Master's thesis

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

Linn Sandvik

Exploring the *in vivo* and *in silico* Behaviour of the Synthetic Methylotroph *Bacillus subtilis* 168 pBV2mp_mdhBm

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

June 2023

NTNU
Norwegian University of
Science and Technology

Linn Sandvik

Exploring the *in vivo* and *in silico* Behaviour of the Synthetic Methylotroph *Bacillus subtilis* 168 pBV2mp_mdhBm

Master’s thesis in MBIOT5
Supervisor: Eivind Almaas
Co-supervisor: Vetle Simensen, 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

Acknowledgements

I would like to thank my supervisor Eivind Almaas, my co-supervisors Vetle Simensen and Christian Schulz, and the rest of the members of Almaas Lab, for help, support and encouragement. Further, I would like to thank Vivien Jessica Klein for providing glycerol stock of the *Bacillus subtilis* 168 pBV2mp_mdhBm strain. I would also like to thank Laura García Calvo, Divyata Vilas Rane and Siri Stavrum for assisting me in the lab. Last but not least, a special thanks goes to my fellow master student Sofie Tande-Petersen, who has been my labpartner (in crime) throughout this whole year.

Abstract

Bacillus subtilis is a bacterial strain that is much used as a cell factory in biotechnological industry [1]. It is typically cultivated with glucose as the carbon source [2], however it has become of interest to find alternative carbon sources that do not compete with food production. A promising alternative is methanol, which can be used as a carbon source by so called methylotrophic microorganisms [3]. *B. subtilis* is not a natural methylotroph, but it has been demonstrated that synthetic methylotrohpy can be induced in *B. subtilis* by making it heterologously express a methanol dehydrogenase (Mdh) gene. However, the synthetic methylotrophic *B. subtilis* strains that have been constructed so far are only able to use methanol as a co-carbon source in addition to e.g. glucose [4]. In our work we have conducted experimental and in silico analyses to confirm that *B. subtilis* is indeed able to use methanol as a carbon source when Mdh is heterologously expressed. This was done by conducting batch fermentations where an Mdh-expressing mutant strain, *B. subtilis* 168 pBV2mp_mdhBm, was cultivated in minimal medium containing glucose and methanol. The phenotype of the mutant was analyzed experimentally by measuring the biomass composition, amino acid distribution, growth rate, and uptake- and secretion rates of nutrients and gas. The experimental data was also used to update the newest available genome-scale metabolic model (GEM) of *B. subtilis*, which allowed us to study the phenotype of the mutant in silico. The in silico analyses that were conducted were flux balance analysis (FBA) and flux variability analysis (FVA), which were used to study the flux distribution through the metabolic network of the mutant strain. The results from the experimental and in silico analyses are in accordance with the findings of Gao et al. [4], and show that the mutant strain can use methanol as a co-carbon source. Our results have also been used to propose further analyses that can be conducted to bring us closer to a mutant *B. subtilis* strain that can use methanol as the only carbon source.

Sammendrag

Bacillus subtilis er en type bakterie som er mye brukt som cellefabrikk i bioteknologisk industri [1]. Bakterien kultiveres vanligvis med glukose som karbonkilde [2], men det har oppstått en interesse for å finne alternative karbonkilder som ikke konkurrerer med matproduksjon. Et lovende alternativ er metanol som kan brukes som karbonkilde av såkalte metylotrofe mikroorganismer [2]. *B. subtilis* er ikke en naturlig metylotrof, men det har blitt vist at syntetisk metylotrofi kan induseres i *B. subtilis* ved å få den til å uttrykke genet methanol dehydrogenase (Mdh). De syntetisk metylotrofe variantene av *B. subtilis* som har blitt fremstilt så langt, er imidlertid bare i stand til å bruke methanol som co-karbonkilde i tillegg til for eksempel glukose [4]. I vårt arbeid har vi utført eksperimentelle og in silico analyser for å bekrefte at *B. subtilis* faktisk er i stand til å bruke metanol som karbon kilde når Mdh er uttrykt. Dette ble gjort ved å utføre såkalte batch fermenteringer, hvor den muterte varianten *B. subtilis* 168 pBV2mp_mdhBm ble kultivert i minimalt medium som inneholdt glukose og metanol. Fenotypen til mutanten ble analysert eksperimentelt ved å måle biomassekomposisjon, aminosyrefordeling, vekstrate og opptaks- og sekresjonsrater av næringsstoffer og gasser. De eksperimentelle dataene ble også brukt til å oppdatere den nyeste tilgjengelige genomskala modellen (GEM) av *B. subtilis*, og den oppdaterte modellen ble brukt til å studere fenotypen til mutanten in silico. In silico analyser som ble utført er analyse av fluksbalanse (FBA) og fluksvariabilitetsanalyse (FVA). Disse analysene ble brukt for å undersøke distribusjonen av fluks gjennom det metabolske nettverket til mutanten. Resultatene fra eksperimentelle og in silico analyser stemmer overens med funnene som ble gjort av Gao et al. [4], og viser at den muterte varianten er i stand til å bruke metanol som co-karbonkilde. Vi har også brukt resultatene våre for å foreslå videre analyser som kan utføres, for å føre oss nærmere en mutert variant av *B. subtilis* som kan bruke metanol som eneste karbonkilde.

Table of Contents

Acknowledgements **i**

Abstract **ii**

Sammendrag **iii**

Table of Contents **vii**

List of Tables **xi**

List of Figures **xv**

Abbreviations **xvi**

1 Introduction **1**

2 Background **3**
- 2.1 Native methanol metabolism pathways 3
- 2.2 Cryptic genes 4
- 2.3 Bioreactor cultivation 5
  - 2.3.1 Batch operation mode 6
  - 2.3.2 Chemostat operation mode 6
- 2.4 Measuring the biomass composition 6
  - 2.4.1 Protein extraction 7
  - 2.4.2 Lipid extraction 7
  - 2.4.3 DNA extraction 7
  - 2.4.4 RNA extraction 7
- 2.5 Chromatography 8
- 2.6 Mass spectrometry 8
- 2.7 Nuclear magnetic resonance (NMR) 8
- 2.8 Uptake- and secretion rates 9
  - 2.8.1 Gas uptake- and secretion rates 9

2.8.2 Uptake- and secretion rates of medium compounds 9
2.9 Genome-scale metabolic modeling 10
2.10 The biomass objective function 13
2.11 Growth associated and non-growth associated maintenance 14
2.12 Genome-scale metabolic models of *Bacillus subtilis* 14

3 Software and methods 15
3.1 The cryptFind algorithm 15
3.1.1 Finding carbon sources that *B. subtilis* can use for growth in vivo and in silico 15
3.2 *Bacillus subtilis* 168 and *B. subtilis* 168 pBV2mp_mdhBm 16
3.3 LB medium 16
3.4 Minimal medium 16
3.4.1 Stock solutions 17
3.4.2 Trace mineral solution (TMS) 17
3.4.3 Phosphate buffer 18
3.4.4 Preparing minimal medium 18
3.5 Glycerol stocks 19
3.6 Precultures 19
3.6.1 Wild type in glucose minimal medium 19
3.6.2 Wild type in minimal medium with glucose and methanol 19
3.6.3 Mutant in glucose minimal medium 20
3.6.4 Mutant in minimal medium with glucose and methanol 20
3.7 Batch culturing in bioreactors 20
3.7.1 Cultivation 20
3.7.2 Sampling 21
3.7.3 Harvesting and freeze-drying 21
3.7.4 Preparation of NMR samples 22
3.8 Biomass composition measurements 22
3.8.1 RNA quantification 22
3.8.2 DNA quantification 23
3.8.3 Lipid quantification 23
3.8.4 Protein hydrolysis and HPLC 24
3.8.5 Calculating the amino acid distribution and total protein content . 24
3.9 Growth rates 25
3.10 Finding the specific O_2 uptake- and CO_2 secretion rates 25
3.11 Finding the uptake- and secretion rates of medium components 26
3.12 Conducting flux balance analysis (FBA) and flux variability analysis (FVA) 26
3.12.1 Creating new synthesis reactions based on experimental data . . . 26
3.12.2 Updating flux bounds based on experimentally measured uptake- and secretion rates 27
3.12.3 Conducting flux variability analysis (FVA) 27
3.13 Adding bacillithiol dependent oxidation of formaldehyde to the model . . 28

4 Results and discussion **29**
4.1 Unraveling contradictions between the in silico and in vivo behaviour of *B. subtilis* . . . 29
4.2 An overview of the batch fermentations that were conducted . . . 30
4.3 The mutant strain grows significantly faster with methanol as co-carbon source . . . 31
4.4 The measured biomass composition varies according to the strain and nutrient environment . . . 33
4.5 The relative amino acid distribution varies according to strain and nutrient environment . . . 36
4.6 Both the wild type and the mutant strain seem to consume methanol . . . 39
4.7 The predicted growth rate is robust to changes in the biomass- and protein synthesis reaction . . . 42
4.8 FVA results suggest that formaldehyde is assimilated into the RuMP pathway when methanol is available . . . 44

5 Conclusion and Outlook **51**

Bibliography **53**
5.1 Growth rate measurements . . . 59
5.2 Biomass composition . . . 59
5.3 Specific uptake- and secretion rates . . . 63
5.4 Using the *B. subtilis* GEM for analysing the phenotype of the organism . 68

List of Tables

3.1 Components used to make LB medium. The components were dissolved in Milli-Q (MQ-H_2O) to obtain the listed concentrations. Supplier and CAS number of each component are provided. 16

3.2 Components used to make stock solutions, and their final concentrations. The components were dissolved in Milli-Q (MQ-H_2O) to obtain the listed concentrations. Supplier and CAS number of each component are provided. 17

3.3 Components used to make trace mineral solution (TMS). The components were dissolved in 5 M HCl to obtain the listed concentrations. Supplier and CAS number of each component are provided. 17

3.4 Stock solutions and components used to make minimal medium. The concentration of the components in stock solution and in minimal medium is given. Milli-Q (MQ-H_2O) was used to obtain the desired concentration of each component. 18

3.5 Components of the trace mineral solution (TMS) and their concentrations (g L^{-1}) in TMS and minimal medium. 19

3.6 OD600 of each bioreactor cultivation when starting the harvesting. 22

4.1 Abbreviations that are assigned to each fermentation based on the *Bacillus subtilis* strain and carbon source(s) that were used. 30

4.2 Average amino acid content (g/gDW) for each fermentation. The total protein content (g/gDW) is calculated as the sum of the content of all 20 amino acids. One-way ANOVA was conducted to compare the content of each amino acid between fermentations, and the p-values are presented. The average content of cys, pro, trp, and met was predicted based on the amino acid distribution derived from the protein coding genes of *B. subtilis*. Only one number was predicted per fermentation, and one-way ANOVA was therefore not conducted for these four amino acids. 37

4.3 Average amino acid content (g/gProtein) for each fermentation. The total protein content (g/gProtein) is calculated as the sum of the content of all 20 amino acids. One-way ANOVA was conducted to compare the content of each amino acid between fermentations, and the p-values are presented. The average content of cys, pro, trp, and met was predicted based on the amino acid distribution derived from the protein coding genes of *B. subtilis*. Only one number was predicted per fermentation, and one-way ANOVA was therefore not conducted for these four amino acids. 38

5.1 Growth rate (h^{-1}) for each biomass sample. The growth rate corresponds to the slope of the regression line found by conducting linear regression of cell dry weight (CDW) on log scale plotted against time (h). Standard deviation (SD), as well as the R^2 value of each regression line are provided. 59

5.2 ANOVA results from a pairwise comparison of growth rate between the four fermentations. The critical value is calculated based on a significance level of 0.95, numerator degrees of freedom (df) = 1 and denominator df = 4. When the F-value is greater than the critical value, there is a significant difference in growth rate between the two fermentations. 59

5.3 The biomass composition for each biomass sample shown as mass fractions (g/gDW) of each biomass component. The amount of protein, RNA, DNA and lipid was measured experimentally, and is shown with standard deviation. The amount of lipoteichoic acid, cell wall, and cofactors & ions are assumed to be the same as for iBsu1147. Furthermore, the sum of the mass fractions is given at the bottom of the table. 60

5.4 p-values from one-way ANOVA analysis comparing the protein, RNA, DNA an lipid content between fermentations. 60

5.5 p-values from one-way ANOVA analysis comparing the scaled values of protein, RNA, DNA, and lipid content between fermentations. The values have been scaled in order to get a sum of one. 61

5.6 The biomass composition for each fermentation after scaling the data to get a sum of one. The content of protein, RNA, DNA, and lipid were measured experimentally and are shown with SD. The contents are presented in g/gDW. 62

5.7 p-values from pairwise comparison of the coefficients of the new BOFs (WTG, WTM, MDHG and MDHM) to those of the original BOF of iBsu1147. The p-values were found by using one-sample t-test. 62

5.8 Density, molar mass, and concentration (N) of O_2 and CO_2. 63

5.9 Average CDW, $mg\ mL^{-1}$) and the common logarithm (log_{10}) of the average CDW over time (h) for *B. subtilis* 168 wild type cultivated in glucose minimal medium. For each time point, triplicate measurements of the CDW were conducted and averaged. Standard deviation (SD) of each triplicate is found by using the log_{10} value of each replicate. 63

5.10 Specific O_2 uptake rate (avg. q_{O2}) based on 10 output gas measurements conducted during the cultivation of *B. subtilis* 168 WT on glucose minimal medium. The 10 measurements are conducted towards the end of the cultivation, before the filter was clogged. Each q_{O_2} has been calculated by using Equation 2.1 and 2.2. The value of $O2_{air}$ needed for Equation 2.1 is based on 10 output gas measurements conducted before inoculation of the reactor, and was found to be 20.895. 64
5.11 Uptake- and secretion rates (mmol/gDW h) of glucose, methanol, L-glu, L-trp, and acetate for each biomass sample. The values presented are the average rates +- SD. 64
5.12 Specific uptake- and secretion rates (mmol/gDW h) of O_2 and CO_2, and respiratory quotient (RQ) for each biomass sample. Negative and positive rates represent uptake and secretion, respectively. 65
5.13 Nutrient environment used to compare the predicted growth rates of iBsu1147, WTG, WTM, MDHG and MDHM. The components that are listed represent the components that were available during experimental cultivations. The corresponding exchange reaction for each component, and the chosen flux bounds are given. A flux bound of (-1000.0, 1000.0) represents unrestricted flux. For D-glucose, l-tryptophan and l-glutamic acid, the uptake rate has been restricted to -1.8, -0.5 and -1.0 respectively. For exchange reactions that are not visualized in the table, the flux bounds were set to (0, 1000.0). 68
5.14 Measured and predicted growth rate (h^{-1}) for each fermentation. Here, ”BOF” means that the BOF and protein synthesis reaction of iBsu1147 were used, while ”mBOF” means that the new BOF and protein synthesis reaction specific for the fermentation was used. 69

List of Figures

2.1 A simplified illustration of the assimilatory RumP pathway gathered and modified from He et al. [5]. The figure shows how methanol is oxidized to formaldehyde before it is incorporated into the RuMP cycle by converting it to Hexulose-6-Phosphate (H-6-P) and further to Fructose-6-Phosphate (F-6-P), which is used to produce biomass and to regenerate Ribulose-5-phosphate (Ru-5-P). Alternatively formaldehyde can be oxidized to CO_2. 4
2.2 A picture of the type of bioreactor that was used in this work. 5

3.1 A flow chart representing the steps conducted before conduction FVA for each fermentation. 27
3.2 Figure: Illustration of the bacillithiol (BSH) dependent formaldehyde oxidation pathway, which is present in *B. subtilis* [6]. The pathway consists of four reactions, and the reaction IDs (ModelSeed or KEGG) are presented above each arrow. Metabolites and reactions that were not already present in iBsu1147 are presented in orange. 28

4.1 Optical density at 600 nm (OD600) over time (h) for each fermentation. The bacteria were harvested at an OD600 of approximately 3. (”Biomass sample” is used the same way as ”fermentation” is used in the rest of the text.) 31
4.2 (A) Cell dry weight (CDW) on log scale plotted against time (h) for each fermentation. Linear regression was used to decide the growth rate (h^{-1}) of each fermentation, which are presented in part B. (B) Estimated growth rate (h^{-1}) for each Fermentation. (”Biomass sample” is used the same way as ”fermentation” is used in the rest of the text.) 32
4.3 Biomass composition for each fermentation, shown as mass fraction (g/gDW) of each component. Only the fractions of the biomass components that were measured are visualized. (”Biomass sample” is used the same way as ”fermentation” is used in the rest of the text.) 34

4.4 Scaled values of the protein, RNA, DNA, and lipid content (g/gCDW) that were used to create a new BOF representing each fermentation. The corresponding coefficients that is used for iBsu1147 are also presented. ("Biomass sample" is used the same way as "fermentation" is used in the rest of the text.) 35

4.5 Specific substrate uptake- and secretion rates (mmol/gDW h) for each fermentation. 39

4.6 Specific uptake- and secretion rates (mmol/gDW h) of O_2 and CO_2 for each fermentation. Negative rates represent uptake rates, while positive rates represent secretion rates. 40

4.7 Predicted growth rate of each model: iBsu1147, WTG, WTM, MDHG and MDHM, when nutrient environment is kept constant. 42

4.8 Experimentally measured and predicted growth rate (h^{-1}) for each fermentation. For the predicted growth rates, "BOF" means that the BOF and protein synthesis reaction of iBsu1147 were used, while "mBOF" means that the new BOF and protein synthesis reaction specific for the fermentation were used. 43

4.9 FVA results for WTG. The figure shows a graphical representation of the assimilatory- and dissimilatory RuMP pathway present in the GEM of *B. subtilis*. The nodes represent metabolites, and the edges represent the interactions between the metabolites. The name of each metabolite and the reaction ID of each reaction is presented. The reactions that have a flux range that does not span 0, are highlighted in pink. The reaction ID and flux range of these reactions are also highlighted. The remaining reactions have a flux range that spans 0. All the reactions that are presented are defined as reversible in the GEM, which is why the reaction arrows point both ways. The direction of the flux follows the direction of the filled arrows. 46

4.10 FVA results for WTM. The figure shows a graphical representation of the assimilatory- and dissimilatory RuMP pathway present in the GEM of *B. subtilis*. The nodes represent metabolites, and the edges represent the interactions between the metabolites. The name of each metabolite and the reaction ID of each reaction is presented. The reactions that have a flux range that does not span 0, are highlighted in pink. The reaction ID and flux range of these reactions are also highlighted. The remaining reactions have a flux range that spans 0. All the reactions that are presented are defined as reversible in the GEM, which is why the reaction arrows point both ways. The direction of the flux follows the direction of the filled arrows. 47

4.11 FVA results for MDHG. The figure shows a graphical representation of the assimilatory- and dissimilatory RuMP pathway present in the GEM of *B. subtilis*. The nodes represent metabolites, and the edges represent the interactions between the metabolites. The name of each metabolite and the reaction ID of each reaction is presented. The reactions that have a flux range that does not span 0, are highlighted in pink. The reaction ID and flux range of these reactions are also highlighted. The remaining reactions have a flux range that spans 0. All the reactions that are presented are defined as reversible in the GEM, which is why the reaction arrows point both ways. The direction of the flux follows the direction of the filled arrows. 48
4.12 FVA results for MDHM. The figure shows a graphical representation of the assimilatory- and dissimilatory RuMP pathway present in the GEM of *B. subtilis*. The nodes represent metabolites, and the edges represent the interactions between the metabolites. The name of each metabolite and the reaction ID of each reaction is presented. The reactions that have a flux range that does not span 0, are highlighted in pink. The reaction ID and flux range of these reactions are also highlighted. The remaining reactions have a flux range that spans 0. All the reactions that are presented are defined as reversible in the GEM, which is why the reaction arrows point both ways. The direction of the flux follows the direction of the filled arrows. 49

5.1 Tukey's HSD test for multiple comparisons was conducted to decide if there are significant differences in protein, RNA, DNA and lipid content between biomass samples. The figure shows the Tukey's HSD test results for A) protein content, B) RNA content, C) DNA content and D) lipid content. The results are presented as 95 % family-wise confidence intervals. Confidence intervals that do not span 0.0 represent significant difference. For instance there is a significant difference in protein content between WTG and MDHG, as the confidence interval does not span 0.0. 61
5.2 Common logarithm (log_{10}) of the average CDWs over time (h) for *B. subtilis* 168 cultivated in glucose minimal medium. A linear regression line has been fitted to the datapoints, and the equation and R^2 are displayed on chart. 64
5.3 Tukey's HSD test for multiple comparisons was conducted to decide if there are significant differences in specific uptake- and secretion rate between fermentations. The figure shows the Tukey's HSD test results for A) specific glucose uptake rate, B) specific acetate secretion rate, C) specific L-glu uptake rate, D) specific L-trp uptake rate, and E) specific methanol uptake rate. The results are presented as 95 % family-wise confidence intervals. Confidence intervals that do not span 0.0 represent a significant difference. 66
5.4 The figure shows the Tukey's HSD test results for A) specific O_2 uptake rate, and B) specific CO_2 secretion rate. The results are presented as 95 % family-wise confidence intervals. Confidence intervals that do not span 0.0 represent a significant difference. 67

Abbreviations

MQ-H_2O = Milli-Q water
TMS = Trace Mineral Solution
RPM = Revolutions Per Minute
dH_2O = Distilled water
OD600 = Optical Density at 600 nm
CDW = Cell Dry Weight
GEM = Genome-Scale Metabolic Model
BOF = Biomass Objective Function
GAM = Growth Associated Maintenance
NGAM = Non-Growth Associated Maintenance
WTG = Wild Type in Glucose medium
WTM = Wild Type in Methanol medium
MDHG = Mutant in Glucose medium
MDHM = Mutant in Methanol medium
SD = Standard Deviation
BSH = Bacillithiol
H-6-P = Hexulose-6-Phosphate
F-6-P = Fructose-6-Phosphate
Ru-5-P = Ribulose-5-phosphate

Chapter 1

Introduction

Bacillus subtilis is a strain of bacteria that is much used as a cell factory in biotechnological industry. This is because *B. subtilis* is able to efficiently produce and secrete heterologous proteins that are necessary for medicine, agriculture, and industrial processes [1]. *B. subtilis* is naturally a chemoheterotroph which uses glucose and malate as its preferred carbon sources [2]. However, with a growing human population it has become of interest to find alternative carbon sources that do not compete with food production. An example of such a compound is the alcohol methanol, which is toxic to humans but can be used as a carbon source by some microorganisms. Furthermore, the production of methanol is becoming more efficient, which makes methanol an affordable carbon source to use for biotechnological industry [3]. *B. subtilis* is not able to use methanol as a carbon- and energy source naturally, and it is therefore of interest to introduce synthetic methylotrophy to the strain. The definition of synthetic methylotrophy is introducing the ability to use C1-compounds (organic compounds with one carbon atom) such as methane or methanol as carbon- and energy sources [3]. Introducing synthetic methylotrophy to *B. subtilis* would make it possible to produce heterologous proteins necessary for medicine, agriculture and industry by using a carbon source that is otherwise considered a waste product.

The alternative to synthetic methylotrophy is to use native methylotrophs as cell factories. Native methylotrophs grow quickly [7], however they typically have low production rate of products of interest, and there are few available tools for genetic modification of these strains [8]. *B. subtilis*, on the other hand, has high production rate, and there are already many tools available for genetic modification of this strain. Furthermore, in theory *B. subtilis* only lacks one gene to be able to use methanol as a carbon source [4]. The missing gene is the methanol dehydrogenase (Mdh) gene. Mdh allows the organism to oxidize methanol to formaldehyde, which is the first step of methanol metabolism. Although *B. subtilis* is not able to convert methanol to formaldehyde, it has the metabolic pathway that is necessary to turn formaldehyde into biomass. One can therefore argue that it is easier to induce synthetic methylotrophy in *B. subtilis* by heterologous expression of Mdh, than to turn a native methylotroph into a cell factory as efficient as *B. subtilis*.

In theory, the Mdh gene is the only missing link between *B. subtilis* and synthetic methylotrophy. In practice, however, it turns out not to be that simple. It was namely demonstrated by Gao et al. that *B. subtilis* is not able to use methanol as a sole carbon-source, even when Mdh is heterologously expressed. The recombinant strain is however able to use methanol when a co-substrate such as glucose is provided [4]. This gives rise to the question of why the recombinant *B. subtilis* strain is still unable to use methanol as the only carbon source? To look into this question we need to know what happens to the methanol after it has been taken up by *B. subtilis*. Is it actually used for biomass production? We know that the methanol is first converted to formaldehyde, as this has been established by Gao et al. [4]. Furthermore, we know that the bacteria need to somehow detoxify the formaldehyde, as formaldehyde in high concentrations is toxic to most cells [9]. However, in *B. subtilis* the formaldehyde can be detoxified in two ways: either it is assimilated, which means that it is incorporated in central metabolism and used for biomass production, or it is dissimilated, which means that it is oxidized to CO_2 and secreted by the cell [10, 11, 6].

To investigate the question of how the formaldehyde is detoxified, we have used a genome-scale metabolic model of *B. subtilis*. A genome-scale metabolic model (GEM) is a mathematical representation of all the chemical reactions that are known to take place in an organism [12]. GEMs serve as useful resources to study the behavior, or phenotype, of an organism under different environmental conditions. The GEM of *B. subtilis* should therefore be able to give us insight to how the recombinant *B. subtilis* strain detoxifies formaldehyde after it has been produced from methanol. However, there is a catch to this plan. The GEM of *B. subtilis* has been built based on data from the wild type strain cultivated with glucose as the carbon source [13]. This model might therefore not accurately reflect the behaviour of a recombinant *B. subtilis* strain with access to methanol as a carbon source. Therefore, we have updated the current GEM of *B. subtilis* so that it better can be used to study the phenotype of the recombinant strain.

To update the existing GEM of *B. subtilis*, we need to know a few things about the recombinant strain, such as the growth rate, biomass composition, and uptake- and secretion rates of nutrients and other metabolites. These data have been gathered by cultivating the recombinant *B. subtilis* strain in medium that contains methanol, and glucose as co-substrate. The bacteria were cultivated in bioreactors, and growth rate, and uptake- and secretion rates were measured. Furthermore, the bacteria were harvested in order to measure the biomass composition. All this data has been used to update the existing GEM of *B. subtilis*, so that it more accurately represents the recombinant strain. Then the updated GEM has been used to investigate how the formaldehyde is detoxified, which hopefully will bring us closer to an answer to how recombinant *B. subtilis* could use methanol as a sole carbon source.

Chapter 2

Background

2.1 Native methanol metabolism pathways

Wild type *B. subtilis* uses glucose as one of its preferred carbon sources [2]. Glucose is used in glycolysis and the tricaboxylic acid (TCA) cycle to create reducing agents for energy production. Furthermore, glucose is used in the pentose phosphate pathway (PPP) to create reducing agents necessary in several anabolic processes, and to form ribose-5-phosphate, a key compound in nucleic acid synthesis. In this way, glucose is used by *B. subtilis* as both a carbon and energy source [14].

Pathways of native methylotrophs are used as inspiration when creating synthetic methylotrophic strains. One of the native pathways of methanol assimilation is the ribulose monophosphate pathway (RuMP) [15]. To use this pathway, the methylotroph first needs to oxidize methanol to formaldehyde, a process catalyzed by the methanol dehydrogenase enzyme (Mdh). Then, the RuMP cycle is used to convert formaldehyde to fructose 6-phosphate (F-6-P), a process catalyzed by two enzymes Hps and Phi. F-6-P is then incorporated in central metabolism to produce energy and biomass. Furthermore, some of the F-6-P is used to regenerate a five carbon sugar needed too keep running the RuMP pathway [16, 17]. The RuMP pathway shares several enzymes with the PPP [18], which is one of the pathways used by *B. subtilis* for sugar metabolism. The enzymes that are not shared are Mdh, Hps and Phi, which are key enzymes of the RuMP pathway [18]. *B. subtilis* expresses the enzymes of the PPP, and it has been found that it also has enzymes with similar enzymatic activity as Hps and Phi. This enables *B. subtilis* to run the RuMP cycle when formaldehyde is present [10]. However, *B. subtilis* is unable to oxidize methanol to formaldehyde as it does not have the Mdh gene. In theory, a mutant *B. subtilis* strain that heterologously expresses the Mdh gene should be able to use methanol as carbon- and energy source [4].

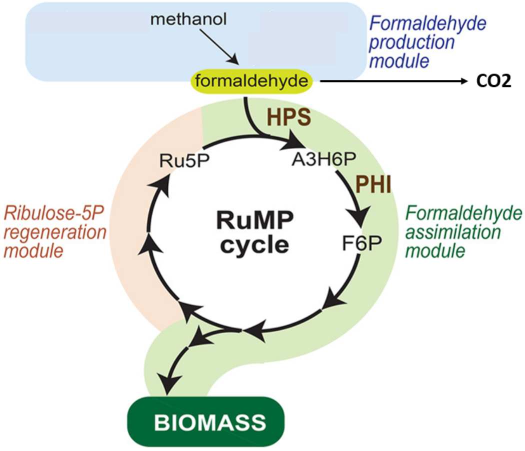

methanol
Formaldehyde production module
formaldehyde
CO2
HPS
Ru5P
A3H6P
PHI
Ribulose-5P regeneration module
RuMP cycle
F6P
Formaldehyde assimilation module
BIOMASS

Figure 2.1: A simplified illustration of the assimilatory RumP pathway gathered and modified from He et al. [5]. The figure shows how methanol is oxidized to formaldehyde before it is incorporated into the RuMP cycle by converting it to Hexulose-6-Phosphate (H-6-P) and further to Fructose-6-Phosphate (F-6-P), which is used to produce biomass and to regenerate Ribulose-5-phosphate (Ru-5-P). Alternatively formaldehyde can be oxidized to CO_2.

It has been demonstrated by Gao et al. [4] in 2022, that it is possible to introduce synthetic methylotrophy in *B. subtilis* by heterologous expression of Mdh. Gao et al. created several synthetic methylotrophic strains, some which could use methanol as carbon source, and some that were methanol dependent. However, none of the strains were able to grow on methanol as a sole carbon source, and they were dependent on a co-carbon source such as glucose, in order to grow and assimilate methanol [4]. This has been the case when inducing synthetic methylotrophy in *E. coli* and *C. glutamicum* as well [18, 19, 20].

Formaldehyde is a toxic compound to cells, and most cells therefore have detoxification systems to prevent formaldehyde from accumulating. *B. subtilis* is able to detoxify formaldehyde via the assimilatory RuMP-cycle , as previously explained. However, *B. subtilis* is also able to detoxify formaldehyde by dissimilation. This can either happen with the dissimilatory RuMP pathway [10, 11], or with the bacillithiol (BSH)-dependent formaldehyde oxidation pathway [6]. In both cases, formaldehyde is oxidized to CO_2. When formaldehyde is dissimilated, some energy is formed, but the carbon is secreted as CO_2 and is thus not used for biomass production [9].

2.2 Cryptic genes

Genes can become inactive due to accumulation of mutations. Different terms exist to describe such genes, such as pseudogenes, which are defined as a duplication of a functional gene, but that has become inactive due to accumulation of mutations. Another example of ”silent” DNA sequences are cryptic genes. The cryptic genes, in similarity to the pseudogenes, are genes that are inactive due to mutations, however, they can become active due to the right mutational events. In this way the cryptic genes may serve as an advantage to the host organism when it needs to adapt to a nutrient environment that requires the activity of the cryptic gene [21, 22].

2.3 Bioreactor cultivation

Bioreactors are used in biotechnological industry to cultivate microorganism. Often the microorganisms are used as cell factories that produce different products of interest. The bioreactor allows for control of the environment of the cells in order to promote production of the products of interest [23]. Bioreactors can be operated in different operation modes [24]. Two of the modes, batch operation mode and chemostat operation mode, will be described in more detail.

A picture of the type of bioreactor that was used is presented in Figure 2.2. A gas pump is connected to the bioreactor to allow a continuous flow of gas through the reactor. The medium is continuously agitated to ensure a uniform mixture of medium and bacterial cells. Furthermore, during aerobic fermentation, the agitation ensures that the levels of dissolved oxygen (DO) remain above a threshold level. The DO is monitored by a DO electrode. The temperature is monitored by a thermometer, and a heating jacket and cooling system is used to sustain the desired temperature. Furthermore, the pH of the medium is monitored by a pH electrode. The pH is regulated by adding acid or base from the acid/base pump.

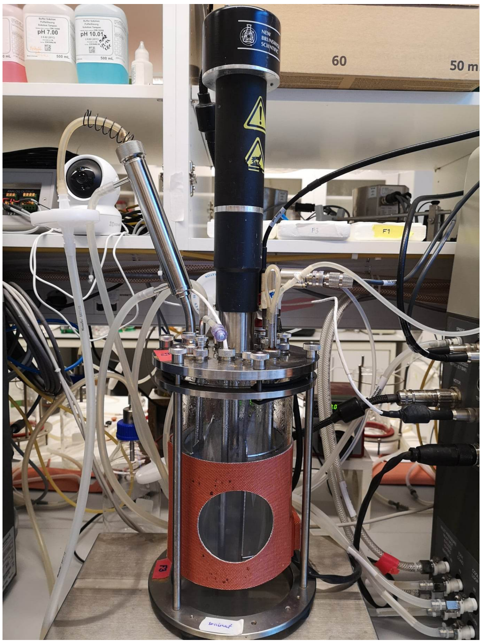


Figure 2.2: A picture of the type of bioreactor that was used in this work.

2.3.1 Batch operation mode

Before the cultivation can start, strain specific nutrient medium is aseptically added to the bioreactor. The cultivation is then started by adding a small volume of inoculum to the bioreactor. In batch operation mode, nothing is added or removed to the bioreactor after the cultivation has started, with three exceptions: a continuous flow of air is allowed to flow through the system, acid or base can be added to regulate the pH, and anti foam reagent can be added to regulate foam formation. The acid or base that is added is typically of high concentration to remain a close to constant volume in the bioreactor. Nothing else can be added or removed from the system, which means that no nutrients are added to the reactor after the cultivation has started [24].

In batch cultures, bacterial growth follows an S-curve. After inoculation, the bacteria need time to adjust to their new environment. This adjustment phase is called the lag-phase, and during this phase the bacteria grow slowly. When the bacteria have adjusted to their environment, they enter the exponential phase, where the bacteria use the available nutrients of the environment to grow exponentially. When the nutrient environment becomes scarce, the growth decreases and the bacterial cells enter the stationary phase. During this phase the division rate and death rate of the cells are equal, and the total biomass is kept constant. If no new nutrients are provided during this stage, the bacterial cells will eventually enter the death phase, where bacterial cells die and lyse faster than they divide. During this phase the biomass decreases. It is typical to harvest the cells during exponential phase or early stationary phase [24].

In our experiments the goal of the batch cultivations was to cultivate *B. subtilis* cells that could be used to analyse the biomass composition of the strain. Although the nutrient environment is not kept constant during a batch cultivation, the changes in nutrient environment are slow enough for the cells to reach metabolic steady state during exponential phase [25]. The biomass composition is therefore also held constant during exponential phase [26]. The bacterial cells were therefore harvested in late exponential phase before the cells reach stationary phase.

2.3.2 Chemostat operation mode

The chemostat operation mode differs from the batch operation mode as fresh nutrient medium is added to the system throughout the cultivation. As new medium is added, bioreactor fluid is removed to keep the volume in the bioreactor constant [24]. The rate at which new medium is added and old medium is removed is called the dilution rate [27]. In chemostat the cells can be kept in exponential phase for far longer than in batch cultivation, due to the constant supply of nutrients.

2.4 Measuring the biomass composition

The biomass composition of cells can be measured experimentally. In the following sections we will look into how the content of protein, lipid, DNA, and RNA can be measured.

2.4.1 Protein extraction

Liquid phase protein hydrolysis can be used to hydrolyze the proteins of a cell sample into their constituent amino acids. During liquid phase protein hydrolysis, boiling HCl is used to break the proteins into free amino acids [28]. After liquid phase protein hydrolysis has been conducted, the amino acids can be quantified with HPLC. The amino acid concentrations are used to calculate the relative amino acid distribution of the cell sample, and to find the total protein content of the cell sample.

During liquid phase protein hydrolysis, boiling HCl is used to break the proteins into free amino acids. The boiling HCl causes conversion of glutamine (gln) to glutamic acid (glu) and asparagin (asn) to aspartate (asp). Furthermore, the cystein (cys), proline (pro), tryptophane (trp) and methionine (met) are destroyed by the boiling HCl [29].

2.4.2 Lipid extraction

The lipid content of *B. subtilis* 168 was determined by using the lipid extraction method developed by Folch et al. in 1951 [30]. Folch's method uses a mixture of chloroform, methanol and water to extract the total lipid content of a cell sample. First, the cell sample is homogenized [30, 31] in a mixture of chloroform and methanol (2:1) (v/v). Both chloroform and methanol are organic solvents, however chloroform is non-polar, while methanol is polar. A mixture of the two solvents allows both non-polar and polar lipids to be dissolved. After the homogenization step, water is added to the mixture, which results in an aqueous phase and an organic phase. Non-lipid components and salts will be dissolved in the aqueous phase, while the lipids are dissolved in the organic phase. The organic phase is extracted, and the organic solvents are evaporated in order to quantify the lipid content.

2.4.3 DNA extraction

The DNA of a cell sample can be extracted by using a mixture of phenol and chloroform (1:1) (v/v). Before the DNA of the cell sample can be extracted, the cells are lysed to release the nucleic acids. When adding the phenol/chloroform mixture, the nucleic acids dissolve in the aqueous phase, while lipids are mainly dissolved in the organic phase. The proteins are found in an interface between the aqueous- and organic phase. The aqueous phase with nucleic acids is extracted, and ethanol and salt is added in order to precipitate the DNA. RNAse can be used to remove RNA from the sample. The concentration of the precipitated DNA can be measured spectrophotometrically [32]. The DNA concentration is used to find the total DNA content of the cells.

2.4.4 RNA extraction

The RNA of a cell sample can be extracted by using perchloric acid ($HClO_4$) [33]. Before the RNA can be extracted, the cells are lysed to release the nucleic acids. When $HClO_4$ is added, RNA dissolves in the $HClO_4$, while DNA and protein stay precipitated as they are not soluble in $HClO_4$. The concentration of the dissolved RNA can be measured spectrophotometrically.

2.5 Chromatography

Chromatography is a technique used to separate the compounds of a mixture. The separation process is based on the compounds interacting differently with a so-called stationary and mobile phase. The stationary phase is always composed of a solid material and is held in place. The mobile phase is composed of either a liquid (liquid chromatography) or a gas (gas chromatography), and migrates through the stationary phase. The compounds to be separated are applied onto the stationary phase and migrate through the stationary phase with the aid of the mobile phase. The compounds can be separated based on their difference in affinity for the stationary phase. Compounds with low affinity for the stationary phase will rapidly migrate through the system and will be detected early. Compounds with higher affinity for the stationary phase take longer to migrate through the system and will be detected later [34].

There are many types of chromatography, one of them being high-pressure liquid chromatography (HPLC). As the name implies, HPLC is a type of liquid chromatography where high pressure is applied to increase the flow rate of mobile phase through the system. This in turn speeds up the separation process, making HPLC an effective way of separating and identifying the compounds of a mixture [34]. In this work, HPLC was used to separate and quantify amino acids.

2.6 Mass spectrometry

Mass spectrometry is another technique that can be used to identify and quantify the compounds of a mixture. In mass spectrometry the compounds of the mixture are first broken down to smaller charged ions. Each ion is characterized by their molecular mass to charge (m/z) ratio. The ions are sorted based on their m/z ratio, and a mass spectrum is created with peaks that represent the m/z ratios that are present in the mixture. The size of a peak is proportional with the quantity of the ions with the given m/z ratio. The resulting mass spectrum can be compared to existing spectra to identify and quantify the compounds of the mixture. In this work, mass spectrometry was used to analyze the composition of output gas from the bioreactors.

2.7 Nuclear magnetic resonance (NMR)

Nuclear magnetic resonance (NMR) is used to identify and quantify the compounds of a sample based on the behaviour of atomic nuclei within a molecule. An NMR sample is first exposed to a strong magnetic field, to align the nuclei of the molecules. Then the sample is exposed to radio frequency pulses, which excites the nuclei of the sample. When the nuclei return to their lower energy state, the absorbed energy is emitted as radiation. The radiation is detected and used to generate an NMR spectra. In an NMR spectra each peak represents a type of nuclei, and the position of the peak reveals information about the chemichal environment surrounding the nuclei. The size of the peak is proportional to the quantity of said nuclei [35]. By comparing the NMR spectra to existing spectra, the compounds of the NMR sample can be identified. In this work, NMR has been used to identify and quantify compounds of the bioreactor fluid during cultivation, in order to calculate secretion and uptake rates of various compounds.

2.8 Uptake- and secretion rates

We differentiate between substrate uptake/secretion rate (u_s) and *specific* substrate uptake/secretion rate (q_s). u_s represents the rate at which substrate is consumed/produced, while q_s represents the substrate uptake/secretion rate per unit CDW. The sign of the rate reveals if the rate is an uptake rate or a secretion rate. A negative sign shows that the substrate is being removed, which means that the rate is an uptake rate. A positive sign shows that the substrate is being produced, which means that the rate is a secretion rate [36]. The sign of the rate is important when the rate is used to setting the constraints of a GEM. However, in the result section, both uptake- and secretion rates are presented as positive values.

2.8.1 Gas uptake- and secretion rates

Air is continuously flowing through the bioreactor during fermentation. The composition of the gas leaving the system (the output gas) can be measured, in order to determine specific uptake- and secretion rates of O_2 and CO_2, respectively. With the software that was used in this work, the composition of the gas leaving the bioreactor is returned as percentages of each gas present. For instance, the software could reveal that there is 20% O_2 in the output gas, meaning that 1% of the air has been taken up as O_2 in this time point (assuming there is 21% O_2 in the air). By converting this 1% to a fraction, and using the flow rate (Q) of air through the system, we can determine the volume of O_2 that is consumed per time unit. This rate can be given in number of moles O_2 consumed per time unit by using the concentration (N) of O_2, as shown in Equation 2.1:

$$u_{\mathrm{O}_2} = \frac{(O2_{out} - O2_{air})}{100} \cdot Q \cdot N \tag{2.1}$$

Here, u_{O_2} represents the O_2 uptake rate, while $O2_{out}$ and $O2_{air}$ represent the percentages of O_2 in output gas and air, respectively.

The specific O_2 uptake rate (q_{O2}) is found by dividing u_{O2}, found with Equation 2.1, by the CDW, as shown in Equation 2.2:

$$q_{\mathrm{O}_2} = \frac{v_{\mathrm{O}_2}}{CDW} \tag{2.2}$$

2.8.2 Uptake- and secretion rates of medium compounds

Substrate uptake/secretion rate of medium compounds u_s is calculated by finding the change in substrate concentration per time unit, as shown in Equation 2.3:

$$u_s = \frac{c_0 - c_1}{T_1 - T_0} \tag{2.3}$$

where c_0 and c_1 are the substrate concentrations at time T_0 and T_1, respectively. The specific substrate uptake rate (q_s) is found by dividing u_s by the average CDW between T0 and T1.

$$q_s = \frac{u_s}{\frac{\sum_{\mathrm{T}0}^{T1} CDW}{2}} \tag{2.4}$$

2.9 Genome-scale metabolic modeling

A metabolic network is a graphical representation of a metabolic pathway. A network or graph consists of nodes connected by links. The nodes and links of a metabolic network represent metabolites and the interactions between them, respectively. A metabolic network reconstruction of a pathway is built in a bottom-up-fashion, where one reaction is added at the time. Metabolic network reconstructions of different pathways can be combined to create more complex pathway maps. When a metabolic reconstruction includes all the known reactions and pathways that arise from the genome of an organism, the reconstruction is considered a genome-scale metabolic network [12].

From the genome-scale metabolic network reconstruction, the genome scale metabolic model (GEM) can be derived. While the genome-scale network reconstruction represents the metabolic network schematically, a GEM represents the information from the network reconstruction mathematically. The information of the network reconstruction is represented mathematically in a stoichiometric matrix, or **S**-matrix. In an **S**-matrix every column represents a chemical reaction, while every row represents a metabolite [12]. For instance, the hypothetical chemical formula A + B -> C, can be converted to the following **S**-matrix:

$$\begin{array}{c} A \\ B \\ C \end{array} \begin{pmatrix} -1 \\ -1 \\ 1 \end{pmatrix} \tag{2.5}$$

In the **S**-matrix above the rows represent the stoichiometric coefficients of A, B, and C, respectively. The negative coefficients of A and B show that they are used to produce C, which has a positive coefficient. A coefficient of zero would imply that the metabolite does not participate in the reaction. In a larger reconstruction, the **S**-matrix has one column for every chemical reaction, and one row for each metabolite. That leaves us with an **S**-matrix of size m × n, where m is the number of metabolites and n is the number of reactions [12].

An example of an arbitrary **S**-matrix with m metabolites and n reactions is shown below:

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & 3 & (...) & \text{n} \end{array} \\ \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ (...) \\ \text{m} \end{array} & \begin{pmatrix} -1 & 1 & -1 & & 0 \\ -1 & 1 & 0 & & 0 \\ 1 & 0 & 1 & & 0 \\ & & & & \\ 0 & -1 & 1 & & -1 \end{pmatrix} \end{array}$$

When the GEM has been derived and is mathematically represented as an **S**-matrix, different modeling approaches can be employed to study the genotype-phenotype relationship of the organism. One approach is constraint-based metabolic modeling. With this approach, a set of constraints are defined, to reflect the in vivo constraints of the system. For example, steady state constraints ensure that the production and consumption of each metabolite is balanced, and mass balance constraints ensure that the mass that enters the system equals the mass that leaves the system. Furthermore, nutrient uptake rates and energy requirements can be defined as constraints [12]. The mass balance of the system is ensured by the equation dx/dt = **S**v. Here, **x** is a vector of length m containing all the metabolite concentrations, **S** is the **S**-matrix, and **v** is a flux vector of length n which holds the flux through each reaction. The steady state is ensured by the equation **Sv** = 0. This equation gives a set of linear equations that must be satisfied for mass balance to be maintained. In a genome scale metabolic reconstruction, there are usually more reactions than metabolites (n>m). The set of linear equations therefore defines a feasible solution space, and there are more than one possible flux vector **v** that satisfy the set of linear equations [37]. An example of what the **Sv** = 0 equation looks like is shown below with an arbitrary **S**-matrix and **v**.

$$\begin{array}{cc} & \begin{array}{ccccc} 1 & 2 & 3 & (...) & \text{n} \end{array} \\ \begin{array}{c} \text{A} \\ \text{B} \\ \text{C} \\ (...) \\ \text{m} \end{array} & \begin{pmatrix} -1 & 1 & -1 & & 0 \\ -1 & 1 & 0 & & 0 \\ 1 & 0 & 1 & & 0 \\ & & & & \\ 0 & -1 & 1 & & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ ... \\ v_n \end{pmatrix} = 0 \end{array}$$

From this equation, the following set of linear constraints can be derived:

$$-v_1 + v_2 - v_3 \; ... = 0$$

$$-v_1 + v_2 \; ... = 0$$

$$v_1 + v_3 \; ... = 0$$

$$...$$

$$-v_2 + v_3 \; ... \; - v_n = 0$$

In addition to the mass balance equation, each reaction has a lower- and upper flux bound that restrains the possible flux through the reaction [38].

When the constraints have been defined, constraint based analysis methods can be employed to answer different questions about the system. Examples of such methods are flux balance analysis (FBA), flux variability analysis (FVA) and minimization of metabolic adjustment (MOMA). FBA answers the question of what the optimal solution of the chosen objective is [12]. To conduct FBA, first a model objective must be chosen. The model objective can either be set to production of a metabolite of interest, or to biomass production. If biomass production is chosen as objective, FBA will find the maximal flux that is possible through the biomass synthesis reaction, in other words the maximal possible growth rate, while maintaining the constraints that have been defined. FBA finds a flux distribution through the metabolic network that yields this optimal solution. The flux distribution is given as a flux vector **v**. As mentioned before, the flux vector returned by FBA is not necessarily unique, as different flux vectors could yield the same optimal solution [37]. To summarize, FBA solves the following linear optimization problem:

$$\text{max } v_{objective}, \text{ subject to:}$$

$$\mathbf{Sv} = 0$$

$$lb_i \leq v_i \leq ub_i$$

Where $v_{objective}$ is the flux of the reaction chosen as objective, **S** is the **S**-matrix, v_i are the fluxes of flux vector **v**, lb_i is the lower bound, and ub_i is the upper bound for each reaction i. FVA on the other hand, answers the question of what the minimal and maximal possible flux of each reaction in the network is. This is done by iteratively solving two optimization problems for each reaction of the network: one minimization and one maximization problem. Before solving the minimization and maximization problem for each reaction, the optimal solution to the original objective function is found with FBA. Then the flux of the objective function is set to a fraction of the optimum, for instance 0.99. Then the objective is iteratively changed to maximize and minimize the flux of each reaction of the network. In that way, FVA finds the flux range of each reaction that would yield the defined fraction of the optimal solution [38]. The FVA problem can be written mathematically as follows:

$$\text{max/min } v_i, \text{ subject to:}$$

$$\mathbf{Sv} = 0$$

$$v_{objective} \geq \gamma z$$

$$lb_i \leq v_i \leq ub_i$$

Where $v_{objective}$ is the flux of the original objective function, z is the original optimal solution and γ is the fraction of the optimum ($0 \leq \gamma < 1$) [38].

According to Segré et al.[39], FBA predictions are more accurate for wild type strains than for mutant- or genetically engineered strains. This is because wild type strains typically have reached an optimal growth state due to evolutionary pressure over time. When a mutation is introduced to a strain, it usually takes several generations for the strain to readjust and reach a new optimal growth state. FBA does not take this into account, and might therefore not be the most suitable tool to study the flux distribution of a mutant- or genetically engineered strain. Segré et al. have addressed this problem and developed a method to find a sub-optimal solution as close to the FBA solution as possible, that takes the mutational constraints into account. Their method is based on the assumption that the metabolic flux of the mutant is initially rearranged as little as possible compared to the wild type, hence the name minimization of metabolic adjustment (MOMA). As previously explained, FBA finds a flux vector v in the feasible solution space that yields the optimal solution to the chosen objective function. When introducing a mutation, the feasible solution space is limited. The goal of MOMA is to find a flux vector x in the new solution space that has minimal Euclidean distance to the flux vector v. Thus, MOMA searches to minimize the following:

$$D(v,x) = \sqrt{\sum_{i=1}^{\mathrm{n}} (v_i - x_i)^2}$$

where D(**v**, **x**) is the Euclidean distance between vectors **v** and **x**. MOMA is solved as a quadratic optimization problem [39].

These tools, and many more, make the GEM a valuable resource to study the genotype-phenotype relationship in both wild type and mutant strains [12].

2.10 The biomass objective function

As explained in the previous section, the objective function can either be set to production of a metabolite of interest, or to biomass production. When choosing biomass production as the objective, FBA can be employed to find the maximal possible flux of the biomass synthesis reaction. The biomass synthesis reaction, also called the biomass objective function (BOF) is a formulation of the content of macro-molecules such as protein, DNA, RNA, and lipid that are required to form biomass. The energy that is required for growth associated processes, are also expressed in the BOF, as ATP hydrolysis. The BOF can be formulated in different ways. One way is to formulate the BOF in one single reaction, where all the macro-moleculre constituents such as amino acids, nucleotide triphosphates, fatty acids etc. are used as substrates to produce biomass directly. Alternatively, the BOF can be formulated with the macro-molecules, such as protein, DNA, RNA, lipid, etc. as substrates. In this case, each macro-molecule is given a corresponding synthesis reaction: protein is synthesized from amino acids, DNA and RNA are synthesized from nucleotide triphosphates, lipids are synthesized from fatty acids, etc. Independently of how the BOF is formulated, the BOF should reflect the experimentally measured biomass composition and energy requirements of the organism [40].

2.11 Growth associated and non-growth associated maintenance

In a GEM, two types of energy requirements are formulated. The energy that is required for growth related processes is called growth associated maintenance (GAM). Examples of growth related processes are macro-molecule synthesis, and cell replication. The GAM is formulated in the BOF as ATP hydrolysis. By connecting the GAM to the BOF, an increase in biomass synthesis will simultaneously require more ATP hydrolysis. The GAM can either be determined experimentally, or it can be estimated based on the energy requirements for macro-molecule synthesis. The other energy requirement is that of non-growth associated processes. This energy requirement is called non-growth associated maintenance (NGAM), and include the energy requirements for processes such as turgor pressure maintenance. The NGAM is represented as an ATP-hydrolysis reaction separate from the BOF. The NGAM can be determined experimentally, or it can be estimated [41].

2.12 Genome-scale metabolic models of *Bacillus subtilis*

There are several available GEMs of *B. subtilis* 168, where three of them will be introduced in this section: iYO844, iBsu1103, and iBsu1147 [13, 42, 43]. iBsu1147 is the newest GEM of *B. subtilis*, and is based on the two previous GEMs, iYO844 and iBsu1103. There are several differences between the three GEMs, however the difference that will be highlighted in this section is the difference in how the BOF has been formulated in the three models.

The first GEM of *B. subtilis* is called iYO844, and was published in 2007 by Oh et al. [13]. The BOF of iYO844 is structured as one single reaction containing all the macro-molecule constituents necessary to synthesize biomass [13]. A later reconstruction of *B. subtilis* was published in 2009 by Henry et al. [42], and is named iBsu1103. The BOF of iBsu1103 is based on the one constructed by Oh et al., but is structured differently. Henry et al. have divided the constituents of the BOF into seven categories: protein, DNA, RNA, lipid, lipoteichoic acid, cell wall, and cofactors and ions. The first six categories represent the macro-molecules that are necessary to synthesize biomass, and have been given each their synthesis reaction. The BOF takes in the products of the six synthesis reactions, as well as the cofactors and ions that are necessary for biomass production and energy maintenance. Henry et al. chose to structure the BOF this way to make it more transparent, as it makes it easier to understand what constituent that is necessary for what macro-molecule [42]. The latest GEM of *B. subtilis*, called iBsu1147, was published in 2013 by Hao et al. iBsu1147 is based on the previous GEM, iBsu1103, in addition to information from KEGG and Uniprot. iBsu1147 uses the same BOF as presented in iBsu1103, but with KEGG IDs where possible [43]. iBsu1147 is the model that has been used in this work to investigate the genotype-phenotype relationship of the wild type and Mdh-expressing mutant, as it is the newest GEM available for *B. subtilis*.

Chapter 3

Software and methods

The supplementary files that are reffered to in this Chapter can be found in my github repository: `https://github.com/linnsandvik/Supplementary-data-Master-thesis.git`.

3.1 The cryptFind algorithm

The cryptFind algorithm developed by Navid and Almaas [44] is developed to find potential cryptic genes of a genome-scale metabolic model (GEM). The algorithm operates in two steps. First the genes that are essential for growth on the selected carbon source are found. Second, the genes that are essential for each of the carbon sources in a growth medium list are found. The algorithm returns the essential genes that are unique for growth on the selected carbon source. The algorithm was implemented as described by Snorre Sulheim in 2017. The code written by Sulheim can be found here. In our case, the selected carbon source is methanol. The growth medium list contains 54 carbon sources that *B. subtilis* is able to use in vivo, and that the iBsu1147 GEM can use in silico. How these were found is explained in the next section

3.1.1 Finding carbon sources that *B. subtilis* can use for growth in vivo and in silico

To use the cryptFind algorithm we need a list of carbon sources that *B. subtilis* grows on in vivo [44]. Oh et al. have tested what carbon sources *B. subtilis* is able to use for growth in vivo [13]. Then we have identified which of these carbon sources are represented in exchange reactions in iBsu1147. A table of carbon sources that *B. subtilis* is able to use for growth in vivo, and that are represented as exchange reactions in iBsu1147 can be found in sheet 2 in file ”C_sources_in_silico.xlsx”. To check which of the carbon sources that can be used for growth by iBsu1147, we iteratively change the nutrient medium to be limited by each of the carbon sources that were found. If the predicted growth rate that is found with FBA is positive, it means that iBsu1147 can use the carbon source for growth.

It was found that iBsu1147 is able to grow on all carbon sources listed except for one, which is C00182: Glycogen.

3.2 *Bacillus subtilis* 168 and *B. subtilis* 168 pBV2mp_mdhBm

The two bacterial strains that were used in this work are *Bacillus subtilis* 168 and *B. subtilis* 168 pBV2mp_mdhBm. The expression plasmid that was used, pBV2mp, is a plasmid of low copy number that originates from *B. subtilis* [45]. The Mdh gene was isolated from the genomic DNA of *Bacillus methanolicus* in accordance to Eikmanns et al. 2013 [46]. In the rest of this work, the term wild type refers to *B. subtilis* 168, and the term mutant refers to *B. subtilis* 168 pBV2mp_mdhBm. The mutant strain was provided by Vivien Jessica Klein, who is an engineer and PhD candidate for the Department of Biology at the Faculty of Natural Sciences at NTNU. She also provided us with the empty plasmid strain *B. subtilis* 168 pBV2mp, which was originally a strain we were going to use as a control in addition to the wild type strain. Unfortunately, we were unable to make *B. subtilis* 168 pBV2mp grow, and it was therefore decided to use only the wild type as control.

3.3 LB medium

LB medium was used to create glycerol stocks, as described in Section 3.5, and precultures, as described in Section 3.6. LB medium was made by dissolving tryptone, NaCl and yeast extract in Milli-Q (MQ-H_2O) to obtain the concentrations listed in Table 3.1. The LB medium was sterilized by autoclavation for 20 minutes at 121°C. When LB medium was used to cultivate *B. subtilis* 168 pBV2mp_mdhBm, 50 mg/mL kanamycin stock solution was added after autoclavation to obtain a final concentration of 5 μg/mL.

Table 3.1: Components used to make LB medium. The components were dissolved in Milli-Q (MQ-H_2O) to obtain the listed concentrations. Supplier and CAS number of each component are provided.

Component	**Concentration (gL^{-1})**	**Supplier**	**CAS number**
Tryptone	10	Sigma-Aldrich	91079-40-2
NaCl	10	VWR chemicals	7647-14-5
Yeast extract	5	Sigma-Aldrich	8013-01-2

3.4 Minimal medium

Minimal media were used for precultures, as described in Section 3.6, and during bioreactor cultivation as described in Subsection 3.7.1. Two types of minimal media were prepared, one with D-glucose as only carbon source, and one with both methanol and D-glucose as carbon sources. Stock solutions of salts and carbon sources, trace mineral solution (TMS) and phosphate buffer were prepared in advance to ease the preparation of the minimal media.

3.4.1 Stock solutions

A stock solution is prepared by dissolving a component in MQ-H_2O to obtain a desired concentration. The components used to make stock solutions, and their final concentration are listed in Table 3.2. The stock solutions were stored at room temperature.

Table 3.2: Components used to make stock solutions, and their final concentrations. The components were dissolved in Milli-Q (MQ-H_2O) to obtain the listed concentrations. Supplier and CAS number of each component are provided.

Component	**Concentration (g L^{-1})**	**Supplier**	**CAS number**
NaCl	50	VWR Chemicals	7647-14-5
NH_4Cl	60	Sigma-Aldrich	12125-02-9
$MgSO_4 \cdot H_2O$	246.5	Sigma-Aldrich	10034-99-8
D-Glucose	400	VWR Chemicals	50-99-7

3.4.2 Trace mineral solution (TMS)

The trace mineral solution (TMS) contains all the minerals that *B. subtilis* needs in order to grow. TMS was made by dissolving the components listed in Table 3.3 in 5 M HCl to obtain the listed concentrations. HCl was used instead of MQ-H_2O to prevent precipitation of the minerals. The TMS was stored in a 1000 mL Reagent Bottle covered in aluminum foil in refrigerator at 4°C.

Table 3.3: Components used to make trace mineral solution (TMS). The components were dissolved in 5 M HCl to obtain the listed concentrations. Supplier and CAS number of each component are provided.

Component	**Concentration (g L^{-1})**	**Supplier**	**CAS number**
$FeCl_2 \cdot 4H_2O$	7.2	Sigma-Aldrich	13478-10-9
$ZnCl_2$	0.5	Sigma-Aldrich	7646-85-7
$CaCl_2 \cdot 2\ H_2O$	0.5	Sigma-Aldrich	10035-04-8
$CuCl_2 \cdot 2\ H_2O$	1	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

3.4.3 Phosphate buffer

Phosphate buffer (PO_4-buffer) is added to the minimal medium to ensure stable pH. PO_4-buffer was made by dissolving disodium phosphate ($Na_2HPO_4 \cdot 7H_2O$) and monopotassium phosphate (KH_2PO_4) in MQ-H_2O to obtain final concentrations of 112 g L^{-1} and 30 g L^{-1}, respectively. 4 M NaOH was used to adjust the pH to 7.2. The PO_4-buffer was stored at room temperature.

3.4.4 Preparing minimal medium

Minimal medium was prepared in two steps. First, components that can be autoclaved without being degraded were dissolved in MQ-H_2O. That includes sodium chloride (NaCl) stock, ammonium chloride (NH_4Cl) stock, PO_4-buffer, L-trp and L-glu. This mixture was sterilized by autoclavation for 20 minutes at 121 °C in the bioreactor. (Assembly of the bioreactor is described in Section 3.7.1.) Secondly, TMS, magnesium sulfate ($MgSO_4 \cdot 7 H_2O$) stock and carbon source(s) were added to the medium in sterile cabinet. Magnesium sulfate- and D-glucose stock were sterile filtered before adding them to the medium. TMS and methanol were not sterile filtered, based on an assumption of no microbial growth in these solutions. The final concentrations of each medium component are listed in Table 3.4 and Table 3.5. When both methanol and D-glucose were used as carbon source, methanol was added in sterile cabinet to obtain a concentration of 200 mM in the final medium. Furthermore, when minimal medium was used to cultivate *B. subtilis* 168 pBV2mp_mdhBm, 50 mg/mL kanamycin stock solution was added in sterile cabinet to obtain a final concentration of 5 μg/mL.

Table 3.4: Stock solutions and components used to make minimal medium. The concentration of the components in stock solution and in minimal medium is given. Milli-Q (MQ-H_2O) was used to obtain the desired concentration of each component.

Stock solution/ component	**Concentration stock (g L^{-1})**	**Concentration medium (g L^{-1})**
NaCl	50	0.5
NH_4Cl	60	0.3
Na_2HPO_4*	112	11.2
KH_2PO_4*	30	3
L-trp**	-	0.05
L-glu**	-	2
TMS	See Table 3.5	See Table 3.5
MgSO4·7H2O	246.5	0.493
D-glucose	400	10

* Na_2HPO_4 and NH_4Cl are dissolved in the PO_4-buffer as described in Section 3.4.3.

** L-trp and L-glu were added as solids.

Table 3.5: Components of the trace mineral solution (TMS) and their concentrations (g L^{-1}) in TMS and minimal medium.

Component	**Concentration TMS (g L^{-1})**	**Concentration medium (g L^{-1})**
$FeCl_2 \cdot 4H_2O$	7.2	$9.6 \cdot 10^{-3}$
$ZnCl_2$	0.5	$6.7 \cdot 10^{-4}$
$CaCl_2 \cdot 2\,H_2O$	0.5	$6.7 \cdot 10^{-4}$
$CuCl_2 \cdot 2\,H_2O$	1	$1.3 \cdot 10^{-3}$
$MnCl_2 \cdot 4\,H_2O$	0.2	$2.7 \cdot 10^{-4}$
$CoCl_2 \cdot 6\,H_2O$	0.05	$7 \cdot 10^{-5}$
$Na_2MoO_4 \cdot 2\,H_2O$	0.01	$1 \cdot 10^{-5}$

3.5 Glycerol stocks

Glycerol stocks were prepared of *B. subtilis* 168. A 250 mL baffled shake flask with 50 mL LB medium was inoculated with *B. subtilis* 168 and incubated for 16 hours at 200 RPM and 37 °C. After incubation, the bacterial cells were harvested by transferring the culture to a sterile 50 mL centrifuge tube and centrifuging it for 5 min at 4000 RCF and 4°C. To create a concentrated cell solution, most of the supernatant was discarded to reach a volume of 16 mL, and the pellet was re-suspended by gentle pipetting. To create glycerol stock from the concentrated cell solution, 4 mL sterile 80 % glycerol was added, which gave a final concentration of 16 % glycerol. The glycerol was mixed into the bacterial culture by inversion. The glycerol stock was then divided over sterile cryotubes (1 mL per tube) and snap-freezed in liquid nitrogen, before it was stored in the freezer at -80 °C.

3.6 Precultures

3.6.1 Wild type in glucose minimal medium

First, preculture using LB medium was made by inoculating 100 mL LB-medium in 250 mL baffled shake flask with 100 μL *B. subtilis* glycerol stock. The preculture was incubated for eight hours at 37 °C and 200 rpm. After incubation, two new 250 mL baffled shake flasks with 100 mL glucose minimal medium were inoculated with 75 and 150 μL LB preculture, respectively. The minimal medium precultures were incubated over night in 37 °C and 200 rpm. The minimal medium preculture inoculated with 75 μL LB preculture was used to inoculate the medium in the bioreactor.

3.6.2 Wild type in minimal medium with glucose and methanol

The precultures of *B. subtilis* 168 WT in minimal medium containing D-glucose and methanol were made by following the same procedure as described in Section 3.6.1. In these precultures, the minimal medium that was used contained methanol in addition to D-glucose.

3.6.3 Mutant in glucose minimal medium

First, preculture using LB medium was made by inoculating 100 mL LB-medium in a 250 mL baffled shake flask with 100 μL glycerol stock of the *B. subtilis* mutant strain. The LB preculture was incubated at 37 °C and 200 revolutions per minute (RPM) for 8.5 hours. Then, a first minimal medium preculture was made by inoculating 100 mL glucose minimal medium with 150 μL of the LB preculture. A 250 mL baffled shake flask was used for the first minimal medium preculture, which was incubated at 37 °C and 200 RPM for 12 hours. A second minimal medium preculture was made by inoculating a 150 mL baffled shake flask containing 75 mL of glucose minimal medium with 750 μL of the first minimal medium preculture. The second minimal medium preculture was incubated at 37 °C and 200 RPM for 7 hours. A third minimal medium preculture was made by inoculating 100 mL glucose minimal medium with 1 mL of the second minimal medium preculture. A 250 mL baffled shake flask was used for the third minimal medium preculture, and it was incubated for 17 hours at 37 °C and 200 RPM. The third minimal medium preculture was used to inoculate the medium in the bioreactor.

3.6.4 Mutant in minimal medium with glucose and methanol

LB preculture was made by inoculating a 250 mL baffled shake flask containing 100 mL LB-medium with 100 μL glycerol stock of the *B. subtilis* mutant. The LB preculture was incubated at 37 °C and 200 RPM for 10 hours. Then, 20 mL LB preculture was centrifuged for 4 minutes at 4 °C and 4500 RPM, and the pellet was used to inoculate 75 mL minimal medium containing glucose and methanol. A 150 mL baffled shake flask was used. The culture was incubated at 37 °C and 200 RPM for 12 hours. The minimal medium preculture was used to inoculate the medium of the bioreactor.

3.7 Batch culturing in bioreactors

3.7.1 Cultivation

The *B. subtilis* cells were cultivated in 3.0 L New Brunswick™ BioFlo® 115 Benchtop Bioreactor system. 1.5 L of minimal medium was used each cultivation. The system was assembled as described in the manufacturer's user manual.

The day before cultivation, the bioreactors were cleaned first with distilled water (dH_2O) and then with MQ-H_2O. The components of the minimal medium that can be autoclaved without being degraded were dissolved in MQ-H_2O and added to the bioreactors, as described in Section 3.4.4. The pH electrodes were calibrated between pH 4 and 7. The dissolved oxygen (DO) electrodes was calibrated to 0 by sparging with nitrogen gas for about 20 minutes. The reactors were then assembled and autoclaved at 121 °C for 20 minutes.

The rest of the medium components were added to the reactors in sterile cabinet the before cultivation. Then, the motor, heat jacket, sparge air, base pump and cooling water were connected to each reactor. Before calibrating the DO-electrode to 100 % oxygen, the agitation was set to 500, air flow to 1.5 L min^{-1} and the temperature was allowed to stabilize at 37 °C. For the rest of the cultivation, the agitation ranged between 200 and 1000. The gas leaving the reactor (output gas) was measured using Thermo Scientific™ Prima BT Benchtop Process Mass Spectrometer and the data was processed by Thermo Scientific™ GasWorks Process Analysis Software. New Brunswick™ BioCommand® Software was used to track the agitation, pH, temperature, and amount of DO. The reactors were inoculated with 50 mL preculture.

3.7.2 Sampling

Samples of the bacterial culture were collected several times throughout each cultivation. This was done to track the cell growth and medium composition. The cell growth was tracked by measuring OD600 and CDW, while the medium composition was analyzed with NMR. The OD600 was measured of 1 mL culture, and minimal medium was used as blank. To analyze the medium composition and determine the cell dry weight (CDW), 3 mL culture was collected from the bioreactor. The sample was centrifuged for 5 minutes at 4 °C and 4500 RCF. The supernatant was sterile filtered and stored in freezer at -20 °C until further NMR analysis, described in Section 3.7.4. The pellet was resuspended in 3 mL Q-H_2O and distributed over three pre-weighed aluminum pans. The aluminum pans with content were dried in oven over night at 100 °C before they were weighed again to determine the CDW.

Furthermore, proteomics samples of each fermentation were prepared when the cells reached OD600 = 2. 3 mL sample was collected and distributed over three 1.5 mL eppendorf tubes and centrifuged at 13.4k RPM for 5 minutes in table centrifuge. The supernatant was discarded and eppendorf tubes with pellets were stored in freezer at -20 °C until further proteomics analyses.

3.7.3 Harvesting and freeze-drying

The cultivation ended with harvesting when the OD600 had reached about 3. The exact OD600 for each bioreactor cultivation is listed in Table 3.6. A syringe was used to empty the bioreactor as much as possible, and the culture was divided over 50 mL centrifuge tubes. The tubes were centrifuged for 4500 RCF for 5 minutes at 4 °C, the supernatant was discarded and the pellet was washed twice with 15 mL 10 g/L NaCl. Both wash steps were conducted by re-suspending the pellet in NaCl solution by vortexing, before the samples were centrifuged and supernatant discarded. After the two NaCl wash steps, a last wash step was conducted the same way, only with 20 mL MQ-H_2O instead of NaCl solution. The cell pellets were stored in the freezer at -80 °C before being lyophilized using Christ Alpha 3-4 LSC basic freeze dryer with for 3 days. Then, the lyophilized cells were stored in -80 °C freezer until further biomass analyses.

Table 3.6: OD600 of each bioreactor cultivation when starting the harvesting.

Bioreactor cultivation	OD600
WT in glucose	3.44
WT in glucose and methanol	4.00
MDH in glucose	2.99
MDH in glucose and methanol	3.00

3.7.4 Preparation of NMR samples

NMR samples were prepared from the frozen media samples that were collected throughout the cultivations (described in Subsection 3.7.2). First, the samples were thawed and vortexed. For each sample, triplicates of 800 μL sample was transferred into new 1.5 mL eppendorf tubes, and 80μL D20-TSP solution was added. Each eppendorf tube was then vortexed before 600 μL of the content was pipetted into NMR-tubes. The further NMR analyses were carried out by Christian Schulz.

3.8 Biomass composition measurements

The biomass composition was measured to find the macromolecular biomass fractions of proteins, RNA, DNA, and lipids.

3.8.1 RNA quantification

For this section, all centrifuge steps were conducted for 10 minutes at 4000 RCF and 4 °C, except for the last centrifuge step where 4000 RPM was used.

Triplicates of 10 mg lyophilized cells were weighed into 15 mL centrifuge tubes. The cells were washed three times with 3 mL 0.7 M $HClO_4$. For each wash step, the cells were re-suspended by vortexing, before the samples were centrifuged and the supernatant was discarded. Then, the pellets were re-suspended by vortexing in 3 mL 0.3M KOH, followed by 1 h incubation in a thermocycler at 37 °and 400 rpm. The samples were cooled down to room temperature before adding 1 mL 3M $HClO_4$. The samples were mixed by inversion. Then the samples were centrifuged and the supernatants were decanted into 50 mL centrifuge tubes. The remaining pellets were washed twice with 4 mL 0.5 M $HClO_4$. For both wash steps, the pellets were re-suspended by vortexing, before the samples were centrifuged and the supernatant was decanted into the 50 mL tubes. Then 3 mL 0.5M $HClO_4$ was added to each 50 mL tube to obtain a final volume of 15 mL, before the samples were centrifuged to get rid of non-visible precipitates of $KClO_4$. The concentration of RNA was measured at 260 nm by using Thermo Scientific™ NanoDrop™ One spectrophotometer with the single-stranded RNA setting. The absorbance ratios of 260 nm over 280 nm and 260 nm over 230 nm were also measured. 0.5 M $HClO_4$ was used as blank.

3.8.2 DNA quantification

Triplicates of 10 mg dry biomass was weighed out and added to 1.5 mL eppendorf tubes. Then 600 μL lysis buffer was added to each eppendorf tube, before the cell samples were incubated at 55 °C for 30 minutes in a thermocycler. The samples were cooled down to room temperature before 600 μL phenol/chloroform (1:1) (v/v) was added to each sample. The samples were inverted until the phases were completely mixed. Then, the samples were centrifuged at 13.4k RPM for 5 minutes, which resulted in a white protein layer between the aqueous and the organic layers. The upper aqueous phase was transferred to new 1.5 mL eppendorf tubes by using 1 mL micropipette where approximately 2 mL of the tip had been cut off, in order to avoid disturbing the protein layer. The steps from and including adding 600 μL phenol/chloroform were repeated twice for the new eppendorf tubes.

To remove phenol, 600 μL of chloroform was added to the tubes, and the tubes were mixed by inversion before they were centrifuged at 13.4k for 5 minutes. Then, the aqueous phase was transferred to a new tube. Then, 40 μ L 3 M sodium acetate and 1 mL ice cold ethanol (99%) was added and the samples were mixed gently by inversion. This resulted in visual DNA precipitation. Then the tubes were stored overnight at -20 °C.

Then the samples were centrifuged at 13400 RPM for 15 minutes, resulting in a transparent DNA-pellet. The supernatant was discarded, and the DNA-pellet was rinsed with 1 mL 70 % ethanol. Then the samples were centrifuged at 13.4k RPM for 2 minutes, the supernatant was carefully discarded, and the DNA-pellets were air-dried for one hour by placing the open tubes in a thermocycler at 37 °C. The dry DNA-pellets were re-suspended in 50 μL TE-buffer by pipetting. Then, 1 μL of RNAse A was added and the samples were incubated for 15 min at 37 °C in a thermocycler. The concentration of DNA at 260 nm, as well as the absorbance ratios of 260 nm over 280 nm and 260 nm over 230 nm were determined using Thermo Scientific™ NanoDrop™ One spectrophotometer and the double-stranded genomic DNA setting. A mix of 50 μL TE-buffer with 1 μL RNAse was used as blank.

3.8.3 Lipid quantification

Triplicates of 40 mg lyophilized biomass was weighed into eppendorf tubes. 150 μL MQ-H_2O was added to each sample, and the samples were vortexed. 0.5 g zirconium beads (1.4 mm) and 400 μL methanol were added to each sample. A homogenizer with settings 20 second interval, 2 cycles, at 5000 rpm were used for all samples, and samples were put on ice in between cycles. 800 μL chloroform was added and the samples were vortexed for 20 min. 100 μL MQ-H_2O was added and the samples were vortexed for 10 min. The samples were centrifuged for 4 min at 13.4k rpm using table centrifuge. As much organic phase as possible was transferred with needle to dark MS-vials. 600 μL chloroform was added to the bead-cell-solution and the samples were vortexed for 10 min, centrifuged for 4 min at 13.4k rpm, before again as much organic phase as possible was transferred to the dark MS-vials. The chloroform was evaporated over the weekend before the MS-vials were weighed to determine the exact weight of the extracted lipids. Supplementary data for the biomass compositions can be found in ”Biomass data.xlsx”.

3.8.4 Protein hydrolysis and HPLC

Triplicates of 10 mg lyophilized cells were weighed into schott flasks. 500 μL 6 M HCl was added to each sample. The flasks were put to boil at 105 °C for 24 hours. After 24 hours the flasks were allowed to cool down to a handling temperature, before the samples were neutralized by adding 500 μL 6 M NaOH. The samples were mixed well before they were filtered and transferred to empty eppendorf tubes by using syringe with filter (0.22 μm) attached. The following dilutions were prepared of each filtered sample: 1:50, 1:100, and 1:250. 200 μL of each of the final dilutions were transferred to HPLC-vials with solid caps. The samples were stored in the freezer (-32 °C) until the samples were analyzed with HPLC by Siri Stavrum.

3.8.5 Calculating the amino acid distribution and total protein content

The amino acid content was measured by acid hydrolysis followed by HPLC. The HPLC-results reveal the concentration (μmol/L) of each amino acid present in the HPLC sample. The dilution factor, total volume of the sample, and the CDW used to prepare the sample were used to calculate the molar fraction (mmol/gDW) of each amino acid.

Acid hydrolysis causes met, cys, pro, and trp to be degraded [29]. The concentration of these amino acids was therefore predicted by using linear regression of the molar fractions of the remaining 16 amino acids calculated from the HPLC results, against molar fractions predicted from a script written by Vetle Simensen. A linear regression was conducted for each fermentation. The script written by Vetle Simensen allows us to predict the molar fractions of each amino acid based on the amino acid distribution derived from the protein coding gene sequences of *B. subtilis* 168 (UniProt Proteome ID UP000001570). The script can be found here.

Furthermore, during boiling in HCl, gln is converted to glu and asn is converted to asp [29]. The HPLC results therefore return a higher concentration of glu and asp, while the concentration of gln and asn is approximately zero. This means that the HPLC concentration of glu actually represents the concentration of glu and gln combined. The same goes for asn and asp. Furthermore, the concentration of glycine (gly) and arginine (arg) is returned as a sum in the HPLC results. The same script was used to predict the molar fraction of these amino acids, and the predicted molar fractions were used to find the experimentally measured molar fractions. For instance, the prediction the molar fraction of asn and asp are 0.30 and 0.40 mmol/gDW, respectively. This means that from the measured Asp concentration, 57 % is asp while 43 % is asn. The molar fraction of asp, asn, glu, gln, gly and arg were found with this strategy. Eventually, the mass fraction (g/gDW) of each amino acid was calculated from the molar fractions (mmol/gDW) by using the molar mass of each amino acid. The total protein content was found by summarizing the mass fractions of all amino acids. The calculations of the amino acid distribution and total protein content are attached in supplementary file ”protein synthesis reactions.xlsx”.

3.9 Growth rates

R version 4.2.2 was used to fit linear regression models and to conduct Analysis of Variance (ANOVA). Triplicate measurements of the CDW were conducted throughout each fermentation. For each fermentation, a linear regression model was fitted where y = CDW on log scale, and x = time (h). The growth rate corresponds to the slope of the linear regression model. R version 4.2.2 was used to fit linear regression models.

3.10 Finding the specific O_2 uptake- and CO_2 secretion rates

Thermo Scientific™ Prima BT Benchtop Process Mass Spectrometer was used to measure the composition of the output gas, and the data was processed by Thermo Scientific™ GasWorks Process Analysis Software. For each fermentation, the flow of air through the system was allowed to stabilize before inoculation of the reactor. The $O2_{air}$ used in Equation 2.1, is calculated for each cultivation by finding the average O_2 content of 10 output gas measurements conducted after the air flow had stabilized and before inoculation. Furthermore, the GasWorks Process Analysis Software returns the flow rate of air through the system for each output gas composition measurement. This flow rate is used for Q, in Equation 2.1. The CO_2 secretion rate ($u_{\mathrm{CO_2}}$) and specific CO_2 secretion rate ($q_{\mathrm{CO_2}}$) is found by using Equation 2.1 and 2.2, where $O2_{out}$ and $O2_{air}$ are replaced with $CO2_{out}$ and $CO2_{air}$, respectively, and by using the concentration (N) of CO_2 in air. The concentration (N) of O_2 and CO_2 that are used in Equation 2.1 are shown in Table 5.8 in Appendix 5.

To find the specific substrate uptake rate (q_s), we need to know the CDW of *B. subtilis* cells in the bioreactor. Triplicate measurements of the CDW were conducted at 6-7 timepoints per fermentation, as described in section 3.7.2. The average of each triplicate, as well as the (log_{10}) of the averages were found. Table 5.9 in Appendix 5 shows the Average CDW and log_{10} of the avg. CDW over time during the cultivation of *B. subtilis* 168 WT in glucose minimal medium, as an example. The log_{10} of the avg. CDWs have been plotted against time, and a linear regression line has been fitted to the data points. Figure 5.2 in Appendix 5 shows the log_{10} of avg. CDWs over time for WT in glucose minimal medium. The equation of the regression line (y = 0.357x - 0.986) can be used to predict the CDW (mg/mL) for a specific time point of the cultivation. By using the volume of minimal medium in the reactor in the given time point, we can find the total CDW (g) in the reactor for that time point. Predicted CDWs were used to determine the specific uptake- and secretion rates as close to the harvesting point as possible, and to compensate for potential errors made during CDW measurements.

The goal is to find the specific O_2 uptake rate ($q_{\mathrm{O_2}}$) and the specific CO_2 secretion rate ($q_{\mathrm{CO_2}}$) for each cultivation as close to the point of harvesting as possible. Therefore, the 10 output gas measurements conducted as close to the start of harvesting as possible were used to determine the $q_{\mathrm{O_2}}$ and $q_{\mathrm{CO_2}}$. For both of the cultivations of the mutant *B. subtilis* strain, it was possible to use the 10 last measurements before harvesting. For the wild type, however, it seems like the filter was clogged towards the end of the cultivation, resulting in no air flow out of the reactor during the last 18 minutes (WT glucose) and 13 minutes (WT in methanol) of the cultivation. For the cultivations of WT we therefore used the 10 measurements closest to harvesting, where the air is still flowing through the system. The q_{O_2} for one cultivation is calculated as the average of the q_{O_2}s found based on these 10 measurements. (Similarily, the q_{CO_2} is calculated as the average of the q_{CO_2}s based on the 10 measurements). Table 5.10 in Appendix 5 shows the 10 time points with $O_{2_{out}}$ measurements, flow rates and predicted CDWs used to calculate the avg. q_{O_2} of *B. subtilis* 168 WT cultivated in glucose minimal medium, as an example. The raw data from the gas composition measurements for each fermentation, as well as the specific gas uptake calculations are shown in Supplementary material ”S2 - Gas uptake rates.xlsx”.

3.11 Finding the uptake- and secretion rates of medium components

Triplicate medium samples were collected at 7-13 time points per cultivation, as described in section 3.7.2, in order to determine the specific uptake- and secretion rates of the carbon sources, L-glu, L-trp, and acetate during exponential growth phase. It is desired to find the specific substrate uptake rates as close to the time of harvesting as possible. Therefore the specific substrate uptake- and secretion rate is found as an average of the three calculations closest to the harvesting time. The raw data from the medium composition measurements for each cultivation, as well as the specific substrate uptake calculations are shown in Supplementary material ”S3 - Medium uptake rates.xlsx”.

3.12 Conducting flux balance analysis (FBA) and flux variability analysis (FVA)

The flow chart presented in Figure 3.1 visualizes the steps that were necessary to conduct FVA. The process illustrated in the flow chart was repeated for each fermentation. The newest GEM available for *B. subtilis*, iBsu1147, was used. In the following sections the steps of the flow chart will be explained in more detail.

3.12.1 Creating new synthesis reactions based on experimental data

The coefficients of the BOF were updated to represent the experimentally measured biomass composition. Furthermore, the coefficients of the protein synthesis reaction were updated to represent the experimentally measured relative amino acid distribution. The coefficients of the remaining synthesis reactions (RNA, DNA, lipid, lipoteichoic acid, and cell wall) were normalized. In addition, a synthesis reaction was made for cofactors and ions. The updated synthesis reactions and BOF were added to the model and the old ones were removed. The coefficients used for the new BOFs and protein synthesis reactions can be found in Table 5.6 and 4.3, respectively.

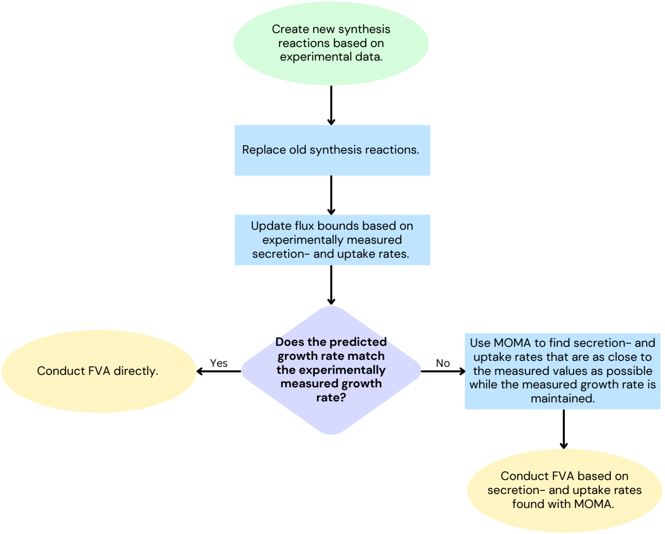

Create new synthesis reactions based on experimental data.
Replace old synthesis reactions.
Update flux bounds based on experimentally measured secretion- and uptake rates.
Does the predicted growth rate match the experimentally measured growth rate?
Yes
Conduct FVA directly.
No
Use MOMA to find secretion- and uptake rates that are as close to the measured values as possible while the measured growth rate is maintained.
Conduct FVA based on secretion- and uptake rates found with MOMA.

Figure 3.1: A flow chart representing the steps conducted before conduction FVA for each fermentation.

3.12.2 Updating flux bounds based on experimentally measured uptake- and secretion rates

Secretion and uptake rates were measured for carbon source, O_2, CO_2, acetate, L-glu, and L-trp. For the corresponding reactions in the GEM, the flux bounds were set to represent the experimentally measured secretion and uptake rates $\pm$ 2 standard deviations (SD). For the remaining compounds that were available in the nutrient environment, the flux bounds were set to be unrestricted (-1000.0, 1000.0). For compounds that were not part of the nutrient environment, only secretion was allowed (0.0, 1000.0).

3.12.3 Conducting flux variability analysis (FVA)

After the flux bounds were updated, FBA was used to predict the maximal feasible growth rate. If the predicted growth rate matched the experimentally measured growth rate, FVA was conducted directly. The flux bounds of the BOF were set to the experimentally measured growth rate $\pm$ 2SD, before FVA was conducted. However, if the predicted growth rate was lower than the experimentally measured growth rate, setting the flux bounds of the BOF to the experimentally measured growth rate would give an infeasible solution. In these cases, FVA could not be conducted directly. Instead, minimisation of metabolic adjustment (MOMA) was conducted first. MOMA finds an optimal flux distribution that is as close to a reference (here, the measured uptake- and secretion rates) as possible, while also maintaining the experimentally measured growth rate. Then the flux bounds of the GEM were updated based on the uptake- and secretion rates found with MOMA, and FVA was conducted.

3.13 Adding bacillithiol dependent oxidation of formaldehyde to the model

The assimilatory- and dissimilatory RuMP pathways are present in iBsu1147, however, the GEM lacks the BSH dependent pathway [6] for formaldehyde oxidation. This pathway was therefore added. MetaCyc was used to find what reactions needed to be added to the model. The BSH dependent oxidation of formaldehyde pathway is illustrated in Figure 3.2. The pathway consists of four reactions: first formaldehyde and bacillithiol (BSH) are converted to S-(hydroxymethyl)-bacillithiol. Then S-(hydroxymethyl)-bacillithiol is converted to S-formyl-bacillithiol. Then S-formylbacillithiol is converted to formate. Eventually formate is oxidized to CO_2. The ModelSeed reaction ID of each reaction is presented above the reaction arrows. The metabolites and reactions that are not already present in iBsu1147 are highlighted in orange. The three first reactions of the pathway, with IDs rxn42075, rxn44708 and rxn46864, and the missing metabolites (bacillithiol, s-(hydroxymethyl)bacillithiol and s-formylbacillithiol) were therefore added to the model. It was verified that flux is allowed through all of the newly added reactions, by iteratively changing the objective to each of the reactions and conducting FBA. The FBA solution was greater than zero for each reaction.

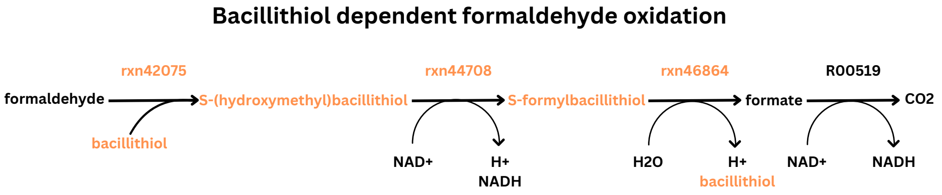

Bacillithiol dependent formaldehyde oxidation
rxn42075
rxn44708
rxn46864
R00519
formaldehyde
S-(hydroxymethyl)bacillithiol
S-formylbacillithiol
formate
CO2
bacillithiol
NAD+
H+
NADH
H2O
H+
bacillithiol
NAD+
NADH

Figure 3.2: Figure: Illustration of the bacillithiol (BSH) dependent formaldehyde oxidation pathway, which is present in *B. subtilis* [6]. The pathway consists of four reactions, and the reaction IDs (ModelSeed or KEGG) are presented above each arrow. Metabolites and reactions that were not already present in iBsu1147 are presented in orange.

Chapter 4

Results and discussion

In the following sections, the measured growth rate, biomass composition, amino acid distribution, and specific uptake- and secretion rates for each fermentation are presented and discussed. Then the the experimentally measured data is used to update the GEM of *B. subtilis* in order to analyze the flux distribution through the formaldehyde metabolism pathways of for each fermentation. Eventually, we look into if our results can bring us closer to an answer of why *B. subtilis* is unable to use methanol as the only carbon source.

4.1 Unraveling contradictions between the in silico and in vivo behaviour of *B. subtilis*

Before delving into the results that are promised above, it was found a discrepancy between the in silico behaviour of iBsu1147 and the in vivo behavior of *B. subtilis*, that we wanted to address. As previously explained, *B. subtilis* is not able to use methanol as the only carbon source, even when expressing Mdh [4]. However, the GEM of the organism is able to. We want to know what causes this contradiction between in vivo and in silico behaviour. The cryptFind algorithm [44] was used to find the cause of the discrepancy. In our case, cryptFind finds the genes that are essential for growth on methanol, but none of the other 54 tested carbon sources. In other words, knocking out the resulting genes would make the GEM unable to grow on methanol, but the GEM would still be able to grow on the remaining carbon sources [44]. Two genes were found with cryptFind: transaldolase and ribulose-phosphate 3-epimerase, suggesting that these two genes could be the reason for the contradictionary behaviour.

The cryptFind algorithm was originally built to find potential cryptic genes in *Yersinia pestis* [44]. However, in our case, the algorithm was used to find the cause of contradictory behaviour of the GEM compared to the organism in vivo. In other words, although transaldolase and ribulose-3-epimerase were found with the cryptFind algorithm, the genes are not necessarily cryptic, which is what the name of the algorithm implies. It is actually unlikely that transaldolase and ribulose-phosphate 3-epimerase are cryptic, since there is evidence that they both are expressed at the transcript and protein level in *B. subtilis* [47]. Both transaldolase and ribulose-phosphate epimerase are listed in F. Kunst et al.'s overview of the protein-coding genes of *B. subtilis*, with the gene names rpe and ywlF, respectively [47]. It could therefore be more likely that the function that are assigned to these genes in the model differs from the true catalytic activity they serve in the organism in vivo. This in turn gives the model the ability to grow with methanol as only carbon source, although this is not (yet) possible in vivo.

4.2 An overview of the batch fermentations that were conducted

We used bioreactors to conduct batch fermentations of both the wild type and mutant strain. Two batch fermentations were conducted for each strain, one where the strain was cultivated in glucose minimal medium, and one where it was cultivated in minimal medium containing methanol and glucose. In total, four batch fermentations were conducted. Each fermentation is assigned an abbreviation based on the strain and carbon source(s) that were used, which are shown in Table 4.1. These abbreviations will be used throughout the rest of the text, first to describe the fermentations, and later to describe the GEMs.

Table 4.1: Abbreviations that are assigned to each fermentation based on the *Bacillus subtilis* strain and carbon source(s) that were used.

Strain	**Carbon source(s)**	**Abbreviation**
B. subtilis 168 wild type	Glucose	WTG
B. subtilis 168 wild type	Methanol and glucose	WTM
B. subtilis 168 pBV2mp_mdhBm	Glucose	MDHG
B. subtilis 168 pBV2mp_mdhBm	Methanol and glucose	MDHM

During each fermentation, the growth rate and specific uptake- and secretion rates of nutrients and gas were measured. Furthermore, the bacteria of each fermentation were harvested at the end of exponential phase. The harvested bacteria were used to measure the biomass composition and amino acid distribution for each fermentation. The experimental data were used directly for phenotype analyses, and also to update the available GEM of *B. subtilis*. For each fermentation, the measured biomass composition and amino acid distribution was used to make a new BOF and protein synthesis reaction. The measured growth rate, and specific uptake- and secretion rates for each fermentation were used to adjust the constraints of the GEM. In total, four updated versions of the GEM iBsu1147 were made, one for each fermentation. The updated GEMs were used to conduct in silico phenotype analyses of each fermentation.

The fermentations and biomass analyses were performed in cooperation with another masters student, Sofie Tande-Petersen. Both of us had to perform 4 to 5 batch-fermentations for our individual projects, and we therefore decided to work together, as there is a significant amount of work in preparation for, and during batch-fermentations. The cooperation allowed for us to exchange ideas and help each other, while still independently planning and carrying out the experiments for our individual projects.

4.3 The mutant strain grows significantly faster with methanol as co-carbon source

The growth rate gives valuable information about the phenotype of each fermentation. Furthermore, growth rate is one of the constraints that are defined before employing constraint-based metabolic modeling on a GEM. The growth rate of each fermentation was therefore measured, to update this constraint. OD600 and CDW was measured throughout each fermentation. The OD600 measurements were mainly used to determine when the bacteria should be harvested. The bacteria were harvested at an OD600 of approximately 3, which is at the end of the exponential growth phase. The OD600 was also plotted against time to get a first idea of the growth rate of each fermentation. The OD600 measurements are presented in Figure 4.1. The growth rate was estimated by conducting linear regression on the CDW measurements on log scale against time. The CDW measurements on log scale are plotted against time in Figure 4.2A, and the estimated growth rates are presented in Figure 4.2B. The estimated growth rates with SD, as well as the R^2 of each regression line are presented in Table 5.1 in Appendix 5.

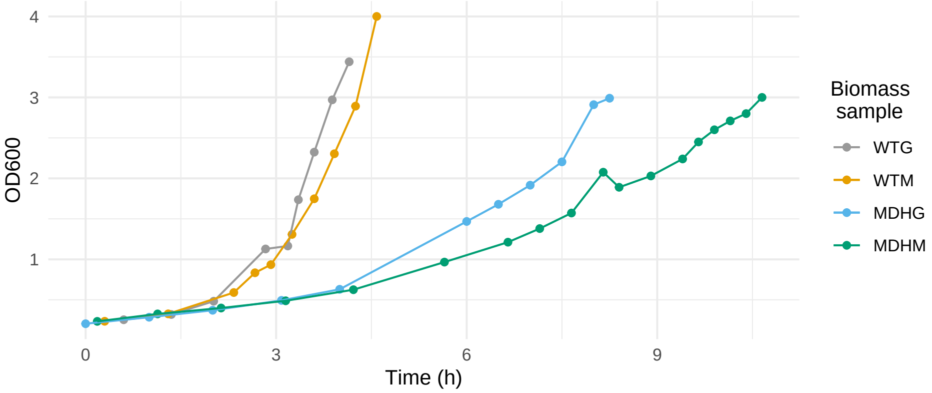


Figure 4.1: Optical density at 600 nm (OD600) over time (h) for each fermentation. The bacteria were harvested at an OD600 of approximately 3. (”Biomass sample” is used the same way as ”fermentation” is used in the rest of the text.)

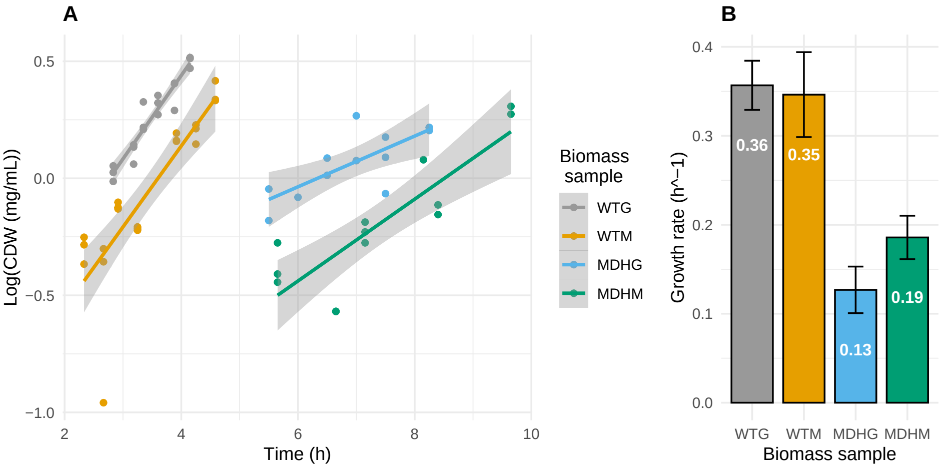


Figure 4.2: (A) Cell dry weight (CDW) on log scale plotted against time (h) for each fermentation. Linear regression was used to decide the growth rate (h^{-1}) of each fermentation, which are presented in part B. (B) Estimated growth rate (h^{-1}) for each Fermentation. ("Biomass sample" is used the same way as "fermentation" is used in the rest of the text.)

The OD600 measurements show that the mutant strain grows slower than the wild type strain. The same observation can be made by looking at the estimated growth rates. The growth rate of the WT strain is 0.36 h^{-1}± 0.03 when cultivated in glucose medium, and 0.35 h^{-1}± 0.05 when cultivated in methanol medium. For the mutant strain however, the growth rate is 0.13 h^{-1}± 0.03 when cultivated in glucose medium, and 0.19 h^{-1}± 0.02 when cultivated in methanol medium. ANOVA was used to conduct a pairwise comparison of the growth rates of the four fermentations, to confirm if the growth rates are significantly different from each other. According to the ANOVA analyses, the growth rates of WTG and WTM are not significantly different from each other, however, the WT grows significantly faster than the mutant. The mutant strain expresses a recombinant plasmid, which can impose an extra metabolic burden on the cell [48, 49]. This could be an explanation for the observed growth rate reduction of the mutant strain compared to that of the WT. Interestingly, the growth rate of MDHM is significantly higher than the growth rate of MDHG. The only difference between the two fermentations is that methanol is added to the medium for MDHM. This suggests that the mutant strain is indeed able to use the methanol for biomass production, which would confirm the findings of Gao et al.[4]. The ANOVA result for each pairwise comparison can be found in Table 5.2 in Appendix 5.

4.4 The measured biomass composition varies according to the strain and nutrient environment

The BOF of a GEM should reflect the biomass composition of the organism that the GEM represents [40]. Therefore, the biomass composition of each fermentation was measured. The biomass components that are present in the BOF of iBsu1147 are protein, RNA, DNA, lipid, lipoteichoic acid, cell wall, and cofactors and ions, which together form one unit of biomass. Out of these components, we experimentally measured the content of protein, RNA, DNA, and lipid. The contents of lipotheichoic acid, cell wall, and cofactors and ions were not measured. For these biomass components, the same coefficients as in iBsu1147 were used (0.030, 0.22 and 0.045 g/gCDW, respectively). The protein, RNA, DNA, and lipid content that was measured for each fermentation is presented in Figure 4.3. The same results are presented textually in Table 5.3 in Appendix 5.

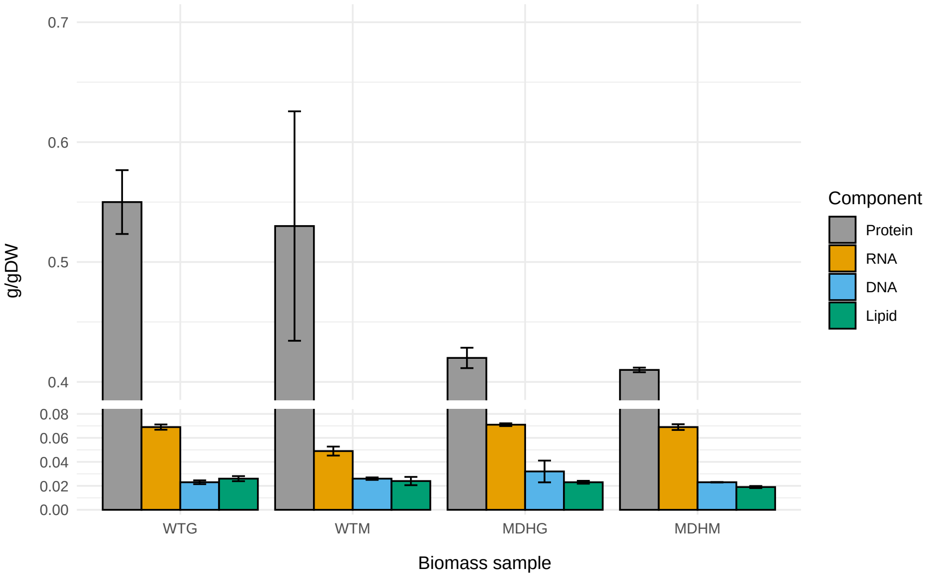


Figure 4.3: Biomass composition for each fermentation, shown as mass fraction (g/gDW) of each component. Only the fractions of the biomass components that were measured are visualized. (”Biomass sample” is used the same way as ”fermentation” is used in the rest of the text.)

One-way ANOVA was used to determine whether there are significant differences in biomass composition between the fermentations. Four ANOVA analyses were conducted, one for each measured biomass component. The resulting p-values from the ANOVA analyses are presented in Table 5.4 in Appendix 5. According to the ANOVA results, there are significant differences in the content of all four biomass components between at least two of the fermentations. The most apparent difference is that of the protein content, as we see that the mutant strain has a significantly lower protein content than the wild type strain. According to Dauner and Sauer, the protein content of *B. subtilis* strain RB50::pRF69 increases with growth rate [50]. Our results comply to this finding, as both the growth rate and protein content is significantly higher for the wild type than for the mutant.

It is worth noting that we have not been able to capture the same amount of the biomass for each fermentation. For the four fermentations, we have been able to measure 67, 63, 54 and 52 % of the biomass, respectively. Moreover, it would be interesting to see if there are significant differences between the relative content of each biomass component. For this, the values have been normalized so that the sum of protein, RNA, DNA, and lipid content for each fermentation equals one. When conducting the same ANOVA analysis on the normalized results, we find significant differneces in RNA and DNA content between at least two of the fermentations, while there is no significant difference in protein and lipid content. The p-values of the ANOVA-analyses on normalized values are presented in Table 5.5 in Appendix 5.

Since there are significant differences in the relative biomass composition between the fermentations, a new BOF was created for each fermentation. When creating a BOF, the sum of all the biomass components must be one. In that way the coefficients of the biomass components of the BOF reflect the relative contribution of each biomass component to biomass production [40]. That means that summarizing the protein, RNA, DNA, lipid, lipoteichoic acid, cell wall, and cofactors and ion content should give one unit of biomass. Therefore, the values were scaled to get a sum of one, for each fermentation. It should be emphasized that only the protein, RNA, DNA, and lipid contents were scaled, as the remaining values are derived from the BOF of iBsu1147, and are already scaled. The scaled values are used as coefficients in the new BOFs, and are presented in Figure 4.4. The figure also shows the coefficients of the BOF of iBsu1147 for comparison. The same results are presented textually in Table 5.6 in Appendix 5.

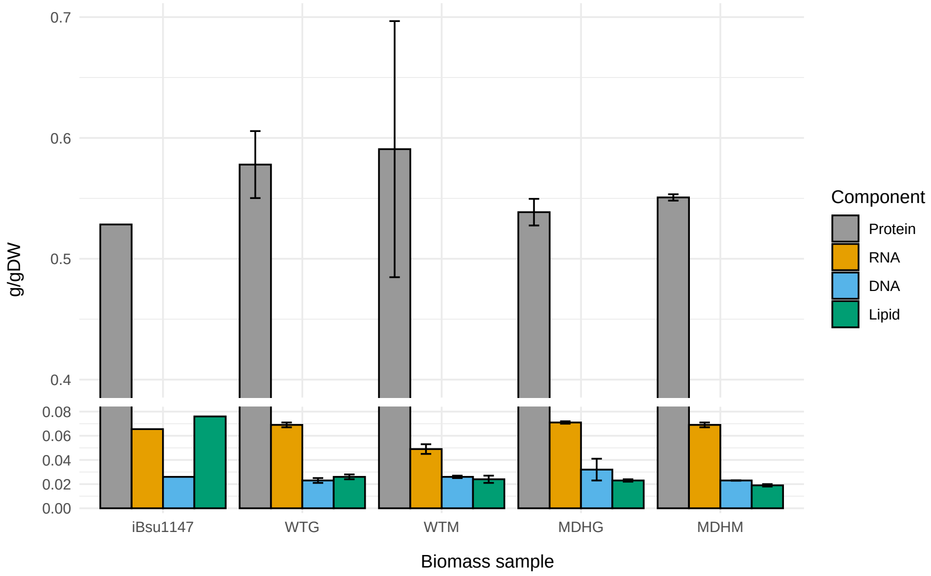


Figure 4.4: Scaled values of the protein, RNA, DNA, and lipid content (g/gCDW) that were used to create a new BOF representing each fermentation. The corresponding coefficients that is used for iBsu1147 are also presented. (”Biomass sample” is used the same way as ”fermentation” is used in the rest of the text.)

Previously, we have argued that the BOF should be updated, to represent the biomass composition of the strain in vivo. However, before discarding the BOF of iBsu1147, we should determine if the new BOF is actually different from the one already present in iBsu1147.One-sample t-test was used to determine if the new coefficients are significantly different from the old coefficients. The coefficients of the BOF of iBsu1147 were pairwise compared to the coefficients of the new BOFs. The p-values from the t-tests are presented in Table 5.7 in Appendix 5. There were not found significant differences in the protein coefficients for any of the pairwise comparisons, however, there were found significant differences in RNA, DNA, and lipid coefficients between all the new BOFs and the BOF of iBsu1147. This suggests that the new BOFs are indeed different from that of iBsu1147

The most noticable difference is the difference in lipid content between our BOFs and the one of iBsu1147. The lipid content is approximately 7 % in iBsu1147, while we measured a lipid content of 2-3 % for all of our fermentations. The decrease in lipid content could be explained by an insufficient lipid extraction. However, if we have been able to collect the whole lipid content, the result highlights the importance of having reproduced the biomass composition measurements, instead of relying on already existing data. Regardless of what explains the difference in lipid content, there were still measured significant differences in RNA and DNA content. Based on these results, it was determined that the new BOFs are indeed different from that of iBsu1147, which supports our decision of discarding the BOF of iBsu1147 and using the new BOFs instead.

4.5 The relative amino acid distribution varies according to strain and nutrient environment

We have seen that there are significant differences in the biomass composition between the fermentations, and when comparing the fermentations to iBsu1147. We also want to analyze potential differences in amino acid distribution. If there are significant differences in amino acid distribution between the fermentations, the coefficients of the protein synthesis reactions should be updated.

The content of all 20 amino acids were measured experimentally by protein hydrolysis followed by HPLC, except for the content of cyt, pro, trp, and met. The four mentioned amino acids are destroyed by the protein hydrolysis process [29], and the content of these amino acids was instead estimated based on the amino acid distribution derived from the protein coding gene sequences of *B. subtilis* 168 (UniProt Proteome ID UP000001570). The amino acid content for each fermentation are presented in Table 4.2. One-way ANOVA was conducted to estimate if the content of each amino acid was significantly different between the fermentations. The resulting p-values are presented in Table 4.2.

Table 4.2: Average amino acid content (g/gDW) for each fermentation. The total protein content (g/gDW) is calculated as the sum of the content of all 20 amino acids. One-way ANOVA was conducted to compare the content of each amino acid between fermentations, and the p-values are presented. The average content of cys, pro, trp, and met was predicted based on the amino acid distribution derived from the protein coding genes of *B. subtilis*. Only one number was predicted per fermentation, and one-way ANOVA was therefore not conducted for these four amino acids.

	Avg. amino acid content (g/gDW)				
Amino acid	**WTG**	**WTM**	**MDHG**	**MDHM**	**p-value**
Asp	0.028	0.025	0.019	0.018	6.8E-06
Glu	0.065	0.047	0.057	0.056	0.33
Asn	0.021	0.019	0.015	0.014	6.8E-06
His	0.009	0.008	0.006	0.005	8.2E-05
Ser	0.019	0.016	0.013	0.011	6.6E-06
Gln	0.034	0.025	0.030	0.029	0.33
Gly	0.032	0.022	0.018	0.017	1.1E-07
Arg	0.045	0.030	0.025	0.024	1.1E-07
Thr	0.030	0.039	0.019	0.021	0.15
Ala	0.046	0.042	0.043	0.043	0.030
Tyr	0.017	0.062	0.012	0.010	0.40
Val	0.033	0.031	0.023	0.022	5.2E-06
Phe	0.022	0.020	0.015	0.021	0.085
Ile	0.029	0.027	0.022	0.024	0.016
Leu	0.030	0.037	0.028	0.026	2.8E-05
Lys	0.044	0.040	0.032	0.030	1.9E-06
Cys	0.004	0.004	0.003	0.003	
Pro	0.018	0.017	0.014	0.014	
Trp	0.009	0.008	0.007	0.007	
Met	0.018	0.017	0.014	0.014	
Sum	0.55	0.53	0.42	0.41	

According to the ANOVA results, there are significant differences in the content of 11 amino acids between at least two fermentations. However, the differences might reflect the difference in total protein content rather than a difference in relative amino acid distribution. To address this, we compared the relative amino acid distribution between the fermentations as well. To find the relative amino acid distribution, the values are normalized so to get a sum of one unit protein. Then the same ANOVA analyses were conducted as before. The relative amino acid distribution for each fermentation, as well as the results from the ANOVA analyses are presented in Table 4.3. When comparing the relative amino acid distribution between fermentations, there were found significant differences in the content of six amino acids, instead of in 11, which was found before normalizing the values.

Table 4.3: Average amino acid content (g/gProtein) for each fermentation. The total protein content (g/gProtein) is calculated as the sum of the content of all 20 amino acids. One-way ANOVA was conducted to compare the content of each amino acid between fermentations, and the p-values are presented. The average content of cys, pro, trp, and met was predicted based on the amino acid distribution derived from the protein coding genes of *B. subtilis*. Only one number was predicted per fermentation, and one-way ANOVA was therefore not conducted for these four amino acids.

	Avg. amino acid content (g/gProtein)				
Amino acid	**WTG**	**WTM**	**MDHG**	**MDHM**	**p-value**
Asp	0.051	0.046	0.046	0.044	0.602
Glu	0.117	0.088	0.138	0.136	0.021
Asn	0.039	0.035	0.035	0.034	0.602
His	0.016	0.014	0.014	0.012	0.163
Ser	0.034	0.030	0.030	0.028	0.161
Gln	0.062	0.046	0.073	0.072	0.021
Gly	0.058	0.041	0.044	0.042	0.003
Arg	0.081	0.056	0.061	0.058	0.003
Thr	0.054	0.073	0.046	0.052	0.494
Ala	0.083	0.079	0.104	0.105	0.001
Tyr	0.030	0.115	0.029	0.023	0.396
Val	0.060	0.057	0.056	0.055	0.861
Phe	0.040	0.038	0.037	0.051	0.135
Ile	0.053	0.050	0.053	0.058	0.345
Leu	0.054	0.070	0.066	0.063	0.012
Lys	0.080	0.075	0.077	0.074	0.963
Cys	0.007	0.007	0.008	0.008	
Pro	0.033	0.031	0.033	0.033	
Trp	0.017	0.016	0.017	0.017	
Met	0.033	0.031	0.033	0.033	
Sum	**1.00**	**1.00**	**1.00**	**1.00**	

Although there were found fewer differences in amino acid distribution when the values were normalized, there were still differences in the relative amino acid distribution between fermentations for glu, gln, gly, arg, ala, and leu. Therefore, it was decided to construct protein synthesis reactions that are specific for each fermentation. The protein synthesis reaction summarizes the content of all 20 amino acids to produce one unit of protein. The relative amino acid distribution, which is presented in Table 4.3, was used to construct new protein synthesis reactions for each fermentation.

4.6 Both the wild type and the mutant strain seem to consume methanol

The specific uptake- and secretion rates of nutrients and gas of an organism can be expressed as constraints to the GEM, in order to conduct constraint-based analyses. Furthermore, the specific uptake- and secretion rates can give us valuable information about the behavior of the organism. The specific uptake rates of glucose, methanol, L-trp and L-glu was measured for each fermentation, as these are the carbon-containing medium compounds. Furthermore, the specific secretion rates of CO_2 and acetate were measured, as these are carbon containing compounds secreted by the organism. Measuring the specific uptake- and secretion rates of the carbon containing compounds, allows us to check if the carbon balance is maintained. In addition, the specific uptake rate of O_2 was measured. The specific uptake rates of glucose, methanol, L-glu, L-trp, and acetate and are shown in Figure 4.5. The specific uptake- and secretion rates of O_2 and CO_2 are presented in Figure 4.6. The results are also presented in Table 5.11 and Table 5.12 in Appendix 5. One-way ANOVA was used to determine if there are significant differences in the specific uptake- and secretion rates between fermentations. Furthermore, Tukey's HSD Test for multiple comparisons was used to do a pairwise comparison. The ANOVA results are presented in Table 5.11 and 5.12 in Appendix 5. The results from the Tukey's HSD Test are presented in Figure 5.3 and 5.4 in Appendix 5.

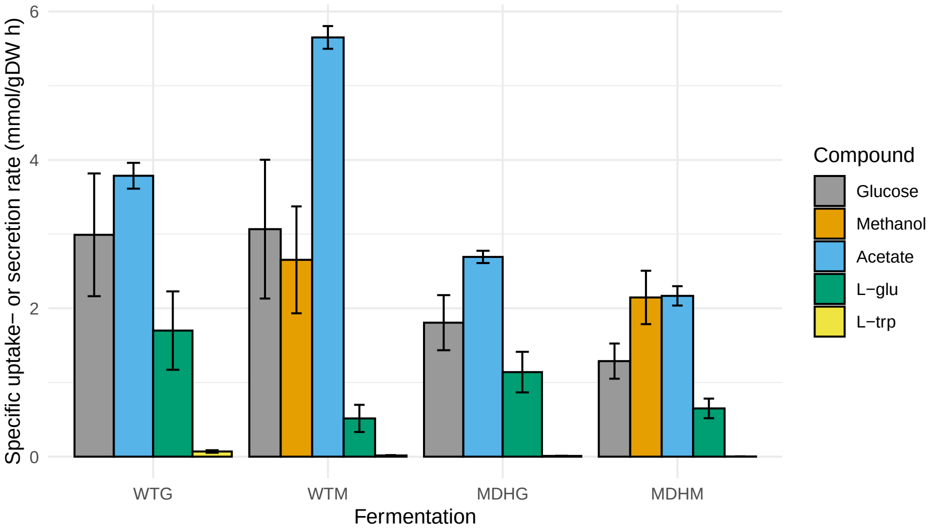


Figure 4.5: Specific substrate uptake- and secretion rates (mmol/gDW h) for each fermentation.

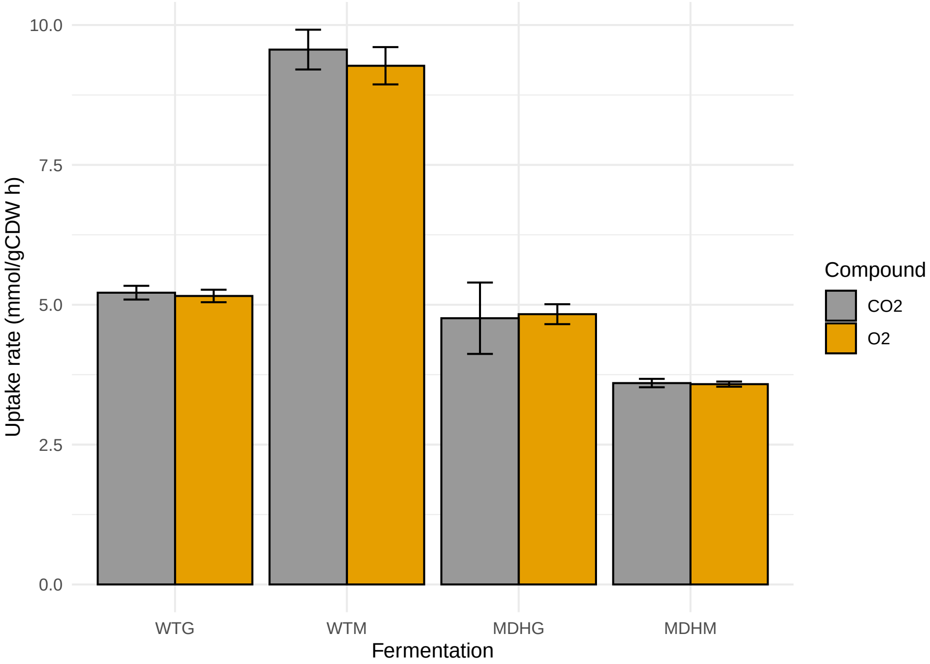


Figure 4.6: Specific uptake- and secretion rates (mmol/gDW h) of O_2 and CO_2 for each fermentation. Negative rates represent uptake rates, while positive rates represent secretion rates.

Previously, it was found that MDHM grows significantly faster than MDHG, and it was speculated that the specific uptake of methanol could be the reason. There was not found significant differences in the specific glucose, L-glu, or L-trp uptake rate between the two fermentations, however, MDHM consumes methanol in addition to glucose. This suggests that MDHM takes up more carbon than MDHG, which in turn could mean that the increase in growth rate is caused by the uptake of the additional carbon from methanol. Furthermore, MDHM does not seem to secrete more carbon than MDHG, as the acetate- and CO_2 secretion rates are not higher than that of MDHG. This suggests that MDHM was able to assimilate the methanol to produce biomass.

A surprising result, however, is that also WTM seem to consume methanol, and that there was not found a significant difference in the specific methanol uptake rate between WTM and MDHM. We expected that MDHM would consume methanol, as it heterologously expresses the Mdh gene. However, we did not expect to measure the same methanol uptake rate for WTM, because the wild type strain does not have the Mdh gene, and should therefore not be able to oxidize methanol to formaldehyde. The unexpected results could suggest that the wild type strain is able to break down methanol, for instance by using an unspecific or promiscuous alcohol dehydrogenase enzyme. Alternatively, the decrease in methanol concentration could be caused by something else than bacterial activity. The shape of the methanol concentration plotted against time can give an indication to what is happening. During the exponential growth phase, we would expect to see exponential decrease in the concentration of the nutrients that the bacteria use [24]. If the decrease in methanol concentration follows an exponential shape, this would suggest that bacterial growth causes the decrease. On the other hand, if the decrease is linear, the decrease might be happening independently of bacterial activity. To address this, we fitted a linear model and an exponential model to the data, to see which model best captures the variance of the data. For both WTM and MDHM, the R^2 values were too similar to determine if one model captures the data better than the other (0.78 and 0.77 for WTM, and 0.89 and 0.88 for MDHM). What should have been done instead is to produce a new control where the same conditions as for WTM and MDHM are used, but without bacteria, which would reveal if the decrease in methanol concentration is dependent on bacterial activity or not.

In addition to the unexpected methanol consuption observed for WTM, we see that WTM has a significantly higher acetate secretion rate than the other fermentations. Similarly, WTM has a higher respiration rate as both the specific O_2 uptake rate and CO_2 secretion rate are significantly higher than for the other fermentations. This could mean that the extra carbon from the methanol has been secreted as CO_2 and acetate, and has not been used for biomass production. That would explain why there is no significant difference in growth rate between WTG and WTM, although WTM seemingly takes up more carbon.

4.7 The predicted growth rate is robust to changes in the biomass- and protein synthesis reaction

Previously, we argued that the new BOFs and protein synthesis reactions are significantly different to those of iBsu1147, and that the new synthesis reactions should be used to conduct in silico phenotype-analyses. However, we have not yet tested if the changes in the BOF and protein synthesis reaction affect the predicted phenotype. To address this question, we compare the predicted growth rates of iBsu1147, WTG, WTM, MDHG, MDHM models, when the models have access to an arbitrary nutrient environment (glucose minimal medium was chosen). The nutrient environment that was used is presented in Table 5.13 in Appendix 5. The predicted growth rates are presented in Figure 4.7.

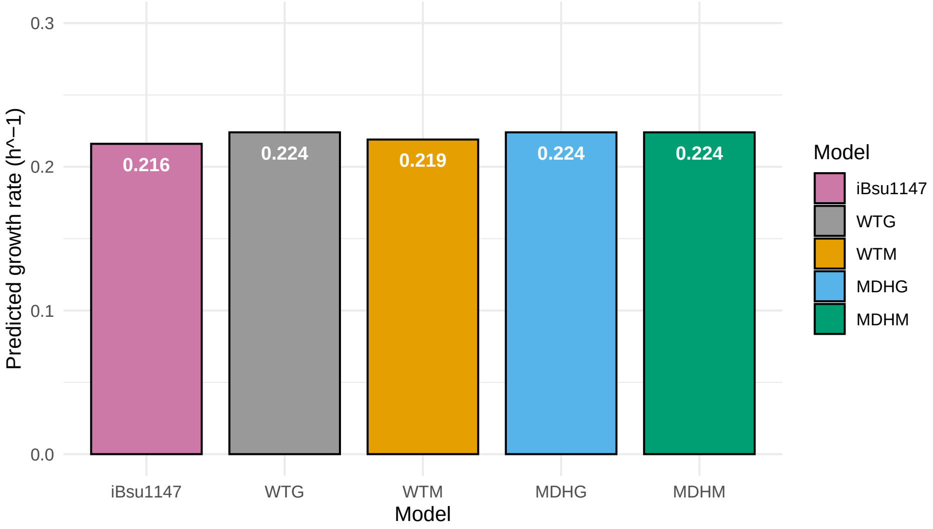


Figure 4.7: Predicted growth rate of each model: iBsu1147, WTG, WTM, MDHG and MDHM, when nutrient environment is kept constant.

The predicted growth rates are close to identical, despite all the five models having different coefficients for their BOF and protein synthesis reaction. This shows that the predicted growth rate is robust to variation in biomass and protein composition. This matches the results of Dinh et al. who found that the predicted growth rate is robust to changes in biomass composition [51].

Furthermore, we have checked if the experimentally measured growth rate of each fermentation is best captured by the BOF and protein synthesis reaction found in iBsu1147, or by the new BOF and protein synthesis reaction which reflect the experimentally measured biomass and protein composition. We compared the predicted growth rate that we get from using the old BOF and protein synthesis reaction, against what is predicted with the new BOF and protein synthesis reaction, for each fermentation. The constraints were updated to match the experimentally measured specific uptake- and secretion rate of each fermentation. The predicted growth rates as well as the experimentally measured growth rate of each fermentation is presented in Figure 4.8. The same results are presented in Table 5.14 in Appendix

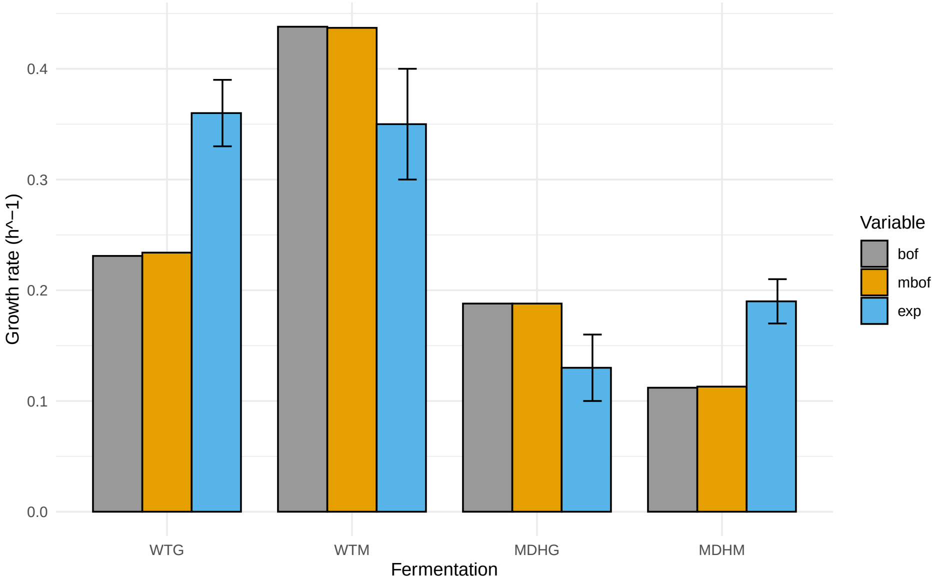


Figure 4.8: Experimentally measured and predicted growth rate (h^{-1}) for each fermentation. For the predicted growth rates, "BOF" means that the BOF and protein synthesis reaction of iBsu1147 were used, while "mBOF" means that the new BOF and protein synthesis reaction specific for the fermentation were used.

Again, we see that the variations in the BOF and protein synthesis reaction has little effect on the predicted growth rate, which complies with the results of Dinh et al. [51]. We could therefore ask ourselves if changing the BOF and protein synthesis reaction is necessary at all? However, it was found by Dinh et al. that although the growth rate predictions seem to be robust to variation in the BOF, the corresponding flux distribution that is predicted is more greatly affected by variation in the BOF. Unfortunately, due to time limitations, the effect of the biomass- and protein composition on flux distribution was not investigated. But, based on the results of Dinh et al., we would expected to see greater variation in the predicted flux distributions than what was observed for the predicted growth rates. Based on this assumption, we choose to conduct the further FVA analyses with updated BOFs and protein synthesis reactions.

The growth rate predictions are seemingly robust to changes in the BOF and protein synthesis reaction, however, the predictions are not necessarily accurate. For WTG and MDHM the predicted growth rate is lower than the measured growth rate. The inaccuracy of the growth rate predictions could be caused by uncertainty in the specific uptake- and secretion rate measurements. When chemostat is used, the bacteria are allowed to adjust to a constant nutrient environment, and the secretion- and uptake rates will stabilize over time [24]. In this work, the bacteria were cultivated in batch operation mode, where the nutrient environment changes throughout the fermentation. It is assumed that growth rate and uptake- and secretion rates stabilize during exponential growth of the batch fermentation [24]. However, using chemostat could likely have given more reliable results of the uptake- and secretion rates. MOMA was used for WTG and MDHM to find the specific uptake- and secretion rates that would allow the measured growth rate. In both cases the specific uptake- and secretion rates had to be increased. Another possible explanation to the predicted growth rates being lower than the measured growth rate could be that the non-growth associated maintenance (NGAM) requirements of the model are too strict. Removing the NGAM requirements for WTG and MDHM results in the predicted growth rates matching the experimentally measured growth rates. However, it is unreasonable to assume that the bacteria do not need any energy for non-growth associated processes. Furthermore, strict NGAM requirements would not explain why the predicted growth rate of WTM and MDHG are higher than the measured growth rates.

Furthermore, we see that the growth rate predicted for WTM is higher than what was measured experimentally. That could be because the GEM is able to use methanol as a carbon source, which increases the growth rate, while that might not be the case in vivo.

4.8 FVA results suggest that formaldehyde is assimilated into the RuMP pathway when methanol is available

B. subtilis has three pathways that can be used for formaldehyde detoxification, which are the assimilatory RuMP-pathway, the dissimilatory RuMP-pathway [10, 11], and the BSH-dependent oxidation of formaldehyde [6]. We want to see where there has to be flux in order to meet the biomass composition, and growth-, uptake-, and secretion rates that were measured for each fermentation. FVA was used to analyze the flux distribution of the formaldehyde detoxification pathways. FVA predicts the minimal and maximal flux that is possible for each reaction, while satisfying the constraints that have been defined. In other words, FVA gives the possible flux range of each reaction [38]. Since we want to know what reactions that must have flux through them, we look at the reactions that have a flux range that does not span zero. The FVA-results for WTG, WTM, MDHG, and MDHM are presented in Figure 4.9, 4.10, 4.11 and 4.12, respectively. The figures show a pathway map built by Vetle Simensen, representing the assimilatory- and dissimilatory RuMP pathway. MetaCyc was used to build the pathway map. The reactions that have a flux range that does not span 0 are highlighted. The FVA results for WTG, WTM, MDHG and MDHM are also attached in the excel file called ”FVA.xlsx”.

According to the FVA results for WTG, methanol is not converted to formaldehyde, and subsequently, formaldehyde is not incorporated into neither the assimilatory- nor the dissimilatory RuMP pathway. This meets our expectations, as no methanol was available during this fermentation. However, we do see that flux is suggested in other parts of the assimilatory RuMP pathway. Flux is suggested in the reactions where D-fructose-6-phosphate (F-6-P) and D-glyceraldehyde-3-phosphate (G-3-P) are used to generate D-ribose-5-phosphate (R-5-P). F-6-P and G-3-P are both intermediates of the glycolytic pathway, while R-5-P is an intermediate and product of the pentose phosphate pathway (PPP). The glycolytic pathway and PPP are both in use when glucose is used as the carbon source. The flux that is observed for WTG in these reactions could therefore likely be explained by activity in these pathways, and that these reactions are shared between the glucose metabolism pathways and the assimilatory RuMP pathway.

For MDHG, however, we do not see the same activity as in WTG, although the nutrient environment is the same. In MDHG, the only reaction where flux is suggested is in the reaction, where D-ribulose-5-phosphate (Ru-5-P) is oxidized to R-5-P. This reaction is shared between the PPP and RuMP pathway, and the suggested activity could yet again be explained by activity in the PPP.

For both WTM and MDHM, on the other hand, flux is suggested through reaction R05338 and R05339 which are the reactions where formaldehyde is incorporated into the assimilatory RuMP pathway by conversion to F-6-P. Furthermore, Ru-5-P is regenerated, which allows more formaldehyde to be incorporated into the assimilatory RuMP pathway. However, according to the FVA results, there does not have to be flux through the dissimilatory RuMP pathway. When we discussed the specific uptake- and secretion rates of WTM, we saw that WTM had a significantly higher secretion rate of acetate and CO_2. We speculated that this could suggest that the extra carbon from methanol was oxidized to CO_2. This does not seem to be captured by the FVA results. However, that does not necessarily mean that there is no flux through this pathway in vivo. But, because biomass production was set as the objective, the formaldehyde must be assimilated into the RuMP pathway in order to maximize biomass production so that the growth rate constraints are satisfied. It could be the case that some of the formaldehyde is dissimilated in vivo. Alternatively, the carbon from formaldehyde can be incorporated into the assimilatory RuMP pathway to produce products that are further fermented, which could explain the increase in CO_2 secretion rate. It should be noted that we for now only can speculate in how the methanol has been used by WTM, as the wild type strain should not be able to consume methanol. For MDHM, we observed an increase in growth rate compared to MDHG, while the secretion rate of acetate and CO_2 remained similar. This suggests that the carbon from methanol was used for biomass production, which seems to be captured by the FVA results as well.

The BSH-dependent oxidation pathway is not presented in the pathway map. This pathway was not originally present in the *B. subtilis* GEM, but was added in order to analyze the possible flux ranges of the reactions in this pathway as well. However, we encountered complications when we tried to conduct FVA after the pathway was added. We could not find an obvious reason for this, as flux is possible through the pathway, and the mass balance for each reaction was verified. Therefore, only the FVA results for the assimilatory- and dissimilatory RuMP pathway are presented. Although the flux through the BSH-dependent pathway was not analyzed, we would expect that the flux of formaldehyde remains through the assimilatory RuMP pathway, as this seems to be necessary to meet the growth rate constraints.

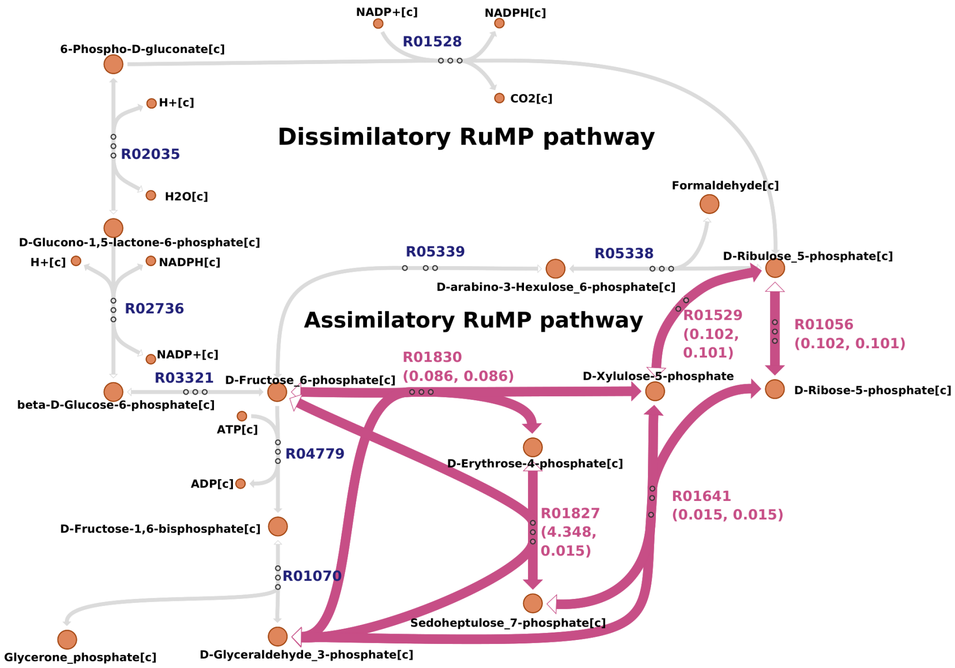

NADP+[c]
NADPH[c]
R01528
6-Phospho-D-gluconate[c]
H+[c]
CO2[c]
Dissimilatory RuMP pathway
R02035
H2O[c]
Formaldehyde[c]
D-Glucono-1,5-lactone-6-phosphate[c]
H+[c]
NADPH[c]
R05339
R05338
D-Ribulose_5-phosphate[c]
D-arabino-3-Hexulose_6-phosphate[c]
R02736
Assimilatory RuMP pathway
R01529
(0.102,
0.101)
R01056
(0.102, 0.101)
NADP+[c]
R01830
(0.086, 0.086)
R03321
D-Fructose_6-phosphate[c]
D-Xylulose-5-phosphate
D-Ribose-5-phosphate[c]
beta-D-Glucose-6-phosphate[c]
ATP[c]
R04779
D-Erythrose-4-phosphate[c]
ADP[c]
R01641
(0.015, 0.015)
R01827
(4.348,
0.015)
D-Fructose-1,6-bisphosphate[c]
R01070
Sedoheptulose_7-phosphate[c]
Glycerone_phosphate[c]
D-Glyceraldehyde_3-phosphate[c]

Figure 4.9: FVA results for WTG. The figure shows a graphical representation of the assimilatory- and dissimilatory RuMP pathway present in the GEM of *B. subtilis*. The nodes represent metabolites, and the edges represent the interactions between the metabolites. The name of each metabolite and the reaction ID of each reaction is presented. The reactions that have a flux range that does not span 0, are highlighted in pink. The reaction ID and flux range of these reactions are also highlighted. The remaining reactions have a flux range that spans 0. All the reactions that are presented are defined as reversible in the GEM, which is why the reaction arrows point both ways. The direction of the flux follows the direction of the filled arrows.

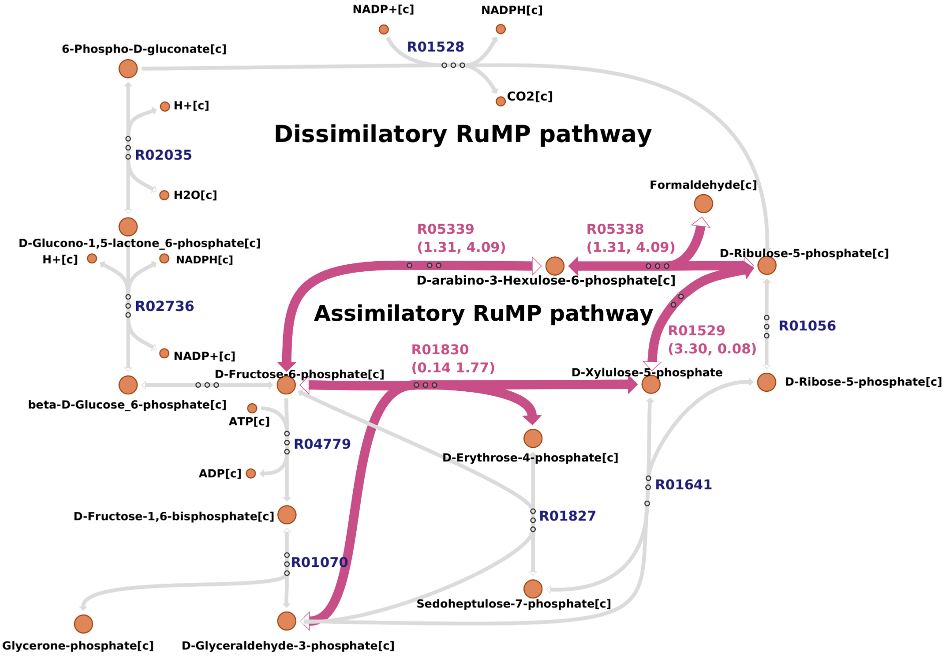

NADP+[c]
NADPH[c]
6-Phospho-D-gluconate[c]
R01528
CO2[c]
H+[c]
Dissimilatory RuMP pathway
R02035
H2O[c]
Formaldehyde[c]
R05339
(1.31, 4.09)
R05338
(1.31, 4.09)
D-Ribulose-5-phosphate[c]
D-Glucono-1,5-lactone_6-phosphate[c]
H+[c]
NADPH[c]
D-arabino-3-Hexulose-6-phosphate[c]
R02736
Assimilatory RuMP pathway
R01529
(3.30, 0.08)
R01056
NADP+[c]
R01830
(0.14 1.77)
D-Fructose-6-phosphate[c]
D-Xylulose-5-phosphate
D-Ribose-5-phosphate[c]
beta-D-Glucose_6-phosphate[c]
ATP[c]
R04779
D-Erythrose-4-phosphate[c]
ADP[c]
R01641
D-Fructose-1,6-bisphosphate[c]
R01827
R01070
Sedoheptulose-7-phosphate[c]
Glycerone-phosphate[c]
D-Glyceraldehyde-3-phosphate[c]

Figure 4.10: FVA results for WTM. The figure shows a graphical representation of the assimilatory- and dissimilatory RuMP pathway present in the GEM of *B. subtilis*. The nodes represent metabolites, and the edges represent the interactions between the metabolites. The name of each metabolite and the reaction ID of each reaction is presented. The reactions that have a flux range that does not span 0, are highlighted in pink. The reaction ID and flux range of these reactions are also highlighted. The remaining reactions have a flux range that spans 0. All the reactions that are presented are defined as reversible in the GEM, which is why the reaction arrows point both ways. The direction of the flux follows the direction of the filled arrows.

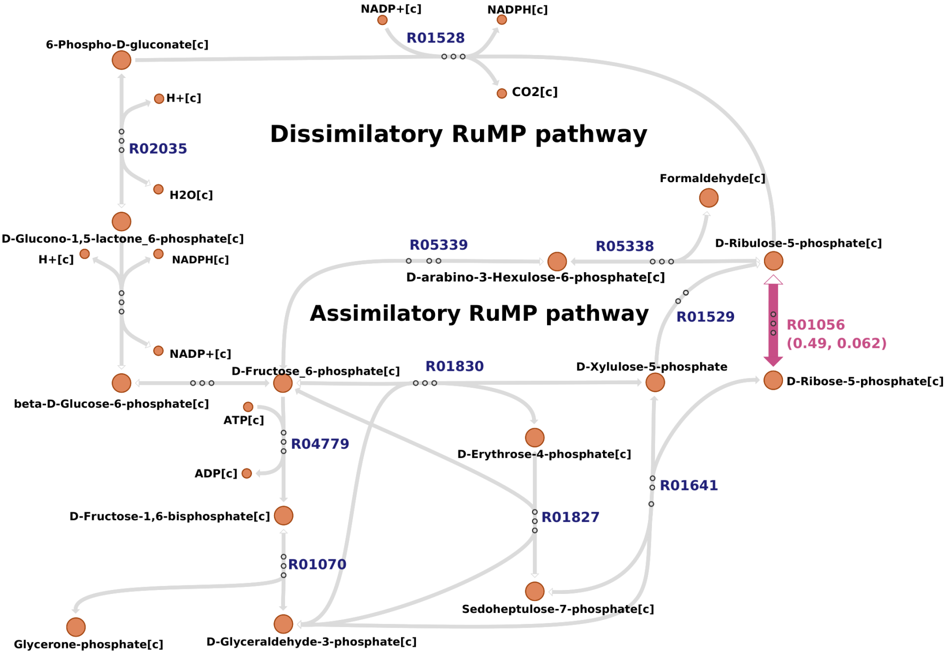

NADP+[c]
NADPH[c]
R01528
6-Phospho-D-gluconate[c]
CO2[c]
H+[c]
R02035
Dissimilatory RuMP pathway
H2O[c]
Formaldehyde[c]
D-Glucono-1,5-lactone_6-phosphate[c]
R05339
R05338
D-Ribulose-5-phosphate[c]
H+[c]
NADPH[c]
D-arabino-3-Hexulose-6-phosphate[c]
Assimilatory RuMP pathway
R01529
R01056
(0.49, 0.062)
NADP+[c]
D-Fructose_6-phosphate[c]
R01830
D-Xylulose-5-phosphate
D-Ribose-5-phosphate[c]
beta-D-Glucose-6-phosphate[c]
ATP[c]
R04779
D-Erythrose-4-phosphate[c]
ADP[c]
R01641
D-Fructose-1,6-bisphosphate[c]
R01827
R01070
Sedoheptulose-7-phosphate[c]
Glycerone-phosphate[c]
D-Glyceraldehyde-3-phosphate[c]

Figure 4.11: FVA results for MDHG. The figure shows a graphical representation of the assimilatory- and dissimilatory RuMP pathway present in the GEM of *B. subtilis*. The nodes represent metabolites, and the edges represent the interactions between the metabolites. The name of each metabolite and the reaction ID of each reaction is presented. The reactions that have a flux range that does not span 0, are highlighted in pink. The reaction ID and flux range of these reactions are also highlighted. The remaining reactions have a flux range that spans 0. All the reactions that are presented are defined as reversible in the GEM, which is why the reaction arrows point both ways. The direction of the flux follows the direction of the filled arrows.

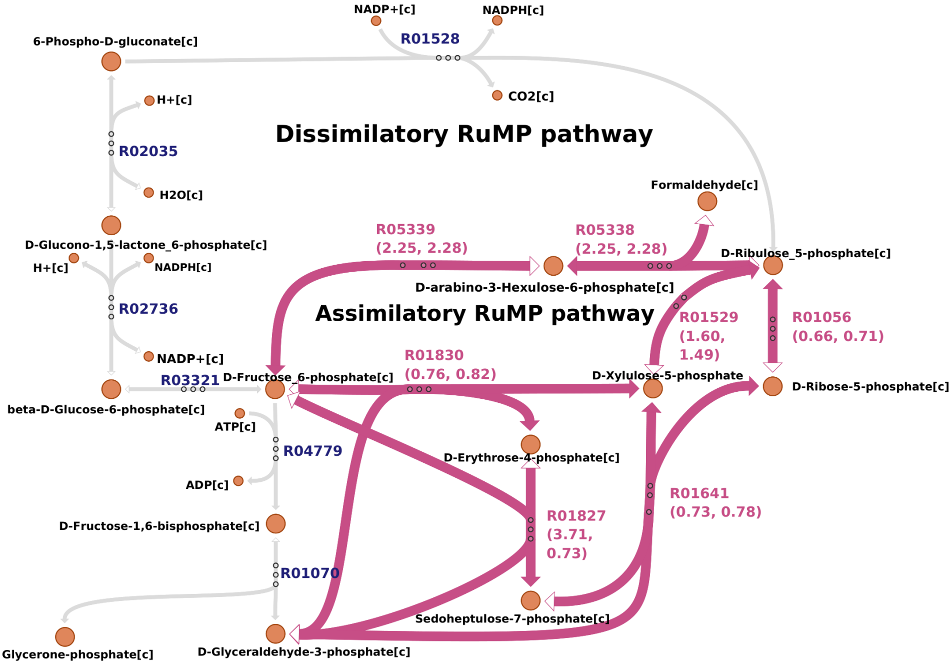

NADP+[c]
NADPH[c]
R01528
6-Phospho-D-gluconate[c]
H+[c]
CO2[c]
Dissimilatory RuMP pathway
R02035
H2O[c]
Formaldehyde[c]
R05339
(2.25, 2.28)
R05338
(2.25, 2.28)
D-Ribulose_5-phosphate[c]
D-Glucono-1,5-lactone_6-phosphate[c]
H+[c]
NADPH[c]
D-arabino-3-Hexulose-6-phosphate[c]
R02736
Assimilatory RuMP pathway
R01529
(1.60,
1.49)
R01056
(0.66, 0.71)
NADP+[c]
R01830
(0.76, 0.82)
R03321
D-Fructose_6-phosphate[c]
D-Xylulose-5-phosphate
D-Ribose-5-phosphate[c]
beta-D-Glucose-6-phosphate[c]
ATP[c]
R04779
D-Erythrose-4-phosphate[c]
ADP[c]
R01641
(0.73, 0.78)
R01827
(3.71,
0.73)
D-Fructose-1,6-bisphosphate[c]
R01070
Sedoheptulose-7-phosphate[c]
Glycerone-phosphate[c]
D-Glyceraldehyde-3-phosphate[c]

Figure 4.12: FVA results for MDHM. The figure shows a graphical representation of the assimilatory- and dissimilatory RuMP pathway present in the GEM of *B. subtilis*. The nodes represent metabolites, and the edges represent the interactions between the metabolites. The name of each metabolite and the reaction ID of each reaction is presented. The reactions that have a flux range that does not span 0, are highlighted in pink. The reaction ID and flux range of these reactions are also highlighted. The remaining reactions have a flux range that spans 0. All the reactions that are presented are defined as reversible in the GEM, which is why the reaction arrows point both ways. The direction of the flux follows the direction of the filled arrows.

Chapter 5

Conclusion and Outlook

So far it has been established that *B. subtilis* is able to use methanol as a carbon source when a co-carbon source such as glucose is provided [4]. These findings are supported by the results of our work. We have analyzed the phenotype of *B. subtilis* 168 pBV2mp_mdhBm, which is a mutant strain that expresses the Mdh gene heterologously. The mutant strain was cultivated in medium containing glucose, and in medium with both glucose and methanol available. The phenotype of the mutant strain was analyzed both experimentally by measuring specific metabolite uptake- and secretion rates, including the methanol uptake rate, and by measuring the growth rate. The results show that the mutant strain takes up methanol and that it grows significantly faster when methanol is available in addition to glucose. This strongly suggests that the methanol that is consumed is assimilated to produce biomass. Furthermore, no increase in CO_2- or acetate secretion rates was observed, which further supports that the carbon from methanol is used for biomass production rather than being secreted.

Furthermore, the biomass composition and amino acid distribution was measured and used to update the BOF and protein synthesis reaction of the GEM of *B. subtilis*. The measured uptake- and secretion rates, as well as the growth rate were defined as constraints. The updated GEM was used to analyze the in silico phenotype of the mutant strain in medium with glucose and methanol. FVA was used to analyze the flux distribution through the assimilatory and dissimilatory RuMP pathways, and the results are in accordance with what was found experimentally: the formaldehyde is incorporated into the assimilatory RuMP pathway and is used to produce biomass.

These results show that *B. subtilis* 168 is able to use methanol for biomass production, when the Mdh gene is heterologously expressed. However, it does not answer our question to why the strain cannot use methanol as the only carbon source. A possible reason could be that not enough Ru-5-P is regenerated to sustain the RuMP pathway, as Ru-5-P is regenerated both in the RuMP pathway itself, but also in the PPP. It is likely that the activity of the PPP is limited when there is no glucose to metabolise. Similarly, other intermediates or products generated by glycolysis, the TCA cycle, or PPP could be limited. If any of the limited compounds are essential to growth, that could explain why *B. subtilis* is unable to grow on methanol without an additional co-substrate to keep these pathways active. To address this question, further analyses of the phenotype of the mutant strain could be conducted. Examples of experimental measurements that could be done are metabolomics measurements, which would reveal if products from glycolysis, the TCA cycle, or PPP are limited when the strain grows with methanol compared to when it grows with only glucose as the carbon source. Furthermore, proteomics analyses would reveal if enzymes of the glucose metabolism pathways are underexpressed, compared to when glucose is used as the carbon source. In addition, further in silico analyses could be conducted to find reactions that are possible bottle-necks of the assimilatory RuMP pathway. For this it would be a good idea to use a GEM where enzyme constraints are implemented. In the GEM that we used, the flux of each reaction is constrained by flux bounds which do not reflect possible constraints imposed based on enzyme efficiency. If any of the reactions of the RuMP-pathway are catalyzed by enzymes of low efficiency, these reactions could be targets to try and increase the flux through the RuMP pathway. Such reactions could be revealed by using an enzyme-constrained metabolic model of *B. subtilis* [52]. Hopefully this would bring us closer to a synthetically methylotrophic *B. subtilis* strain which can use methanol as the only carbon source.

Although the synthetic methylotrophic strains of *B. subtilis* that have been generated so far cannot use methanol as the sole carbon source [4], using methanol as a co-carbon source will still ensure that less glucose is needed in production [53]. The next chapter would then be to start using the synthetic methylotrophic *B. subtilis* strain as a cell factory for production of products of interest.

Bibliography

[1] Yuan Su, Chuan Liu, Huan Fang, and Dawei Zhang. Bacillus subtilis: a universal cell factory for industry, agriculture, biomaterials and medicine. *Microbial Cell Factories*, 19(1):173, 12 2020.

[2] Roelco J. Kleijn, Joerg M. Buescher, Ludovic Le Chat, Matthieu Jules, Stephane Aymerich, and Uwe Sauer. Metabolic Fluxes during Strong Carbon Catabolite Repression by Malate in Bacillus subtilis. *Journal of Biological Chemistry*, 285(3):1587–1596, 1 2010.

[3] William B Whitaker, Nicholas R Sandoval, Robert K Bennett, Alan G Fast, and Eleftherios T Papoutsakis. Synthetic methylotrophy: engineering the production of biofuels and chemicals based on the biology of aerobic methanol utilization. *Current Opinion in Biotechnology*, 33:165–175, 6 2015.

[4] Bo Gao, Ning Zhao, Jieying Deng, Yang Gu, Shiru Jia, Ying Hou, Xueqin Lv, and Long Liu. Constructing a methanol-dependent Bacillus subtilis by engineering the methanol metabolism. *Journal of Biotechnology*, 343:128–137, 1 2022.

[5] Hai He, Christian Edlich-Muth, Steffen N. Lindner, and Arren Bar-Even. Ribulose Monophosphate Shunt Provides Nearly All Biomass and Energy Required for Growth of ¡i¿E. coli¡/i¿. *ACS Synthetic Biology*, 7(6):1601–1611, 6 2018.

[6] Gerald L Newton, Mamta Rawat, James J La Clair, Vishnu Karthik Jothivasan, Tanya Budiarto, Chris J Hamilton, Al Claiborne, John D Helmann, and Robert C Fahey. Bacillithiol is an antioxidant thiol produced in Bacilli. *Nature Chemical Biology*, 5(9):625–627, 9 2009.

[7] Maciek R Antoniewicz. Synthetic methylotrophy: Strategies to assimilate methanol for growth and chemicals production. *Current Opinion in Biotechnology*, 59:165–174, 10 2019.

[8] Lennart Schada von Borzyskowski, Mitja Remus-Emsermann, Ramon Weishaupt, Julia A. Vorholt, and Tobias J. Erb. A Set of Versatile Brick Vectors and Promoters for the Assembly, Expression, and Integration of Synthetic Operons in

¡i¿Methylobacterium extorquens¡/i¿ AM1 and Other Alphaproteobacteria. *ACS Synthetic Biology*, 4(4):430–443, 4 2015.

[9] Hiroya Yurimoto, Nobuo Kato, and Yasuyoshi Sakai. Assimilation, dissimilation, and detoxification of formaldehyde, a central metabolic intermediate of methylotrophic metabolism. *The Chemical Record*, 5(6):367–375, 2005.

[10] Hisashi Yasueda, Yoshio Kawahara, and Shin-ichi Sugimoto. ¡i¿Bacillus subtilis yckG¡/i¿ and ¡i¿yckF¡/i¿ Encode Two Key Enzymes of the Ribulose Monophosphate Pathway Used by Methylotrophs, and ¡i¿yckH¡/i¿ Is Required for Their Expression. *Journal of Bacteriology*, 181(23):7154–7160, 12 1999.

[11] Ryoji Mitsui, Yoko Kusano, Hiroya Yurimoto, Yasuyoshi Sakai, Nobuo Kato, and Mitsuo Tanaka. Formaldehyde Fixation Contributes to Detoxification for Growth of a Nonmethylotroph, ¡i¿Burkholderia cepacia¡/i¿ TM1, on Vanillic Acid. *Applied and Environmental Microbiology*, 69(10):6128–6132, 10 2003.

[12] Bernhard O. Palsson. *Systems Biology: Constraint-based Reconstruction and Analysis*. Cambridge University Press, 2nd ed. edition, 2015.

[13] You-Kwan Oh, Bernhard O. Palsson, Sung M. Park, Christophe H. Schilling, and Radhakrishnan Mahadevan. Genome-scale Reconstruction of Metabolic Network in Bacillus subtilis Based on High-throughput Phenotyping and Gene Essentiality Data. *Journal of Biological Chemistry*, 282(39):28791–28799, 9 2007.

[14] Oliver Schilling, Oliver Frick, Christina Herzberg, Armin Ehrenreich, Elmar Heinzle, Christoph Wittmann, and Jorg Stulke. 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. *Applied and Environmental Microbiology*, 73(2):499–507, 1 2007.

[15] Julia Vorholt. Cofactor-dependent pathways of formaldehyde oxidation in methylotrophic bacteria. *Archives of Microbiology*, 178(4):239–249, 10 2002.

[16] Jonas E. N. Müller, Tonje M. B. Heggeset, Volker F. Wendisch, Julia A. Vorholt, and Trygve Brautaset. Methylotrophy in the thermophilic Bacillus methanolicus, basic insights and application for commodity production from methanol. *Applied Microbiology and Biotechnology*, 99(2):535–551, 1 2015.

[17] Anh Duc Nguyen, In Yeub Hwang, Jeon Young Chan, and Eun Yeol Lee. Reconstruction of methanol and formate metabolic pathway in non-native host for biosynthesis of chemicals and biofuels. *Biotechnology and Bioprocess Engineering*, 21(4):477–482, 8 2016.

[18] Chang-Ting Chen, Frederic Y.-H. Chen, Igor W. Bogorad, Tung-Yun Wu, Ruoxi Zhang, Abraxa S. Lee, and James C. Liao. Synthetic methanol auxotrophy of Escherichia coli for methanol-dependent growth and production. *Metabolic Engineering*, 49:257–266, 9 2018.

[19] Fabian Meyer, Philipp Keller, Johannes Hartl, Olivier G. Gröninger, Patrick Kiefer, and Julia A. Vorholt. Methanol-essential growth of Escherichia coli. *Nature Communications*, 9(1):1508, 4 2018.

[20] Philibert Tuyishime, Yu Wang, Liwen Fan, Qiongqiong Zhang, Qinggang Li, Ping Zheng, Jibin Sun, and Yanhe Ma. Engineering Corynebacterium glutamicum for methanol-dependent growth and glutamate production. *Metabolic Engineering*, 49:220–231, 9 2018.

[21] Barry G. Hall, Shozo Yokoyama, and David H. Calhoun. Role of cryptic genes in microbial evolution. *Molecular Biology and Evolution*, 2 1984.

[22] Mitali Mukerji and S. Mahadevan. Cryptic genes: Evolutionary puzzles. *Journal of Genetics*, 76(2):147–159, 8 1997.

[23] Michele Rigon Spier, Luciana Porto de Souza Vandenberghe, Adriane Bianchi Pedroni Medeiros, and Carlos Ricardo Soccol. APPLICATION OF DIFFERENT TYPES OF BIOREACTORS IN BIOPROCESSES, 2011.

[24] Ali Cinar, Satish J. Parulekar, C. Ündey, and G. Birol. *Batch Fermentation: Modeling, Monitoring, and Control.* Marcel Dekker, Inc., New York.

[25] John Villadsen, Jens Nielsen, and Gunnar Lidén. *Bioreaction Engineering Principles.* Springer US, Boston, MA, 2011.

[26] Pauline Doran. *Bioprocess Engineering Principles.* Oxford: Elsevier, 2nd edition, 2012.

[27] P. F. Stanbury, A. Whitaker, and S. J. Hall. *Principles of Fermentation Technology.* Elsevier Science Ltd., 2nd edition, 1984.

[28] J. E. Noble, A. E. Knight, A. J. Reason, A. Di Matola, and M. J. A. Bailey. A Comparison of Protein Quantitation Assays for Biopharmaceutical Applications. *Molecular Biotechnology*, 37(2):99–111, 9 2007.

[29] Alison J. Darragh, Dorian J. Garrick, Paul J. Moughan, and Wouter H. Hendriks. Correction for Amino Acid Loss during Acid Hydrolysis of a Purified Protein. *Analytical Biochemistry*, 236(2):199–207, 5 1996.

[30] Jordi. Folch, I. Ascoli, M. Lees, J.A. Meath, and F.N. LeBaron. PREPARATION OF LIPIDE EXTRACTS FROM BRAIN TISSUE. *Journal of Biological Chemistry*, 191(2):833–841, 8 1951.

[31] Warren M. Sperry and Florence C. Brand. THE DETERMINATION OF TOTAL LIPIDES IN BLOOD SERUM. *Journal of Biological Chemistry*, 213(1):69–76, 3 1955.

[32] Elizabeth van Pelt-Verkuil, Alex van Belkum, and John P. Hays. The Different Types and Varieties of Nucleic Acid Target Molecules. In *Principles and Technical Aspects of PCR Amplification*, pages 25–61. Springer Netherlands, Dordrecht, 2008.

[33] S. Benthin, J. Nielsen, and J. Villadsen. A simple and reliable method for the determination of cellular RNA content. *Biotechnology Techniques*, 5(1):39–42, 1991.

[34] Ozlem Coskun. Separation Tecniques: CHROMATOGRAPHY. *Northern Clinics of Istanbul*, 2016.

[35] Joseph B. Lambert and Eugene P. Mazzola. *Nuclear Magnetic Resonance Spectroscopy: An Introduction to Principles, Applications, and Experimental Methods*. Prentice Hall, 2003.

[36] K. Miyanaga and H. Unno. Reaction Kinetics and Stoichiometry. In *Comprehensive Biotechnology*, pages 33–46. Elsevier, 2011.

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

[38] Steinn Gudmundsson and Ines Thiele. Computationally efficient flux variability analysis. *BMC Bioinformatics*, 11(1):489, 12 2010.

[39] Daniel Segrè, Dennis Vitkup, and George M. Church. Analysis of optimality in natural and perturbed metabolic networks. *Proceedings of the National Academy of Sciences*, 99(23):15112–15117, 11 2002.

[40] Adam M Feist and Bernhard O Palsson. The biomass objective function. *Current Opinion in Microbiology*, 13(3):344–349, 6 2010.

[41] Ines Thiele and Bernhard Ø Palsson. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nature Protocols*, 5(1):93–121, 1 2010.

[42] Christopher S Henry, Jenifer F Zinner, Matthew P Cohoon, and Rick L Stevens. iBsu1103: a new genome-scale metabolic model of Bacillus subtilis based on SEED annotations. *Genome Biology*, 10(6):R69, 2009.

[43] Tong Hao, Binbin Han, Hongwu Ma, Jing Fu, Hui Wang, Zhiwen Wang, Bincai Tang, Tao Chen, and Xueming Zhao. In silico metabolic engineering of Bacillus subtilis for improved production of riboflavin, Egl-237, (R,R)-2,3-butanediol and isobutanol. *Molecular BioSystems*, 9(8):2034, 2013.

[44] Ali Navid and Eivind Almaas. Genome-scale reconstruction of the metabolic network in Yersinia pestis, strain 91001. *Molecular BioSystems*, 5(4):368, 2009.

[45] Marta Irla, Tonje M. B. Heggeset, Ingemar Nærdal, Lidia Paul, Tone Haugen, Simone B. Le, Trygve Brautaset, and Volker F. Wendisch. Genome-Based Genetic Tool Development for Bacillus methanolicus: Theta- and Rolling Circle-Replicating Plasmids for Inducible Gene Expression and Application to Methanol-Based Cadaverine Production. *Frontiers in Microbiology*, 7, 9 2016.

[46] B. J. Eikmanns, N. Thum-Schmitz, L. Eggeling, K.-U. Ludtke, and H. Sahm. Nucleotide sequence, expression and transcriptional analysis of the Corynebacterium glutamicum gltA gene encoding citrate synthase. *Microbiology*, 140(8):1817–1828, 8 1994.

[47] F. Kunst, N. Ogasawara, I. Moszer, A. M. Albertini, G. Alloni, V. Azevedo, M. G. Bertero, P. Bessières, A. Bolotin, S. Borchert, R. Borriss, L. Boursier, A. Brans, M. Braun, S. C. Brignell, S. Bron, S. Brouillet, C. V. Bruschi, B. Caldwell, V. Capuano, N. M. Carter, S.-K. Choi, J.-J. Codani, I. F. Connerton, N. J. Cummings, R. A. Daniel, F. Denizot, K. M. Devine, A. Düsterhöft, S. D. Ehrlich, P. T. Emmerson, K. D. Entian, J. Errington, C. Fabret, E. Ferrari, D. Foulger, C. Fritz, M. Fujita, Y. Fujita, S. Fuma, A. Galizzi, N. Galleron, S.-Y. Ghim, P. Glaser, A. Goffeau, E. J. Golightly, G. Grandi, G. Guiseppi, B. J. Guy, K. Haga, J. Haiech, C. R. Harwood, A. Hénaut, H. Hilbert, S. Holsappel, S. Hosono, M.-F. Hullo, M. Itaya, L. Jones, B. Joris, D. Karamata, Y. Kasahara, M. Klaerr-Blanchard, C. Klein, Y. Kobayashi, P. Koetter, G. Koningstein, S. Krogh, M. Kumano, K. Kurita, A. Lapidus, S. Lardinois, J. Lauber, V. Lazarevic, S.-M. Lee, A. Levine, H. Liu, S. Masuda, C. Mauël, C. Médigue, N. Medina, R. P. Mellado, M. Mizuno, D. Moestl, S. Nakai, M. Noback, D. Noone, M. O'Reilly, K. Ogawa, A. Ogiwara, B. Oudega, S.-H. Park, V. Parro, T. M. Pohl, D. Portetelle, S. Porwollik, A. M. Prescott, E. Presecan, P. Pujic, B. Purnelle, G. Rapoport, M. Rey, S. Reynolds, M. Rieger, C. Rivolta, E. Rocha, B. Roche, M. Rose, Y. Sadaie, T. Sato, E. Scanlan, S. Schleich, R. Schroeter, F. Scoffone, J. Sekiguchi, A. Sekowska, S. J. Seror, P. Serror, B.-S. Shin, B. Soldo, A. Sorokin, E. Tacconi, T. Takagi, H. Takahashi, K. Takemaru, M. Takeuchi, A. Tamakoshi, T. Tanaka, P. Terpstra, A. Tognoni, V. Tosato, S. Uchiyama, M. Vandenbol, F. Vannier, A. Vassarotti, A. Viari, R. Wambutt, E. Wedler, H. Wedler, T. Weitzenegger, P. Winters, A. Wipat, H. Yamamoto, K. Yamane, K. Yasumoto, K. Yata, K. Yoshida, H.-F. Yoshikawa, E. Zumstein, H. Yoshikawa, and A. Danchin. The complete genome sequence of the Gram-positive bacterium Bacillus subtilis. *Nature*, 390(6657):249–256, 11 1997.

[48] Dave Siak-Wei Ow, Peter Morin Nissom, Robin Philp, Steve Kah-Weng Oh, and Miranda Gek-Sim Yap. Global transcriptional analysis of metabolic burden due to plasmid maintenance in Escherichia coli DH5α during batch fermentation. *Enzyme and Microbial Technology*, 39(3):391–398, 7 2006.

[49] Zhijun Wang, Li Xiang, Junjie Shao, Alicja Węgrzyn, and Grzegorz Węgrzyn. Effects of the presence of ColE1 plasmid DNA in Escherichia coli on the host cell metabolism. *Microbial Cell Factories*, 5(1):34, 12 2006.

[50] Michael Dauner and Uwe Sauer. Stoichiometric growth model for riboflavin-producingBacillus subtilis. *Biotechnology and Bioengineering*, 76(2):132–143, 9 2001.

[51] Hoang V. Dinh, Debolina Sarkar, and Costas D. Maranas. Quantifying the propagation of parametric uncertainty on flux balance analysis. *Metabolic Engineering*, 69:26–39, 1 2022.

[52] Pavlos Stephanos Bekiaris and Steffen Klamt. Automatic construction of metabolic models with enzyme constraints. *BMC Bioinformatics*, 21(1):19, 12 2020.

[53] Wenming Zhang, Ting Zhang, Sihua Wu, Mingke Wu, Fengxue Xin, Weiliang Dong, Jiangfeng Ma, Min Zhang, and Min Jiang. Guidance for engineering of synthetic

methylotrophy based on methanol metabolism in methylotrophy. *RSC Advances*, 7(7):4083–4091, 2017.

Appendix - Supplementary data

5.1 Growth rate measurements

The growth rate was estimated for wild type and mutant cultivated in glucose and methanol minimal medium, by conducting linear regression on CDW measurements on log scale plotted against time. The estimated growth rates are presented in Table 5.1.

Table 5.1: Growth rate (h^{-1}) for each biomass sample. The growth rate corresponds to the slope of the regression line found by conducting linear regression of cell dry weight (CDW) on log scale plotted against time (h). Standard deviation (SD), as well as the R^2 value of each regression line are provided.

Biomass sample	**Growth rate**	**SD**	**R^2**
WTG	0.36	0.03	0.91
WTM	0.35	0.05	0.73
MDHG	0.13	0.03	0.59
MDHM	0.19	0.02	0.78

ANOVA was used to conduct a pairwise comparison of growth rate between the four fermentations, and the results are presented in Table 5.2.

Table 5.2: ANOVA results from a pairwise comparison of growth rate between the four fermentations. The critical value is calculated based on a significance level of 0.95, numerator degrees of freedom (df) = 1 and denominator df = 4. When the F-value is greater than the critical value, there is a significant difference in growth rate between the two fermentations.

Compared fermentations		**Critical value**	**F-value**
WTG	WTM	7.708647	0.108
MDHG	MDHM	7.708647	8.107
WTG	MDHG	7.708647	109.44
WTM	MDHG	7.708647	48.703
WTG	MDHM	7.708647	64.421
WTM	MDHM	7.708647	26.857

5.2 Biomass composition

The biomass composition was measured for each fermentation, and the results are presented in Table 5.3.

Table 5.3: The biomass composition for each biomass sample shown as mass fractions (g/gDW) of each biomass component. The amount of protein, RNA, DNA and lipid was measured experimentally, and is shown with standard deviation. The amount of lipoteichoic acid, cell wall, and cofactors & ions are assumed to be the same as for iBsu1147. Furthermore, the sum of the mass fractions is given at the bottom of the table.

	Cellular Content (g/gDW)			
	WT Glucose	**WT Methanol**	**MDH Glucose**	**MDH Methanol**
Protein	0.55 ± 0.027	0.53 ± 0.096	0.42 ± 0.008	0.41 ± 0.002
RNA	0.069 ± 0.002	0.049 ± 0.004	0.071 ± 0.001	069 ± 0.002
DNA	0.023 ± 0.002	0.026 ± 0.001	0.032 ± 0.009	0.023 ± 0.0002
Lipid	0.026 ± 0.002	0.024 ± 0.003	0.023 ± 0.001	0.019 ± 0.001
Lipoteichoic acid	0.030	0.030	0.030	0.030
Cell wall components	0.22	0.22	0.22	0.22
Cofactors & ions	0.045	0.045	0.045	0.045
Sum	0.97	0.93	0.84	0.82

One-way analysis of variance (ANOVA) was used to decide if there are significant differences in biomass composition between the four biomass samples. Four ANOVA analyses were conducted, one for each of the biomass components that were measured: protein, RNA, DNA and lipid. The p-values from the ANOVA analyses are presented in Table 5.4 Tukey's HSD Test for multiple comparisons was used to find what groups that are significantly different from each other. The Tukey's HSD test results for protein, RNA, DNA and lipid content are presented in Figure 5.1 A, B, C and D, respectively.

Table 5.4: p-values from one-way ANOVA analysis comparing the protein, RNA, DNA an lipid content between fermentations.

Biomass component	**p-value**
Protein	0.0119
RNA	<2e-16
DNA	0.000385
Lipid	0.0289

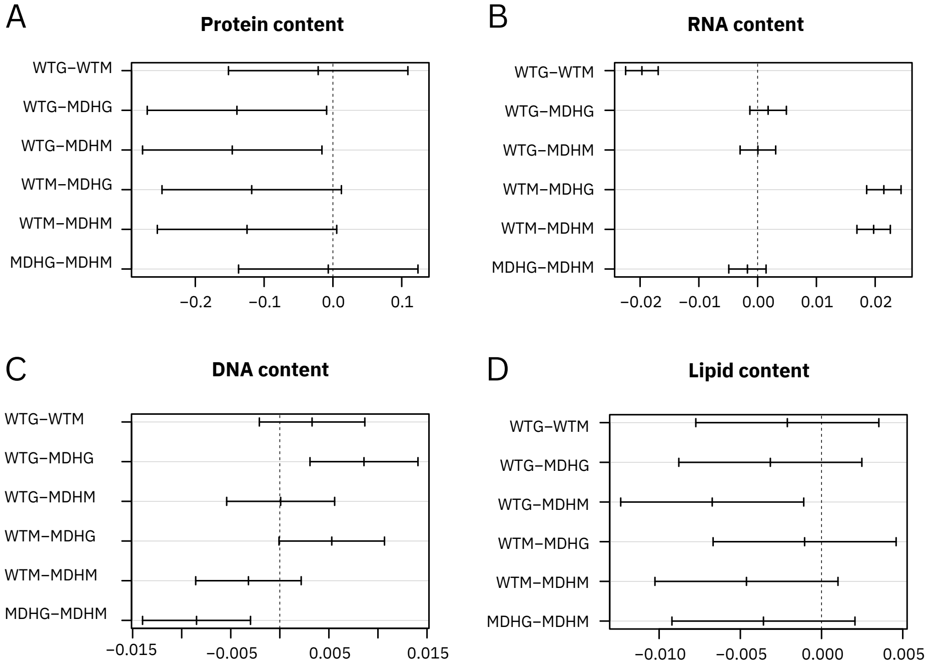


Figure 5.1: Tukey's HSD test for multiple comparisons was conducted to decide if there are significant differences in protein, RNA, DNA and lipid content between biomass samples. The figure shows the Tukey's HSD test results for A) protein content, B) RNA content, C) DNA content and D) lipid content. The results are presented as 95 % family-wise confidence intervals. Confidence intervals that do not span 0.0 represent significant difference. For instance there is a significant difference in protein content between WTG and MDHG, as the confidence interval does not span 0.0.

The measured content of protein, RNA, DNA, and lipid was scaled to get a sum of one for each fermentation. One-way ANOVA was used to compare the relative content of the biomass components, and the results are presented in Table 5.5.

Table 5.5: p-values from one-way ANOVA analysis comparing the scaled values of protein, RNA, DNA, and lipid content between fermentations. The values have been scaled in order to get a sum of one.

Biomass component	p-value
Protein	0.65
RNA	<2e-16
DNA	1.86e-06
Lipid	0.325

The measured biomass composition was used to create a new BOF. This was done for each of the four fermentations. First, the contents of the biomass components was scaled to give a sum of one unit of biomass. The scaled values are presented in table 5.6. The scaled values (without SD) are used as coefficients in the new BOFs. The coefficients used in iBsu1147 are also presented in Table 5.6, for comparison.

Table 5.6: The biomass composition for each fermentation after scaling the data to get a sum of one. The content of protein, RNA, DNA, and lipid were measured experimentally and are shown with SD. The contents are presented in g/gDW.

	Cellular Content (g/gDW)				
	iBsu1147	**WTG**	**WTM**	**MDHG**	**MDHM**
Protein	0.53	0.58 ± 0.03	0.59 ± 0.11	0.54 ± 0.01	0.55 ± 0.003
RNA	0.066	0.072 ± 0.002	0.054 ±0.004	0.092 ± 0.002	0.093 ± 0.003
DNA	0.026	0.024 ± 0.002	0.029 ± 0.001	0.041 ± 0.012	0.031 ± 0.0003
Lipid	0.076	0.027 ± 0.002	0.026 ± 0.004	0.029 ± 0.002	0.026 ± 0.001
Lipoteichoic Acid	0.030	0.030	0.030	0.030	0.030
Cell Wall Components	0.22	0.22	0.22	0.22	0.22
Cofactors & Ions	0.045	0.045	0.045	0.045	0.045
Sum	1.00	1.00	1.00	1.00	1.00

The coefficients of the new BOFs were pairwise compared to the coefficients of the original BOF of iBsu1147. One-sample t-test was used for the comparison, and the resulting p-values are presented in Table 5.7.

Table 5.7: p-values from pairwise comparison of the coefficients of the new BOFs (WTG, WTM, MDHG and MDHM) to those of the original BOF of iBsu1147. The p-values were found by using one-sample t-test.

	p-value			
Biomass component	**WTG**	**WTM**	**MDHG**	**MDHM**
Protein	0.1047	0.6019	0.1206	0.05312
RNA	1.39E-06	2.04E-08	1.08E-12	7.55E-11
DNA	3.87E-14	2.20E-16	0.0001046	2.20E-16
Lipid	0.001081	0.003139	0.0006226	0.0002856

5.3 Specific uptake- and secretion rates

The density, molar mass, and concentration of O_2 and CO_2 are shown in Table 5.8, and were used to calculate the specific uptake- and secretion rates of O_2 and CO_2, respectively.

Table 5.8: Density, molar mass, and concentration (N) of O_2 and CO_2.

Compound	**Density (g/L)**	**Molar mass (g/mol)**	**N (mol/L)**
O2	1.43	32.00	0.0447
CO2	1.98	44.01	0.0450

Table 5.9 shows the average CDW measurements of the wild type cultivated in glucose minimal medium. The CDW measurements were used to find the growth rate by linear regression. The linear regression line is presented in Figure 5.2. The growth rate was in turn used to predict the CDW in time points used for calculations of uptake- and secretion rates of O_2 and CO_2.

Table 5.9: Average CDW, $mg\ mL^{-1}$) and the common logarithm (log_{10}) of the average CDW over time (h) for *B. subtilis* 168 wild type cultivated in glucose minimal medium. For each time point, triplicate measurements of the CDW were conducted and averaged. Standard deviation (SD) of each triplicate is found by using the log_{10} value of each replicate.

Time (h)	**Average CDW** ($mg\ mL^{-1}$)	log_{10} **Average CDW**	**SD**
2.83	1.05	0.0226	0.0333
3.18	1.30	0.114	0.0450
3.35	1.80	0.254	0.0653
3.60	2.08	0.317	0.0415
3.88	2.35	0.370	0.0668
4.15	3.16	0.500	0.0255

Table 5.10 shows the 10 time points with $O2_{out}$ measurements, flow rates and predicted CDWs used to calculate the avg. q_{O_2} of *B. subtilis* 168 WT cultivated in glucose minimal medium.

The uptake- and secretion rates of nutrients and gases were measured for each fermentation. The uptake- and secretion rates for carbon source, L-glutamic acid, L-tryptophan and acetate are presented in Table 5.11. The uptake rates of O_2 and the secretion rates of CO_2, as well as the respiratory quotient (RQ) are presented in Table 5.12.

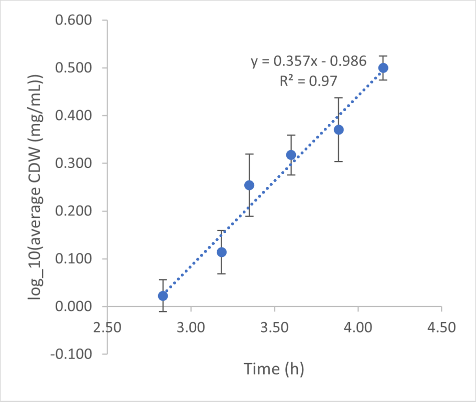


Figure 5.2: Common logarithm (log_{10}) of the average CDWs over time (h) for *B. subtilis* 168 cultivated in glucose minimal medium. A linear regression line has been fitted to the datapoints, and the equation and R^2 are displayed on chart.

Table 5.10: Specific O_2 uptake rate (avg. q_{O2}) based on 10 output gas measurements conducted during the cultivation of *B. subtilis* 168 WT on glucose minimal medium. The 10 measurements are conducted towards the end of the cultivation, before the filter was clogged. Each q_{O_2} has been calculated by using Equation 2.1 and 2.2. The value of $O_{2_{air}}$ needed for Equation 2.1 is based on 10 output gas measurements conducted before inoculation of the reactor, and was found to be 20.895.

Time (h)	**Predicted CDW (g)**	**Flow rate (mL min^{-1})**	$O_{2_{out}}$ (%)	q_{O_2} **(mmol/gCDW h)**	**avg.** q_{O_2}	SD
3.75	3.32	659	19.9203	-5.18	-5.14	0.110
3.78	3.39	659	19.9040	-5.17		
3.80	3.44	649	19.8887	-5.09		
3.82	3.50	661	19.8678	-5.20		
3.85	3.57	649	19.8476	-5.10		
3.87	3.64	637	19.8251	-5.01		
3.89	3.71	663	19.8187	-5.15		
3.92	3.78	689	19.7980	-5.35		
3.94	3.86	656	19.8088	-4.95		
3.97	3.94	646	19.7271	-5.14		

Table 5.11: Uptake- and secretion rates (mmol/gDW h) of glucose, methanol, L-glu, L-trp, and acetate for each biomass sample. The values presented are the average rates +- SD.

	Substrate uptake rate (mmol/gDW h)				
Compound	**WTG**	**WTM**	**MDHG**	**MDHM**	**p-value**
Glucose	3.0 ± 0.83	3.1 ± 0.94	1.8 ± 0.37	1.3 ± 0.24	0.0262
Methanol	0	2.7 ± 0.72	0	2.1 ± 0.36	0.337
L-glu	1.7 ± 0.53	0.5 ± 0.18	1.1 ± 0.27	0.7 ± 0.13	0.00706
L-trp	0.070 ± 0.02	0.016 ± 0.005	0.010 ± 0.002	0.003 ± 0.001	6.07e-05
Acetate	3.8 ± 0.17	5.7 ± 0.15	2.7 ± 0.08	2.2 ± 0.13	6.69e-09

Table 5.12: Specific uptake- and secretion rates (mmol/gDW h) of O_2 and CO_2, and respiratory quotient (RQ) for each biomass sample. Negative and positive rates represent uptake and secretion, respectively.

	Specific substrate uptake rate (mmol/gDW h)				
Compound	**WTG**	**WTM**	**MDHG**	**MDHM**	**p-value**
O2	-5.16 ± 0.11	-9.27 ± 0.33	-4.83 ± 0.18	-3.58 ± 0.05	<2e-16
CO2	5.22 ± 0.12	9.56 ± 0.36	4.76 ± 0.64	3.60 ± 0.08	<2e-16
RQ	-1.01	-1.03	-0.99	-1.01	

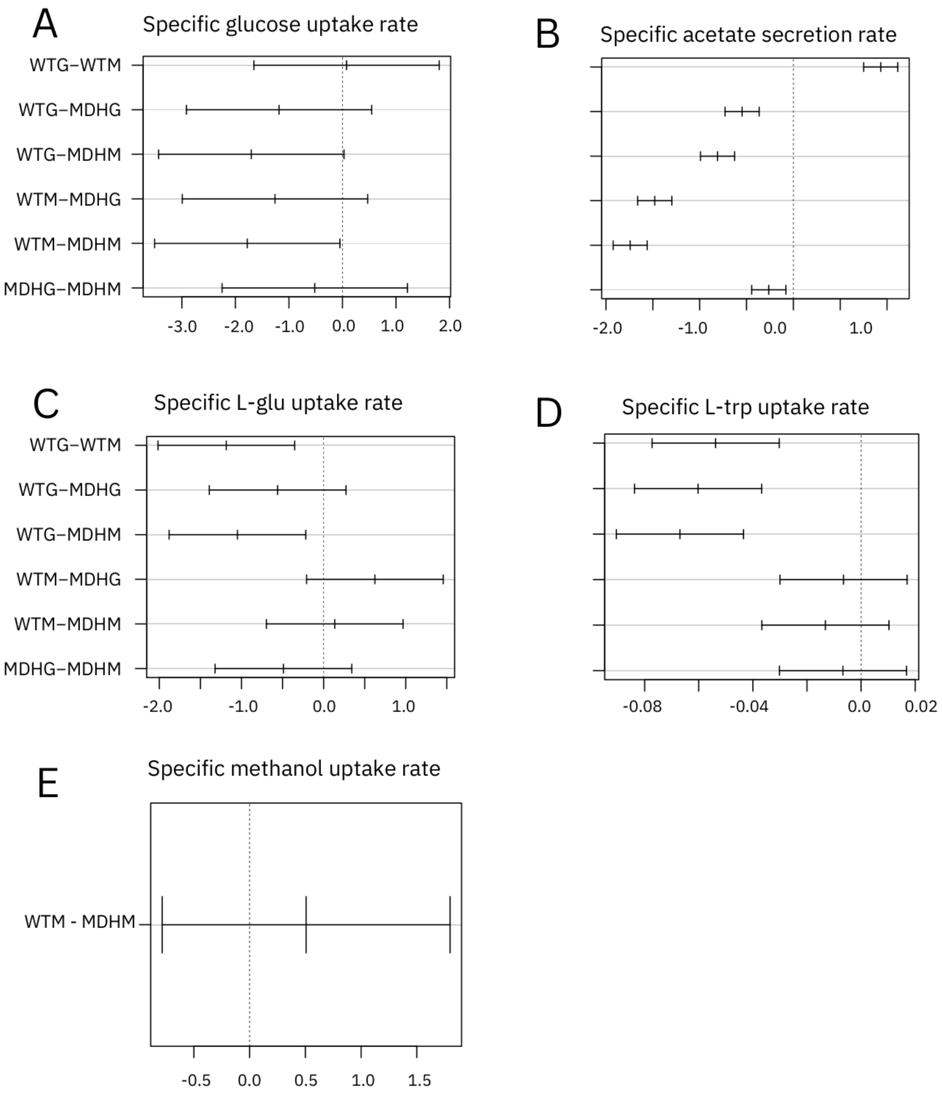


Figure 5.3: Tukey's HSD test for multiple comparisons was conducted to decide if there are significant differences in specific uptake- and secretion rate between fermentations. The figure shows the Tukey's HSD test results for A) specific glucose uptake rate, B) specific acetate secretion rate, C) specific L-glu uptake rate, D) specific L-trp uptake rate, and E) specific methanol uptake rate. The results are presented as 95 % family-wise confidence intervals. Confidence intervals that do not span 0.0 represent a significant difference.

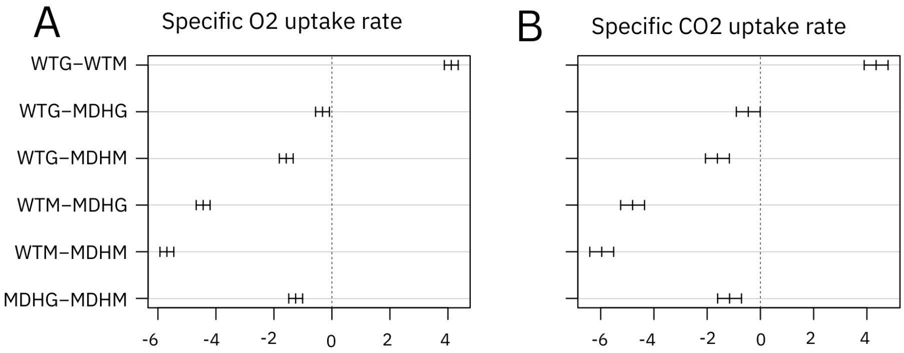


Figure 5.4: The figure shows the Tukey's HSD test results for A) specific O_2 uptake rate, and B) specific CO_2 secretion rate. The results are presented as 95 % family-wise confidence intervals. Confidence intervals that do not span 0.0 represent a significant difference.

5.4 Using the *B. subtilis* GEM for analysing the phenotype of the organism

The predicted growth rate of iBsu1147, WTG, WTM, MDHG, and MDHM were measured and compared, and the nutrient environment that was used is presented in Table 5.13.

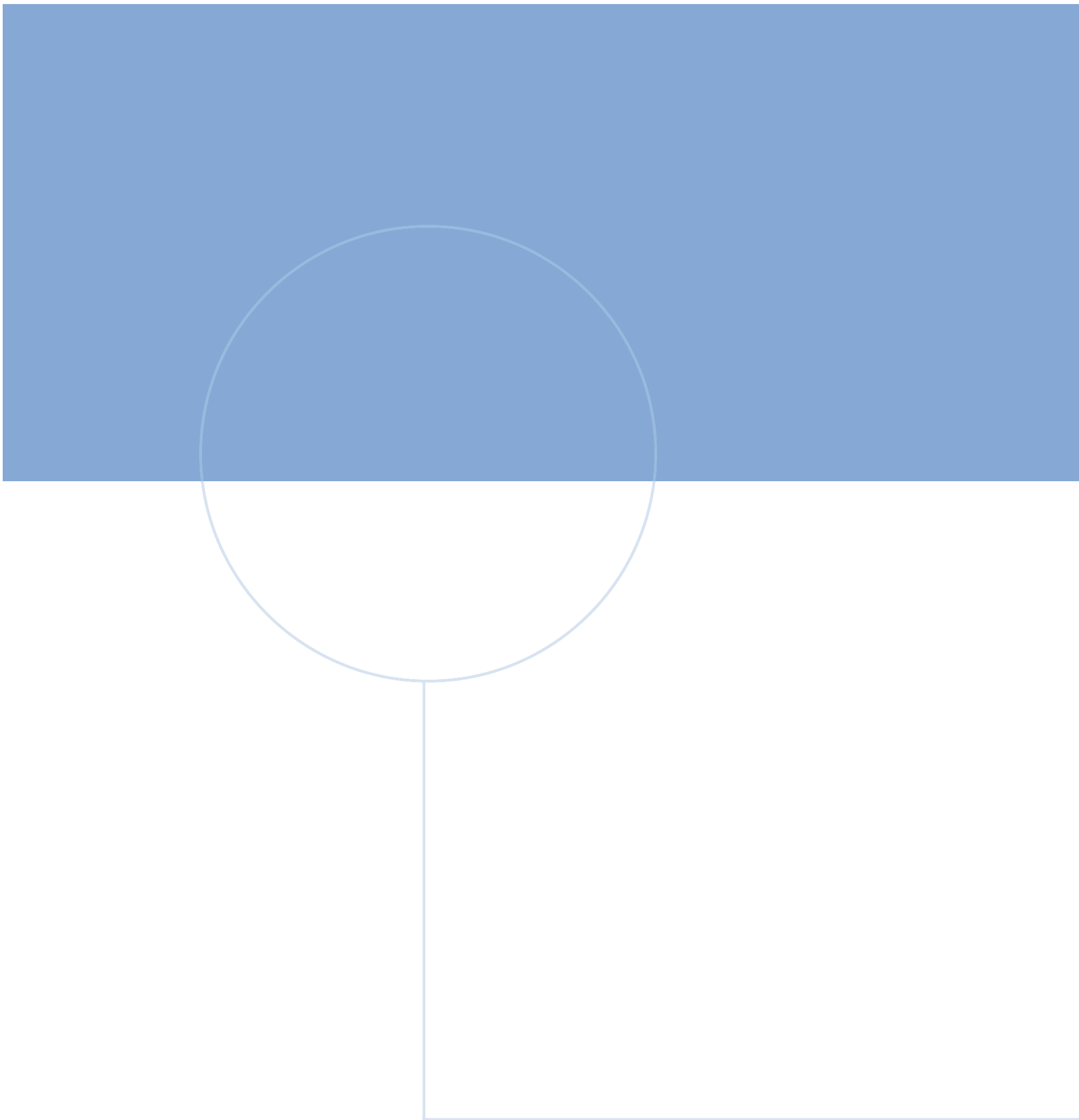
Table 5.13: Nutrient environment used to compare the predicted growth rates of iBsu1147, WTG, WTM, MDHG and MDHM. The components that are listed represent the components that were available during experimental cultivations. The corresponding exchange reaction for each component, and the chosen flux bounds are given. A flux bound of (-1000.0, 1000.0) represents unrestricted flux. For D-glucose, l-tryptophan and l-glutamic acid, the uptake rate has been restricted to -1.8, -0.5 and -1.0 respectively. For exchange reactions that are not visualized in the table, the flux bounds were set to (0, 1000.0).

Component	**Exchange reaction ID**	**Lower bound**	**Upper bound**
Na+	E00160	-1000	1000
NH4+	E00006	-1000	1000
HPO42-	E00003	-1000	1000
K+	E00084	-1000	1000
Mg2+	E00103	-1000	1000
$SO_4{}^{2-}$	E00023	-1000	1000
Fe2+	E00217	-1000	1000
Zn2+	E00015	-1000	1000
Ca2+	E00030	-1000	1000
Cu2+	E00027	-1000	1000
Mn2+	E00013	-1000	1000
Co2+	E00067	-1000	1000
MoO4$\hat{2}$-	E00200	-1000	1000
O2	E00002	-1000	1000
H2O	E00001	-1000	1000
H+	E00033	-1000	1000
D-glucose	E00096	-1.8	1000
L-tryptophane	E00032	-0.5	1000
L-glutamic acid	E00009	-1.0	1000

Furthermore, we checked if the experimentally measured growth rate of each fermentation was better explained by the old BOF and protein synthesis reaction from iBsu1147, or by the fermentation-specific BOF and protein synthesis reaction. The predicted growth rate for each fermentation when using the old BOF and protein reaction compared to the new BOF (mBOF) and protein reaction are presented in Table 5.14. The experimentally measured growth rate for each fermentation is also presented.

Table 5.14: Measured and predicted growth rate (h^{-1}) for each fermentation. Here, ”BOF” means that the BOF and protein synthesis reaction of iBsu1147 were used, while ”mBOF” means that the new BOF and protein synthesis reaction specific for the fermentation was used.

	Growth rate (h^{-1})		
Fermentation	**Predicted (BOF)**	**Predicted (mBOF)**	**Measured**
WTG	0.231	0.234	0.36 ± 0.03
WTM	0.438	0.437	0.35 ± 0.05
MDHG	0.188	0.188	0.13 ± 0.03
MDHM	0.122	0.113	0.19 ± 0.02



NTNU
Norwegian University of
Science and Technology