

Acknowledgements

First of all, I would like to thank my supervisors Eivind Almaas and Christian Schulz for giving me the opportunity to work on this exciting project. I am truly grateful for the exceptional guidance you have given me and for everything you have thought me. The completion of this work would not have been possible without your support.

I would also like to express my gratitude to the rest of the SystBioLab group for all valuable input on my project. Especially Emil Karlsen and Vetle Simensen, who have given me lots of advice and guidance in my work.

Thanks, also, to Lilja Brekke Thorfinnsdottir for helping with setting up and running the bioreactors, to Siri Stavrum for running multiple HPLC analyses for the protein assay and to Jochen Schmidt and the people at The Technical University of München for performing the carbohydrate fingerprint analysis and elemental analysis by combustion.

Finally, I would like to thank my friends for all the support they have given me and for making my time in Trondheim a truly unique experience. My thoughts go to Eirik, whom I wish was still here with us.

Abstract

The field of systems biology has enjoyed increasing popularity and recognition over the last two decades. The systems biology approach is holistic in that it studies complex biological systems, such as cells or organisms, as a whole instead of reduced to their isolated parts. This approach requires a multidisciplinary skill base including biology, informatics, and statistics. An important subfield of systems biology is constraint-based modelling, which has become a widely used research tool within medicine and biotechnology. A particularly important ingredient in constraint-based analysis is genome-scale metabolic models (GEMs). GEMs are mathematical reconstructions of metabolic networks where the presence of each metabolic reaction is based on the organism's genome. All reactions and metabolites are represented in a stoichiometric framework. A key part of GEMs is the biomass objective function (BOF), which imitates biomass production by consuming energy, macromolecule monomers, and other molecules that make up the dry biomass composition of a cell. Biomass generation, i.e., the growth of an organism, is not just an important output in constraint-based analysis: growth maximisation is assumed to be the biologically logical goal, thus, the BOF is commonly used as the objective function.

The stoichiometry of a BOF is defined by the amount of each metabolite that goes into 1 g of dry biomass, i.e., all compounds that the organism needs to grow. Consequently, the BOF has a great impact on prediction results and should therefore be an accurate representation of the biomass composition of the studied organism. Due to a lack of publications on complete molecular biomass compositions, BOFs are usually imported from previous GEMs or GEMs of related organisms. However, there can be a great variation in biomass composition, even between closely related organisms, which would cause inaccurate predictions in a GEM. Another approach commonly used in modelling is to base the content of various parts of the BOF on various publications. For example, by importing the protein content and composition from a proteomics study and the lipids from a lipidomics study. Even if the studies are using the same organism strain, an issue arising from this approach is that the growth conditions usually vary. As biomass composition also varies greatly between growth conditions for the same organism, this will also result in potential inaccuracies.

The principal aim of this thesis is to construct condition-specific biomass objective functions for *E. coli* K-12 MG1655 based on experimental data. Consequently, we have grown *E. coli* under various controlled conditions and applied a set of analytical methods to measure the complete macromolecular composition. Proteins, carbohydrates, lipids, DNA, and RNA were quantified with an average total biomass recovery of 86 % for all cultures. The monomer distribution in proteins and carbohydrates was also measured directly achieving an increased resolution of carbohydrate monomers compared to what has been previously reported. Furthermore, our measurements are comparable to the gold-standard composition reported in literature for *E. coli*. The experimental pipeline was also applied to *S. cerevisiae*, *S. salar*, and *A. thaliana* to assess the potential for use on other organisms.

The BOFs identified for *E. coli* were implemented to the iML1515 model – the leading GEM for *E. coli* K-12 MG1655. The BOF constructed from our measurements of *E. coli* growing exponentially had a 5 % increase in prediction accuracy compared to the more general wild-type BOF shipped with the model. Simulation of the remaining BOFs predicted biologically reasonable uptake rates of limiting nutrients at given growth rates. Our results show that there is a potential for increasing the prediction accuracy of GEMs by directly measuring the biomass composition of the modelled organism. Further, we show that the constructed measurement pipeline can easily be applied to determine such BOFs relevant for GEMs.

Sammendrag

Systembiologi er et felt som har fått økende oppmerksomhet gjennom de siste to tiårene. Feltet baserer seg på en holistisk tilnærming, der komplekse biologiske systemer studeres i sin helhet i stedet for å reduseres til enkeltdele. Dette er en interdisiplinær tilnærming som blant annet kombinerer biologi, statistikk og informatikk. Et viktig felt innen systembiologi er restriksjonsbasert modellering – en teknikk som nå er mye brukt innen medisin og bioteknologi. En populær type modell som brukes i restriksjonsbasert modellering er genomskala metabolske modeller (GEMer). GEMer er matematiske rekonstruksjoner av metabolske nettverk, der alle reaksjonene er basert på den studerte organismens genom. Alle reaksjoner og metabolitter er representert i et støkiometrisk rammeverk i en slik modell. Biomasseobjektivfunksjonen (BOF) er en viktig del av GEMer. Denne reaksjonen imiterer vekst ved å forbruke energi, makromolekylmonomerer og andre molekyler som utgjør biomassesammensetningen til en celle. Biomassegenerering, dvs. veksten til en organisme, er ikke bare en viktig del av resultatet i restriksjonsbasert analyse: maksimering av vekst anses å være det biologisk logiske målet, så en BOF brukes ofte som objektivfunksjon i restriksjonsbasert modellering.

Støkiometrien til en BOF defineres av mengden av hver metabolitt som forbrukes i produksjonen av 1 g biomasse; dvs. alle stoffene en organisme trenger for å vokse. Dermed har BOFen stor innvirkning på prediksjonsresultater, noe som betyr at den bør være en mest mulig presis representasjon av biomassesammensetningen til den studerte organismen. På grunn av mangel på publikasjoner av den komplette biomassesammensetningen til organismer, importeres ofte BOFer fra tidligere GEMer, eller fra GEMer for beslektede organismer. Det er imidlertid store variasjoner i biomassesammensetning, selv mellom nært beslektede organismer, hvilket medfører unøyaktige prediksjoner i en GEM. En annen tilnærming som brukes mye i modellering er å basere innholdet av ulike deler av BOFen på ulike publikasjoner. For eksempel ved å importere proteininnhold og -sammensetning fra en proteomstudie og lipider fra en lipidstudie. Selv om ulike studier skulle bruke samme stamme av organismen, vil det ofte være variasjoner i vekstbetingelser. Siden biomassesammensetning varierer betydelig med varierende vekstbetingelser, vil dette også resultere i potensielt unøyaktige prediksjoner.

Hovedmålet med denne oppgaven er å konstruere vekstbetingelsesspesifikke biomasseobjektivfunksjoner for *E. coli* K-12 MG1655. For å gjøre dette, har vi dyrket *E. coli* under ulike kontrollerte vekstbetingelser og brukt et sett analytiske metoder til å måle hele den makromolekylære sammensetningen. Proteiner, karbohydrater, lipider, DNA og RNA er kvantifisert og vi har målt gjennomsnittlig 86 % av den totale biomassesammensetningen i alle kulturene. For proteiner og karbohydrater har vi også målt monomerfordelingen. På denne måten har vi oppnådd mer detaljerte målinger av karbohydrater enn tidligere publikasjoner. Videre, er målingene våre sammenlignbare med den biomassesammensetningen av *E. coli* i litteraturen som anses som gullstandard. Sett av eksperimentelle metoder er også testet på *S. cerevisiae*, *S. salar* og *A. thaliana* for å undersøke potensialet for å bruke metodene på andre organismer.

Vi implementerte BOFene vi konstruerte for *E. coli* i modellen iML1515 – den ledende GEMen for *E. coli* K-12 MG1655. BOFen som var konstruert fra våre målinger av *E. coli* i eksponentiell vekst oppnådde 5 % mer nøyaktige prediksjoner enn en mer generell villtype BOF fra modellen. Simuleringer med de resterende BOFene ga biologisk rimelige prediksjoner for opptaksrater av begrensende næringsstoffer ved ulike vekstrater. Resultatene våre tilsier at prediksjonsnøyaktigheten til GEMer kan forbedres ved å direkte måle biomassesammensetningen til organismen som modelleres. Videre har vi vist at settet av metoder vi har foreslått er en enkel måte å bestemme biomassesammensetning for bruk i restriksjonsbasert modellering.

Table of contents

Acknowledgements.....	I
Abstract.....	II
Sammendrag.....	III
Abbreviations.....	IV
1 Introduction.....	1
2 Theory.....	4
2.1 Models in systems biology.....	4
2.1.1 Linear programing.....	4
2.1.2 Flux balance analysis.....	5
2.1.3 The biomass objective function.....	6
2.2 Escherichia coli metabolism and biomass composition.....	8
2.2.1 Protein.....	9
2.2.2 Carbohydrate.....	9
2.2.3 Lipid.....	10
2.2.4 DNA.....	11
2.2.5 RNA.....	11
2.3 Experimental methods.....	11
3.2.1 Culturing in bioreactor.....	11
3.2.2 Chromatography and mass spectrometry.....	12
3.2.3 Liquid-liquid phase extraction.....	12
3.2.4 Spectrophotometry.....	13
3 Materials and Methods.....	14
3.1 Culture conditions and biomass harvest.....	14
3.1.1 Exchange rate determination.....	15
3.2 Biomass composition analysis.....	15
3.2.1 Protein.....	15
3.2.2 Carbohydrate.....	15
3.2.3 Lipid.....	16
3.2.4 DNA.....	16
3.2.5 RNA.....	16
3.2.6 Elemental analysis.....	17
3.3 Biomass objective functions.....	17
4 Results and Analysis.....	18
4.1 Growth and exchange rates.....	18
4.1.1 Growth and exchange rates of <i>E. coli</i> U-lim.....	18
4.1.2 Growth rates of <i>E. coli</i> cultures grown in chemostat.....	20
4.2 Biomass composition analysis.....	20
4.2.1 Protein.....	21
4.2.2 Carbohydrate.....	23
4.2.3 Lipid.....	25
4.2.4 DNA.....	27
4.2.5 RNA.....	28
4.2.6 Total biomass composition.....	29
4.3 Biomass objective functions.....	31
4.4 Flux balance analysis.....	32
5 Discussion.....	37

6 Conclusion and Outlook.....	44
References.....	46
Appendix A: NMR based quantification of glucose, lactate and glutamine in cell culture media.....	53
Appendix B: Amino acid distribution.....	58
Appendix C: Ribose comparison between carbohydrates and RNA.....	59
Appendix D: Protein and carbohydrate functions and chemical formulas.....	60
Appendix E: Compounds implemented from <i>iML1515</i>	62
Appendix F: Complete biomass objective functions.....	64

Abbreviations

<i>A. thaliana</i>	=	<i>Arabidopsis thaliana</i>
BOF	=	Boimass objective Function
CDW	=	Cell dry weight
COBRA	=	Constraint-based reconstruction and analysis
DO	=	Dissolved oxygen
<i>E. coli</i>	=	<i>Escherichia coli</i>
EDTA	=	Ethylenediaminetetraacetic acid
ESI-MS	=	Electrospray ionisation mass spectrometry
FAME	=	Fatty acid methyl ester
FBA	=	Flux balance analysis
GAM	=	Growth associated maintenance
GC-VUV	=	Gas chromatography vacuum ultraviolet spectroscopy
GEM	=	Genome-scale metabolic model
GPR	=	Gene-protein reaction
HPLC	=	High performance liquid chromatography
KDO	=	3-deoxy-D-manno-2octulosonic acid
LLE	=	Liquid-liquid extraction
LP	=	Linear programming
LPS	=	Lipopolysaccharide
NGAM	=	Non-growth associated maintenance
NMR	=	Nuclear magnetic resonance
OD	=	Optical density
OPA	=	O-phthalaldehyde
PhPP	=	Phenotype phase plane
PMP	=	1-phenyl-3methyl-5-pyrazolene
RQ	=	Respiratory quotient
<i>S. cerevisiae</i>	=	<i>Saccharomyces cerevisiae</i>
<i>S. salar</i>	=	<i>Salmo salar</i>
SDS	=	Sodium dodecyl sulphate
TE	=	Tris-EDTA
THF	=	Tetrahydrofuran
UHPSCF-MS	=	Ultra-high performance supercritical fluid chromatography tandem mass spectrometry

1 Introduction

Traditionally, reductionism has dominated life sciences. During the last two decades, however, the holistic approach of systems biology has received much attention (1, 2). Systems biology aims to understand whole biological systems and not just their isolated parts (3). Thusly, the systems biology approach exposes emergent properties that are often overlooked in reductionism (4). In the past, these holistic approach applications have been limited by a lack of data and tools. But now, as new technology improving computational power is developed and increasing amounts of large-scale omics data are made available, many scientists look to the systems biology approach to make sense of the complexity of biological processes(5). This approach has led to advances in a wide range of disciplines, such as immunology, biochemistry, and cancer research, to mention a few (6-8).

An important subfield within systems biology is constraint-based modelling, which has been recognised as an important tool in fields such as medicine and food technology (9-11) Constraint-based modelling has contributed to unravelling the principles underlying metabolic networks by considering physical, enzymatic topological constraints regulating their phenotype (12). According to Resendis-Antonio (12), the process of constraint-based modelling can be divided into four steps:

1. Metabolic reconstruction of an organism
2. Mathematical representation of the metabolic network
3. In silico analysis
4. Experimental assessment of computational predictions.

The first two steps are accounted for by the construction of a genome-scale metabolic model (GEM) (13, 14). GEMs are widely used in biology as a tool to utilise large-scale omics data to predict metabolic phenotypes, discover gene functions and predict gene modification targets (15). A GEM is a mathematical representation of the metabolic network of a cell, a tissue, or a whole multicellular organism, in which all known reactions and metabolites are integrated into a stoichiometric framework and balanced for mass and energy (16). Furthermore, the gene-protein-reaction (GPR) relationships are annotated, meaning that each reaction in the model is linked to the protein(s) that catalyse it, and the proteins are linked to their respective encoding genes. The GPR annotation makes it possible to integrate both proteomic and transcriptomic data into one model (16). This possibility to combine, organize and utilise large-scale omics data might be why constraint-based modelling has received enormous popularity in the last decade. According to Gu *et al.* (13), GEMs had been constructed for 6239 different organisms in February 2019, and there is no reason to believe that this trend will turn any time soon.

Representing metabolic models as GEMs allows researchers to run in silico analyses that cannot be performed using general metabolic databases. GEMs can be analysed using various COBRA (COstraints Based Reconstruction and Analysis) methods, one of the simplest being flux balance analysis (FBA) (17). In an FBA, the flow of metabolites through the metabolic network is calculated by optimizing an objective function within a solution space limited by the assumption of steady-state and a set of applied flux constraints. The objective function can be any linear combination of fluxes, but often the biologically logical objective is to maximise growth (18, 19). To simulate this, a pseudo-reaction called the biomass objective function (BOF) is used as the objective function.

A BOF imitates growth by consuming metabolic precursors of proteins, carbohydrates, lipids, and nucleic acids, and in many cases also metabolites such as vitamins, minerals, and some high-energy metabolites (20). The amount of each metabolite consumed in a BOF should reflect the content of each metabolite that is needed to produce 1 g of the studied organism's biomass (21).The BOF is not balanced for mass, in the sense that it usually only produces some lower energy metabolites, protons, and inorganic phosphates. The "lost" mass is assumed to make up the biomass of the organism. The

multidimensional direction of the objective function has a great impact on the solution in an FBA, thus the BOF must be an utmost precise representation of biomass generation in the studied organism (21-26). Hence, knowledge about the biomass composition of the studied organism is essential in the construction of a BOF.

Due to the lack of complete biomass composition data for various organisms, many BOFs today are either imported from previous GEMs of the same organism or GEMs constructed for related organisms (21, 27, 28). Another common approach is to construct BOFs based on biomass compositions of model organisms in literature (21). However, publications of complete biomass compositions are scarce, even for well-studied organisms. Hence, various parts of the BOF are often based on various publications (29, 30). While proteomic, lipidomic etc. studies provide detailed compositions of their respective macromolecules, there are often variations in the strains used between studies. Furthermore, there are many factors affecting organism growth, and the chance of finding separate studies with the exact same growth conditions are small (31). As biomass composition can vary considerably with strain and growth conditions, basing the BOF on separate studies can result in inaccurate predictions (22, 23, 31).

Considering condition-specific variations in biomass composition, the complete biomass should optimally be measured in one culture grown in conditions matching the simulation. However, a “standard” approach to do this is yet to be developed (32). Attempts have been made to measure the biomass composition of *E. coli* for genome-scale metabolic modelling purposes. Beck *et al.* (33) have reviewed various methods for measuring protein, carbohydrate, lipid, DNA, and RNA and experimentally determined the biomass composition of *E. coli*, *Alicyclobacillus acidocaldarius*, and *Synechococcus* sp. PCC 7002 and for computational purposes. However, they could only recover up to 65.2 % of the cell dry weight (CDW), which gives a high uncertainty when the composition is normalized to 100 % for BOF applications. In 2014 Long and Antoniewicz (34) developed a gas chromatography/mass spectrometry method for quantifying the content of proteins, lipids, RNA, and glycogen. They applied the method on three *E. coli* strains and recovered 82 % of the CDW of wild-type *E. coli*. Even though the recovery percentage is high, the method has issues such as the requirement that the organism must be completely C13 labelled. Besides, the method does not measure the content of DNA and carbohydrates other than glycogen.

In this work, we suggest a pipeline of analytical methods for determination of complete macromolecular composition. This pipeline involves amino acid quantification by HPLC; carbohydrate quantification by HPLC-MS; lipid extraction and quantification according to Folchs method; DNA extraction and spectrophotometric quantification; and RNA extraction and spectrophotometric quantification. The methods are relatively simple and time-efficient, and require no carbon labelling. We applied the methods to seven cultures of *E. coli* K-12 strain MG1655 and achieved an average biomass recovery of 86 %. For proteins and carbohydrates, we did not only measure the total content, but also the composition. The carbohydrate quantification method we applied resulted in an increased resolution of carbohydrate monomers compared to what has been previously reported. Furthermore, our measured values are comparable to the gold-standard *E. coli* composition reported by Neidhardt *et al.* (35). Additionally, some of the methods in our pipeline were tested on *Saccharomyces cerevisiae* cells, *Salmo salar* muscle, and *Arabidopsis thaliana* leaves to assess their potential for other organisms.

The workflow in this thesis involves cultivation of the organism, experimental quantification of each macromolecule group, BOF construction, and BOF evaluation through FBA analysis. The *E. coli* cultures we have analysed were grown in bioreactors to ensure an aerobic environment and controlled growth conditions. Furthermore, growth rates and some exchange rates have been measured during cultivation, allowing experimental assessment of the computational predictions made using our BOFs. BOFs were constructed based on the experimentally determined macromolecule composition and

implemented to the *E. coli* GEM *iML1515*. To evaluate the BOFs, FBA analysis was performed with restrictions corresponding to the measured exchange rates. The simulation results showed potential for increasing prediction accuracy by using experimentally determined, condition-specific BOFs.

The principal aim of this thesis is to construct condition-specific biomass objective functions for *E. coli* based on experimental data. To do this we have grown *E. coli* under controlled conditions and combined already existing methods of macromolecule quantification to measure the complete macromolecule composition of the cultures.

2 Theory

The problem addressed in this thesis is tripartite. The first part is analytical and consists of the experimental determination of an organism's biomass composition. The second part is computational and consists of implementing the experimentally measured biomass into a computational model followed by simulation. The third part is the biological aspect, which should be considered when designing experiments, constructing biomass objective functions, and evaluating the simulation results by comparison to in vivo phenotypes.

In this chapter, I will first describe the principles behind the computational methods used. To avoid getting lost in the plethora of computational techniques used in systems biology, I have chosen to focus solely on the methods used directly in this work. Then I will address the biological aspect briefly, focusing on the biomass composition of *E. coli*. Lastly, I will touch upon the analytical methods used in this thesis. As the goal of this thesis is not to develop new methods but rather to combine relatively simple methods effectively, I will not go into the principles behind the methods or the newest technology.

2.1 Models in systems biology

Models are a central part of systems biology because biological systems are far too complex to be described without any degree of simplification (3). To make sense of the large amounts of data describing such systems, it must be put into a framework. Computational models provide frameworks into which various kinds of omics data can be combined (16). In constraint-based modelling, experimental data is incorporated into a model as constraints that will increase the accuracy of the models' predictions (36). Various COBRA methods have been developed to analyse metabolic models, the most commonly used being FBA (18). FBA is a method for predicting the flux through a metabolic network at a steady state. It is a simple and efficient technique that is based on the principles of linear programming.

2.1.1 Linear programming

Linear programming or linear optimization is a technique in the field of operations research that historically has been used in company management, but relatively recently it has become an indispensable tool in systems biology as well (37). The linear programming (LP) problem consists of a function to be optimized (maximised or minimised) subject to certain linear constraints (38). The function to be optimized (Z) is called the objective function and is a linear function of a set of n decision variables x_j :

$$Z = c_1x_1 + c_2x_2 + \dots + c_nx_n. \quad 2-1$$

The constraints are linear equations or inequalities of the same decision variables as the objective function. Equation 2-2 shows the equation/inequality of a given constraint i .

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \begin{cases} \leq \\ = \\ \geq \end{cases} b_i \quad 2-2$$

Inequalities can easily be converted mathematically from "not less than" to "not greater than" and vice versa by multiplying b by -1 , and each equation can simply be replaced with a set of two opposite inequalities. In this way the LP problem can be formulated with all restrictions as "less-than" inequalities, which is the standard form (38). The standard form of the LP problem with n decision variables and m constraints is shown in Equation 2-3, where \bar{c} is a vertical vector of the constants in the objective function, \bar{x} is a vertical vector of the decision variables x_j , \bar{b} is a vertical vector of all the

constants b_i that constraint i is bound to be “not greater than” and $\bar{\mathbf{A}}$ is an $[m \times n]$ matrix containing the constants a_{ij} of the constraints. In addition to the constraints defined by $\bar{\mathbf{A}}$, all decision variables must be non-negative in standard form.

$$\begin{aligned} & \text{maximise} && \bar{\mathbf{c}}^\top \bar{\mathbf{x}} && 2-3 \\ & \text{subject to} && \begin{cases} \bar{\mathbf{A}}\bar{\mathbf{x}} \leq \bar{\mathbf{b}} \\ \bar{\mathbf{x}} \geq \mathbf{0} \end{cases} \end{aligned}$$

Any solution for $\bar{\mathbf{x}}$ that satisfies all the constraints is called a feasible solution (38). Out of all feasible solutions, the solution(s) that gives the objective function the most optimal value is/are the optimal solution(s). This is illustrated for an example LP problem with only two decision variables and three constraints in Figure 2.1.

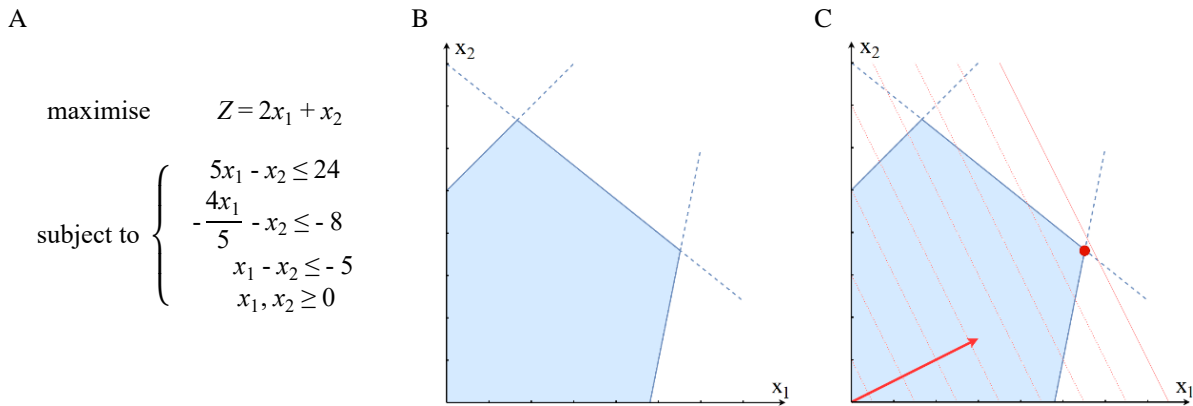


Figure 2.1: Solving an LP problem. **A:** The LP problem in standard form. **B:** The constraints (blue lines) limit the solution space (blue area) of the problem. Any solution within this area is feasible. **C:** The optimal solution (red point) is the solution within the feasible area that maximises the value of the objective function (red arrow).

All the constraints in an LP problem are linear, which means that at least one of the optimal solutions of the problem will be in a corner of the solution space. Furthermore, any local optimum in an LP problem is also a global optimum. Hence, LP problems can be solved efficiently with the right algorithms. One popular algorithm for solving LP problems is the SIMPLEX algorithm, which solves LP problems by iteratively investigating the corners of the solution space, moving from one corner to the next along the edge which optimized the objective function the most (38). When there is no edge along which the algorithm can move to further optimize the value of the objective function, it has found an optimal solution.

2.1.2 Flux balance analysis

Flux balance analysis (FBA) is a method that utilises the principles of linear programming to analyse GEMs and other metabolic reconstructions (18). In simplified terms, a GEM consists of a set of n chemical reactions with m different metabolites that represent all known reactions in the metabolism of the modelled system or organism. In FBA, these reactions and metabolites are represented mathematically in a stoichiometric m times n matrix $\bar{\mathbf{S}}$, where S_{ij} is the stoichiometric coefficient of metabolite M_i in reaction R_j . Furthermore, when running an FBA, each reaction will have a flux v_j . The change in the concentration of metabolite M_i can be described as follows (39):

$$\frac{dM_i}{dt} = \sum_{j=1}^n S_{ij}V_j \quad 2-4$$

The FBA approach relies on the steady state approximation (SSA), in which it is assumed that the system has fully adapted to the environment, and the concentrations of all metabolites are constant (39). In terms of linear programming SSA can be translated to the following set of constraints:

$$\bar{S} \bar{v} = \bar{0}, \quad 2-5$$

where \bar{v} is a vertical vector containing the fluxes through all the n reactions. By assuming steady state, the FBA approach circumnavigates the need of information on enzyme kinetics and metabolite concentration. Hence, the metabolic fluxes through the system can be predicted based on only the stoichiometric coefficients in the metabolic network (18).

In addition to the constraints imposed by SSA, each flux v_j is constrained by a lower bound a_j and an upper bound b_j (18). Adjusting these bounds on uptake reactions, that is, reactions that produce metabolites such as glucose, oxygen, or ammonia without consuming anything, is a common way to simulate specific growth conditions. The whole LP problem in FBA is shown in Equation 2-6:

$$\begin{aligned} &\text{maximise} && Z = \bar{c}^T \bar{v} && 2-6 \\ &\text{subject to} && \begin{cases} \bar{S} \bar{v} = \bar{0} \\ \bar{\alpha} \leq \bar{v} \leq \bar{\beta} \end{cases} \end{aligned}$$

Here, $\bar{\alpha}$ and $\bar{\beta}$ are vertical vectors containing the lower and upper bounds for the fluxes in \bar{v} , respectively. The objective function Z is a linear combination of the fluxes in \bar{v} , specified by the constants in \bar{c} . Usually, the \bar{c} is a single-entry vector, meaning that the objective function is one of the reactions in the model. By running an FBA, the flux distribution that optimizes this objective function are calculated.

While the FBA solution provides much information about the metabolic state of the studied system it does not say much about the factors determining the solution. One way to investigate which restrictions are limiting, is by calculation of shadow prices (40). The shadow price of a constraint i is defined mathematically as:

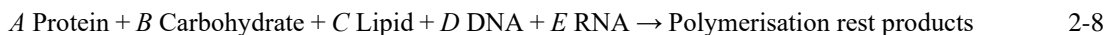
$$\gamma_i = \frac{-\partial Z}{\partial b_i}, \quad 2-7$$

where b_i is the i -th element in the vector \bar{b} , which defines the right-hand side of the constraints in an LP problem (see Equation 2-3) (41). In FBA, the shadow prices are interpreted as the change in the objective function (Z) at an optimal point, given a change in the exchange rate of metabolite (42). This means that if the shadow price of an exchange reaction is 0, changing the exchange rate would not change the optimal location and the exchanged metabolite is not limiting.

2.1.3 The biomass objective function

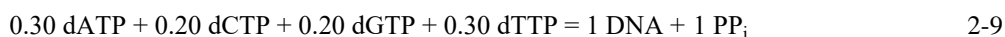
When simulating an organism, a natural choice of objective is to maximise growth (18, 19). This can be done by using a pseudo-reaction called the biomass objective function (BOF) as the objective function. A BOF imitates growth by consuming metabolites that are substrates in the production of biomass components. Feist and Palsson (20) classify BOFs into three levels of detail: basic, intermediate, and advanced.

A basic level BOF is formulated by defining the weight fractions of the different macromolecules in the cell and the distribution of monomers within each macromolecule group (20). Equation 2-8 shows the outline of the reaction equation of a basic level BOF, where the coefficients A , B , C , D , and E correspond to the amount of each macromolecule in the biomass composition of the modelled organism. Polymerisation rest products are typically water ADP and inorganic phosphates, but can also include other side products of macromolecule polymerisation reactions.



Beck *et al* (33) have suggested an approach for constructing BOFs from experimental data of macromolecule composition. First, separate functions are calculated for each macromolecule group. These macromolecular functions were constructed to imitate the synthesis or polymerisation reactions of the macromolecules, and include all metabolites consumed and produced. For protein, carbohydrate,

DNA, and RNA the functions represent the elongation of a polymer by the addition of one average monomer. To construct these functions the monomer distribution must be known. As an example, the function of hypothetical DNA consisting of 60 % AT and 40 % GC is shown in Equation 2-9:



In the case of lipids, the macromolecule function represents the production of one average lipid molecule with an average fatty acid distribution. The chemical formula of the main product of the macromolecule functions is calculated as shown for the example DNA in Table 2.1.

Table 2.1: Calculation of DNA macromolecule formula.

Compound	Stoichiometry	Chemical formula	C	H	N	O	P	Charge
dATP	0.30	$\text{C}_{10}\text{H}_{12}\text{N}_5\text{O}_{12}\text{P}_3^{-4}$	3.0	3.6	1.5	3.6	0.9	-1.2
dCTP	0.20	$\text{C}_9\text{H}_{12}\text{N}_3\text{O}_{13}\text{P}_3^{-4}$	1.8	2.4	0.6	2.6	0.6	-0.8
dGTP	0.20	$\text{C}_{10}\text{H}_{12}\text{N}_5\text{O}_{13}\text{P}_3^{-4}$	2.0	2.4	1.0	2.6	0.6	-0.8
dTTP	0.30	$\text{C}_{10}\text{H}_{13}\text{N}_2\text{O}_{14}\text{P}_3^{-4}$	3.0	3.9	0.6	4.2	0.9	-1.2
PP _i	-1	$\text{HO}_7\text{P}_2^{-3}$	0	-1	0	-7	-2	3
DNA	1	$\text{C}_{9.8}\text{H}_{11.3}\text{N}_{3.7}\text{O}_6\text{P}^{-1}$	9.8	11.3	3.7	6	1	-1

According to Beck *et al.* (33), the separate macromolecular functions can be incorporated into the total BOF by following the 4 steps described by:

1. Calculate mass fractions as g of macromolecule per g of dry cell weight. And scale the fractions to make up 100 %
2. Calculating the molar weight of each macromolecule from the elemental content (shown in Table 2.1 for DNA)
3. Dividing the mass fraction of each macromolecule by its molar mass to calculate the stoichiometric coefficients (*A*, *B*, *C*, *D*, and *E* in Equation 2-8) given in mmol/gCDW.
4. Incorporating the molar coefficients into the BOF by multiplying each macromolecule reaction by its respective stoichiometric factor.

A BOF constructed according to this procedure would be of the basic level according to Feist and Palsson (20). A BOF in the intermediate level has all the information from the basic level, but it also includes energy requirements in the form of polymerisation energy and maintenance energy (20). In a BOF, energy is accounted for as consumption of nucleotide triphosphates and production of their corresponding nucleotide diphosphates, inorganic phosphate, and other rest products. Polymerisation energy can be calculated based on measured amounts of macromolecules and knowledge about their synthesis pathways, as done for DNA in Table 2.1. Maintenance energy, on the other hand, must be measured, e.g., by calorimetry (43), or estimated (26). One way to estimate maintenance energy is by plotting experimentally determined growth rates against maximal predicted ATP production at the same growth conditions (27). ATP production rates can be predicted using a BOF lacking any energy requirements and locking the growth rate to the experimentally determined values. By examining various growth conditions, one can make a linear regression, where the non-growth associated maintenance (NGAM) and the growth associated maintenance (GAM) can be estimated to be the y-line intercept and slope, respectively. When NGAM and GAM are incorporated into a GEM, only NGAM is included in the BOF. GAM is not growth-rate dependent and is therefore implemented as a separate reaction (27). This approach for maintenance energy estimation gives good predictions, however, it requires detailed phenotypic data.

The BOF's level of resolution can be further increased from the intermediate to the advanced level by including essential core components and necessary vitamins, elements, and cofactors needed for growth (20). A BOF containing all measured metabolites, both essential and non-essential, are often

referred to as a wild-type BOF. Another advanced BOF, called the core BOF, consists of only those metabolites that are essential for cell survival and growth, not all metabolites found in biomass. According to Feist *et al.* (24), a core BOF can have increased accuracy in the prediction of essential genes, metabolites, and reactions, compared to a wild-type BOF. When it comes to growth prediction by FBA, both types of BOFs predict similar growth rates, but the wild-type BOF predicts a more complex set of fluxes than the core BOF (24). Feist and Palsson (20) states that a BOF of at least the intermediate level is needed for predicting growth rate and the addition of the essential metabolites of an advanced BOF will increase the prediction accuracy.

Aside from the level of detail, the organism- and condition- specificity is an important attribute of a BOF that affects prediction accuracy (21-26, 33). This was demonstrated for *S. cerevisiae* by Dikicioglu *et al.* (23) when they simulated the Yeast v7.0 model (44) with 72 in silico generated BOFs in addition to its original BOF under various limiting conditions. They found that the flux distribution varied greatly between BOFs, especially in some of the limited growth conditions. On the other hand, Szélioia *et al.* (45) have measured the biomass composition of various Chinese hamster ovary cell lines to make specific BOFs, but the variation in biomass composition had only a small impact on the predicted growth rate.

The BOFs constructed in this thesis will be implemented to the model *iML1515* for evaluations. This model is considered the leading GEM for *E. coli* K-12 MG1655 per today (June 2021). It accounts for 1,515 open reading frames and 2,719 reactions involving 1,192 metabolites (27). *iML1515* has two built-in BOFs of the advanced level: one wild-type BOF and one core BOF. Both BOFs are based on the gold-standard *E. coli* biomass composition reported by Neidhardt *et al.* (35). The maintenance energy term is calculated by plotting experimentally measured growth rates against predicted of ATP production, as described above. In the evaluation of one of the BOFs constructed in this thesis, the wild-type BOF from *iML1515* is used for comparison of prediction results.

2.2 *Escherichia coli* metabolism and biomass composition

E. coli is used as a model organism for all bacteria and is one of the most studied and best understood organisms today. It is a gram-negative heterotroph that can grow on various carbon sources (46, 47). *E. coli* is a facultative anaerobe meaning that it is capable of both aerobic and anaerobic growth (48). When oxygen is scarce it ferments glucose and produces various products, such as acetate, formate, and succinate in a process called mixed acid fermentation. The amount of ATP produced in mixed acid fermentation varies depending on the product. However, aerobic respiration produces more ATP per molecule of glucose and is therefore prioritized when oxygen is available (48).

The primary habitat of *E. coli* is in the gastrointestinal tract of warm-blooded animals, but it can also survive and even grow in dramatically different habitats, such as soil and water (49). The key to *E. coli*'s survival in alternative habitats is the flexibility of its metabolism, which gives it an ability to adapt to various environmental conditions (46). A wide variety of *E. coli* strains can be found in nature, some are beneficial and even essential for humans and other are pathogenic (49). *E. coli* strain K-12 MG1655 is the first strain of *E. coli* that had its genome sequenced (27). This strain is naturally found in the human intestine, where it breaks down complex carbon compounds and facilitates digestion.

In the five following subsections (2.2.1 – 2.2.5) I will describe the five macromolecule groups: proteins, carbohydrates, lipids, DNA, and RNA, with emphasis on their location and synthesis in *E. coli* and growth condition-dependent variations in their concentration and composition.

2.2.1 Protein

Proteins are essential to all life forms. They perform a broad range of cellular tasks within structure, transport, and catalysation of chemical reactions and can be found in almost all parts of the bacterial cell (50). Proteins consist of one or more long, structured chains of amino acids that are synthesized according to a recipe encoded in the genome and transcribed as mRNA. Elongation of these polypeptide chains can be summarized into 4 steps: one step of amino acid activation in cytosol followed by three steps of elongation in ribosomes (Figure 2.2) (51).

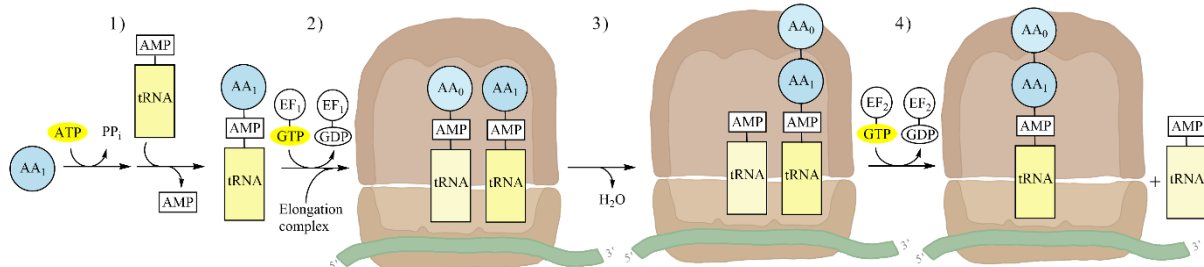
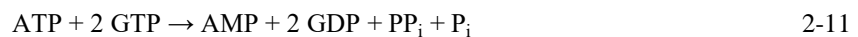


Figure 2.2: 4 steps of protein elongation. 1) Amino acid activation. 2) Decoding (binding to translation elongation complex). 3) Peptide bond formation. 4) Translocation and release of tRNA. AA = amino acid, EF = elongation factor. High energy metabolites are highlighted in yellow.

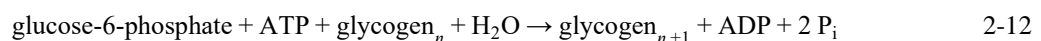
From Figure 2.2 one can see that total energy spent when elongating a polypeptide chain by one amino acid is:



The energy released in the reaction described in Equation 2-11 equals 4 molecules of ATP, which is higher than the polymerisation cost of the other biological polymers (52). Furthermore, the synthesis energy of many amino acids is high, making proteins energetically expensive molecules (53). The high energy cost associated with protein production might be one reason why gene expression is highly regulated. Growth conditions such as nutrient availability have a big impact on which proteins are produced (54, 55). Li *et al.* (56) report that the protein content is generally lower in *E. coli* grown in nitrogen-limited conditions than in carbon-limited conditions. This was partly explained by stalling of ribosome activity caused by glutamine deficiency. Bipatnath *et al.* (57) reported that the total protein concentration per mass in *E. coli* growing exponentially on unlimited media decreases with increasing growth rate. On the other hand, Li *et al.* found the protein levels to be stable with varying growth rates, both in carbon and nitrogen-limited growth (56). According to Neidhardt *et al.* (35), the dry biomass of *E. coli* is made up of 55 % protein. This percentage is based on analyses of *E. coli* strain B/r in balanced growth at 37 °C in a glucose minimal medium with a growth rate of 1.5 h⁻¹.

2.2.2 Carbohydrates

Storage carbohydrates play an important role in microbial life. They serve as quickly accessible sources of energy in times of starvation. In *E. coli* the main storage carbohydrate is glycogen (58). Glucose molecules are stored in these polysaccharides when there is an excess of energy in a process that can be summarised as follows (59):



When there is a depletion of energy, glucose molecules are released from glycogen and utilised. Glycogen levels are typically low in fast-growing *E. coli*, but when growth is limited by another nutrient than the carbon source, glycogen can contribute up to around 20 % of the dry biomass of a cell (58, 60).

In addition to storage, carbohydrates serve important functions in the cellular envelope of bacteria. N-acetylglucosamine and N-acetylmuramic acid make up the glycan strains in peptidoglycan

and provide structure and rigidity to bacterial cells (61). Carbohydrates can also be found in lipopolysaccharides (LPS) in the outer membrane of gram-negative bacteria. LPS consists of a hydrophobic domain called lipid A, a core oligosaccharide and a repeating polysaccharide called the O antigen (62). Both carbohydrate domains of LPS vary among strains and growth conditions.

2.2.3 Lipids

Lipids are a diverse group of highly reduced hydrophobic molecules that are used for energy storage in many organisms (63). In *E. coli*, however, most lipids are glycerophospholipids found in the membranes of the cellular envelope (64). Glycerophospholipids consist of a glycerol backbone with two fatty acids and one polar head group and are synthesized as shown in Figure 2.3 (65).

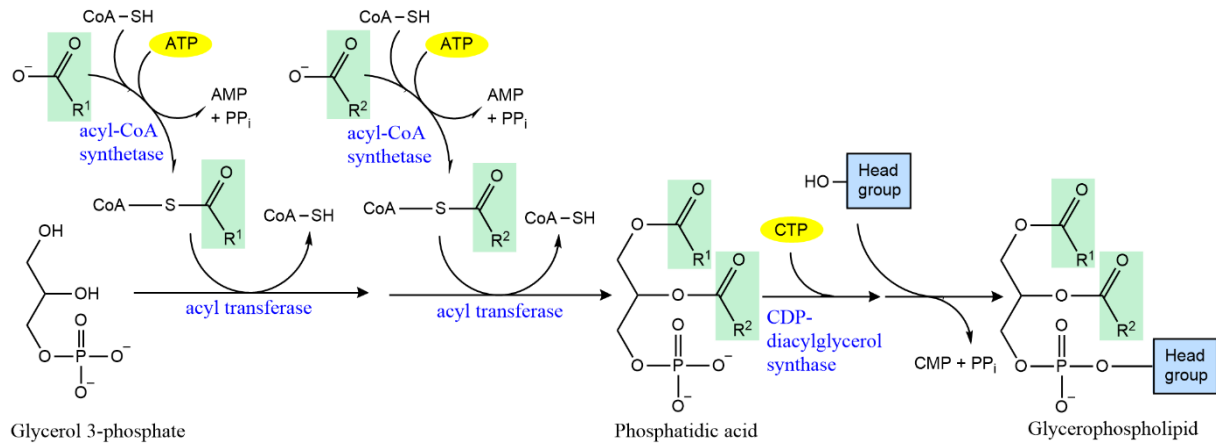


Figure 2.3: Biosynthesis of glycerophospholipids. Fatty acids are highlighted in green, high energy metabolites are highlighted in yellow, enzymes are blue.

According to Neidhardt *et al.* (35), the lipids of *E. coli* are dominated by three classes of phospholipids: phosphatidylethanolamine, phosphatidylglycerol, and cardiolipin. The head group of these glycerophospholipids is attached as shown in Figure 2.4 (65).

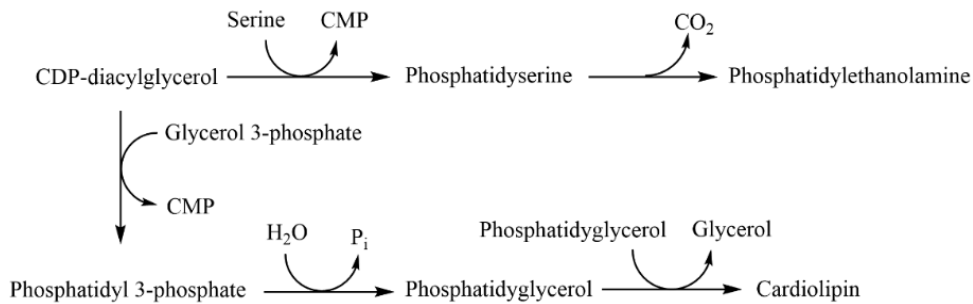


Figure 2.4: Phospholipid head group attachment of phosphatidylethanolamine, phosphatidylglycerol, and cardiolipin in *E. coli*.

The distribution of lipid classes varies between strains, between the inner and outer membrane, and between growth temperatures (64). Neidhardt *et al.* (35) report that there are three dominating fatty acids in *E. coli* membranes: palmitic (16:0), palmitoleic (16:1), and *cis*-vaccenic (18:1) acids. There are also strain-dependent variations in fatty acid composition and experiments with different growth temperatures show that the amount of saturated fatty acids increases and the amount of unsaturated fatty acids decreases with increasing temperature (64).

2.2.4 DNA

The genome of *E. coli* consists of one circular chromosome that forms the nucleoid. In addition, a small amount of DNA is found in plasmids (66). The chromosome is mostly found in a supercoiled state, but during DNA replication, it is uncoiled, and the double helix is opened. In DNA replication, new DNA is synthesized by DNA polymerases using old DNA as a template. The polymerisation reaction of DNA is summarised in Equation 2-13 (66):



where dNTP represent any of the four deoxynucleotide triphosphates: dATP, dCTP, dGTP and dTTP. In addition to the energy directly spent in polymerisation, energy is spent on unwinding the double helix, removing supercoils, ligating, and proofreading (35). Because DNA replication is an important step in microbial growth, faster-growing cells will contain more nucleoids. Still, the relative concentration of DNA in cells is stable with varying growth rate (35).

2.2.5 RNA

RNA is similar to DNA in its structure, but instead of deoxyribonucleotides, it consists of ribonucleotides. Furthermore, it is mostly single-stranded and instead of thymine, it has uracil. RNA is synthesized by RNA polymerase using one DNA strand as a template. The polymerisation reaction can be summarised to (66):



where NTP represents ATP, CTP, GTP, or UTP.

While the known functions of DNA are limited to information storage and transmission, there is a broad range of RNA types performing various functions. Neidhardt *et al.* (35) have reported that the RNA in *E. coli* is made up of 81% ribosomal RNA (rRNA), 8.6 % transfer RNA (tRNA), and 2.4 % (mRNA). However, other classes of RNA, such as small RNA (sRNA) and CRISPR RNA (crRNA) are also found in *E. coli* (67, 68). Since ribosomes are required for translation, the number of ribosomes affects the rate at which protein can be synthesised. In fast-growing cells there is a big demand for proteins, and hence, ribosomes. A linear relationship between the RNA/protein ratio and the specific growth rate has been observed (56, 69).

2.3 Experimental methods

2.3.1 Culturing in bioreactor

When studying the behaviour of microorganisms, having a controlled and observable environment is of essence. A bioreactor provides the scientist with just that – a stable and controllable environment in which microorganisms be grown and observed. A bioreactor is a controlled system and everything that goes into it can be regulated, and everything that goes out can be monitored (70). Some factors, such as temperature, pH, and gas supply are typically kept stable in the system while others are allowed or even induced to change. The culturing process is often classified by the way the substrate is added to the system. Two common types of culturing processes are batch culture and chemostat.

In a batch process, all the nutrients are added at the beginning of the cultivation without adding any more during the process that follows (19). This kind of cultivation results in a cell concentration that increases (exponential growth) as long as there are sufficient amounts of substrate in the media. The culture will reach a stationary phase when a substrate becomes limiting and a death phase when the nutrient is depleted (19).

A chemostat is an open system in which the culture is kept in a steady-state of exponential growth limited by a substrate (71). This is done by regulating the inflow of new media containing the

substrate and removal of old media containing cells (the dilution rate). In a chemostat, the concentration of cells depends on the concentration of the limiting substrate in the fresh media (72). While the culture is in a steady-state, the growth rate is equal to the dilution rate and independent of the limiting substrate concentration in the fresh media. However, if the concentration of the limiting substrate in the fresh media is too high or too low, the culture might drop out of the steady state (72).

2.3.2 Chromatography and mass spectrometry

In chromatography different compounds are separated based on their retention time in a column (73). This technique relies on the continuous differential distribution of compounds between a stationary phase and a mobile phase. The stationary phase is either a solid or a liquid sprayed onto a solid surface and the mobile phase is a liquid (liquid chromatography) or a gas (gas chromatography). Molecules with a higher affinity for the stationary phase will spend more time traveling through the column and be detected later (73).

Chromatography techniques can be classified based on the separation mechanism used. In ion-exchange chromatography, compounds are retained because their charged groups are temporarily bound to groups with the opposite charge that are covalently bound to the solid phase (74). Positively charged columns retain anions and negatively charged columns retain cations. The mobile phase contains ions that compete with the substrate in binding the stationary phase charged group. This kind of chromatography is a common choice in carbohydrate quantification (75-77).

In partition chromatography, various compounds are separated based on their differential distribution between two liquids: one polar and one nonpolar. The liquid that serves as the stationary phase is bound to the column in the form of a thin film. In reversed-phase partitioning chromatography, the stationary phase is nonpolar and the mobile phase is relatively polar (74). Hence, hydrophobic compounds will have a longer retention time than hydrophilic. This kind of chromatography has proven to differentiate between amino acids based on the polarity of their side chains (78). Other separation mechanisms used in chromatography are adsorption, affinity, and size exclusions.

Chromatography is often coupled with mass spectrometry. In mass spectrometry, compounds are converted to a charged state and differentiated based on their mass/charge ratio (79). By coupling this technique to chromatography, the sensitivity, accuracy, and speed are increased (74, 79).

2.3.3 Liquid-liquid phase extraction

Liquid-liquid extraction (LLE) is a classic technique used to extract an analyte from a solution. LLE uses two immiscible liquid phases, typically one organic and one aqueous, where the analyte is more soluble in one than in the other (80). The phases are added to the sample containing the analyte, mixed, and then separated, e.g., by centrifugation. After phase separation, the analyte will be distributed between the two phases based on its relative solubility in each phase. The extraction may be repeated multiple times depending on the distribution equilibrium to extract the total amount of analyte (80).

After extraction, the analyte can be further analysed. The concentration of analyte can be determined by various techniques. One quantification method that can be combined with LLE is gravimetric analysis, i.e., determination of concentration based on mass (81). One example of this is the Blight and Dyer method for lipid quantification (82). In this method lipids are extracted include LLE with chloroform followed by chloroform evaporation and weighting. Another technique that can be combined with LLE is spectrophotometric quantification. This is a common method for the quantification of DNA (83).

2.3.4 Spectrophotometry

Spectrophotometry is an analytical technique that is used to quantify compounds by sending electromagnetic radiation through a sample and measuring the attenuation (84). In UV-Vis spectrophotometry, the radiation sent through the sample is either UV or visible light (85). Some electrons involved in molecular bonds can absorb the energy of electromagnetic radiation of a specific wavelength and increase their energy state. This results in an attenuation when light is sent through a sample. The concentration is then calculated from the attenuation according to the Beer-Lambert law:

$$A = -\log_{10} \left[\frac{I}{I_0} \right] = kcl, \quad 2-15$$

where A is the absorbance, which is defined as the negative logarithm of the ratio between initial intensity I_0 and intensity measured after the light has passed through the sample I . c is the concentration of the sample, l is the travel length of the light through the sample and k is a proportionality constant (85).

UV spectroscopy is a popular way to measure the concentration of nucleic acids (86). Since pyrimidine and purine bases have absorption maxima at 260 nm, the concentration of nucleic acids is determined based on absorbance at this wavelength. The traditional way to assess the purity of samples measured this way is by calculating the A_{260}/A_{280} ratio, i.e., the ratio between absorbance at 260 nm and 280 nm (87). This ratio should be around 2.00 for pure DNA and around 1.80 for pure RNA. Lower ratios indicate pollution, as proteins absorb at 280 nm. Another ratio that is used to assess DNA and RNA purity is A_{260}/A_{230} (88).

Another value measured by spectroscopy is optical density (OD) (89). While absorbance is defined as the attenuation due to absorption, OD is defined as attenuation due to absorption and scattering of light. Still, the Beer-Lambert law (Equation 2-15) holds for this kind of measurement if some assumptions are made (89). This spectrophotometry method is commonly used to measure cell concentration in a sample because there is a linear relationship between OD and concentration up to some threshold OD value (19). For *E. coli* a conversion factor of 0.396 can be used from OD to concentration in the linear area (89). Some common wavelengths used to measure bacterial concentration are 480, 540, 600, and 660 nm (19). In this work, OD is measured at 600 nm, shown as OD_{600} .

3 Materials and Methods

The methods described in the first two sections of this chapter (3.1 and 3.2) are also described by Simensen *et al.* (90). The *E. coli* culture described by Simensen *et al.* is the same as the sample referred to as *E. coli* U-lim in this thesis and all measurement results for this culture are the same in these two works.

3.1 Culture conditions and biomass harvest

Three cultures of *E. coli* strain K-12 MG 1655 were grown in 1.5 L Eppendorf NewBrunswick BioFlo 115 bioreactors. All cultures were grown aerobically in an M9 minimum salts media containing 1 mM MgSO₄, 18.7 mM NaCl, 22.0 mM KH₂PO₄, 33.7 mM Na₂HPO₄, 0.2 % trace mineral solution and various concentrations of glucose and NH₄Cl, which are listed in Table 3.1. The pH was constantly kept at pH 7 by automatic titration of 4 M NaOH. The pH probe was calibrated in pre-mixed solutions of pH 4 and pH 7. The dissolved oxygen (DO) level was measured using an oxygen electrode that was calibrated to 0 % by flushing with nitrogen and to 100 % in the fermenter after 30 minutes of 500 ml min⁻¹ airflow and 500 rpm stirring at 37 °C. Exchange gases were sterile filtered with 0.2 µm filters. The organisms were pre-cultured overnight in shake flasks at 37 °C in an M9 medium of the same composition.

One culture (*E. coli* U-lim) was grown in a batch setup. This culture was grown in an unlimited growth medium to capture the unlimited exponential growth. The batch bioreactor was inoculated with a pre-culture to a start OD₆₀₀ of 0.059 and the sample (*E. coli* U-lim) was harvested during exponential growth at an OD₆₀₀ of 6.7. Several supplementary samples were taken at different time points and the OD₆₀₀ of these samples was measured to determine the growth rate.

The remaining two cultures were first cultivated in batch setup and later kept at a stable OD₆₀₀ by a chemostat. Carbon limited medium was used in one of the bioreactors with continuous culture and nitrogen-limited media was used in the other (Table 3.1). Samples were taken from each of the chemostats at three growth rates: 0.1 h⁻¹, 0.2 h⁻¹, and 0.4 h⁻¹. The growth rates were assumed to be equal to the dilution rates (72). As the dilution rates were decreased, less of the limiting nutrient (glucose or ammonium) was added to the media. Between every sampling, the bioreactors were kept at a constant dilution rate until the whole volume of the bioreactor was exchanged three times.

Table 3.1: Glucose and NH₄Cl concentrations in the media used to grow *E. coli* samples in three bioreactors.

Bioreactor	Sample	Growth rate	Glucose conc. [% w/w]	NH ₄ Cl conc. g/L
1	<i>E. coli</i> C-lim 0.4	0.4 h ⁻¹	1.40	5.0
	<i>E. coli</i> C-lim 0.2	0.2 h ⁻¹	0.70	5.0
	<i>E. coli</i> C-lim 0.1	0.1 h ⁻¹	0.35	5.0
2	<i>E. coli</i> N-lim 0.4	0.4 h ⁻¹	2.00	1.0
	<i>E. coli</i> N-lim 0.2	0.2 h ⁻¹	2.00	1.0
	<i>E. coli</i> N-lim 0.1	0.1 h ⁻¹	2.00	1.0
3	<i>E. coli</i> U-lim	0.71 h ⁻¹	2.00	5.0

To collect samples, the fermenter was drained with a 50 mL syringe. The aliquots were centrifuged at 4500 rpm and 4 °C for 5 minutes and washed twice in 0.9 % NaCl solution. Then they were washed once with MQ water. The samples were pooled until each pellet had a volume of approximately 7.5 mL. The pellets were frozen at -20 °C and lyophilised for three days before they were stored in -20 °C.

3.1.1 Exchange rate determination

The glucose uptake rate of *E. coli* U-lim was calculated from the glucose concentration and OD₆₀₀ measurements of samples taken at multiple time points. Glucose concentration was measured by nuclear magnetic resonance (NMR). Media samples of 2.5 mL were taken at multiple time points, lyophilised, and rehydrated in D₂O. The samples were analysed in an NMR spectrometer to create a ¹H spectra and glucose was quantified using the α -glucose doublet. Creatine was used as the external standard. For detailed protocol see Appendix A, for calculations see supplementary material S1.

O₂ and CO₂ exchange rates of *E. coli* U-lim were determined from off-gas analyses and OD measurements (see supplementary material S1). OD₆₀₀ measurements were converted to concentration by multiplying with a conversion factor of 0.396 g/L (89). The off-gas was analysed by an Eppendorf DASGIP GA4 gas analyser.

3.2 Biomass composition analysis

In each method described in this section, three parallels of each sample were made as technical replicates, unless stated otherwise. All methods were previously tested on the sample *E. coli* U-lim and published in (90). Due to lack of sample, DNA content was not measured for *E. coli* C-lim 0.2 and *E. coli* C-lim 0.1. The DNA extraction procedure was also not tested for *S. salar* and *A. thaliana*. The RNA extraction procedure was not tested for *A. Thaliana*.

3.2.1 Protein

To quantify proteins, ~1 mg dried biomass was suspended in 500 μ L 6M HCl in a glass vial. The vials were sealed with caps and the samples were boiled for 24 h at 110 °C. After boiling, the samples were neutralized with the 500 μ L 6M NaOH, and MQ water was added up to 1 ml if there were a visible loss of volume. The samples were filtered using a syringe with a 0.2 polyethersulphone membrane filter. Then the samples were and diluted to a concentration of approximately 0.02 mg biomass/L. 200 μ L were transferred to an HPLC vial and stored at -20 °C until further analysis.

The sample preparation protocol was modified for sample *E. coli* U-lim. Instead of suspending ~1 mg in 500 μ L 6M HCl, ~10 mg of dried biomass was suspended in 5 mL 6M HCl in a glass vial. After the boiling step, the sample was neutralised with 5 mL 6M NaOH. Also, instead of adding water to account for the volume lost during boiling, the volume loss was estimated by weighing the filled vials before and after boiling. Apart from this, the sample preparation procedure of *E. coli* U-lim was identical to that of the remaining samples.

The samples were then analysed by reversed-phase partition HPLC with pre-column O-phthalaldehyde (OPA)-derivatisation. A Waters Nova-Pak C18 4 μ m, column was coupled to an ULTIMATE.3000.WP Injector with two mobile phases: methanol and sodium acetate buffer with tetrahydrofuran (THF). The standard that was used was a premade standard from Sigma Aldrich diluted to 10 nmol/L. An RF2000-DIGITAL detector was used to detect the UV signal. The HPLC analysis was performed by a lab technician.

3.2.2 Carbohydrate

The total carbohydrate content was measured according to the protocol described in Rühmann *et al.* (91) at the TU München. In brief, the samples were hydrolysed by boiling in 4 M trifluoroacetic acid at 121 °C for 90 minutes and then neutralised with a 3.2 % ammonia solution. Then they were derivatised with 1-phenyl-3-methyl-5-pyrazolone (PMP) and incubated for 100 minutes at 70 °C. The samples were then filtered, diluted, and analysed by HPLC with UV detection coupled with ESI-MS.

3.2.3 Lipid

Lipids were extracted from the biomass by liquid-liquid extraction using a chloroform/methanol extraction protocol based on Folch's method (92). ~40 mg of dried cells were weighted out in an Eppendorf tube and rehydrated by vortexing shortly with 0.15 mL of water. 0.5 g zirconium beads and 0.4 mL methanol were added to the samples and then the samples were homogenised in two 20 second intervals at 6,500 rpm with a pause of 2 minutes where the samples were kept on ice. After homogenisation, 0.8 mL chloroform was added and the samples were vortexed for 20 minutes, before 0.1 mL water was added and the sample were vortexed again for 10 minutes. The samples were then centrifuged for 4 minutes in a table centrifuge at 13,400 rpm. The organic phase was collected using a syringe with a needle and transferred to a weighted vial with a lid. 0.6 mL chloroform was added to the beads-cell solution in the Eppendorf tube and the Eppendorf tubes were vortexed for 10 minutes and centrifuged for 4 minutes. Again, the organic phase was collected with a syringe with a needle and transferred to the same glass vial. The glass vials were weighted before the chloroform was evaporated in the fume hood. When the chloroform had evaporated the vials were weighed again and the lipid content was calculated. The potential loss of the organic phase was calculated from the difference in weight before and after chloroform evaporation.

All the steps described above were also performed on a blank sample, which was then subtracted from the other samples. Total lipid content of all samples was measured this way, but due to insufficient amounts of sample, no technical replicates were made for *E. coli* 2, *E. coli* 3, *S. cerevisiae*, *S. salar*, and *A. thaliana*. For the other samples, three technical replicates were used.

3.2.4 DNA

DNA was extracted from the lyophilised biomass by phenol/chloroform extraction as described by Green and Sambrook consisting(93). A lysis buffer was prepared from 9.34 ml Tris-EDTA (TE) buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA pH 8.0), 600 μ L 10 % sodium dodecyl sulphate (SDS) and 60 μ L proteinase K (20 mg/L). ~10 mg of dried cells were suspended in 600 μ L lysis buffer and incubated for 30 minutes at 55 °C. When the samples had cooled down to room temperature, 600 μ L phenol/chloroform (1:1 v/v) was added and the sample was mixed by inverting the tube. The sample was then centrifuged in a table centrifuge at 13,400 rpm. The aqueous phase was transferred to a new tube using a 1 mL micropipette. The extraction process, consisting of chloroform addition, centrifugation, and extraction of the aqueous phase, was repeated twice. Then, 600 μ L chloroform was added to the collected aqueous phases and the samples were centrifuged at 13,400 rpm for 5 minutes. The entire aqueous phase was then transferred to a new tube. The DNA was precipitated from the aqueous phase by addition of 40 μ L 3 M sodium acetate (pH 5.5) and 1 mL ice-cold 99 % ethanol and incubating overnight at -20 °C. Then the samples were centrifuged for 15 minutes at 4 °C at 9,000 rpm. The supernatant was discarded and the pellet was washed by adding 1 mL 70 % ethanol, centrifuging for 2 minutes and discarding the supernatant. The pellet was then resuspended in 50 μ L TE buffer, 1 μ L RNase A was added and the suspension was incubated for 15 minutes at 37 °C. The concentration of dsDNA was measured by UV-visible spectroscopy against a TE-buffer blank using the Thermo Scientific NanoDrop One set to measure dsDNA (94).

3.2.5 RNA

Total cellular RNA was quantified by the protocol described by Benthin *et al.* (95). To degrade the cell walls, the lyophilised biomass (~ 10 mg) was washed three times with 3 mL 0.7 M perchloric acid (HClO₄). The samples were centrifuged for 10 minutes at 4000 rpm at 4 °C and the supernatant was discarded between washes. The samples were resuspended in 0.3 M potassium hydroxide and incubated

at 37 °C in a water bath for 1 h, shaking every 15 minutes to lyse the cells. Then the samples were cooled and 1 mL 3 M HClO₄ was added. The samples were centrifuged for 10 minutes at 4000 rpm at 4 °C and the supernatant was decanted into a new polypropylene centrifuge tube. The pellet was washed twice with 4 mL 0.5 M HClO₄ by resuspending and mixing, centrifuging, and decanting the supernatant into the polypropylene tube. Another 3 mL of 0.5 M HClO₄ was added to the collected supernatants and then they were centrifuged again to remove any precipitates of KClO₄. The concentration of RNA was measured by UV-visible spectroscopy against a 0.5 HClO₄ blank using the Thermo Scientific NanoDrop One set to measure RNA (94).

3.2.6 Elemental analysis

Elemental analysis by combustion was performed by a lab technician at TU München for all *E. coli* samples. The total carbon, hydrogen, nitrogen, oxygen, and sulphur content was determined in all *E. coli* cultures, *S. salar* and *A. thaliana*.

3.3 Biomass objective functions

BOFs were constructed as described in section 2.1.3. Essentially, the separate macromolecule functions were constructed based on measured or estimated monomer distributions. Then they were combined into one biomass function (calculations are shown in supplementary material S2). The maintenance energy term was obtained from *iML1515*, a genome-scale reconstruction of *E. coli* strain K-12 MG1655 (27). The macromolecules were assumed to make up 96.1 % of the cell dry weight and the remaining 3.9 % was accounted for by cofactors, vitamins, and minerals. The distribution of non-macromolecule biomass components was implemented from the *iML1515* model and scaled.

The BOFs were evaluated by implementing them into the *iML1515* model and running FBA. For the BOF constructed for *E. coli* U-lim, constraints corresponding to their growth conditions were applied. For the remaining BOFs, the uptake rate of the limiting nutrient was tweaked until the experimentally determined growth rate was reached and then the predicted exchange rates were evaluated based on literature (see section 4.4 for more details). The model was downloaded from BiGG Models (available at <http://bigg.ucsd.edu/models/iML1515> March 2021) and simulated using Python and the pyCOBRA scientific package (96, 97).

4 Results and Analysis

The workflow of this thesis can be divided into four steps as shown in Figure 4.1. During the cultivation step, media samples were taken to determine growth and exchange rates. The results from these analyses are described in section 4.1. The second step is a comprehensive biomass composition analysis, which is described in section 4.2. In the biomass analysis, each macromolecule group is quantified separately and the results are described in separate subsections (4.2.1-5). The experimentally determined macromolecule composition is then used in the third step – BOF construction, which is described in section 4.3. Finally, the BOFs are implemented into the model *i*ML1515 and evaluated by FBA in section 4.4. The FBA results are compared to the experimentally determined growth and exchange rates.

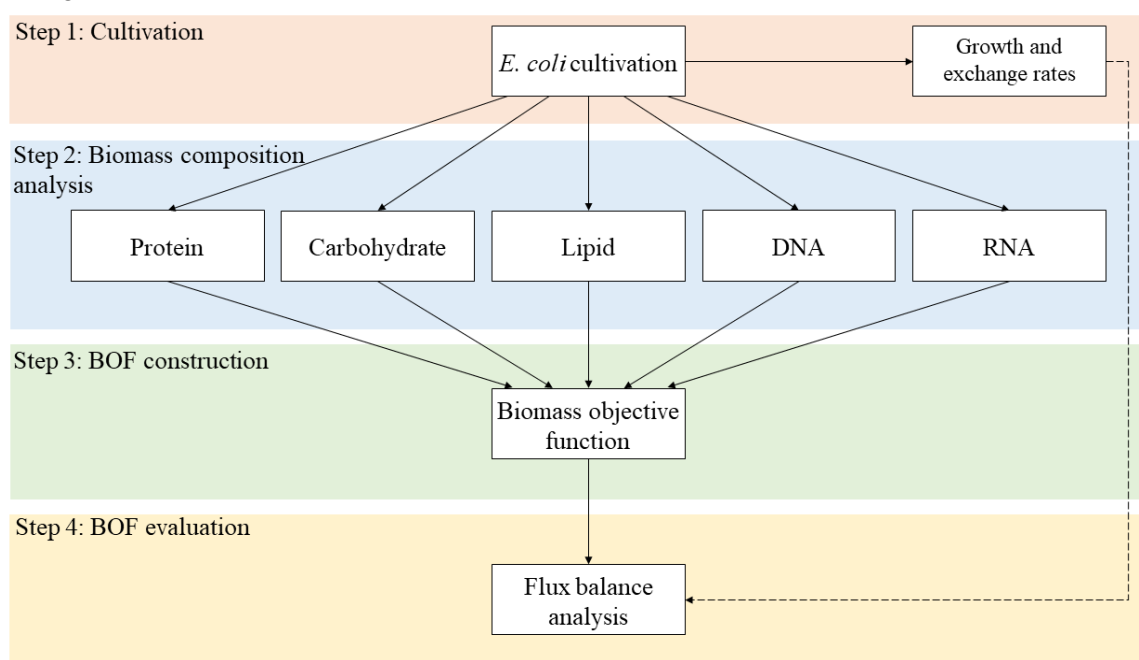


Figure 4.1: Schematic diagram showing the workflow of this thesis.

4.1 Growth and exchange rates

When constructing growth condition-specific BOFs, knowing the growth conditions is of obvious importance. Because bioreactors were used for cultivation, it was possible to monitor some of the culture conditions that would have been unknown for cultures grown in shake flasks. This information and additional information from media analysis are of great value when evaluating the BOFs.

All the *E. coli* samples were cultured by Christian Schulz, Emil Karlsen and Vetle Simensen. During cultivation, media samples were taken at multiple time points to determine OD_{600} . The *E. coli* U-lim samples were analysed with NMR to determine the glucose concentration. Further, O_2 and CO_2 content in the off-gas from *E. coli* U-lim were determined with a gas analyser. All calculated rates in this subsection are based on these measurements.

4.1.1 Growth and exchange rates of *E. coli* U-lim

To determine the growth rate of *E. coli* U-lim, the OD_{600} measurements from the exponential phase were plotted against time. The growth rate was estimated to be 0.71 h^{-1} by fitting the exponential function to the points that were thought to be in the exponential phase (Figure 4.3). See supplementary material S1 for calculations.

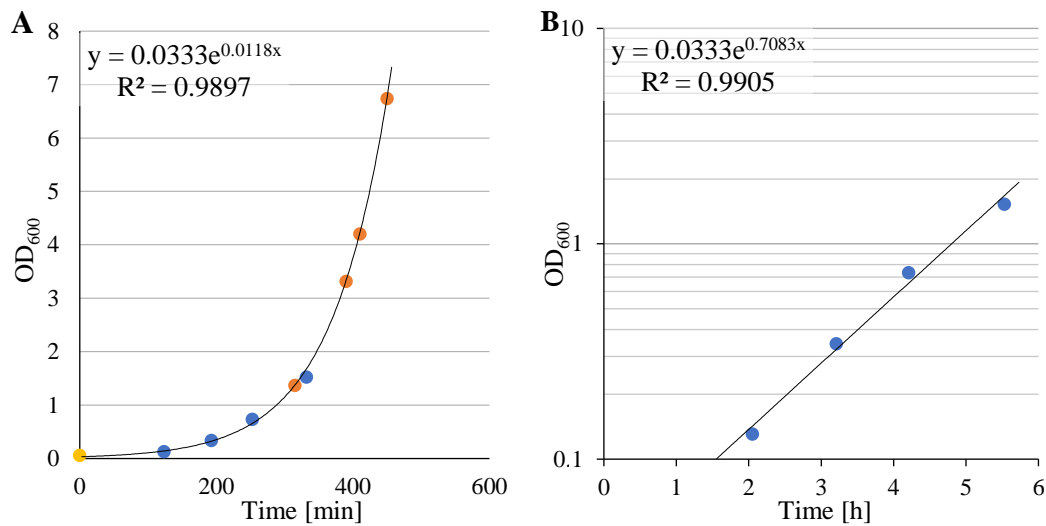


Figure 4.2: Estimation of *E. coli* U-lim growth rate. **A:** OD₆₀₀ measurements plotted against time [minutes]. The equation and R-value of the exponential regression (black line) are shown. Orange points are estimated from the regression line equation and used in the calculation of exchange rates. The yellow point is not considered to be in the exponential phase and is therefore excluded from the regression (B). **B:** OD₆₀₀ values from the exponential phase plotted against time. The OD₆₀₀ measurements are shown on a logarithmic scale. The equation of the regression (black line) and the R-value are shown. The constant in the exponent of the regression equation is the estimated growth rate given in [h⁻¹].

The off-gas analysis showed an exponential increase in CO₂ content over time, and a simultaneous decrease in O₂ content (Figure 4.3). The respiratory quotient (RQ) at various time points was calculated as shown in Equation 4-1 (98):

$$RQ = -\frac{\Delta CO_2}{\Delta O_2}, \quad 4-1$$

where ΔCO_2 and ΔO_2 are the changes in CO₂ and O₂ outflow, respectively, given in mmol/min. RQ is an indicator of the metabolic state of an organism and is commonly used to determine the carbon source of microorganisms. In Figure 4.4, Calculated RQ values are plotted against time. At the time when the *E. coli* U-lim biomass sample was taken (~446 minutes), the RQ was stable around 1, which is expected for bacteria grown aerobically with carbohydrates as the carbon source (98, 99).

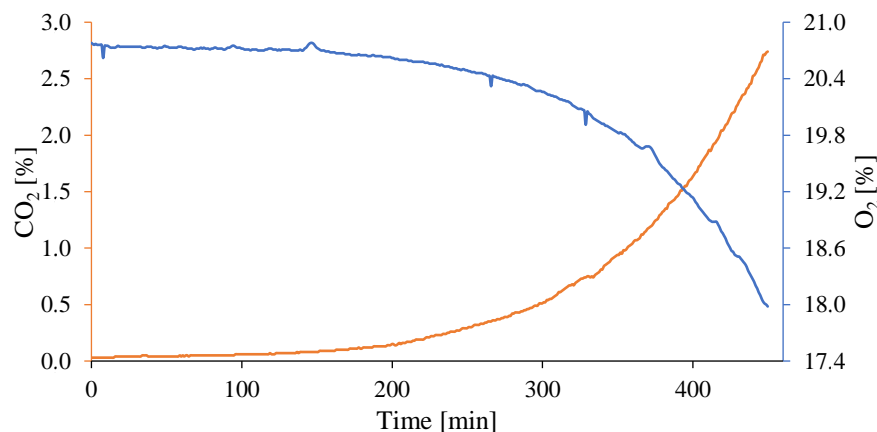


Figure 4.3: Off-gas analysis of *E. coli* U-lim. CO₂ (orange) and O₂ (blue) content in off-gas over time.

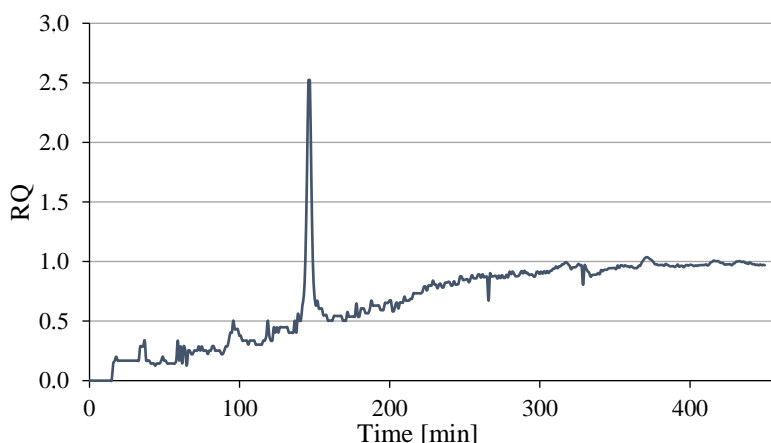


Figure 4.4: RQ of *E. coli* U-lim over time.

The glucose, O₂, and CO₂ exchange rates of *E. coli* U-lim are shown in Table 4.1. The glucose uptake rate is calculated from the glucose concentration measured by NMR at various time points and the estimated amount of cell dry weight at those time points. An attempt was made to measure the acetate, formate, and ethanol concentration by NMR to determine their respective exchange rates. However, as these compounds are volatile, they evaporated during lyophilisation and could not be quantified (data not shown). Cell dry weight was calculated from OD₆₀₀ values that were either directly measured or estimated as shown in Figure 4.2A. A conversion factor of 0.39 was used (89). The oxygen and carbon dioxide exchange rates were calculated from the gas analysis measurements and the estimated dry weight concentrations at multiple time points. For calculations, see supplementary material S1.

Table 4.1: Exchange rates and of *E. coli* U-lim. Negative rates are uptake, positive rates are secretion.

Compound	Rate [$\mu\text{mol/gCDW h}^{-1}$]
Glucose	-9.03
O ₂	-10.09
CO ₂	9.12

4.1.2 Growth rates of *E. coli* cultures grown in chemostat

No measurement of exchange rates was done for the remaining six *E. coli* cultures. The growth rates were determined assuming that the growth rate was equal to the dilution rate. For this assumption to hold, the culture must be in a steady-state (72). Each time concentration of the limiting nutrient in the media is decreased, there is a chance that the cell concentration might drop below steady-state levels. If this happens, the OD will stabilize towards a steady state over time. According to Wides and Milo (72), the difference in OD from the steady-state OD will decrease by a factor of e^{-1} for every chemostat turnover. Since the entire volume of the bioreactors was exchanged three times for each time the nutrient concentration was changed, it is reasonable to assume that the cultures were in a steady state at the time of biomass harvest. Furthermore, media samples were taken at multiple time points and the measured OD values were stable (data not shown).

4.2 Biomass composition analysis

In this section, the results from the second step in our workflow – the biomass composition analysis, will be described (Figure 4.1). The separate macromolecule quantification results will be presented in subsections 4.2.1–4.2.5. In subsection 4.2.6 the complete experimentally determined macromolecular composition and the elemental composition determined by combustion analysis are shown. The biomass compositions described in this section are measured in lyophilised samples that

have been stored at $-20\text{ }^{\circ}\text{C}$. All samples were harvested in early March 2020, but because of the Covid-19 pandemic, the first biomass analyses could not be performed before September 2020. It is uncertain to what degree the prolonged storage has affected the samples.

4.2.1 Protein

In the protein assay, the detected concentrations of glutamine and asparagine were low, suggesting that these amino acids were deaminated during acid hydrolysis and therefore were detected as glutamate and aspartate, respectively (100). Furthermore, glycine and arginine shared a common top in the HPLC spectra (Figure 4.5), and only the sum concentration, not separate concentrations, could be measured. To determine the concentration of the paired amino acids separately, the sum concentration was multiplied by the codon ratio in the genome of *E. coli* strain K-12 MG1655 (GenBank accession number U00096.3 (101)) (see Supplementary material S2 for calculations).

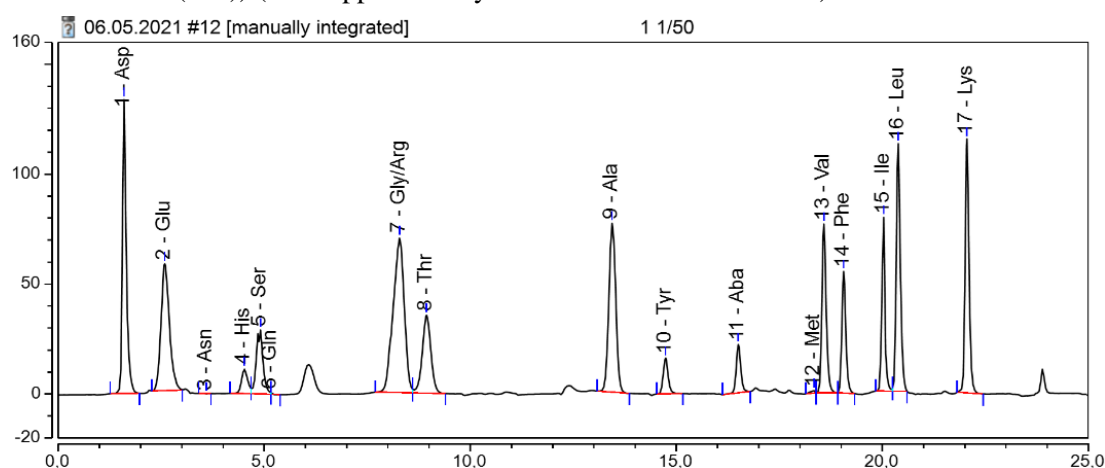


Figure 4.5: HPLC spectra from analysis of *E. coli* U-lim I. Concentrations were determined from the area under each peak by Siri Stavrum. The red lines mark the lower bounds of the peak area.

The concentration of cysteine proline and tryptophan was not measured in the HPLC analysis. Therefore, the concentration of these amino acids was estimated by making a linear regression between the measured amino acid concentrations and the distribution of codons in the genome of *E. coli* strain K-12 MG1655. Since glycine and arginine, glutamine and glutamate, and asparagine and aspartate were measured pairwise one data point per pair was made for these amino acids. The linear regression made for one parallel measurement of *E. coli* U-lim is shown in Figure 4.6. For calculations, see Supplementary material S2.

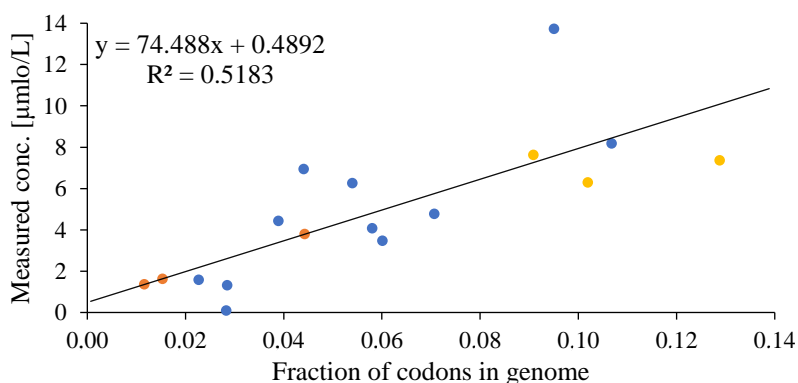


Figure 4.6: Linear regression between the measured amino acid concentrations of *E. coli* U-lim parallel I and the number of codons of the amino acids in the genome of *E. coli* strain K-12 MG1655. The equation of the regression line and the R^2 value are shown. Amino acids that were measured pairwise are marked in yellow, estimated amino acids are marked in orange. See supplementary material S2 for calculations.

The measured amino acid distribution and total protein content of all *E. coli* cultures and *S. cerevisiae*, *S. salar*, and *A. thaliana* are shown in Figure 4.7. No estimations were made for *S. cerevisiae*, *S. salar*, and *A. thaliana*, hence glycine and arginine are shown as one amino acid and cysteine, proline, and tryptophan are not included for these organisms. The exact amino concentrations in all *E. coli* cultures are shown in Appendix B.

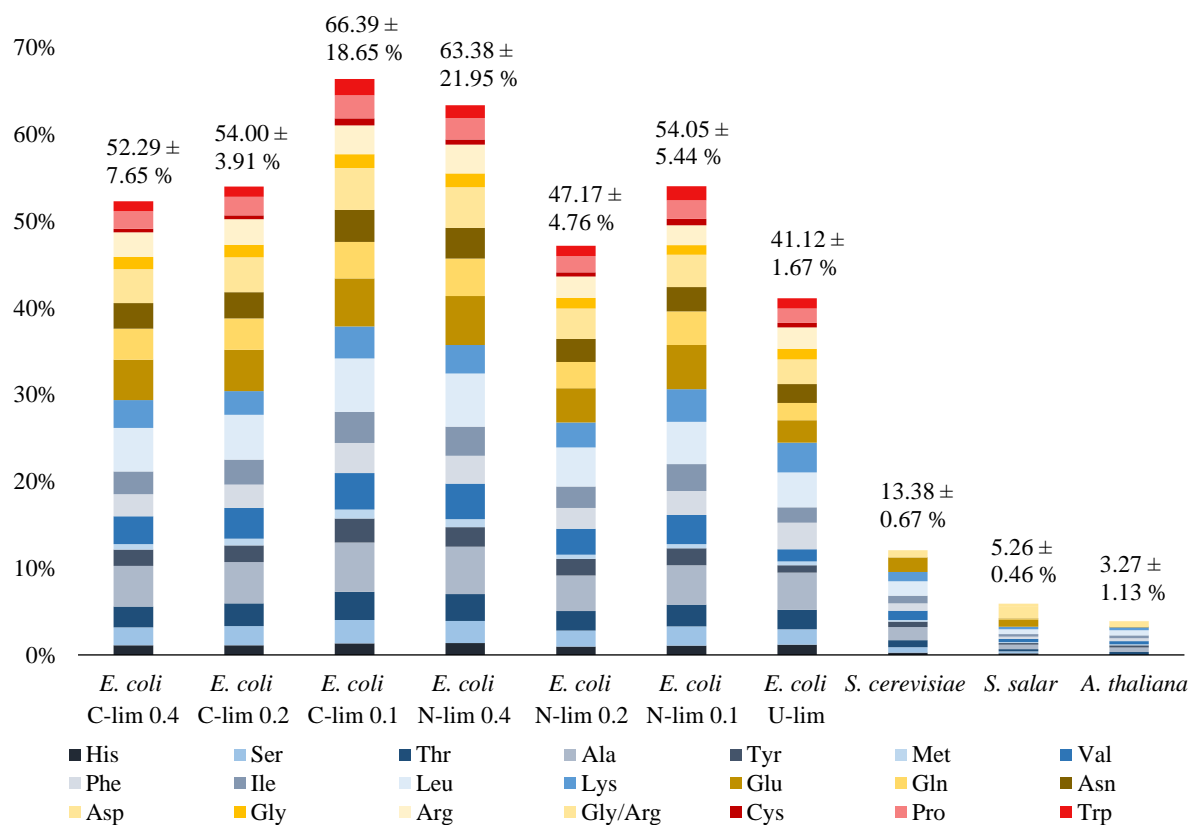


Figure 4.7: Amino acid and protein content of all *E. coli* cultures, *S. cerevisiae*, *S. salar*, and *A. thaliana* measured in mass % of total CDW. The total protein content that is shown is the average and standard deviation of three technical replicates. Directly measured amino acids are blue, pair-wise measured amino acids are yellow, and estimated amino acids are red. Average total protein percentages based on three technical replicates are shown above each column. For exact contents of separate amino acids, see Appendix B or supplementary material S2.

A variation in total protein content was observed between different growth conditions. However, there is no clear pattern between the protein content and the growth rates or the limiting nutrient. The protein content measured in the *S. cerevisiae*, *S. salar*, and *A. thaliana* is much lower than in *E. coli*. For *A. thaliana* this was expected, as the protein content in *A. thaliana* leaves has been reported to be between 1.0 and 2.5 % (102). The protein in *S. cerevisiae* and *S. salar*, on the other hand, were strongly underdetermined compared to previously reported measurements. Between 58 % and 63 % protein have been reported in *S. cerevisiae* (103) and around 76 % in *S. salar* muscle tissue (104). One possible reason for the underdetermined protein content in *S. salar* is that the samples were not homogenised prior to hydrolysis. Both the *S. salar* sample and the *A. thaliana* sample were lyophilised in larger cellular groups, which might have made the protein fraction inaccessible in the quantification.

An issue with the protein quantification was that the veils used during acid hydrolysis were leaking, which led to a loss of volume during the boiling step. The volume that was lost in *E. coli* U-lim was measured by weighting the veils before and after boiling and assuming that the same amount of water and HCl evaporated. One of the parallels was calculated to have a volume loss of 4.1 % while the other two had volume losses of 0.3 and 0.1 % (for calculations see Supplementary material S2). A

volume loss of 4.1 % corresponds to an overdetermination of only 0.08 %, but because the remaining samples were boiled in smaller volumes than *E. coli* U-lim, the volume loss percentage from these samples is assumed to be much greater. To account for visible volume loss, the samples were approximately filled to 1 mL again, using the 1 mL marking on the vials. However, the added water volume is assumed to be inaccurate, hence there would still be a considerable volume loss. A loss in volume during boiling will result in an over-quantification of total amino acid content. This might be one explanation for the high protein contents measured in *E. coli* C-lim 0.1 and *E. coli* N-lim 0.4. The great variation in the amount of volume that leaks out can partly explain the great standard deviations (shown in Figure 4.7) in some of the samples.

In Figure 4.8 the columns representing *E. coli* cultures are normalised to 100%, highlighting the amino acid distribution. The estimated amino acids are not included in Figure 4.8 and the pairwise measured amino acids are shown as sums. The exact relative contents of all amino acids are shown in Appendix B.

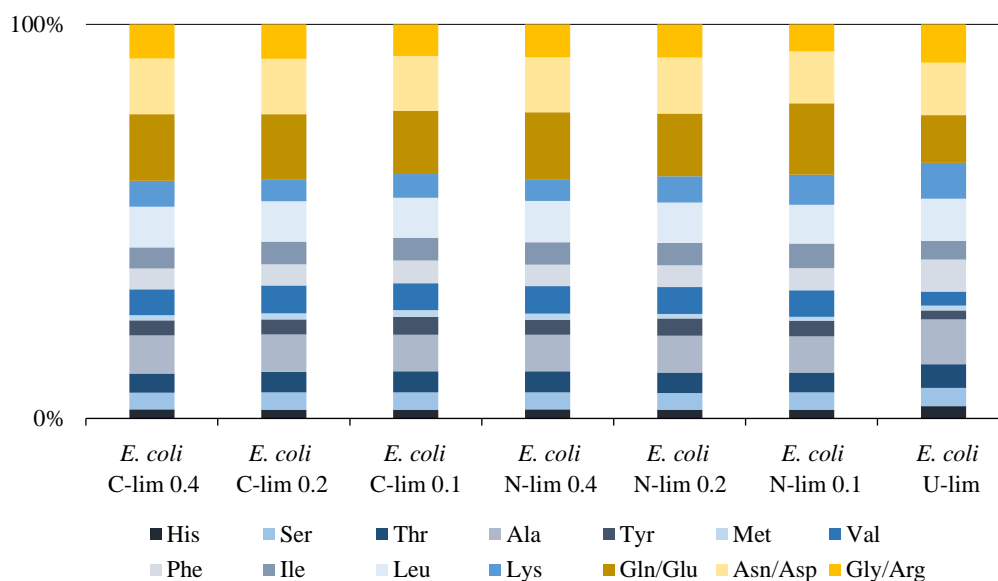


Figure 4.8: Amino acid distribution in all *E. coli* cultures. Separately measured amino acids are blue, pairwise measured amino acids are yellow. Cysteine, proline, and tryptophan are excluded. The shown distributions are based on three technical replicates. See Appendix B for exact relative distribution.

From Figure 4.8 we can see that the amino acid distribution is stable even though the total protein content vary between conditions. The methionine content is very low in all samples. This is possibly because oxygen was not removed from the vials before acid hydrolysis. The presence of oxygen during acid hydrolysis may lead to methionine being oxidised to methionine sulfoxide or methionine sulfone. One way this might had been avoided is if the vials would have been flushed with nitrogen prior to hydrolysis. However, when considering the addition of steps to a procedure the trade-off between precision and efficiency should be assessed.

4.2.2 Carbohydrate

The carbohydrate analysis was done externally at the Technical University of München in Germany. By coupling HPLC to ESI-MS, the sugars could be distinguished not only based on retention time, but also on mass to charge ratio (91). Hence, a large variety of monosaccharides and disaccharides were quantified. The carbohydrates that were measured in the *E. coli* samples, excluding ribose are shown in Figure 4.9.

The carbohydrate content was significantly higher in the samples grown in Nitrogen limitation, than in the other *E. coli* samples. This was expected as *E. coli* tend to accumulate glycogen during

nitrogen-limited growth (60). The carbohydrates in all samples were dominated by glycogen. Still, other sugars make a notable contribution to the total carbohydrate content that would not have been detected in a classic glycogen assay (105). Considerable amounts of rhamnose were measured in *E. coli* N-lim 0.1 and some were also measured in *E. coli* C-lim 0.1. Rhamnose can be found in lipopolysaccharides (LPS) in the outer membrane of *E. coli* (106). The enzymes mediating the incorporation of rhamnose and some other sugars to LPS are highly regulated and differ not only between strains, but also growth conditions (62, 107, 108). Growth rate related variations may be the reason why the rhamnose content in five of the *E. coli* samples was too low to be detected. This can also be the reason why glucuronic acid (GlcUA) was only detected in one sample (*E. coli* N-lim 0.1). Glucosamine and galactose are two components of LPS that were detected in all samples.

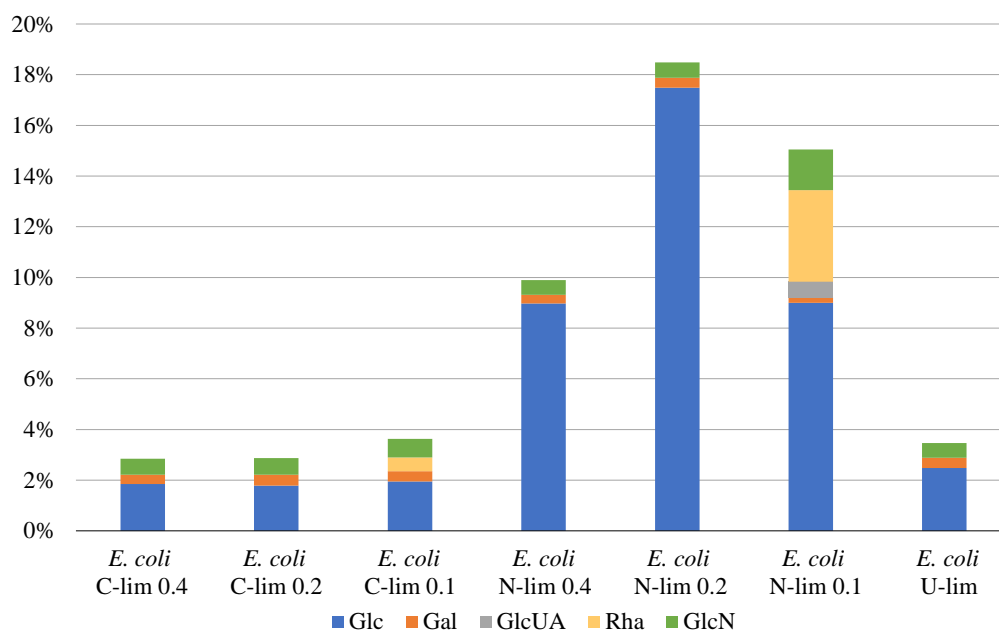


Figure 4.9: Measured carbohydrates in all *E. coli* samples. The averages and standard deviations of three technical replicates are shown. Concentrations are given in mass % of total CDW. The exact values are shown in Table 4.2. Glc = glucose, Gal = galactose, GlcUA = glucuronic acid, Rha = rhamnose, GlcN = glucosamine

The carbohydrate analysis results from all *E. coli* samples are also shown in Table 4.2. The measured carbohydrate content of *S. cerevisiae*, *S. salar*, and *A. thaliana* are also included. In addition to the substances listed in the table, the method also accounted for concentrations of galactosamine, N-acetylgalactosamine, cellobiose, lactose, and gentiobiose, however, these concentrations were too low to be measured in all samples and were therefore assumed to be zero. The ribose content was also measured and is shown in Table 4.2, but since all cellular ribose is assumed to be in RNA, it is not included as a part of the total carbohydrate content of the cells.

No N-acetylglucosamine (GlcNAc) was detected during measurements. This is most likely a result of complete deacetylation during hydrolysis, which would mean that all cellular GlcNAc is detected as glucosamine (GlcN) (109). The GlcNAc and GlcN concentrations provided in Table 4.2 are estimated from the measured GlcN content and the ratio between GlcN and GlcNAc reported by Neidhardt *et al.* (35). For calculations, see Supplementary material S2. The content of N-acetylmuramic acid (MurNAc) was not measured in the carbohydrate assay. Peptidoglycan is composed of equal amounts of GlcNAc and MurNAc, so one can hypothesise that the MurNAc content might be similar to the content of GlcNAc (35). However, since the GlcNAc content was not measured directly, and because some GlcNAc may originate from LPS, I decided to not estimate MurNAc this way (62, 110). Hence, MurNAc is not included in the BOFs constructed in this thesis.

Table 4.2: Carbohydrate fingerprint and total carbohydrate composition. Concentrations are given in average mass % of CDW based on three technical replicates. Glc = glucose, Man = mannose, GlcUA = glucosuronic acid, GalUA = galacturonic acid, Rha = rhamnose, Fuc = fucose, GlcN = glucosamine, GlcNAc = N-Acetylglucosamine, Xyl/Ara = xylose and arabinose, Rib = ribose. Standard deviations of monomers are not included in this table due to lack of space, see supplementary material S2.

Sample	Glc	Man	Gal	GlcUA	GalUA	Rha	Fuc	GlcN	GlcNAc	Xyl/ Ara	Total	Rib
<i>E. coli</i> C-lim 0.4	1.85	0	0.37	0	0	0	0	0.23	0.41	0	2.85 ± 0.09	5.67
<i>E. coli</i> C-lim 0.2	1.78	0	0.43	0	0	0	0	0.23	0.42	0	2.87 ± 0.11	3.23
<i>E. coli</i> C-lim 0.1	1.95	0	0.40	0	0	0.55	0	0.26	0.46	0	3.63 ± 0.24	3.25
<i>E. coli</i> N-lim 0.4	8.97	0	0.35	0	0	0	0	0.21	0.37	0	9.90 ± 0.33	5.13
<i>E. coli</i> N-lim 0.2	17.5	0	0.40	0	0	0	0	0.22	0.38	0	18.5 ± 0.5	4.03
<i>E. coli</i> N-lim 0.1	9.00	0	0.20	0.63	0	3.62	0	0.58	1.02	0	15.05 ± 0.50	2.10
<i>E. coli</i> U-lim	2.48	0	0.40	0	0	0	0	0.21	0.37	0	3.47 ± 0.11	5.75
<i>S. cerevisiae</i>	3.40	9.64	0	0	0	0	0	0.07	0.13	0	13.24 ± 1.18	0
<i>S. salar</i>	0.15	0	0	0	0	0	0	0	0	0	0.15 ± 0.04	0.69
<i>A. thaliana</i>	29.6	0.94	4.48	0	4.12	1.82	0.71	0	0	5.61	47.2 ± 2.1	0

From the standard deviations shown in Table 4.2, we can see that the measuring method has high precision, both for *E. coli* samples and the remaining organisms. On the other hand, ribose has not been detected in *S. cerevisiae* and *A. thaliana*, implying that not all carbohydrates have been quantified in the eucaryotes.

The carbohydrate content was expected to be low in *S. salar* muscle as this tissue mostly consists of proteins and lipids (104). The total carbohydrate content of *S. cerevisiae* is underdetermined compared to previously reported values (43.6 %) (111). Furthermore, no ribose was detected in *S. cerevisiae* and *A. thaliana* indicating that the carbohydrate content of these organisms is underdetermined. On the other hand, *A. thaliana* leaves have been reported to contain 22.44 % carbohydrates, which is significantly less than what was measured in our sample (112). However, carbohydrate content in *A. thaliana* leaves varies greatly with growth conditions (113).

In the carbohydrate analysis used for this thesis, carbohydrates are marked with 1-phenyl-3-methyl-5-pyrazolone (PMP) pre-column (91). Because PMP reacts with the aldehyde group of reducing sugars, it will not mark ketoses (114). Hence ketoses will remain undetected in this carbohydrate assay. In bacteria, ketoses are mainly found in exopolysaccharides, and during planktonic growth they will not contribute much to total carbohydrate content (115, 116). Still, it is worth considering this when measuring other organisms known to produce ketoses, e.g., plants (117).

4.2.3 Lipid

An attempt was made to determine the lipid class distribution of sample *E. coli* U-lim by resuspending the dried lipids in dichloromethane and isopropyl alcohol (2:1 v/v) before analysing them using ultrahigh performance super-critical fluid chromatography with mass spectrometry (UHPSFC-MS), as described by Bartosova et al. (118). However, the results could not be used since the lipids were partly hydrolysed and oxidised. Significant fatty acid oxidation and lipid hydrolysis occur during prolonged storage at -20 °C (119). The hydrolysis is potentially caused by enzymatic activity, as some bacterial lipases are active in frozen media (120). The 6 months storage at -20 °C is considered the most likely reason why the results from the lipid class assay could not be used.

Since lipid class and fatty acid distributions were not measured, they were estimated from previously reported values. Temperature is one of the environmental factors that have the strongest impact on both lipid class, and fatty acid distribution (64, 121). The lipid class and fatty acid distributions reported by Neidhardt *et al.* (35) are measured in *E. coli* grow at the same temperature as our cultures, so these distributions could be used for estimation. According to Neidhardt *et al.*, the lipids of an average

E. coli cell are 75 % phosphatidyl-ethanolamine, 18 % phosphatidylglycerol, 5 % cardiolipin, and traces of phosphatidylserine. The fatty acid composition is 45 % palmitic (16:0) acid, 33 % palmitoleic (16:1) acid, and 25% *cis*-vaccenic (18:1) acid (35). The content of each lipid class in the *E. coli* samples, estimated based on these values, are shown in Table 4.3: Total lipid content and percentage loss of organic phase of all *E. coli* cultures, *S. cerevisiae*, *S. salar*, and *A. thaliana* measured in mass % of total CDW and estimated lipid class content of *E. coli*. Red values are measurements without technical replicates, the remaining values are based on three technical replicates. along with the total lipid content, which was measured directly. The lipid content measured in *S. cerevisiae*, *S. salar*, and *A. thaliana* is also shown.

Table 4.3: Total lipid content and percentage loss of organic phase of all *E. coli* cultures, *S. cerevisiae*, *S. salar*, and *A. thaliana* measured in mass % of total CDW and estimated lipid class content of *E. coli*. Red values are measurements without technical replicates, the remaining values are based on three technical replicates. The percentage loss is given in volumetric %. PE = phosphatidylethanolamine, PG = phosphatidylglycerol, CL = cardiolipin. See Supplementary material S2 for all measurements.

Sample	PE	PG	CL	Total	Loss of organic phase
<i>E. coli</i> C-lim 0.4	4.5	1.1	0.3	5.91 ± 0.16	9.63 %
<i>E. coli</i> C-lim 0.2	2.1	0.5	0.1	2.75	10.3 %
<i>E. coli</i> C-lim 0.1	4.7	1.1	0.3	6.11	8.94 %
<i>E. coli</i> N-lim 0.4	4.5	1.1	0.3	5.90 ± 0.26	14.0 %
<i>E. coli</i> N-lim 0.2	4.0	1.0	0.3	5.17 ± 0.34	12.0 %
<i>E. coli</i> N-lim 0.1	3.8	0.9	0.3	4.94 ± 2.80	11.6 %
<i>E. coli</i> U-lim	4.1	1.0	0.3	6.32 ± 0.21	11.4 %
<i>S. cerevisiae</i>	-	-	-	1.64	16.0 %
<i>S. salar</i>	-	-	-	22.7	10.4 %
<i>A. thaliana</i>	-	-	-	2.80	40.6 %

The total lipid content measured in *E. coli* is relatively stable between all cultures, except *E. coli* C-lim 0.2, which has significantly lower lipid content. As this measurement was done without technical replicates, it is hard to say if the lipid content in the culture was low or if it is an outlier measurement. From the standard deviations, we can see that the results have been stable between technical replicates for all cultures except *E. coli* N-lim 0.1.

To assess the amount of organic phase that might have been lost in the extraction steps of the lipid quantification, the samples were weighed both before and after chloroform evaporation (see subsection 3.2.3). The loss percentage of the organic phase is shown in Table 4.3 (see Supplementary material S2 for calculations). There are two causes for loss of organic phase: inadequate extraction and chloroform evaporation. Because lipids are only lost in inadequate extraction, and not in chloroform evaporation, it would be inaccurate to scale the lipid content up by the loss percentage. Therefore, the directly measured values for total lipid content (marked in bold in Table 4.3) are used in the construction of BOFs.

Based on the scale of the loss of organic phase it is reasonable to assume that some lipids have been lost in extraction in all samples. This can be a part of the explanation for why our lipid measurements are lower than the content reported by Neidhardt *et al.* (9.1 %) (35). On the other hand, the lipid content of *S. cerevisiae*, *S. salar*, and *A. thaliana* does not seem to have been underdetermined. Although the organic phase loss was high in the *A. thaliana* sample, the measured lipid content was higher than previously reported values (< 1.0 %) (122). The measured lipid content of both *S. cerevisiae* and *S. salar* concur with previously reported values (1.21 % and 21.03 %, respectively) (104, 123). It should be noted, however, that the lipid content in these organisms was measured without technical replicates.

4.2.4 DNA

The genome of *E. coli* strain K-12 MG1655 has 50.8 % GC and 49.2 % AT (GenBank accession number U00096.3 (101)). The mass percentages of each nucleotide listed in Table 4.4 are calculated from this distribution and the total DNA content that was measured. For calculations, see Supplementary material S2. The total DNA contents of *E. coli* C-lim 0.2 and *E. coli* C-lim 0.1 were not measured because we did not have enough dried biomass of these samples. The DNA and nucleotide content of these two cultures that are shown in Table 4.4: DNA content and estimated composition given in mass % of CDW. are the concentrations used when constructing BOFs for these two cultures and were estimated as the average of all *E. coli* cultures.

Table 4.4: DNA content and estimated composition given in mass % of CDW. Red values were estimated from the average of all the measured values for *E. coli*. The average A_{260}/A_{280} ratio and A_{260}/A_{230} ratio are included for evaluation of purity. The total DNA content that is shown is the average of three technical replicates that were measured three times each. See Supplementary material S2 for all measurements.

Sample	dAMP	dCMP	dGMP	dTMP	Total	A_{260}/A_{280}	A_{260}/A_{230}
<i>E. coli</i> C-lim 0.4	1.61	1.54	1.74	1.56	6.45 ± 2.29	1.88	1.79
<i>E. coli</i> C-lim 0.2	1.02	0.98	1.10	0.99	4.09	-	-
<i>E. coli</i> C-lim 0.1	1.02	0.98	1.10	0.99	4.09	-	-
<i>E. coli</i> N-lim 0.4	1.33	1.28	1.44	1.29	5.34 ± 1.28	1.89	1.68
<i>E. coli</i> N-lim 0.2	1.49	1.43	1.61	1.45	5.97 ± 1.95	1.80	1.41
<i>E. coli</i> N-lim 0.1	0.33	0.32	0.36	0.32	1.34 ± 0.27	1.94	1.46
<i>E. coli</i> U-lim	0.33	0.32	0.36	0.32	1.33 ± 0.32	2.12	1.25
<i>S. cerevisiae</i>	-	-	-	-	1.41 ± 0.45	2.14	2.18

To assess the purity of the extracted DNA samples, A_{260}/A_{280} and A_{260}/A_{230} ratios are included in Table 4.4. Pure DNA has an A_{260}/A_{280} ratio of ~ 1.8 (124). A high A_{260}/A_{280} ratio can be caused by the presence of RNA since uracil has a higher A_{260}/A_{280} ratio than thymine (125). However, the presence of RNA is unlikely since the samples were treated with RNase (see section 3.2.4). A more likely explanation for the increased A_{260}/A_{280} is that the DNA was suspended in TE buffer, which is slightly basic. According to Willfinger *et al.* (126), the A_{260}/A_{280} ratio can be overrepresented by up to 0.2-0.3 points in basic solutions. High A_{260}/A_{280} ratios do not necessarily mean that there is a problem with the sample (127). Since tryptophan and tyrosine sidechains have high absorbance at 280 nm, high A_{260}/A_{280} ratios can indicate low contamination by proteins.

The A_{260}/A_{230} ratio is another indicator of protein contamination because peptide bonds absorb at 230nm. Nucleic acids have an absorbance minimum at 230nm, thus low A_{260}/A_{230} ratios can be an indication of contamination by substances with high absorbance at 230 nm (124). Out of the six samples we measured, five have an A_{260}/A_{230} ratio < 2.0 , which is considered an indication of pollution(128). Since protein contamination would result in low A_{260}/A_{280} ratios, an alternative cause for the low A_{260}/A_{230} ratios is phenol contamination from the phenol-chloroform extraction (see section 3.2.4)(124). As phenol absorbs light at 260 nm, phenol pollution can cause overdetermined concentration (129).

DNA content in *E. coli* was expected to remain stable with varying growth rates (130). The variations we can see in Table 4.4 might be a result of inadequate extraction, overdetermination due to pollution, or both. One way to assess the accuracy of the DNA quantifications would be to spike the samples with a DNA standard (131). Another way to improve DNA measurements would be to apply a quantification method without any extraction step, e.g., the method suggested by Sandaa *et al.* (132) for quantification of DNA in soil. In this method, the DNA is marked with PicoGreen, a fluorescent dye that binds specifically to double-stranded DNA and quantified in a fluorometer.

4.2.5 RNA

According to Neidhardt *et al.* (35), ribosomal RNA make up around 81 % of the cellular RNA in *E. coli*. Therefore, the concentration of each ribonucleotide was estimated based on the nucleobase distribution in ribosomal RNA, as proposed by Beck *et al.* (33). For this, the DNA sequences of S5, S16, and S23 RNA of *E. coli* strain K-12 (available at European Nucleotide Archive (ENA)) were used. The combined sequence of S5, S16, and S23 RNA consists of 25.7 % AMP, 22.4 % CMP, 31.5 % GMP, and 20.3 % UMP. The nucleotide concentrations in Table 4.5 are estimated from this distribution and the total measured RNA contents. For calculation see Supplementary material S2. A_{260}/A_{280} and A_{260}/A_{280} ratios were not noted during RNA quantification.

Table 4.5: RNA content of all *E. coli* cultures, *S. cerevisiae*, and *S. salar*, and estimated nucleotide composition in *E. coli*. All compositions are given as mass % of CDW. The total RNA content that is shown is the average of three technical replicates that were measured three times each. See Supplementary material S2 for all measurements.

Sample	AMP	CMP	GMP	UMP	Total
<i>E. coli</i> C-lim 0.4	4.58	3.72	5.86	3.38	17.5 ± 0.6
<i>E. coli</i> C-lim 0.2	3.09	2.51	3.96	2.29	11.9 ± 0.8
<i>E. coli</i> C-lim 0.1	3.10	2.52	3.97	2.29	11.9 ± 0.4
<i>E. coli</i> N-lim 0.4	4.46	3.62	5.71	3.29	17.1 ± 0.7
<i>E. coli</i> N-lim 0.2	3.65	2.96	4.67	2.69	14.0 ± 0.4
<i>E. coli</i> N-lim 0.1	2.29	1.86	2.93	1.69	8.76 ± 1.14
<i>E. coli</i> U-lim	4.96	4.03	6.36	3.66	19.0 ± 0.5
<i>S. cerevisiae</i>	-	-	-	-	3.12 ± 0.14
<i>S. salar</i>	-	-	-	-	0.45 ± 0.03

From the standard deviations in Table 4.5, we can see that the precision of the measuring method is good. The accuracy of the method could have been assessed by spiking the samples with a standard, as suggested for the DNA quantification (131).

Another way to evaluate our method is by comparison to results from other measuring methods. Ribose content was measured for all *E. coli* cultures as a part of the carbohydrate quantification (Subsection 3.2.2). The total ribose content can also be calculated from the total RNA content and nucleotide distribution. Ribose content measured as carbohydrates and as RNA is compared in Figure 4.10.

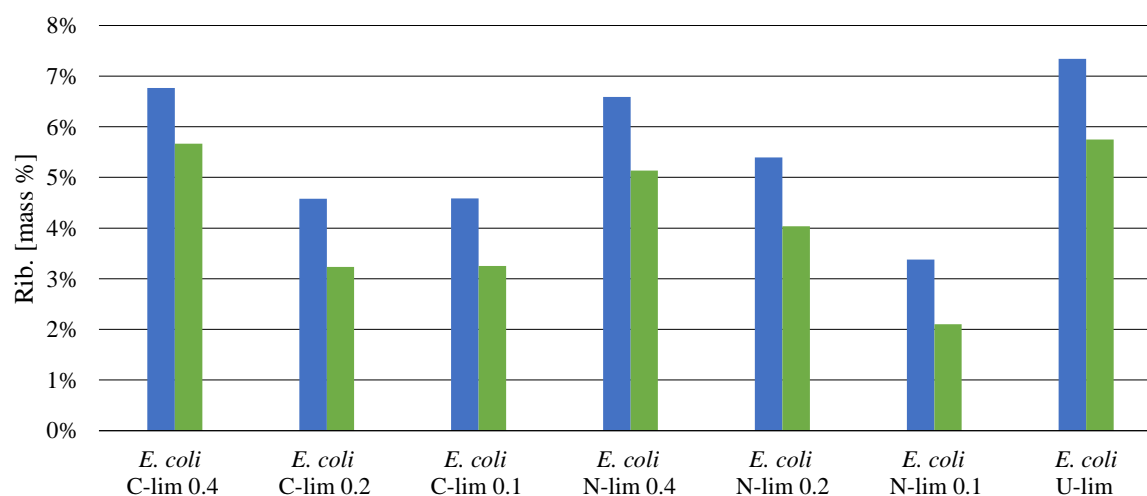


Figure 4.10: Comparison of ribose content in all *E. coli* cultures measured as carbohydrates and RNA. Blue = measured as RNA, green = measured as carbohydrates.

The ribose concentration measured as carbohydrates is consequently lower than the concentration measured as RNA. Still, there is a clear correlation between the measured concentrations (see Appendix C). This suggests that either the ribose content is consequently underdetermined in the carbohydrate assay, or the RNA content is consequently overdetermined in the RNA quantification. Overdetermination of RNA could potentially have been caused by protein pollution in the extracted samples.

Previously reported measurements of RNA content in *E. coli* suggest a linear relationship between the RNA/protein ratio and the specific growth rate (69). Measurements on *E. coli* grown under limitation of various nutrients show that this linear relationship is applicable for both carbon and nitrogen-limited cultures (56). Figure 4.11 shows a plot of RNA/proteins ratios, calculated from measured RNA and protein content, at different growth rates. We can observe a clear linear correlation between RNA/protein ratio and growth rate in the samples.

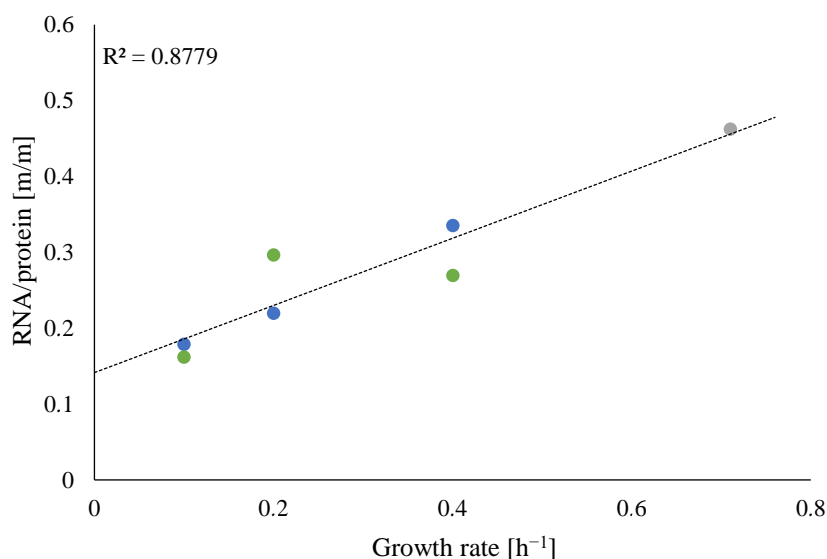


Figure 4.11: Measured RNA/protein ratio in *E. coli* at various growth rates. Carbon limited samples are represented by blue points, nitrogen-limited samples are green, and the grey point represents *E. coli* U-lim.

The estimation method used to quantify separate ribonucleotides relies on the ribonucleotide distribution of rRNA. Because the nucleotide sequence and hence distribution, is constant and equal in all ribosomes, this is a reliable method to determine the distribution in rRNA (133). On the other hand, the method only accounts for 81 % of the total cellular RNA (35). A better way to determine the ribonucleotide distribution would be to measure it directly, for example, by digesting isolated RNA and then quantifying nucleosides by liquid chromatography tandem mass spectrometry, as suggested by Thüring *et al.* (134).

The RNA content measured in *S. salar* is close to previously reported values (0.71 %) (135). Ertugay and Hamamci have reported between 0.93 % and 1.13 % RNA in *S. cerevisiae* (136), which is considerably lower than what we measured. This does not necessarily mean that the RNA is overdetermined in the *S. cerevisiae* sample, as RNA production in yeast is strongly dependent on growth conditions (137).

4.2.6 Total biomass composition

All *E. coli* samples were sent to the Technical University of München in Germany for elemental analysis by combustion. The determined concentration of carbon, hydrogen, nitrogen, oxygen, and sulphur in all is shown in Table 4.6. It was reported that there was some trouble with the nitrogen measurement in some samples resulting in low measurements of *E. coli* C-lim 0.1, *E. coli* N-lim 0.2, *E.*

coli N-lim 0.1, *S. salar*, and *A. thaliana*. Three technical replicates were made for each sample, however, there was some trouble with some of the parallel measurements, thus not all values given in Table 4.6 are based on three parallel measurements.

Table 4.6: Elemental composition of *E. coli*, *S. salar*, and *A. thaliana* measured by combustion analysis. The number of parallel measurements used to determine the average and standard deviation are shown in the right-most column. The nitrogen measurements that were affected by trouble during the analyses are highlighted in red.

Sample	C	H	N	S	O	Total	# of parallels
<i>E. coli</i> C-lim 0.4	43.279 ± 0.304	6.468 ± 0.106	12.747 ± 0.173	0.309 ± 0.006	26.230 ± 1.495	89.032 ± 2.085	2
<i>E. coli</i> C-lim 0.2	45.496 ± 0.899	6.734 ± 0.240	12.862 ± 0.201	0.358 ± 0.008	24.273 ± 1.853	89.724 ± 3.201	3
<i>E. coli</i> C-lim 0.1	44.599 ± 0.299	6.835 ± 0.012	2.561 ± 0.614	0.320 ± 0.017	24.701 ± 1.888	79.015 ± 2.830	3
<i>E. coli</i> N-lim 0.4	43.863	6.589	12.921	0.318	27.526	91.217	1
<i>E. coli</i> N-lim 0.2	41.886 ± 0.241	6.521 ± 0.049	1.086 ± 0.656	0.237 ± 0.037	28.814 ± 1.333	78.543 ± 2.316	2
<i>E. coli</i> N-lim 0.1	41.706 ± 0.315	6.323 ± 0.111	3.147 ± 1.199	0.200 ± 0.070	27.633 ± 1.965	79.009 ± 3.661	3
<i>E. coli</i> U-lim	43.978 ± 0.011	6.708 ± 0.022	14.094 ± 0.378	0.323 ± 0.015	25.969 ± 0.875	91.070 ± 1.301	2
<i>S. salar</i>	47.159 ± 4.697	8.012 ± 0.707	0.000 ± 0.000	0.089 ± 0.086	22.474 ± 1.641	77.733 ± 7.132	3
<i>A. thaliana</i>	31.164 ± 0.936	5.673 ± 0.041	0.000 ± 0.000	0.000 ± 0.000	49.624 ± 1.311	86.461 ± 2.288	3

The total macromolecular composition measured in all *E. coli* cultures, *S. cerevisiae*, *S. salar*, and *A. thaliana* is shown in Table 4.7. The grey numbers for DNA content are not measured, but rather estimated as the average of the remaining *E. coli* samples. These numbers are included in the results because they were used for the construction of BOFs.

Table 4.7: Macromolecular compositions of all *E. coli* cultures, *S. cerevisiae*, *S. salar*, and *A. thaliana* given in mass % of CDW. Red numbers are measured without technical replicates. Grey values are estimated as the average of the DNA content measured in *E. coli* samples. The remaining values are based on three technical replicates.

Sample	Protein	Carbohydrate	Lipid	DNA	RNA	Total
<i>E. coli</i> C-lim 0.4	52.29 ± 7.65	2.58 ± 0.09	5.91 ± 0.16	6.45 ± 2.29	17.5 ± 0.6	85.0 ± 10.8
<i>E. coli</i> C-lim 0.2	54.00 ± 3.91	2.87 ± 0.11	2.75	4.09	11.9 ± 0.8	75.6 ± 4.8
<i>E. coli</i> C-lim 0.1	66.39 ± 18.65	3.63 ± 0.24	6.11	4.09	11.9 ± 0.4	92.1 ± 19.3
<i>E. coli</i> N-lim 0.4	63.38 ± 21.95	9.90 ± 0.33	5.90 ± 0.26	5.34 ± 1.28	17.1 ± 0.7	101.6 ± 24.5
<i>E. coli</i> N-lim 0.2	47.17 ± 4.76	18.5 ± 0.5	5.17 ± 0.34	5.97 ± 1.95	14.0 ± 0.4	90.8 ± 8.0
<i>E. coli</i> N-lim 0.1	54.05 ± 5.44	15.05 ± 0.50	4.94 ± 2.80	1.34 ± 0.27	8.76 ± 1.14	84.1 ± 10.1
<i>E. coli</i> U-lim	41.12 ± 1.67	3.47 ± 0.11	5.32 ± 0.21	1.33 ± 0.32	19.0 ± 0.5	70.3 ± 2.8
<i>S. cerevisiae</i>	13.38 ± 0.67	13.24 ± 1.18	1.64	1.41 ± 0.45	3.12 ± 0.14	32.8 ± 2.4
<i>S. salar</i>	5.26 ± 0.46	0.15 ± 0.04	22.7	-	0.45 ± 0.03	28.6 ± 0.5
<i>A. thaliana</i>	3.27 ± 1.13	47.2 ± 2.1	2.80	-	-	53.3 ± 3.2

The BOFs constructed in this thesis are defined to take in all the components (measured in mmol) needed to produce 1 g of dry biomass. To make this kind of BOF we need to assume that we have measured the entire biomass of the organisms. In this thesis this is done by scaling all macromolecules by the same factor so that they sum up to 96.5% of total CDW, that is 0.96 g of biomass (Table 4.8). The remaining 3.9 % are assumed to be various metabolites, vitamins, cofactors, and inorganic ions as reported by Neidhardt *et al.* (35).

Table 4.8: Macromolecular composition of *E. coli*. Values are given in mass % of total CDW and scaled to make up a total of 96.5 % of the CDW. Red numbers are based on measurements without technical replicates. Gray values are based on the average DNA content measured in *E. coli* samples.

Sample	Protein	Carbohydrate	Lipid	DNA	RNA	Total
<i>E. coli</i> C-lim 0.4	59.1	3.22	6.68	7.29	19.81	96.1
<i>E. coli</i> C-lim 0.2	68.7	3.65	3.50	5.20	15.08	96.1
<i>E. coli</i> C-lim 0.1	69.3	3.78	6.38	4.26	12.40	96.1
<i>E. coli</i> N-lim 0.4	50.0	9.36	5.58	5.05	16.15	96.1
<i>E. coli</i> N-lim 0.2	49.9	19.6	5.47	6.32	14.80	96.1
<i>E. coli</i> N-lim 0.1	61.7	17.2	5.64	1.53	10.01	96.1
<i>E. coli</i> U-lim	56.3	4.74	7.27	1.82	26.01	96.1

One issue with scaling the macromolecule contents this way is that the upscaling will not be proportional to the relative loss in each group. For example, the proteins, that might be overdetermined already will be scaled by the same factor as carbohydrates, which most likely are underdetermined. However, as we did not know the relative loss for the macromolecules, this was considered the most logical way to scale the macromolecule content.

4.3 Biomass objective functions

After measuring the complete macromolecular composition of the *E. coli* samples, we could proceed to the third step of our workflow – construction of condition-specific BOFs (Figure 4.1). The macromolecule component of our BOFs was constructed based on the method used by Beck *et al.* (33) (described in subsection 2.1.3). The macromolecule function equations and chemical formulas of *E. coli* U-lim are shown in Table 4.9. The compositions of lipids, DNA, and RNA were not directly measured, and the functions and chemical formulas of these macromolecules are equal for all *E. coli* samples. Protein and carbohydrate functions and chemical formulas for all *E. coli* samples are shown in Appendix D. All macromolecular functions were balanced for mass and charge to make sure that the mass balance constraint of stoichiometric modelling was satisfied. For calculations see Supplementary material S2.

Table 4.9: Macromolecule function equations and chemical formulas of *E. coli* U-lim. Lipid, DNA, and RNA functions and formulas are the same for all the *E. coli* samples.

Molecule	Equation	Chemical formula
Protein	0.023 His + 0.053 Ser + 0.069 Thr + 0.157 Ala + 0.015 Tyr + 0.009 Met + 0.035 Val + 0.055 Phe + 0.041 Ile + 0.092 Leu + 0.070 Lys + 0.053 Glu + 0.041 Gln + 0.050 Asn + 0.065 Asp + 0.055 Gly + 0.041 Arg + 0.013 Cys + 0.045 Pro + 0.017 Trp + 2 GTP + ATP → H ₂ O + 2 GTP + 2 P _i + AMP + PP _i + protein	C _{4.85} H _{7.60} N _{1.32} O _{1.45} S _{0.02}
Carb.	0.734 Glc1P + 0.118 Gal1P + 0.005 GlcN1P + 0.007 GlcNAc1P + ATP → ADP + PP _i + carbohydrate	C _{6.2} H _{10.4} N _{0.1} O _{4.9}
Lipid	0.92 palmitic acid + 0.68 palmitoleic acid + 0.46 <i>cis</i> -vaccenic acid + 2.05 ATP + 1.26 G3P + 1.03 CTP + 0.79 Ser + 0.24 H ₂ O + 0.79 H ⁺ → 2.05 AMP + 3.08 PP _i + 1.03 CMP + 0.79 CO ₂ + P _i + 0.03 glycerol + lipid	C _{39.1} H _{76.2} N _{0.8} O _{8.6} P _{1.0}
DNA	0.246 dATP + 0.254 dCTP + 0.254 dGTP + 0.246 dTTP → PP _i + DNA	C _{9.7} H _{11.2} N _{3.8} O ₆ P
RNA	0.257 ATP + 0.224 CTP + 0.315 GTP + 0.203 UTP → PP _i + RNA	C _{9.6} H _{10.8} N _{3.9} O _{6.9} P

Instead of scaling the macromolecule coefficients to 100 %, like Beck *et al.* (33), our coefficients were scaled to 96.1 % (Table 4.8). To account for the remaining 3.9 % of the total CDW, all metabolites present in the *i*ML1515 wild type BOF, that are not directly included in the production

of macromolecules, were scaled to a mass of 0.039 g and incorporated in the BOFs. A list of these metabolites and their stoichiometric coefficients are shown in Appendix E.

Maintenance energy was also imported from *iML1515*. In *iML1515*, the growth associated maintenance (GAM) and the non-growth associated maintenance (NGAM) were estimated to be 75.55 mmol ATP/gCDW h⁻¹ and 6.86 mmol ATP/gCDW, respectively (27). As NGAM was already accounted for as a reaction in *iML1515*, only GAM was incorporated in our BOFs. When GAM was incorporated into our BOFs, the energy already accounted for in the polymerisation reactions was subtracted (Supplementary material S2). The complete BOFs for all *E. coli* samples can be found in Appendix F.

By integrating maintenance energy and non-macromolecule components into our BOFs we increased their detail level from basic to advanced. According to Feist and Palsson (20), a BOF of advanced level should be sufficient to predict growth rates.

4.4 Flux balance analysis

The last step in our workflow is the evaluation of the experimentally determined BOFs (Figure 4.1). The BOFs were evaluated by running FBA for the *E. coli* strain K-12 MG1655 GEM *iML1515* with our BOFs as objectives. As glucose was the sole carbon source during cultivation, all other carbon uptake fluxes were locked to 0. To evaluate the BOF of *E. coli* U-lim, additional restrictions corresponding to the exchange rates determined in subsection 4.1.1 were applied to the model. The goal growth rate was the growth rate estimated for *E. coli* U-lim (0.71 h⁻¹). Simulations were run with the experimentally determined *E. coli* U-lim BOF and with the built-in wild-type BOF from *iML1515* (BIOMASS_Ec_iML1515_WT_75p37M) for comparison. The restrictions and goal growth rate are shown in Table 4.10 and the simulation results are shown in Table 4.11

Table 4.10: Experimentally determined exchange rates used within the model and goal growth rate for evaluation of the *E. coli* U-lim BOF.

Restricted process	Restricted reaction	Lower bound	Upper bound	Growth rate [h ⁻¹]
Glucose uptake	EX_glc_D_e	-9.06	0	0.71
Oxygen uptake	EX_o2_e	-10.09	0	
CO ₂ secretion	EX_co2_e	0	9.12	

Table 4.11: Simulation results for the evaluation of the *E. coli* U-lim BOF. Exchange rates are given in mmol/gCDWh⁻¹. The exchange rates of inorganic ions, water, and protons are not shown.

Objective function	Uptake rates					Secretion rates			Shadow prices			Predicted growth rate
	Glc	NH ₄	O ₂	P _i	SO ₄	CO ₂	Acet.	Form.	O ₂	Glc	CO ₂	
<i>E. coli</i> U-lim BOF	9.06	5.50	10.09	0.48	0.06	6.93	9.85	7.05	-0.034	-0.026	~0	0.519 h⁻¹
<i>E. coli</i> WT BOF	9.06	5.68	10.09	0.73	0.12	9.12	9.63	3.88	-0.032	-0.024	0	0.487 h⁻¹

From Table 4.11 we can see that the growth rate predicted by our BOF is closer to the goal growth rate than what was predicted by the wild-type BOF. However, both predicted growth rates are notably lower than the experimentally determined growth rate (the predicted growth rates using our BOF and the wild-type BOF are 73.1 and 68.6 % of the measured growth rate, respectively). There are also big variations in secretion rates between the two BOFs. While the wild-type BOF maximise CO₂ secretion, and hence, respiration, our BOF relies more on fermentation. The relative flow of carbon in both BOFs is shown in Figure 4.12

This work

WT model

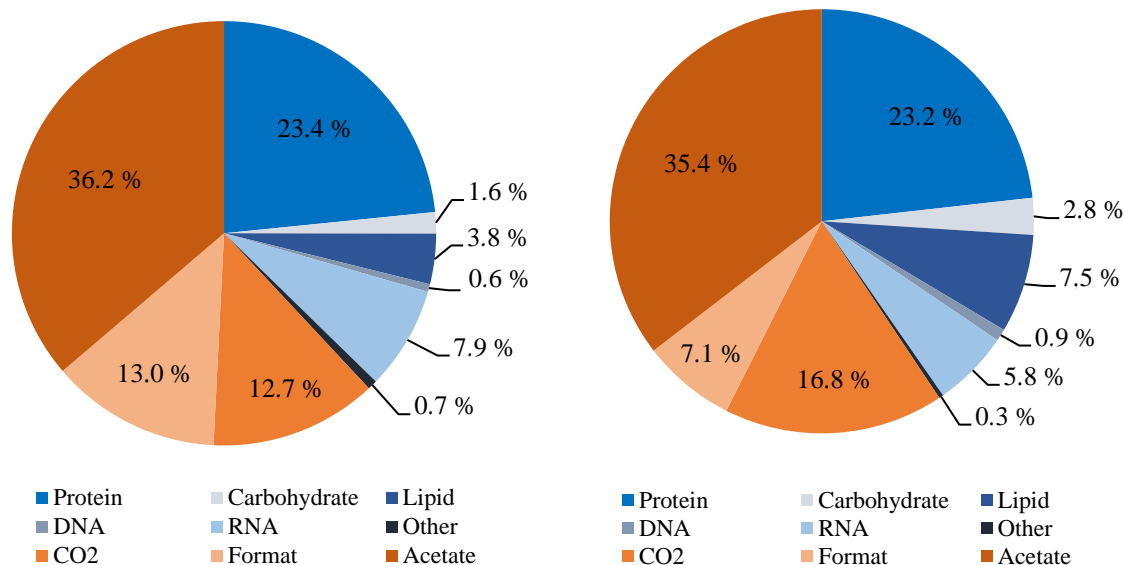


Figure 4.12: Relative carbon flow in the simulation of *E. coli* U-lim BOF and the wild-type BOF from *iML1515*.

From Figure 4.12 we can see that the relative carbon flow to biomass is similar for both BOFs (62:38 % with our BOF and 59:41 % in the wild-type BOF). The distribution between energy metabolism end products, more specifically between CO₂ and formate, differs between both BOFs. The energy production is greater in the wild-type BOF since more ATP is produced per glucose molecule in respiration than in fermentation (48). Higher energy demand in the wild-type BOF, potentially caused by the big lipid fraction, can be the explanation for these results.

There is also variation between the uptake rates when simulating both BOFs. The wild-type BOF has twice the sulphate uptake of the experimentally determined BOF and a notably higher uptake of inorganic phosphate. One possible explanation for the higher sulphate uptake in the wild-type BOF can be the methionine consumption, which is 0.15 mmol/g in the wild-type BOF and 0.05 mmol/g the experimentally determined BOF. The difference in phosphate uptake can be caused by the difference in phospholipid or DNA content.

The predicted RQ values also differ between the two BOFs. While the RQ value predicted using the wild-type BOF (0.90) matches the experimentally measured RQ value, the RQ value predicted by the experimentally determined BOF is considerably lower (0.69). This is far below the expected RQ value (1.0) for bacteria growing aerobically on glucose and is considered an indication of fermentation (138).

To further investigate the effects of glucose and oxygen uptake on the growth rate, a phenotype phase plane (PhPP) was made for our BOF (Figure 4.13). From the PhPP we can see that growth is limited by both glucose and oxygen uptakes, which is also confirmed by the shadow prices in Table 4.11. Our solution lies slightly beneath the line of optimality (Figure 4.13 A), which means that carbon is not fully oxidised and hence not fully utilised in energy production (42). This also shows that the experimentally determined BOF predicts fermentation in the given growth conditions

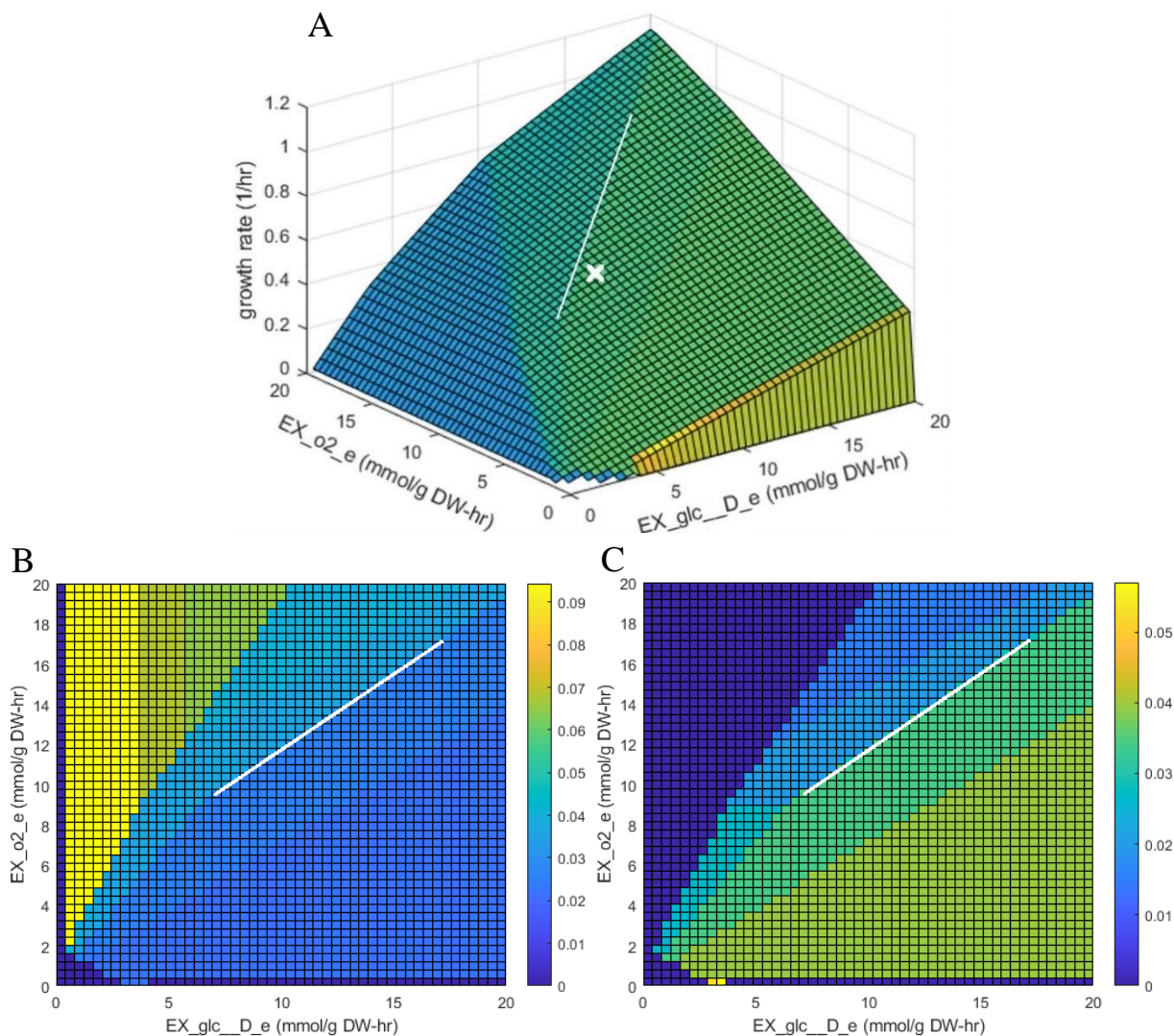


Figure 4.13: A: Glucose and oxygen phenotype phase plane for *E. coli* U-lim BOF. B: Glucose shadow prices. C: Oxygen shadow prices. White lines = line of optimality, White X = FBA solution applying restrictions in Table 4.10 (A).

The six remaining BOFs were evaluated by systematically tweaking the uptake rate of the limiting nutrient (ammonia or glucose) to evaluate if the goal growth rate can be achieved with reasonable uptake rates. The restricted reactions and goal growth rates are listed in Table 4.12 and the simulation results are listed in Table 4.13. All other uptakes were made non-limiting.

Table 4.12: Restrictions and goal growth rates for the evaluation of BOFs of *E. coli* C-lim 0.4, 0.2, and 0.1, and *E. coli* N-lim 0.4, 0.2 and 0.1.

Objective function	Restricted process	Restricted reaction	Goal growth rate [h ⁻¹]
<i>E. coli</i> C-lim 0.4 BOF	Glucose uptake	EX_glc_D_e	0.4
<i>E. coli</i> C-lim 0.2 BOF	Glucose uptake	EX_glc_D_e	0.2
<i>E. coli</i> C-lim 0.1 BOF	Glucose uptake	EX_glc_D_e	0.1
<i>E. coli</i> N-lim 0.4 BOF	Ammonium uptake	EX_nh4_e	0.4
<i>E. coli</i> N-lim 0.2 BOF	Ammonium uptake	EX_nh4_e	0.2
<i>E. coli</i> N-lim 0.1 BOF	Ammonium uptake	EX_nh4_e	0.1

Table 4.13: Simulation results for the evaluation of the BOFs of BOFs of *E. coli* C-lim 0.4, 0.2, and 0.1, and *E. coli* N-lim 0.4, 0.2, and 0.1. Exchange rates are given in mmol/gCDWh⁻¹. Exchange rates of inorganic ions, water, and protons are not shown.

Objective function	Uptake rates					Secretion rates		Shadow prices		Predicted growth rate
	Glc	NH ₄	O ₂	P _i	SO ₄	CO ₂	Acet.	Glc	NH ₄	
<i>E. coli</i> C-lim 0.4 BOF	4.58	4.41	11.0	0.42	0.04	11.2	0	-0.093	0	0.400 [h ⁻¹]
<i>E. coli</i> C-lim 0.2 BOF	2.44	2.24	6.32	0.15	0.03	6.42	0	-0.093	0	0.200 [h ⁻¹]
<i>E. coli</i> C-lim 0.1 BOF	1.39	1.08	3.97	0.07	0.02	4.12	0	-0.091	0	0.100 [h ⁻¹]
<i>E. coli</i> N-lim 0.4 BOF	4.61	4.11	10.8	0.33	0.04	11.1	0.13	~0	-0.097	0.400 [h ⁻¹]
<i>E. coli</i> N-lim 0.2 BOF	2.42	1.81	6.27	0.16	0.02	6.40	0.05	~0	-0.111	0.200 [h ⁻¹]
<i>E. coli</i> N-lim 0.1 BOF	1.39	0.92	3.92	0.05	0.01	4.08	0.03	~0	-0.108	0.100 [h ⁻¹]

According to Schulze and Lipe (139) the growth rate, the oxygen uptake rate, and the uptake of the limiting nutrient are directly proportional in aerobic continuous cultures. In Figure 4.14 the predicted growth rate (panel A) and oxygen uptake (panel B) are plotted against the uptake of limiting nutrient. We can see that there is a strong linear correlation between the uptake of limiting nutrient and both growth rate and oxygen uptake. The relationship between the uptake of limiting nutrient and growth seems to be proportional, but from the intercept in Figure 4.14 B, it appears oxygen uptake is not proportional to the uptake of limiting nutrient. The high intersect value of the linear regressions indicates that the predicted oxygen uptake is higher than what is expected in continuous cultures.

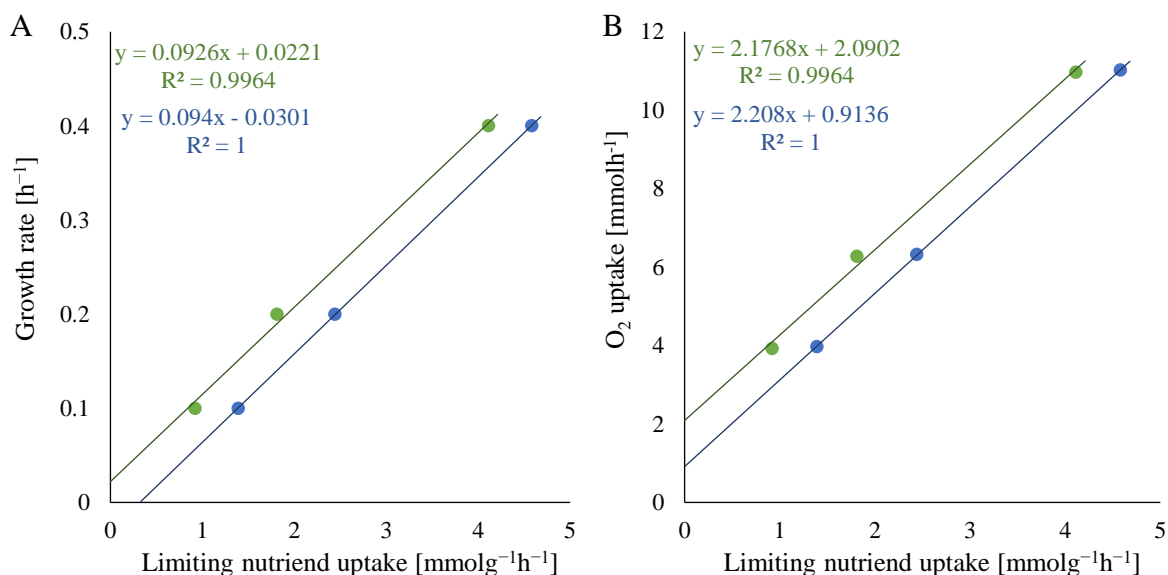


Figure 4.14: Growth (A) and oxygen uptake (B) over limiting nutrient uptake rate. Green points represent nitrogen-limited samples and red points represent carbon-limited samples. The equations and R values of the linear regressions are shown.

The predicted oxygen uptake rates in the fast-growing cultures (*E. coli* C-lim 0.4 and N-lim 0.4) are higher than what was measured for the culture growing exponentially in unlimited media (*E. coli* U-lim). Theoretically, the oxygen uptake of this culture should only be limited by the organism itself, i.e., the protein's ability to transport and utilise oxygen. Based on this, and the linear regression in Figure 4.14 B, it appears that the oxygen requirement is somewhat overestimated by the experimentally determined BOFs. This might also be the case in *E. coli* U-lim BOF, which predicted considerably more fermentation than what was suggested by the experimentally determined RQ value. It should be noted, however, that the growth conditions of *E. coli* U-lim might not have been entirely aerobic. Even in agitated aerobic fermenters, the DO level is not completely homogenous, and there will be areas where the access to oxygen is low (140).

Kayser *et al.* (141) have published a study where they measured the glucose uptake of the *E. coli* K-12 strain TG1, growing aerobically at different growth rates on a glucose-limited medium. Our predicted values are shown alongside their measurements in Table 4.14. Compared to the values reported by Kayser *et al.*, glucose requirements were consequently overestimated in our simulations. A part of the difference can be a result of different strains as both the measured biomass composition and the GEM used in simulations are for *E. coli* K-12 strain MG1655. Monk *et al.* (27) have reported variations in the glucose uptake rate of different *E. coli* strains growing aerobically at similar growth rates. Furthermore, the growth conditions differ between this work and what Kayser *et al.* report. The culture studied by Kayser *et al.* was grown at pH 6.6 and 28 °C, while the cultures in this work were grown at pH 7.0 and 37 °C. More importantly, the media used by Kayser *et al.* contained EDTA and acetate in addition to glucose (141). The presence of alternative carbon sources can contribute to an increased growth rate as these are often utilised during glucose limitation (142). The oxygen uptake predicted by our experimentally determined BOF also seems to be somewhat overdetermined.

Table 4.14: Comparison of predicted glucose and oxygen uptake to values measured on *E. coli* K-12 strain TG1 (141). Growth rates are given in h^{-1} , uptake rates are given in $\text{mmol g}^{-1}\text{h}^{-1}$.

BOF	Simulation predictions			Measured values		
	Growth rate	Glucose uptake	O ₂ uptake	Growth rate	Glucose uptake	O ₂ uptake
<i>E. coli</i> C-lim 0.4	0.400	4.58	11.0	0.397	4.01	8.23
<i>E. coli</i> C-lim 0.2	0.200	2.44	6.32	0.203	2.08	5.49
<i>E. coli</i> C-lim 0.1	0.100	1.39	3.97	0.134	1.40	3.40

E. coli is capable of fast adaptation to changes in nutrient availability. Therefore, it is natural to expect the *E. coli* C-lim 0.1 BOF and *E. coli* N-lim 0.1 BOF to predict faster growth in carbon and nitrogen-limited conditions, respectively, than an “unspecialised” BOF (143). Figure 4.15 shows the growth rate predicted by *E. coli* C-lim 0.1 BOF, *E. coli* N-lim 0.1 BOF and the wild-type BOF from *iML1515* at strong carbon and nitrogen limitations (uptake rate $1.0 \text{ mmol g}^{-1} \text{ h}^{-1}$). The growth rate predicted for strong carbon limitation is stable for all three BOFs, but under strong nitrogen limitation, the *E. coli* N-lim 0.1 BOF predicted a notably higher growth rate. This is most likely because the *E. coli* N-lim 1.0 BOF have a large proportion of carbohydrates (17.19 %).

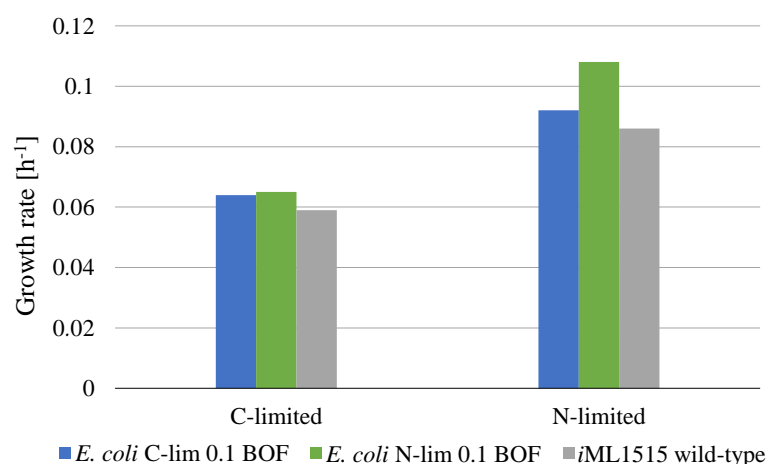


Figure 4.15: Predicted growth rates of *E. coli* C-lim BOF, *E. coli* N-lim BOF, and the wild-type BOF from *iML1515* when restricting glucose uptake to $1.0 \text{ mmol g}^{-1} \text{ h}^{-1}$ (C-limited) and when restricting ammonia uptake to $1.0 \text{ mmol g}^{-1} \text{ h}^{-1}$ (N-limited).

5 Discussion

In this chapter I will evaluate the experimental results further. First, by comparison to the elemental analysis results, then by comparison to the *E. coli* biomass composition that has been previously reported by Neidhardt *et al.* (35) and Beck *et al.* (33). I will also briefly discuss our biomass recovery of *S. cerevisiae*, *S. salar*, and *A.thaliana*. Finally, I will address some general observations regarding the simulation results.

One way to evaluate the macromolecule composition measured in this work is by comparison to the total elemental composition measured by combustion. Figure 5.1 shows the content of carbon, hydrogen, nitrogen, and oxygen, measured as macromolecules and by combustion, in all *E. coli* cultures. Elemental composition measured by combustion analysis is also shown in Table 4.6.

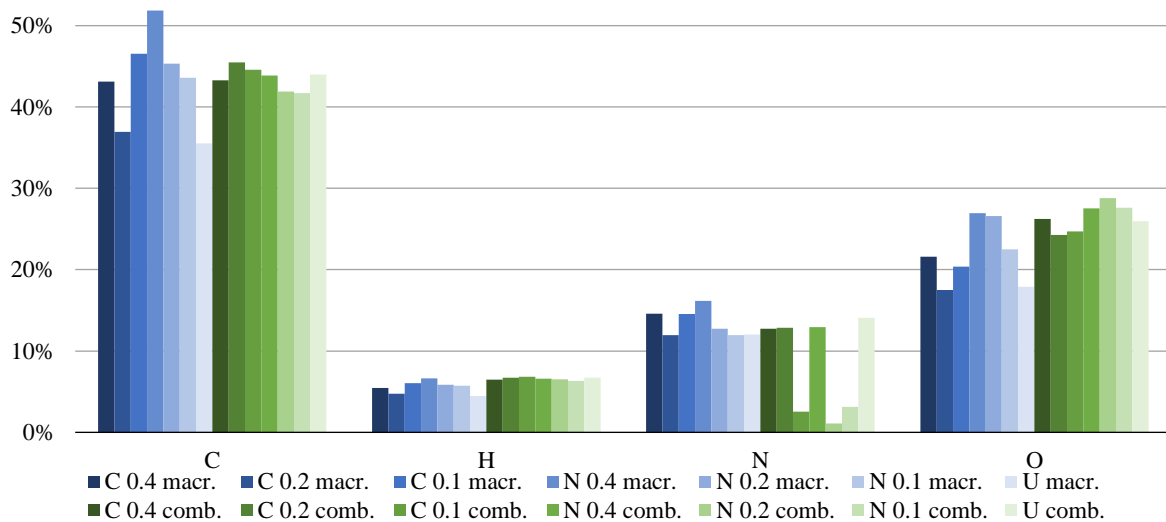


Figure 5.1: Elemental contents given as mass % of CDW in all *E. coli* cultures. Blue columns are measured as macromolecules and green columns are measured by elemental analysis by combustion.

The measured nitrogen content in *E. coli* C-lim 0.1, *E. coli* N-lim 0.2, and *E. coli* N-lim 0.1 are not considered when assessing the results as there was trouble measuring nitrogen content in these samples. The elemental distribution is generally more stable when measured by combustion, which is not surprising, as it is based on fewer measurements. Because nitrogen content could not be measured in some of the samples, it is hard to say if there is a difference in nitrogen content between nitrogen and carbon-limited growth. From the oxygen content, we can see that the nitrogen-limited samples contain more oxygen, though this is clearer in the macromolecule measurements than in the elemental analysis. The increased oxygen content is most likely a result of carbohydrate accumulation, as carbohydrates generally have high oxygen content.

The sulphur contents measured as macromolecules and by combustion were too low to be visualised along with the other elements in Figure 5.1 and are therefore shown in Figure 5.2. The only sulphur-containing compound measured in macromolecule assays is methionine. In combustion analysis, on the other hand, all sulphur-containing cellular components are accounted for. These can be components such as biotin, coenzyme A and sulphate, to mention a few. Since most sulphur-containing components are not included in macromolecule measurements, one would expect the sulphur content measured by combustion to be higher. The high macromolecular sulphur contents in Figure 5.2, can partly be explained by overdetermined protein contents. However, the sulphur content might also be underdetermined in the combustion analysis.

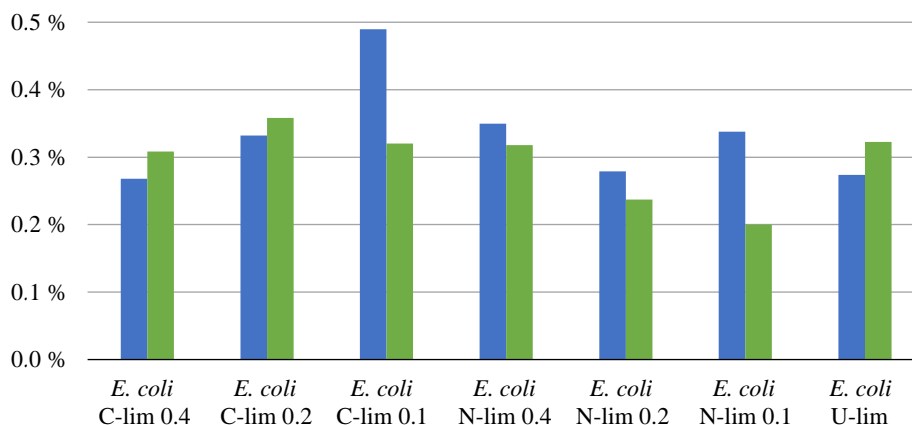


Figure 5.2: Sulphur content given as mass % of CDW in all *E. coli* cultures. Blue columns are measured as macromolecules and green columns are measured by elemental analysis by combustion.

The average elemental composition of all *E. coli* samples, measured as macromolecules and by combustion analysis, is shown in Figure 5.3. Here we can see that the average elemental distribution is very similar between measuring methods. The only exception is oxygen content which is higher in the combustion analysis. Low oxygen concentration can imply that the carbohydrates are somewhat underdetermined in our macromolecule quantification. This is also supported by the comparison of ribose measured as carbohydrates and as RNA (Figure 4.10).

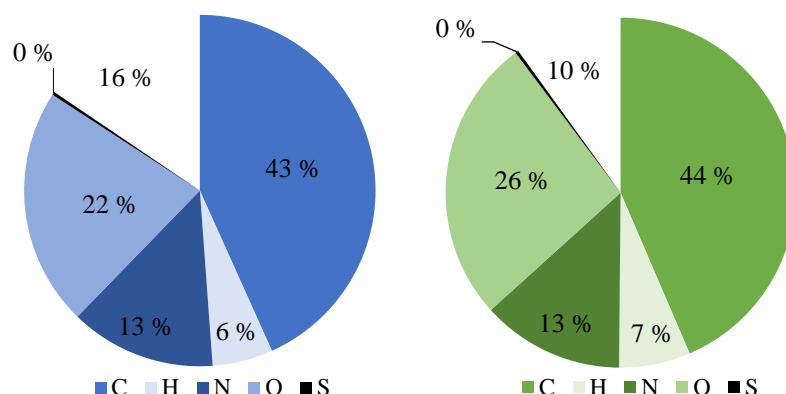


Figure 5.3: Average elemental composition of *E. coli* measured as macromolecules (blue) and by combustion (orange). Nitrogen values from *E. coli* C-lim 0.1, N-lim 0.2, and N-lim 0.1 measured by combustion are excluded.

In Table 5.1 the *E. coli* biomass compositions measured in this work are compared to measurements done by Beck *et al.* (33) and to the gold-standard *E. coli* biomass composition reported by Neidhardt *et al.* (35). Of our *E. coli* cultures, *E. coli* U-lim has the growth conditions that are closest to both Neidhardt *et al.* and Beck *et al.* The composition data from Neidhardt *et al.* are based on *E. coli* strain B grown in glucose minimal medium at 37 °C and harvested at a growth rate of 1.5 h⁻¹. Beck *et al.* measured the biomass composition of *E. coli* strain K-12 MG1655 grown in a glucose minimal medium at 37 °C and harvested during exponential growth. The composition reported by Neidhardt is scaled up to 100 % and the “other” group (see Table 5.1) consists of non-carbohydrate monomers of lipopolysaccharides and peptidoglycan as well as vitamins, building block metabolites, and inorganic ions.

Table 5.1: Comparison of measured macromolecule and monomer composition of *E. coli* to the compositions found by Beck *et al.* (33) and Neidhardt *et al.* (35). All concentrations are given in mass % of total CDW. “-” means not measured, “n.d.” means not detected. Red numbers are measured without technical replicates. The remaining measurements listed in this work are the average of three technical replicates. Standard deviations are not included due to lack of space. For measurements and standard deviations see tables in the previous chapter or Supplementary material S2.

Compound	<i>E. coli</i> C-lim 0.4	<i>E. coli</i> C-lim 0.2	<i>E. coli</i> C-lim 0.1	<i>E. coli</i> N-lim 0.4	<i>E. coli</i> N-lim 0.2	<i>E. coli</i> N-lim 0.1	<i>E. coli</i> U-lim	Beck <i>et al.</i>	Neidhardt <i>et al.</i>
Protein	52.3	54.0	66.4	63.4	47.2	54.0	41.1	35.2	55.0
Alanine	4.7	4.8	5.7	5.5	4.1	4.5	4.3	2.5	3.4
Arginine	2.8	3.0	3.3	3.3	2.5	2.3	2.5	2.4	4.3
Asparagine/Aspartate	6.9	7.1	8.5	8.2	6.2	6.5	5.0	3.5	5.2
Cystine	0.4	0.5	0.8	0.6	0.5	0.7	0.5	0.4	0.9
Glutamine/Glutamate	8.2	8.4	9.7	10.0	7.0	9.0	4.6	5.0	6.4
Glycine	1.4	1.4	1.6	1.6	1.2	1.1	1.2	2.3	3.3
Histidine	1.1	1.1	1.3	1.4	1.0	1.1	1.2	0.7	1.2
Isoleucine	2.6	2.9	3.5	3.4	2.5	3.1	1.8	1.7	3.1
Leucine	5.0	5.1	6.2	6.2	4.5	4.9	4.0	2.6	4.8
Lysine	3.2	2.7	3.7	3.3	2.9	3.8	3.4	2.4	4.2
Methionine	0.6	0.7	1.0	0.9	0.5	0.5	0.4	1.1	1.9
Phenylalanine	2.6	2.7	3.5	3.2	2.4	2.8	3.0	1.8	2.6
Proline	2.0	2.1	2.7	2.5	1.9	2.2	1.7	1.3	2.0
Serine	2.1	2.2	2.7	2.5	1.9	2.2	1.8	1.4	1.8
Threonine	2.4	2.6	3.2	3.1	2.3	2.5	2.3	1.7	2.4
Tryptophan	1.1	1.2	1.9	1.5	1.2	1.6	1.2	0.8	1.0
Tyrosine	1.9	1.9	2.8	2.3	1.9	2.0	0.9	1.4	2.1
Valine	3.2	3.5	4.2	4.2	3.0	3.3	1.4	2.2	4.0
Carbohydrate	2.9	2.9	3.6	9.9	18.48	15.05	3.47	4.2	5.3
Glucose	1.8	1.8	2.0	9.0	17.5	9.0	2.5	4.2	2.7
Galactose	0.4	0.4	0.4	0.4	0.4	0.2	0.4	-	-
Glucuronic acid	n.d.	n.d.	n.d.	n.d.	n.d.	0.6	n.d.	-	-
Rhamnose	n.d.	n.d.	0.6	n.d.	n.d.	3.6	n.d.	-	-
Glucosamine	0.6	0.6	0.7	0.6	0.6	1.6	0.6	-	0.3
N-acetylglucosamine	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	-	0.6
N-acetylmuramic acid	-	-	-	-	-	-	-	-	0.8
KDO	-	-	-	-	-	-	-	-	0.6
Heptose	-	-	-	-	-	-	-	-	0.5
Lipid	5.9	2.8	6.1	5.9	5.2	4.9	5.3	6.7	9.1
DNA	6.5	-	-	5.3	6.0	1.3	1.3	1.0	3.1
dAMP	1.6	-	-	1.3	1.5	0.3	0.3	0.2	0.8
dCMP	1.5	-	-	1.3	1.4	0.3	0.3	0.2	0.7
dGMP	1.7	-	-	1.3	1.4	0.3	0.3	0.3	0.8
dTMP	1.6	-	-	1.3	1.5	0.3	0.3	0.2	0.7
RNA	17.5	11.9	11.9	17.1	14.0	8.8	19.0	17.2	20.5
AMP	4.6	3.1	3.1	4.5	3.7	2.3	5.0	4.5	5.4
CMP	3.7	2.5	2.5	3.6	3.0	1.9	4.0	3.6	3.8
GMP	5.9	4.0	4.0	5.7	4.7	2.9	6.4	5.8	7.0
UMP	3.4	2.3	2.3	3.3	2.7	1.7	3.7	3.3	4.2
Others	-	-	-	-	-	-	-	-	7.0
Total	85.0	75.6	92.1	101.6	90.8	84.1	70.3	64.3	100

From Table 5.1 we can see that the macromolecular measurements sum up to an average of over 100 % of the dry weight of *E. coli* N-lim 0.4. This overdetermination is thought to mainly be caused by an over-quantification of proteins. Several of the protein measurements may be overdetermined due to volume loss during hydrolysis (as discussed in subsection 3.2.1). Both *E. coli* N-lim 0.4 and *E. coli* C-lim 0.1 have a great standard deviation in protein measurements and are suspected to have considerably overdetermined protein contents (see Table 4.7). Of all *E. coli* samples, the total macromolecule recovery was lowest in *E. coli* U-lim. Due to modifications of the protein quantification, the protein content in this culture was not significantly overdetermined, which might be the reason for the higher precision in this sample. Generally, the precision is high for the measurements of carbohydrates, lipids, and RNA, and since DNA contributes only a small portion of the total biomass, most of the uncertainty results from protein measurements (Table 4.7).

Compared to the value reported by Neidhardt *et al.* (35) our protein contents are generally low (Table 5.1). The exceptions are *E. coli* C-lim 0.1 and *E. coli* N-lim 0.4 which are suspected to be overdetermined based on their high standard deviation (Table 4.7). The protein content measured in *E. coli* U-lim, which is the most reliable of our protein measurements, is considerably lower than Neidhardt *et al.*, but still higher than what was measured by Beck *et al.* (33) (Table 5.1). Increasing the amount of biomass used for protein determination from ~1 mg to ~10 mg (as done for *E. coli* U-lim) would be an improvement of the method we have used. This would not only reduce the chance of overdetermination, but also minimise the potential loss caused by biomass sticking to the walls of the glass vials. By weighting the samples before and after hydrolysis one could also assess the potential overdetermination from volume loss.

The relative amino acid distribution is quite stable between our *E. coli* cultures (see Figure 4.8 or Appendix B). The distribution is also similar to what was reported by Neidhardt *et al.* (35) and Beck *et al.* (33), but there are some differences worth pointing out. The relative methionine content measured in our cultures is almost 3 times lower than what is previously reported. This indicates that methionine most likely was oxidised during the HPLC sample preparation as discussed in section 4.2.1. The content of glycine measured in our samples is also less than half of the previously reported glycine content. This is partly because we estimated glucose content from the overall content of glycine and arginine and the genomic codon ratio (see Supplementary material S2). However, the arginine content in our samples is also somewhat lower than what was found by Neidhardt *et al.* and slightly less than what Beck *et al.* found, meaning that the total glycine and arginine content we measured is lower than previously reported. The relative content of alanine and glutamine/glutamate are considerably higher in all our samples than what was reported by Neidhardt *et al.* and Beck *et al.* Furthermore, the relative asparagine/aspartate is also notably higher in all our samples, except *E. coli* U-lim, than what Neidhardt *et al.* reports, but similar to what reported by Beck *et al.* As Neidhardt *et al.* analysed another *E. coli* strain than what we and Beck *et al.* did, the variation in asparagine/aspartate content can be explained by strain-dependent variations.

Our carbohydrate measurements are lower than what was found by Neidhardt *et al.* (35) and Beck *et al.* (33) in all samples that were not grown in nitrogen limitation. As carbohydrate content is expected to be much higher in nitrogen-limited cultures, these samples are not comparable to Neidhardt *et al.* and Beck *et al.* An underdetermination of carbohydrate content has already been suggested from the comparison of ribose measured as carbohydrates and RNA (Figure 4.10), and from the comparison of total oxygen content measured as macromolecules and by combustion analysis (Figure 5.3). Based on this, and that our carbohydrate content is notably lower than previously reported values, we can conclude that the carbohydrate content is most likely somewhat underdetermined. On the other hand, the carbohydrate assay we used has an impressive improvement of resolution compared to carbohydrate assays traditionally used in biomass composition analysis (33, 105). In addition to glucose, we have

quantified galactose, glucuronic acid, rhamnose, and glucosamine, though glucuronic acid and rhamnose were not detected in all samples. In contrast, Beck *et al.* assumed all carbohydrates were glucose, completely ignoring carbohydrates found in peptidoglycan and LPS. Because N-acetyl-glucosamine was deacetylated during hydrolysis and detected as glucosamine, we could not quantify peptidoglycan in our samples (109) (discussed in subsection 3.2.2). Other carbohydrates that were quantified by Neidhardt *et al.* and not by us are N-acetylmuramic acid, heptoses, and 3-deoxy-d-manno-2-octulosonic acid (KDO). All these compounds are expected to be present in *E. coli*, as they are essential parts of peptidoglycan (N-acetylmuramic acid) or common components of the core oligosaccharide in LPS (heptoses and KDO) (61, 62).

The lipid content measured in our samples is lower than reported by Neidhardt *et al.* (35) and slightly lower than reported by Beck *et al.* (33)(Table 5.1). A part of the lipids was most likely lost due to inadequate extraction (Table 4.3). If the lipids content is scaled up to account for this loss (~10 % assuming no chloroform evaporation) the values are similar to the measurements done by Beck *et al.*, but still lower than what Neidhardt *et al.* reported. Adding more repetitions of the extraction step to the method could potentially increase the lipid recovery, but not enough to reach the lipid level reported by Neidhardt *et al.* The remaining difference can be a consequence of variations between strains (144). The prolonged storage of the *E. coli* samples, due to the Covid-19 pandemic is a likely reason why lipid class distribution could not be measured directly. In the future, the UHPSFC-MS method described by Bartosova *et al.* (118) should be applied to quantify lipid classes. As the fatty acid distribution of *E. coli* is strain- and condition-dependent (145), this also should be measured directly. This could be done by FAME (fatty acid methyl ester) analysis using gas chromatography-vacuum ultraviolet spectroscopy (GC-VUV) (146).

When comparing our DNA measurements to Neidhardt *et al.* (35), it appears that the DNA content has been overdetermined in some samples and underdetermined in others (Table 5.1). The samples were not pure and the overdetermination can be a result of contamination (Table 4.4). Underdetermination is most likely caused by inadequate extraction. Since both these sources of error are linked to the isolation step, one should consider changing to a method where isolation is not needed. E.g., by PicoGreen marking and fluorometric spectrophotometry, as suggested by Sandaa *et al.* (132). Our RNA measurements are in coherence with both Neidhardt *et al.* and Beck *et al.* (33) (Table 5.1). Because all our samples, except *E. coli* U- lim, had low growth rates, they were expected to have significantly lower RNA content than what Neidhardt *et al.* and Beck *et al.* measured (69). The RNA content in *E. coli* U-lim is also somewhat lower than in Neidhardt *et al.*, but this is most likely also a result of different growth rates. Optimally, the nucleotide distribution of both DNA and RNA should be measured directly instead of estimated. A method that shows great promise for this task is the HPLC method proposed by Huang *et al.* (147).

In addition to the macromolecules that we have measured, Neidhardt *et al.* (35) have reported the content of LPS, peptidoglycan, precursor molecules of monomers, vitamins, cofactors, and inorganic ions. We account for a part of the lipopolysaccharide in our carbohydrate assay, but the rest of these molecules are not measured in our method. Hence, we are not aiming for 100 % biomass recovery with our method. These molecules that remain undetected do not necessarily contribute a great part of the total biomass. Neidhardt *et al.* report that soluble metabolites and inorganic ions make up a total of 3.9 % of the dry weight. Nevertheless, they are essential for cell function and growth and should therefore be considered when constructing a BOF. In the future measurement of these compounds might be added to the biomass composition analysis.

The total biomass recovery in this work is considerably higher than what was reported by Beck *et al.* (33) (Table 5.1). Even when the standard deviation is subtracted (Table 4.7), all *E. coli* cultures have a greater recovery than what Beck *et al.* reported. High recovery percentages are important in BOF

determination because they leave less room for error associated with scaling the content to 1 g. As discussed in subsection 4.2.6, the scaling factor is not necessarily proportional to the loss of recovery of each macromolecule group.

Our measurements show that the biomass composition of *E. coli* varies greatly with growth conditions (Table 5.1). This is maybe most evident from the glycogen and total carbohydrate content. The measured glycogen content is more than four times higher in the strongly nitrogen-limited cultures than in the carbon-limited cultures. Also, a considerable amount of rhamnose was detected in *E. coli* N-lim 0.1 and some were found in *E. coli* C-lim 0.1, while the remaining cultures did not contain detectable levels of this sugar. Furthermore, the measured RNA content is dependent on the growth rate. There also appear to be some variations in the total protein and DNA content. However, the uncertainty of some of these measurements is high, making it hard to conclude if these are relevant biological variations or just variations due to measurement error.

The measured macromolecule composition of *S. cerevisiae*, *S. salar*, and *A. thaliana* are shown in Table 5.2 alongside previously reported values. Even though we measured all macromolecule groups of *S. cerevisiae*, only 32.8 % of the biomass was recovered. The loss is mainly in proteins and carbohydrates, as both these macromolecule classes are assumed greatly underdetermined. In *S. salar* muscle, DNA is not expected to make up a big proportion of the total biomass, hence, the recovery of 28.6 is considered low. The loss is assumed to mainly be caused by the underdetermination of proteins as the other macromolecules are in coherence with previously reported values.

For *A. thaliana* 53.3 % of the dry weight content was recovered, by measuring proteins, carbohydrates, and lipids. DNA and RNA are not expected to contribute much to the total dry weight. Still, the biomass recovery is promising. It is hard to say where the biomass is “lost”, as all measured macromolecule groups were either higher than or at coherence with previously reported values. A part of the remaining biomass is accounted for by soluble metabolite pools, cofactors, vitamins, inorganic ions, and pigments.

Generally, the methods for protein and carbohydrate determination would need some modifications to work well on eukaryotes. The addition of a homogenisation step could potentially increase the recovery of both proteins and carbohydrates, especially for multicellular organisms, as this would break down complex structures and make the proteins and carbohydrates more available for measurement.

Table 5.2: Comparison of the measured macromolecular composition of *S. cerevisiae*, *S. salar*, and *A. thaliana* to previously reported values. Values are given in mass % of total CDW. “-“ means not measured or no reference. Red values are measured with no technical replicates. The remaining values are the average of three technical replicates.

Macromolecule	<i>S. cerevisiae</i>		<i>S. salar</i>		<i>A. thaliana</i>	
	This work	Reference	This work	Reference	This work	Reference
Protein	13.38 ± 0.67	58 – 63 ^a	5.26 ± 0.46	~ 76 ^e	3.27 ± 1.13	1 – 2.5 ^g
Carbohydrate	13.24 ± 1.18	43.6 ^b	0.15 ± 0.04	-	47.2 ± 2.1	22.44 ^h
Lipid	1.64	1.21 ^c	22.7	21.03 ^e	2.80	< 1.0 ⁱ
DNA	1.41 ± 0.45	-	-	-	-	-
RNA	3.12 ± 0.14	0.93 – 1.13 ^d	0.45 ± 0.03	0.71 ^f	-	-
Total	32.8 ± 2.4	-	28.6 ± 0.5	-	53.3 ± 3.2	-

^a Albers *et al.* (103)

^b Plata *et al.* (111)

^c Arneborg *et al.* (123)

^d Ertugay and Hamamci (136)

^e Bell *et al.* (104)

^f McKee *et al.* (135)

^g Li *et al.* (102)

^h Teng *et al.* (112)

ⁱ Gigon *et al.* (122)

The FBA simulation using our experimentally determined BOFs demonstrate that the BOF affects prediction results. Our experimentally determined BOF for exponentially growing *E. coli* achieved better predictions of growth rate than the wild-type BOF from *iML1515*, showing that more

condition-specific BOFs can increase prediction accuracy (Table 4.11). The difference in predictions was, however, even more evident from the predicted exchange fluxes, as the predictions made by the wild-type BOF from the model were dominated by respiration while the experimentally determined BOF predicted more fermentation. Furthermore, simulations of the wild-type BOF and the BOFs constructed for *E. coli* C-lim 0.1 and *E. coli* N-lim 0.1 during strong nitrogen limitation accentuate how various BOFs affect growth rate (Figure 4.15).

There are some indications that the BOFs we have constructed for the continuous cultures tend to overestimate the oxygen requirement. Comparison of predictions made by our C-limited BOFs to measured uptake rates of a related *E. coli* shows that the predicted oxygen uptake is consequently higher than the measured uptake (Table 4.14). Even though the strain and growth conditions do not match our BOF completely, this pattern should be noted. Furthermore, the high intersects of the linear regressions between oxygen and limiting nutrient uptake predicted by these BOFs, also suggest overestimated oxygen uptake (Figure 4.13 B). On the other hand, the predicted uptake rates of limiting nutrient are reasonable, both when considering the relationship to growth rate and compared to previously reported values.

By growing and measuring some of the exchange rates of *E. coli* U-lim during cultivations, we have obtained a basis for evaluation of the prediction results. While the experimentally determined BOF for *E. coli* U-lim predicted a more accurate growth rate than the wild-type BOF from the model, the wild-type BOF made more accurate predictions of CO₂ secretion (Table 4.11).

Table 4.11). One way to further compare the accuracies of the BOFs would be to analyse experimental measurements of fermentation products. By avoiding lyophilisation of media samples, the fermentation product secretion rates could be determined by NMR in the future. Measurement of the exchange rates of the continuous cultures would support further assessment of the BOFs constructed for these samples. This would be especially useful for the nitrogen-limited cultures, as no published material was found on the ammonia uptake rate of *E. coli* during nitrogen-limited growth. Furthermore, such measurements would allow us to estimate the maintenance energy requirements of our cultures in a similar way to what was done in the *iML1515* model.

6 Conclusion and Outlook

The principal aim of this thesis was to construct condition-specific BOFs for *E. coli* based on experimentally determined data. The biomass composition of seven *E. coli* cultures was successfully measured and implemented in the *iML1515* model for phenotypic evaluation. Biomass recovery within this work was at 86 % on average. This is considerably more than what Beck *et al.* (33) reported (64.3 %) and matches the impressive 82 % achieved by Long and Antoniewicz (34) using GC/MS with ¹³C labelled samples.

The simulation results using the constructed BOFs are promising. The growth rate prediction accuracy was increased by 5 % using the BOF constructed for *E. coli* U-lim compared to *E. coli* wild-type BOF from *iML1515*. Furthermore, the BOFs constructed for the remaining *E. coli* cultures also predicted reasonable uptake rates of limiting nutrients, both compared to experimentally measured values and based on theoretical relationships between growth and uptake (139, 141). This shows a potential for improving GEMs by constructing condition-specific BOFs. However, more detailed measurements of exchange rates should be done during cultivation in the future, as this would give a better basis for the evaluation of the simulation results.

The pipeline of experimental methods used in this work is time efficient and has allowed us to measure the complete macromolecule content in *E. coli*. All measurements are performed on the same sample of *E. coli* avoiding any variations in the phenotypic state between the macromolecule groups. Furthermore, our experimental results are comparable to the gold-standard biomass composition of *E. coli* reported by Neidhardt *et al.* (35). Our measured carbohydrate content suggests a somewhat lower recovery compared to previously reported values. On the other hand, the resolution achieved using this method is superior to what Beck *et al.* (33) and Long and Antoniewicz (34) have reported. The increased resolution is of great value since the method measures components of peptidoglycan and LPS – two important contributors to bacterial biomass (35).

Carbohydrate, lipid, and RNA quantification show high precision. For some of the *E. coli* cultures, the protein quantification had great variation between technical replicates, however, the modifications made to the procedure, that were tested on *E. coli* U-lim, showed great improvement of precision. The lipid, DNA, and RNA quantification methods in this work, all include extraction steps. Inadequate extraction is a potential cause of the underdetermination of these macromolecules. Especially in the case of DNA, there appears to be considerable loss during extraction. For DNA and RNA quantification, methods without isolation steps should be considered in the future. Another alternative would be to use an extraction standard in these procedures to evaluate loss during sample preparation, such as labelled lipids for lipid quantification (148). By using a more refined MS method for quantification of DNA and RNA, labelled nucleotides could be used as a standard in these quantifications (149).

Implementing the lipid class and fatty acid distribution reported by Neidhardt *et al.* into our BOFs makes them less strain and growth condition-specific and should therefore be avoided in the future. The lipid class assay using UHPSFC-MS described by Bartosova *et al.* (118) can be used if prolonged storage of the samples is avoided. Additionally, a FAME analysis could be performed to determine the fatty acid distribution. By measuring the nucleotide distribution in DNA and RNA directly, we could assess whether it varies significantly with growth conditions or deviates much from values estimated from the genome. Direct measurements of monomer distribution are especially important if this pipeline of methods is to be applied to organisms that are not as extensively studied as *E. coli*.

The specificity of the BOFs could be increased further by extending our pipeline to include methods for quantification of ions, vitamins, cofactors, and other molecules that do not fit into the five

macromolecule groups. Even though the contribution from these compounds to the total biomass of *E. coli* is small, other organisms may produce considerably larger amounts of these molecules. For example, hypersaline species of the microalgae *Dunaliella* can accumulate up to 14 % β -carotene (150). Furthermore, measuring these compounds would increase the biological relevance for environment-specific pathways. Another important contribution to the BOFs, that strongly affects simulation predictions, is the maintenance energy term (151). By measuring exchange rates during cultivation, specific maintenance energy could be calculated for various organisms (152). It should be considered to do this in the future, especially for less studied organisms than *E. coli*.

The pipeline of methods used in this work has shown great potential for construction of condition-specific *E. coli* BOFs. The application range can be broadened to include more organisms by addition of the steps discussed above. Measurements of *S. cerevisiae*, *S. salar*, and *A. thaliana* biomass show that the methods for protein and carbohydrate quantification must be modified for construction of eucaryote BOFs, potentially by the addition of homogenisation steps. Furthermore, all methods in the pipeline should be optimised for high-throughput in the future.

References

1. Delker RK, Mann RS. From Reductionism to Holism: Toward a More Complete View of Development Through Genome Engineering. *Adv Exp Med Biol.* 2017;1016:45-74.
2. Kienle G, Kiene H. From Reductionism to Holism: Systems-oriented Approaches in Cancer Research. *Glob Adv Health Med.* 2012;1(5):68-77.
3. Kitano H. Systems Biology: A Brief Overview. *Science.* 2002;295(5560):1662.
4. Petrasek D. Systems biology: the case for a systems science approach to diabetes. *J Diabetes Sci Technol.* 2008;2(1):131-4.
5. Robinson SW, Fernandes M, Husi H. Current advances in systems and integrative biology. *Computational and Structural Biotechnology Journal.* 2014;11(18):35-46.
6. Zielinski DC, Jamshidi N, Corbett AJ, Bordbar A, Thomas A, Palsson BO. Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism. *Scientific Reports.* 2017;7(1):41241.
7. Hansen ASL, Lennen RM, Sonnenschein N, Herrgård MJ. Systems biology solutions for biochemical production challenges. *Current opinion in biotechnology.* 2017;45:85-91.
8. Pezeshki A, Ovsyannikova IG, McKinney BA, Poland GA, Kennedy RB. The role of systems biology approaches in determining molecular signatures for the development of more effective vaccines. *Expert Review of Vaccines.* 2019;18(3):253-67.
9. Yasemi M, Jolicoeur M. Modelling Cell Metabolism: A Review on Constraint-Based Steady-State and Kinetic Approaches. *Processes.* 2021;9(2).
10. Oberhardt MA, Yizhak K, Ruppin E. Metabolically re-modeling the drug pipeline. *Current Opinion in Pharmacology.* 2013;13(5):778-85.
11. Rau MH, Zeidan AA. Constraint-based modeling in microbial food biotechnology. *Biochemical Society Transactions.* 2018;46(2):249-60.
12. Resendis-Antonio O. Constraint-based Modeling. In: Dubitzky W, Wolkenhauer O, Cho K-H, Yokota H, editors. *Encyclopedia of Systems Biology.* New York, NY: Springer New York; 2013. p. 494-8.
13. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. *Genome Biology.* 2019;20(1):121.
14. Jensen KG, S.; Herregård, M., J. Enhancing Metabolic Models with Genome-Scale Experimental Data. In: Rajewsky NJ, S.; Barciszewski, J., editor. *Systems biology: Springer Nature;* 2018. p. 337-48.
15. Kim WJ, Kim HU, Lee SY. Current state and applications of microbial genome-scale metabolic models. *Current Opinion in Systems Biology.* 2017;2:10-8.
16. Zhang C, Hua Q. Applications of genome-scale metabolic models in biotechnology and systems medicine. *Frontiers in physiology.* 2016;6:413.
17. O'Brien EJ, Monk JM, Palsson BO. Using Genome-scale Models to Predict Biological Capabilities. *Cell.* 2015;161(5):971-87.
18. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol.* 2010;28(3):245-8.
19. Madigan M, Martinko J, Bender K, Buckley D, Sathl D. *Microbial Growth and Control. Brock Biology of Microorganisms.* 14 ed: Pearson Education Limited; 2015. p. 167-206.
20. Feist AM, Palsson BO. The biomass objective function. *Current Opinion in Microbiology.* 2010;13(3):344-9.
21. Bernstein DB, Sulheim S, Almaas E, Segrè D. Addressing uncertainty in genome-scale metabolic model reconstruction and analysis. *Genome Biology.* 2021;22(1):64.
22. Pramanik J, Keasling JD. Effect of *Escherichia coli* biomass composition on central metabolic fluxes predicted by a stoichiometric model. *Biotechnology and Bioengineering.* 1998;60(2):230-8.
23. Dikicioglu D, Kırdar B, Oliver SG. Biomass composition: the “elephant in the room” of metabolic modelling. *Metabolomics.* 2015;11(6):1690-701.

24. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, et al. A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Molecular Systems Biology*. 2007;3(1):121.
25. Yuan H, Cheung CY, Hilbers PAJ, van Riel NAW. Flux balance analysis of plant metabolism: the effect of biomass composition and model structure on model predictions. *Frontiers in plant science*. 2016;7:537.
26. Lachance J-C, Lloyd CJ, Monk JM, Yang L, Sastry AV, Seif Y, et al. BOFdat: Generating biomass objective functions for genome-scale metabolic models from experimental data. *PLoS Comput Biol*. 2019;15(4).
27. Monk JM, Lloyd CJ, Brunk E, Mih N, Sastry A, King Z, et al. iML1515, a knowledgebase that computes *Escherichia coli* traits. *Nat Biotechnol*. 2017;35(10):904-8.
28. Henry CS, Zinner JF, Cohoon MP, Stevens RL. iBsu1103: a new genome-scale metabolic model of *Bacillus subtilis* based on SEED annotations. *Genome Biology*. 2009;10(6):R69.
29. Nogales J, Mueller J, Gudmundsson S, Canalejo FJ, Duque E, Monk J, et al. High-quality genome-scale metabolic modelling of *Pseudomonas putida* highlights its broad metabolic capabilities. *Environmental microbiology*. 2020;22(1):255-69.
30. Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R. Genome-scale reconstruction of metabolic network in *Bacillus subtilis* based on high-throughput phenotyping and gene essentiality data. *J Biol Chem*. 2007;282(39):28791-9.
31. Hecht A, Filliben J, Munro SA, Salit M. A minimum information standard for reproducing bench-scale bacterial cell growth and productivity. *Communications Biology*. 2018;1(1):219.
32. Xavier JC, Patil KR, Rocha I. Integration of Biomass Formulations of Genome-Scale Metabolic Models with Experimental Data Reveals Universally Essential Cofactors in Prokaryotes. *Metabolic engineering*. 2017;39:200-8.
33. Beck AE, Hunt KA, Carlson RP. Measuring Cellular Biomass Composition for Computational Biology Applications. *Processes*. 2018;6(5).
34. Long CP, Antoniewicz MR. Quantifying Biomass Composition by Gas Chromatography/Mass Spectrometry. *Analytical Chemistry*. 2014;86(19):9423-7.
35. Neidhardt FC II, Schaechter M. *Physiology of the Bacterial Cell: a Molecular Approach*. USA: Sinauer Associates, Inc.; 1990.
36. Dunn MJ, Little PFR, Subramaniam S, Jorde LB. *Encyclopedia of genetics, genomics, proteomics and bioinformatics*: Wiley; 2005.
37. Banga JR. Optimization in computational systems biology. *BMC Systems Biology*. 2008;2(1):47.
38. Vanderbei RJ. *Linear programming*. 4 ed: Springer; 2015.
39. Schilling CH, Edwards JS, Letscher D, Palsson BØ. Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. *Biotechnology and Bioengineering*. 2000;71(4):286-306.
40. Reznik E, Mehta P, Segrè D. Flux imbalance analysis and the sensitivity of cellular growth to changes in metabolite pools. *PLoS computational biology*. 2013;9(8):e1003195-e.
41. Aucamp DC, Steinberg DI. The Computation of Shadow Prices in Linear Programming. *The Journal of the Operational Research Society*. 1982;33(6):557-65.
42. Edwards JS, Ramakrishna R, Palsson BO. Characterizing the metabolic phenotype: A phenotype phase plane analysis. *Biotechnology and Bioengineering*. 2002;77(1):27-36.
43. Luong JHT, Yerushalmi L, Volesky B. Estimating the maintenance energy and biomass concentration of *Saccharomyces cerevisiae* by continuous calorimetry. *Enzyme and Microbial Technology*. 1983;5(4):291-6.
44. Aung HW, Henry SA, Walker LP. Revising the Representation of Fatty Acid, Glycerolipid, and Glycerophospholipid Metabolism in the Consensus Model of Yeast Metabolism. *Industrial Biotechnology*. 2013;9(4):215-28.
45. Szélliová D, Ruckerbauer DE, Galleguillos SN, Petersen LB, Natter K, Hanscho M, et al. What CHO is made of: Variations in the biomass composition of Chinese hamster ovary cell lines. *Metabolic Engineering*. 2020;61:288-300.

46. Ihssen J, Egli T. Global physiological analysis of carbon- and energy-limited growing *Escherichia coli* confirms a high degree of catabolic flexibility and preparedness for mixed substrate utilization. *Environmental Microbiology*. 2005;7(10):1568-81.
47. Unden G, Bongaerts J. Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 1997;1320(3):217-34.
48. Kim BH, Gadd GM. Anaerobic fermentation. *Bacterial Physiology and Metabolism*. USA: Cambridge University Press; 2008. p. 252-97.
49. van Elsas JD, Semenov AV, Costa R, Trevors JT. Survival of *Escherichia coli* in the environment: fundamental and public health aspects. *The ISME Journal*. 2011;5(2):173-83.
50. Alberts B, Johnson A, Lewis J, Morgan D, Raff M, Roberts K, et al. *Proteins*. *Molecular Biology of the Cell*. 6 ed. USA: Garland Science; 2015. p. 109-72.
51. Rodnina MV. Translation in Prokaryotes. *Cold Spring Harb Perspect Biol*. 2018;10(9):a032664.
52. Kaleta C, Schäuble S, Rinas U, Schuster S. Metabolic costs of amino acid and protein production in *Escherichia coli*. *Biotechnology Journal*. 2013;8(9):1105-14.
53. Akashi H, Gojobori T. Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. *Proceedings of the National Academy of Sciences*. 2002;99(6):3695.
54. Gyaneshwar P, Paliy O, McAuliffe J, Popham DL, Jordan MI, Kustu S. Sulfur and nitrogen limitation in *Escherichia coli* K-12: specific homeostatic responses. *J Bacteriol*. 2005;187(3):1074-90.
55. Brown DR, Barton G, Pan Z, Buck M, Wigneshweraraj S. Nitrogen stress response and stringent response are coupled in *Escherichia coli*. *Nature Communications*. 2014;5(1):4115.
56. Li SH-J, Li Z, Park JO, King CG, Rabinowitz JD, Wingreen NS, et al. *Escherichia coli* translation strategies differ across carbon, nitrogen and phosphorus limitation conditions. *Nat Microbiol*. 2018;3(8):939-47.
57. Bipatnath M, Dennis PP, Bremer H. Initiation and velocity of chromosome replication in *Escherichia coli* B/r and K-12. *J Bacteriol*. 1998;180(2):265-73.
58. Wilkinson JF. Carbon and Energy Storage in Bacteria. *Microbiology*. 1963;32(2):171-6.
59. Berg J, Tymoczko J, Stryer L. Glycogen Is Synthesized and Degraded by Different Pathways. In: 5, editor. *Biochemistry*. New York: W. H. Freeman; 2002.
60. Holme T, Palmstierna H. Changes in Glycogen and Nitrogen-Containing Compounds in *Escherichia coli* B During Growth in Deficient Media. 1. Nitrogen and Carbon Starvation. *Acta Chemica Scandinavica*. 1956;10(4):578-86.
61. Vollmer W, Blanot D, De Pedro MA. Peptidoglycan structure and architecture. *FEMS Microbiology Reviews*. 2008;32(2):149-67.
62. Bertani B, Ruiz N. Function and Biogenesis of Lipopolysaccharides. *EcoSal Plus*. 2018;8(1):10.1128/ecosalplus.ESP-0001-2018.
63. Stumpf P, Barber G. Comparative mechanisms for fatty acid oxidation. *Comparative Biochemistry V1: A Comprehensive Treatise*. 1: Academic press inc.; 1960. p. 75-103.
64. Lugtenberg EJJ, Peters R. Distribution of lipids in cytoplasmic and outer membranes of *Escherichia coli* K12. *Biochimica et Biophysica Acta (BBA) - Lipids and Lipid Metabolism*. 1976;441(1):38-47.
65. Nelson DL, Cox MM. Biosynthesis of Membrane Phospholipids. *Lehninger Principles of Biochemistry*. 7 ed. New York: W. H. Freeman and Company; 2017. p. 830-5.
66. Madigan M, Martinko J, Bender K, Buckler DR, Stahl D. *Molecular Microbiology*. *Brock Biology of Microorganisms*. 14 ed: Pearson Education Limited; 2015. p. 25-48.
67. Luo ML, Jackson RN, Denny SR, Tokmina-Lukaszewska M, Maksimchuk KR, Lin W, et al. The CRISPR RNA-guided surveillance complex in *Escherichia coli* accommodates extended RNA spacers. *Nucleic Acids Res*. 2016;44(15):7385-94.
68. Waters LS, Storz G. Regulatory RNAs in Bacteria. *Cell*. 2009;136(4):615-28.
69. Bremer H, Dennis PP. Modulation of chemical composition and other parameters of the cell by growth rate. *Escherichia coli and Salmonella: cellular and molecular biology*. 1996;2(2):1553-69.

70. Wang B, Wang Z, Chen T, Zhao X. Development of Novel Bioreactor Control Systems Based on Smart Sensors and Actuators. *Frontiers in Bioengineering and Biotechnology*. 2020;8:7.
71. Ziv N, Brandt NJ, Gresham D. The use of chemostats in microbial systems biology. *J Vis Exp*. 2013(80):50168.
72. Wides A, Milo R. Understanding the dynamics and optimizing the performance of chemostat selection experiments. *arXiv preprint arXiv:180600272*. 2018.
73. Coskun O. Separation techniques: Chromatography. *North Clin Istanbul*. 2016;3(2):156-60.
74. Hortin GL, Goldberger BA. Chromatography and Extraction. In: Burtis CA, Ashwood ER, Bruns DE, editors. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. 5 ed: Elsevier; 2014.
75. Bruggink C, Maurer R, Herrmann H, Cavalli S, Hoefler F. Analysis of carbohydrates by anion exchange chromatography and mass spectrometry. *Journal of chromatography A*. 2005;1085(1):104-9.
76. Hardy MR, Townsend RR, Lee YC. Monosaccharide analysis of glycoconjugates by anion exchange chromatography with pulsed amperometric detection. *Analytical Biochemistry*. 1988;170(1):54-62.
77. Cataldi TRI, Campa C, Angelotti M, Bufo SA. Isocratic separations of closely-related mono- and disaccharides by high-performance anion-exchange chromatography with pulsed amperometric detection using dilute alkaline spiked with barium acetate. *Journal of Chromatography A*. 1999;855(2):539-50.
78. Valkó K, Bevan C, Reynolds D. Chromatographic Hydrophobicity Index by Fast-Gradient RP-HPLC: A High-Throughput Alternative to log P/log D. *Analytical Chemistry*. 1997;69(11):2022-9.
79. Pitt JJ. Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry. *Clin Biochem Rev*. 2009;30(1):19-34.
80. Nickelson B, Colón I. *Liquid-Liquid and Solid-Phase Extraction Techniques. Sample Preparation of Pharmaceutical Dosage Forms*. Boston, MA: Springer; 2011.
81. Stanley F. Principles of Basic Techniques and Laboratory Safety. In: Burtis CA, Ashwood ER, Bruns DE, editors. *Tietz Textbook of Clinical Chemistry and Molecular Diagnostics*. 5 ed: Elsevier; 2014. p. 207-32.
82. Sündermann A, Eggers LF, Schwudke D. Liquid Extraction: Bligh and Dyer. In: Wenk MR, editor. *Encyclopedia of Lipidomics*. Dordrecht: Springer Netherlands; 2016. p. 1-4.
83. García-Alegría AM, Anduro-Corona I, Pérez-Martínez CJ, Corella-Madueño MAG, Rascón-Durán ML, Astiazaran-García H. Quantification of DNA through the NanoDrop Spectrophotometer: Methodological Validation Using Standard Reference Material and Sprague Dawley Rat and Human DNA. *International Journal of Analytical Chemistry*. 2020.
84. Campbell IT, Brand U. *Spectrophotometry. Geochemistry*. Dordrecht: Springer Netherlands; 1998. p. 587-8.
85. Akash MSH, Rehman K. Ultraviolet-Visible (UV-VIS) Spectroscopy. In: Akash MSH, Rehman K, editors. *Essentials of Pharmaceutical Analysis*. Singapore: Springer Singapore; 2020. p. 29-56.
86. Nwokeoji AO, Kilby PM, Portwood DE, Dickman MJ. Accurate Quantification of Nucleic Acids Using Hypochromicity Measurements in Conjunction with UV Spectrophotometry. *Analytical Chemistry*. 2017;89(24):13567-74.
87. Kim SK, Byun SH, Lee BM. Effects of Chemical Carcinogens and Physicochemical Factors on the UV Spectrophotometric Determination of DNA *Journal of Toxicology and Environmental Health*. 2005;68:2081-95.
88. Laws GM, Adams SP. Measurement of 8-OHdG in DNA by HPLC/ECD: the importance of DNA purity. *BioTechniques*. 1996;20(1):36-8.
89. Myers JA, Curtis BS, Curtis WR. Improving accuracy of cell and chromophore concentration measurements using optical density. *BMC Biophysics*. 2013;6(1):4.
90. Simensen V, Schulz C, Karlsen E, Bråtelund S, Burgos I, Thorfinnsdottir LB, et al. Quantification of macromolecular biomass composition for constraint-based metabolic modeling (manuscript in preparation). 2021.

91. Rühmann BS, Jochen; Sieber, Volker. Automated Modular High Throughput Exopolysaccharide Screening Platform Coupled with Highly Sensitive Carbohydrate Fingerprint Analysis. *JoVE*. 2016(110):e53249.
92. Eggers LF, Schwudke D. Liquid Extraction: Folch. In: Wenk MR, editor. *Encyclopedia of Lipidomics*. Dordrecht: Springer Netherlands; 2016. p. 1-6.
93. Green MR, Sambrook J. Isolation of High-Molecular-Weight DNA Using Organic Solvents. *Cold Spring Harbor protocols*. 2017;2017(4):pdb.prot093450.
94. Desjardins PC, Deborah. NanoDrop Microvolume Quantitation of Nucleic Acids. *JoVE*. 2010(45):e2565.
95. Benthin S, Nielsen J, Villadsen J. A simple and reliable method for the determination of cellular RNA content. *Biotechnology Techniques*. 1991;5(1):39-42.
96. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nature Protocols*. 2019;14(3):639-702.
97. van Rossum G, Drake FL. *Python 3 Reference Manual*.: Scotts Valley, CA: CreateSpace.; 2009.
98. Romero-Kutzner V, Packard TT, Berdalet E, Roy SO, Gagné JP, Gómez M. Respiration quotient variability: bacterial evidence. *Marine Ecology Progress Series*. 2015;519:47-59.
99. Rodrigues RMNV, Williams PJIB. Heterotrophic bacterial utilization of nitrogenous and nonnitrogenous substrates, determined from ammonia and oxygen fluxes. *Limnology and Oceanography*. 2001;46(7):1675-83.
100. Mustatea G, Ungureanu E, Iorga E. Protein Acidic Hydrolysis for Amino Acid Analysis in Food - Progress over Time: a Short Review. *Juornal of Hygienic Engineering and Design*. 2019;26:81-8.
101. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res*. 2016;44(D1):D67-D72.
102. Li L, Nelson CJ, Trösch J, Castleden I, Huang S, Millar AH. Protein Degradation Rate in *Arabidopsis thaliana* Leaf Growth and Development. *Plant Cell*. 2017;29(2):207-28.
103. Albers E, Larsson C, Lidén G, Niklasson C, Gustafsson L. Influence of the Nitrogen Source on *Saccharomyces cerevisiae* Anaerobic Growth and Product Formation. *Applied and environmental microbiology*. 1996;62:3187-95.
104. Bell JG, Henderson RJ, Tocher DR, McGhee F, Dick JR, Porter A, et al. Substituting Fish Oil with Crude Palm Oil in the Diet of Atlantic Salmon (*Salmo salar*) Affects Muscle Fatty Acid Composition and Hepatic Fatty Acid Metabolism. *The Journal of Nutrition*. 2002;132(2):222-30.
105. Roe JH, Dailey RE. Determination of glycogen with the anthrone reagent. *Analytical Biochemistry*. 1966;15(2):245-50.
106. Boman HG, Monner DA. Characterization of lipopolysaccharides from *Escherichia coli* K-12 mutants. *J Bacteriol*. 1975;121(2):455-64.
107. Klein G, Raina S. Regulated Control of the Assembly and Diversity of LPS by Noncoding sRNAs. *Biomed Res Int*. 2015;2015:153561-.
108. Wang L, Wang Q, Reeves PR. The variation of O antigens in gram-negative bacteria. *Sub-cellular biochemistry*. 2010;53:123-52.
109. Rühmann B, Schmid J, Sieber V. Fast carbohydrate analysis via liquid chromatography coupled with ultra violet and electrospray ionization ion trap detection in 96-well format. *Journal of Chromatography A*. 2014;1350:44-50.
110. Whitfield C, Kaniuk N, Firdich E. Molecular insights into the assembly and diversity of the outer core oligosaccharide in lipopolysaccharides from *Escherichia coli* and *Salmonella*. *Journal of endotoxin research*. 2003;9(4):244-9.
111. Plata M, Koch C, Wechselberger P, Herwig C, Lendl B. Determination of carbohydrates present in *Saccharomyces cerevisiae* using mid-infrared spectroscopy and partial least squares regression. *Analytical and bioanalytical chemistry*. 2013;405.
112. Teng N, Wang J, Chen T, Wu X, Wang Y, Lin J. Elevated CO₂ induces physiological, biochemical and structural changes in leaves of *Arabidopsis thaliana*. *New Phytologist*. 2006;172(1):92-103.

113. Calenge F, Saliba-Colombani V, Mahieu S, Loudet O, Daniel-Vedele F, Krapp A. Natural variation for carbohydrate content in *Arabidopsis*. Interaction with complex traits dissected by quantitative genetics. *Plant Physiol.* 2006;141(4):1630-43.
114. Bailly C, Hecquet P-E, Kouach M, Thuru X, Goossens J-F. Chemical reactivity and uses of 1-phenyl-3-methyl-5-pyrazolone (PMP), also known as edaravone. *Bioorganic & Medicinal Chemistry.* 2020;28(10):115463.
115. Tsuneda S, Aikawa H, Hayashi H, Yuasa A, Hirata A. Extracellular polymeric substances responsible for bacterial adhesion onto solid surface. *FEMS Microbiology Letters.* 2003;223(2):287-92.
116. Kostakioti M, Hadjifrangiskou M, Hultgren SJ. Bacterial Biofilms: Development, Dispersal, and Therapeutic Strategies in the Dawn of the Postantibiotic Era. *Cold Spring Harbor Perspectives in Medicine.* 2013;3(4).
117. Ebskamp MJM, van der Meer IM, Spronk BA, Weisbeek PJ, Smeeckens SCM. Accumulation of Fructose Polymers in Transgenic Tobacco. *Bio/Technology.* 1994;12(3):272-5.
118. Bartosova Z, Gonzalez SV, Voigt A, Bruheim P. High Throughput Semiquantitative UHPSFC–MS/MS Lipid Profiling and Lipid Class Determination. *Journal of Chromatographic Science.* 2021.
119. Baron CP, Kjærsgård IVH, Jessen F, Jacobsen C. Protein and Lipid Oxidation during Frozen Storage of Rainbow Trout (*Oncorhynchus mykiss*). *Journal of Agricultural and Food Chemistry.* 2007;55(20):8118-25.
120. Hazlewood GP, Dawson RM. A phospholipid-deacylating system of bacteria active in a frozen medium. *Biochem J.* 1976;153(1):49-53.
121. Ishinaga M, Kanamoto R, Kito M. Distribution of Phospholipid Molecular Species in Outer and Cytoplasmic Membranes of *Escherichia coli*. *The Journal of Biochemistry.* 1979;86(1):161-5.
122. Gigon A, Matos A-R, Laffray D, Zuily-Fodil Y, Pham-Thi A-T. Effect of Drought Stress on Lipid Metabolism in the Leaves of *Arabidopsis thaliana* (Ecotype Columbia). *Annals of Botany.* 2004;94(3):345-51.
123. Arneborg N, Høy C-E, Jørgensen OB. The effect of ethanol and specific growth rate on the lipid content and composition of *Saccharomyces cerevisiae* grown anaerobically in a chemostat. *Yeast.* 1995;11(10):953-9.
124. O'Neill M, McPartlin J, Arthure K, Riedel S, McMillan ND. Comparison of the TLDA with the Nanodrop and the reference Qubit system. *Journal of Physics: Conference Series.* 2011;307.
125. T042-Technical Bulletin NanoDrop Spectrophotometers. 260/280 and 260/230 Ratios: Thermo Fisher Scientific - NanoDrop products; 2009.
126. Wilfinger WW, Mackey K, Chomczynski P. Effect of pH and Ionic Strength on the Spectrophotometric Assessment of Nucleic Acid Purity. *BioTechniques.* 1997;22(3):474-81.
127. T123 – Technical Bulletin NanoDrop Lite. Interpretation of Nucleic Acid 260/280 Ratios: Thermo Scientific NanoDrop Products; 2012.
128. Matlock B. Assessment of Nucleic Acid Purity. Technical Note 52646: Thermo Fisher Scientific; 2015.
129. Łabudzińska A, Gorczyńska K. The UV difference spectra as a characteristic feature of phenols and aromatic amines. *Journal of Molecular Structure.* 1995;349:469-72.
130. Brunschede H, Dove TL, Bremer H. Establishment of exponential growth after a nutritional shift-up in *Escherichia coli* B/r: accumulation of deoxyribonucleic acid, ribonucleic acid, and protein. *J Bacteriol.* 1977;129(2):1020-33.
131. Quantitation of residual DNA in final product biopharmaceuticals using the Threshold Total DNA assay system. Assay validation and determination of detection limit. Threshold Immunoassay System - DNA Application note: molecular devices; 2010.
132. Sandaa RA, Enger Ø, Torsvik V. Rapid method for fluorometric quantification of DNA in soil. *Soil Biology and Biochemistry.* 1998;30(2):265-8.
133. Nomura M, Morgan EA, Jaskunas SR. Genetics of bacterial ribosomes. *Annual review of genetics.* 1977;11(1):297-347.

134. Thüring K, Schmid K, Keller P, Helm M. Analysis of RNA modifications by liquid chromatography–tandem mass spectrometry. *Methods*. 2016;107:48-56.
135. McKee MJ, Knowles CO, Buckler DR. Effects of aluminum on the biochemical composition of Atlantic salmon. *Archives of Environmental Contamination and Toxicology*. 1989;18(1):243-8.
136. Ertugay N, Hamamci H. Continuous cultivation of bakers' yeast: change in cell composition at different dilution rates and effect of heat stress on trehalose level. *Folia microbiologica*. 1997;42(5):463-7.
137. Deutch C. Analysis of the RNA Content of the Yeast *Saccharomyces cerevisiae*. *The American Biology Teacher*. 2008;70:537-45.
138. Schulz C, Kumelj T, Karlsen E, Almaas E. Genome-scale metabolic modelling when changes in environmental conditions affect biomass composition. *PLOS Computational Biology*. 2021;17(5):e1008528.
139. Schulze KL, Lipe RS. Relationship between substrate concentration, growth rate, and respiration rate of *Escherichia coli* in continuous culture. *Archiv für Mikrobiologie*. 1964;48(1):1-20.
140. Gogate PR, Beenackers AACM, Pandit AB. Multiple-impeller systems with a special emphasis on bioreactors: a critical review. *Biochemical Engineering Journal*. 2000;6(2):109-44.
141. Kayser A, Weber J, Hecht V, Rinas U. Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. I. Growth-rate-dependent metabolic efficiency at steady state. *Microbiology*. 2005;151(3):693-706.
142. Kolkman A, Daran-Lapujade P, Fullaondo A, Olsthoorn MMA, Pronk JT, Slijper M, et al. Proteome analysis of yeast response to various nutrient limitations. *Molecular Systems Biology*. 2006;2(1):2006.0026.
143. Vasilakou E, van Loosdrecht MCM, Wahl SA. *Escherichia coli* metabolism under short-term repetitive substrate dynamics: adaptation and trade-offs. *Microbial Cell Factories*. 2020;19(1):116.
144. Chang Y-Y, Cronan JE. Membrane cyclopropane fatty acid content is a major factor in acid resistance of *Escherichia coli*. *Molecular Microbiology*. 1999;33(2):249-59.
145. Dubničková M, Řezanka T, Koščová H. Adaptive changes in fatty acids of *E. coli* strains exposed to a quaternary ammonium salt and an amine oxide. *Folia microbiologica*. 2006;51(5):371-4.
146. Santos IC, Smuts J, Choi WS, Kim Y, Kim SB, Schug KA. Analysis of bacterial FAMES using gas chromatography - vacuum ultraviolet spectroscopy for the identification and discrimination of bacteria. 2018(1873-3573 (Electronic)).
147. Huang Q, Kaiser K, Benner R. A simple high performance liquid chromatography method for the measurement of nucleobases and the RNA and DNA content of cellular material. *Limnology and Oceanography: Methods*. 2012;10(8):608-16.
148. Rampler E, Coman C, Hermann G, Sickmann A, Ahrends R, Koellensperger G. LILY-lipidome isotope labeling of yeast: in vivo synthesis of ^{13}C labeled reference lipids for quantification by mass spectrometry. *Analyst*. 2017;142(11):1891-9.
149. Matsuda S, Kasahara T. Simultaneous and absolute quantification of nucleoside triphosphates using liquid chromatography-triple quadrupole tandem mass spectrometry. *Genes Environ*. 2018;40:13-.
150. Metting FB. Biodiversity and application of microalgae. *Journal of Industrial Microbiology*. 1996;17(5):477-89.
151. Széliová D, Štor J, Thiel I, Weinguny M, Hanscho M, Lhota G, et al. Inclusion of maintenance energy improves the intracellular flux predictions of CHO. *bioRxiv*. 2020:2020.12.21.423738.
152. Monk JM, Koza A, Campodonico MA, Machado D, Seoane JM, Palsson BO, et al. Multi-omics Quantification of Species Variation of *Escherichia coli* Links Molecular Features with Strain Phenotypes. *Cell systems*. 2016;3(3):238-51.e12.

Appendix A: NMR based quantification of glucose, lactate and glutamine in cell culture media

Purpose:

Quantify glucose, lactate and glutamine in cell culture media by NMR. This is interesting as glucose and glutamine are the main carbon and energy sources in cell culture, and a proportion of the glucose carbon is excreted as lactate. Samples of 2,5 mL are collected during metabolome sampling, and stored at -20°C. The samples are concentrated by freeze-drying, reconstitution in d₂O and quantified from an external standard using the processing software Topspin.

Materials:

- Freeze-dryer
- Vortex
- Pipette and pipette tips (1000 µL)
- NMR tubes (5 mm)
- Analytical scale
- 15 mL centrifuge tubes
- Computer with TopSpin installed

Chemicals:

- d₂O
- Creatine

Wet lab protocol:

Prepare samples for NMR

- 1) Pre-cool freeze-dryer
- 2) Collect samples from freezer: 2,5 mL of sample in 15 mL centrifuge tube
- 3) Replace the lids with lids with holes
- 4) Place tubes in liquid nitrogen
- 5) Place tubes in pre-cooled freeze dryer. Leave until dry
- 6) Add 600 μL d_2O to freeze-dried sample, vortex thoroughly
- 7) Transfer 500 μL of sample to labeled NMR tube, cap the tube
- 8) Store the samples cold (4°C)

Prepare external standard for NMR

Creatine (131,13 g/mol) is used as an external standard to allow for quantification. Prepare a 70 mM solution with d_2O :

$$0.070 \text{ mol/L} \times 131.13 \text{ g/mol} \times 0.005 \text{ L} = 0.0458 \text{ g} = 45.9 \text{ mg} \quad (\text{A-1})$$

- 1) Weigh out 45,9 mg creatine
- 2) Dissolve in 5 mL d_2O in 15 mL centrifuge tube, vortex thoroughly
- 3) Transfer 500 μL of standard to labeled NMR tube, cap the tube

Record NMR spectra

Record spectra: samples and the external standard on 400 MHz instrument.

- 1) Log into Icon NMR to program recording:
 - Choose available slots (remove and delete samples that are “finished”)
 - Type name of sample
 - Choose method: “N PROF_1H” (= noesyggpr1d, 32 scans)
 - Choose solvent: d_2O
 - Program run in randomized order
- 2) Wipe NMR tubes with paper wipes, place in blue “spinners” using the depth meter
- 3) Place samples with spinner in the slots programed in Icon NMR
- 4) Icon NMR: Submit the samples
- 5) When spectra is recorded, store the samples cold (4°C)

Dry lab protocol:

Process NMR spectra using TopSpin. Topspin download:

<https://www.bruker.com/nc/service/support-upgrades/software-downloads/nmr/free-topspin-processing.html>

Quantify peaks using quantification tool “ERETIC 2”. Table **A.1** lists which peaks to use for quantification, the number of atoms/signal, molar masses and which region to integrate.

TopSpin: Process parameter settings

Must be performed for all spectra

- 1) Copy raw spectra from \\felles.ansatt.ntnu.no\nt-unix\felles\nmrdata\nmr400ny
- 2) Open spectra in TopSpin
- 3) Command line: Write “efp”: Fourier transforms spectrum with line broadening 0,3 Hz
- 4) Command line: Write “apk”: Phases spectrum. Can be performed manually if automatic phasing is bad (External standard: use first and second creatine peak as 0 and 1):
 - Enter “Process” menu → Enter “Adjust phase” menu → choose α -glucose peak at ~5,2 ppm → Press “0”, adjust phase by moving cursor → Choose lactate peak at ~1,3 ppm → Press “1”, adjust phase by moving cursor → Press “Return and save phased spectrum”
- 5) Command line: Write “abs n”: Corrects baseline

TopSpin: ERETIC 2 calibration

- 1) Open creatine spectrum (processed)
- 2) Enter “Process” menu → Enter “Integrate” menu → Press “Define new region using cursor” → Integrate signal
- 3) Right click → Press “Eretic” → Press “Define as Eretic Reference”
 - Define concentration, number of atoms, region start and end, sample volume, molar mass (Table **A.1**)
- 4) Press “Return, save regions”

TopSpin: Quantification with external standard

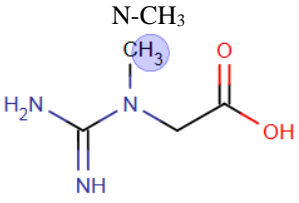
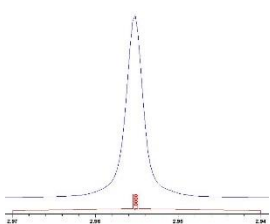
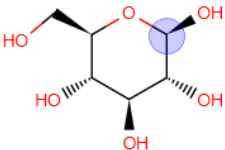
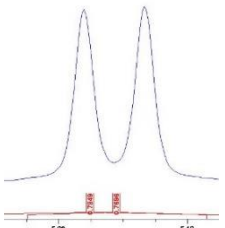
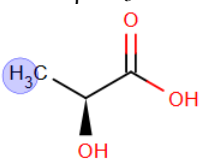
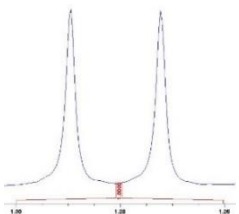
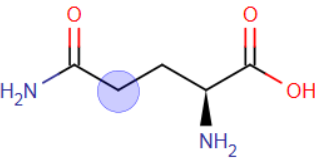
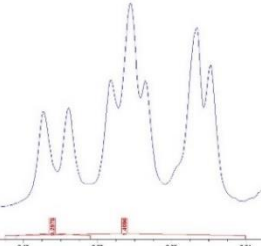
- 1) Open processed spectra
- 2) Zoom in on glucose, lactate or glutamine peak
- 3) Perform manual baseline correction close to each peak of interest:
 - Enter “Process” menu → Enter “Advanced” menu → Press “Correct Baseline” → Press “Manual correction mode” → Press “A” “B” and/or “C” to adjust baseline by moving cursor → Press “Return, save regions”
- 4) Enter “Process” menu → Enter “Integrate” menu → Press “Define new region using cursor” → Integrate signal
- 5) Right click → Press “Eretic” → Press “Calculate concentration”
 - Add new compound by clicking “+”
 - Define sample volume, number of atoms, region start and end, molar mass (Table **A.1**)
- 6) Press “Return, save regions”
- 7) Press the “Integrals” tab, record the concentration (Eretic)
 - Right click on purple header, tic “Concentration (Eretic)” to view this option

Calculate original concentration

Appendix A

- 1) Perform corrections according to Table **A.1**
- 2) Correct for sample being concentrated: 2,5 mL reconstituted in 0,6 mL:
Original concentration = Measured concentration/(2,5/0,6)

Table A.1: TopSpin Eretic 2 quantification rules for cell culture media samples

Compound	Molar mass (g/mol)	Protons for quantification	Peak for quantification (~ppm)	Integrate region	Atoms/signal	Corrections
Creatine	131,1332	 <p>N-CH₃</p>	3	 <p>From middle of signal: 0,015 ppm to each side</p>	3	
Glucose	180,1559	 <p>C1</p>	5, 2	 <p>From middle of signal: 0,02 ppm to each side</p>	1	Correct for anomers: - α -glucose peak is integrated - α -glucose: 36% - Measured concentration = [α -glucose] / 36 * 100
Lactate	90,0795	 <p>βCH₃</p>	1,3	 <p>From middle of signal: 0,02 ppm to each side</p>	3	
Glutamine	146,1445	 <p>γCH₂</p>	2,4	 <p>From middle of two peaks of multiplet with highest ppm: 0,01 ppm to each side</p>	2	Correct for integrating two peaks of the multiplet: - The two peaks with highest ppm: no overlap with other signals - Check what proportion of the multiplet the two peaks with highest ppm constitute - Measured concentration = [Two peaks integrated] / percentage of multiplet integrated * 100

Appendix B: Amino acid distribution

The measured amino acid concentration in all *E. coli* samples are shown in Table B.1. And the relative amino acid distribution is shown in Table B.2

Table B.1: Measured amino acid concentration in all *E. coli* samples. The concentrations shown are averages of three technical replicates. Concentrations are given in mass % of total

Amino acid	<i>E. coli</i> C-lim 0.4	<i>E. coli</i> C-lim 0.2	<i>E. coli</i> C-lim 0.1	<i>E. coli</i> N-lim 0.4	<i>E. coli</i> N-lim 0.2	<i>E. coli</i> N-lim 0.1	<i>E. coli</i> U-lim
Aspartate	3.91 ± 0.85	4.01 ± 0.46	4.84 ± 1.45	4.67 ± 1.89	3.51 ± 0.22	3.72 ± 0.50	2.86 ± 0.39
Glutamate	4.67 ± 0.63	4.74 ± 0.57	5.52 ± 1.54	5.66 ± 1.82	3.93 ± 0.38	5.08 ± 0.44	2.59 ± 0.33
Asparagine	2.97 ± 0.65	3.04 ± 0.35	3.68 ± 1.10	3.54 ± 1.43	2.67 ± 0.16	2.82 ± 0.38	2.17 ± 0.30
Histidine	1.11 ± 0.20	1.09 ± 0.08	1.34 ± 0.40	1.35 ± 0.40	0.96 ± 0.08	1.09 ± 0.17	1.18 ± 0.11
Serine	2.08 ± 0.16	2.22 ± 0.31	2.70 ± 0.88	2.55 ± 1.01	1.86 ± 0.20	2.19 ± 0.22	1.76 ± 0.26
Glutamine	3.57 ± 0.48	3.63 ± 0.44	4.22 ± 1.18	4.33 ± 1.39	3.01 ± 0.29	3.89 ± 0.34	1.98 ± 0.25
Glycine	1.39 ± 0.15	1.44 ± 0.16	1.60 ± 0.47	1.60 ± 0.52	1.21 ± 0.16	1.11 ± 0.23	1.21 ± 0.06
Arginine	2.85 ± 0.31	2.95 ± 0.33	3.29 ± 0.97	3.30 ± 1.06	2.48 ± 0.33	2.28 ± 0.46	2.47 ± 0.12
Threonine	2.38 ± 0.26	2.62 ± 0.26	3.24 ± 0.88	3.13 ± 1.20	2.26 ± 0.28	2.51 ± 0.31	2.26 ± 0.29
Alanine	4.70 ± 1.13	4.76 ± 0.48	5.69 ± 1.70	5.45 ± 1.99	4.09 ± 0.26	4.54 ± 0.59	4.30 ± 0.93
Tyrosine	1.87 ± 0.36	1.94 ± 0.17	2.75 ± 0.77	2.27 ± 0.86	1.91 ± 0.20	1.99 ± 0.40	0.86 ± 0.18
Methionine	0.64 ± 1.09	0.79 ± 0.71	1.04 ± 0.86	0.92 ± 1.46	0.47 ± 0.73	0.48 ± 0.77	0.44 ± 0.50
Valine	3.20 ± 0.24	3.53 ± 0.26	4.20 ± 1.18	4.11 ± 1.25	2.98 ± 0.43	3.33 ± 0.31	1.37 ± 1.20
Phenylalanine	2.56 ± 0.41	2.69 ± 0.11	3.49 ± 0.94	3.18 ± 1.15	2.41 ± 0.22	2.77 ± 0.29	3.07 ± 1.31
Isoleucine	2.64 ± 0.16	2.90 ± 0.26	3.55 ± 1.01	3.37 ± 1.05	2.48 ± 0.41	3.11 ± 0.16	1.78 ± 0.12
Leucine	5.00 ± 0.82	5.15 ± 0.47	6.17 ± 1.74	6.12 ± 2.40	4.49 ± 0.42	4.86 ± 0.68	4.03 ± 1.46
Lysine	3.20 ± 0.88	2.74 ± 1.21	3.70 ± 1.29	3.28 ± 0.65	2.90 ± 1.30	3.76 ± 0.97	3.44 ± 1.44
Cysteine	0.42 ± 0.12	0.46 ± 0.12	0.82 ± 0.19	0.59 ± 0.22	0.49 ± 0.10	0.72 ± 0.03	0.52 ± 0.51
Proline	2.03 ± 0.31	2.11 ± 0.13	2.65 ± 0.74	2.49 ± 0.87	1.86 ± 0.19	2.16 ± 0.21	1.66 ± 0.15
Tryptophan	1.12 ± 0.25	1.19 ± 0.17	1.89 ± 0.47	1.49 ± 0.53	1.19 ± 0.19	1.62 ± 0.09	1.19 ± 0.85
Total	52.29 ± 7.65	54.00 ± 3.91	66.40 ± 18.65	63.38 ± 21.95	47.17 ± 4.76	54.05 ± 5.44	41.12 ± 1.67

Table B.2: Relative amino acid distribution in all *E. coli* samples given in mass % of total protein.

Amino acid	<i>E. coli</i> C-lim 0.4	<i>E. coli</i> C-lim 0.2	<i>E. coli</i> C-lim 0.1	<i>E. coli</i> N-lim 0.4	<i>E. coli</i> N-lim 0.2	<i>E. coli</i> N-lim 0.1	<i>E. coli</i> U-lim
Aspartate	7.47	7.43	7.30	7.36	7.45	6.88	6.95
Glutamate	8.92	8.78	8.31	8.93	8.34	9.40	6.29
Asparagine	5.67	5.64	5.54	5.59	5.65	5.22	5.28
Histidine	2.11	2.03	2.01	2.14	2.04	2.01	2.87
Serine	3.97	4.10	4.07	4.02	3.94	4.06	4.27
Glutamine	6.83	6.72	6.36	6.83	6.38	7.20	4.81
Glycine	2.65	2.66	2.42	2.53	2.56	2.06	2.93
Arginine	5.45	5.47	4.96	5.20	5.26	4.23	6.02
Threonine	4.55	4.85	4.88	4.94	4.78	4.64	5.49
Alanine	8.99	8.81	8.57	8.60	8.68	8.40	10.45
Tyrosine	3.57	3.59	4.14	3.58	4.06	3.67	2.09
Methionine	1.22	1.46	1.56	1.45	0.99	0.89	1.08
Valine	6.12	6.54	6.32	6.48	6.33	6.17	3.34
Phenylalanine	4.89	4.97	5.26	5.02	5.10	5.13	7.45
Isoleucine	5.05	5.37	5.34	5.31	5.26	5.75	4.33
Leucine	9.57	9.54	9.30	9.66	9.52	9.00	9.79
Lysine	6.12	5.08	5.57	5.17	6.15	6.96	8.37
Cysteine	0.81	0.85	1.23	0.94	1.04	1.33	1.26
Proline	3.89	3.91	4.00	3.92	3.95	4.00	4.04
Tryptophan	2.14	2.21	2.84	2.35	2.52	3.00	2.88
Total	100	100	100	100	100	100	100

Appendix C: Ribose comparison between carbohydrates and RNA

The average ribose content in all *E. coli* cultures, measured as carbohydrates and as RNA are shown in Table C.1. Figure C.1 shows a linear regression between ribose measured as carbohydrates and as RNA. For calculations see supplementary material S2.

Table C.1: Ribose content in all *E. coli* cultures measured as RNA and carbohydrates. Values are the average of three replicates given in mass % of total CDW.

Sample	Measured as RNA	Measured as carbohydrate
<i>E. coli</i> C-lim 0.4	6.77 ± 0.24	5.67 ± 0.21
<i>E. coli</i> C-lim 0.2	4.58 ± 0.29	3.23 ± 0.15
<i>E. coli</i> C-lim 0.1	4.59 ± 0.15	3.25 ± 0.03
<i>E. coli</i> N-lim 0.4	6.59 ± 0.26	5.13 ± 0.19
<i>E. coli</i> N-lim 0.2	5.40 ± 0.17	4.03 ± 0.13
<i>E. coli</i> N-lim 0.1	3.38 ± 0.44	2.10 ± 0.22
<i>E. coli</i> U-lim	7.34 ± 0.17	5.75 ± 0.38

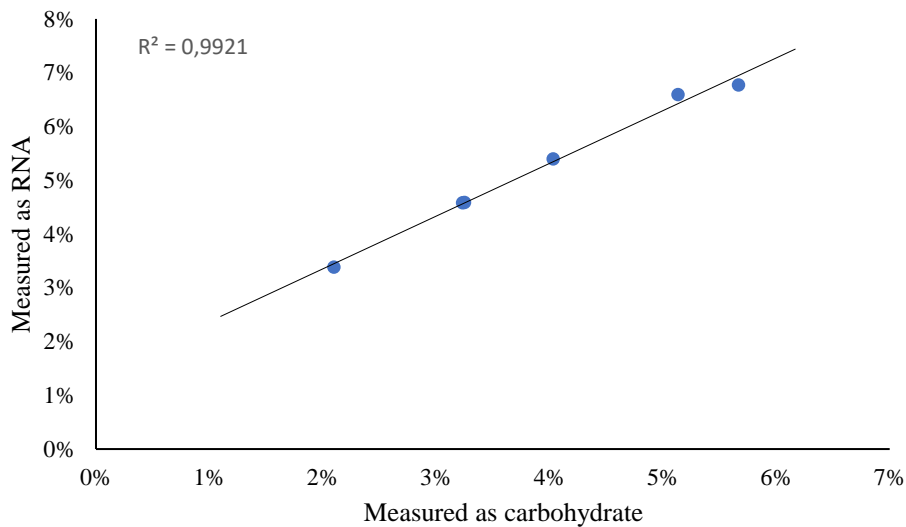


Figure C.1: Linear regression between average ribose content measured as RNA and carbohydrate in all *E. coli* cultures.

Appendix D: Protein and carbohydrate functions and chemical formulas

The general protein function is shown in Equation D-1 and the general protein chemical formula is shown in Equation D-2. The values of the variables $a - t$ and $u - v$ for each *E. coli* sample are shown in Table D.1.

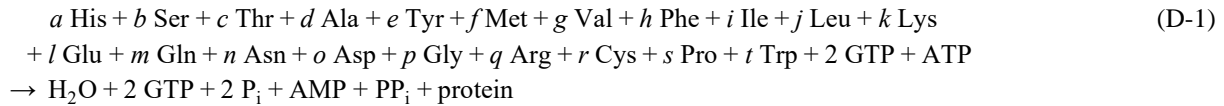


Table D.1: Stoichiometric coefficients of protein functions.

	<i>E. coli</i> C-lim 0.4	<i>E. coli</i> C-lim 0.2	<i>E. coli</i> C-lim 0.1	<i>E. coli</i> N-lim 0.4	<i>E. coli</i> N-lim 0.2	<i>E. coli</i> N-lim 0.1	<i>E. coli</i> U-lim
<i>a</i>	0.0698	0.0696	0.0686	0.0685	0.0704	0.0652	0.0653
<i>b</i>	0.0750	0.0733	0.0700	0.0754	0.0701	0.0799	0.0526
<i>c</i>	0.0535	0.0533	0.0526	0.0524	0.0539	0.0499	0.0500
<i>d</i>	0.0167	0.0161	0.0163	0.0171	0.0161	0.0160	0.0226
<i>e</i>	0.0497	0.0507	0.0505	0.0495	0.0490	0.0509	0.0527
<i>f</i>	0.0578	0.0565	0.0540	0.0581	0.0540	0.0616	0.0406
<i>g</i>	0.0506	0.0503	0.0459	0.0483	0.0486	0.0392	0.0553
<i>h</i>	0.0379	0.0378	0.0344	0.0362	0.0364	0.0294	0.0415
<i>i</i>	0.0580	0.0615	0.0625	0.0623	0.0608	0.0595	0.0691
<i>j</i>	0.1358	0.1338	0.1306	0.1306	0.1328	0.1288	0.1574
<i>k</i>	0.0261	0.0263	0.0306	0.0262	0.0299	0.0271	0.0153
<i>l</i>	0.0089	0.0117	0.0119	0.0088	0.0090	0.0071	0.0092
<i>m</i>	0.0673	0.0713	0.0693	0.0714	0.0690	0.0681	0.0354
<i>n</i>	0.0359	0.0366	0.0389	0.0369	0.0376	0.0381	0.0548
<i>o</i>	0.0488	0.0512	0.0513	0.0512	0.0502	0.0557	0.0411
<i>p</i>	0.0914	0.0909	0.0893	0.0917	0.0913	0.0867	0.0924
<i>q</i>	0.0523	0.0437	0.0486	0.0482	0.0511	0.0598	0.0702
<i>r</i>	0.0084	0.0091	0.0131	0.0098	0.0109	0.0142	0.0132
<i>s</i>	0.0433	0.0435	0.0448	0.0437	0.0441	0.0451	0.0448
<i>t</i>	0.0124	0.0129	0.0167	0.0136	0.0147	0.0176	0.0167
<i>v</i>	4.85	4.85	4.91	4.87	4.89	4.93	4.85
<i>w</i>	7.60	7.58	7.62	7.62	7.62	7.69	7.56
<i>x</i>	1.32	1.31	1.31	1.32	1.32	1.31	1.35
<i>y</i>	1.45	1.45	1.43	1.45	1.44	1.453	1.38
<i>z</i>	0.017	0.021	0.025	0.000	0.020	0.021	0.022

Appendix D

The general carbohydrate function is shown in Equation D-3 and the general carbohydrate chemical formula is shown in Equation D-4. The value of the variables $A - F$ and $V - Y$ for each *E. coli* sample are shown in Table D.2.

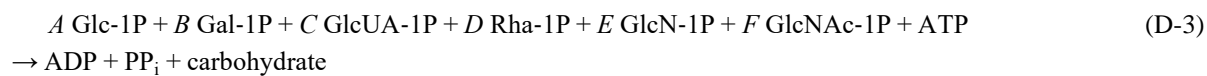


Table D.2: Stoichiometric coefficients of carbohydrate functions.

	<i>E. coli</i> C-lim 0.4	<i>E. coli</i> C-lim 0.2	<i>E. coli</i> C-lim 0.1	<i>E. coli</i> N-lim 0.4	<i>E. coli</i> N-lim 0.2	<i>E. coli</i> N-lim 0.1	<i>E. coli</i> U-lim
A	0.670	0.643	0.543	0.913	0.950	0.591	0.734
B	0.133	0.156	0.111	0.036	0.022	0.013	0.118
C	0	0	0	0	0	0.038	0
D	0	0	0.172	0	0	0.267	0
E	0.083	0.084	0.073	0.021	0.012	0.038	0.062
F	0.114	0.117	0.101	0.030	0.016	0.052	0.086
V	6.23	6.23	6.20	6.06	6.03	6.11	6.17
W	10.5	10.5	10.4	10.1	10.1	10.1	10.4
X	0.197	0.201	0.173	0.051	0.028	0.090	0.148
Y	4.92	4.92	4.76	4.98	4.99	4.73	4.94

Appendix E: Compounds implemented from *iML1515*

All compounds that were implemented from the *iML1515* wild-type BOF are shown in Table E.1. The stoichiometric coefficients are scaled to sum up to 0.039 g.

Table E.1: Compounds implemented from *iML1515* with scaled stoichiometric coefficients. Stoichiometric coefficients are given in [$\mu\text{mol/g CDW h}^{-1}$]. Negative stoichiometric coefficients mean that the compounds are consumed in the BOF.

Compound ID	Name	Stoichiometry
10fthf_c	10-Formyltetrahydrofolate	-0.0489
2fe2s_c	[2Fe-2S] iron-sulphur cluster	-0.0057
2ohph_c	2-Octaprenyl-6-hydroxyphenol	-0.0489
4fe4s_c	[4Fe-4S] iron-sulphur cluster	-0.0543
amet_c	S-Adenosyl-L-methionine	-0.0489
btn_c	Biotin	-0.0004
ca2_c	Calcium	-1.0851
cl_c	Chloride	-1.0851
coa_c	Coenzyme A	-0.0368
cobalt2_c	Co ²⁺	-0.0055
cu2_c	Copper	-0.1477
fad_c	Flavin adenine dinucleotide oxidized	-0.0489
fe2_c	Fe ²⁺ mitochondria	-1.3998
fe3_c	Iron (Fe ³⁺)	-1.6276
k_c	Potassium	-40.689
mg2_c	Magnesium	-1.8084
mlthf_c	5,10-Methylenetetrahydrofolate	-0.0489
mn2_c	Manganese	-0.1442
mobd_c	Molybdate	-0.0015
nad_c	Nicotinamide adenine dinucleotide	-0.3916
nadh_c	Nicotinamide adenine dinucleotide - reduced	-0.0099
nadp_c	Nicotinamide adenine dinucleotide phosphate	-0.0245
nh4_c	Ammonium	-2.7125
ni2_c	Nickel	-0.0673
pheme_c	Protoheme C ₃₄ H ₃₀ FeN ₄ O ₄	-0.0489
pydx5p_c	Pyridoxal 5'-phosphate	-0.0489
ribflv_c	Riboflavin C ₁₇ H ₂₀ N ₄ O ₆	-0.0489
sheme_c	Siroheme C ₄₂ H ₃₆ FeN ₄ O ₁₆	-0.0489
so4_c	Sulfate	-0.9041
succoa_c	Succinyl-CoA	-0.0215
thf_c	5,6,7,8-Tetrahydrofolate	-0.0489
thmpp_c	Thiamine diphosphate	-0.0489
udcpdp_c	Undecaprenyl diphosphate	-0.0121
zn2_c	Zinc	-0.0710
2dmmql8_c	2-Demethylmenaquinol 8	-0.0489
5mthf_c	5-Methyltetrahydrofolate	-0.0489
accoa_c	Acetyl-CoA	-0.0611
chor_c	Chorismate	-0.0489
enter_c	Enterochelin	-0.0489
gthrd_c	Reduced glutathione	-0.0489
hemeO_c	Heme O C ₄₉ H ₅₆ FeN ₄ O ₅	-0.0489
lipopb_c	Lipoate (protein bound)	-0.0007
malcoa_c	Malonyl CoA C ₂₄ H ₃₃ N ₇ O ₁₉ P ₃ S	-0.0068

Appendix E

mococdp_c	Molybdopterin cytosine dinucleotide	-0.0015
mocogdp_c	Molybdopterin guanine dinucleotide	-0.0015
mql8_c	Menaquinol 8	-0.0489
nadph_c	Nicotinamide adenine dinucleotide phosphate - reduced	-0.0734
ptrc_c	Putrescine	-7.2902
q8h2_c	Ubiquinol-8	-0.0489
spmd_c	Spermidine	-1.4778

Appendix F: Complete biomass objective functions

The stoichiometric coefficients of all compounds in the complete BOFs are listed in Table F.1.

Table F.1: Stoichiometric coefficients of all compounds in BOFs for all *E. coli* cultures. Stoichiometry is given in mmol/gDWh⁻¹.

Compound ID	<i>E. coli</i> C-lim 0.4	<i>E. coli</i> C-lim 0.2	<i>E. coli</i> C-lim 0.1	<i>E. coli</i> N-lim 0.4	<i>E. coli</i> N-lim 0.2	<i>E. coli</i> N-lim 0.1	<i>E. coli</i> U-lim
Substrates							
h2o_c	-12.72687	-12.76823	-11.14658	-9.22905	-11.31859	-10.48690	-12.72687
asp_L_c	-0.44219	-0.43749	-0.38107	-0.32404	-0.36828	-0.34140	-0.44219
glu_L_c	-0.46573	-0.44602	-0.41929	-0.32268	-0.45150	-0.27514	-0.46573
asn_L_c	-0.33859	-0.33499	-0.29179	-0.24813	-0.28200	-0.26142	-0.33859
his_L_c	-0.10225	-0.10402	-0.09512	-0.07437	-0.09059	-0.11827	-0.10225
ser_L_c	-0.36294	-0.39578	-0.33983	-0.28907	-0.35304	-0.36012	-0.36294
gln_L_c	-0.35911	-0.34391	-0.32330	-0.24881	-0.34814	-0.21215	-0.35911
gly_c	-0.32001	-0.29269	-0.26888	-0.22372	-0.22143	-0.28931	-0.32001
arg_L_c	-0.24003	-0.21953	-0.20168	-0.16780	-0.16609	-0.21699	-0.24003
thr_L_c	-0.39103	-0.39801	-0.34687	-0.28015	-0.33618	-0.36152	-0.39103
ala_L_c	-0.85054	-0.83262	-0.72646	-0.61173	-0.72780	-0.82347	-0.85054
tyr_L_c	-0.16738	-0.19529	-0.14555	-0.13766	-0.15327	-0.07994	-0.16738
met_L_c	-0.07418	-0.07565	-0.04921	-0.04166	-0.04031	-0.04789	-0.07418
val_L_c	-0.45338	-0.44163	-0.39742	-0.31785	-0.38449	-0.18492	-0.45338
phe_L_c	-0.23253	-0.24817	-0.20503	-0.17335	-0.21516	-0.28667	-0.23253
ile_L_c	-0.32547	-0.32702	-0.28508	-0.23097	-0.31494	-0.21485	-0.32547
leu_L_c	-0.57820	-0.56915	-0.51004	-0.42047	-0.48968	-0.48352	-0.57820
lys_L_c	-0.27804	-0.30970	-0.26800	-0.23510	-0.33767	-0.36700	-0.27804
cys_L_c	-0.05757	-0.08372	-0.05456	-0.05039	-0.08002	-0.06891	-0.05757
pro_L_c	-0.27653	-0.28523	-0.24329	-0.20299	-0.25463	-0.23417	-0.27653
trp_L_c	-0.08224	-0.10643	-0.07583	-0.06759	-0.09969	-0.08753	-0.08224
atp_c	-56.40009	-56.15580	-58.66535	-61.53551	-58.34995	-59.64840	-56.40009
ctp_c	-0.16849	-0.19125	-0.20802	-0.19605	-0.16173	-0.30931	-0.16849
gtp_c	-12.87736	-12.87979	-11.30136	-9.36960	-11.40693	-10.74213	-12.87736
utp_c	-0.10499	-0.08629	-0.11241	-0.10299	-0.06964	-0.18107	-0.10499
datp_c	-0.04619	-0.03789	-0.04488	-0.05621	-0.01359	-0.01615	-0.04619
dctp_c	-0.04770	-0.03913	-0.04635	-0.05804	-0.01404	-0.01668	-0.04770
dgtp_c	-0.04759	-0.03904	-0.04625	-0.05791	-0.01400	-0.01664	-0.04759
dttp_c	-0.04613	-0.03785	-0.04483	-0.05614	-0.01357	-0.01613	-0.04613
g1p_c	-0.14040	-0.12562	-0.52361	-1.14219	-0.63325	-0.21008	-0.14040
gal1p_c	-0.03412	-0.02577	-0.02044	-0.02613	-0.01407	-0.03384	-0.03412
glcur_c *	0.00000	0.00000	0.00000	0.00000	-0.04088	0.00000	0.00000
rml1p_c	0.00000	-0.03986	0.00000	0.00000	-0.28625	0.00000	0.00000
gam1p_c	-0.01842	-0.01682	-0.01226	-0.01411	-0.04053	-0.01777	-0.01842
acgam1p_c	-0.02549	-0.02326	-0.01697	-0.01952	-0.05607	-0.02458	-0.02549
hdca_c	-0.04698	-0.08571	-0.07495	-0.07355	-0.07576	-0.09774	-0.04698
hdcea_c	-0.03474	-0.06339	-0.05543	-0.05439	-0.05603	-0.07228	-0.03474
ocdcea_c	-0.02354	-0.04295	-0.03756	-0.03685	-0.03796	-0.04897	-0.02354
glyc3p_c	-0.06472	-0.11809	-0.10327	-0.10133	-0.10438	-0.13466	-0.06472
Substrates implemented form <i>i</i>ML1515 wild-type BOF							
10fthf_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
2fe2s_c	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001
2ohph_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005

Appendix F

4fe4s_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
amet_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
btn_c	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
ca2_c	-0.00109	-0.00109	-0.00109	-0.00109	-0.00109	-0.00109	-0.00109
cl_c	-0.00109	-0.00109	-0.00109	-0.00109	-0.00109	-0.00109	-0.00109
coa_c	-0.00004	-0.00004	-0.00004	-0.00004	-0.00004	-0.00004	-0.00004
cobalt2_c	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001
cu2_c	-0.00015	-0.00015	-0.00015	-0.00015	-0.00015	-0.00015	-0.00015
fad_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
fe2_c	-0.00140	-0.00140	-0.00140	-0.00140	-0.00140	-0.00140	-0.00140
fe3_c	-0.00163	-0.00163	-0.00163	-0.00163	-0.00163	-0.00163	-0.00163
k_c	-0.04069	-0.04069	-0.04069	-0.04069	-0.04069	-0.04069	-0.04069
mg2_c	-0.00181	-0.00181	-0.00181	-0.00181	-0.00181	-0.00181	-0.00181
mlthf_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
mn2_c	-0.00014	-0.00014	-0.00014	-0.00014	-0.00014	-0.00014	-0.00014
mobd_c	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
nad_c	-0.00039	-0.00039	-0.00039	-0.00039	-0.00039	-0.00039	-0.00039
nadh_c	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001
nadp_c	-0.00002	-0.00002	-0.00002	-0.00002	-0.00002	-0.00002	-0.00002
nh4_c	-0.00271	-0.00271	-0.00271	-0.00271	-0.00271	-0.00271	-0.00271
ni2_c	-0.00007	-0.00007	-0.00007	-0.00007	-0.00007	-0.00007	-0.00007
pheme_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
pydx5p_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
ribflv_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
sheme_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
so4_c	-0.00090	-0.00090	-0.00090	-0.00090	-0.00090	-0.00090	-0.00090
succoa_c	-0.00002	-0.00002	-0.00002	-0.00002	-0.00002	-0.00002	-0.00002
thf_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
thmpp_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
udcpdp_c	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001
zn2_c	-0.00007	-0.00007	-0.00007	-0.00007	-0.00007	-0.00007	-0.00007
2dmmql8_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
5mthf_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
accoa_c	-0.00006	-0.00006	-0.00006	-0.00006	-0.00006	-0.00006	-0.00006
chor_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
enter_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
gthrd_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
hemeO_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
lipopb_c	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
malcoa_c	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001	-0.00001
mococdp_c	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
mocogdp_c	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000	-0.00000
mql8_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
nadph_c	-0.00007	-0.00007	-0.00007	-0.00007	-0.00007	-0.00007	-0.00007
ptrc_c	-0.00729	-0.00729	-0.00729	-0.00729	-0.00729	-0.00729	-0.00729
q8h2_c	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005	-0.00005
spmd_c	-0.00148	-0.00148	-0.00148	-0.00148	-0.00148	-0.00148	-0.00148
Products							
glyc_c	0.00138	0.00252	0.00221	0.00216	0.00223	0.00288	0.00138
adp_c	49.80467	49.48154	52.79160	56.63541	52.44258	53.96956	49.80467
amp_c	6.46265	6.56513	5.73159	4.76985	5.81929	5.44986	6.46265
cmp_c	0.05263	0.09602	0.08397	0.08239	0.08487	0.10950	0.05263
gdp_c	12.71478	12.74616	11.12728	9.21011	11.29908	10.46173	12.71478

Appendix F

pi_c	62.31312	62.01844	63.36490	64.66250	62.69012	64.17019	62.31312
ppi_c	7.43751	7.47067	7.12386	6.78889	7.37285	6.80147	7.43751
co2_c	0.04053	0.07396	0.06467	0.06346	0.06537	0.08433	0.04053
h_c	68.61788	68.29550	68.84457	69.18516	68.25479	69.29156	68.61788

* Glucuronate-1-phosphate was added to the BOF as glucuronate because the model did not have a metabolite for glcU1p_c.