

Appendix D - One sample T-test comparing nBOF coefficients to oBOF

A one-sample T-test was used to compare the resulting coefficients in the nBOFs to the coefficients in the original BOF in *i*Bsu1147. The script used to perform the one sample T-test can be found in the Statistical analyses script. The P-values from the analysis are listed below. The raw data values are listed in the script for all of the fermentation analyses. The expected value in the test was the coefficient from the model.

Glucose

Coefficient	Expected value	New coefficient value	T-statistic:	P-value
DNA	0.0260	0.0234	-5.176	0.00058
RNA	0.0655	0.0697	6.707	$3.34 \cdot 10^{-5}$
Lipid	0.076	0.0311	-44.82	0.00049
protein	0.528	0.572	2.721	0.11

Glycerol

Coefficient	Expected value	New coefficient value	T-statistic:	P-value
DNA	0.0260	0.0245	-3.213	0.012
RNA	0.0655	0.0784	24.19	$1.31 \cdot 10^{-14}$
Lipid	0.076	0.0301	-36.11	0.00076
Protein	0.528	0.559	14.31	0.0048

Xylose

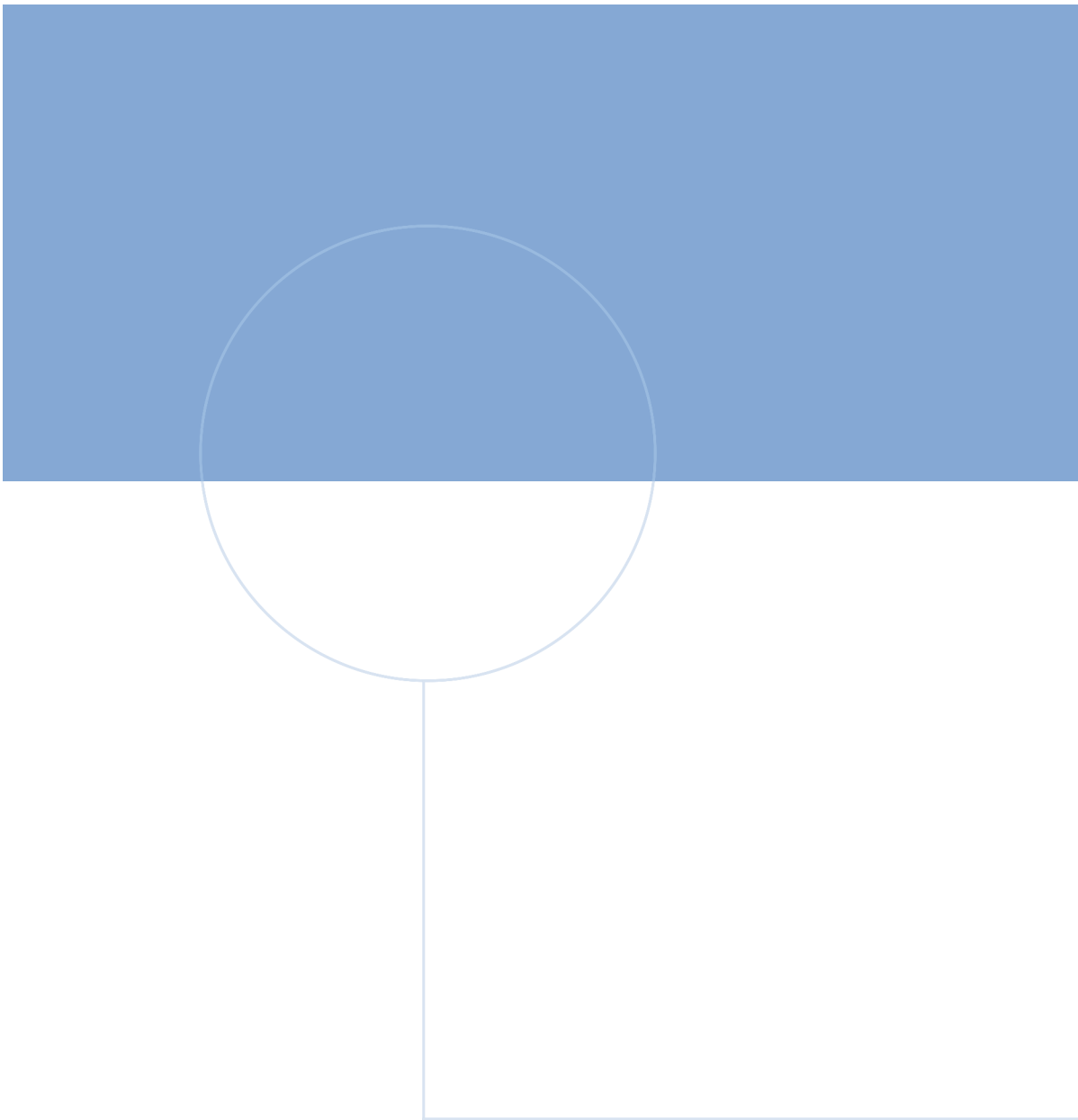
Coefficient	Expected value	New coefficient value	T-statistic:	P-value
DNA	0.0260	0.0145	-108.3	$5.92 \cdot 10^{-14}$
RNA	0.0655	0.0642	-1.81	0.097
Lipid	0.076	0.0369	-35.85	0.00077
Protein	0.528	0.576	26.017	0.0015

Mannitol

Coefficient	Expected value	New coefficient value	T-statistic:	P-value
DNA	0.0260	0.0484	32.82	$8.09 \cdot 10^{-10}$
RNA	0.0655	0.0899	49.25	$2.87 \cdot 10^{-13}$
Lipid	0.076	0.0336	-22.01	0.002
Protein	0.528	0.519	-0.7992	0.508

Succinate

Coefficient	Expected value	New coefficient value	T-statistic:	P-value
DNA	0.0260	0.0282	7.94	$2.35 \cdot 10^{-5}$
RNA	0.0655	0.0490	-17.9	$4.71 \cdot 10^{-11}$
Lipid	0.076	0.0369	-96.7	0.00010
Protein	0.528	0.578	6.52	0.023



NTNU

Norwegian University of
Science and Technology