Lydia HANGELMANN

Correction, Refinement and Comparison of Genome Scale Metabolic Models of Different *Escherichia coli* Strains

Master Thesis

Institute for Genomics and Bioinformatics

Graz University of Technology

Petersgasse 14, 8010 Graz

Supervisor and Evaluator:

Dipl.-Ing. Dr.techn. Gerhard Thallinger

Graz, March 2014

EIDESSTATTLICHE ERKLÄRUNG

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Graz, am …….…… ……..
(Unterschrift)

STATUTORY DECLARATION

I declare that I have authored this thesis independently, that I have not used other than the declared sources / resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

…… ……..
date (signature)

Abstract

The systems biological view on microorganisms using genome scale metabolic models enables inference on physiological capacities which can be used for biotechnological purposes e.g. to increase the yield of a certain desired product or decrease unwanted byproduct formation. Furthermore metabolic models give hints on essential genes which are useful for metabolic engineering.

In the course of this thesis, genome scale metabolic models of five biotechnologically-relevant *Escherichia coli* strains - MG1655, W3110, HMS174, RV308 and BL21(DE3)- were corrected, refined and extended. In parallel, each of the strains was subjected to batch cultivations, whereas substrate consumption rates and biomass formation were measured. These values in turn were used as input parameters for model simulation. Furthermore, knock out simulations were performed in order to find genes, which are essential to only one of the strains despite high homology of the genome.

Depending on the strain roughly 50 genes were added to the network reconstructions, equaling an extension of the gene lists of 3 %, encoding for around 30 new reactions. Furthermore incorrect predictions regarding gene essentiality was corrected. Using these improved models, a strain-specific essentiality of RV308 for, *folD*, a gene involved in folate metabolism, was found. Batch cultivations on minimal medium supplemented with glucose yielded similar growth rates for strains MG1655, W3110 and BL21(DE3), whereas HMS174 grew roughly 20 % slower and RV308 grew significantly faster (22 %). As substrate uptake rates were the only input parameters or model simulation, simulated growth rates were more than twice as high as the experimental ones.

Future focus should therefore be put on a more extensive analysis of cultivation (by-)products to have additional input parameters at hand. Additionally, strain-specific differences in the biomass composition should be investigated in order to fundamentally improve in silico predictions.

Keywords: Genome scale metabolic models, *Escherichia coli cultivation*, Knock out simulation

Zusammenfassung

Die systembiologische Betrachtung von Mikroorganismen unter der Verwendung von Genom-basierenden metabolischen Modellen ermöglicht Rückschlüsse auf deren metabolische Kapazität. Dieses Wissen kann für biotechnologische Anwendungen herangezogen werden z.B. zur Verbesserung der Produktausbeute, aber auch zur Senkung der Produktion von ungewollten Nebenprodukten. Weiters liefern diese Modelle wichtige Hinweise zur Rolle gewisser Gene für den Organismus, welche in Klonierungsstrategien Eingang finden.

Im Laufe dieser Arbeit wurden die Modelle von fünf biotechnologisch-relevanten *Escherichia coli* Stämmen - MG1655, W3110, HMS174, RV308 und BL21(DE3)- korrigiert und erweitert. Parallel dazu wurde jeder Stamm in Erlenmeyerkolben kultiviert, wobei Wachstumsraten und Substratverbrauch gemessen wurden. Des Weiteren wurden Knock-out Simulationen durchgeführt, um stammesspezifische essentielle Gene zu finden.

In Abhängigkeit des Stammes wurden circa 50 Gene den metabolischen Netzwerken hinzugefügt. Dies entspricht einer Vergrößerung der Genlisten um 3 % beziehungsweise 30 neuen Reaktionen. Außerdem wurden inkorrekten Vorhersagen bezüglich essentieller Gene ausgebessert. Im Zuge von Simulationen der verbesserten Modelle wurde eine stammesspezifische Letalität von RV308 für *folD*, einem Gen, das in den Folatmetabolismus involviert ist, gefunden. *Escherichia coli* Kultivierungen in Minimalmedium mit Glucose lieferten ähnliche Wachstumsraten für MG1655, W3110 und BL21(DE3), wobei HMS174 rund 20 % langsamer und RV308 22 % schneller anwuchsen. Da jedoch einzig Substrataufnahmeraten als Inputparameter zur Modellsimulation verwendet wurden, waren die simulierten Wachstumsraten mehr als doppelt so hoch wie die experimentell bestimmten.

Der Schwerpunkt zukünftiger Arbeiten sollte daher auf die Analyse von Fermentationsbeiprodukten gelegt werden, um zusätzliche Inputparameter parat zu haben. Außerdem sollte die Biomassezusammensetzung der einzelnen Stämme näher untersucht werden, um *in silico* Vorsagen grundlegend zu verbessern.

Stichwörter: Metabolische Modelle, *Escherichia coli* Kulturen, Knock out Simulation

Glossary

3,5 DNS	Dinitrosalicylic acid
ac	Acetate
ACS	American Chemical Society
ATCC	American Type Culture collection
ATP	Adenosine triphosphate
BiGG	Biochemical Genetic and Genomic knowledgebase
BOF	Biomass Objective Function
c	Objective weight vector
C	Substrate concentration
ca2	Calcium
CaCl2	Calcium chloride
cbl1	Cob(I)alamin
cl	Chloride
co2	Carbon dioxide
cobalt2	Cobalt
COBRA	Constraint-based Reconstruction and Analysis
comp	Compartment
CPU	Central processing unit
cu2	Copper
CV	Coefficient of variation
cytd	Cytidine
DNA	Deoxyribonucleic acid
DOS	Disk Operating System
DW	Dry weight
E. coli	*Escherichia coli*
EC	Enzyme Commission
EFM	Elementary Flux Modes
FBA	Flux Balance Analysis
fe2+, fe3+	Iron
FVA	Flux Variability Analysis
GB	Gigabyte
GC	Gas chromatography
glc	Glucose
GOLD	Genomes OnLine Database
gpr	Gene protein relationship

grRules	Gene relationship Rules
h	Proton
h2o	Water
HPLC	High-performance liquid chromatography
indel	insertion-deletion
ISO	International Organization for Standardization
k	Kalium
KEGG	Kyoto Encyclopedia of Genes and Genomes
KO	Knock out
LB	Lysogeny broth
lb	lower bound
M_10fthf[c]	10-Formyltetrahydrofolate
M_2mcacn[c]	cis-2-Methylaconitate
M_2mcit[c]	2-Methylcitrate
M_2obut[c]	2-Oxobutanoate
M_3hcinnm[c]	3-hydroxycinnamic acid
M_3hcinnm[p]	3-hydroxycinnamic acid
M_3hpppn[c]	3-(3-hydroxy-phenyl)propionate
M_3hpppn[p]	3-(3-hydroxy-phenyl)propionate (cytosolic)
M_4h2opntn[c]	4-Hydroxy-2-oxopentanoate (periplasmatic)
M_5mthf[c]	5-Methyltetrahydrofolate
M_acald[c]	Acetaldehyde
M_accoa[c]	Acetyl -Coenzyme A
M_adp[c]	Adenosine diphosphate
M_akg[c]	2-Oxoglutarate
M_amp[c]	Adenosine monophosphate
M_atp[c]	Adenosine triphosphate
M_cechddd[c]	cis-3-(3-carboxyethyl)-3,5-cyclohexadiene-1,2-diol
M_cenchddd[c]	cis-3-(3-carboxyethenyl)-3,5-cyclohexadiene-1,2-diol
M_cinnm[c]	trans-Cinnamate
M_co2[c]	Carbon dioxide
M_coa[c]	Coenzyme A
M_dhap[c]	Dihydroxyacetone phosphate
M_dhcinnm[c]	2,3-dihydroxicinnamic acid
M_dhpppn[c]	3-(2,3-Dihydroxyphenyl)propanoate
M_for[c]	Formate
M_fum[c]	Fumarate

M_g3p[c]	Glyceraldehyde 3-phosphate
M_galt[p]	Galactitol
M_galt1p[c]	Galactitol 1-phosphate
M_gln-L[c]	L-Glutamine
M_gly[c]	Glycine
M_h[c]	Proton
M_h2o[c]	Water
M_hkndd[c]	2-Hydroxy-6-oxonona-2,4-diene-1,9-dioate
M_hkntd[c]	2-Hydroxy-6-ketononatrienedioate
M_methf[c]	5,10-Methenyltetrahydrofolate
M_micit[c]	Methylisocitrate
M_mlthf[c]	5,10-Methylenetetrahydrofolate
M_nad[c]	NAD
M_nadh[c]	NADH
M_nadp[c]	NADP
M_nadph[c]	NADPH
M_nh4[c]	Ammonium
M_oaa[c]	Oxaloacetate
M_op4en[c]	2-Oxopent-4-enoate
M_pep[c]	Phosphoenolpyruvate
M_pi[c]	Phosphate
M_ppcoa[c]	Propanoyl-CoA
M_pppn[c]	Phenylpropanoate
M_pppn[p]	Phenylpropanoate
M_pyr[c]	Pyruvate
M_succ[c]	Succinate
M_succoa[c]	Succinyl-CoA
M_tag6p-D[c]	D-Tagatose 6-phosphate
M_tagdp-D[c]	D-Tagatose
M_thf[c]	5,6,7,8-Tetrahydrofolate
met	Metabolite
mg2	Magnesium
MgSO4	Magnesium sulfate
mn2	Mangan
mobd	Molybdate
MoMA	Minimisation of Metabolic Adjustment
MSDS	Material safety data sheet

NAD	Nicotinamide adenine dinucleotide
NADP	Nicotinamide adenine dinucleotide phosphate
NaOH	Sodium hydroxide
NCBI	National Center for Biotechnology Information
nh4	Ammonium
ni2	Nickel
o2	Oxygen
p.a.	Per analysis
pi	Phosphate
$\mathbf{q_{O2}}$	Specific oxygen uptake rate
$\mathbf{q_s}$	Specific substrate uptake rate
R_3HCINNMH	3-hydroxycinnamate hydroxylase
R_3HPPPNH	3-(3-hydroxy-phenyl)propionate hydroxylase
R_CINNDO	Cinnamate dioxygenase
R_DHCIND	2,3-dihydroxycinnamate dehydrogenase
R_DHCINDO	2,3-dihydroxycinnamate 1,2-dioxygenase
R_DHPPD	2,3-dihydroxyphenylpropionate dehydrogenase
R_FOMETRi	Aminomethyltransferase
R_FORt2pp	formate transport via proton symport (uptake only, periplasm)
R_FORtppi	formate transport via diffusion (cytoplasm to periplasm)
R_FTHFD	formyltetrahydrofolate deformylase
R_FTHFLi	formate-tetrahydrofolate ligase
R_GALTptspp	Galactitol transport via PEP:Pyr PTS (periplasm)
R_GLTPD	Galactitol-1-phosphate dehydrogenase
R_GLYCL	Glycine Cleavage System
R_HCINNMt2rpp	3-Hydroxycinnamic acid transport via proton symport, reversible (periplasm)
R_HKNDDH	2-Hydroxy-6-ketonona-2,4-dienedioic acid hydrolase
R_HKNTDH	2-Hydroxy-6-ketononotrienedioate hydrolase
R_HOPNTAL	4-Hydroxy-2-oxopentanoate aldolase
R_HPPPNDO	2,3-Dihydroxypheylpropionate 1,2-dioxygenase
R_HPPPNt2rpp	3-(3-Hydroxyphenyl)propionate transport via proton symport, reversible (periplasm)
R_MCITD	2-methylcitrate dehydratase
R_MCITL2	Methylisocitrate lyase
R_MCITS	2-Methylcitrate synthase
R_MICITDr	2-Methylisocitrate dehydratase
R_MTHFC	Methenyltetrahydrofolate cyclohydrolase

R_MTHFD	Methylenetetrahydrofolate dehydrogenase (NADP)
R_MTHFR2	5,10-Methylenetetrahydrofolate reductase (NADH)
R_OBTFL	2-Oxobutanoate formate lyase
R_OP4ENH	2-oxopent-4-enoate hydratase
R_PFK2	Phosphofructokinase
R_PFL	Pyruvate formate lyase
R_PPPNDO	Phenylpropanoate Dioxygenase
R_PPPNt2rpp	3-phenylpropionate transport via proton symport, reversible (periplasm)
R_TGBPA	Tagatose-bisphosphate aldolase
R_THFAT	Tetrahydrofolate aminomethyltransferase
RAM	Random-Access Memory
RAVEN	Reconstruction, Analysis, and Visualization of Metabolic Networks
rev	reversible
RNA	Ribonucleic acid
ROOM	Regulatory on/off minimization
rxn	Reaction
S	Stoichiometric matrix
SBML	Systems Biology Markup Language
SBO	Systems Biology Ontology
SD	Standard deviation
so4	Sulfate
succ	Succinate
t	Time
TCA	Tricarboxylic acid cycle
thymd	Thymidine
TU	Technical University
TUX	Total unconstraint flux
ub	Upper bound
URL	Uniform Resource Locator
USB	Universal Serial Bus
USP	United States Pharmacopeia
UV	Ultraviolet
v	Weight vector
WT	Wild type
x	Biomass concentration
Z	Optimization solution (real number)
zn2	Zinc

Table of Contents

Chapter I Introduction

High-throughput technologies paved the way for new emergent fields in life science

Thanks to the improvement of high-throughput technologies, the available knowledge of the physiology of organisms has expanded rapidly. Major tools are “Next Generation Sequencing” methods, which enable the functional and cross-species study of genomes, as well as microarray technology or mass spectrometry. Those achievements finally led to a shift in scope from an isolated view on selected reactions to a systemic sight on biological processes, thus creating the new field of systems biology [1]. A second important stream in life science emerged from the desire for optimization of host strains for biotechnological applications: Metabolic engineering focuses on optimization of organisms by redirecting fluxes in certain pathways through knockouts or heterologous pathways by means of synthetic biology [2]. Systems biology can therefore serve as “scaffold” for metabolic engineering strategies, aiding the discovery of possible pathway targets [3].

The wealth in data and complexity made it furthermore necessary to enrich also the (bio-) informatical side by developing data management tools and powerful algorithms.

The following sections will be shading light on both the mathematical and the biological side of systems biological approaches to metabolic modeling.

Section A: Formal approaches to study biochemical processes

Having evolved from chemical reaction kinetics, mathematical models have a long tradition e.g. in studying enzyme mechanics and kinetics, such as Michaelis- Menten kinetics [4]. There are three major groups of mathematical descriptions of biological processes, which differ in detail and purpose. These are the detailed kinetic models, the graph-based approach and stoichiometric models.

Kinetic models describe dynamic phenomena such as enzyme mechanistic, such as inhibition effects, on a very detailed level, requiring a high level of quantitative information of time dependent processes, which are formalized as ordinary differential equations. Kinetic models are inherently very complex and demanding in terms of molecular biology labor since they need biological information for proper parameter estimation. Therefore they are limited to well-studied pathways, such as glycolysis [5].

The graph-based approach in turn tries to explain dependencies and interconnections within the network to study the robustness of certain pathways and does not require any kinetic information. This approach therefore provides only qualitative information, such as pathway lengths, co-participation of metabolites and co-connectivity of reactions [6].

Stoichiometric network analysis in turn uses reaction stoichiometries of the network and applies certain assumptions/constraints (see below), so that it is possible to infer quantitative information of the system, despite data scarcity [7].

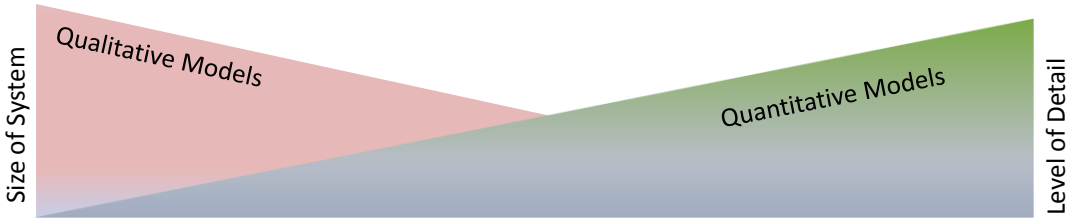
The most popular methods applied to stoichiometric networks are elementary flux mode analysis and flux balance analysis:

An elementary flux mode (EFM) [8] is a non-decomposable steady-state pathway, which represents minimal entities of a network and therefore cannot be further simplified. For analysis, all possible EFM of a system are calculated, which imposes a great demand on computational hardware. Still, all possible solutions are at hand and can be put into consideration for metabolic engineering strategies.

Flux balance analysis, on the other hand [9] organizes metabolites and reactions in form of a matrix of stoichiometric coefficients. The objective is calculated by constraining the possible flux space to a certain maximum or minimum. Since it is the method used in this thesis, it will be discussed in further detail in the following section.

It should further be noted, that there exist also “intermediate” modeling approaches, seeking to bridge stoichiometric and kinetic modeling. Steuer *et al* [10] for example proposed an

approach called structural kinetic modeling. Thereby, a set of local linear models is created at each parameter point in the model space, which can be subjected to further statistical analysis.



Network Analysis	**Stoichiometric Analysis**	**Structural Kinetic Modeling**	**Kinetic Modeling**
Static description		Dynamic description	
No kinetic parameters			Kinetic parameters
Topological properties	Quantitative predictions	Local linear models	Differential equations

Figure 1 Overview of major modeling approaches.
The choice of an appropriate modeling approach is a compromise based on system size and demand on wealth of detail. Since network analysis and stoichiometric analysis do not require kinetic information they are suited for analysis of large scale systems. In contrary kinetic models require measured parameters, but still allow most precise simulations (as extracted from Steuer *et al.* [11]).

This classification of modeling approaches according to their underlying paradigm given above (Figure 1) is not comprehensive, as modeling of metabolism is a vast and fast growing field. A similar and yet more detailed comparison of model methods is given by Tomar and De [12].
In conclusion, it should be mentioned that *Karr et al.* recently published a whole-cell model of *Mycoplasma genitalium* [13], which is a superposition of various modelling approaches. The organism is divided into modules, representing independent models, which are simulated using different modeling strategies (see above). These models are aggregated to a greater whole, since the output values of one model are used as input parameters for another.

Flux Balance Analysis

Probably the most prominent method applied to stoichiometric metabolic models is flux balance analysis (FBA) [14].

The basis of FBA is a matrix composed of stoichiometric coefficients of metabolites participating in reactions. These reactions in the network are bounded by constraints, which are imposed in two ways: as reaction equations, balancing input and output metabolites, and as inequalities of upper and lower boundaries to the system, which reflect enzymatic capacities or thermodynamic constraints etc. However, it is impossible to measure every reaction flux rate, thereby leaving this system underdetermined. Therefore there exists a solution space (Figure 2), having multiple feasible solutions, instead of only one. The flux space is subjected to various optimization procedures such as linear programming or mixed linear integer programming, depending on the context. To this purpose, an objective flux needs to be defined, which is optimized, hence minimized or maximized. Intuitively, this objective of optimization must meet cellular purposes in order to obtain meaningful predictions. Most often, biomass production is chosen as objective for (micro-)organisms, since they strive for optimal growth. However, also other objectives might be chosen. For example for finding optimal sets of reactions for energy efficiency such as minimization of ATP consumption [15].

The stoichiometric matrix **S** has a size of **m** x **r**. Each row (**m**) denotes a metabolite whereas the columns **r** represent the reactions. In case the metabolite is consumed within a reaction, it has a negative coefficient and vice versa. If one metabolite is not used in the reaction, the coefficient is zero. As only few metabolites participate in a given reaction, it is a sparse matrix, consisting mostly of zeros. This matrix is applied for the calculation of fluxes according to Equation 1 that is based on the principle of mass conservation.

$$\frac{\mathrm{dC}}{\mathrm{dt}} = \mathrm{Sv} - \mu\mathrm{C} \tag{1}$$

Where C [mmol/gDW] is the concentration vector of metabolites (m) per gram dry weight of cells (gDW), v (mmol/(gDW*h)) the reaction rate flux vector of n reactions and μ (1/h) is the specific dilution rate which leads to volume change in the system.

As the rate of dilution in the system is insignificant compared to the rate at which metabolites are being consumed or produced, the term μC is negligible.

Equation 1 can be further simplified by assuming, that at steady-state, no metabolites are accumulated, and therefore the metabolite concentration is constant over time (Equation 2).

$$0 = Sv \tag{2}$$

As mentioned previously, many fluxes are inaccessible to biochemical investigation, so that these fluxes are not fixed to a certain flux rate, but are constraint within bounds by different constraints (Table 1).

Table 1 Overview on constraints imposed on the system

Some constraints are dependent on the environmental conditions, such as available substrates, hence representing adjustable constraints, whereas other constraints rely on physical laws (such as stoichiometry) and are nonadjustable. "m" denotes measureable parameters such as fluxes and" k" kinetic constants (adapted from Llaneiras and Picó [7]).

Constraint	Type	Mathematical formulation
Systematic stoichiometry	nonadjustable	$0 = S \times v$
Enzyme/transport capacities	nonadjustable	$v < v_{max}$
(ir)reversibility of reaction fluxes	adjustable	$v > 0$
Measurable fluxes	adjustable	$v = v_m$ or $v_{min} < v < v_{max}$
Regulatory constraints	adjustable	e.g. $v_1 = 0$ if $v_2 = 0$
Kinetic constants	adjustable	$V = k * C_m$

The solution space can be calculated by optimizing – either minimizing or maximizing- the objective.

$$Z = c^T v \tag{3}$$

The scalar Z results from multiplication of the flux vector with the transposed objective weight vector c which ranges from 0 to 1 (by convention), denoting how much a reaction contributes to a certain phenotype. For example, if optimal growth shall be simulated the biomass producing function is weighted as 1.

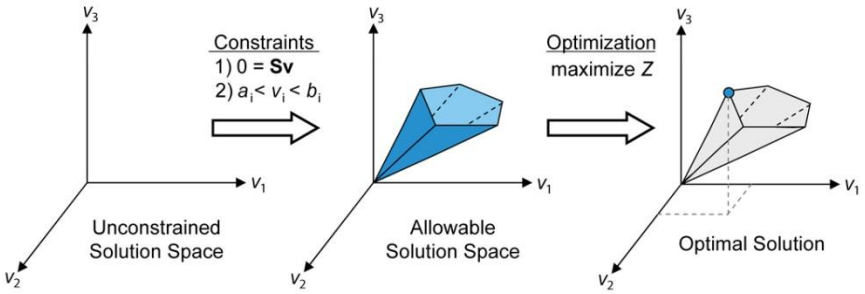


Figure 2 Geometrical representation of constraint-based modeling approach.

Here, a coordinate system is spanned by three flux vectors v1-3. Applying constraints such as lower (a_i) and upper bounds (b_i) and by assuming of steady state, a flux cone is crated representing the allowable solution space. Within this space, the solution (Z) is retrieved by optimizing a given objective [9].

Redundant pathways, leading to the same metabolite, grant stability, allowing an organism to survive through perturbancies, such as gene deletions or other stressful situations. However, it also implies that there might not be a single best solution, as returned by the optimization procedure, but a variety of flux distributions leading to the same optimum. Therefore it is highly misleading to assume that the flux distribution obtained by mathematical optimization of the flux space has any biological significance as long there is no deeper study on flux variability [16]. This issue will be further expanded in the following chapters.

Apparently, the model is prone not to reflect a desired metabolic status of the studied organism once any constraint is set incorrectly or alternative optimal solutions are not considered for their feasibility. However, the applicability of FBA is literally constraint by many assumptions: Firstly, there is the assumption of temporal homogeneity as in Equation 2, which restricts the predictive capacity to continuous cultivations. Furthermore, the biomass function [17], implies homogeneity of biomass production within cells of a population and over time.

Therefore many derivatives of FBA have been developed which seek to improve the predictive power e.g. by adding constraints. Some examples are given in Table 2.

Table 2 A (non-) comprehensive overview of some FBA derivates

Since FBA is a widely applied method, it is steadily improved. This overview lists some arbitrarily chosen derivates, as also discussed in [18].

Method	Description	Reference
flexFBA	A flexible proportion of reactants and product stoichiometry is assumed for the biomass function accounting for varying biomass composition.	[19]
tFBA	Time linked simulation accounts for varying byproduct proportions over time.	[19]
FBrAtio	Flux ratios weight the contribution of convergent pathways to a given product pool.	[20]
MD-FBA	The simulation also accounts for dilution of intermediate metabolite pools.	[21]
coupledFBA	Constraints are applied to link a non-metabolic network to the S matrix, e.g. for linking the transcriptional/translation framework to the metabolic matrix.	[22]
dFBA	Changing flux constraints account for dynamic change in nutrient concentration during short time intervals.	[23]
rFBA	Transcriptional regulation via a boolean vector is imposed on the reaction network (reactions are either on or off).	[24]

Genome Scale Metabolic Models

Network reconstruction process and examples

The reconstruction process of (genome-based) metabolic networks should be taken with great care of as the quality of reconstruction determines the outcome, hence a predictive, reliable simulation. Among published procedures [25, 26], the standard procedure for model reconstruction proposed by Thiele and Palsson in 2010 [27] is the richest in detail: This 96 step protocol designed as standard operation procedure explains all necessary stages starting with the assembly of a draft reconstruction to the dissemination of the final model.

Overall, the process can be roughly divided into four stages:

Firstly, a *draft reconstruction* based on the sequenced genome is set up. Nowadays, the genome is generated at high speed and low cost using "Next Generation Sequencing" methods. Still it needs to be well annotated and lined with biochemical information. Although it seems tedious to organize several instances, such as reactions, metabolites and genes, the draft reconstruction process can be automated to a larger part and needed information can be extracted from various databases, such as MetaCyc[1], NCBI[2] and KEGG[3] [28]. Even through this appears to be comfortable, one should always keep in mind, that it is difficult to maintain the quality within vast data sets and comparability among the databases [29]. Liberal and Pinney [30] for instance, tackled missannotations in genome sequences by developing a method, which is able to predict annotation accuracy. Interestingly, they observed a clear negative correlation between annotation accuracy in prokaryotic network reconstructions and phylogenetic distance to *Escherichia coli,* a popular model organism. This emphasizes the importance of biochemical background information on an organism for the quality of a model.

Once a draft genome network representation is available, the reconstruction is manually refined and reconciled. Since not all genes found in databases suit the scope of the model e.g. proteins without any metabolic function, the network reconstruction needs to be revised. This refining process should not be underestimated as the results from the first step should be reconsidered. Further literature research is indispensable, as databases are not complete. After the reconstruction has been reviewed, it can be converted into a mathematical model which can be simulated using various toolboxes. For example, the RAVEN toolbox, is a novel toolbox, which largely automates the annotation of reconstructions whereas the COBRA toolbox [31] has been established as "golden standard" in many laboratories worldwide, as it contains a vast collection of functions for model analysis network reconstruction refinement. To date, there are software solutions available based on different environments, e.g. in R (Sybil package), the DOS

[1] http://www.metacyc.org/
[2] http://www.ncbi.nlm.nih.gov/
[3] http://www.genome.jp/kegg/

command line (FASIMU [32]), MATLAB (RAVEN Toolbox, COBRA Toolbox) or Java (Optflux [33]). Some of these tools will be further discussed in Chapter 2.

Furthermore models need to be evaluated experimentally to check its predictions. This can be done with regard to growth on various substrates or by creation of mutants. This data in turn can be used for iterative model correction.

The final models can be stored and compared in repositories such as BiGG[4] or MetaNetX[5] [34].

Genome Scale Metabolic Models of *Escherichia coli*

Metabolic network reconstructions of *Escherichia coli* actually date back to the pre-genome area. The first model was built by Varma and coworkers, published in 1993. It was based purely on biochemical reactions, not on genetic information. Nevertheless, it was possible to compute growth rate, amino acid production as well as the impact of acetate and other by-products [35, 36]. This model has been refined by Pramanik and Keasling in 1998 [37, 38] by updating the biomass objective function, and thorough analysis of biomass composition. By 1997, the first genome sequence of *Escherichia coli* was published [39], estimating 4288 protein coding genes. The first genome based reconstruction for *Escherichia coli* links the genetic information with the biochemical reaction network, which allowed gene deletion studies [40]. Within the next years, further refined reconstructions have been published which did not only contained more genes, but is also richer in scope and model aspects [43]. In 2007 an extended network reconstruction, iAF1260, was published, which has not only been updated for further components (reactions, metabolites and genes), but also for thermodynamic consistency of the reactions [42]. Thus, this reconstruction accounts for 48 % of the characterized ORFs [43]. As noticed with these models, they often failed to correctly predict certain growth phenotypes due to lacking enzyme information. Therefore Orth *et al.* [44, 45] published a new model, iJO1366, in which the biomass objective function (BOF) was updated and some network gaps were closed. The BOF is an artificial reaction, which calculates the growth from substrates based on elementary composition [17]. Further remaining “orphan enzymes”, meaning metabolic enzymes lacking gene sequence link, where assigned from sequence inference with metagenomic sequence information [46].

[4] http://bigg.ucsd.edu/
[5] http://metanetx.org/

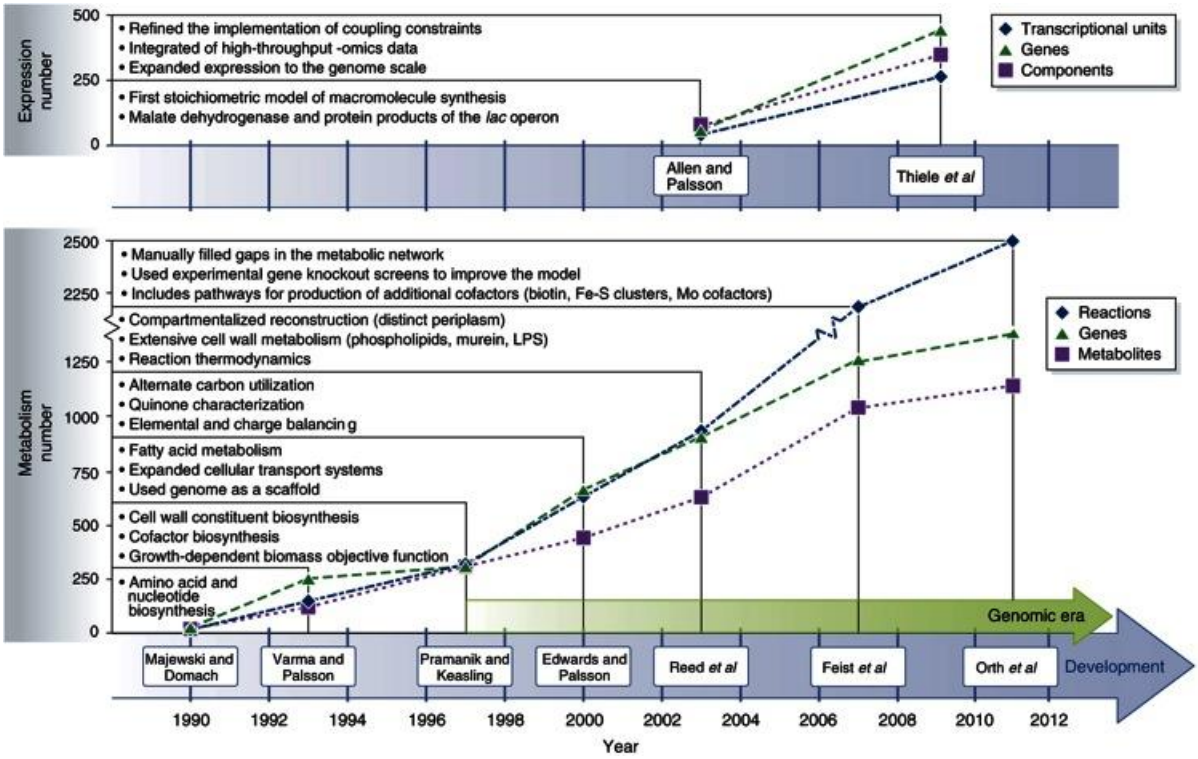


Figure 3 Timeline of model network reconstruction of *Escherichia coli.*
Starting over twenty years ago, network reconstructions of *Escherichia coli* have been steadily improved through addition of pathways and constraints through experimental evaluation (reprinted from [47]).

Furthermore, Baumler *et al.* [48] studied the evolution of *Escherichia coli* phylogeny using genome scale network reconstructions. In this study, iAF1260 was improved by adding new gene-protein-relations (gpr) through sequence homology and several models of related pathogenic and apathogenic strains were built by comparing genome annotations.

So far, metabolic models have been applied in metabolic engineering, such as optimization production of lactic acid or amino acids [49] as well as for basic research on finding new functional enzyme-gene associations from homology [50]. In 2013, Monk *et al.* [51] published a cross-strain comparison of 55 different pathogenic and apathogenic *Escherichia coli* strains, investigating the functional differences concerning auxotrophies and growth behaviour: A discrimination between the strains based on growth behaviour was predicted correctly in 80 % of the cases.

In parallel, a “second” generation of metabolic networks has been developed: in 2012 Thiele *et al.* [52] published another network reconstruction that was more than ten times larger with respect to reaction entries than all the others before. The so-called “ME”-matrix couples the transcriptional and translational machinery, hence “E”-matrix (“expression”), with the metabolic

“M”- matrix. This made it possible to predict the evolutionary evolved codon usage. However this wealth in detail comes with costs: First of all it is time-intensive to simulate and secondly since it has not been built for classical genotype-phenotype predictions it does not contain any gene-reaction association and can therefore not be subjected to gene essentiality simulations.

Another reconstruction of this kind has been published in October 2013 by O’Brien *et al*. [53] which is even refined for the M-matrix as well as richer in scope and detail. Similarly, this model is hard to compute and needs to be accessed via COBRApy (a “sister toolbox” to the MATLAB-based COBRA Toolbox, programmed in Phyton. This model accounts for roughly 80 % of the growth-associated proteome, which enables refined growth predictions through setting of proteomic constraints.

Section B: *Escherichia coli* host strains

As outlined in the previous section, it is of great importance to have a sound knowledge of the organism(s) of which the genome scale metabolic model should be built. The following paragraphs will therefore give background information on *Escherichia coli* and especially will portray the strains in use.

Escherichia coli was first brought to public attention by Theodor Escherich in the late 19th century as *Bacterium coli commune,* which he isolated during microbiological examinations of infant feces [54]. Since then, this remarkable Gram negative, rod-shaped, facultative anaerobic bacterium has advanced to one of the most popular study objects for both fundamental and applied research, such as DNA recombination, pathogenesis mechanisms and genetic engineering; detailed examples will be given below. This extensive use, especially in the "early days" of fundamental research, lead to a rapid diversification of the original wild type, due to naive handling or intentional by UV treatments or phage infections. In 1968, Bachmann and coworkers [55] rigorously classified strains in use at that time which were descendants of *E. coli* K-12 and thus allowed back tracking of different strains. Nowadays, there are five *E. coli* strains amongst all the isolates which have gained popularity in laboratories, mostly because they are generally regarded as safe to work with. These are *Escherichia coli* B, C, Crooks, W and K-12 [56].

Attempts to cover the genetic diversity of *Escherichia coli* lead to a classification of six phylogenetic groups A,B1,B2,D,C,E based on multilocus-enzyme-electrophoresis [57], which clusters the K-12, B, C in group A and *E. coli* W in B1 [54]. Although there is an undeniable genetic diversity, it could be shown, that the Escherichia "core" genome can be reduced to roughly 2000 genes, which is less than half of the postulated number of genes that K-12 MG1655 possesses [59].

To date[6], there are 2127 *Escherichia coli* genomes registered in the Genomes Online database[7]. For this work five frequently biotechnologically applied *Escherichia coli* strains, the K-12 strains MG1655, W3110, RV308 and HMS174 as well as the *E coli* B strain BL21(DE3) have been used. The genomes of MG1655, W3110 and BL21(DE3) have already been sequenced and compared extensively [60, 61]. In contrary, RV308 and HMS174 have been sequenced, but not yet published.

Below, strains in use are portrayed for their background more extensively.

[6] As accessed on 09/12/2013
[7] http://genomesonline.org/cgi-bin/GOLD/index.cgi

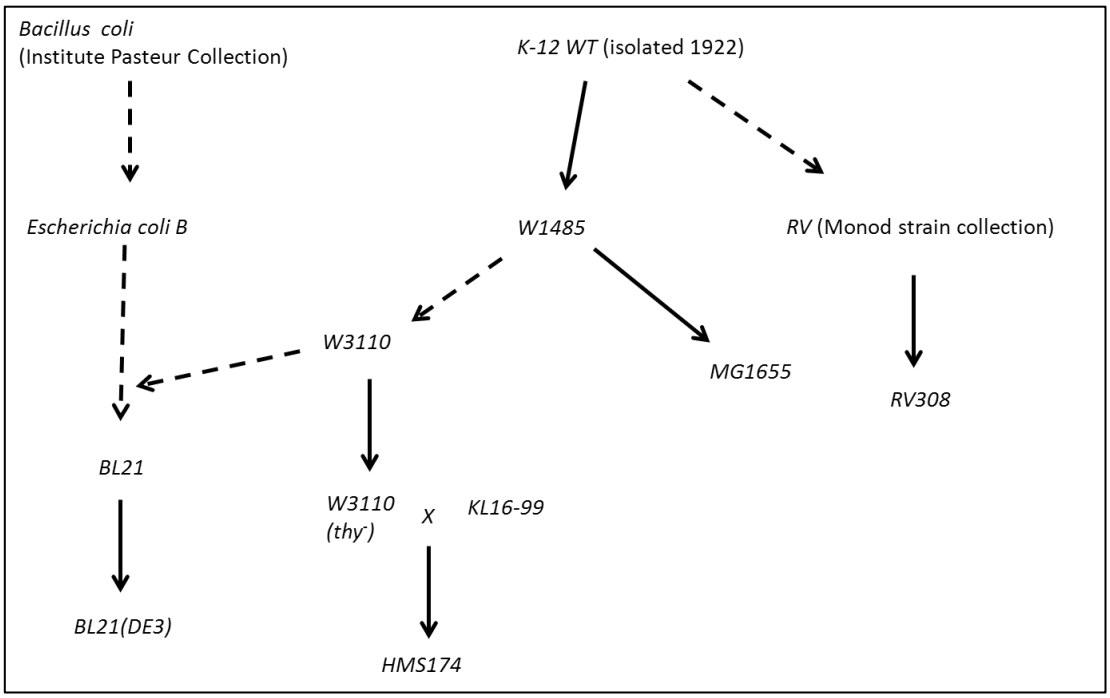


Figure 4 Ancestral relationships between the strains used for this work.
Relationships between the strains have been reconstructed after Bachmann *et al.*[55] and Daegelen *et al.* [62]. The solid lines indicate a direct ancestry/offspring relationship, whereas dashed lines indicate that one or more "intermediate" strains where omitted in the figure for arrangement reasons. The length of pointers is chosen arbitrarily and does not reflect any phylogenetic distance. All strains used in this work are related to a greater or lesser extent, the (reported) mutagenic agents exerted can be retrieved from literature cited in the main text.

MG1655

MG1655 has gained reputation as "wild type" strain in use because of very little strain alterations from the archetype. In an early application, this strain was used for genetic mapping which was performed by transposition, phage infection and combination of antibiotic resistances [63]. Soupene and coworkers elucidated physiological characteristics of MG1655 by rigorously investigating growth defects and cross regulations [64]. In a more recent study; MG1655 was genetically engineered for selective geraniol production by systematic knock outs of target enzymes, which lead to a twofold increase of geraniol production [65].

W3110

As can be seen in Figure 4, W3110 is a K-12 derivate created through relatively little interventions. The first record in Pubmed dates back to 1966 when Paigen [66] published his work on the role of galactose in the galactose pathway of different wild type and mutant *Escherichia coli* strains.

Akinterinwa and Cirino [67] demonstrated, that the replacement of xylose kinase xylB of E coli (producing toxic products) and by heterologous expression of xylokinase *P. stipites* leads to production of xylitol without impairing toxic byproducts.

HMS174

The HMS174 strain was originally created by Campbell and coworkers for physical mapping of bacteriophage T7 DNA. The HMS174 was a product of recombination between W3110(*thy*$^-$) and KL-16-99 [68]. Apart from initial studies on phage and DNA recombination analysis the scope of study has shifted to heterologous protein production. For example HMS174 derivate HMS174(DE3) pLys was applied for the expression and characterization of an immunogenic lipoprotein derived from *Fracisella Ipp3*. As could be shown, the heterologous expressed protein was correctly palmyotylated which resembled the native protein [69]. Most recently a knockout strain HMS174(DE3) was used to design a new addition system that ensures plasmid stability without the need of antibiotic supply [70].

RV308

The *Escherichia coli* K-12 strain RV308 was published by the group of Maurer to identify the role of the right operator region for the gene regulation of phage lambda. RV308 was derived from RV through P1 transduction. RV308 was used in β-galactosidase assays to report for promotor activity in the promotor region of bacteriophage lambda [71]. The parent strain RV was derived from the J. Monod strain Collection. More recently, RV308 was used as expression host for anti-HER2, since it is capable of growing to high cell densities [72].

BL21(DE3)

Different to the other strains mentioned above, this strain is not derived from the K-12 line but originates from a *Bacillus coli* recorded in the Pasteur collection 1918 which was used for phage investigations. Studier et al reconstructed the history of E coli B similar to Bachmann's effort for K-12 strains [62]. BL21, the direct progenitor of BL21(DE3), was generated by UV and P1 transduction from both B strain as well as from K-12 W3110 along its ancestors. BL21(DE3) was constructed by Studier by superinfection of phage DE3, thereby creating the lysogen BL21(DE3) that carries the inducible T7 RNA polymerase [62]. This was the starting point of a unique success story of BL21(DE3) – and mutant variations - in biotechnology. This strain has become one of the most popular protein expression hosts, especially because of the ability to grow to high cell densities , absence of proteases as well as the lack of genes for flagella biosynthesis [73].

Objective

Genome-scale metabolic models are providing a systemic view on networks, thus helping to understand the physiology of an organism.
Since this is a “primer” project, the objective of this thesis is to refine existing genome scale network reconstructions of five biotechnologically-relevant *Escherichia coli* strains: MG1655, W3110, HMS174, RV308 as well as BL21(DE3). In parallel, different available toolboxes shall be tested.
A further aim is to find strain-specific pathway differences from knock out simulations which may allow conclusions on specific metabolic capacities.
In parallel, batch cultivations on minimal medium of these strains are envisaged, whereas growth rates and substrate uptake rates shall be determined. These values shall subsequently serve as input parameters for model simulation.

Chapter II Materials and Methods

Section A: Computational Environment

This section contains all the necessary installation information needed. Furthermore used toolboxes are reviewed and compared, whereas the focus is put on RAVEN toolbox and COBRA toolbox respectively.

Table 3 Hardware

Computer	Intel® Core™2Duo CPU E6750 @ 2.66GHz 2,00 GB of RAM
Operating System	Microsoft Windows XP Professional Version 2002 Service Pack 3

Since this thesis is a "primer" project, a selection of programs and toolboxes for genome scale metabolic modeling have been installed and tested. The software has been installed according to the installation information provided on web sites and manuals, respectively.

Table 4 Software

Software	Version	Comment	Source/ URL
MATLAB	R2012b (network license)	Simulation environment	MathWorks http://www.mathworks.de/products/matlab/
COBRA Toolbox	Version 2.0.5	MATLAB based	http://sourceforge.net/projects/opencobra/files/cobra/ [31]
RAVEN Toolbox	Version 1.07	MATLAB based	http://129.16.106.142/tools.php?c=raven [74]
SBMLToobox	Version 4.1.0	Dependency for COBRA and RAVEN toolbox	http://sbml.org/Software/SBMLToolbox [75]
libSBML	Version 5.6.0_libxml2-x86	Library for SBML Toolbox	http://sbml.org/Software/libSBML [76]
Paint4Net	Version 1.3	extension for COBRA toolbox, used for network visualization	http://biosystems.lv/index.php/software/paint4net [77]

Table 4 Software (continued)

Optflux	Version 3.0.3 ("Optflux3")	Java	http://www.optflux.org/ [33]
Cytoscape	Version 2.8.3	Java	http://www.cytoscape.org/
CyFluxViz Toolbox	Version .093	extension for Cytoscape, used for network visualization	http://www.charite.de/sysbio/people/koenig/software/cyfluxviz3/ [78]
Gurobi	Version 5 (free academic license)	Numerical solver	http://www.gurobi.com/
Mosek	Version 6.0.0.148 (free academic personal license)		http://www.mosek.com/
Tomlab (cplex)	Version 7.9 (R7.9.0) 2-Jul-2012 (trial licence)		http://tomopt.com/tomlab/products/
R	R i386 3.0.1	Simulation environment	http://cran.r-project.org/
Sybil	Version 1.2.5	Systems biology toolbox, R based	http://cran.r-project.org/web/packages/sybil/index.html [79]

General elements of a genome scale metabolic network reconstruction

Summarizing briefly, network reconstructions used herein contain information on reactions, including reaction equations, allowed bounds and involved enzymes (genes). Similarly, the list of metabolites gives information on elementary composition and localization. Information on annotated genes is also provided as enzyme names and identifies ("b-numbers"). Moreover, cellular compartments are listed for localization of reactions. Lastly, the reconstruction also comprises information on the organism to be modeled as well as on the author of the reconstruction.
This listing is not exhaustive, further information on entries is given at a later point in this chapter. Network reconstructions used in this thesis can be found in the digital appendix.

A common model representation: the systems biology markup language

As mentioned already in Chapter 1, there are many different software packages available for flux balance analysis, which address different problems in computational systems biology. To allow interoperability, a standardized data interchange format for biological models has been developed, called systems biology markup language (SBML) [80] Typically, the SBML file, based on the extensible markup language (XML) format, contains information on the SBML version and general information about the network reconstruction such as species and author information. Furthermore, flux units are specified, which are typically mmol/(gDW*h).
Below, examples are given for compartment, metabolite, and reaction representation. A more detailed discussion on the SBML format is provided by David Ruckerbauer [81].

For *Escherichia coli* three physiological compartments are defined, namely the cytosol (c), the periplasma (p) and an extracellular space (e). Below, the notation for the periplasmatic compartment is given. As can be seen, also the relationship between the different compartments is supplied, as "outside of the cytoplasma".

```
<listOfCompartments>
<compartment id="C_p" name="periplasm" outside="C_c" spatialDimensions="3">
</compartment>
.
.
.
</listOfCompartments>
```

Metabolites are defined in a similar way. For example glucose-6-phosphate can occur in all three cellular compartments (cytosol, periplasm and extracellular). An additional compartment, boundary (b), is specified which indicates the borders of the system to be modeled.
Therefore, for such metabolite entries, also called exchange metabolites, the "boundaryCondition" (see below) is set to "true".

```
<listOfSpecies>
.
.
.
<species id="M_M_g6p_c" name="D-Glucose 6-phosphate_C6H11O9P"
compartment="C_c" boundaryCondition="false">
</species>
<species id="M_M_g6p_e" name="D-Glucose 6-phosphate_C6H11O9P"
compartment="C_e" boundaryCondition="false">
</species>
<species id="M_M_g6p_b" name="D-Glucose 6-phosphate_C6H11O9P"
compartment="C_b" boundaryCondition="true">
</species>
<species id="M_M_g6p_p" name="D-Glucose 6-phosphate_C6H11O9P"
compartment="C_p" boundaryCondition="false">
</species>
.
.
.
<\listOfSpecies>
```

Similarly, an example reaction, hexokinase, is provided below. The involved metabolites are listed separately depending on their role in the reaction, either as substrate or product. Additional information on stoichiometry, bounds and reversibility is also comprised in the entry.

```
<listOfReactions>
.
.
.
<reaction id="R_R_HEX1" name="hexokinase (D-glucose:ATP)" reversible="false">
  <notes>
    <html:p>GENE_ASSOCIATION: b2388</html:p>
    <html:p>SUBSYSTEM:Glycolysis/Gluconeogenesis</html:p>
    <html:p>PROTEIN_CLASS: 2.7.1.1</html:p>
  </notes>
  <listOfReactants>
    <speciesReference species="M_M_atp_c" stoichiometry="1"/>
    <speciesReference species="M_M_glc_DASH_D_c" stoichiometry="1"/>
  </listOfReactants>
  <listOfProducts>
    <speciesReference species="M_M_adp_c" stoichiometry="1">
    </speciesReference>
    <speciesReference species="M_M_g6p_c" stoichiometry="1">
    </speciesReference>
    <speciesReference species="M_M_h_c"stoichiometry="1">
    </speciesReference>
  </listOfProducts>
  <kineticLaw>
  <math xmlns="http://www.w3.org/1998/Math/MathML">
    <apply>
    <ci> LOWER_BOUND </ci>
    <ci> UPPER_BOUND </ci>
    <ci> OBJECTIVE_COEFFICIENT </ci>
    </apply>
  </math>
  <listOfParameters>
    <parameter id="LOWER_BOUND" value="     0.00000000"/>
    <parameter id="UPPER_BOUND" value="  1000.00000000"/>
    <parameter id="OBJECTIVE_COEFFICIENT" value="     0.00000000"/>
  </listOfParameters>
  </kineticLaw>
</reaction>
.
.
.
<listOfReactions>
```

Simulation tools

The "basis" toolbox used in this thesis is the RAVEN toolbox -**R**econstruction **A**nalysis and **V**isualization of M**e**tabolic **N**etworks- developed by Rasmus Agren *et al.* [74] with the main purpose to help the semi-automatic network reconstruction of genome scale metabolic models based on protein homology, but also network analysis and visualization. The RAVEN toolbox provides two approaches for creating a network reconstruction. In case similar networks are available, these can be used as template for the new model, whereas the reconstruction is created from protein ontology. The second approach does not depend on existing models, but makes extensive use of information gathered from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. Metabolites and reactions are extracted from KEGG and coupled to genes based on protein homology. Another backbone of the software is the analysis of genome scale metabolic models, which contains major basic FBA analysis tools such as optimization of objectives, gene essentiality simulation and gap identification. Another interesting feature of this toolbox is its visualization capacity. Opposite to COBRA toolbox, which uses rendered pathway maps using Java and Perl, the RAVEN toolbox allows the incorporation of maps drawn using CellDesigner [82], a popular tool in systems biology. The calculated fluxes can be plotted onto such maps.

The **Co**nstraint **B**ased **R**econstruction **A**nalysis toolbox (COBRA) has been created by the Palsson laboratory [31]. Meanwhile, it has established itself as golden standard for flux balance analysis and contains a wide range of functions for Flux balance analysis such as optimization problems, knockout studies, ^{13}C analysis, as well as many add-ons such as Paint4Net [77]. Recently, the COBRA toolbox has also been implemented in Phyton [83], which was necessary due to the emerging model size and complexity.

Opposite to the MATLAB toolboxes mentioned above, Optflux is an open source software package implemented in Java and therefore provides a user friendly graphical user interface, which provides laboratory scientists easy model handling. Like RAVEN and COBRA toolboxes, it spans basic flux balance analysis, but also provides strain optimization functions like Optknock [84]. The downside is the speed, as outlined below.

Furthermore, SBML -models can be analyzed in R [79]. This Sybil package, developed in 2013 by Dietrich *et al.* [79] is far more time-efficient than similar calculations performed in MATLAB due to algorithm architecture and implementation. According to the authors, flux variability analysis (FVA) is 53 times faster with Sybil than with COBRA and over hundredfold faster than Optflux. There are a myriad of available software tools, a good overview is given by Tomar *et al.* [12].

RAVEN versus COBRA Toolbox

The format of choice for network representation is Excel since it is comparably easy to handle.

In analogy to the SBML file, the Excel workbook gives information on the network reconstruction and contains reaction, metabolite and gene lists as well as information on the compartments included in the reconstruction. Excel files of all network reconstructions can be found in the appendix or supplied along with published network reconstructions, such as iAF1260 [42].

For simulation, the “raw” network reconstructions, “stored” in Excel files in turn can be converted into SBML file. Although SBML enables the free exchange of XML files between different software packages, this freedom of exchange is restricted: Although both COBRA and RAVEN toolbox support an Excel format for “network storage”, which can be imported into as model into MATLAB workspace, these models have slight differences in there informational content as illustrated in Table 5.

As the network reconstructions in Excel have a slightly different informational content, so do the subsequent SBML files, which are imported into MATLAB workspace. Since COBRA is taken as “golden standard”, other toolboxes expect the SBML files to contain species information according to COBRA toolbox conventions. Once this “standard” is disregarded by lacking of distinct compulsory species, successful interchange to other toolboxes is impeded (Chapter 3, function xls2Rmodel.m).

Table 5 Comparison of features from a SBML file which can be accessed within a MATLAB workspace. Here, the reconstruction (here iJO1366_raven.V1.xml) is loaded into MATLAB workspace by either RAVEN or COBRA toolbox as a structure. Although the structure contains common core variables, regardless of the toolbox, both toolboxes may contain additional (optional) variables. A summary of structure variables and their classes are given below, albeit a detailed description of each variable can be found in the appendix, as well as the reconstructions themselves. Bold variables are obligatory input variables for simple FBA analysis.

COBRA Toolbox	Common core features	RAVEN Toolbox
metCharge: [1843x1 int32]	**rxns: {2628x1 cell}**	comps: {4x1 cell}
confidenceScores: {2628x1 cell}	**mets: {1843x1 cell}**	compNames: {4x1 cell}
rxnReferences: {2628x1 cell}	**S: [1843x2628 double]**	compOutside: {4x1 cell}
rxnNotes: {2628x1 cell}	**rev: [2628x1 double]**	metComps: {1843x1 cell}
metChEBIID: {1843x1 cell}	**lb: [2628x1 double]**	compMiriams: {4x1 cell}
metKEGGID: {1843x1 cell}	**ub: [2628x1 double]**	metMiriams: {1843x1 cell}
metPubChemID: {1843x1 cell}	**c: [2628x1 double]**	geneMiriams: {1416x1 cell}
metInChIString: {1843x1 cell}	**description**	
	genes: {1416x1 cell}	
	rxnGeneMat: [2628x1416 double]	
	grRules: {2628x1 cell}	
	rxnGeneMat: [2628x1416 double]	
	grRules: {2628x1 cell}	
	subSystems: {2628x1 cell}	
	rxnECNumbers: {2628x1 cell}	
	rxnNames: {2628x1 cell}	
	metFormulas: {1843x1 cell}	
	b: [1843x1 double]	

As can be seen, the core structure does not require genetic information for basic flux analysis, though it is essential for certain simulations. One major difference between the toolboxes is information on metabolite charge. Most noticeable, the COBRA toolbox contains metabolite formulas and charges, as can be accessed in PubChem or ChEBI databases, whilst the RAVEN toolbox does not. It should be therefore noted, that models, as they are used for this thesis cannot be subjected to energy balancing. This issue will be further discussed in Chapter 4.

Network reconstructions

As pointed out in the introduction, the network reconstruction process is a laborious task. However, many network reconstructions versions of *Escherichia coli* K-12 MG1655 exist already. Prior to this thesis, the latest reconstruction of MG1655 as prepared by Yamada *et al.* [46] has therefore been used as template to adapt all other strains by modifying gene lists and corresponding reactions based on a genome differences. Unlike other published reconstructions, which have been converted into models and simulated using the COBRA Toolbox, the network reconstructions used herein are formatted to suit RAVEN toolbox "convention".
The initial network reconstructions are listed in Table 6. A detailed description of the revision process is provided in Chapter 3.

Table 6 initial network reconstructions.
Initial network reconstructions were created prior to this thesis. They have different metabolite lists and reaction-to-gene associations. This will further explained in Chapter 3. As for the nomenclature: according to previous reconstructions: „i" stands for *in silico,* „W" is the initial of the strain and 1407 is the total number of genes in the network reconstruction.

Strain	Network reconstruction	Version	Metabolites	Reaction-to-Gene Association
MG1655	iJO1366-novel-air-GLC_raven.xls	6/11/2012 19:36	charged	correct
W3110	iW1407_raven.xls	31/10/2012 21:55	uncharged	partially incorrect
HMS174	iHMS1391_raven.xls	6/11/2012 22:05	charged	correct
RV308	iRV1394_raven.xls	31/10/2012 21:59	uncharged	partially incorrect

Section B Microbial Experiments

Materials

Table 7 Escherichia coli host strains

Strain	Source	Genotype
MG1655	TU Graz strain collection	*F^- lambda$^-$ ilvG$^-$ rf^{b-}50 rph$^-$1*
W3110	TU Graz strain collection	*F^- lambda$^-$ IN(rrnD-rrnE)1 rph$^-$1*
HMS174[8]	ATCC – 47011	*F^- recA1 hsdR(rK12$^-$ mK12$^+$) (Rif^R)*
RV308[8]	ATCC – 31608	*lacIq$^-$ su$^-$ ΔlacX74 gal IS II::OP308, strA*
BL21(DE3)	TU Graz strain collection	*F^- ompT gal dcm lon hsdS$_B$(r_B^- m_B^-) λ(DE3 [lacI lacUV5-T7 gene 1 ind1 sam7 nin5])*

Table 8 Instruments

Instrument	Model/Supplier
Benchtop centrifuge	Biofuge „pico“, Hereaus
Centrifuge	Eppendorf Centrifuge 5810 R
Laboratory scale	Kern EW (0.5 – 1500 g)
Analytical scale	Sartorius, BL1205 (max 120 g, d = 0,1 mg)
Orbital sShaker	Infors HT Orbitron
Spectrophotometer I	Eppendorf BioPhotometer plus
Spectrophotometer II (Softmax Pro 4.3 LS)	Spectramax Plus 384, Molecular Devices
Firesting[9]	PyroScience
Needle -type microsensor (Orange Fiber)[9]	PyroScience
pH-meter	WTW Series inolab pH720
Microtiter plates	Greiner
Deepwell plates	Greiner
Vortex	Ika Vortex Genius 3
Syringe (20 mL)	Braun, Inject(R)
Cannula (G 20 x 1 1/2" / ø 0,90 x 40 mm, yellow	Braun, Sterican(R)

[8] Strains retrieved from Jürgen Mairhofer, Institute for Applied Microbiology, University of Natural Resources and Life Sciences, Vienna

[9] Kindly provided by the group of Georg Gübitz (IFA Tulln/TU Graz)

Table 8 Instruments (continued)

Magnetic plate	Variomag MonoKomet
Cuvettes (SemiMicro)	Greiner, Bioone
Filter (diameter 0.2 µm)	Sterileo exp.
100 mL Erlenmeyer shaking flasks	Schott Duran
300 mL Erlenmeyer shaking flasks	Schott Duran
1 L Erlenmeyer shaking flasks	SigmaX
1 L Erlenmeyer shaking flasks(baffled)	SigmaX
2 L Erlenmeyer shaking flaskss	SigmaX
2 L Erlenmeyer shaking flasks (baffled)	SigmaX
Water bath	GFL
Bunsen burner	Campingaz

Table 9 Chemicals

Chemical	**Specifications**	**Supplier**	**Cat #**
Glycerol	≥98 %, Ph.Eur., anhydrous	Roth	7530.4
Sodium hydroxide	≥99 %, p.a., ISO, NaOH	Roth	6771.1
Disodium hydrogen phosphate	≥98 %, Ph.Eur., USP, Na_2HPO_4 anhydrous	Roth	T876.2
Monopotassium phosphate	≥99 %, p.a., ACS, KH_2PO_4	Roth	3904.1
Sodium chloride	≥99,5 %, p.a., ACS, ISO NaCl	Roth	3957.1
Magnesium Sulfate	≥99 %, p.a., ACS Heptahydrat	Roth	P027.1
Calcium chloride	≥99 %, Ph.Eur., USP $CaCl_2$Dihydrat	Roth	T885.1
Glucose	α-D(+)-Glucose Monohydrat	Roth	6887.1
Sodium sulfite	p.a., ACS reagent, , ≥98.0% anhydrous	Fluka (Sigma-Aldrich)	71989

Table 9 Chemicals (continued)

Chemical	Specifications	Supplier	Cat #
Agar	General purpose, bacteriological grade	Lab M Limited	MC006 (CAS#: 9002-18-0)
LB – Medium	Luria-Miller	Roth	X968.3
LB - Agar	Lennox	Roth	X965.3
3,5-Dinitrosalicylic acid	≥98%	Sigma-Aldrich	D 0550-10G
Potassium sodium tartrate tetrahydrate	$C_4H_4KNaO_6 \cdot 4\ H_2O$ ≥99 %, p.a., ACS, ISO	Roth	8087.2

Methods

Preparation of glycerol stocks

20 mL LB medium in a 100 mL Erlenmeyer flask are inoculated with a single colony, and incubated overnight at 37 °C at 110 rpm orbital motion. 400 µl of this culture are mixed thoroughly with 1 mL of 50 % (w/w) glycerol in cryovials and stored at – 80 °C until use.

Escherichia coli cultivation

20 µl of *Escherichia coli* glycerol stock are streaked out on M9 agar plates supplemented with 2 g/L glucose and incubated for two to three days at 37 °C or at room temperature over the weekend. If not used, the plate is sealed with parafilm and stored at 4°C for as long as one month maximally.
50 mL M9 medium in 300 mL Erlenmeyer flasks are inoculated with a *Escherichia coli* single colony. This preculture is incubated at 37 °C at 110 rpm orbital motion for at least 20 hours. The main culture is set up as follows:
Fresh medium is inoculated with preculture so that the main culture has an initial OD_{600} of 0.05 in 200 mL culture volume. The main culture is "default" cultured in 1 L nonbaffled Erlenmeyer flasks at 37 °C and 110 rpm orbital motion; deviations of this default setting are noted separately. In parallel, an uninoculated negative control (20 mL of medium in 100 mL Erlenmeyer flasks) is incubated. Samples are taken regularly, roughly every half an hour, until the culture reaches stationary phase. One millilitre is taken aseptically and measured in the spectrophotometer. Furthermore, the sample is centrifuged and the supernatant is stored at -20 °C for 3,5-dinitrosalicylic acid assay analysis.
When the cells have reached stationary phase, 50 mL of culture is spun down in falcon tubes for 30 min at 4000 rpm (this is done in duplicates). In case the pH is measured, the supernatant is filter-sterilized prior to the measurement. The pellet is dried at 100 °C overnight. Thus, the cell dry weight (DW) is determined from the weight difference of the empty falcon and after the incubation at 100 °C. The ratio of OD_{600}- to-cell dry weight is calculated as follows

$$\frac{DW[mg]}{(OD_{600} * dried\ volume\ [mL])} \quad (4)$$

DW denotes the net bacterial dry weight, which is calculated per final OD_{600} (determined before harvesting) as well as the pelleted culture volume, typically 50 mL.

As the cultivation is performed without any antibiotics, control dilution streak outs on M9 agar plates supplemented with 2 g/L glucose are prepared from a culture in stationary phase to compare the colony morphology.

M9 Minimal Medium

In order to retrieve valid simulation results, it is crucial to use a minimal medium of known composition. The M9 minimal medium used herein[10], is based on Maniatis *et al.* [85]. It is one of the most widely applied minimal medium. However many variations exist regarding e.g. vitamin and trace element supplementation. Even research done for development of *Escherichia coli* models is inconsistent in the composition of M9 medium used, regarding vitamins and trace elements supplementation [36, 37, 45, 86, 87].

Preparation

The medium is composed of four components, which are prepared separately, autoclaved or filter sterilized and finally mixed aseptically:

A. M9 salts (stock solution, either five or tenfold)
B. 1 M $MgSO_4$
C. 1 M $CaCl_2$
D. Substrate solution (40 % glucose)

Components A-C are stored at room temperature, the Substrate solution is stored at 4 °C.

Table 10 Composition of 5 x M9 Stock Solution (pH 7.4)

Component	Concentration in stock solution [g/L]	Final stock concentration [mM]
NaH_2PO_4	30.0	250.0
KH_2PO_4	15.0	110.2
NaCl	2.5	42.8
NH_4Cl	5.0	93.5

Each component is mixed until fully diluted and the pH is adjusted with NaOH pellets.

One litre of M9 minimal medium is prepared by aseptically mixing 20 mL of autoclaved M9 salts, 2 mL of 1 M $MgSO_4$ (final concentration: 2 mM) and 0.1 mL of 1 M $CaCl_2$ (viz. 0.1 mM final concentration). Furthermore the substrate, glucose, is added to a desired final concentration of 2 g/L. The medium is always prepared freshly before use.

[10] http://www.cmdr.ubc.ca/bobh/recipes/MEDIA%20M1.pdf

3,5 Dinitrosalicylic acid assay

In order to monitor the substrate consumption (glucose) during cultivation of *Escherichia coli*, the supernatant was assayed using the 3,5 dinitrosalicylic acid assay according to Miller [88]. *In principio*, 3,5 dinitrosalicylic acid is reduced by reducing moieties of saccharides such as fructose, glucose or galactose and di- or oligosaccharides such as lactose, respectively. Thus, under alkaline conditions, 3-amino-5-nitro salicylic acid is formed, which causes the solution to turn from yellow to various shades of red depending on the extent of reaction. Thus the concentration of reducing sugars can be spectroscopically quantified at 540 nm (Figure 5). This assay is the procedure of choice since it provides an easy handling at low cost. Despite being a relatively simple analytical method when compared to HPLC, it still has a sensitivity of at least 500 nmol of analyte [89]. It should be noted that this method cannot distinguish between different reducing monosaccharides or mixtures of oligosaccharides, thus narrowing the spectrum of possible analytes [89].

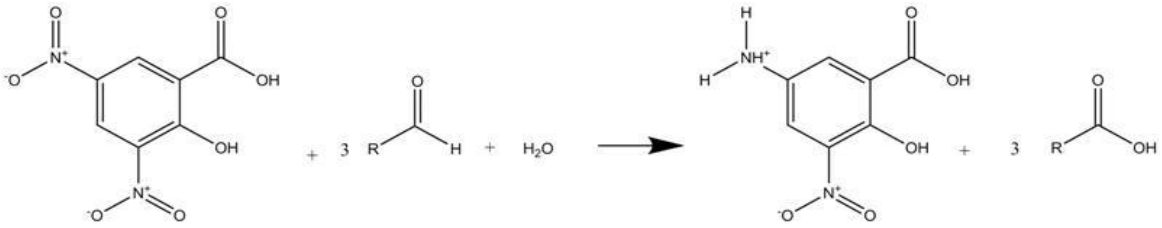


Figure 5 Reaction of 3,5- dinitrosalicylic acid (DNS) with reducing moieties (such as reducing sugars). Under alkaline conditions, 3,5- dinitrosalicylic acid is reduced to 3- amino -5-nitrosalicylic acid, which can be detected at 540 nm. The reducing moiety in turn is oxidized to the respective carboxylic acid.

Preparation

For 100 mL solution, 1 g of 3,5-DNS is dissolved in 20 mL of 2 N NaOH and 50 mL of dH_20. When it is completely dissolved, 30 g of sodium potassium tartrate tetrahydrate are added and the solution is filled up with dH_20. to a total volume of 100 mL.

The solution is stored in a glass vessel covered with aluminium foil to avoid bleaching effects. For disposal, the plates are rinsed with ethanol and disposed into the organic waste flask according to the MSDS. Calibration curves are prepared in the range of 4 g to 0 g glucose. The dilutions are prepared in 1 x M9 medium. In order to quantify reducing sugar moieties, viz. the concentration of substrate in the medium, the DNS assay is carried out as follows:

The culture supernatant, which is stored at -20 °C, is thawn and centrifuged for ten minutes at 8000 rpm. 100 µl of the sample is applied to a deep well plate and 100 µl of DNS reagent is added, whereas each sample is measured in triplicates. The outermost slots of the deep well plates are left empty to get a relative homogenous exposure to heat in the next step. The plate is covered with aluminium foil and placed in a boiling water bath at 98 - 100°C for six minutes.

Once the reaction mix is cooled down to room temperature, 500 µl of dH_20 are added to each slot. Each triplicate was transferred twice to a microtiter plate. Hence, one culture sample is analysed sixfold, to compensate for eventual pipetting mistakes etc. The plate is placed in the spectrophotometer and read out at 540 nm (Spectrophotometer II).

Statistical analysis of data

The specific growth rate µ [hr^{-1}] is calculated by plotting the cell mass, as determined from measured OD_{600} over time. The growth rate µ is readily readable from the slope of the resulting curve, albeit only those data points are considered, where the substrate is not yet exhausted (exponential phase). The specific substrate uptake rate, $q_{s,}$ [mmol/(gDW*h)] is calculated through two different methods, the average and the logarithmic method.

For the average method, the substrate concentration S [mM] is plotted against the biomass concentration X [gDW/L], albeit only those values are considered, that have been included for the estimation of the specific growth rate. The slope of this line is multiplied by the specific growth rate to determine the specific substrate uptake rate q_s.

$$q_S = \frac{\Delta S}{\Delta X} * \mu \qquad (5)$$

The logarithmic method plots the logarithmic ratio of substrate [g/L] per biomass concentration [gDW/L] against the time points. Due to the applied logarithm, the linear range is smaller than that of the average method. The specific substrate consumption rate is derived by calculating the exponent of the slope from the linearization.

$$q_S = \exp\left(\frac{\Delta \ln\left(\frac{S}{X}\right)}{\Delta t}\right) \qquad (6)$$

Determination of the oxygen uptake rate

For determination of oxygen uptake, a fiberoptic oxygen meter connected to a retractable-needle-type microsensor (OXR50) is used.

The detection is based on fluorescence quenching. In principle, a small luminophor spot at the tip of the senor reacts with oxygen molecules, as illustrated in Figure 6. The interaction of triplet

oxygen molecules with the indicator triggers a quenching of infrared emission, which can be detected.

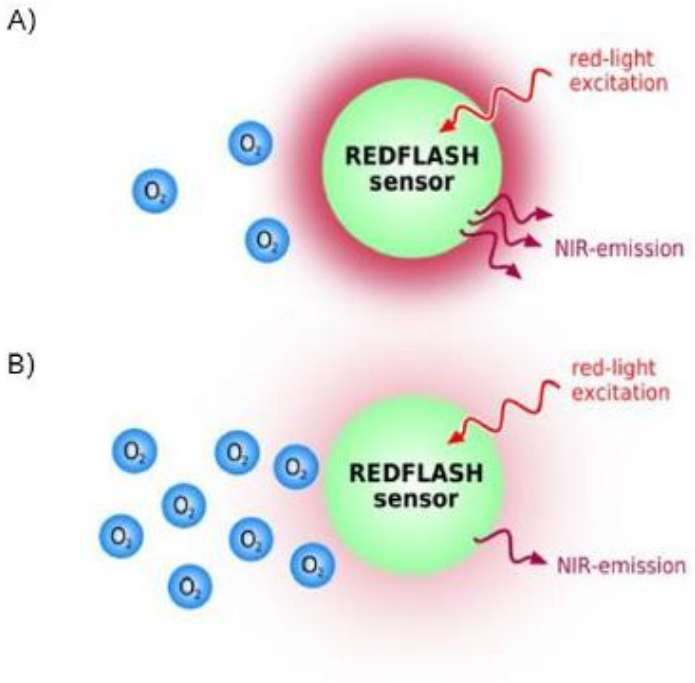


Figure 6 Detection principle of oxygen molecules by the redFlash sensor.
A) If the concentration of oxygen is low the fluorescence emission of the indicatorfluorescence of the luminophore is not quenched, thus emitting in infrared light.
B) At high concentration of oxygen molecules, the indicator emission is quenched.
(figure taken from instructions manual[5])

According to the manual[11], the sensor emits a red excitation light, whereas the resulting emission is recorded due to phase shift. This shift is converted into oxygen concentration or saturation based on Stern Vollmer Theory.

In order to measure the oxygen consumption rate, a 4 mL vial is filled to the top with culture and closed tightly with a cap. The needle-type sensor is inserted into the vial through a tiny hole in the cap and the oxygen decay is recorded. The sensor is calibrated according to the instructions using 2-point calibration, whereas 30 g/L sodium sulfite serves as zero point and environmental air (at known ambient pressure, temperature and humidity) is taken as 100 % point. After usage, the sensor is first rinsed with 70 % ethanol and afterwards with ddH_20.

[11] http://www.pyro-science.com/media/files/Manual%20Firesting%20O2.pdf as accessed 2013-12-20

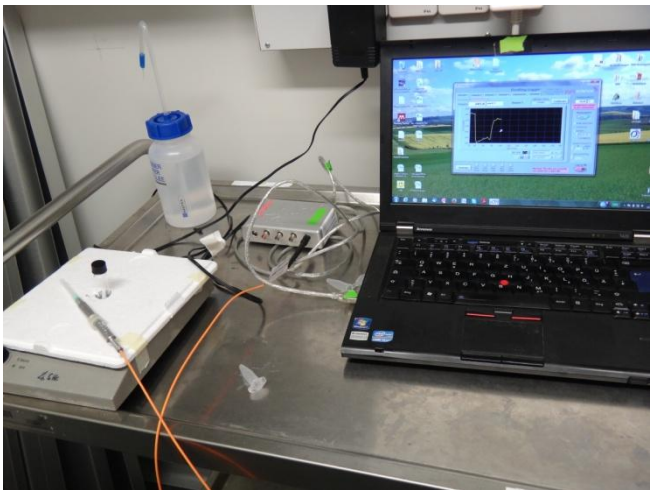


Figure 7 General setup for determination of oxygen uptake rates.
The measurements are undertaken in the 37 °C room (+/- 1 °C), where also the cultivation takes place.
At the left: magnetic plate with styropor and the 4 mL vial placed stably in the middle. In front of the vial lies the microsensor which is connected (orange fiber) to the Firesting oxygen meter (at the back). The oxygen meter in turn is connected via USB to the laptop where the software records incoming signals.

Chapter III Results

In this chapter, the outcome of the thesis work is presented. The first part of this chapter deals with the correction and refinement of the network reconstructions. At the end of this part, they are compared and consequences of strain-specific gene deletions are visualized. Furthermore, a single gene deletion simulation is presented. In addition, *Escherichia coli* cultivations are compared and the gathered data is used as input parameters for model simulation. The last part is dedicated to the influence of different solvers (purchased from different companies) on the optimization result. Raw data from batch fermentations as well as source code written can be retrieved from the (digital) appendix. Data used for plots is tabulated at the end of this thesis. All pathway maps in the following passages are drawn using a slightly modified version of draw_by_rxn.m (*draw_by_rxnMOD.m)* from Paint4Net, a COBRA Toolbox extension [77] (Chapter 2).

Network reconstruction revision process and comparison of final network reconstructions

Revision Process of network reconstructions of *E. coli* K-12 strains

Initially, the network reconstructions have been incomplete, with regard to missing compulsory entries such as bounds and compartment assignments, or the appearance of special characters within some entries, such as “&”,”<”,”>”. This impedes a conversion into SBML files from Excel files or an invalid SBML file is returned. Once corrected, the network reconstructions have been converted into models, which were used for growth phenotype comparison. Apparently, some strains exhibited false-negative growth behaviour, since some gene entries in the network reconstruction of MG1655 were removed to adapt the gene list to other strains. In the course of this, corresponding reactions were also deleted regardless of other isoenzymes annotated to this reaction. Subsequently, a literature recherche using databases including EcoCyc[12], PathCase[13], Pubmed[14], has been conducted to resolve gaps in the network due to missing information. Thereby, a new pathway, aerobic L-ascorbate degradation [90], has been added to the network. Moreover, exchange reactions for acetyl-maltose, a product of overflow metabolism when fed on maltose, and for gamma-hydroxybutyrate, a fermentation byproduct, have been included.
In a third revision, the network reconstructions of MG1655 and W3110 have been compared to published reconstructions of Baumler *et al*. [48].

[12] http://ecocyc.org/
[13] http://dblab.case.edu/PathCaseRCMN/web/About.aspx
[14] http://www.ncbi.nlm.nih.gov/pubmed

The fourth and last revision undertaken in this thesis has fundamentally improved the predictive capacity of gene essentiality analysis: The conversion to "RAVEN format" has confused isoenzymes (gene names separated by a ";") and enzyme complexes (separation by ":"), respectively in W3110 and RV308 network reconstructions. The functional consequences of this confusion is immense, since the deletion of an enzyme within a complex erases the reaction within the pathway, if no further alternative isoenzyme is annotated. Thus, the gpr lists, containing information gene-protein relationships, needed to be revised. Furthermore exchange metabolites have been assigned a distinct compartment: "boundary". This indicates metabolites crossing system boundaries according to RAVEN convention.
The revision process is summed up in Figure 8, whereas a detailed list of changes to the reconstructed models can be found in the digital appendix.

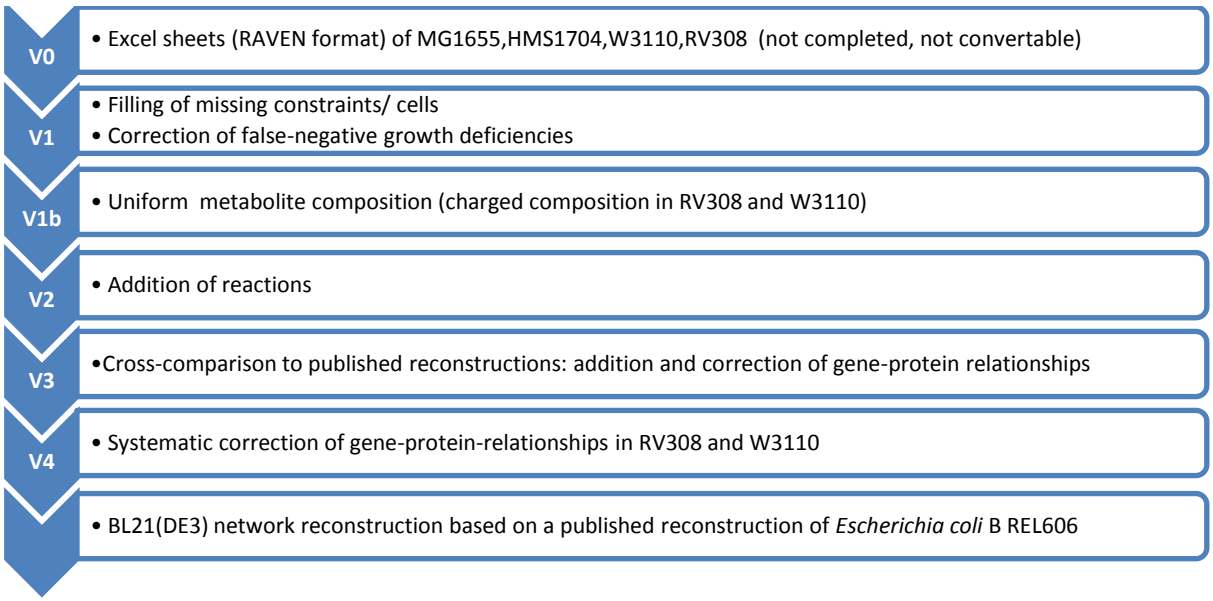


Figure 8 Overview of changes made to the initial network reconstructions at a glance.
During the course of this thesis, the network reconstructions have been corrected and extended. Changes made to the networks are summarized as described in the main text. These changes have been substantial and crucial for functional performance.

Network reconstruction of BL21(DE3)

Additionally, a BL21(DE3) reconstruction has been set up by adapting small changes from the REL606 network reconstruction published by Yoon *et al.* [73] and a sequence comparison between BL21(DE3) and REL606[15].

Besides differences in flagella biosynthesis or proteases which are not included in the network, the genome sequence also revealed, that there are also metabolic differences such as an arabinose uptake system as well as a B-specific cluster for degradation or aromatic compounds, such as degradation of 3- and 4-hydroxy phenyl acetic acid. Contrarily, REL606 lacks evidence for synthesis of certain oligosaccharides of the lipopolysaccharides as well as fatty acids and O-antigen synthesis. These reactions have been removed from the network reconstruction.

The BL21(DE3) network reconstruction is based on the suggested changes mentioned Yoon *et al*. However, while Yoon *et al*. used the iAF1260 network reconstruction of MG1655, the basis network reconstruction used herein is iJO1366, as mentioned in Chapter 2. Further genome comparisons revealed only little metabolic differences between REL606 and BL21(DE3).

Accordingly, small differences regarding galactose degradation or iron-sulfur cluster biosynthesis have been implemented. More detailed information on the changes can be found in the corresponding file.

Comparison of final network reconstructions of K-12 strains

Roughly 50 genes (depending on the strain) have been added to the network reconstructions, equaling an extension of the gene lists of 3 %, encoding for 30 new reactions. However these "raw" reconstructions are not consistent, since they contain gaps due to lacking source or sink metabolites and therefore cannot carry flux. Taking such cases out of account, the consistent reconstructions are roughly 13 % smaller.

Table 11 Overview of final network reconstructions of MG1655 and W3110.

Entry numbers of reactions, metabolites and genes are listed for the full reconstruction, containing also entries, which are never used. The numbers for the consistent network reconstructions as well as numbers of unused entries are also listed.

Strain	MG1655			W3110		
Model	**Raw**	**Consistent**	**Unused**	**Raw**	**Consistent**	**Unused**
Reactions	2664	2407	257	2657	2387	270
Metabolites	1826	1606	220	1826	1593	233
Genes	1462	1325	137	1455	1314	141

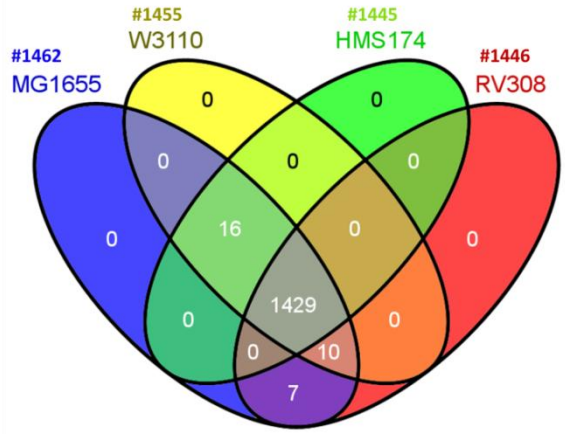
[15] Files of insertions/ deletions (indels) between REL606 and BL21(DE3) were kindly provided by Peter Krempl (Austrian Center for Industrial Biotechnology ACIB)

Table 12 Overview of final network reconstructions of HMS174 and RV308 and BL21(DE3)
Entry numbers of reactions, metabolites and genes are listed for the full reconstruction, containing also entries, which are never used. The numbers for the consistent network reconstruction and numbers of unused entries are also listed.

Strain	HMS174			RV308			BL21(DE3)		
Model	**Raw**	**Consistent**	**Unused**	**Raw**	**Consistent**	**Unused**	**Raw**	**Consistent**	**Unused**
Reactions	2651	2374	277	2647	2361	286	2646	2305	341
Metabolites	1824	1590	234	1820	1574	246	1833	1539	294
Genes	1445	1298	147	1446	1298	148	1431	1243	188

Figure 9 illustrates, that none of the K-12 strains compared herein possesses unique strain-specific genes. Once, the network reconstructions are reduced to the active entries, the total number of genes decreases, indicating, that many genes are not connected to the general metabolism. Furthermore, the number of “none-core” genes increases, as upon gene deletion according to sequence comparisons, so that further “collateral” genes in a pathway are inactivated. The deletions of genes due to assumed sequence differences lead to the deletion of whole pathways as artificial dead-end and no-source reactions, respectively.

A



B

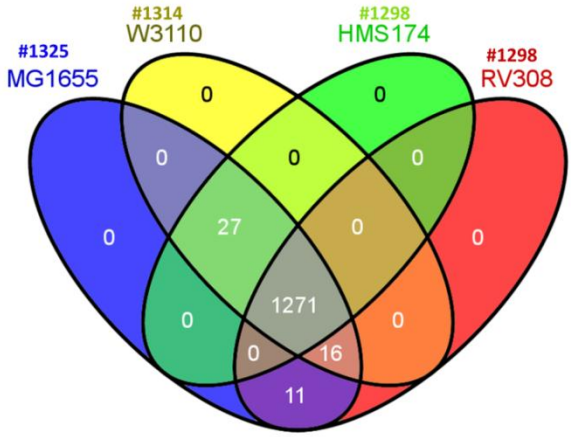


Figure 9 Comparison of gene list entries in *Escherichia coli* K-12 network reconstructions
Some reactions (and associated genes) not connected with the network. Some are never used due to gaps in the network, leaving it partly inconsistent. Therefore strain-specific gene deletions "passively" inactivate other genes in the affected pathways. As a gene deletion may inactivate genes in the common core set for the number of genes shared by only two or three of the strains increases.
A) Venn diagram of numbers of genes in the full/inconsistent network reconstruction
B) Venn diagram of genes in the consistent network reconstruction.
The Venn Diagrams have been generated using VENNY (by Juan C. Oliveros)
http://bioinfogp.cnb.csic.es/tools/venny/ as accessed February 10th, 2014

Comparing the Venn diagrams (Figure 9), a difference of four genes among common genes of MG1655 and RV308 becomes obvious: The deletion of a galactitol transporter (reaction "R_GALTptspp" b2092:b2093:b2094) leads to the inactivation of consecutive reactions (Figure 10). Since this transporter is the only source for galactitol-1 phosphate, it will therefore no longer be processed to glyceraldehyde 3-phosphate and dihydroxyacetone phosphate, both metabolites of central carbon metabolism. Another example is allose uptake, which is entirely missing in W3110 and HMS174 (pathway is not shown here).
An additional example is 2-Oxopent-4-enoate (M_op4en[c]), which has two influx branches: one from 3-(2,3-Dihydroxyphenyl)-propanoate and the other one from 2,3-dihydroxicinnamic acid. As the genes b0347, b0348 and b0349 do not exist in the genome sequence of RV308, the reactions "R_3HCINNMH","R_3HPPPNH", „R_DHCINDO", „R_HPPPNDO", „R_HKNDDH", „R_HKNTDH" as illustrated in Figure 10 are erased from the pathway and so are all upstream reactions, having 2-Oxopent-4-enoate as single sink metabolite, respectively.

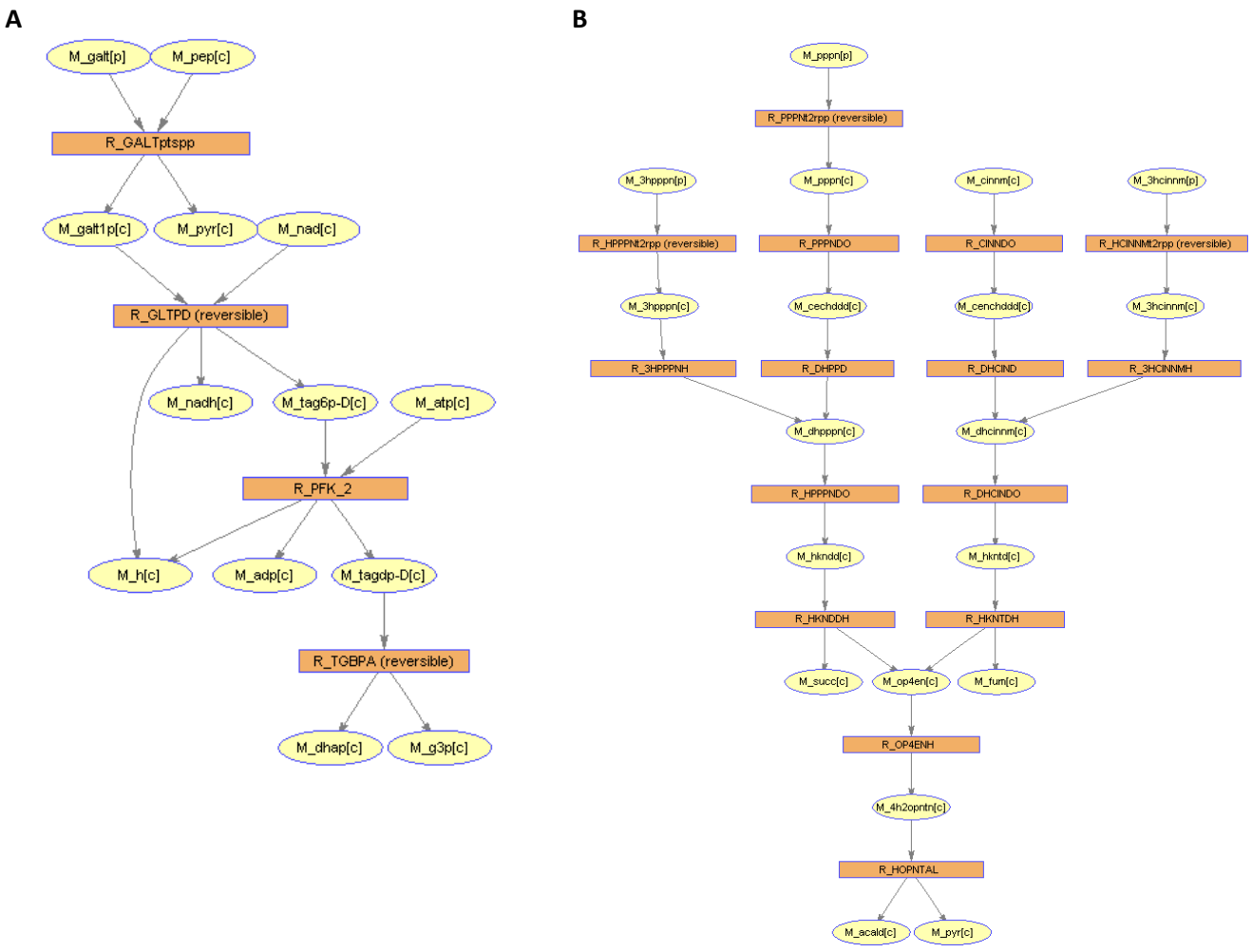


Figure 10 Pathways which are inactive in network reconstructions of W3110 and HMS174.
A: Galactitol catabolism. B: 3-phenylpropionate and 3-(3-hydroxyphenyl)propionate degradation Some bulk metabolites such as H_2O, and NADP/NADPH have been removed to reduce complexity of the figure. Details on the reactions can be retrieved directly from the network reconstructions.

Regarding the function of the inactivated genes, Figure 11 reveals that mostly genes are affected which are involved in the biosynthesis of cofactors and prosthetic groups. Since the network reconstruction of MG1655 lacks recycling reactions, creating dead ends that cannot carry flux. Another noticeable proportion of genes is involved in transport processes or is not assigned to any specific pathway.

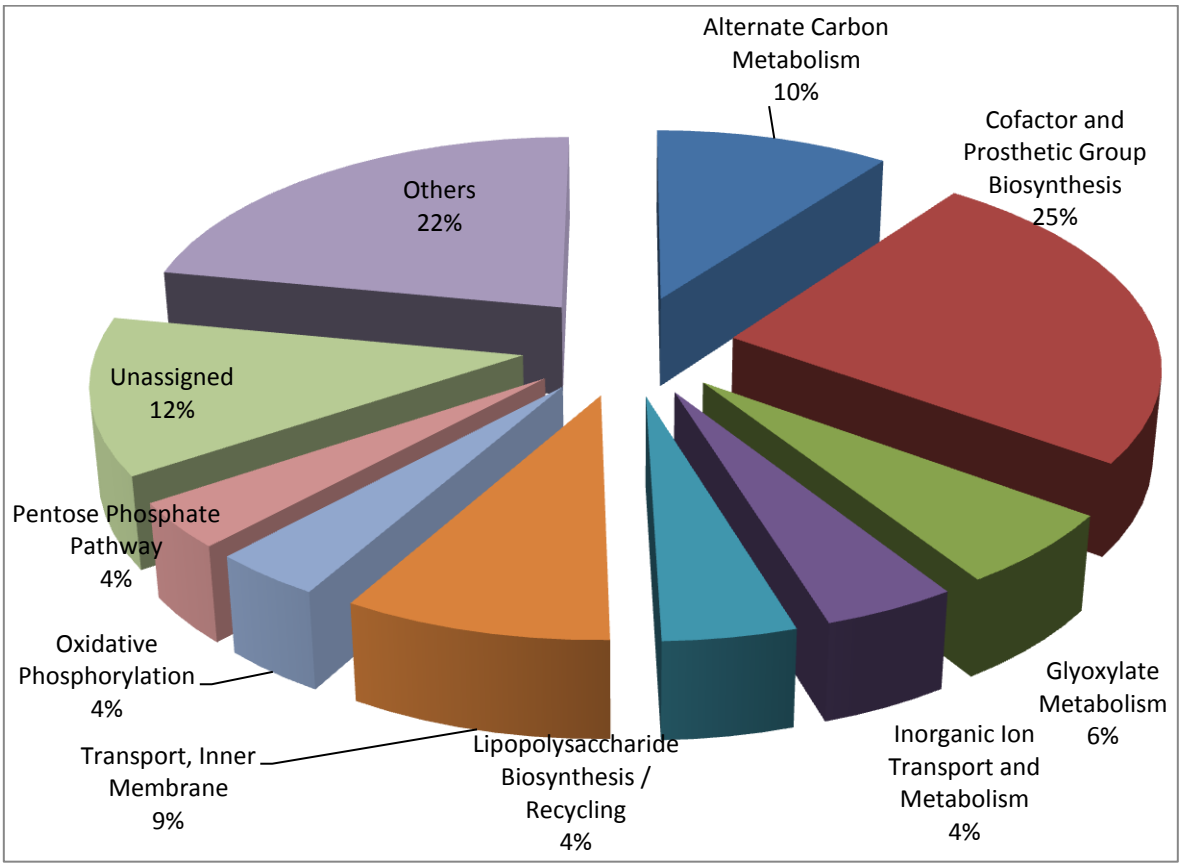


Figure 11 Inactive genes sorted by subsystem affiliation in the consistent model of MG1655. Subsystems are a way to functionally cluster genes. In this figure, the inactive genes found in the consistent model of MG1655 (iJO1366_raven.V4.xml) are clustered per subsystem affiliation. Most of the inactive genes belong to cofactor biosynthesis pathways. Furthermore genes involved in transport reactions or without distinct classification are affected. Smaller fractions (below 3%) such as metabolism of amino acids and nucleotides are cumulated (22%). A similar proportion of affected subsystems could also be observed for the other models (W3110, HMS174, RV308 and BL21(DE3)).

Data concerning model (in-) consistency and inactive genes can be found in the appendix ("ConsistentModel.xlsx").

Gene essentiality analysis

In order to investigate major physiologic differences, the *Escherichia coli* models have been subjected to a series of *in silico* single gene deletions. The algorithm used for calculation is FBA, however, there exist more elaborate algorithms such as regulatory on/off minimization (ROOM) [91] or minimization of metabolic adjustment [92], which predict mutant growth rates from gene knock outs more precisely than FBA. However, these algorithms are also more time-intensive.
The objective herein is, to find unique essential genes, which are lethal in only one of the models, despite only little differences in the gene lists.

Table 13 Effect of *folD* (b0529) knock out under aerobic conditions.
The crosses indicate a fatal knock out. This knock out has been performed under aerobic conditions using FBA algorithm. The script KOv5.m can be found in the appendix. This gene deletion is not lethal, when performed under anaerobic conditions (therefore not shown here).

substrates/models	MG1655	W3110	HMS174	RV308	BL21(DE3)
acetate				x	
acetaldehyde				x	
α-ketoglutarat				x	
ethanol				x	
D – fructose				x	
fumarate				x	
D-glucose	x	x	x	x	x
L –glutamine				x	
L- glutamate				x	
D-lactate				x	
L-malate				x	
pyruvate				x	
succinate				x	

Upon single gene deletion, one interesting essential gene has been found (Table 13), namely *foldD* (b0529), which is involved in folate metabolism. Folates are essential cofactors that mediate the transfer of one-carbon units from donor molecules in many pathways. Thus, they are involved in amino acid metabolism, such as methionine biosynthesis, interconversion of serine to glycine and vice versa, and histidine catabolism [93].

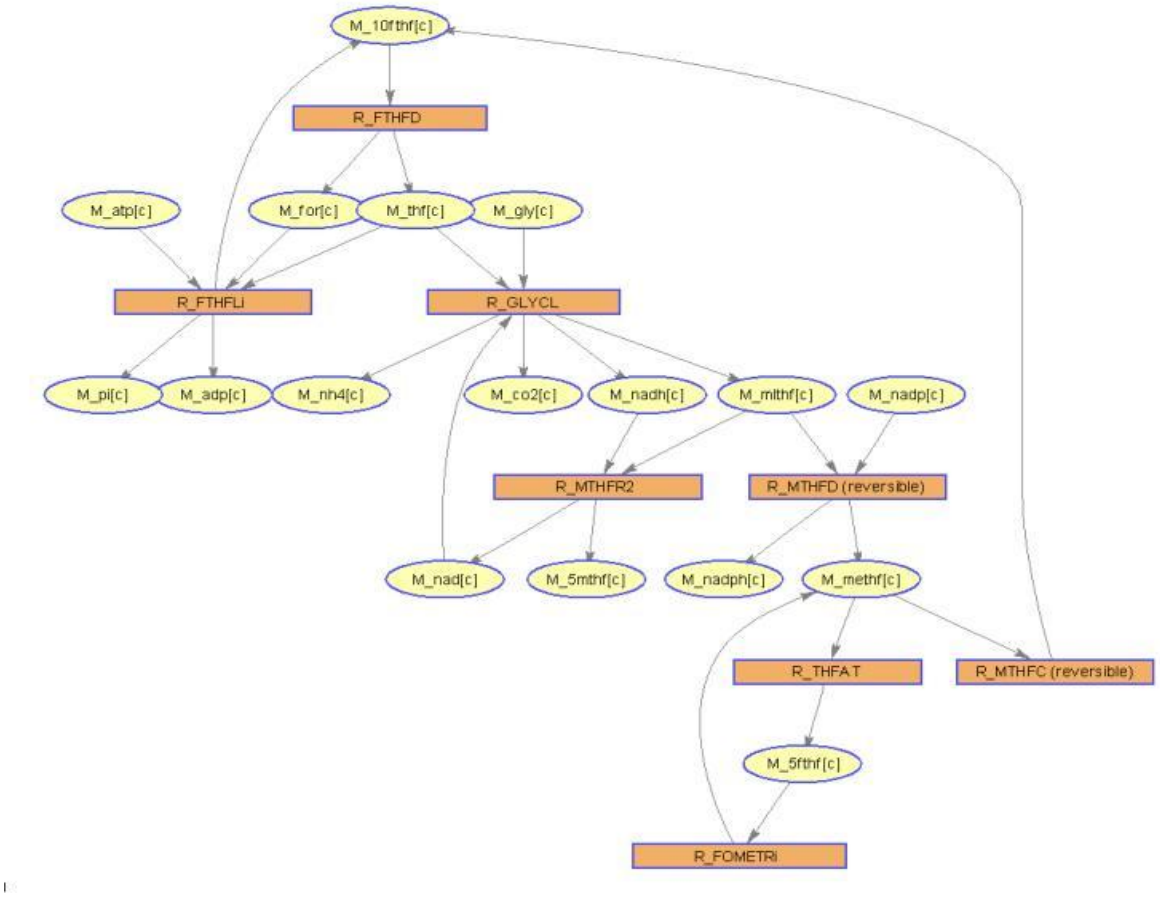


Figure 12 Formyltetrahydrofolate metabolism in *Escherichia coli* K-12.
Briefly, derivates of folate can be easily interconverted into one another. The formyltetrahydrofolate deformylase (R_FTHFD; top of the figure) produces tetrahydrofolate (M_thf[c]) from 10-Formyltetrahydrofolate (M_10fthf[c]), which can be back-converted under ATP consumption by formate-tetrahydrofolate ligase (R_FTHFLi). 10-Formyltetrahydrofolate can be furthermore produced from 5,10-Methenyltetrahydrofolate (M_methf[c]) which in turn could also be metabolized to 5-Formyltetrahydrofolate (M_5fthf[c]).
The gene of interest, *folD* (b0529), encodes a bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase (R_MTHFC, R_MTHFD), which catalyzes the interconversion of 5,10-methenyltetrahydrofolate to 10-formyltetrahydrofolate and the NADP dependent interconversion of 5,10-methylenetetrahydrofolate to 5,10-Methenyltetrahydrofolate.
(Drawn according to subsystem "folate metabolism" of the network reconstruction; protons and H_2O are not shown in this figure).

Figure 12 illustrates a part of the folate metabolism, more precisely: metabolism of folate derivates. The gene, which is knocked out in this *in silico* experiment is *folD* (b0529) which encodes a bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase involved in reactions "R_MTHFC" and "R_MTHFD". It catalyzes the interconversion of 5,10-methenyltetrahydrofolate to 10-formyltetrahydrofolate and the NADP-dependent interconversion of 5,10-methylenetetrahydrofolate to 5,10-Methenyltetrahydrofolate. Knocking out *folD* hence leads to a fatal lack of supply in 10-formyltetrahydrofolate and 5,10-methylenetetrahydrofolate.

The absence of the 2- methylcitrate cycle (Figure 13) sets *Escherichia coli* RV308 apart from the other network reconstructions:

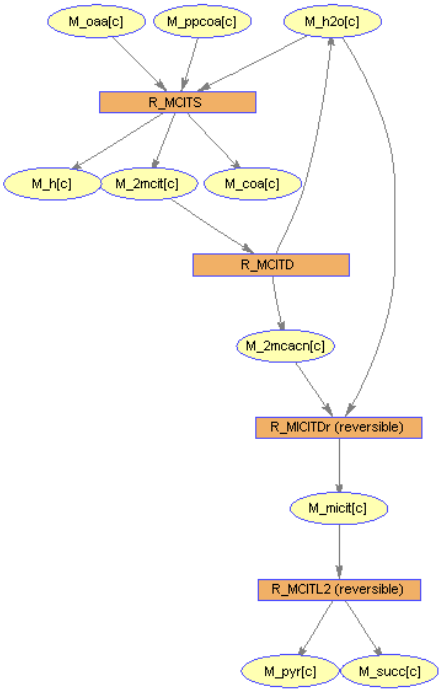


Figure 13 2-methylcitrate cycle of *Escherichia coli.* Oxaloacetate and propanoyl-CoA are transformed to methylcitrate and coenzymeA by 2-methylcitrate synthase (b0333). Methylcitrate in turn is converted into cis-2-methylaconitate by 2-methylcitrate dehydratase (b0334). 2-methylisocitrate dehydratase catalyzes the reversible conversion between 2- methylaconitate and methyliscocitrate (b0118), which can methylisocitrate lyase (b0331) then catabolized to succinate and pyruvate, two bulk metabolites to fuel central carbon metabolism [94].

This pathway seemingly has to have a compensating effect, when taking into account, that for the other models, growth was predicted for 12 out of 13 substrates (Table 13). This raises the question, why the knockout of *folD* has a broader fatal effect, if the 2- methylcitrate cycle is absent under aerobic growth condition. Therefore a series of control simulations have been performed, which are collected in the script clarifyKO.m, and following further observations have been made:

In the simulation, RV308ΔfolD does not grow under aerobic conditions. Similarly, the other models also do not grow, if one of the reactions of the cycle is deleted. However, pyruvate and

succinate substitution (artificial drains) do not restore growth, strengthening the hypothesis, that this pathway might be redundant. Substitution of an intermediate byproduct, acetyl-CoA, does not lead to restored growth either. Therefore the bottleneck has to be at the beginning of the pathway, rather than its end: This pathway is a sink for oxaloacetate and propanoyl-CoA, albeit the first one named, is a bulk metabolite, whilst the other one is not.
The removal of one single reaction of the pathway creates a dead end, and therefore inactivates all reactions involved in the biosynthesis of propanoyl-CoA. Interestingly, growth is also restored by enabling free exchange of formate, which is a fermentation byproduct and further metabolized in formyltetrahydrofolate metabolism as illustrated in Figure 12. These findings suggest that the inactivation of reactions due to the dead end at propanoy-lCoA has a cross-effect on the metabolisation of formate. When aerobic conditions are simulated, reactions regulating the production of (anaerobic) fermentation byproducts are disabled, so that there is no further sink for formate except for non-aerobically regulated reactions such as folate biosynthesis.

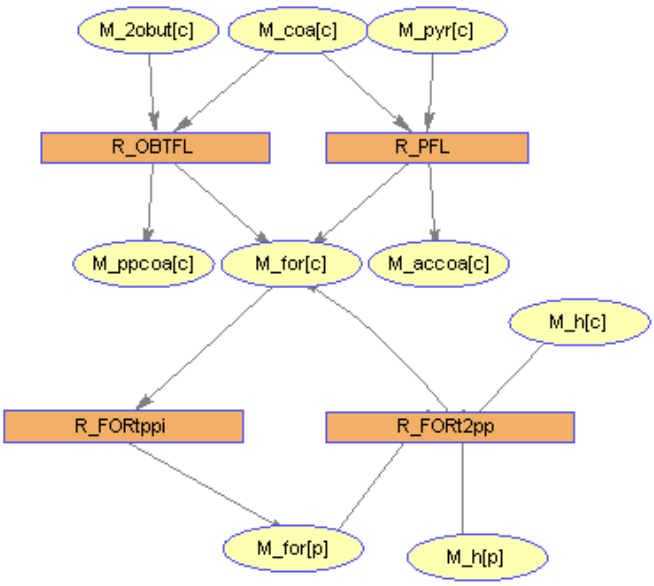


Figure 14 Crucial reactions affected by oxygen regulation.

Upon oxygen regulation, 152 are shut off (when growth on glucose is simulated; regulateModelv3.m). However, only those listed beside are crucial for *in silico* vitality.
All reactions in the network need a source and a sink in order to carry flux. In case of *folD* knockout, there is no further source for formate which is unaffected by oxygen regulation. Furthermore there has to be a sink for propanoyl-CoA as well (see Figure 13).
2-oxobutanoate formate lyase (R_OBTFL) metabolizes 2-Oxobutanoate (M_2obut[c]) and CoA (M_coa[c]) to formate (M_for[c]) and propanoyl-CoA (M_ppcoa[c]). Alternatively, formate is synthesized by pyruvate-formate-lyase (R_PFL) from CoA and pyruvate (M_pyr[c]), thereby producing Acetyl CoA (M_accoa[c]). Furthermore formate can also be im-or exported by formate transporters (R_FORtppi,R_FORt2pp).

Summarizing, the lethality of *folD* is of systemic nature and not due to a trivial metabolite essentiality. Moreover, when glucose is used as substrate for growth simulation, a more stringent regulation is applied on the models, according to Covert *et al.* [95]. Hence, the compensational effect of the 2-methylcitrate cycle (as sink for propanoyl-CoA) gets extinct as formate becomes a dead end metabolite as well and therefore the knockout of *folD* is lethal for all models on glucose under aerobic conditions.

Escherichia coli cultivations

The different *Escherichia coli* K-12 and B strains have been cultivated and analyzed as described in Chapter 2. The growth rates have been determined by OD_{600} measurements and dry weight determinations. Thereby a conversion factor OD_{600}-to-gDW/L (gram dry weight per litre) determined is used for all batches and all samples, which has also been mentioned in the literature[36]. The glucose uptake rate has been measured using the DNS-assay. The sample values can be retrieved from the appendix.

First of all should be noted, that the calculation of the specific substrate uptake rate q_s [mmol/(gDW*L)] is error prone, if calculated incorrectly. Figure 15 illustrates the effect of different calculation methods on q_s. The "average method", where the inversed growth yield per substrate is multiplied with the growth rate μ, is often used because of its simplicity. The "logarithmic method" as proposed by Görgens *et al.* [96], calculates the consumption rate from a logarithmic linearization. The detailed formulas for calculation can be found in Chapter 2.

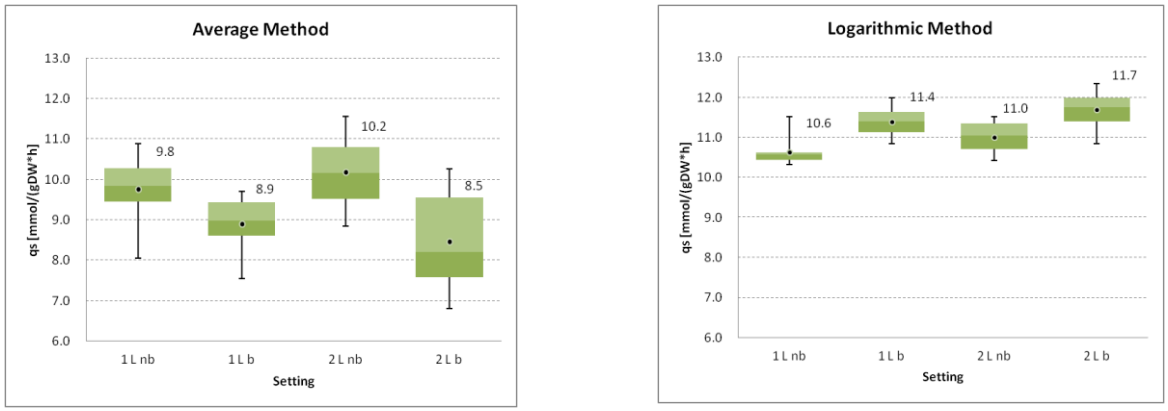


Figure 15 Comparison of calculation methods for determination of specific glucose uptake rates.
The two determination methods (as described in Chapter 2) are compared for MG1655 at different settings. The strain has been cultivated in 1 litre non-baffle "nb" (10 batches), baffled "b" (7 batches) shaking flasks, as well as in 2 L unbaffled flasks (4 batches) and baffled flasks (11 batches) at 37 °C and 110 rpm orbital motion. The average values are marked black and displayed beside each data series. The dark green area represents the 25th percentile and the light green one the 75th percentile.

As can be seen in Figure 15, q_s values vary depending on the calculation method used. While the values calculated with the logarithmic method show a trend towards higher rates with higher flask volume and baffle, the average method returns values that have a higher coefficient of variation and are even contradictory to the other method. Therefore all substrate uptake rates further used in this thesis are calculated using the logarithmic method. Switching the cultivation setting to baffled flasks and bigger flask volume influences the substrate uptake rate by improved mixing and better aeration thereby causing higher growth rates. However the overall increase when changing the settings from 1 L unbaffled to 2 litre baffles are relatively small.

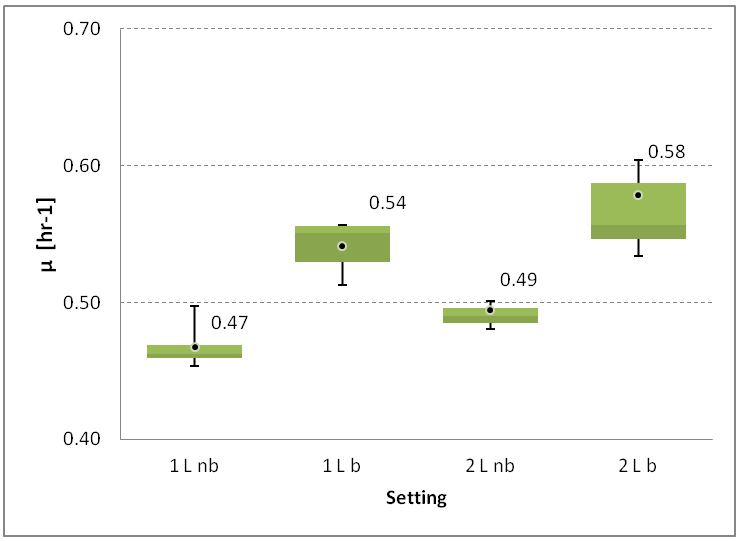


Figure 16 Comparison of growth rates μ at different settings
Escherichia coli MG1655 has been cultivated at different setting (see Figure 14) in M9 minimal medium (supplemented with 2 g/L glucose) at 37 °C and 110 rpm. Tabulated values can be retrieved from the appendix (starting with Table 19). The dark green area represents the 25th percentile and the light green one the 75th percentile.

As expected, the growth rates show a similar tendency, towards faster growth, when the setting is switched to baffled flasks or 2 litre flasks, through the increase seems to be more obvious than observed with the substrate consumption rates.

In order to further address strain-specific growth differences, MG1655, W3110, HMS174, RV308 and BL21(DE3) have been cultivated at default settings, 1 litre nonbaffled, on 2 g/L initial glucose concentration in M9 minimal medium as described in Chapter 2. The resulting growth rates and substrate uptake rates are compared in Figure 17 and Figure 18, respectively.

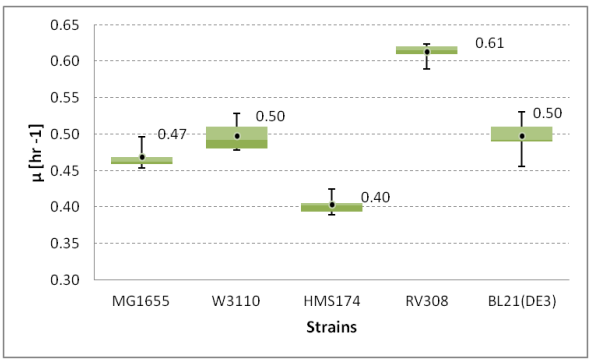


Figure 17 Growth rates (μ) of different *Escherichia coli* strains. Strains cultured in 200 mL M9 supplemented with 2 g/L glucose, at 37 °C and 110 rpm orbital motion.

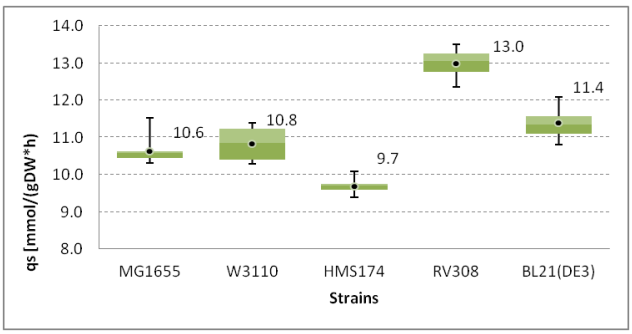


Figure 18 Specific substrate uptake rates (q_s) of different *Escherichia coli* strains. Strains cultured in 200 mL M9 supplemented with 2 g/L glucose, at 37 °C and 110 rpm orbital motion.

Both growth rate and glucose uptake rate show similar tendencies: the closely related K-12 strains MG1655. W3110 show similar growth and glucose uptake rates. While exhibiting a similar growth rate, BL21(DE3) has an slightly elevated glucose uptake rate. Contrarily, HMS174 and RV308 both grow at lower and higher rates, respectively. Tabulated values can be found in the appendix.

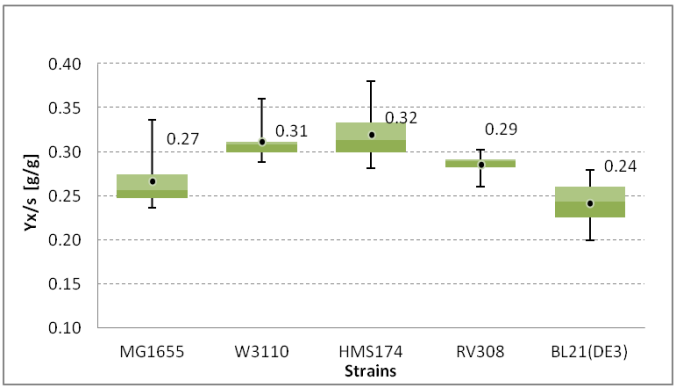


Figure 19 Comparison of biomass yield from glucose between the strains.
Regarding the biomass yield per substrate $Y_{x/s}$ [g/g] – determined by plotting biomass concentration over glucose consumption, there seems no clear tendency, given the high variances (up to 11 %) of the data.

Additionally to substrate supply, the oxygen consumption plays a crucial role for bacterial growth rates. Since the simple shake flask setting does not allow for continuous monitoring or supply of oxygen, the oxygen consumption has been determined by means of a microsensor as described in Chapter 2. Since oxygen measurements during the course of cultivation (and sample taking) would have come at the expense of sample accuracy (regarding glucose consumption), the oxygen uptake rate has been measured in separate cultivations. Therefore, one measurement has been undertaken for MG1655 at exponential growth phase for each of the following settings: 1 L non baffled and baffled flask and 2 L baffled.

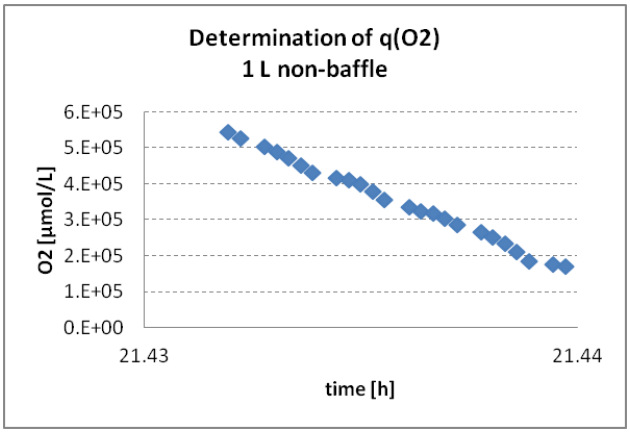


Figure 20 Oxygen consumption of 1 L non-baffled culture cells in exponential growth phase. Settings: M9 supplemented with glucose, 37°C orbital motion. At an OD_{600} of 1.8 (0,54 gDW/L) , a oxygen uptake rate of 25,7 mmol/[gDW*h] was retrieved (R^2 = 0,9962)

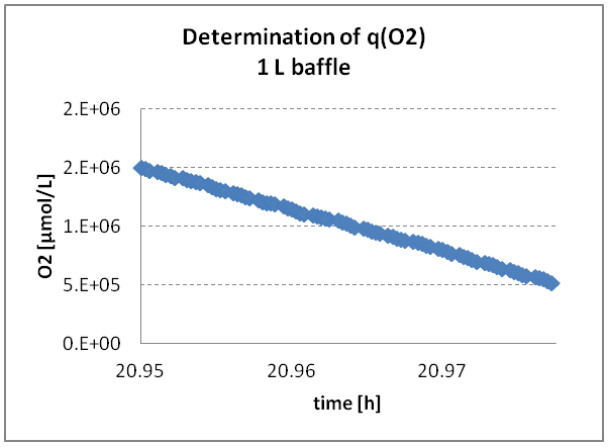


Figure 21 Oxygen consumption of 1 L baffled culture cells in exponential growth phase. Settings: M9 supplemented with glucose, 37°C orbital motion. At an OD_{600} of 1.4 (0,42 gDW/L) , a oxygen uptake rate of 26,5 mmol/[gDW*h] was retrieved (R^2 = 0,9997)

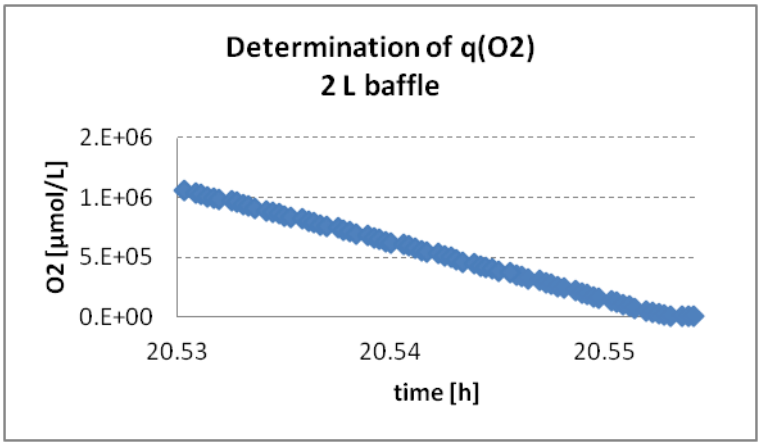


Figure 22 Oxygen consumption of 2 L nonbaffled cultured cells in exponential growth phase.
Settings: M9 supplemented with glucose, 37°C orbital motion. At an OD_{600} of 1.8 (0,54 gDW/L) , a oxygen uptake rate of 25.7mmol/[gDW*h] was retrieved (R^2 = 0,9995)

In Figure 20 – Figure 22 the oxygen concentration is illustrated for each setting. The resulting oxygen decay over time is summarized in Table 28 (appendix) for the different settings.

As can be seen above, the oxygen consumption is similar among the settings. This somehow "contradicts" the growth trend exhibited at different settings (see previous figures). However the values are in agreement with published values. Andersen and Meyenburg [97] for example determined a maximal oxygen uptake rate of roughly 20 mmol/(h*gDW) for *Escherichia coli* B/r

NF790, when grown on minimal medium in batches with different oxygenation rates. This publication was already used for studies on the influence of varying oxygenation rates on the metabolic model of *Escherichia coli* by Varma and Palsson [35]. As the values retrieved from the measurements lie above the postulated maximal rate, the oxygen uptake rate is set unconstraint for model simulation (see following paragraphs).

Model evaluation – comparison of experimental and simulated results

In this part, the model is compared to the experimental results. Detailed information on the function can be retrieved from the individual toolboxes or from the appendix, in case the function has been written in course of this thesis work. All of the following calculations have been performed using the Gurobi5 solver.
As to evaluate the model with the experimental results, the gathered values were used as constraints for the model simulation: Oxygen is set unconstraint, glucose uptake rates (q_s) are set accordingly.

Dynamic FBA simulation of MG1655 cultivation

First of all, it is important to elicit, which of the growth conditions fit best to the simulation.
To this purpose, the uptake rates are modified in the model and the simulated growth rate μ [1/h] is compared to the experimental one. As can be seen in Table 14 below, the simulated result is more than twice the value as the corresponding experimental setting.

Table 14 Comparison of experimental and simulated growth rates at different substrate uptake rates (q_s). Experimentally determined values have been used as input parameters for growth simulation. Settings: model: iJO1366_raven.V4.xml; oxygen bounds unconstraint, q_s set accordingly, function used for simulation: optimizeCbModel.m (COBRA Toolbox).

Setting	Cultivation μ [1/h]	Simulation μ [1/h]	Relative difference (%) (cultivation and simulation)
1L nb	0.47	1.03	75
1L b	0.54	1.11	69
2L nb	0.49	1.07	74
2L b	0.58	1.13	64

Apparently the growth rates do not vary to the same extent with changing settings as the experimental rates. This furthermore indicates that the growth rates are not exclusively based on changing substrate uptake rates but also on other parameters such as oxygen supply. Also the difference between simulation and experiment is smallest for the cultivation in 2 litre baffled flasks. This supports the assumption that this setting provides the best growth conditions through improved mixing and oxygen supply among all settings tested.

A ruled out previously, the measured oxygen uptake rates do not fit to the growth trend observed. As the measurements have not been undertaken in situ but treated like isolated samples, which possibly lead to a falsification. Therefore the oxygen supply variability is tested, albeit experimental determined biomass and glucose values serve as input parameters and oxygen supply is minimized and maximized.

According to the optimization (optimizeCbModel.m), the oxygen uptake rate can vary in between 7.2 and 294.6 mmol/(gDW*h), at the given glucose supply for the observed growth rate. Therefore the experimentally determined values (Figures 20 - 22) lie in full range.

At the end of cultivation, decay in pH (roughly 0.5 pH units) was observed, which indicate the production of acidic byproducts. These byproducts have not been measured, since the emphasis of this thesis laid on refinement of the reconstructions, the laboratory setup was therefore kept simple.

Table 15 lists initial medium composition as well as major secretion fluxes during growth simulation the medium components. The simulated flux rates are retrieved from the exponential phase which assumes “optimal conditions”.

Table 15 feed composition.
The bounds of uptake fluxes (substrates) are mostly unconstraint. Still, the range is not fully utilized for optimal growth. Model: iJO1366_raven.V4.xml.Constraints as below, biomass objective function (c = 1) to the experimental growth rate algorithm:optimizeCbModel.m.

Substrate	Uptake bounds	Uptake fluxes
ca2	-1000	-0.0030
cbl1	-0.01	0.0000
cl	-1000	-0.0030
co2	-1000	8.3950
cobalt2	-1000	0.0000
cu2	-1000	-0.0004
fe2+	-1000	-0.0048
fe3+	-1000	-0.0045
glc	-11.6	-11.6000
h	-1000	19.7888
h2o	-1000	23.6456
k	-1000	-0.1128
mg2	-1000	-0.0050
mn2	-1000	-0.0004
mobd	-1000	-0.0001
nh4	-1000	-6.2385
ni2	-1000	-0.0002
o2	-7.16	-7.1600
pi	-1000	-0.5573
so4	-1000	-0.1457
zn2	-1000	-0.0002

Table 16 Major secreted metabolites.
Simulated growth rate and growth byproducts are listed below. Settings as in Table 15 feed composition.

Secreted metabolite	Flux rate [mmol/(gDW*h)]
h2o	52.95
co2	33.57
h	14.86
succ	0.61
biomass growth [h-1]	0.58
ac	0.49
cyt	0.22
thymd	0.19

Table 17 Minimal and maximal fluxes [mmol/(gDW*h)] for secreted metabolites
Flux variability analysis for secreted metabolites as listed in Table 16 retrieved feasible minimal and maximal simulated flux rates.

Secreted metabolite	Min flux [mmol/(gDW*h)]	Max flux [mmol/(gDW*h)]
h2o	22.25	602.87
co2	1.53	101.48
h	-1000	535.74
succ	0	8.65
biomass growth [h-1]	0.58	0.58
ac	0	17.49
cytd	0	3.62
thymd	0	3.03

In Table 16 major secretion product rates are listed. However, more than 300 secretion fluxes have been computed, of which the secretion rate was negligible, e.g. fmol/(gDW*h) and rather a byproduct of fierce numerical operations than of significance in this case.
Besides biomass and carbon dioxide, water and acetate, also cytidine and thymidine have been produced. Although every internal metabolite could be theoretically found in the media e.g. upon necrosis [98], the rates of cytidine and thymidine are unexpectedly high. Therefore flux

variability analysis (FVA) has been conducted to investigate the allowed flux span. As can be seen in
Table 17, the secreted values vary substantially. However it is allowed for acetate to reach also higher levels than returned initially (Table 16).

Comparison of growth predictions of different models

Similar to the simulations of MG1655 introduced previously, biomass growth has also been simulated for different strains, oxygen supply is left unconstraint and the specific substrate consumption rate q_s is used as input parameter, the biomass objective is set to 1. This setting assumes that no further byproduct is produced to a larger extent since the biomass flux is optimized, Even though this setting is incorrect if no further constraints for secretion byproducts are set, it is accepted in this case due to lacking data.

According to the simulations (Table 18) relative difference between simulation and cultivation is smallest for RV308 and biggest for BL21(DE3). Moreover the high discrepancies have been expected as already ruled out previously, that these settings are not optimal.
It should furthermore pointed out, that the acetate secretion rate of BL21(DE3) is significantly lower than for the other strains [99]. It should be furthermore mentioned that a uniform biomass composition was for all strains has been assumed for matter of simplification, which is unrealistic as there exist obvious physiological differences within the strains.

Regarding these simulations, different biomass objective functions of different network reconstructions of MG1655 have been used. For example the “core” biomass function only contains a reduced common core set of metabolites used in *Escherichia coli* cells. When the more detailed “wild type” biomass objective function is applied, the difference even increases. Most interestingly, the relative difference is slightly decreased, when the BOF of iAF1260 is used, since it is an precursor reconstruction of iJO1366 and the biomass composition has been refined for iJO1366.

Table 18 Comparison of simulated and experimentally determined growth rates using different biomass objective functions

The biomass objective function has an impact on simulated growth rates. The core biomass function is a reduced a version of a wild type composition, necessary to account for common components among average *Escherichia coli* strains [42]. Settings as used in Table 15, glucose uptake rates as in Figure 18.
Models: iJO1366_raven.V4.xml (MG1655), iW1407_raven.V4.xml (W3110), iHMS1391_raven.V4.xml (HMS174), iRV1394_raven.V4.xml (RV308), iBL_raven.V4.xml (BL21(DE3)).

Biomass objective function from network reconstruction		iJO1366				iAF1260			
		Core		Wild type		Core		Wild type	
Models	**μ cultivation**	**μ simulation**	**Relative difference (%)**	**μ simulation**	**Relative difference (%)**	**μ simulation**	**Relative difference (%)**	**μ simulation**	**Relative difference (%)**
MG1655	0.47	1.03	75	1.04	76	1.01	73	1.01	73
W3110	0.50	1.06	72	1.07	73	1.04	71	1.04	71
HMS174	0.40	0.93	77	0.94	78	0.91	76	0.92	76
RV308	0.61	1.25	69	1.26	69	1.23	67	1.24	68
BL21(DE3)	0.50	1.11	78	0*		1.09	76	0*	

* This model does not grow, since it does not account for membrane components that are included in the "extended" MG1655 wild type function

Since biomass composition is a crucial issue in metabolic modeling, it will be discussed in more detail in the next chapter.

Influence of different solvers on the computed results

As stated in the introduction, the linear optimization yields a flux cone (flux space), with many different feasible flux distributions. Most noticeable, the result is dependent on the solver applied to the problem. Although the optimized objective, e.g. the growth rate, is unique, there exist many different optimal flux solutions leading to optimal growth.

Solver performance is often measured by time efficiency, but they are also different for internal initial parameters and the way they itinerate through the matrix during optimization - and so do some results.

Shadow prices, a derivative of the objective function, indicate how much the addition of a metabolite increases or decreases the objective, such as maximal ATP yield. For example, if the shadow price of protons (h[c]) is -0.5, viz. two additional mol protons/mol glucose (assuming that the influx of glucose is 1 mmol/(gDW*h)) would reduce the ATP yield by 1 mol ATP/mol glucose. These shadow prices vary depending on the solver applied to the optimization problem (Figure 23)

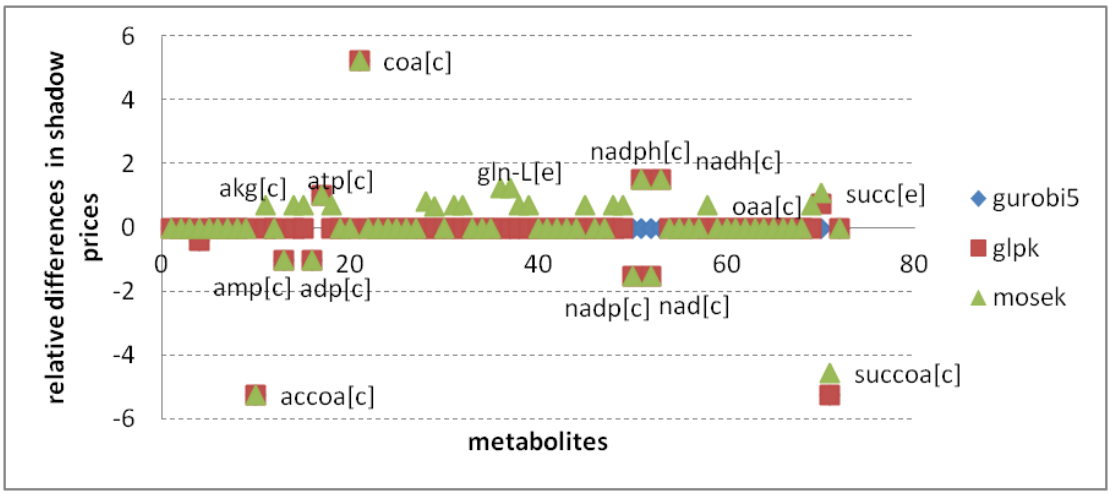


Figure 23 Relative differences in shadow prices for Gurobi, GLPK and Mosek solver relative to Tomlab solver.

The *Escherichia coli* core model[16] is used for simulation of optimal growth using optimizeCbModel.m similar to the tutorial supplied by Orth *et al.* [9]. Additionally to the solver used in the tutorial (tomlab_cplex), various other solvers are applied (mosek, glpk and gurobi5).

Whilst the optimized objective retrieved is the same, they shadow prices for other metabolites vary. The results of different solvers are compared in terms of difference in value compared to the tomlab solver [abs(solver X) – abs(tomlab)]. The absolute values are taken, as the signs of shadow price values between tomlab and the other solvers are always opposed.

[16] http://systemsbiology.ucsd.edu/Downloads/EcoliCore

Most noticeable, some bulk metabolites vary tremendous such as ATP or NAD and NADP, though this seems to be rather incidental. In any case, the results should be taken with great care. As already pointed out in the previous examples, a flux variability analysis should be performed, as only the objective is fixed to its optimum whereas all other calculated fluxes are allowed to vary.

Chapter IV Discussion and Outlook

In the previous chapter the results were presented, dealing with many interesting aspects of network reconstructions and metabolic simulations. The following passages will therefore discuss these issues and will give possible future directions.

Setting Standards

The first part of the previous chapter deals with formal model discrepancies due to different formats of toolboxes. So far, many standards have been set to facilitate exchange of models. Gene annotation is standardized with b-numbers[17], which have been assigned by the Blattner group during sequencing of the first *Escherichia coli* genome [39]. Contrarily, details in nomenclature for reactions and metabolite identifiers are not clearly defined which makes comparisons between reconstruction versions of different groups tedious. Recently, there are several initiatives undertaken to find a common namespace for metabolites and reactions, summarized by Bernard *et al.* [100]. Thereby reconciliation of metabolite lists is done in several steps using structural information such as the Simplified Molecular Input Line Entry System (SMILES) and IUPAC International Chemical Identifiers (InChI) [101]. Furthermore these reconciliation algorithms also make use of chemical nomenclature. Moreover, reactions are then refined by matching shared metabolites and through comparison of cross references from other sources, e.g. Enzyme Commission (EC.) numbers[18].

Another important issue is the informational content. The RAVEN toolbox does not contain information on the metabolite charge which impedes the calculation of the energy balance. However, for previous network reconstructions of MG1655 (Figure 3), a standard pH of 7.2 has been assumed for charged metabolites [27]. This in turn is a simplification of reality, since the pH may vary within cell compartments [102]. This issue is especially important when eukaryotic organisms shall be modeled, which exhibit a very complex compartmentalization. However it is a conventional decision to use either charged or uncharged metabolite lists. Here, it has been decided to use charged metabolite lists in the network reconstructions. As no further information regarding charge is provided from the "RAVEN-formatted" network reconstructions, the compulsory entries for metabolite charges are automatically set to zero when network reconstructions are analyzed using the COBRA Toolbox. This is essentially wrong and misleading. The metabolite list should therefore be switched to uncharged versions.

[17] B- number : "bxxxx", whereas "xxxx" is a four digit number.
[18] http://www.chem.qmul.ac.uk/iubmb/enzyme/

Constraining metabolic models by regulation

In Chapter 3, a gene essentiality simulation is presented, whereas an interesting candidate gene, *folD*, which is essential in the model of RV308 has been found. This essentiality though is not based on missing metabolic links but rather on the constraining of fluxes due to oxygen regulation. Still, the importance of these findings remains doubtful, since the regulation implied to simulate aerobic conditions is boolean and not graduated.

The regulation applied herein is based on Covert *et al.* [95] observations made from microarray analysis. They examined knock out strains of global transcription factors *(fnr, arcA* and *cydAB*) and established regulation rules based on oxygen deprivation which they found to be correct in 80 % of the cases. Furthermore, 18.3% of the phenotypes tested were only correct, if the regulatory effects were considered. Using mRNA profiling, the first proposed regulatory network was refined so that there were no false-positive predictions.

As mentioned previously, the lethality of *folD* knock out in RV308 is based on oxygen regulation. This is insofar interesting, as it highlights the functions of pathway not only for their supply for biomass components to sponsor growth, but also as integral essential part of a network. However, these pathways seem to be rather minimalistic with respect to the occurrence of isoenzymes. Hence it seems likely, that this is a false positive result. Still, for evaluation of this prediction, the strains should be subjected to knockout of *folD* and checked for growth behaviour.

Currently, no RV308 knock out strain for *folD* is reported in the literature. However other investigations of lethality in K-12 strains for *folD* under aerobic conditions are conflicting, according to Baba *et al.* [103], BW25113ΔfolD does not grow on LB medium. Gerdes *et al* [104] on the contrary observed, that MG1655ΔfolD is able to grow on LB medium[19].
Therefore this knock out requires further investigation. Moreover RNA microarray analysis need to be performed in order to clarify strain specific differences in regulation [105].

So far studies of the proteome of the strains used herein exist already, e.g. carried out by Marisch *et al.* [99]. However, major focus was put on pathogenicity factors such as motility and iron supply and should be expanded to other pathways too, e.g. cofactor biosynthesis.

[19] Remark: this information has been retrieved from EcoCyc (search enquiry : „folD"), as the web pages for supplementary material stated in the paper were no longer existent.

Comparison of experimental and simulated microbial growth

In order to collect input parameters for model simulation, growth curves of different *Escherichia coli* strains were conducted. Even through a weak trend towards better growth conditions when using baffled flasks and 2 L flasks, is observable, the oxygen measurements do not support this hypothesis. A possible reason for this might lie in the cultivation setting itself (Erlenmeyer shaking flasks in orbital motion), which does not allow for in situ oxygen measurements. Therefore, for future experiments, a more advanced cultivation setup should be chosen to avoid measurement artifacts/errors.
Besides oxygen supply, a key role is also played by secretion products. Secretion rates have been simulated but have not been experimentally verified here. The simulated range for acetate secretion for example would match reported values. Shiloach *et al*. [106] for example determined an experimental acetate secretion rate of 0.25 g acetate per 32 gDW per hour or Zhuang *et al.* [107] reported a secretion rate of 6 mmol/gDW per hour at a growth rate of 0.6 per hour. These values however, should be handled with care, since the conditions are not entirely identical e.g. different strain or cultivation setup.

Through comparison of growth phenotypes of *Escherichia coli*, it could be observed, that the simulated growth rates do not resemble the experimental ones. Baumler *et al.*[48] retrieved similar glucose uptake rates despite higher growth rates when cultivating W3110 and MG1655. Still, the discrepancy between experimental values and simulation according to Baumler *et al.* was roughly 40 % relative difference.
Therefore, the biomass composition is a major influencing factor in simulations.

The biomass objective function (BOF) is an artificial reaction accounting for biomass formation. Thus, it lists major cellular building blocks including nucleic acids, amino acids, saccharides and lipids [17]. The BOF furthermore contains vitamins and cofactors. Additionally, the BOF also accounts for growth associated energy cost, therefore ATP hydrolysis is also integrated within the reaction.

The principal biomass composition *of Escherichia coli* leads back to Neidhardt *et al* [108]. Accordingly, the dry weight is comprised of 55% protein, 20.5 % RNA, 3.1 % DNA and 9.1 % lipids. The biomass of *Escherichia coli* is furthermore composed of various lipopolysaccharides and peptidoglycans, as well as glycogen, polyamines, cofactors and ions.

Pramanik and Keasling [38] investigated the impact of the assumed biomass composition on the simulation and found, that the biomass composition is heavily influencing the simulation result. The predicted fluxes differed 16 % from the measured ones, in case an experimentally evaluated biomass composition was used. Noteworthy, the variation between simulation and experiment was observed to be up to 80 % in case a simplified BOF was applied.

Especially the lipid ratio is varying between strains and depending on the cultivation conditions since it is important for physiological maintenance [109] and therefore plays a crucial role for the quality of the BOF, even though lipids are not the biggest mass fraction within the cell. Unfortunately, the analysis of the lipid composition is a laborious task and requires expensive equipment such as GC-MS analysis.

Therefore an algorithm has been developed by Senger *et al* [110] to relax the need of expensive measurements. They implemented a genetic algorithm which adjusts the stoichiometry of the biomass objective function according to the observed exchange fluxes, such as the influx of media components and the secretion of fermentation byproducts including acetate, lactate and carbon dioxide. Another group of exchange reactions is defined as TUX, total unconstraint exchange fluxes, which shall be minimized, since it is assumed that the major exchange fluxes are observable. The biomass coefficients between lower and upper bounds are mixed and combined several times. Afterwards, a combination of bounds which reflects lowest TUX values is accepted as new BOF.

Metabolic Modeling –Quo vadis?

As mentioned in the introduction FBA is a good method to find results despite of lacking kinetic information and data scarcity, respectively. Still the methods themselves have some (partially hidden) drawbacks: In the previous chapter, the flux variability analysis was applied manifold: To investigate, whether the oxygen consumption measured lies in a feasible range and secondly to see which are the rates at which secretion products are built. Still one should keep in mind, that FVA minimizes and maximizes only one reaction at a time and therefore does not consider correlation between reaction fluxes, which makes feasible ranges broader than they are in reality [16].

An additional obstacle is the lack of information on differentially expressed genes which resemble the “weight” of a reaction in the network. Even though it is possible to predict growth behavior of one species, it is impossible to account for strain-specific differences in gene expression or enzyme promiscuity [47], that is playing a key role for strain– specific growth rates. However, this can be resolved by extending and enriching the models with proteome and microarray data, as remarked already.

A further unexpected finding concerns the connectivity of the networks. As a general ambition in metabolic modeling is not only to predict or analyze physiological behaviour, but also to collect and integrate all existing information on a certain organism. However, there are certain reactions that lack connectivity because they lie out of the scope of the model. For example, although the transcriptional and translational machinery is well described in *Escherichia coli*, the network reconstructions used herein do not include this information.

The aminoacyl-tRNA synthetase reactions in the reconstructions used herein are therefore not connected to the network. Furthermore certain cell wall constituents such as lipopolysaccharides or O-antigen, which are not included in the biomass function, are dead-end metabolites and therefore the reaction is shut off. This blemish of lacking connectivity is getting smaller with every extended reconstruction being published. For example, recently a whole-cell model of *Mycoplasma genitalium* has been published by Karr *et al.* [13, 111]. This model is integrating several modeling strategies, since there is no all-round modeling solution for every aspect of a complex living system. Even though it might be a long way to establish a whole cell model for *Escherichia coli* too, since the organism is more complex, the framework could be used as scaffold to guide future approaches of *Escherichia coli* models.

Conclusion

This has been a “primer” project to metabolic network reconstructions of different *Escherichia coli* strains. Major emphasis has therefore been put on refinement and correction of network reconstructions as well as to establish a modeling environment by testing existing toolboxes and writing functions and scripts for data management and customization of simulations. Thus, the cultivation setup was kept as simple as possible. The data gathered indicated strain specific growth differences. Still, these could not be mirrored by model simulations, because too little constraints have been set.

Now that the models are corrected, the next step should therefore be focused on a more elaborate laboratory setup, such as a detailed metabolome analysis for biomass elucidation as well as transcriptomics studies to elucidate regulatory differences between the strains. Furthermore, as these reconstructions are corrected, they are a good “wild type” basis for further simulations as there are steadily new algorithms available.

Metabolic models are a promising tool for aiding fundamental and applied research. However, even though much effort has been put into the implementation of toolboxes to facilitate and elaborate simulation, the models should still be handled conscientiously, since one single entry could falsify a simulation.

References

1. Kitano H: **Systems biology: a brief overview.** *Science* 2002, **295**:1662–4.

2. Andrianantoandro E, Basu S, Karig DK, Weiss R: **Synthetic biology: new engineering rules for an emerging discipline.** *Mol. Syst. Biol.* 2006, **2**:2006.0028.

3. Sagt CMJ: **Systems metabolic engineering in an industrial setting.** *Appl. Microbiol. Biotechnol.* 2013, **97**:2319–26.

4. Johnson KA: **A century of enzyme kinetic analysis, 1913 to 2013.** *FEBS Lett.* 2013, **587**:2753–66.

5. Smallbone K, Messiha HL, Carroll KM, Winder CL, Malys N, Dunn WB, Murabito E, Swainston N, Dada JO, Khan F, Pir P, Simeonidis E, Spasić I, Wishart J, Weichart D, Hayes NW, Jameson D, Broomhead DS, Oliver SG, Gaskell SJ, McCarthy JEG, Paton NW, Westerhoff H V, Kell DB, Mendes P: **A model of yeast glycolysis based on a consistent kinetic characterisation of all its enzymes.** *FEBS Lett.* 2013, **587**:2832–41.

6. Ma H, Zeng A: **Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms.** *Bioinformatics* 2003, **19**:270–277.

7. Llaneras F, Picó J: **Stoichiometric modelling of cell metabolism.** *J. Biosci. Bioeng.* 2008, **105**:1–11.

8. Zanghellini J, Ruckerbauer DE, Hanscho M, Jungreuthmayer C: **Elementary flux modes in a nutshell: properties, calculation and applications.** *Biotechnol. J.* 2013, **8**:1009–16.

9. Orth JD, Thiele I, Palsson BØ: **What is flux balance analysis?** *Nat. Biotechnol.* 2010, **28**:245–8.

10. Steuer R, Gross T, Selbig J, Blasius B: **Structural kinetic modeling of metabolic networks.** *Proc. Natl. Acad. Sci. U. S. A.* 2006, **103**:11868–73.

11. Steuer R: **Computational approaches to the topology, stability and dynamics of metabolic networks.** *Phytochemistry* 2007, **68**:2139–51.

12. Tomar N, De RK: **Comparing methods for metabolic network analysis and an application to metabolic engineering.** *Gene* 2013, **521**:1–14.

13. Karr J, Sanghvi J, Macklin D, Gutschow M, Jacobs J, Bolival BJ, Assad-Garcia N, Glass J, Covert M: **A Whole-Cell Computational Model Predicts Phenotype from Genotype**. *Cell* 2013, **150**:389–401.

14. Varma A, Palsson BØ: **Metabolic Flux Balancing: Basic Concepts, Scientific and Practical Use**. *Bio/Technology* 1994, **12**:994–998.

15. Schuetz R, Kuepfer L, Sauer U: **Systematic evaluation of objective functions for predicting intracellular fluxes in Escherichia coli.** *Mol. Syst. Biol.* 2007, **3**:119.

16. Mahadevan R, Schilling CH: **The effects of alternate optimal solutions in constraint-based genome-scale metabolic models**. *Metab. Eng.* 2003, **5**:264–276.

17. Feist AM, Palsson BØ: **The Biomass Objective Function**. *Curr Opin Microbiol.* 2010, **13**:344–349.

18. Lewis NE, Nagarajan H, Palsson BØ: **Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods.** *Nat. Rev. Microbiol.* 2012, **10**:291–305.

19. Birch EW, Udell M, Covert MW: **Incorporation of flexible objectives and time-linked simulation with flux balance analysis.** *J. Theor. Biol.* 2014, **345**:12–21.

20. McAnulty MJ, Yen JY, Freedman BG, Senger RS: **Genome-scale modeling using flux ratio constraints to enable metabolic engineering of clostridial metabolism in silico.** *BMC Syst. Biol.* 2012, **6**:42.

21. Benyamini T, Folger O, Ruppin E, Shlomi T: **Flux balance analysis accounting for metabolite dilution.** *Genome Biol.* 2010, **11**:R43.

22. Thiele I, Fleming RMT, Bordbar A, Schellenberger J, Palsson BØ: **Functional characterization of alternate optimal solutions of Escherichia coli's transcriptional and translational machinery.** *Biophys. J.* 2010, **98**:2072–81.

23. Mahadevan R, Edwards JS, Doyle FJ: **Dynamic flux balance analysis of diauxic growth in Escherichia coli.** *Biophys. J.* 2002, **83**:1331–40.

24. Covert MW, Schilling CH, Palsson BØ: **Regulation of gene expression in flux balance models of metabolism.** *J. Theor. Biol.* 2001, **213**:73–88.

25. Baart GJE, Martens DE: **Genome-Scale Metabolic Models: Reconstruction and Analysis**. *Methods Mol Biol.* 2012, **799**:107–126.

26. Feist AM, Herrgård MJ, Thiele I, Reed JL, Palsson BØ: **Reconstruction of biochemical networks in microorganisms**. *Nat Rev Microbiol.* 2011, **7**:129–143.

27. Thiele I, Palsson BØ: **A protocol for generating a high-quality genome-scale metabolic reconstruction.** *Nat. Protoc.* 2010, **5**:93–121.

28. Hamilton JJ, Reed JL: **Software platforms to facilitate reconstructing genome-scale metabolic networks.** *Environ. Microbiol.* 2014, **16**:49–59.

29. Fearnley L, Davis M, Ragan M, Nielsen L: **Extracting reaction networks from databases-opening Pandora's box.** *Brief. Bioinform.* 2013.

30. Liberal R, Pinney JW: **Simple topological properties predict functional misannotations in a metabolic network.** *Bioinformatics* 2013, **29**:154–61.

31. Schellenberger J, Thiele I, Orth JD: **Quantitative prediction of cellular metabolism with constraint-based models : the COBRA Toolbox v2.0**. *Nat. Protoc.* 2011, **6**:1290–1307.

32. Hoppe A, Hoffmann S, Gerasch A, Gille C, Holzhütter H: **FASIMU: flexible software for flux-balance computation series in large metabolic networks.** *BMC Bioinformatics* 2011, **12**:28.

33. Rocha I, Maia P, Evangelista P, Vilaça P, Soares S, Pinto JP, Nielsen J, Patil KR, Ferreira EC, Rocha M: **OptFlux: an open-source software platform for in silico metabolic engineering.** *BMC Syst. Biol.* 2010, **4**:45.

34. Pabinger S, Trajanoski Z: **Genome Scale Model Management and Comparison**. *Syst. Metab. Eng.* 2013, **985**:3–16.

35. Varma A, Boesch B, Palsson BØ: **Stoichiometric interpretation of Escherichia coli glucose catabolism under various oxygenation rates.** *Appl. Environ. Microbiol.* 1993, **59**:2465–73.

36. Varma A, Palsson BØ: **Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110.** *Appl. Environ. Microbiol.* 1994, **60**:3724–31.

37. Pramanik J, Keasling JD: **Effect of Escherichia coli biomass composition on central metabolic fluxes predicted by a stoichiometric model.** *Biotechnol. Bioeng.* 1998, **60**:230–8.

38. Pramanik J, Keasling JD: **Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements.** *Biotechnol. Bioeng.* 1997, **56**:398–421.

39. Blattner F, Plunkett G, Bloch C, Perna N, Burland V, Riley M, Collado-Vides J, Glasner J, Rode C, Mayhew G, Gregor J, Davis N, Kirkpatrick H, Goeden M, Rose D, Mau B, Shao Y: **The Complete Genome Sequence of Escherichia coli K-12**. *Science.* 1997, **277**:1453–1462.

40. Edwards JS, Palsson BØ: **The Escherichia coli MG1655 in silico metabolic genotype: its definition, characteristics, and capabilities.** *Proc. Natl. Acad. Sci. U. S. A.* 2000, **97**:5528–33.

41. Reed JL, Vo TD, Schilling CH, Palsson BØ: **An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR).** *Genome Biol.* 2003, **4**:R54.

42. Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, Karp PD, Broadbelt LJ, Hatzimanikatis V, Palsson BØ: **A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information.** *Mol. Syst. Biol.* 2007, **3**:121.

43. Feist AM, Palsson BØ: **The Growing Scope of Applications of Genome-scale Metabolic Reconstructions: the case of E. coli**. *Nat. Biotechnol.* 2011, **26**:659–667.

44. Orth JD, Conrad TM, Na J, Lerman J, Nam H, Feist AM, Palsson BØ: **A comprehensive genome-scale reconstruction of Escherichia coli metabolism--2011.** *Mol. Syst. Biol.* 2011, **7**:535.

45. Orth JD, Palsson BØ: **Gap-filling analysis of the iJO1366 Escherichia coli metabolic network reconstruction for discovery of metabolic functions.** *BMC Syst. Biol.* 2012, **6**:30.

46. Yamada T, Waller AS, Raes J, Zelezniak A, Perchat N, Perret A, Salanoubat M, Patil KR, Weissenbach J, Bork P: **Prediction and identification of sequences coding for orphan enzymes using genomic and metagenomic neighbours.** *Mol. Syst. Biol.* 2012, **8**:581.

47. McCloskey D, Palsson BØ, Feist AM: **Basic and applied uses of genome-scale metabolic network reconstructions of Escherichia coli.** *Mol. Syst. Biol.* 2013, **9**:661.

48. Baumler DJ, Peplinski RG, Reed JL, Glasner JD, Perna NT: **The evolution of metabolic networks of E. coli.** *BMC Syst. Biol.* 2011, **5**:182.

49. Xu C, Liu L, Zhang Z, Jin D, Qiu J, Chen M: **Genome-scale metabolic model in guiding metabolic engineering of microbial improvement.** *Appl. Microbiol. Biotechnol.* 2013, **97**:519–39.

50. Kharchenko P, Chen L, Freund Y, Vitkup D, Church GM: **Identifying metabolic enzymes with multiple types of association evidence.** *BMC Bioinformatics* 2006, **7**:177.

51. Monk JM, Charusanti P, Aziz RK, Lerman J, Premyodhin N, Orth JD, Feist AM, Palsson BØ: **Genome-scale metabolic reconstructions of multiple Escherichia coli strains highlight strain-specific adaptations to nutritional environments.** *Proc. Natl. Acad. Sci. U. S. A.* 2013, **110**:20338–43.

52. Thiele I, Fleming RMT, Que R, Bordbar A, Diep D, Palsson BØ: **Multiscale modeling of metabolism and macromolecular synthesis in E. coli and its application to the evolution of codon usage.** *PLoS One* 2012, **7**:e45635.

53. O'Brien EJ, Lerman J, Chang RL, Hyduke DR, Palsson BØ: **Genome-scale models of metabolism and gene expression extend and refine growth phenotype prediction.** *Mol. Syst. Biol.* 2013, **9**:693.

54. Escherich T: **Die Darmbakterien des Säuglings und ihre Beziehungen zur Physiologie der Verdauung**. *Habil., Medizinische Fak. München* 1886.

55. Bachmann BJ: **Pedigrees of some mutant strains of Escherichia coli K-12.** *Bacteriol. Rev.* 1972, **36**:525–57.

56. Archer CT, Kim JF, Jeong H, Park JH, Vickers CE, Lee SY, Nielsen LK: **The genome sequence of E. coli W (ATCC 9637): comparative genome analysis and an improved genome-scale reconstruction of E. coli.** *BMC Genomics* 2011, **12**:9.

57. Chaudhuri RR, Henderson IR: **The evolution of the Escherichia coli phylogeny.** *Infect. Genet. Evol.* 2012, **12**:214–26.

58. Clermont O, Bonacorsi S, Bingen E: **Rapid and simple determination of the Escherichia coli phylogenetic group.** *Appl. Environ. Microbiol.* 2000, **66**:4555–8.

59. Tenaillon O, Skurnik D, Picard B, Denamur E: **The population genetics of commensal Escherichia coli.** *Nat. Rev. Microbiol.* 2010, **8**:207–17.

60. Hayashi K, Morooka N, Yamamoto Y, Fujita K, Isono K, Choi S, Ohtsubo E, Baba T, Wanner BL, Mori H, Horiuchi T: **Highly accurate genome sequences of Escherichia coli K-12 strains MG1655 and W3110.** *Mol. Syst. Biol.* 2006, **2**:2006.0007.

61. Studier FW, Daegelen P, Lenski RE, Maslov S, Kim JF: **Understanding the differences between genome sequences of Escherichia coli B strains REL606 and BL21(DE3) and comparison of the E. coli B and K-12 genomes.** *J. Mol. Biol.* 2009, **394**:653–80.

62. Daegelen P, Studier FW, Lenski RE, Cure S, Kim JF: **Tracing ancestors and relatives of Escherichia coli B, and the derivation of B strains REL606 and BL21(DE3).** *J. Mol. Biol.* 2009, **394**:634–43.

63. Grossman AD, Erickson JW, Gross CA, Dove W, Jaacks KJ, Grossman AD, Erickson JW, Gross CA: **A collection of strains containing genetically for genetic mapping of Escherichia coli . A Collection of Strains Containing Genetically Linked Alternating Antibiotic Resistance Elements for Genetic Mapping of Escherichia coli**. *Microbiol Rev.* 1989, **53**:1–24.

64. Soupene E, Heeswijk WC Van, Plumbridge J, Stewart V, Bertenthal D, Lee H, Prasad G, Paliy O, Charernnoppakul P, Kustu S: **Physiological Studies of Escherichia coli Strain MG1655 : Growth Defects and Apparent Cross-Regulation of Gene Expression**. *J. Bacteriol.* 2003, **185**:5611–5626.

65. Zhou J, Wang C, Yoon S-H, Jang H-J, Choi E-S, Kim S-W: **Engineering Escherichia coli for selective geraniol production with minimized endogenous dehydrogenation.** *J. Biotechnol.* 2014, **169**:42–50.

66. Paigen K: **Role of the galactose pathway in the regulation of beta-galactosidase.** *J Bacteriol.* 1966, **92**:1394–403.

67. Akinterinwa O, Cirino PC: **Heterologous expression of D-xylulokinase from Pichia stipitis enables high levels of xylitol production by engineered Escherichia coli growing on xylose.** *Metab. Eng.* 2009, **11**:48–55.

68. Campbell JL, Richardson CC, Studier FW: **Genetic recombination and complementation between bacteriophage T7 and cloned fragments of T7 DNA.** *Proc. Natl. Acad. Sci. U. S. A.* 1978, **75**:2276–80.

69. Parra MC, Shaffer S a, Hajjar AM, Gallis BM, Hager A, Goodlett DR, Guina T, Miller S, Collins CM: **Identification, cloning, expression, and purification of Francisella lpp3: an immunogenic lipoprotein.** *Microbiol. Res.* 2010, **165**:531–45.

70. Kroll J, Steinle A, Reichelt R, Ewering C, Steinbüchel A: **Establishment of a novel anabolism-based addiction system with an artificially introduced mevalonate pathway: complete stabilization of plasmids as universal application in white biotechnology.** *Metab. Eng.* 2009, **11**:168–77.

71. Maurer R, Meyer B, Ptashne M: **Gene regulation at the right operator (OR) bacteriophage lambda. I. OR3 and autogenous negative control by repressor.** *J. Mol. Biol.* 1980, **139**:147–61.

72. Nellis DF, Ekstrom DL, Kirpotin DB, Zhu J, Andersson R, Broadt TL, Ouellette TF, Perkins SC, Roach JM, Drummond DC, Hong K, Marks JD, Park JW, Giardina SL: **Preclinical manufacture of an anti-HER2 scFv-PEG-DSPE, liposome-inserting conjugate. 1. Gram-scale production and purification.** *Biotechnol. Prog.* 2005, **21**:205–20.

73. Yoon SH, Han MJ, Jeong H, Lee CH, Xia XX, Lee DH, Shim JH, Lee SY, Oh TK KJ: **Comparative multi-omics systems analysis of Escherichia coli strains B and K-12.** *Genome Biol.* 2012, **13**:R37.

74. Agren R, Liu L, Shoaie S, Vongsangnak W, Nookaew I, Nielsen J: **The RAVEN toolbox and its use for generating a genome-scale metabolic model for Penicillium chrysogenum.** *PLoS Comput. Biol.* 2013, **9**:e1002980.

75. Keating SM, Bornstein BJ, Finney A, Hucka M: **SBMLToolbox: an SBML toolbox for MATLAB users.** *Bioinformatics* 2006, **22**:1275–7.

76. Bornstein BJ, Keating SM, Jouraku A, Hucka M: **LibSBML: an API library for SBML.** *Bioinformatics* 2008, **24**:880–1.

77. Kostromins A, Stalidzans E: **Paint4Net: COBRA Toolbox extension for visualization of stoichiometric models of metabolism.** *Biosystems.* 2012, **109**:233–9.

78. König M, Holzhütter H: **FluxViz - Cytoscape plug-in for visualization of flux distributions in networks**. *Genome informatics* 2010, **24**:96–103.

79. Gelius-Dietrich G, Amer Desouki A, Fritzemeier CJ, Lercher MJ: **Sybil--efficient constraint-based modelling in R.** *BMC Syst. Biol.* 2013, **7**:125.

80. Hucka M, Finney a., Sauro HM, Bolouri H, Doyle JC, Kitano H, Arkin a. P, Bornstein BJ, Bray D, Cornish-Bowden a., Cuellar a. a., Dronov S, Gilles ED, Ginkel M, Gor V, Goryanin II, Hedley WJ, Hodgman TC, Hofmeyr J-H, Hunter PJ, Juty NS, Kasberger JL, Kremling a., Kummer U, Le Novere N, Loew LM, Lucio D, Mendes P, Minch E, Mjolsness ED, Nakayama Y, Nelson MR, Nielsen PF, Sakurada T, Schaff JC, Shapiro BE, Shimizu TS, Spence HD, Stelling J, Takahashi K, Tomita M, Wagner J, Wang J: **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models**. *Bioinformatics* 2003, **19**:524–531.

81. Ruckerbauer D: **Reparation and Verification of the Yeast Consensus Model**. *Master´s Thesis* 2012:Karl Franzens University Graz.

82. Funahashi BA, Matsuoka Y, Jouraku A, Morohashi M, Kikuchi N, Kitano H: **A Versatile Modeling Tool for Biochemical Networks**. *Proc. IEEE* 2008, **96**:1254–1265.

83. Ebrahim A, Lerman J, Palsson BØ, Hyduke DR: **COBRApy: COnstraints-Based Reconstruction and Analysis for Python.** *BMC Syst. Biol.* 2013, **7**:74.

84. Burgard AP, Pharkya P, Maranas CD: **Optknock: a bilevel programming framework for identifying gene knockout strategies for microbial strain optimization.** *Biotechnol. Bioeng.* 2003, **84**:647–57.

85. Maniatis T, Fritsch E, Sambrook J: **M9 Minimal Medium**. *J. Mol. Cloning A Lab. Manual, 3 edn.,Cold Spring Harb. Lab. Press* 2001:A2.2.

86. Joyce AR, Reed JL, White A, Edwards R, Osterman A, Baba T, Mori H, Lesely SA, Palsson BØ, Agarwalla S: **Experimental and computational assessment of conditionally essential genes in Escherichia coli.** *J. Bacteriol.* 2006, **188**:8259–71.

87. Edwards JS, Ibarra RU, Palsson BØ: **In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data.** *Nat. Biotechnol.* 2001, **19**:125–30.

88. Miller G: **Use of dinitrosalicylic acid reagent for determination of reducing sugar**. *Anal. Chem.* 1959, **31**:426–428.

89. Moretti R, Thorson J: **A comparison of sugar indicators enables a universal high-throughput sugar-1-phosphate nucleotidyltransferase assay**. *Anal. Biochem.* 2008, **377**:251–258.

90. Campos E, Montella C, Garces F, Baldoma L, Aguilar J, Badia J: **Aerobic L-ascorbate metabolism and associated oxidative stress in Escherichia coli.** *Microbiology* 2007, **153**:3399–408.

91. Shlomi T, Berkman O, Ruppin E: **Regulatory on/off minimization of metabolic flux changes after genetic perturbations.** *Proc Natl Acad Sci U S A.* 2005, **102**:7695–700.

92. Segrè D, Vitkup D, Church G: **Analysis of optimality in natural and perturbed metabolic networks**. *Proc Natl Acad Sci U S A.* 2002, **99**:15112–7.

93. Harvey RJ DI: **Regulation in the folate pathway of Escherichia coli.** *Adv Enzym. Regul.* 1975, **13**:99 –124.

94. Textor S, Wendisch VF, De Graaf AA, Müller U, Linder MI, Linder D, Buckel W: **Propionate oxidation in Escherichia coli: evidence for operation of a methylcitrate cycle in bacteria.** *Arch. Microbiol.* 1997, **168**:428–36.

95. Covert M, Knight E, Reed J, Herrgard M, Palsson BØ: **Integrating high-throughput and computational data elucidates bacterial networks**. *Nature.* 2004, **429**:92–6.

96. Görgens JF, Van Zyl WH, Knoetze JH: **Reliability of methods for the determination of specific substrate consumption rates in batch culture**. *Biochem. Eng. J.* 2005, **25**:109–112.

97. Andersen KB, von Meyenburg K: **Are growth rates of Escherichia coli in batch cultures limited by respiration?** *J. Bacteriol.* 1980, **144**:114–23.

98. Paczia N, Nilgen A, Lehmann T, Gätgens J, Wiechert W, Noack S: **Extensive exometabolome analysis reveals extended overflow metabolism in various microorganisms.** *Microb. Cell Fact.* 2012, **11**:122.

99. Marisch K, Bayer K, Scharl T, Mairhofer J, Krempl PM, Hummel K, Razzazi-Fazeli E, Striedner G: **A comparative analysis of industrial Escherichia coli K-12 and B strains in high-glucose batch cultivations on process-, transcriptome- and proteome level.** *PLoS One* 2013, **8**:e70516.

100. Bernard T, Bridge A, Morgat A, Moretti S, Xenarios I, Pagni M: **Reconciliation of metabolites and biochemical reactions for metabolic networks.** *Brief. Bioinform.* 2014, **15**:123–35.

101. O'Boyle NM: **Towards a Universal SMILES representation - A standard method to generate canonical SMILES based on the InChI.** *J. Cheminform.* 2012, **4**:22.

102. Wilks JC, Slonczewski JL: **pH of the cytoplasm and periplasm of Escherichia coli: rapid measurement by green fluorescent protein fluorimetry.** *J. Bacteriol.* 2007, **189**:5601–7.

103. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko K, Tomita M, Wanner BL, Mori H: **Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.** *Mol. Syst. Biol.* 2006, **2**:2006.0008.

104. Gerdes SY, Scholle MD, Campbell JW, Bala G, Ravasz E, Daugherty MD, Somera AL, Kyrpides NC, Anderson I, Gelfand MS, Bhattacharya A, Kapatral V, Souza MD, Baev M V, Grechkin Y, Mseeh F, Fonstein MY, Overbeek R, Baraba A, Oltvai ZN: **Experimental Determination and System Level Analysis of Essential Genes in Escherichia coli MG1655**. *J Bacteriol.* 2003, **185**:5673–5684.

105. Nagy PL, Marolewski A, Benkovic SJ, Zalkin H, Nagy PL, Marolewski A, Benkovic SJ: **Formyltetrahydrofolate hydrolase, a regulatory enzyme that functions to balance pools of**

tetrahydrofolate and one-carbon tetrahydrofolate adducts in Escherichia coli. *J Bacteriol.* 1995, **177**:1292–8.

106. Shiloach J, Kaufman J, Guillard a S, Fass R: **Effect of glucose supply strategy on acetate accumulation, growth, and recombinant protein production by Escherichia coli BL21 (lambdaDE3) and Escherichia coli JM109.** *Biotechnol. Bioeng.* 1996, **49**:421–8.

107. Zhuang K, Vemuri GN, Mahadevan R: **Economics of membrane occupancy and respiro-fermentation.** *Mol. Syst. Biol.* 2011, **7**:500.

108. Neidhardt FC: **Chemical Composition of Escherichia coli**. In *Escherichia coli Salmonella typhimurium.Cellular Mol. Biol.* Vol. 1. edited by Neidhardt FC, Ingraham, J.L, Magasanik, B., Brooks Low, K., Schaechter, M, Umbarger E Waschington,D.C.: American Society for Microbiology; 1987, **1**:3–6.

109. De Siervo AJ: **Alterations in the phospholipid composition of Escherichia coli B during growth at different temperatures.** *J. Bacteriol.* 1969, **100**:1342–9.

110. Senger R, Nazem-Bokaee H: **Resolving cell composition through simple measurements, genome-scale modeling, and a genetic algorithm.** *Methods Mol Biol.* 2013, **985**:85–101.

111. Karr JR, Sanghvi JC, Macklin DN, Arora A, Covert MW: **WholeCellKB: model organism databases for comprehensive whole-cell models.** *Nucleic Acids Res.* 2013, **41**:D787–92.

Appendix

Escherichia coli cultivations

Below, results from *Escherichia coli* cultivations, which were presented in Chapter 3, are tabulated. Batch raw data can be found in the "digital" appendix.

Table 19 Overview on MG1655 batch cultivation (in 1 L baffled Erlenmeyer flask)

Main culture settings: 200 mL M9 supplemented with 2 g/L glucose in 1 L nonbaffled Erlenmeyer flask at 37 °C and 110 rpm orbital motion

9

						Average Method		Logarithmic Method	
Date of cultivation	flask index	μ [1/h]	R^2 (μ)	Yx/s	R^2(Yx/s)	qs [mmol/(gDW*h)]	R^2 (qs)	qs [mmol/(gDW*h)]	R^2 (qs)
24/07/2013	I	0.47	0.9957	0.28	0.9922	9.3	0.9922	10.5	0.9896
	II	0.46	0.9947	0.26	0.9958	9.8	0.9958	10.6	0.9842
25/07/2013	I	0.46	0.9931	0.25	0.9895	9.9	0.9902	10.4	0.9965
	II	0.46	0.9959	0.29	0.9912	8.8	0.9912	10.6	0.9962
26/07/2013	I	0.5	0.9967	0.34	0.9802	8.1	0.9802	10.3	0.997
	II	0.49	0.9978	0.27	0.9683	9.9	0.9683	10.3	0.9906
20/09/2013	I	0.45	0.9962	0.26	0.9948	9.8	0.9948	10.8	0.9893
	II	0.46	0.9966	0.25	0.9884	10.4	0.9884	10.6	0.9876
	III	0.47	0.9927	0.24	0.9938	10.9	0.9938	11.5	0.9823
	IV	0.46	0.9976	0.24	0.9943	10.6	0.9943	10.5	0.9932
Average		0.47		0.27		9.8		10.6	
SD		0.01		0.03		0.8		0.3	
CV (%)		3.1		11.1		8.6		3.3	

Table 20 Overview on MG1655 batch cultivation (1 L baffled Erlenmeyer flask)

Main culture settings: 200 mL M9 supplemented with 2 g/L glucose in 1 L baffled Erlenmeyer flask at 37 °C and 110 rpm orbital motion

						Average Method		Logarithmic Method	
Date of cultivation	**flask index**	**μ [1/h]**	**R^2 (μ)**	**Yx/s**	**R^2(Yx/s)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**
26/09/2013	I	0.56	0.9989	0.33	0.9951	9.2	0.9951	11.4	0.9951
	II	0.56	0.9984	0.32	0.9968	9.6	0.9968	11.5	0.9911
	III	0.56	0.9982	0.32	0.9968	9.7	0.9968	11.1	0.9894
11/10/2013	I	0.52	0.9993	0.32	0.9912	8.8	0.9912	11.2	0.9910
	II	0.51	0.9990	0.32	0.9968	7.5	0.9947	10.8	0.9948
	III	0.54	0.9966	0.36	0.9943	8.4	0.9943	12.0	0.9934
	IV	0.55	0.9987	0.34	0.9914	9.0	0.9914	11.8	0.9950
Average		0.54		0.33		8.9		11.4	
SD		0.02		0.01		0.8		0.4	
CV (%)		3.4		4.1		8.5		3.4	

Table 21 Overview on MG1655 batch cultivation (2 L nonbaffled Erlenmeyer flask)

Main culture settings: 200 mL M9 supplemented with 2 g/L glucose in 2 L nonbaffled Erlenmeyer flask at 37 °C and 110 rpm orbital motion

						Average Method		**Logarithmic Method**	
Date of cultivation	**flask index**	**μ [1/h]**	**R^2 (μ)**	**Yx/s**	**R^2(Yx/s)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**
24/07/2013	I	0.49	0.9963	0.25	0.9972	9.8	0.9972	10.8	0.9918
	II	0.51	0.9960	0.26	0.9904	10.6	0.9904	11.3	0.9904
	III	0.50	0.9939	0.24	0.9960	11.6	0.9960	11.5	0.9980
	IV	0.49	0.9941	0.24	0.9921	8.8	0.9921	10.4	0.9969
Average		0.49		0.25		10.2		11.0	
SD		0.01		0.01		1.2		0.5	
CV (%)		1.8		4.4		11.4		4.5	

Table 22 Overview on MG1655 batch cultivation (2 L baffled Erlenmeyer flask)

Main culture settings: 200 mL M9 supplemented with 2 g/L glucose in 2 L baffled Erlenmeyer flask at 37 °C and 110 rpm orbital motion

						Average Method		Logarithmic Method	
Date of cultivation	**flask index**	**μ [1/h]**	**R^2 (μ)**	**Yx/s**	**R^2(Yx/s)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**
05/10/2013	I	0.55	0.9990	0.31	0.9916	9.6	0.9916	11.2	0.9909
	II	0.56	0.9993	0.32	0.9905	9.7	0.9905	11.1	0.9935
	III	0.56	0.9978	0.33	0.9944	9.5	0.9944	10.8	0.9898
08/11/2013	I	0.60	0.9967	0.42	0.9946	7.7	0.9946	11.9	0.9924
	II	0.62	0.9984	0.41	0.9955	8.2	0.9955	11.6	0.9924
	III	0.61	0.9955	0.49	0.9952	6.8	0.9952	11.8	0.9816
	IV	0.60	0.9925	0.48	0.9868	6.8	0.9868	11.8	0.9847
17/10/2013	I	0.55	0.9958	0.36	0.9800	10.3	0.9800	12.0	0.9960
	II	0.58	0.9961	0.40	0.9865	8.0	0.9865	12.0	0.9818
	III	0.57	0.9979	0.42	0.9890	7.5	0.9890	12.2	0.9917
	IV	0.57	0.9978	0.35	0.9953	8.9	0.9953	12.4	0.9833
Average		0.58		0.39		8.4		11.6	
SD		0.03		0.06		1.3		0.5	
CV (%)		4.4		16.1		15.1		3.9	

Table 23 Overview on W3110 batch cultivation

Main culture settings: 200 mL M9 supplemented with 2 g/L glucose in 1 L nonbaffled Erlenmeyer flask at 37 °C and 110 rpm orbital motion

						Average Method		Logarithmic Method	
Date of cultivation	**flask index**	**μ [1/h]**	**R^2 (μ)**	**Yx/ s**	**R^2(Yx/s)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**
17/09/2013	I	0.48	0.9960	0.30	0.9930	8.8	0.9930	10.6	0.9949
	II	0.48	0.9972	0.30	0.9970	9.0	0.9970	10.4	0.9905
	III	0.48	0.9979	0.32	0.9902	8.2	0.9902	10.3	0.9931
	IV	0.48	0.9975	0.29	0.9951	9.2	0.9951	10.4	0.9912
31/07/2013	I	0.50	0.9925	0.31	0.9858	8.9	0.9858	11.3	0.9936
01/08/2013	I	0.51	0.9970	0.31	0.9884	9.1	0.9884	11.1	0.9952
	II	0.51	0.9964	0.31	0.9849	9.1	0.9849	11.2	0.9907
	I	0.53	0.9904	0.36	0.9805	8.1	0.9805	11.4	0.9964
Average		0.50		0.31		8.8		10.8	
SD		0.02		0.02		0.4		0.5	
CV (%)		3.8		7.0		4.8		4.2	

Table 24 Overview on RV308 batch cultivation

Main culture settings: 200 mL M9 supplemented with 2 g/L glucose in 1 L nonbaffled Erlenmeyer flask at 37 °C and 110 rpm orbital motion

		6				Average Method		Logarithmic Method	
Date of cultivation	**flask index**	**μ [1/h]**	**R^2 (μ)**	**Yx/s**	**R^2(Yx/s)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**
13/09/2013	I	0.62	0.9935	0.30	0.9899	11.3	0.9899	13.2	0.9922
	II	0.62	0.9919	0.29	0.9961	11.8	0.9961	13.5	0.9938
	III	0.61	0.9950	0.26	0.9804	12.7	0.9804	12.7	0.9880
	IV	0.59	0.9947	0.29	0.9805	11.1	0.9805	12.4	0.9894
19/09/2013	I	0.62	0.9937	0.28	0.9806	11.7	0.9805	12.9	0.9988
	II	0.61	0.9918	0.29	0.9867	11.1	0.9853	13.2	0.9968
Average		0.61		0.29		11.6		13.0	
SD		0.01		0.01		0.6		0.4	
CV (%)		2.0		5.0		5.2		3.3	

Table 25 Overview on HMS174 batch cultivation

Main culture settings: 200 mL M9 supplemented with 2 g/L glucose in 1 L nonbaffled Erlenmeyer flask at 37 °C and 110 rpm orbital motion

						Average Method		Logarithmic Method	
Date of cultivation	**flask index**	**μ [1/h]**	**R^2 (μ)**	**Yx/s**	**R^2(Yx/s)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**
13/08/2013	I	0.42	0.9957	0.31	0.9834	7.5	0.9834	9.7	0.9858
14/08/2013	I	0.40	0.9909	0.29	0.9781	7.6	0.9781	9.7	0.9806
	II	0.40	0.9937	0.29	0.9866	7.4	0.9866	9.4	0.9892
16/08/2013	I	0.42	0.9953	0.31	0.9813	7.5	0.9813	9.7	0.9861
06/11/2013	I	0.41	0.9974	0.33	0.9932	7.0	0.9932	9.6	0.9960
	II	0.40	0.997	0.32	0.9893	7.0	0.9893	9.6	0.9959
	III	0.39	0.9953	0.35	0.9795	6.1	0.9795	9.5	0.9935
	IV	0.40	0.995	0.34	0.9865	6.4	0.9865	9.6	0.9934
07/11/2013	I	0.39	0.9925	0.28	0.9902	7.6	0.9902	10.1	0.9866
	II	0.39	0.9927	0.31	0.9919	6.9	0.9919	9.8	0.9933
	III	0.40	0.9933	0.38	0.9825	5.7	0.9825	9.8	0.9978
Average		0.40		0.32		7.0		9.7	
SD		0.01		0.03		0.6		0.2	
CV (%)		2.9		9.2		9.3		1.9	

Table 26 Overview on BL21(DE3) batch cultivation

Main culture settings: 200 mL M9 supplemented with 2 g/L glucose in 1 L nonbaffled Erlenmeyer flask at 37 °C and 110 rpm orbital motion

						Average Method		Logarithmic Method	
Date of cultivation	**flask index**	**μ [1/h]**	**R^2 (μ)**	**Yx/s**	**R^2(Yx/s)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**	**qs [mmol/(gDW*h)]**	**R^2 (qs)**
20/08/2013	I	0.49	0.9908	0.28	0.9721	9.5	0.9721	10.8	0.9939
	II	0.49	0.9908	0.25	0.9835	10.0	0.9835	11.1	0.9789
21/08/2013	I	0.49	0.9916	0.22	0.9798	11.9	0.9798	11.3	0.9917
	II	0.51	0.9914	0.23	0.9801	11.9	0.9801	11.6	0.9881
22/08/2013	I	0.46	0.9933	0.20	0.9822	13.0	0.9838	12.1	0.9781
	II	0.49	0.9910	0.20	0.9817	14.4	0.9817	12.0	0.9778
05/11/2013	I	0.48	0.9949	0.24	0.9743	10.9	0.9743	11.0	0.9915
	II	0.53	0.9918	0.26	0.9688	11.0	0.9688	11.5	0.9960
	III	0.52	0.9962	0.26	0.9734	10.7	0.9734	11.4	0.9927
	IV	0.51	0.9965	0.27	0.9728	10.4	0.9859	11.1	0.9859
Average		0.50		0.24		11.4		11.4	
SD		0.02		0.03		1.5		0.4	
CV (%)		4.3		10.9		13.0		3.7	

Table 27 Average glucose uptake rates q_s. of MG1655 at different settings

All batches where performed at 37 °C and 100 rpm motion, with 2 g/L glucose initial concentration. The relative difference is calculated using the default setting (1 L none-baffle "nb") as reference. This tabulation is the basis for Figure 15, logarithmic method.

setting	**batches**	**average q_s [mmol/(gDW*h)**	**SD**	**CV (%)**	**relative difference (%)**
1 L nb	10	10.6	0.3	3.3	
1 L b	7	11.4	0.4	3.4	7
2 L nb	4	11	0.5	4.5	3
2 L b	11	11.6	0.5	3.9	9

Table 28 Overview of oxygen measurements using a needle-type microsensor

MG1655 cultures have been prepared at different settings (M9/ glucose supplement 110 rpm 37 °C) and the oxygen consumption has been measured.

setting	gDW/L*	qO_2 [µmol/(L*s)]	R^2	qO_2 [mmol/(L*h)]	qO_2 [µmol/(gDW*h)]
1 L nb	0.54	5.00E+07	0.9962	13.9	25.7
1 L b	0.48	4.00E+07	0.9997	11.1	24.7
2 L b	0.54	5.00E+07	0.9995	13.9	25.7

* the cell dry weight concentration has been calculated from OD_{600} measurements and conversion with a factor of 0.3 (OD_{600}/gDW)

Table 29 Overview of *Escherichia coli* batch cultivations

Main culture: 200 mL culture (M9 2 g/L glucose) in 1 L unbaffled Erlenmeyer flask at 110 rpm orbital motion at 37 °

		q_s [mmol/(gDW*h)]			µ [1/h]			$Y_{x/s}$		
Strain	**batches**	**Average**	**SD**	**CV (%)**	**Average**	**SD**	**CV (%)**	**Average**	**SD**	**CV (%)**
MG1655	10	10.6	0.3	3.3	0.00	0.01	3.1	0.27	0.03	11.1
W3110	8	10.8	0.5	4.2	0.50	0.01	2.0	0.31	0.02	7.0
HMS174	11	9.7	0.2	1.9	0.40	0.01	2.9	0.32	0.03	9.2
RV308	6	13.0	0.4	3.3	0.61	0.01	2.0	0.29	0.01	5.0
BL21(DE3)	10	11.4	0.4	3.7	0.50	0.02	4.3	0.24	0.03	10.9

Oxygen Measurements

Cultivation Settings		
date of cultivation		**strain**
17.12.2013		**MG1655**
main culture parameters		
substrate	**glucose**	**M9 minimal medium**
flask	**1L non-baffle**	**110 rpm**

Sensor Settings	
Firestubg Logger	**V2.37**
Sensor (Channel)	**1**
Calibration	**2-Point**
Temperature [°C]	**36**
Pressure (hPa)	**986**
Humidity (%)	**44**
Oxygen Unit	**µmol/L**
Sensor Code	**ZA7-546-218**
Sensor Type	**Z : Normal Range O2 - Micro/Minisensor - Orange Fiber**

OD600	1.8	
gDW/L	0.54	
qO2 µmol/(L*s)	5.00E+07	R^2 = 0,9962
qO2 µmol/(L*h)	1.39E+04	
qO2 mmol/(gDW*h)	25.7	

bold values are taken for calculation

time [h:m:s]	time [decimal]	O_2 [µmol/L]
21:25:55	**21.43**	**5.45E+05**
21:25:56	**21,43**	**5.26E+05**
21:25:58	**21.43**	**5.04E+05**
21:25:59	**21.43**	**4.89E+05**
21:26:00	**21.43**	**4.71E+05**
21:26:01	**21.43**	**4.53E+05**
21:26:02	**21.43**	**4.31E+05**
21:26:04	**21.43**	**4.17E+05**
21:26:05	**21.43**	**4.11E+05**
21:26:06	**21.44**	**3.99E+05**
21:26:07	**21.44**	**3.80E+05**
21:26:08	**21.44**	**3.55E+05**
21:26:10	**21.44**	**3.37E+05**
21:26:11	**21.44**	**3.26E+05**
21:26:12	**21,44**	**3.19E+05**
21:26:13	**21,44**	**3,05E+05**
21:26:14	**21.44**	**2.87E+05**
21:26:16	**21.44**	**2.65E+05**
21:26:17	**21.44**	**2.52E+05**
21:26:18	**21.44**	**2.35E+05**
21:26:19	**21.44**	**2.11E+05**
21:26:20	**21.44**	**1.87E+05**
21:26:22	**21.44**	**1.76E+05**
21:26:23	**21.44**	**1.71E+05**
21:26:24	21.44	1.71E+05
21:26:25	21.44	1.74E+05
21:26:26	21.44	1.81E+05
21:26:28	21.44	2.43E+05
21:26:29	21.44	2.82E+05
21:26:30	21.44	3.51E+05
21:26:31	21.44	3.52E+05
21:26:32	21,44	4.01E+05
21:26:34	21.44	3.83E+05

Cultivation Settings		
date of cultivation		strain
17.12.2013		MG1655
main culture parameters		
substrate	glucose	M9 minimal medium
flask	1L baffle	110 rpm
Sensor Settings		
Firestubg Logger		V2.37
Sensor (Channel)		1
Calibration		2-Point
Temperature [°C]		36
Pressure (hPa)		986
Humidity (%)		44
Oxygen Unit		µmol/L
Sensor Code		ZA7-546-218
Sensor Type		Z : Normal Range O2 - Micro/Minisensor – Orange Fiber

OD600	1.5	
gDW/L	0.48	
qO2 µmol/(L*s)	4.00E+07	R^2 = 0.9997
qO2 µmol/(L*h)	1.11E+04	
qO2 mmol/(gDW*h)	24.7	

bold values are taken for calculation

time [h:m:s]	time [decimal]	O_2 [µmol/L]	(continued)		
			time [h:m:s]	time [decimal]	O_2 [µmol/L]
20:56:50	20.95	1.60E+06			
20:56:52	20.95	1.60E+06	20:57:29	20.96	1.21E+06
20:56:53	20.95	1.58E+06	20:57:30	20.96	1.20E+06
20:56:54	20.95	1.57E+06	20:57:31	20.96	1.20E+06
20:56:55	20.95	1.56E+06	20:57:32	20.96	1.19E+06
20:56:56	20.95	1.55E+06	20:57:34	20.96	1.17E+06
20:56:57	20.95	1.53E+06	20:57:35	20.96	1.16E+06
20:56:59	20.95	1.52E+06	20:57:36	20.96	1.15E+06
20:57:00	20.95	1.50E+06	20:57:37	20.96	1.13E+06
20:57:01	20.95	1.49E+06	20:57:38	20.96	1.11E+06
20:57:02	20.95	1.48E+06	20:57:39	20.96	1.10E+06
20:57:04	20.95	1.47E+06	20:57:41	20.96	1.09E+06
20:57:05	20.95	1.45E+06	20:57:42	20.96	1.08E+06
20:57:06	20.95	1.44E+06	20:57:43	20.96	1.08E+06
20:57:07	20.95	1.43E+06	20:57:44	20.96	1.07E+06
20:57:08	20.95	1.42E+06	20:57:45	20.96	1.06E+06
20:57:10	20.95	1.41E+06	20:57:47	20.96	1.05E+06
20:57:11	20.95	1.40E+06	20:57:48	20.96	1.04E+06
20:57:12	20.95	1.39E+06	20:57:49	20.96	1.02E+06
20:57:13	20.95	1.38E+06	20:57:50	20.96	1.01E+06
20:57:14	20.95	1.37E+06	20:57:51	20.96	9.92E+05
20:57:16	20.95	1.35E+06	20:57:53	20.96	9.83E+05
20:57:17	20.95	1.33E+06	20:57:54	20.97	9.72E+05
20:57:18	20.96	1.32E+06	20:57:55	20.97	9.58E+05
20:57:19	20.96	1.31E+06	20:57:56	20.97	9.47E+05
20:57:20	20.96	1.30E+06	20:57:57	20.97	9.37E+05
20:57:22	20.96	1.29E+06	20:57:59	20.97	9.24E+05
20:57:23	20.96	1.28E+06	20:58:00	20.97	9.10E+05
20:57:24	20.96	1.27E+06	20:58:01	20.97	8.94E+05
20:57:25	20.96	1.25E+06	20:58:02	20.97	8.85E+05
20:57:26	20.96	1.24E+06	20:58:03	20.97	8.74E+05
20:57:28	20.96	1.22E+06	20:58:05	20.97	8.67E+05

Cultivation Settings

date of cultivation		**strain**
17.12.2013		**MG1655**
main culture parameters		
substrate	**glucose**	**M9 minimal medium**
flask	**2L baffle**	**110 rpm**

Sensor Settings

Firestubg Logger	**V2.37**
Sensor (Channel)	**1**
Calibration	**2-Point**
Temperature [°C]	**36**
Pressure (hPa)	**986**
Humidity (%)	**44**
Oxygen Unit	**µmol/L**
Sensor Code	**ZA7-546-218**
Sensor Type	**Z : Normal Range O2 - Micro/Minisensor - Orange Fiber**

OD600	1.8	
gDW/L	0.54	
qO2 µmol/(L*s)	5.00E+07	R^2 = 0.9995
qO2 mmol/(L*h)	1.39E+01	
qO2 mmol/(gDW*h)	25.7	

bold values are taken for calculation

time [h:m:s]	time [decimal]	O_2 [µmol/L]	(continued)		
20:31:57	**20.53**	**1.06E+06**	**time [h:m:s]**	**time [decimal]**	**O_2 [µmol/L]**
20:31:59	**20.53**	**1.04E+06**	**20:32:41**	**20.54**	**5.13E+05**
20:32:00	**20.53**	**1.02E+06**	**20:32:42**	**20,55**	**4,97E+05**
20:32:01	**20.53**	**1.01E+06**	**20:32:43**	**20.55**	**4,79E+05**
20:32:02	**20.53**	**9.97E+05**	**20:32:44**	**20.55**	**4,63E+05**
20:32:03	**20.53**	**9.85E+05**	**20:32:46**	**20,55**	**4,47E+05**
20:32:05	**20.53**	**9.73E+05**	**20:32:47**	**20.55**	**4,30E+05**
20:32:06	**20.54**	**9.58E+05**	**20:32:48**	**20.55**	**4,15E+05**
20:32:07	**20.54**	**9.43E+05**	**20:32:49**	**20.55**	**3,99E+05**
20:32:08	**20.54**	**9.25E+05**	**20:32:50**	**20.55**	**3,83E+05**
20:32:09	**20.54**	**9.07E+05**	**20:32:52**	**20.55**	**3.67E+05**
20:32:11	**20.54**	**8.89E+05**	**20:32:53**	**20.55**	**3.52E+05**
20:32:12	**20.54**	**8.74E+05**	**20:32:54**	**20.55**	**3.38E+05**
20:32:13	**20.54**	**8.63E+05**	**20:32:55**	**20.55**	**3.20E+05**
20:32:14	**20.54**	**8.48E+05**	**20:32:57**	**20.55**	**3.03E+05**
20:32:15	**20.54**	**8.35E+05**	**20:32:58**	**20.55**	**2.87E+05**
20:32:17	**20.54**	**8.20E+05**	**20:32:59**	**20.55**	**2.71E+05**
20:32:18	**20.54**	**8.06E+05**	**20:33:00**	**20.55**	**2.57E+05**
20:32:19	**20.54**	**7.89E+05**	**20:33:01**	**20.55**	**2.40E+05**
20:32:20	**20.54**	**7.74E+05**	**20:33:03**	**20.55**	**2.23E+05**
20:32:21	**20.54**	**7.59E+05**	**20:33:04**	**20.55**	**2.06E+05**
20:32:23	**20.54**	**7.45E+05**	**20:33:05**	**20.55**	**1.90E+05**
20:32:24	**20.54**	**7.30E+05**	**20:33:06**	**20.55**	**1.74E+05**
20:32:25	**20.54**	**7.14E+05**	**20:33:07**	**20.55**	**1.56E+05**
20:32:26	**20.54**	**6.96E+05**	**20:33:09**	**20.55**	**1.38E+05**
20:32:28	**20.54**	**6.79E+05**	**20:33:10**	**20.55**	**1.22E+05**
20:32:29	**20.54**	**6.64E+05**	**20:33:11**	**20.55**	**1.05E+05**
20:32:30	**20.54**	**6.48E+05**	**20:33:12**	**20.55**	**8.88E+04**
20:32:31	**20.54**	**6.33E+05**	**20:33:13**	**20.55**	**7.16E+04**
20:32:32	**20.54**	**6.21E+05**	**20:33:15**	**20.55**	**5.49E+04**
20:32:34	**20.54**	**6.08E+05**	**20:33:16**	**20.55**	**3.82E+04**
20:32:35	**20.54**	**5.93E+05**	**20:33:17**	**20.55**	**2.41E+04**
20:32:36	**20.54**	**5.75E+05**	**20:33:18**	**20.56**	**1.39E+04**
20:32:37	**20.54**	**5.59E+05**	**20:33:19**	**20.56**	**8.95E+03**
20:32:38	**20.54**	**5.43E+05**	20:33:21	20.56	7.32E+03
20:32:40	**20.54**	**5.30E+05**	20:33:22	20.56	7.03E+03

Source Code

Basically, all functions and scripts written are customizations of COBRA Toolbox functions or used for data management This source code suits “RAVEN-formatted” files (Excel or SBML). However, on a case to case basis they are also executable on COBRA-“format” (or others).

Source code for model simulation

RegulateModelv3.m	The function regulateModelv3.m applies basic (boolean) regulation on constraints depending on the substrate (glucose/non-glucose), or aerobic and anaerobic cultivation.
GrowthFBA.m	The function growthFBA.m calculates the biomass growth rate [hr^-1] under specified condition using FBA. A table of optimization results is printed % to the screen. Additionally, a bar diagram is returned. The table might also be written to an Excel file if desired.
GrowthFBAv1s.m	The script growthFBAv1s.m calculates the steady-state growth rate of the five network reconstructions in use on a broad range of substrates; the optimization solutions are written to an output file.This script is based on the similar function growthFBA.m
KOv5.m	The script KOv5.m is used to simulate gene deletions for a set of models on different substrates. The models are compared for growth behaviour and “unique lethal genes”, viz. genes, that are only lethal in one model, but not in the others. A summary on retrieved genes is written. At the moment the script performs single gene deletion using and the growth rates of the mutants is calculated using FBA, however it can be easily customized such as for double gene deletion or on a combination of substrates too.
Xls2Rmodel.m	The function xls2Rmodel.m converts Excel files of network reconstructions to SBML files, which are analyzable using sybil package in R. A simple example script in R (KOexample.r) is also provided in the source code section.

Source code for data management

Inter.m	The function inter.m is called during execution of KOv5.m. It is used to find intersecting essential genes among different models in order to find the unique ones.
RetrieveV3.m	The function retrieveV3.m retrieves the growth rations (wild type/mutant) for unique essential genes of all models. This function is called during execution of KOv5.m.
CommonMatrixV2.m	The function commonMatrixV2.m is called during KOv5.m. Only essential genes are extracted. These genes are subjected to inter.m to find a unique lethal gene.
FlogicMaskV3.m	The function flogicMaskV4.m is called by supdatemodel4.m. It creates a logic mask (0 or 1) of all the entries to compare differences within common entries of two network reconstructions.
FdevdecV3 .m	The function fdevdecV4.m is called by supdatemodel4.m to find deviations within common entries according to the logic mask (as returned from flogicMaskV4.m).
ConsistentModel.m	Consistent models are composed only of active reactions and genes, respectively. Since network reconstructions used herein serve also as knowledge base for "all" known (metabolic) reactions and genes, not all of them are indeed connected to the network due to (knowledge) gaps in the system. This function is based on the COBRA function reduceModel.m. Consistent models are generated and lists of inactive and active genes are retrieved.

Source code for comparison and extension of network reconstructions

Supdatemodel4.m	The script supdatemodel4.m can be applied to two different network reconstructions to find differences between them. A directory is created, that stores source files as well as reports on the differences, such as deviations among common entries as well as unique entries to the reconstructions. Specially, the "results.xlsx"-file can be used as input for fupdatemodelsheets2.m, to "update" network reconstructions based on the differences encountered.

Fupdatemodelsheets2.m	This function updates an existing ("RAVEN-formatted") network reconstruction by incorporating changes according to an "update" Excel file, such as addition, deletion or change of entries. The resulting Excel file contains an updated network reconstruction which can be directly transformed into a model for simulation.
AddSBO.m	The function addSBO.m adds the Systems Biology Ontology (SBO) terms to a model structure. SBO terms contribute a new semantic layer to the model and aims to facilitate the categorization of metabolic processes. There are many different ways how to categorize the occurring reactions in models e.g. by function, by reaction mechanism. This function, whatsoever adds SBO terms according to the existing subsystem, e.g. "alternate carbon metabolism", entry for this reaction in the model. Upon completion of this function a list of SBO terms is returned that can be copied into the Excel file. Extensive documentation on SBO terms can be found online[20].

Source code for demonstrative purposes

COBRAtut.m	The script COBRAtut.m is based on the tutorial on FBA supplied with Orth *et a*l. [9]. Results of different solvers applied on the same problem are compared.
clarifyKO.m	The script clarifyKO.m is used to stepwise clarify the reason for unique lethal gene discovered using KOv5.m. Growth on different substrates is simulated to find crucial reactions in the network.
ThesisExample.m	The script ThesisExample.m contains the source code for various simulations presented in Chapter 3 such as comparison of strain specific growth behaviour.

[20] http://www.ebi.ac.uk/sbo/main

Description of model features

Feature	Description
rxns	Cell vector of rxnIDs {'R_HEX1','R_GK1',…}
mets	Cell vector of metabolite IDs {M_glc_c, M_h2o_c,…}
S	Sparse matrix of reaction and metabolites
rev	Information on reaction reversibility
lb ub	Lower bound and upper bound of allowed fluxes
c	Weight coefficient vector for objective function
genes	Lists gene IDs e.g "b0002"
metCharge	metabolite charge in physiological environment (not in RAVEN format)
rules or gprRules	Information on gene protein relations
rxnGeneMat	Matrix containing reactions and genes according to gpr.
subSystems	Returns pathway classification e.g. "Amino Acid Metabolism"
confidenceScores	Extent of reliability of the data; e.g. from database, evaluated in the laboratory.
rxnReferences, rxnNotes	literature references and notes
rxnECNumbers	EC number of enzymes involved
rxnNames metNames	Long name of enzyme(rxn) or metabolite e.g. R_HEX1_c hexokinase
metFormulas	Charged metabolite composition metFormula + metCharge = netComposition as retrieved from KEGG DB
metChEBIID,metKEGGID, PubChemID	Links metabolite information from various databases (ChEBI, KEGG, PubChem)
metInChIString	IUPAC International Chemical Identifier (InChI)
b	Right Hand site vector needed for solving the linear programming problem since A*x =b
description	Model identifier
CompNames	Compartment names such as "c" for cytosolic
CompOutside	Relationship of compartments e.g. cytosol lies within periplasm
metComps	Metabolite compartments e.g. M_glc_c for cytosolic glucose
Comp-met-gene Miriam	Minimum Information Required In the Annotation of Models)

Supplementary Files (CD)

A CD is appended to this thesis containing experimental raw data (batches) as well as source code and files concerning network reconstructions.

Reconstructions

a. V0
    i. iJO1366-novel-air-GLC_raven.xls
    ii. iW1407_raven.xls
    iii. iRV1394_raven.xls
    iv. iHMS1391_raven.xls
b. V1
    i. iJO1366_raven.V1.xls
    ii. iW1407_raven.V1.xls
    iii. iRV1394_raven.V1.xls
    iv. iHMS1391_raven.V1.xls
    v. HMSupdateV0_01.xlsx
    vi. Wupdate_V0_01.xlsx
    vii. JOupdate_V0_01.xlsx
    viii. RVpdate_V0_01.xlsx
c. V1b
    i. clarification.txt
    ii. iJO1366_raven.V1b.xls
    iii. iW1407_raven.V1b.xls
    iv. iRV1394_raven.V1b.xls
    v. iHMS1391_raven.V1b.xls
    vi. HMSupdateV1_1b.xlsx
    vii. Wupdate_V1_1b.xlsx
    viii. JOupdate_V1_1b.xlsx
    ix. RVpdate_V1_1b.xlsx

d. V2
    i. iJO1366_raven.V2.xls
    ii. iW1407_raven.V2.xls
    iii. iRV1394_raven.V2.xls
    iv. iHMS1391_raven.V2.xls
    v. HMSupdateV1b_2.xlsx
    vi. Wupdate_V1b_2.xlsx
    vii. JOupdate_V1b_2.xlsx
    viii. RVpdate_V1b_2.xlsx
e. V3
    i. iJO1366_raven.V3.xls
    ii. iW1407_raven.V3.xls
    iii. iRV1394_raven.V3.xls
    iv. iHMS1391_raven.V3.xls
    v. HMSupdateV2_3.xlsx
    vi. Wupdate_V2_3.xlsx
    vii. JOupdate_V2_3.xlsx
    viii. RVpdate_V2_3.xlsx
f. V4
    i. COMMENTS_V4.txt
    ii. iJO1366_raven.V4.xls
    iii. iW1407_raven.V4.xls
    iv. iRV1394_raven.V4.xls
    v. iHMS1391_raven.V4.xls
    vi. HMSupdateV3_4.xlsx
    vii. Wupdate_V3_4.xlsx
    viii. JOupdate_V3_4.xlsx
    ix. RVpdate_V3_4.xlsx
g. BL21(DE3)
    i. BupdateJO_BL.xlsx
    ii. iBL_raven.V4.xls
    iii. REL606-BL21(DE3)_indels.xlsx

Growth curves

a. F_MG1655_1L.xls
b. F_MG1655_1Lbaffle.xls
c. F_MG1655_2L.xls
d. F_MG1655_2Lbaffle.xls
e. F_W3110.xls
f. F_HMS174.xls
g. F_RV308.xls
h. F_BL21(DE3).xls
i. 20120627_DNS.Calibration.xls
j. 20120926_DNS.Calibration.xls
k. 20121030_DNS.Calibration.xls
l. 20130908_DNS.Calibration.xls

Source Code

a. AddSBO.m
b. clarifyKO.m
c. COBRAtut.m
d. CommonMatrixV2.m
e. ConsistentModel.m
f. FdevdecV3 .m
g. FlogicMaskV3.m
h. Fupdatemodelsheets2.m
i. GrowthFBAv1s
j. Inter.m
k. KOexample.r
l. KOv5.m
m. RegulateModelv3.m
n. RertieveV3.m
o. Supdatemodel4.m
p. ThesisExample.m
q. Xls2Rmodel.m
r. growthFBAv1.m

Further files

further files containing raw data for tables and figures presented in this thesis

- O2 measurements
    - 1Lbaffle.txt
    - 2Lbaffle.txt
    - 1Ldefault.txt
    - results.xlsx
- ConsistentModel.xlsx
- CultivationResults.xlsx
- EcoreTutSolvers.xlsx