Organization and Integration of Large-scale Datasets for Designing a Metabolic Model and Re-annotating the Genome of *Mycoplasma pneumoniae*

An Application of the Systems Biology Approach to a Minimal Bacterium

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

Dr. rer. nat
im Fach Biophysik

eingereicht an der
Mathematisch-Naturwissenschaftlichen Fakultät I
Humboldt-Universität zu Berlin

von

M. Sc. Judith Andrea Heidrun Wodke

Präsident der Humboldt-Universität zu Berlin:
Prof. Dr. Jan-Hendrik Olbertz

Dekan der Mathematisch-Naturwissenschaftlichen Fakultät I:
Prof. Dr. Stefan Hecht

Gutachter:
1. Prof. Dr. Dr. h.c. Edda Klipp
2. Prof. Dr. Luis Serrano
3. Prof. Dr. Hermann-Georg Holzhütter

eingereicht am: 09.08.2012
Tag der mündlichen Prüfung: 29.10.2012

Ich widme diese Arbeit dem Leben selbst,
in all seinen vielfältigen Facetten

Abstract

Mycoplasma pneumoniae, one of the smallest known self-replicating organisms, is a promising model organism in systems biology when aiming to assess understanding of an entire living cell. One of the key steps towards this goal is the design of mathematical models that describe the cellular processes by connecting the involved components and that allow to unravel underlying mechanisms. For *Mycoplasma pneumoniae*, a wealth of genome-wide datasets on genomics, transcriptomics, proteomics, and the metabolism has been produced in a collaborative research project during the past years. However, a proper system facilitating easy exchange of the generated information and mathematical models to integrate the different datasets and to fill remaining knowledge gaps were lacking. In addition, the analysis of the metabolome was somewhat lagging behind and different *in vivo* observations of metabolic behavior remained unexplained. This thesis presents a combinatorial approach to design a metabolic model for *Mycoplasma pneumoniae*. First, we developed a database, **MyMpn**, in order to provide access to structured and organized data, a prerequisite for successful modeling of biological systems. Second, we built a predictive, genome-scale, constraint-based metabolic model and, in parallel, we explored the metabolome *in vivo* reaching unprecedented coverage of the *in silico* predicted metabolome. We defined the biomass composition of a *Mycoplasma pneumoniae* cell, corrected the metabolic wiring diagram, showed that *Mycoplasma pneumoniae* dedicates a large proportion of its energy to cellular homeostasis, and analyzed the metabolic behavior under different growth conditions and in deleterious mutants. Finally, as suggested from the correction of the functional annotation of metabolic key enzymes and the *in silico* knock-out predictions, we manually re-annotated the genome of *Mycoplasma pneumoniae*. The database, despite not yet being released to the public, is internally already used for data analysis, integration and visualization, as well as for the design of different mathematical models. The definition of the biomass composition of *Mycoplasma pneumoniae* provides the basis for *in silico* growth simulations of wall-less bacteria. Unraveling the principles governing energy metabolism and adaptive capabilities upon gene deletion facilitates the development of engineering tools and dynamic models for metabolic sub-systems. Our results highlight the impact of the reductive genome evolution on the metabolism and especially the cellular energy balancing in a minimal bacterium which reflects the high degree of adaptation to a relatively unchanging niche, the human lung. Furthermore, we revealed that the degree of complexity in which the genome of *Mycoplasma pneumoniae* is organized far exceeds what has been considered possible so far. We showed that different genes can be transcribed from the same genomic region and identified numerous small RNAs. The 32 newly identified genes together with the corrections of previously annotated genes allowed us to correct in total a 10% error rate in the annotation.

Keywords: Constraint-Based Modeling, Database Design, Genome Re-annotation, Metabolism, Mycoplasma pneumoniae

Zusammenfassung

Mycoplasma pneumoniae, einer der kleinsten lebenden Organismen, ist ein erfolgversprechender Modellorganismus der Systembiologie, um eine komplette lebende Zelle zu verstehen. Ein wichtiger Schritt dahingehend ist die Konstruktion mathematischer Modelle, die zelluläre Prozesse beschreiben, indem sie die beteiligten Komponenten vernetzen. Diese ermöglichen es, zugrundeliegende Mechanismen zu entschlüsseln. Für *Mycoplasma pneumoniae* wurden in einem kooperativen Projekt während der letzten Jahre diverse genomweite Datensätze für Genomics, Transcriptomics, Proteomics und Metabolomics produziert. Allerdings fehlten sowohl ein System zum effizienten Informationsaustausch als auch mathematische Modelle, um die vorhandenen Daten zu integrieren und verbleibende Wissenslücken zu füllen. Außerdem waren das Metabolome noch nicht detailliert *in vivo* untersucht worden und verschiedene Beobachtungen im metabolischen Verhalten ungeklärt. Diese Dissertation präsentiert einen kombinatorischen Ansatz zur Entwicklung eines metabolischen Modells für *Mycoplasma pneumoniae*. Zuerst haben wir eine Datenbank, **MyMpn**, entwickelt, um Zugang zu strukturierten, organisierten Daten zu schaffen - eine Grundvoraussetzung für erfolgreiche Modellierung biologischer Systeme. Als nächstes haben wir ein genomweites, Constraint-basiertes metabolisches Modell mit Vorhersagekapazitäten konstruiert und parallel dazu das Metabolome in beispielloser Abdeckung des *in silico* Vorhergesagten experimentell charakterisiert. Wir haben die Zusammensetzung einer *Mycoplasma pneumoniae* Zelle definiert, das metabolische Netzschema korrigiert, gezeigt, dass ein Grossteil der produzierten Energie auf zelluläre Homeostase verwendet wird, und das Metabolismusverhalten unter verschiedenen Wachstumsbedingungen und in Gen-Knockout-Mutanten analysiert. Schließlich, suggeriert durch die Korrektur der funktionalen Annotation metabolischer Schlüsselenzyme und die *in silico* Knockoutvorhersage, haben wir manuell das Genom von *Mycoplasma pneumoniae* reannotiert. Die Datenbank, obwohl noch nicht für die Öffentlichkeit zugänglich, wird bereits intern für Analyse, Integration und Visualisierung von experimentellen Daten und in der Modellierung genutzt. Die Definition der Zusammensetzung der Biomasse *Mycoplasma pneumoniaes* legt den Grundstein für Wachstumssimulationen zellwandloser Bakterien. Die Entdeckung der Prinzipien, die den Energiemetabolismus und die Anpassungsfähigkeiten bei Genausfall kontrollieren, erleichtert die Entwicklung von Manipulationstechniken und dynamischen Modellen metabolischer Teilsysteme. Unsere Ergebnisse heben den Einfluss der reduktiven Genomevolution auf den Metabolismus und speziell die Energiebilanzierung eines Minimalbakteriums hervor, der die hochgradige Anpassung an einen gleich bleibenden Lebensraum, die menschliche Lunge, widerspiegelt. Überdies haben wir aufgedeckt, dass die Genomorganisation in *Mycoplasma pneumoniae* komplexer ist als bisher für möglich gehalten. Wir haben gezeigt, dass mehrere unterschiedliche Gene aus der gleichen Genomregion transkribiert werden, und eine Vielzahl small RNAs identifiziert. Die Entdeckung von 32 bisher nicht annotierten Genen und die Korrektur von diversen bereits annotierten Genen resultieren in der Korrektur einer 10%igen Fehlerrate.

Schlagwörter: Constraint-basierte Modellierung, Datenbankentwicklung, Genomreannotation, Metabolismus, Mycoplasma pneumoniae

Contents

List of Figures

List of Tables

1. Introduction

1.1. Outline

1.1.1. Objectives and Introduction into Research Projects

Systems biology is a rapidly emerging research field aiming to understand biological systems by bringing together knowledge from the different natural sciences [Auyang, 1999, Kitano, 2001]. It tackles the general aim of biology, science of life and living organisms, by applying the holism approach, i.e. by putting together information on different system components rather than taking them apart (reductionism approach) [Auyang, 1999, Oshry, 2007]. Thus, probably the greatest challenge within systems biology is the understanding and computational simulation of a complete organism. Despite the technological advances on the field of experimental methods and computational analysis tools, understanding of a multicellular organism, even a small one, is still not amenable. In contrast, the analysis of bacteria, prokaryotic organisms, on a genome-scale has become possible, providing insight into the basic principles of life common to all cells, regardless of the organism they compose or belong to. In addition to the putatively easier analysis when compared to more complex cells such as yeast, new treatment strategies can be developed based on the study of bacterial cells, pathogenic for humans, leading to advances in health care [Gallagher et al., 2007].

In 1962 Morowitz and Tourtellotte published an article on mycoplasmas, cell wall-less minimal bacteria, as the smallest living organisms [Morowitz and Tourtellotte, 1962]. Since then the research interest in mycoplasmas as minimal model organisms has been increasing constantly and gained special attention when the first whole-genome sequencing techniques were developed [Fraser et al., 1995, Himmelreich et al., 1996] and with the discovery of their ability to invade host cells [Andreev et al., 1995]. *Mycoplasma pneumoniae* (*M. pneumoniae*), an obligate human parasite preferentially colonizing the pulmonary epithelium and associated with a variety of diseases [Chiner et al., 2003, Waites and Talkington, 2004], has a genome size of 816,394 base pairs coding for only 689 proteins [Himmelreich et al., 1996, Dandekar et al., 2000]. Contrary to other mycoplasmas, it can be grown relatively easy under laboratory conditions without addition of host cells, thus facilitating its experimental exploration.

In a joint attempt to understand an organism in its entirety, a multidisciplinary research project applying the systems biology approach to the minimal model organism *M. pneumoniae* has been initiated. Several genome-wide datasets on genomics, transcriptomics, proteomics, and the metabolism have been produced during the past years [Yus et al., 2012, Güell et al., 2009, 2011, Kühner et al., 2009, Maier et al., 2011, van Noort et al., 2012, Yus et al., 2009]. The diversity of those large-scale datasets offers

a great opportunity to obtain systems level understanding of this minimal bacterium and to approach computational simulation of a whole cell. To this end, the different cell components, such as DNA, RNAs, proteins, and metabolites, have to be identified, if possible quantified, and placed in their dynamic network context [Kitano, 2001]. A wide variety of different mathematical modeling approaches, for example Boolean networks, constraint-based approaches, and dynamic approaches using partial or ordinary differential equations (PDEs or ODEs), can be employed for this aim [Klipp et al., 2005]. However, information exchange between the research groups involved in the 'mycoplasma project' was not optimized and mathematical models for the different cellular sub-systems, connecting the wealth of available data, were lacking.

One of the most important cellular sub-systems is the metabolism, which can be defined as the particular sets of biochemical reactions that, taking place in the cells, sustain life of every living organism [Alberts et al., 2008]. The constant performance of biochemical reactions allows the cell to create and maintain order in a universe that continuously evolves towards greater disorder. For survival and proliferation the cell needs to produce energy and to newly synthesize and maintain the different cell components. Therefore, the biochemical reactions form mostly linear series, the metabolic pathways, in which the product of one reaction is the substrate of the following reaction. Two general types of pathways can be distinguished: catabolic and anabolic pathways. Catabolic pathways break down nutrients into smaller molecules, thus providing cell building blocks and energy, while anabolic pathways use the produced energy for the synthesis of cell components. The different pathways are linked to each other, thus forming the metabolic network of a cell [Alberts et al., 2008].

The reduced genome of *M. pneumoniae* is accompanied by a lean metabolic network [Himmelreich et al., 1996, Dandekar et al., 2000, Yus et al., 2009]. The lack of nearly all anabolic pathways known from more complex organisms highly facilitates the direct connection of extracellular nutrient depletion to cellular processes [Yus et al., 2009]. In addition, most building blocks for the cell components, such as nucleobases, amino acids and fatty acids, have to be taken up from the environment. For energy production *M. pneumoniae* relies on glycolysis and organic acid fermentation due to the absence of a citric acid cycle and a functional respiratory chain [Himmelreich et al., 1996, Dandekar et al., 2000, Yus et al., 2009]. As a result, this reduced metabolic network limits the chemical diversity of metabolites and the small cell size of *M. pneumoniae* limits the abundances of all molecules.

Advances in analytical techniques monitoring cellular macromolecules, for example mRNAs, proteins or metabolites, allow the determination and quantification of such molecules on a large scale, establishing the "-omics" approach for the analysis of biological systems. The high-throughput analysis of proteins (proteomics) and mRNAs (transcriptomics) facilitated new insights into cellular processes, such as transcription, translation, protein regulation, and metabolism. The study of cellular metabolites on a global scale (metabolomics), despite having shown to provide valuable information on the phenotypic state of biological systems [Cornish-Bowden and Cárdenas, 2000, Nicholson and Lindon, 2008, Buescher et al., 2012], is lagging behind due to several factors. In contrast to proteins and mRNAs which are produced from a limited set of build-

ing blocks, metabolites are chemically diverse and of very different abundances. Thus, it is technically impossible to address all of them using a single analytical approach [Goodacre et al., 2004, Liberman et al., 2012]. In addition, metabolites are often unstable or short-lived due to rapid processing in the cellular context, thus posing great challenges for sample preparation and processing [Scalbert et al., 2009, van Gulik, 2010]. Following the general trend in "-omics" analyses of other organisms, the metabolome of *M. pneumoniae*, contrary to the genome, the transcriptome, and the proteome, has neither been characterized nor quantified yet.

The decision which modeling approach to use depends on the size of the analyzed system, i.e. the balance between model complexity (manageability) and model granularity (detail), and the specific question(s) the model is aimed to answer [Klipp et al., 2005]. Constraint-based modeling is a static modeling approach that is used for the genome-scale reconstruction of biological networks [Fell and Small, 1986, Savinell and Palsson, 1992a,b, Oberhardt et al., 2009, Feist et al., 2009]. Information about the network structure, the connectivity of the network and the changes in network fluxes under different conditions can be extracted. Flux balance analysis (FBA) is a mathematical method that determines a set of metabolic fluxes fulfilling the steady state condition for a given set of available nutrients [Kauffman et al., 2003, Varma and Palsson, 1994b, Reed and Palsson, 2003]. Constraint-based modeling has been applied to different organisms and cells, among them *Escherichia coli*, *Pseudomonas aeruginosa* and *Homo sapiens* [Edwards and Palsson, 2000, Oberhardt et al., 2008, Duarte et al., 2007, Gille et al., 2010, Rolfsson et al., 2011], and was used for example to predict mutant phenotypes [Edwards and Palsson, 2000], pathway activity [Covert et al., 2001], or metabolic flux distributions [Oberhardt et al., 2008].

To tackle the comprehensive examination of the metabolism of *M. pneumoniae* we designed a multi-layer approach integrating different experimental and computational analyses (Figure 1.1). First, to guarantee fast and easy information exchange within the mycoplasma research community and to provide a basis for the successful design of a genome-scale metabolic model, we developed a database with interactive web interface for *M. pneumoniae*. In addition to the data access, this web interface also supplies several data analysis and visualization tools, some of which can be run locally or be applied with unpublished results by using a temporary memory. Furthermore, during the data collection and organization process, we obtained a systematic overview about the available knowledge as well as gaps in our understanding of *M. pneumoniae* metabolism, leading to the questions a metabolic model could answer.

As the main project of this thesis, we designed a predictive, genome-scale, constraint-based metabolic model for *M. pneumoniae*, *i*JW145 ("*i* + initials of the model builder + number of genes"). This model was built to validate the reconstructed network, to analyze the central carbon metabolism responsible for energy homeostasis, and to explore the metabolic behavior under different conditions. The identification and preferentially quantification of system components is one of the pre-requisites for the correct representation of biological networks by mathematical models. To complement the available data and to address the complicated metabolomics analysis, we applied a combinatorial approach joining different analysis techniques, namely nuclear magnetic

resonance spectroscopy (NMR), gas chromatography coupled to mass spectrometry (GC-MS), and liquid chromatography coupled to mass spectrometry (LC-MS), to characterize the metabolome of *M. pneumoniae in vivo*. Iteratively integrating *in silico* growth sim-

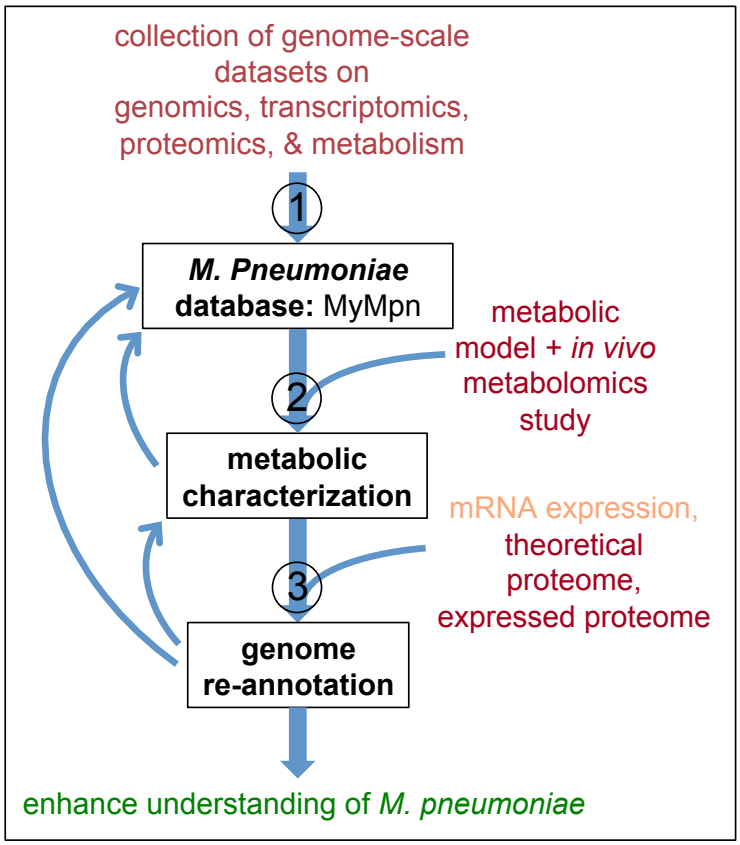

Figure 1.1.: Workflow for the thesis: Aiming to enhance our understanding of *M. pneumoniae*, specifically of the metabolism, we conducted several sub-projects: 1. We designed and implemented a database, **MyMpn**, thereby structuring the available data. 2. We comprehensively characterized the metabolism of *M. pneumoniae* by integrating the development and application of a constraint-based metabolic model with an *in vivo* metabolomics study. 3. We characterized the genome of *M. pneumoniae* by manually re-annotating the genome and subsequently defining its essential regions.

ulations with experimental validations and literature information, enabled us to correct the metabolic network structure and to unravel principles governing metabolic flux distributions in *M. pneumoniae* under a variety of conditions. Amongst other results, this comprehensive analysis pointed out several errors in the genome annotation of *M. pneumoniae*.

The fundamental question which essential functions underlie sustained cellular growth and proliferation, has been addressed by several organism-wide gene essentiality studies in bacteria [Hutchison et al., 1999, Gerdes et al., 2003, Kobayashi et al., 2003, Glass et al.,

2006, French et al., 2008, Dybvig et al., 2010, Christen et al., 2011]. However, several problems prevented the final answering of this question up to now. One of the major challanges arises from the fact that automatic genome annotations based on sequence similarities have been shown to be highly error prone [Casari et al., 1995, Brenner, 1999]. To assure a genome annotation as accurate as possible, we manually curated the annotation based on the integration of the theoretical coding capabilities of the *M. pneumoniae* genome with experimental data on transcriptomics [Güell et al., 2009, 2011] and proteomics.

All results presented have been iteratively integrated amongst each other and are incorporated into the **MyMpn** database to contribute to the attempt to understand the minimal bacterium *M. pneumoniae*, one of the most promising model organisms in systems biology (Figure 1.1).

I present in this thesis the development of the *M. pneumoniae* database, **MyMpn**, and the metabolic model, *i*JW145, as well as the related projects of the experimental exploration of the *M. pneumoniae* metabolome and of the genome re-annotation, all of which contributing to the attempt to understand *M. pneumoniae*. Based on the applied systems biology approach, all of these projects have been carried out together with other researchers, combining knowledge from different research fields. Therefore, at the beginning of each chapter, my contributions to the presented results are detailed.

1.1.2. Scope of the Thesis

With the technological advances for genome-scale analysis of different cellular components and the emergence of the systems biology field, for the first time it became amenable to understand organisms as an entity, not only as a collection of largely independent sub-systems that can be analyzed in separation [Auyang, 1999, Oshry, 2007]. While the established biological model organisms, yeast and *E. coli*, have been studied for decades, a new group of such model organisms emerged with the discovery of the genome-reduced mycoplasmas, the smallest self-replicating organisms known today [Morowitz and Tourtellotte, 1962]. In a combined effort several research groups from the Center of Genomic Research (CRG), Barcelona, and the European Molecular Biology Laboratory (EMBL), Heidelberg, initiated a scientific project aiming to gain full-detailed understanding of a living cell, *M. pneumoniae*. In the first phase of this project, several high-throughput data sets have been produced to analyze the major cellular building blocks by a combination of complementary experimental techniques [Güell et al., 2009, Kühner et al., 2009, Yus et al., 2009, Maier et al., 2011, Güell et al., 2011, van Noort et al., 2012, Yus et al., 2012]. However, to understand a biological system, the goal of all systems biologists, theoretical formalisms, such as mathematical models that allow to place the system components into their dynamic network context, are indispensable [Kitano, 2001, 2002b].

For *M. pneumoniae*, not only a possibility to easily exchange information, the basis of all successful multidisciplinary research projects, but also mathematical models that combine the diverse large-scale datasets intending to explain the underlying structural and regulatory mechanisms were lacking. A database to store the produced experimental

and computational data as well as to further enhance the information exchange amongst the different groups involved in the mycoplasma project, was needed. Taking advantage of the expertise in database development in the Bioinformatics Core Facility at the CRG, we designed and implemented a database for *M. pneumoniae* that facilitates data storage and exchange: **MyMpn**. In addition, via the web interface we provide a working platform to analyze, integrate and visualize the available data. One effect of the associated data organisation was the supply of structured data, necessary for the design of mathematical models.

With the manual reconstruction of the metabolic network based on experimental data [Yus et al., 2009] a thorough basis for metabolic research in *M. pneumonoiae* had been generated. However, the resulting wiring diagram (the metabolic map) had not been evaluated and several observations during growth curve experiments, as for example the slow dubling times of *M. pneumoniae* grown in batch culture, remained unexplained. For the design of genome-scalemetabolic models, static modeling approaches, such as constraint-based modeling, are generally applied. Building such a genome-scale constraint-based model (*i*JW145), we completed the wiring diagram by adding necessary transport reactions and correcting mis-annotated reactions.

Furthermore, the metabolome had not yet been characterized *in vivo*, while for transcriptome and proteome quantitative data for various conditions was available. Several analysis techniques, such as NMR, GC-MS, or LC-MS have been applied to study metabolites in different organisms [Soga et al., 2003, van der Werf et al., 2007, 2008, t'Kindt et al., 2010, Liebeke et al., 2011]. However, the coverage of the *in silico* predicted metabolome reached in those studies amounted to only 9 - 25%. Combining the established analysis techniques, we obtained complementary results raising the obtained coverage to over 50%. The iterative combination of the experimental metabolome exploration and the development of the metabolic model enabled us to semi-quantitatively define the biomass composition of an average *M. pneumoniae* and thus to simulate growth for this organism *in silico*. Subsequently conducting growth simulations with *i*JW145, we unraveled general principles underlying energy balancing and metabolic adaptation to different environmental conditions or in knock-out mutants.

Taking into account the available data on transcriptomics and proteomics as well as the small size of *M. pneumoniae*, a manual re-annotation of the genome appeared to be feasible. Such an experimental validation would allow to complement for the high error rate of an automatically obtained genome annotation, reported since the first sequenced genome [Casari et al., 1995, Brenner, 1999]. We applied a combinatorial approach considering the theoretical coding capabilities of the genome, analyzing available transcriptomics data, and conducting a proteomics study which focused on the size determination of the produced proteins in order to manually refine the genome annotation of *M. pneumoniae*. Our results highlight the importance of an experimental validation of genome annotations that have been obtained automatically based on sequence alignments. Most importantly, we showed that bacterial genomes are highly structured and organized in a complexity far exceeding what has been assumed possible for prokaryotes in general and especially for minimal bacteria.

Taken together, the presented findings significantly increase understanding of *M. pneu-*

moniae and prove that even minimal cells are finetuned on all cellular levels, thereby reflecting evolution, environment, and the specific cellular properties.

1.1.3. Thesis Organization

In systems biology expertise from different research fields as well as experimental and theoretical data is combined to analyze complex biological processes. Thus, knowledge about the biology of the examined system, about the experimental exploration possibilities, about the modeling methodology, and about the underlying mathematics is indispensable to successfully understand a biological system. The first chapter of this thesis, apart from the thesis outline, contains an introduction into the biological, the methodological (computational and experimental) and the mathematical background for the presented work (sections 1.2, 1.3, and 1.4, respectively). In section 1.2 the analyzed organism *M. pneumoniae*, its reduced genome, and the peculiarities of its metabolism are described. In the methodological background section (1.3) the research field systems biology (section 1.3.1) and computational modeling in biology (section 1.3.2) are generally introduced. Additionally, the metabolomics approach (section 1.3.3) and genome annotation and essentiality studies (1.3.4), including the analysis techniques applied in the presented work, as well as biological databases are introduced. The mathematical theories underlying the applied modeling approach are outlined in section 1.4.

Chapters 2 to 4 describe the different research projects, each containing a short introduction, a section for material and methods, a results section, and a discussion. The first project, relating to the development of **MyMpn**, the database and working platform for *M. pneumoniae*, is described in chapter 2. Since the database itself is the result of this project, the different steps towards this result are outlined, namely the database development process (section 2.2.1), the database implementation (section 2.2.2), the data integration and maintenance (section 2.2.3), and the implementation of the web interface (section 2.2.4) are briefly described. Furthermore, short descriptions of selected data analysis and visualization tools are provided (section 2.2.5).

The second and main project of this thesis, detailed in chapter 3, involves the *in vivo* exploration of the metabolite space and the development of a genome-scale constraint-based model for *M. pneumoniae* metabolism. In the Material and Methods section, the conducted computational (section 3.2.1) and experimental (section 3.2.2) procedures are outlined. The results section represents the iterative combination of experimental and computational analyses that allowed us to characterize key features of the *M. pneumoniae* metabolism. Subsections of the results describe the model building process (section 3.3.1), the identification (section 3.3.2) and quantification (section 3.3.3) of metabolites, and the definition of the *M. pneumoniae* biomass composition (section 3.3.4). Furthermore, the model refinement process (section 3.3.5) leading to the correction of the wiring diagram for metabolism and of the functional annotation of metabolic key enzymes, as well as the model validation by the qualitative determination of *in silico* growth capabilities on different carbon sources and by an *in silico* knock-out study (section 3.3.6) are acquainted. Finally, the model has been applied for the *in silico* prediction of double mutant phenotypes and of metabolic flux distributions along

the exponential growth phase (section 3.3.7). The subsequent integration of those flux predictions with *in vivo* monitoring of qualitative central carbon fluxes allowed us to calculate *in vivo* catalytic rates for glycolytic key enzymes. The results are summarized and placed in the context of actual literature in a broad discussion (section 3.4).

Chapter 4, presents the third project covering the in-depth analysis of the mycoplasma genome, which resulted in its re-annotation. To this end, we first analyzed available mRNA expression data (section 4.3.1. Second, we designed an *in silico* peptide library reflecting the theoretical coding capabilities of the *M. pneumoniae* genome (section 4.3.2). Finally, we combined our resuts with a proteomics analysis to integrate information about the correct size of translated genomic regions (section 4.3.3). The discussion of this chapter (section 4.4) apart from the presented results focuses on their ongoing integration with a genome-wide *in vivo* essentiality study

A summarizing discussion is presented in chapter 5 connecting the different projects to each other, interpreting the detailed results with respect to the state of the art in *M. pneumoniae* research and applied methodologies, and providing and outlook on ongoing an future projects related to the findings of this thesis.

1.2. Biological Background

1.2.1. *Mycoplasma pneumoniae*

Mycoplasmas, the smallest self-replicating organisms known [Morowitz and Tourtellotte, 1962], belong to the class *Mollicutes* (lat: molis - soft, cutis - skin). These wall-less bacteria evolved from more conventional progenitors in the Firmicutes taxon by a process of degenerative or reductive genome evolution [Razin et al., 1998]. Mycoplasmas are obligate parasites of humans, mammals, reptiles, fish, and plants, living in relatively unchanging niches that require little adaptive capacity [Razin, 1978, Razin et al., 1998]. They are composed of the minimum set of organelles: a plasma membrane, ribosomes and a circular DNA molecule [Razin et al., 1998]. The total number of known mycoplasma species, despite counting already close to 400, is constantly increasing. Their small size makes them interesting candidates to apply genome-scale analyses and already for 46 mycoplasma species the complete genome sequences can be found at the National Center of Biotechnology Information (NCBI) [Tatusova et al., 1999]. In addition, they promise to be the perfect candidates to assess the basic cellular functions of prokaryotes.

M. pneumoniae, one of the smallest mycoplasmas, preferentially colonizes human lung epithelial cells and is involved in a wide variety of diseases in children and adults [Waites and Talkington, 2004]. Historically, it was considered a virus rather than a bacterium due to problems in detection and laboratory cultivation [Eaton et al., 1945] and thereafter referred to as "Eaton's agent" until in 1962 Hayflick and colleagues identified it as a mycoplasma and named it *Mycoplasma pneumoniae* [Chanock et al., 1962a,b]. Due to the lack of a cell wall, *M. pneumoniae* is resistant to antibiotics targeting the synthesis of the peptidoglycan layer, such as penicillin [Eaton et al., 1945].

With a volume of only 0.067 femtoliters it has an oval shape with an attachment organelle that is used for movement and attachment to cells and other surfaces [Seybert

et al., 2006, Yus et al., 2009]. In its natural environment, the human lung, it grows attached to epithelial or other *M. pneumoniae* cells, under laboratory conditions clumps sticking to the bottom of the growth flasks are found [Seybert et al., 2006, Yus et al., 2009]. It can divide in about 8 hours (maximum speed observed [Seybert et al., 2006]), although in batch culture growth experiments doubling times of about 20 hours during the exponential growth phase have been observed [Yus et al., 2009]. *M. pneumoniae* M129, the *M. pneumoniae* strain used in our analyses, has a genome size of 816,394 base pairs encoding for only 689 proteins [Himmelreich et al., 1996, Dandekar et al., 2000]. The reduced genome is accompanied by a lean metabolic network lacking most anabolic pathways involved in cell building block synthesis [Pollack et al., 1997, Yus et al., 2009]. For ATP synthesis, it relies on glycolysis and organic acid fermentation due to the lack of a functional respiratory chain and a citric acid cycle [Dandekar et al., 2000, Yus et al., 2009].

M. pneumoniae is an ideal organism for systems biology studies due to a number of outstanding properties. The simple cell structure allows better analysis of cell components, for example with electron tomography [Seybert et al., 2006] or mass spectrometry [Maier et al., 2011], than would be possible in larger organisms. The small cell size also limits the abundance space of cell components to a minimum. mRNA expression data has been analyzed for a large number of different conditions facilitating information on the ability of *M. pneumoniae* to react on environmental perturbations, amongst them cold shock, heat shock, osmotic stress and starvation [Güell et al., 2009]. The proteome spans only 3 orders of magnitude in abundance and 60% of all supposed proteins in *M. pneumoniae* have been quantified by mass spectrometry, including 78.6% of all metabolic proteins [Maier et al., 2011]. In addition the relation between mRNA expression and protein abundance has been studied suggesting complex regulatory mechanism for gene regulation and protein synthesis. The minimal genome, the small number of encoded proteins, and the simple metabolic network allow the construction of genome-scale models for gene regulation and transcription, for the protein interaction network, and for the metabolism. Due to the lack of many pathways, *in vivo* metabolite measurements can be related directly to catabolic activity. Despite this apparent simplicity, *M. pneumoniae* shows a differentiated response to a variety of stress conditions similar to more complex bacteria [Güell et al., 2009]. In addition, in contrast to the smallest mycoplasma, *Mycoplasma genitalium*, it can be cultivated relatively easy under laboratory conditions without host cells, predestining it as a model organism for systems biology.

1.2.2. The Metabolism of *M. pneumoniae*

The metabolism of a cell, i.e. the respective set of biochemical reactions taking place in it, allows to take up nutrients from the environment and to process them into energy and cellular building blocks (Figure 1.2). Thereby, pro- and eukaryotes have many pathways in common, for example central carbon metabolism or nucleotide metabolism pathways [Alberts et al., 2008]. Other pathways exist only in either pro- or eukaryotes, only in a few organisms or even in only a few cell lines allowing the respective cells to accomplish specific functions, as for example functions related to specific organs in multi-cellular

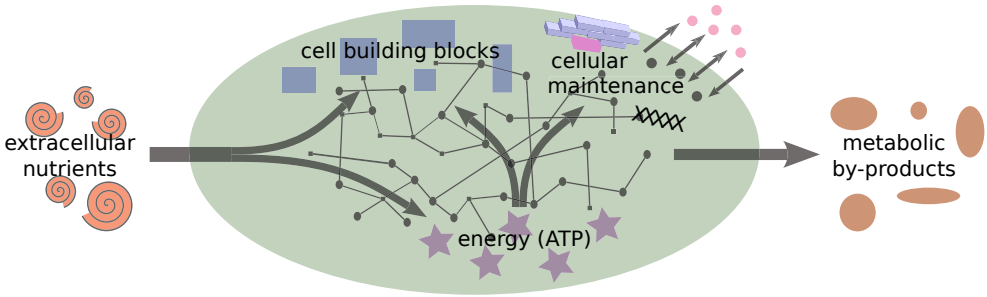

Figure 1.2.: Schematic representation of the metabolism of a cell: Nutrients (red spirals) are taken up from the environment and processed into cell building blocks (blue rectangles) and energy (purple stars). The energy is used for growth (production of cell building blocks) and cellular maintenance functions, such as DNA repair or maintenance of a favorable proton gradient across the membrane. Finally, metabolic by-products the cell cannot use (brown ellipses), are secreted to the environment. The metabolic network is represented by the grey dots and lines in the background.

organisms or to virulence in simple parasites. In case of *M. pneumoniae* the massive genome reduction resulted in a mainly linear metabolic network lacking most anabolic and energy producing pathways known from larger organisms [Pollack et al., 1997, Yus et al., 2009]. While even in *E. coli* more than 500 reactions span the metabolic network [Neidhardt, 1996], in *M. pneumoniae* only about 250 metabolic reactions are possibly taking place [Yus et al., 2009].

Glycolysis, pyruvate metabolism and arginine metabolism are the only energy producing pathways and the contribution of the arginine pathway to the total energy yield of a cell is negligible, providing maximally 1 ATP if no energy has to be consumed to provide the precursor arginine [Yus et al., 2009]. In addition to glycolysis and the pyruvate metabolism, *M. pneumoniae* disposes four main metabolic pathways, namely nucleotide metabolism, amino acid metabolism (including arginine metabolism), lipid metabolism and the pentose phosphate pathway (PPP). Finally, a number of cofactor processing pathways, providing secondary metabolites such as vitamins, CoA, NAD+, NADH or folic acid derivatives, and up-take systems for alternative sugar and carbon sources, such as fructose, mannose, mannitol, ribose, glycerol, G3P, and phosphatidylcholine, complete the metabolic network [Yus et al., 2009]. Interestingly, transport reactions to take up sugars, nucleobases, amino acids, fatty acids, vitamins, and other cofactors, and to export (toxic) metabolic by-products, amongst them organic acids and peroxide, make up about 30% of all reactions, reflecting the parasitic life of *M. pneumoniae*. The six main metabolic pathways account for more than half of all reactions, while the remaining reactions (17.5%) are related to the processing of alternative sugar sources or belong to secondary metabolites pathways, i.e. to CoA, folate, and cofactor metabolism (11.8%) [Yus et al., 2009].

M. pneumoniae takes up sugars (preferentially glucose, alternatively fructose, mannose, mannitol, ribose, glycerol, glycerol 3-phosphate (G3P), ascorbate and glycero-3-phosphocholine (G3PC)) from the environment. These sugars are processed through glycolysis and the produced pyruvate converted into either lactic or acidic acids. The generated acids are exported from the cell and the energy obtained during their synthesis is used for cellular maintenance functions and proliferation. During the design of a defined medium for *M. pneumoniae*, which from now on will be referred to as minimal medium, it has been shown that for RNA and DNA synthesis the two purine bases adenine and guanine and the ribosylated pyrimidine base cytidine have to be provided with the medium to allow growth [Yus et al., 2009]. Additionally, different fatty acids, amino acids (also in form of short peptides), and the precursors of all cofactor processing pathways have to be provided [Yus et al., 2009]. Mycoplasma lipids and the membrane composition have been studied already a long time ago [McElhaney and Tourtellotte, 1969, Pollack et al., 1970, 1973] showing that the lipid composition of mycoplasmas varies according to the fatty acids provided with the growth medium.

Despite the identification and quantification of major cell building blocks, such as the DNA, the total protein content, or the mRNAs, for most metabolites detailed information is lacking and the exact composition of an *M. pneumoniae* cell remains unknown. In addition, the contributions of many cellular maintenance processes, such as ATPase function, protein folding or DNA repair, on energy homeostasis are not known. Hence, a detailed characterization of the *M. pneumoniae* metabolism by combining mathematical modeling with *in vivo* analyses could provide the necessary information complementing the available data.

1.2.3. Genomics in *M. pneumoniae*

The essential genes of an organism are those genes that compose the minimum gene complement allowing growth and replication under the most favorable external conditions [Koonin, 2003]. However, approaches to determine the minimal genome sustaining life have not yet been successful which at least in part can be attributed to the errors in genome annotations not validated experimentally [Brenner, 1999].

The genome of *M. pneumoniae* has been sequenced twice and, in the current annotation, contains 689 protein-coding genes [Himmelreich et al., 1996, Dandekar et al., 2000]. Nevertheless, mistakes in genome annotations have been made since the first genome annotation of *Haemophilus influenza* [Fleischmann et al., 1995] for which just one month after the initial publication 148 amendments have been released [Casari et al., 1995]. Once established, those annotation errors spread around very fast among newly annotated organisms, since in absence of detailed experimental information many genes are annotated based on sequence similarity analyses. For *M. genitalium* the annotation error rate has been estimated to be 8% [Brenner, 1999]. The errors in the functional annotation of several genes detected due to sequence alignments and experimental results, as well as the three wrongly annotated genes detected with our constraint-based model suggest that the error rate in *M. pneumoniae* is not much smaller (chapter 3) [Yus et al., 2009].

In addition to protein and RNA coding genes, other genomic regions have gained attention during the past years. sRNAs, for instance, have been shown to account for up to 10 to 20% of the bacterial transcriptome [Sorek and Cossart, 2010, Güell et al., 2011] and the number of sRNAs in some cases equates the total amount of cellular transcription factors [Hershberg et al., 2003, Irnov et al., 2010]. tRNAs, rRNAs, and sRNAs have been annotated for *M. pneumoniae* [Himmelreich et al., 1996, Dandekar et al., 2000, Güell et al., 2009] and for *M. genitalium* the existence of antisense RNA has been proven [Lluch-Senar et al., 2007].

A genome annotation of high quality, preferentially based on experimental evidence, would facilitate the further analysis of cellular functions of *M. pneumoniae* and significantly improve the success probabilities for the determination of the minimal essential genome.

1.3. Methodological Background

1.3.1. Systems Biology

The origins of biology (greek: *bios* - life and *-logia* - study of) can be traced back to ancient times when Aristotle (384-322 BC) classified living things into categories, some of which in slightly altered definitions are still valid nowadays [Bohn, 1862]. The term biology for the first time appeared in German (as *Biologie*) at the end of the 18th century [Avila, 1995]. One of the first discovered principles underlying life was the theory of evolution proposed by Darwin [Darwin, 1859]. Modern biology encompasses many different disciplines attempting to study different aspects of life and living organisms [Avila, 1995]. In the middle of the 20th century the field of molecular biology arose, applying the so-called reductionist approach by studying the single molecules comprising living organisms are composed of [Auyang, 1999, Oshry, 2007]. Watson and Crick in 1965 discovered the structure of the DNA double helix laying the foundation for genes and genomes [Watson and Crick, 1953a,b]. Probably the youngest biological discipline, systems biology, emerged at the beginning of the 21st century when more and more large-scale experimental datasets became available and the limitations of the reductionist approach for the ambition to understand biological complexity became obvious [Kitano, 2001, Friboulet and Thomas, 2005]. Systems biology, assuming that understanding of a system is only possible by looking at it as a whole, brings together concepts and knowledge from all natural sciences and approaches unraveling of fundamental principles and properties of complex biological systems by relating the interactive properties of single system components to systemic functions [Noble, 2008, Westerhoff et al., 2009].

Establishing a fruitful collaboration between experimental and theoretical researchers is probably one of the biggest challenges within systems biology, since the research concepts and also the terminology used differ significantly from one scientific discipline to the next. However, to obtain system-level understanding of biological processes the development of mathematical models formally describing the scrutinized system and the integration of different experimental data, presumably provided by different researchers, is indispensable. Consequently, when designing experiments and mathematical models

in close collaboration with all scientists involved in a project, the possible knowledge gain far exceeds the simple sum of findings obtained by analyzing the different experimental datasets on their own.

For me, systems biology offers the great opportunity to profit from very diverse scientific expertise and the application of different research techniques and analysis tools in the aim to understand living organisms as a whole.

1.3.2. Mathematical Modeling in Biology

The integration of diverse experimental data into suitable computational models is of upmost importance when aiming to understand complex biological systems. Using mathematical formalisms to describe biological processes allows to unravel general principles as well as specific details of an examined system that are not amenable to experimental research. Additionally, when combining mathematical models with different experimental data it is possible to iteratively draw and validate hypotheses concerning the behavior of the scrutinized system [Kitano, 2002b].

In general, mathematical models in the biological sciences are designed to answer one or more specific questions about a biological process. In the first step, an appropriate mathematical approach has to be selected taking into account the size of the system, the type(s) of experimental data that will be integrated, and the question(s) the model is aimed to answer. This is not a trivial task, since each of the multiple available modeling approaches has different properties and restrictions that while perfectly granting the description of one process can frustrate knowledge gain over another. The different modeling approaches can be generally categorized according to the following criteria:

- Static modeling approaches describe a system under steady state conditions while dynamic approaches include information about the time-dependent changes of the system components.
- In deterministic models every state is uniquely defined by parameters and initial values, in contrast to stochastic models which include randomness, i.e. the variable states are described by probability distributions.
- Distributed parameter systems assume an infinite-dimensional state space, while in lumped approaches spatially distributed field variables are represented as single characters.
- In discrete modeling approaches formulas with discrete variables, often recurrence relations such as $f(x+1) = y * f(x)$, are used to fit data mostly obtained from point measurements, whereas continuous approaches normally apply differential equations to fit data obtained from serial or sustained measurements.

The method of choice depends on the combination of the above-mentioned properties necessary to describe the biological process of interest and the size of the modeled network in order to allow balancing of model complexity and model granularity. In addition,

the type(s) of experimental data that can be used for the model building and the question(s) the model is aimed to answer have to be taken into account in order to select an applicable mathematical formalism.

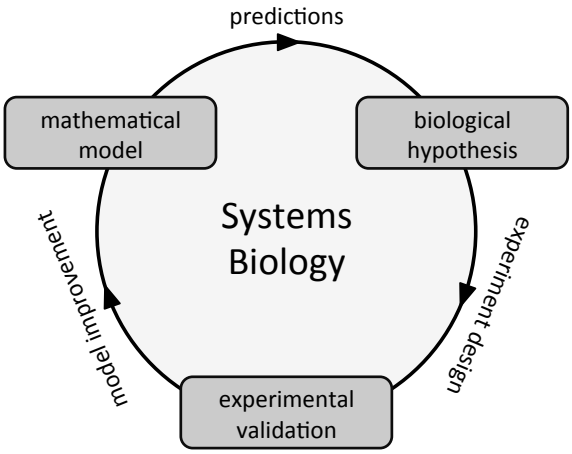

Figure 1.3.: Schematic representation of an idealized integrative systems biology approach: Model predictions are used to draw biological hypothesis that are validated experimentally, resulting in an improved model, a new hypothesis and so on.

Despite that in reality science is neither straightforward nor exactly cyclic [Alon, 2009], the model, the experimental data, and the drawn hypotheses should be refined in an iterative cycle until the model is able to accurately reproduce experimental findings of the described biological system (Figure 1.3). To this end, the parameters of the model have to be either measured directly or fitted to experimental data. , new experiments should be designed to allow the validation of *in silico* predictions and the subsequent adaptation of the model. The integration of different experimental data obtained by studying different aspects of the scrutinized system, for example data on proteins and metabolites when modeling a cellular signaling or metabolic pathway, grants higher accuracy then using only data of one kind.

Once the model is able to reproduce experimental findings, the predictive capacity of the model should be evaluated using experimental data not included in the model generation process. Models with high accuracy and specificity in predictions can provide information about properties of the examined system that have not been explored experimentally or that cannot be assessed due to lacking experimental techniques.

For this work, the constraint-based modeling approach is of particular interest and, therefore, outlined in section 1.4.1. It is important to note that in most of the cases different modeling approaches could be used to model the same biological process when focusing on different aspects of such process. No preset mathematical method exists to scientifically explore biological systems ”the right way”. So, one of the most important

principles for the usage of mathematical models in biology has been introduced by Box and Draper [1987]:

"Remember that all models are wrong; the practical question is how wrong they have to be to not be useful."

1.3.3. Metabolomics

The study of metabolomics deals with the high-throughput analysis of cellular metabolites. Despite their relevance for the phenotypic state of a biological system [Cornish-Bowden and Cárdenas, 2000, Fiehn, 2002, Nicholson and Lindon, 2008, Dunn et al., 2011, Buescher et al., 2012], metabolomics studies are lagging behind successes in adjacent fields, such as transcriptomics and proteomics. When examining the different steps of a metabolomics study (Figure 1.4), several reasons for this disequilibrium can be identified.

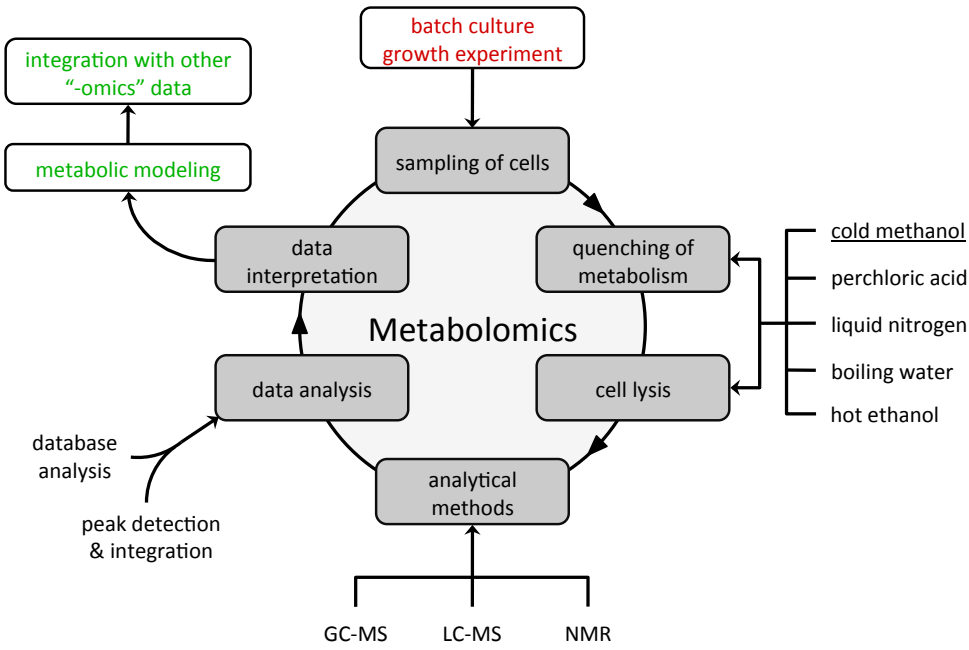

Figure 1.4.: Schematic representation of the different steps of a metabolomics analysis: First, samples are taken from the cell culture and then prepared for analysis, involving a metabolism quenching and a cell lysis step. In this study cold methanol (underlined) has been used for metabolism quenching. Second, the prepared samples can be analyzed by different analytical methods, such as NMR, LC-MS, or GC-MS. Finally, the obtained data needs to be analyzed and subsequently can be used for the design of mathematical models and further integrated with complementary information on the examined system.

Proteins and mRNAs are composed of homogeneous building blocks (either amino acids or nucleotides) and thus detectable by a single technical approach. Metabolites, in

contrast, are difficult to assess experimentally on a large scale when applying a single analytical method due to their chemical diversity and vastly different cellular abundance [Goodacre et al., 2004, Creek et al., 2011, Geier et al., 2011, Liberman et al., 2012]. In addition, they are not encoded by cellular DNA and their presence largely depends on transient cellular requirements. Finally, metabolites are often either unstable or short-lived, posing great challenges for sample preparation (Figure 1.4, cell lysis and metabolite quenching) and processing prior to analysis [Scalbert et al., 2009, van Gulik, 2010].

The different techniques established for the identification and quantification of cellular components are NMR and mass spectroscopy (MS), whereupon MS is normally coupled either to gas chromatography (GC) or to liquid chromatography (LC). When aiming to identify and quantify metabolites on a genome-scale, the combination of different approaches is recommendable in order to cover the diverse metabolite space of an organism. The three techniques commonly applied for metabolomics analysis are introduced below.

Nuclear Magnetic Resonance Spectroscopy (NMR)

Nuclear magnetic resonance is a physical phenomenon based on the magnetic properties of the nuclei of atoms that dunked in a static magnetic field are exposed to another, oscillating magnetic field. The subatomic particles, i.e. protons, electrons, and neutrons, have quantum mechanic spin. In some atoms, amongst them ^{12}C and ^{16}O, paired spins counterbalance each other (net spin equal to zero), but most atoms, for example ^{1}H, ^{13}C, and ^{15}N, do possess an overall spin. When a particle has a net spin, it can absorb a photon of frequency ν if placed in a magnetic field B. ν thereby depends on the specific gyromagnetic ratio γ, defined as the proportionality constant between the nuclear magnetic moment and the nuclear angular momentum, of the respective particle and, thus, resulting in a particle-specific NMR signal. When combining the NMR signals obtained from different particles of a complex molecule, it is possible to obtain information about the chemical and physical properties of such molecule [Atta-Ur-Rahman, 1986].

NMR is a technique that without requiring complex sample preparation steps allows to directly identify and quantify metabolites. Furthermore it allows to determine unknown components based on structural information. However, the application of NMR also has some disadvantages. It is not very sensitive, thus especially low abundant compounds are difficult to detect, and the technical equipment is expensive.

Chromatography

Chromatography is a laboratory technique to separate mixtures, such as cellular samples, and to determine the relative proportions of the different components [Harwood and Moody, 1989]. This separation is based on the components specific affinities towards two immiscible solvents, also called phases. When dissolved in one phase, the so-called mobile phase, the sample is transported through the second phase, denominated the stationary phase. Based on their specific chemical partition coefficients, the components of the sample travel at different speeds causing them to separate. Analysis of the retention times in the stationary phase grants the fast database powered identification of the

sample components. In the oldest chromatography technique, the paper chromatography, the sample to examine is placed on a cellular chromatography paper which is then set with the tip into a solvent. While the solvent rises through the paper the different components of the sample are taken along with different efficiency.

In **liquid chromatography (LC)** the sample under examination is dissolved in a fluid, with which it is moved through a column containing the stationary phase [Snyder et al., 2010]. LC is applicable to volatile and non-volatile compounds and a high number of different detectors can be chosen. In high-performance liquid chromatography (HPLC) for example, the liquid phase is conducted through the stationary phase by a high-pressure pump. However, imprecisions can result from ion suppression and quantification is only amenable with isotope-labeled reference compounds. Furthermore, LC alone does not positively identify all components since low abundant compounds can be hidden behind high abundant ones that have the same retention time.

Gas chromatography (GC) is a chromatography type able to separate and analyze (identify and quantify) biochemical molecules that can be vaporized without decomposition based on their volatility [Pavia et al., 2005]. The sample thereby is added to the liquid phase of a column and this column is heated. For the different sample components the retention time is measured and compared to retention times obtained from pure compounds. Disadvantages of GC are the limited range of available detectors and that sample derivatization is often required, thus introducing a higher experimental error with every preparation step.

Mass Spectrometry (MS)

Mass spectrometry (MS), an analytical technique measuring the mass-to-charge ratio of charged particles [Sparkman, 2000], is commonly used for the identification of molecules or sample compositions. To this end, the sample compounds are ionized and subsequently separated by an electromagnetic field. Usually, the resulting ion signals are detected quantitatively and then processed into so-called mass spectra. Those mass spectra are evaluated based on database information, which can be complemented by the determination of the mass spectra of known pure compounds.

In general, MS is conducted in combination with gas or liquid chromatography (GC-MS or LC-MS). By this combination the two techniques complement for their limitations in complex sample analysis, thus allowing to separate even highly similar molecules and to better identify (and quantify) the different sample components than would be possible by using only one method. Apart from analyzing biological samples for scientific purposes, GC-MS for example is also applied in drug testing, environmental analyses, or fire investigations.

1.3.4. Genomics

Survival, growth, and reproduction of cells depend on their ability to store, retrieve, and maintain the required genetic instructions [Alberts et al., 2008]. The genetic information is inherited from mother to daughter cells and the genes, sections of the genome encoding

proteins or functional RNAs, and their associated regulatory upstream regions constitute the mayor information-containing fraction of each genome. The field of genetics emerged at the beginning of the 20th century, but the mechanism by which cells are able to copy their genetic material literally thousands of times while maintaining it largely unchanged remained obscure until Watson and Crick [1953a,b] discovered the structure of DNA.

When towards the end of the 20th century, the sequencing of whole genomes became amenable [Fleischmann et al., 1995, Fraser et al., 1995], the analysis of genes and their functions on a global scale (genomics) became popular. Once sequenced, a genome has to be annotated, i.e. the different protein and RNA coding genes should be determined to allow the study of genes, of their expression, and of their translation into proteins in the particular organism. The annotation of sequenced genomes is conducted mainly based on sequence similarity to other, already annotated genes in other organisms. One of the major challenges in the annotation of entire genomes is putative transfer of annotation errors from other organisms, which can spread easily along newly annotated genomes due to the lack of experimental evidence for the annotated functions. Especially for larger genomes, an experimental validation of such automatically generated annotations is merely impossible due to the simple amount of genes that can be or cannot be expressed under varying conditions.

For the reduced genome of *M. pneumoniae* such a manual curation of the automatic annotation could reveal the impact of error propagation between the annotations of different species. However, to this end experimental data on genomic products, i.e. all produced transcripts and the translated proteins, has to be taken into account. Several high-throughput sequencing methods have been developed, e.g. Solexa sequencing or deep sequencing techniques, that allow to sequence DNA and also all expressed transcripts (transcriptomics). Other techniques, such as DNA microarrays and tiling arrays, provide information about the expression levels of mRNAs. Thus, it is possible to determine alternative transcriptions start sites (TSSs) inside operons but also inside annotated genes. While the transcriptome had been monitored under a sufficient amount of different growth conditions providing a comprehensive quantitative picture [Güell et al., 2009, 2011], for the proteome the available quantitative data did not allow to distinguish between different isoforms of the same protein [Maier et al., 2011].

1.3.5. Proteomics

Proteins, cellular molecules build from polypeptides, comprise the major fraction of the cellular dry weight and are responsible for nearly all biological functions [Alberts et al., 2008]. Their final physical shape, the conformation, is defined by four aspects. The amino acid sequence of each polypeptide, also called primary structure, is defined by the sequence of a gene and folds into the secondary structure, i.e. into regularly repeating local structures based on chain-internal hydrogen bonds, such as α-helices or β-chains. The tertiary structure describes the 3-dimensional shape of a single protein molecule and the quarternary structure defines the conformation composed of different protein molecules, which in general represents a minimum in the energy landscape. Nevertheless, proteins are not rigid but able to change their conformation, for example

upon interaction with binding partners, allowing them to carry out the diverse cellular functions [Alberts et al., 2008].

Proteomics is the large-scale analysis of the proteins of an organism, i.e. of the proteome, aiming to gain information on the translated fraction of the genome under different conditions, the protein quantities, structures, and functions [Wilkins et al., 1996, Anderson and Anderson, 1998, Blackstock and Weir, 1999]. As in the analysis of metabolites, proteins are commonly quantified by MS and sequenced with high-throughput methods just as genomic sequences or transcripts. Thereby, sample preparation can significantly increase the knowledge gain. A separation of the different proteins of a sample by weight using western blots, can allow to determine the actually translated ORFs, thus revealing organism-specific versions for some proteins or the existence of different isoforms. An example for the practical application of proteomics is the identification of putative drug targets based on the determination of desease-related proteins. Besides, in proteogenomics proteomic analysis techique are employed for the improvement of gene annotations and have been shown to facilitate the discovery of post-translational modifications [Gupta et al., 2007].

1.3.6. Biological Databases

Information exchange, and to this end data storage and accessibility, is one of the fundamental principles for scientific knowledge gain, not exclusively but especially in the electronic era. In former times the information had to be stored physically at specific places and thus, access to it was only possible through direct physical contact, i.e. going for example to a library, an archive, or a museum, or written request transmitted by mail. In contrast, today the world wide web provides the possibility to make information accessible to the whole humanity at once and online databases collect information about every imaginable topic. Some biological databases for example provide access to general information on biological numbers, the Bionumbers database [Milo et al., 2010], about enzymes, BRENDA [Scheer et al., 2011], about genes and pathways, KEGG [Kanehisa and Goto, 2000], or biological models, BioModel database [Li et al., 2010]. Others focus on information related to a specific organism, such as EcoCyc covering genomic and metabolomic information about *E. coli* [Keseler et al., 2011] or SubtiWiki for *B. subtilis* [Flórez et al., 2009].

Generally, the database development process can be divided into three main steps [Churcher, 2007]:

- The design of the database structure taking into account the information to be stored and the desired accessibility.
- The implementation of the database tables.
- The insertion of the data.

For the design of the database a relational scheme, also called unified modeling language (UML) class diagram or entity-relation (ER) diagram, is designed, which displays the database structure. This structure is defined by different tables of the database and

their internal connections by foreign keys. In addition, information on the unique key(s) of the table, that allow to distinguish the different database entries, and foreign keys used to connect data of different tables are indicated. Usually, also the type of data contained in each column of a table, for instance if the stored data will be a number or a string (series of characters), is included in this formal description of the database, since the data type dictates the size of the storage that is reserved for the database entries.

In most of the cases, subsequently or in parallel to the database development process an interface that allows and defines access to the stored data is implemented. Thereby, it is important to note that often a huge effort is made to provide intuitive access to and interaction with the information, thus granting the possible usage of the database to a diversity of people, not all being experts in database usage or familiar with the topic.

When aiming to obtain full understanding of an entire, even though single cell organism, a properly designed database facilitating the exchange of results, methods, and tools between an increasing number of research groups, is crucial. Specifically for the development of mathematical models, such easy access to well annotated and structured data is of upmost importance, since this data is not only pinpointing the question(s) a model is designed to answer but mainly dictating the applicable mathematical formalisms. Wrong or sparse data could prevent the successful model design, if a formalism is chosen that cannot reliably reproduce the biological properties and behavior of the modeled system.

1.4. Mathematical Background

Mathematical modelling offers a great variety of different approaches to determine specific aspects of the investigated system [Klipp et al., 2005]. Static modeling approaches describe a system under steady state conditions, thus allowing to model large systems without requirement for information about time dependent quantitative changes of the system components. In metabolic modeling, the constraint-based modeling approach has been established during the past two decades [Fell and Small, 1986, Savinell and Palsson, 1992a,b, Oberhardt et al., 2009, Feist et al., 2009]. However, static approaches are of limited viability when aiming to obtain insights into regulatory processes or to understand the functional mechanisms of cellular sub-systems in detail. Instead, dynamic modeling approaches, for example based on ODEs are utilized for studying time-dependent changes of biological systems [Klipp et al., 2005].

Independent of the modeling approach employed, some general properties apply to all mathematical models. The numeric relations of the system components to the biochemical reactions taking place in the examined system can be described by the stoichiometric matrix N.

$$N = \begin{pmatrix} n_{1,1} & \dots & n_{1,n} \\ \vdots & & \vdots \\ n_{m,1} & \dots & n_{m,n} \end{pmatrix} \tag{1.1}$$

The entries $n_{i,j}$ of N describe the quantitative involvement of component i into reaction j. The stoichiometric matrix can be used to extract information about the modeled system. For example, by calculating the rank of N, one can determine the linear independent components of a system, i.e. those components that cannot be described by a multiple of one or the combination of several other components.

1.4.1. Constraint-based Modeling

Constraint-based modeling is a static modeling approach applicable for large-scale metabolic networks. A constraint-based reconstruction is a union of (i) a stoichiometrically balanced metabolic model, (ii) a set of constraints for metabolic fluxes, and (iii) the list of genes responsible for the catalysis of reactions included in the model. The pseudo steady state-assumption, which states that the concentrations of the metabolites do not change over a certain period of time such that every discrete time point can be simulated as if the system would be in a true steady state. Therefore, to build a constraint-based model the knowledge of only the stoichiometry and the (ir)reversibility of the reactions that can occur within the modeled system is indispensable. Detailed knowledge on species quantities, reaction mechanisms and the respective kinetic parameters (which are usually unknown) is not required. The genes are connected with the reactions by logical expressions and do not mandatorily have to be defined. The relation of system components m and reactions n is defined by the stoichiometric matrix, a matrix of size mxn. Under steady state conditions the concentrations of the network components do not change, i.e.

$$N \cdot \nu = 0 \tag{1.2}$$

with $\nu = (v_1, ..., v_n)$ being the vector of reaction velocities fulfilling the steady state condition.

The constraints limit the metabolic fluxes and can be based on diverse experimental data, such as *in vivo* flux determinations, gene expression data or experimentally quantified metabolite conversions. Constraints commonly limit the available nutrients by setting maximum values for the source or the uptake reactions for sparse or known growth limiting nutrients. In addition, constraints can be used to define essential functions of the modeled networks, such as detoxification from metabolic by-products or the turnover of cellular entities such as mRNAs or proteins.

A number of analysis methods exist that allow to extract various features of the metabolic network, such as maximal growth yields or gene essentiality, and enable predictions about the flux distribution and the resulting behavior in different situations, some of which are described below.

Flux Balance Analysis (FBA)

FBA is an analysis method that predicts flux distributions of a metabolic reconstruction and is applicable to genome scale networks [Varma and Palsson, 1994a,b, Orth et al., 2010]. To this end, linear programming is used to optimize this flux distribution for a given set of nutrients and minimal requirements (defined by constraints $A \cdot \nu \leq b$) and

towards one or several objective functions:

$$F = c^{\top} \cdot \nu \tag{1.3}$$

with $c^{\top}$ being a vector of weights indicating the contribution of each reaction to the objective function. Common objective functions are biomass or energy production (maximization of the respective flux) or detoxification (minimization of the synthesizing flux and/or maximization of degrading flux). If the formulated FBA problem is feasible, i.e. the provided nutrients allow to match all requirements defined by the constraints of the model, an optimal flux distribution is predicted. However, this may not necessarily be the only optimal solution. Indeed, the analysis of flux distribution spaces showed that, especially in more complex networks with many branching components (i.e. components that do interconnect different routes of the modeled system), it is highly probable that more than one optimal solution with respect to the objective function and the simulated conditions exist [Mahadevan et al., 2002].

Growth Simulations

The prediction of flux distributions can also be used to assess growth rates for the simulated organism, if the FBA is optimized towards growth. However, to this end the biomass composition of the respective organism, i.e. the different cellular components, ranging from macromolecules, such as DNA, RNA, or proteins, to simple metabolic molecules, has to be determined. The biomass composition of mycoplasmas differs significantly from those of higher prokaryotes, such as *E. coli*, due to the lacking cell wall and the reduced cell size and genome. In a general form, the biomass equation defining the average macromolecular cell composition of *M. pneumoniae* and putatively the other mycoplasmas reads:

$$\begin{array}{l} DNA + RNA + proteins + lipids + freebases + \\ free\ amino\ acids + fatty\ acids + cofactors \end{array} \longrightarrow Biomass \tag{1.4}$$

The different cellular components have to be identified and at least in their majority quantified to allow a realistic reproduction of the metabolic processes involved in their uptake and synthesis.

Once the biomass composition of the modeled organism is known, FBA can be used to simulate growth and determine *in silico* doubling times. If growth, represented by biomass production, is the objective function for the FBA problem, the resulting objective value ov can be directly related to the growth rate t_{doub}. When exponential growth is simulated, this relation is in general described by:

$$t_{doub} = \frac{ln(2)}{ov}. \tag{1.5}$$

If the cell population is maintained at a constant size, i.e. if the steady state assumption also applies to the total amount of cells simulated as assumed for the model presented

in this thesis, then this relation simply reads:

$$t_{doub} = \frac{1}{ov}. \tag{1.6}$$

In Silico Knock-outs

To simulate gene knock-outs *in silico*, all reactions catalyzed by the gene that shall be knocked out are silenced (i.e. their maximum flux is set to zero). FBA with growth as objective function is used to determine the knock-out effect on the system's behavior and the flux distribution. Thereby, the objective value of the knock-out simulation ov_{ko} when compared to that of the wild type ov_{wt} allows to distinguish between different mutant phenotypes. If ov_{ko} equals ov_{wt}, the knocked out gene has no effect on the growth rate of the organism under the simulated conditions. When $0 < ov_{ko} < ov_{wt}$, then the respective gene knock-outs produces a reduced fitness phenotype. An $ov_{ko} = 0$ represents growth arrest but not death of the respective mutant and if no ov_{ko} is given, the *in silico* gene deletion rendered the FBA problem infeasible, i.e. at least one of the minimum requirements specified by the constraints cannot be matched by the knock-out mutant. When silencing the reactions of two different genes at the same time double knock-out phenotypes can be predicted.

1.4.2. Enzyme Kinetics

As mentioned beforehand, several mathematical formalisms exist to describe different aspects of biochemical reactions when aiming to analyze time-dependent changes of a biological system [Klipp et al., 2005]:

- **Boolean Rules** describe the qualitative time-dependent relation between the different system components in form of "if-then" statements for discrete steps, such as $C(t+1) = A(t) + B(t)$ with A, B, and C being the system components and t the time. The components of a Boolean model have two possible states: they are either present and active or not active (regardless of if they are present or not) with respect to the modeled process.

- The **Law of Mass Action** states that the velocity v of a reaction j is proportional to the probability that the involved reactants X_i meet, i.e. to the product of the concentrations of all reactants.

- **Michaelis-Menten-like Kinetics** are usually employed for the description of enzyme-catalyzed reactions. They base on an enzymatic mechanism introduced by Brown [1902] for irreversible one substrate (S) reactions without effectors. This mechanism assumes that in such enzyme-catalyzed reactions a reversible formation of an enzyme-substrate complex (ES) is followed by the irreversible release of the product (P):

$$\mathrm{E} + \mathrm{S} \underset{\mathrm{k}_{-1}}{\overset{\mathrm{k}_1}{\rightleftharpoons}} \mathrm{ES} \longrightarrow \mathrm{E} + \mathrm{P} \tag{1.7}$$

To build a model that is able to reproduce experimentally observed behavior and possibly even has predictive power the kinetic laws describing the biochemical processes under determination have to be chosen with caution. Apart from the question the model is aimed to answer, quantity and quality of the available data can influence on this decision, since the appropriate kinetic law should avoid the problem of overfitting described in section 1.4.3. In the simplest case, i.e. when the reaction velocity v only depends on one system component, the Law of Mass Action is applied in its simplest form:

$$v_j = k \cdot X \tag{1.8}$$

When the reaction velocity depends on more than one system component, higher order kinetics, considering the probability that the two components meet, have to be applied:

$$v_j = k \cdot \prod^{i} X_i \tag{1.9}$$

For enzyme catalyzed reactions, enzyme kinetics are usually employed. In general, the maximum velocity V_{max} of a reaction is equal to the product of the enzyme concentration E and its catalytic rate k_{cat}.

$$V_{max} = E \cdot k_{cat} \tag{1.10}$$

Probably the most common kinetic applied in models for biological processes is the Michaelis Menten Kinetic, that for a simple reaction as described by Equation 1.7 reads:

$$v = \frac{V_{max} \cdot S}{K_m + S} \tag{1.11}$$

with K_m being the Michaelis constant that represents the substrate concentration at half maximum velocity.

However, the more detailed the mathematical formalism, the less straight forward the parametrization of the included variables. For large biological systems, such as metabolic networks, dynamic modeling approaches are difficult if not impossible to apply since they require detailed experimental information on the majority of the system components. Hence, a possible extraction of information about the network dynamics already from static metabolic models, provides a great advantage for the design of dynamic models for the different metabolic sub-systems. To this end, the integration of knowledge on enzyme kinetics with experimental data provides a powerful method, allowing, for example, to calculate kinetic parameters directly from *in vivo* data by adapting Equation 1.10.

1.4.3. Model Parametrization

The model parametrization is one of the greatest challenges during the model building process [Kitano, 2002b]. Parameters describe specific properties of the examined system, e.g. catalytic rates of enzymes, inhibitory and activatory regulation, binding or dissociation constants or the velocity of a reaction. They have to be calibrated carefully to allow the model to accurately describe the system properties and to reproduce experimental findings, which is indispensable when aiming to obtain biological knowledge gain from

model predictions.

However, in most of the cases, the experimental data available is either sparse or not suitable for the applied modeling approach. Furthermore, this data reflects the behavior of the examined biological system under the specific conditions in which the experiment has been accomplished. Indeed, when the physical and chemical conditions, for example the temperature or the pH of the growth medium in cell culture, are changed, the cellular behavior changes and thus do the parameters describing the modeled cellular sub-system. In addition, experimental data is generally not accurate due to experimental errors introduced during sample preparation and analysis techniques. In contrast, mathematical equations are exact (based on their respective definitions), except from rounding errors that result from limited accuracy in computational calculations. They can only consider a defined error, for instance by allowing a deviation for specific parameters or by considering probabilities as in stochastic modeling approaches. Therefore, it is of upmost importance that the chosen mathematical formalism reflects the available data in order to avoid overfitting when many parameters are fitted to only few data points [Draper and Smith, 1998]. The term overfitting describes the determination of a parameter set for a mathematical model that despite being able to reproduce the data used for fitting does not represent the actual biological properties of the system but only a possible mathematical formalism to describe the training data. In general, those parameter sets do not reliably predict independent evaluation data, not used during the fitting process.

In this work, the constraint-based modeling approach and FBA have been applied. Based on the assumed steady state, no kinetic parameters as known from ODE based approaches are included. Instead, only the stoichiometric information, the reaction reversibilities, and the effective reaction velocities define the prediction results. The stoichiometric information and the reversibility base on the inherent biochemical properties of the modeled organism (genes, proteins, cellular composition) and have to be defined carefully during the model reconstruction process. If possible, they should be based on *in vivo* metabolite measurements, either directly identifying the metabolic intermediates or proving pathway activity based on end product quantifications, or on labeled isotope tracing experiments. The effective reaction velocities reflect the simulated growth conditions. To this end, they are limited by so-called constraints, either defining a minimum or a maximum flux for the respective reaction, in order to allow the *in silico* reproduction of nutrient conditions and metabolic functions, such as detoxification or cellular homeostasis.

The modeling platform ToBiN (**To**olbox for **Bi**ological **N**etworks, the source code is available at http://github.com/miguelgodinho/tobin), employed for the presented metabolic model, uses constraints of the unit mmol*g^{-1}*h^{-1}. Therefore, all experimental data, for example the number of glucose molecules taken up per cell and second, has to be converted into this unit. Once the constraints have been determined correctly, the FBA solver optimizes the flux distribution towards the defined objective function, growth in form of biomass production for *M. pneumoniae*. Changing sets of constraints allow to simulate different growth conditions and the resulting flux distributions can provide information about the metabolic behavior, the network connectivity, its adap-

tive capacities towards conditional changes, and on the importance of single network components, especially of metabolic enzymes.

The one correct formalism for the calculation all flux constraints of a constraint-based model, which are needed to allow the correct description of the simulated network behavior, does not exist. Nevertheless, maximum constraints are often calculated from catalytic rates or *in vivo* metabolite depletion or synthesis rates. Minimum constraints, for instance, can be determined based on turnover rates or reflect qualitatively essential functions, such as the degradation or secretion of toxic metabolic by-products. Finally, it is important to note that as for any other mathematical modeling approach, the data used for evaluating the hypotheses drawn from the model should be independent of the training data used to determine the model parameters.

2. An Interactive Database for *M. pneumoniae*: MyMpn

This chapter aims to introduce into *M. pneumoniae* research by describing the development of an interactive database and working platform for *M. pneumoniae*, **MyMpn**, aimed to be released at the end of 2012. The database design and development have been finished and the manuscript is in preparation. However, additional data and a comparative analysis of *M. pneumoniae* and *M. genitalium* shall be included before the public release. This can lead to changes on the content of the database and on the different sections of the interface when compared to the state of the art described herein. Since the design process of a database itself is the major result, there is no separate Material and Methods section contained in this chapter. Instead, the applied methods are mentioned and detailed throughout the results section.

I am responsible for the design of the database structure, the data collection and formatting, the coding of statistical analysis tools, the design of the clickable metabolic map, and the coordination of the database project. I further contributed to the development of most included analysis tools by accomplishing functional tests. The implementation of the database and the web interface, including the advanced search tool BioMart, the genome browser, the interactive tool for the metabolic map, the comparative alignments of M. pneumoniae proteins to other organisms, and the embedding of the statistical and visualization tools has been accomplished by the Bioinformatics core facility at the CRG, Barcelona.

2.1. Introduction

Mathematical modeling of biological systems requires comprehensive (experimental) data that is properly annotated, well sorted, and easy accessible. During the past years a wealth of organism-wide datasets for the genome, the transcriptome, the proteome and the metabolism of *M. pneumoniae* have been produced by the different groups involved in the mycoplasma project [Güell et al., 2009, Lluch-Senar et al., manuscript in preparation, Kühner et al., 2009, Maier et al., 2011, van Noort et al., 2012, Yus et al., 2009]. This data was locally stored in the different research groups and exchanged upon request. Obviously, this is not an optimal solution for information exchange based on data sharing, since each researcher has to know if the needed data is available and who produced it. Furthermore, the data has to be requested each time what is time-consuming, especially if personal availability is taken into account. However, when aiming to study an organism applying the systems biology approach, it is of upmost importance to know

what kind and quantity of data is available, since this information directly influences on the decision which modeling approaches to apply for the different cellular subsystems.

We developed and implemented a database for *M. pneumoniae* that contains the available data on genomics, transcriptomics, proteomics, metabolomics, phylogenomics, and regulomics, as well as data analysis and visualization tools aimed to supply comprehensive information about one of the most promising model organisms in systems biology.

2.2. Results

We developed an interactive database, **MyMpn**, providing a data storage for *M. pneumoniae* data produced by different research groups and a working platform for researchers interested in mycoplasmas (http://mycoplasma.crg.es). In addition, the cataloging of the data and the greatest possible standardization of its storage format facilitates the development of mathematical models for *M. pneumoniae*. The database development process can be divided into five main steps (Figure 2.1):

1. the collection of the experimental data to be included and the subsequent definition of the general database structure,
2. the design and implementation of the database tables,
3. the entering of the available data into the database,
4. the implementation of the web interface,
5. the integration of existing and newly developed analysis and visualization tools,

Those steps are outlined in the following subsections.

2.2.1. Data Collection and Structural Design of the Database

In a first step towards the design of a database for *M. pneumoniae*, we collected the experimental data from the different research groups to get an overview about the data types and the possible interconnecting features. We obtained mRNA expression data [Güell et al., 2009], growth curve measurements [Yus et al., 2009], metabolite assay data [Yus et al., 2009], protein quantification data [Maier et al., 2011], results from phylogenomic studies (unpublished results) and the screening of a transposon library [Lluch-Senar et al., manuscript in preparation], information about protein complexes [Kühner et al., 2009] and about post-translational modification sites [van Noort et al., 2012], functional annotation data [Güell et al., 2009, Yus et al., 2009], sequencing results [Yus et al., 2009, Güell et al., 2009], the metabolic network [Yus et al., 2009], and pictures and videos obtained by light microscopy (unpublished results) and nuclear magnetic resonance tomography [Seybert et al., 2006]. Due to the different format of of the data, varying from text files over vector-based and other graphics to movies, and the differing size and content of the experimental datasets (e.g. different numbers of samples and

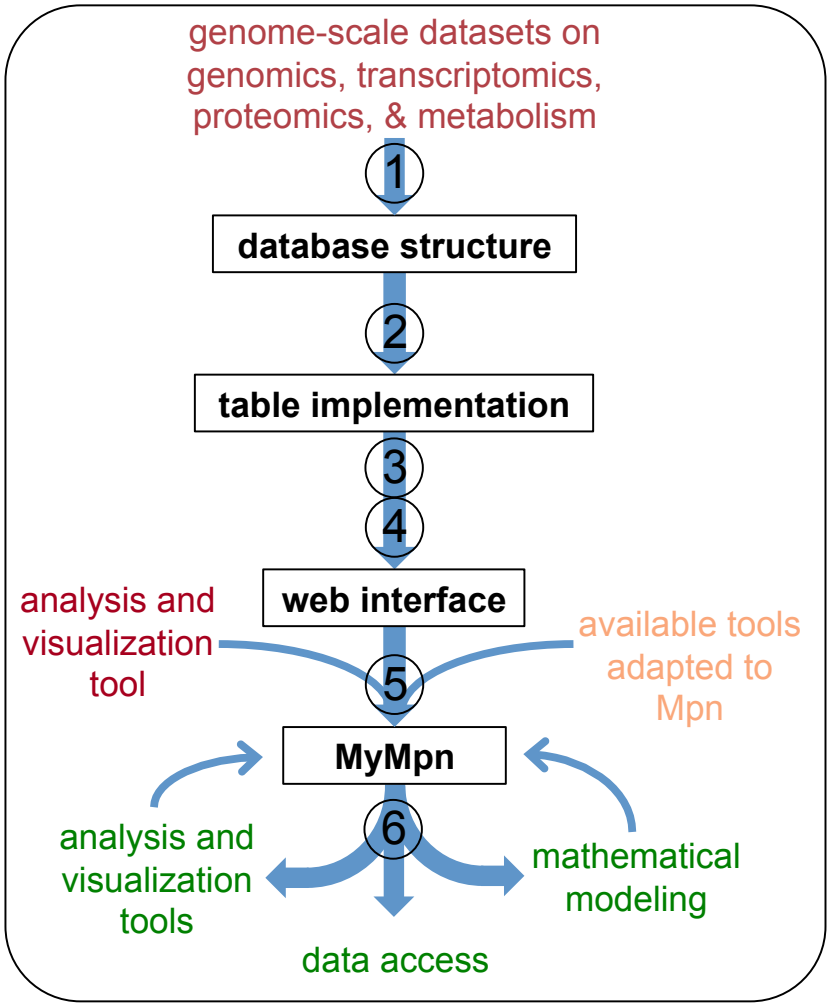

Figure 2.1.: Workflow for the design of the *M. pneumoniae* database, **MyMon**. 1: Data collection and definition of the database structure; 2: Implementation of the database tables; 3: Entering of the data; 4: Development of the web interface to provide data access; 5: Integration of data analysis and visualization tools; 6: Applications of the **MyMpn** database. Database input from the mycoplasma project has red fonts, existing analysis tools that have been adapted to *M. pneumoniae* (Mpn) orange fonts and the actual applications of the database green fonts.

replicates), the definition of appropriate tables that allow space efficient data storage and fast data access is not a trivial task.

To provide organism-wide information, the database cannot put any restrictions on the data types to be included, neither for the accessible data nor for data that may become available in the future. The different experimental results have been structured allowing not only the simple storage but also the interconnection of different experimental results that provide information on the same aspect of interest, for example mRNA and protein abundances for the same gene, or all genes that have been specifically regulated under a certain perturbation condition. To this end, whenever possible data and information have been sorted associated to MPN IDs. Those IDs refer to different sections on the genome, amongst them genes, non-coding RNAs, and regulatory regions, as well as to the encoded proteins or RNAs. Data which cannot be assigned to different MPN IDs, for example metabolite identification and quantification data, has to be organized in form of raw and processed data sheets which can later be used online for visualization or be downloaded for further analysis.

In conclusion of the data collection process, we identified two general rules according to which the database has to be designed: i) experimental data needs to be stored allowing as well access to the raw data as also to connect results from different experiments and ii) to provide practical tools for data analysis, a possibility to temporarily upload and store unpublished data in the genome browser ('GBrowse') without thereby providing free access to this data is required. While the first rule seems to be obvious, the second rule would provide the possibility to analyze newly produced data in a standardized way, what implicates two major advantages. The results from different experiments accomplished by different researchers are comparable and the integration of new data into the existing database structure is facilitated.

2.2.2. Design and Implementation of the Database Tables

In collaboration with the Bioinformatics Core Facility at the CRG Barcelona, we designed the structure of the **MyMpn** database based on the available experimental data, but taking into account that more and different data will be produced in the future (UML class diagram: Figure 2.2). To further allow the implementation of an advanced search tool, able to extract and combine information from the different data tables, we designed all data tables according to some general principles:

- Usage of separate tables for the different cellular entities, i.e. genes, mRNAs, proteins, etc.
- Usage of database internal IDs to uniquely identify the different database entries.
- Connection of information from different tables if possible based on the database internal 'gene_id' assigned to each annotated gene, which accordingly, should be included as a foreign key in all tables which store gene, RNA, or protein related information.

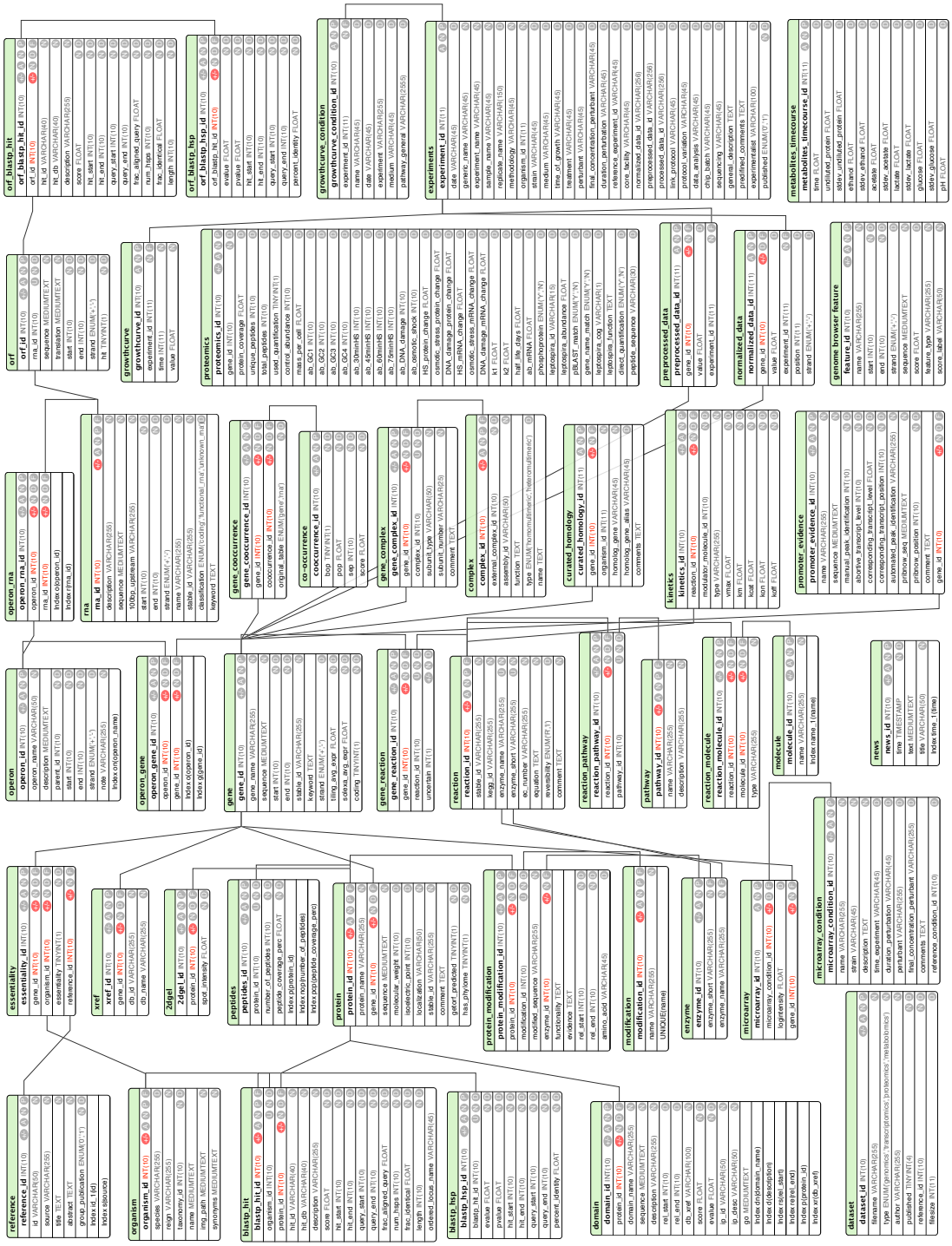

Figure 2.2.: UML class diagram of **MyMpn** displaying the different database tables, the format of the contained data and the connection of that data amongst the different tables. This figure is provided in large format for take out at the very back of this thesis.

- Experimental results are stored in form of the normalized raw data, and in pre-processed as well as processed format for different analyses. Each experiment is indexed by a unique experiment ID and stored in form of separate samples whenever biological or technical replicates have been obtained. Those samples, in the general experiment table (identified by their unique sample ID and the experiment ID, which is the same for all the replicates of one experiment), regardless of the experiment type, are described based on the applied analysis technique and parameters.
- Integration of visual results, amongst them microscope pictures, western blots, or electrophoresis gels in the web interface and for published data with a download possibility.

In total, as of today, the database contains 46 tables (Figure 2.2). The gene table could be considered as the central table of the database, since almost all tables are either directly or indirectly connected to this table via the internal gene ID. This gene ID, as all the other internal IDs used to uniquely identify most entities included in the database, is not visible to the public and used to connect the different database tables in form of foreign keys. The information about protein complexes, for instance, is stored in three different tables (tables 'gene', 'protein', and 'complex' in Figure 2.2), connected amongst each other by an intermediate table ('gene_complex' table) connecting a complex ID to a gene ID, which is also used to relate to the protein table, using an internal 'gene_complex ID' apart from the gene ID to connect the gene and the protein table. The final structure of the *MyMpn* database was implemented by the Bioinformatics Core Facility at the CRG Barcelona using the open source relational database management system MySQL (http://MySQL.com).

2.2.3. Data Incorporation

To allow not only entering but also simple updating of already included data, a parser has been programmed by the Bioinformatics Core Facility at the CRG, Barcelona. This parser gets as an input a collection of tables containing experimental data, but also general information such as annotation information on sequences or functions, as well as different identifiers connecting our data to commonly used biological databases, such as KEGG [Kanehisa and Goto, 2000] or the PDB [Berman et al., 2000], and sorts the contained information into its corresponding tables. Thus we assure that the information can be maintained updated by different researchers, while for updating the database a database administrator can control the correct format of the newly provided or updated data sheets.

To assure consistency in data formats for experimental data, new data has to be provided in standardized formats and with additional information describing the experimental conditions applied. In a general experiments table, each experimental sample is identified by a unique sample name and the experiment is described, for example by detailing amongst others the used organism, the time of growth at which the sample has been taken. The raw and pre-processed data is provided in tabular format preferentially

MPN ID based, and otherwise depending on the data. Metabolite identification and quantification data, for example, is stored based on the metabolite name and the unique experiment (sample) ID.

2.2.4. The Web Interface

The web interface aims to provide data access in an intuitive way (Figure 2.3). To this

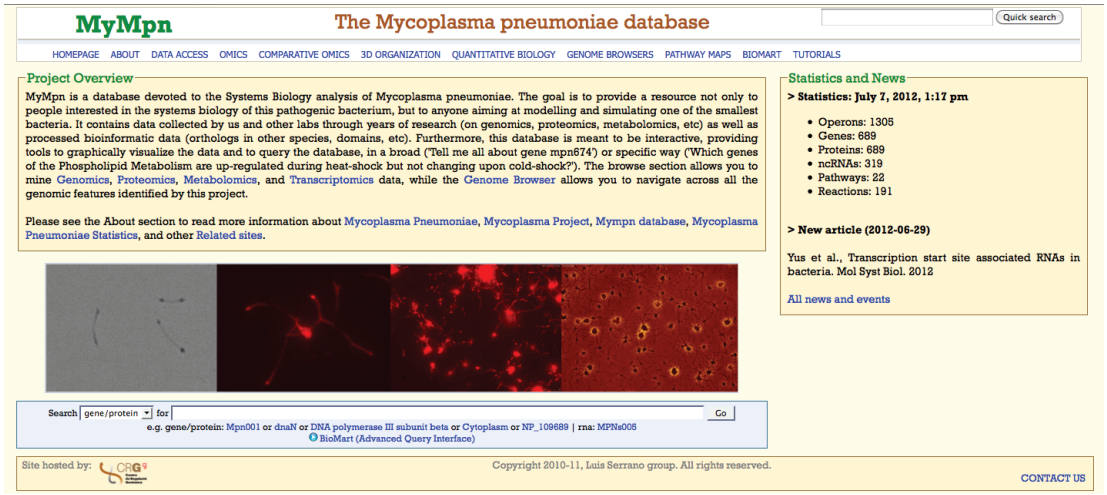

Figure 2.3.: **MyMpn** homepage. A short introduction into the aims of the database, relevant statistics and news, such as recent publications, some nice microscopy pictures of*M. pneumoniae* as well as the different sections of the database in the menu bar are shown.

end, the information provided by the **MyMpn** database is assorted into different sections, namely "homepage", "about", "data access", "omics", "comparative omics", "3D organization", "quantitative biology", "genome browser", "pathway maps", "biomart", and "tutorials". The "homepage" describes the general aim of the **MyMpn** database and lists the latest news, such as recent publications of the group. In the "about" section, *M. pneumoniae*, the "Mycoplasma project", the **MyMpn** database, the involved research groups (related sites), and important literature are introduced. Thereby, descriptions of the data contained in the database and information on the applied experimental or computational techniques to produce this data are given in the '**MyMpn** database' subsection.

To access specific information, for example information about a gene or protein in *M. pneumoniae*, the "data access" section provides different simple search possibilities. Amongst others those include a section for mathematical models, where *M. pneumoniae* models will be provided for download upon publication. The genome-scale datasets are amenable through the "omics" section, which is further subdivided into 'Genomics', 'Transcriptomics', 'Proteomics', 'Metabolomics', and 'Regulomics', including amongst others regularly updated lists of homologous proteins in other organisms ('Genomics' →

'Homology based on pBLAST'), a visualization of gene expression ('Transcriptomics' → 'Gene expression graphs'), protein quantification results ('Proteomics' → 'Protein quantification (Mass spectrometry)'), and information about metabolites ('Metabolomics' → 'Metabolites'). The 'Regulomics' subsection will once available provide access to the Chip-seq data analysis currently conducted.

The "comparative omics" section contains an analysis tool for synteny, and three subsections for comparative genomics, transcriptomics, and proteomics, that are still under construction. The synteny tool, aligning genomic regions of *M. pneumoniae* and *M. genitalium*, facilitates information about common genomic features of the two closely related organisms. In section "3D organization" ER tomography results [Seybert et al., 2006] and light microscopy studies visualizing different properties of *M. pneumoniae* cells are shown. Statistical information on the biophysical properties of *M. pneumoniae* and for selected other bacteria can be found in section "quantitative biology".

Sections "genome browser", "pathway maps", and "biomart" provide access to three important data analysis tools described in detail in the next section (2.2.5). The "tutorials" section is aimed to teach the user the usage of the database, but also of the different analysis and visualization tools either by short indicative descriptions or by video tutorials that step by step guide through the putatively more difficult applications of the **MyMpn** database.

2.2.5. Analysis and Visualization Tools

Several data analysis and visualization tools have been incorporated into the **MyMpn** database. As examples, the genome browser and the clickable metabolic map, are outlined in detail.

Genome Browser

The **MyMpn** database provides two different genome browser. The 'MyGBrowser' allows to browse the *M. pneumoniae* genome, thereby displaying a selection of genomic features, such as operons, genes, ncRNAs, transcription start and transcription termination sites (TSS and TTS), Pribnow boxes, as well as DNA and RNA hairpins for both strands. In contrast, the *Mycoplasma pneumoniae* genome browser ('GBrowse' in the drop down menu of "genome browsers") does not only provide tracks for additional selection, for example Chip-seq profiles or tiling array results, but also for uploading personal data into a temporary storage (Figure 2.4). Since each time the 'GBrowse' section is selected, a personal version opens in a new window, tracks that are added for analysis by a researcher are not available for other people, thus providing an online possibility to integrate new and unpublished data with the data already available in the **MyMpn** database (Figure 2.4, select and customize tracks in the upper menu bar). In addition, this genome browser enables the user to customize the view, for example by highlighting specific features in the 'Preferences' section.

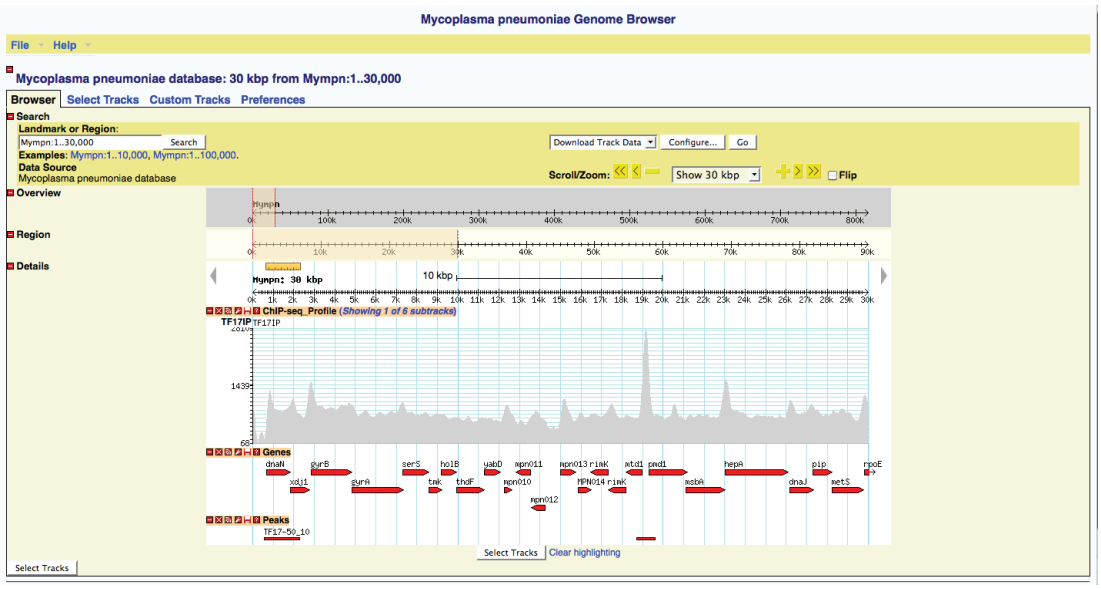

Figure 2.4.: As an example for the features of the *Mycoplasma pneumoniae* Genome Browser a cutout of 30000 base pairs of the *M. pneumoniae* genome is displayed with a selected Chip-seq profile, the annotated genes of this genomic region and the peaks indicating transcription start sites identified by this Chip-seq profile.

Clickable Metabolic Map

The clickable metabolic map has been developed based on the reaction list of the final metabolic model *i*JW145. The map was designed using CellDesigner 4.1 [Kitano et al., 2005]. The positional information stored in the xml file produced by CellDesigner allows to connect information contained in the database to the different model species (proteins, metabolites) and reactions. To this end, we used CellPublisher, a visualization tool for cellular networks [Flórez et al., 2010].

The metabolic map provides information about the reactions catalyzed by the different metabolic enzymes and provides links to the gene table and other databases, such as KEGG [Kanehisa and Goto, 2000]. Thus, a potential user can not only navigate easily through the metabolic network of *M. pneumoniae* but also find further related information or interpret *in silico* growth simulation results. In addition, the lists of different pathways, enzymes, and metabolites provide a possibility to have a visual impression for the entities involved in specific metabolic processes. Furthermore, this allows to connect metabolic species (enzymes and metabolites) that for better clearness are represented by several copies in the map.

2.3. Discussion

We developed an interactive database that serves as a data storage and analysis platform for the mycoplasma community. The database contains data on genomics, transcriptomics, proteomics, metabolomics, and phylogenomics reflecting the state of the art of published data about *M. pneumoniae* and further data allowing to draw biologically relevant conclusions for the database release. This data is connected by internal identifiers allowing to assign all information available about a gene and its encoded protein to the MPN ID. Simple access to the data is provided by search masks for different topics, amongst others keywords, genes, and ncRNAs. Furthermore, the advanced search tool BioMart (which is still under construction to include the additional data being prepared for inclusion at the moment) will facilitate the connection of different data types, for example if one would like to know all the genes upregulated upon heat shock.

Selected data can be analyzed with different statistical and visualization tools. The *Mycoplasma pneumoniae* Genome Browser provides a framework for advanced sequence analysis as well as integration and comparison of different experimental data, for example on mRNA expression and protein sequencing. Despite not yet being released to the public, the database is already used by members of the mycoplasma project to analyze and visualize experimental and computational results as well as to access the genome-wide datasets produced from other researchers.

Information access is one of the most important prerequisites for the successful design of mathematical models for biological systems. The available experimental data not only dictates the questions most mathematical models are aimed to answer but also directly influences on the decision which mathematical formalism to chose in each case. For *M. pneumoniae*, the collection of the available experimental data pinpointed open questions in our understanding of the metabolism of this minimal organism, especially with respect to the cell composition, the essential metabolic functions, as well as the regulation of the central carbon metabolism responsible for energy homeostasis.

3. Metabolome Analysis and Characterization of *M. pneumoniae* Metabolism

This chapter addresses the experimental and computational analysis of the metabolism of *M. pneumoniae*, the major project of this thesis, presented in:

1. **J.A.H. Wodke**, J. Puchalka, M. Lluch-Senar, M. Godinho, V. Martins dos Santos, L. Serrano, E. Klipp, T. Maier: "Metabolic modeling and quantitative biomass and energy balancing in *Mycoplasma pneumoniae*", *in preparation for re-submission to MSB*

2. T. Maier, J. Marcos, **J.A.H. Wodke**, M. Liebeke, R. Gutiérrez-Gallego, L. Serrano: "Comprehensive metabolome analysis and quantitative integration with proteomics data in *Mycoplasma pneumoniae*", *in preparation for submission to PNAS*

I was involved in the project design and development, I built the metabolic model, conducted the integration of the experimental data, defined the biomass composition, and I analyzed the in silico results. In addition, I carried out the bioinformatic analysis of the experimental results and was involved in figure and table generation for the second paper. I wrote the manuscript of the modeling paper (with help of Tobias Maier) and also commented on the experimental manuscript, especially with respect to the integration of experimental and computational results.

3.1. Introduction

When aiming to understand an organism in its entirety, accurate descriptions of the biochemical composition of the respective organism and the reaction network responsible for the uptake and subsequent processing of nutrients into energy and cell building blocks, i.e. of the metabolism, are indispensable. Detailed experimental data on genomics, transcriptomics, proteomics, and metabolomics have been published [Yus et al., 2012, Güell et al., 2009, Kühner et al., 2009, Maier et al., 2011, van Noort et al., 2012, Yus et al., 2009] and integrated into our *M. pneumoniae* database, MyMpn. Analyzing the available data for its feasibility with respect to the development of mathematical models for *M. pneumoniae*, we found that quantitative information on metabolites and the exact composition of a *M. pneumoniae* cell remain unknown. Hence, the experimental exploration of the metabolome and the design of a metabolic model allowing to verify the reconstructed metabolic network and to assess the energy balancing and the presumably

related slow growth in batch culture, occurred to be a key step towards understanding of this minimal bacteria.

In a joint attempt of experimental and computational research we explored and characterized the metabolism of *M. pneumoniae*. On one hand, we identified and quantified all cellular components assessable through NMR, GC-MS, and LC-MS and studied the central carbon flux using ^{13}C-glucose tracer experiments. We accomplished a comparative analysis of absolute and relative metabolite abundances amongst different species. Our results indicate that metabolic pathways are regulated as functional units, thus allowing the simplification of adaptive responses. On the other hand, we developed a predictive genome-scale constraint-based model of the metabolic network of *M. pneumoniae*: *i*JW145. We defined the biomass composition of an average *M. pneumoniae* cell based on quantitative experimental data to allow the performance of growth simulations. Based on an iterative process of model predictions, their evaluation with experimental findings and consequential refinement of the model, we were able to correct the annotation of the metabolic network and, subsequently, also the functional annotation of key metabolic enzymes. To validate our final reconstruction, we qualitatively predicted the metabolic capabilities of *M. pneumoniae* when grown on alternative carbon source *in silico*. In addition, we conducted an *in silico* knock-out study, prediction gene essentiality with high accuracy (96%) and specificity (98%), thus proving the predictive capacity of the model. The resulting mutant phenotypes have been analyzed providing insight into pathway regulation and adaptive capacities of *M. pneumoniae*.

We applied the validated model to first, predict double mutant phenotypes, which so far are difficult to analyze experimentally. Second, we quantitatively dissected the *M. pneumoniae* energy metabolism, showing that *M. pneumoniae*, in contrast to other bacteria, at least under laboratory conditions uses most of its energy for cellular maintenance and not growth. Finally, applying simple mathematical fittings to *in vivo* metabolite concentration data and maintenance costs determined *in silico*, we calculated constraint sets for different points of growth. This enabled us to analyze time-dependent changes in the metabolic behavior of *M. pneumoniae* without necessity to determine the exact underlying kinetic parameters. Furthermore, integrating metabolite abundances and carbon flux data with quantitative proteomics data, we were able to calculate *in vivo* catalytic parameters for several glycolytic enzymes.

3.2. Material and Methods

3.2.1. Computational Procedures

Metabolic Reconstruction

We used the reconstruction and modeling platform ToBiN (**To**olbox for **Bi**ological **N**etworks, the source code is available at http://github.com/miguelgodinho/tobin). The initial reconstruction was based on the reaction network published by Yus et al. [2009]. To keep atoms and charges in the model balanced and to cope with reactions that cannot be represented directly in stoichiometric models, for example DNA and RNA elongation

reactions, some changes had to be introduced (see below). To allow the simulation of compound exchange with the environment, so-called source and sink reactions have been defined, respectively, for all metabolites known to be taken up and/or secreted by *M. pneumoniae* (Appendix A, Table A.1). Reaction reversibilities and minimum and maximum constraints were defined based on experimental data and literature (Appendix A, Table A.2).
The applied FBA solver uses the simplex algorithm, a numeric optimization procedure that after a finite number of steps ascertains either an exact solution or the infeasibility of the problem. For the model visualization we used CellDesigner 4.1 [Kitano et al., 2005] and the CellPublisher [Flórez et al., 2010] for the clickable online version included in the **MyMpn** database (Figure A.1, http://mycoplasma.crg.es/pathways.php). All abbreviations used on the model map can be found in the List of Abbreviations at the beginning of this thesis.

Curve Fittings

Metabolite Fittings: We used KaleidaGraph 4.0 to fit sigmoidal curves to the glucose consumption, acetic acid production, lactic acid production and protein synthesis determined *in vivo* (Appendix A, Figure A.2, and Figure 3.13). The general equation for sigmoidal curves reads:

$$f(x) = a + \frac{b - a}{1 + c^{(d - x)}} \tag{3.1}$$

For glucose consumption this equation has been adapted to represent the inverted curve progression:

$$f(x) = a + \frac{b - a}{1 + c^{(d - (180 - x))}} \tag{3.2}$$

To describe the progression of maintenance costs we fitted a logarithmic function to the manually determined maintenance costs at different times of growth:

$$f(x) = a \cdot x^3 + b \cdot x^2 + c \cdot x + d \tag{3.3}$$

The variable values for each fitted curve can be found in Appendix A, Table A.8. The resulting curves allow to calculate constraints for $0 \leq x \leq 180$ with x being the growth time of a four days batch culture in hours. Still, we recommend to only use the presented model with biomass production as single objective function for $24 \leq x \leq 60$ since this is the determined exponential growth phase (Figure reffig:fittings).

Fittings for Heavy Isotope Labeling of Metabolites: We used Prism5 to fit one-phase and two-phase exponential decay function to data on heavy isotope labeling in glycolytic intermediates.The one-phase exponential decay function is defined:

$$f(x) = span * e^{(-K * x)} + plateau \tag{3.4}$$

The two-phase exponential decay function reads:

$$f(x) = span_1 * e^{(-K_1 * x)} + span_2 * e^{(-K_2 * x)} + plateau \tag{3.5}$$

with $span_1 = (Y_0 - plateau) * part_1$, $span_2 = (Y_0 - plateau) * (1 - part_1)$, and $part_1 \in [0,1]$, defining the first exponential phase as fraction of the time until reaching *plateau*.The parameters for fitting the curves to the glycolytic intermediates can be found in Appendix A, Tables A.16 & A.17.

Definition of Flux Constraints

For the initial simulations, used to refine the network structure of the model, we calculated maximum constraints for glucose consumption and acetic production based on the concentration changes measured for glucose, acetic acid, and lactic acid in the growth medium (Appendix A, Figure A.2). To this end, the glucose uptake and the lactic and acetic acid synthesis rates for one cell have been calculated from the population measurements by the parallel determination of protein increase in the population. According to Yus et al. [2009], one *M. pneumoniae* cell contains 10 fg of protein which allows the determination of the described rates by converting the units from mmol*ml^{-1} of medium to molecules per cell. For instance, when assuming that 1.339969489 mM of glucose have been depleted from the medium by a *M. pneumoniae* colony on average comprised of 1878886249.46534 cells during six hours, than one cell consumed 19882.97753 molecules of glucose per second. By division of this glucose consumption rate by the Avogadro constant (6.022E+23) and multiplication with 3600 (for the hour) and 1000 (to get mmole, not mole), the uptake of glucose per cell and hour amounts to 1.19E-13 mmole*h^{-1}. When taking into account the assumed total cell mass of 16.13 fg (Results, section 3.3.4), one gram of *M. pneumoniae* comprises 6.20E+13 cells, resulting in a final model constraint of 7.369004196 mmole*g^{-1}*h^{-1}. Those initial constraints are not shown, since for later simulations refined constraints have been calculated for glucose availability, and acetic acid production as described below.

We fitted sigmoidal curves (Equations 3.1 & 3.2) to the *in vivo* concentration courses of external metabolites (Figure 3.13B-D). In addition, after manually fitting the minimum constraint for cellular maintenance costs based on the integration of *in vivo* doubling times, we fitted a logarithmic function (Equation 3.3) to the constraints determined for different time points (Figure 3.13F). Based on the fitted functions, we calculated the ratio of lactic acid to acidic acid and, subsequently, the maximum constraints for glucose uptake and acetic acid production, as well as the minimum constraint for cellular maintenance costs (Appendix A, Table A.9). Glycerol and G3P were limited to 2.5% of the respective glucose constraint based on the minimal medium composition in [Yus et al., 2009]. Ribose was limited to 2.5% of the respective glucose constraint, too, in order to account for ribosylated bases in rich medium. All other sugar sources were silenced (set to zero), since it is known that bacteria generally use up one sugar source before switching to another [Monod, 1966] and proteins involved in uptake and processing of alternative sugars have been shown to be low abundant or not detectable [Maier et al., 2011].

Arginine availability was constrained to 0.25 $mmol^{*}g^{-1*}h^{-1}$ (value chosen arbitrarily due to lack of experimental data) in order to prevent unlimited energy production from arginine. mRNA and protein turnover have been accounted for by minimum constraints on the respective degradation reactions based on experimentally determine mRNA and protein half-lives [Maier et al., 2011]. To represent known detoxification events, minimum constraints were set on dihydroxyacetone production (spontaneous) and 5-formyl tetrahydrofolate (regulatory function). Based on the facts that the conversion of DHAP into G3P is very slow (Results, section 3.3.7) and that glycerol is essential for growth in minimal medium [Yus et al., 2009] we constrained the conversion of DHAP into G3P to 0.25 $mmol^{*}g^{-1*}h^{-1}$ (value chosen arbitrarily) in order to allow the model to reproduce the experimental data. The ATPase reaction and two reactions describing the activity of the chaperones DnaK/DnaJ/GrpE and GroEL/GroES have been included in the model for completeness but have not been constrained, since there was no information available about their exact contributions to energy consumption. Those and other not yet quantified energy sinks are accounted for by the maintenance energy. The constraint sets for different times of batch culture growth in rich medium and for minimal medium can be found in Appendix A, Table A.9. For the simulation of growth on alternative sugars, glucose has been limited to zero and the respective alternative sugar source was limited to provide the same amount of carbon as when using the constraint for glucose for 36 hours of batch culture growth.

Growth Simulations

Growth of an organism is defined as the production, leading in final consequence to the duplication, of cellular material by the mother cell, that is then sub-divided into separate daughter cells. Therefore, to simulate growth *in silico* with a constraint-based metabolic model, the maximization of biomass production, i.e. the synthesis of all cellular building blocks as required for the biomass composition, is chosen as objective function for the FBA problem. The resulting objective value *ov* gives information about the doubling time t_{doub} of an average *M. pneumoniae* cell as described by Equation 1.6. Thus, it is possible to distinguish between growth (*ov* larger than zero), catabolic activity (growth arrest) (*ov* equal to zero), and cell death (infeasibility of the FBA). The FBA is considered infeasible if at least one of the minimum requirements specified cannot be satisfied under the given nutrient conditions, both defined by the respective constraints.

Gene Essentiality Prediction

The gene-protein-relationship has been determined for all reactions for which the catalyzing enzyme is known. In each *in silico* knock-out all reactions catalyzed by the corresponding gene product have been limited to zero flux. A gene is considered essential when its knock-out leads to an objective value of zero (no growth but minimum constraints can be matched) or the infeasibility of the FBA (minimum constraints are not fulfilled). Genes coding for proteins that catalyze DNA degradation, protein folding,

and the ATPase reaction have been excluded from the essentiality prediction since their corresponding functions have not been modeled explicitly. All simulations of this section have been performed using rich medium conditions for 36 hours growth time (Appendix A, Table A.9). Maintenance expenses have been neglected, to prevent that knock-outs leading to significant slower energy production can result in infeasibility of the FBA. Subsequently, the obtained objective values give no information about the absolute doubling times but only about the relative changes in the growth rate between wild type and knock-out simulation.

For the prediction of double mutant phenotypes, we applied the same strategy as for the single *in silico* gene knock-outs, but simultaneously silenced the reactions catalyzed by two different non-essential proteins at a time. Double knock-outs resulting in reduced fitness, i.e. the objective value is smaller than for each of the two single knock-outs alone, or in cell death, i.e. the objective value equals zero or the FBA is infeasible, were considered for the analysis of synthetic lethal and sick interactions (Figure 3.11).

For the statistical analysis of accuracy and specificity of the gene essentiality prediction, we evaluated the prediction results based on a genome-wide transposon study in *M. genitalium* [Glass et al., 2006], transposon screens in *M. pneumoniae* (this work) and the simulation conditions. Computationally and experimentally essential genes are considered true positives, true negatives are computationally and experimentally not essential, computationally essential and experimentally non-essential genes are defined as false positives and computationally non-essential and experimentally essential genes accordingly false negative hits.

Comparison of Qualitative Changes in Fluxes and Protein Abundances

First, linear fittings to the *in silico* reaction fluxes obtained at t=24h, 36h, 48h, 60h and to protein abundances measured at t=24h, 36h, 48h, 72h during batch culture growth experiments *in vivo* [Maier et al., 2011], have been conducted. Second, we determined the qualitative overall change of fluxes and protein abundances during the exponential growth phase, considering proteins to change only if the measured abundance difference exceeds 25% of the abundance at t=24h, thus accounting for the reported experimental error that would otherwise have a high impact especially on the changes of low-abundant proteins [Maier et al., 2011]. Finally, we aligned protein concentration changes with the change of the sum of fluxes of reactions catalyzed by the respective protein (Figure 3.14B).

Sequence Comparison

All sequence analyses have been performed using the Basic Local Alignment Search Tool for proteins (pBLAST, algorithm pblast) [Altschul et al., 1997]. pBLAST was used, since *M. pneumoniae* uses the TGA codon to encode for tryptophan instead of indicating the end of a gene as in most other organisms. Protein sequences of related organisms (ordered for preference: other mycoplasmas, *B. subtilis*, *L. lactis*, *E. coli*) were obtained from KEGG [Kanehisa and Goto, 2000] or the National Center of Biotechnology Information

(NCBI) [Tatusova et al., 1999] and used to perform pBLAST against the *M. pneumoniae* proteome. Alternatively, *M. pneumoniae* protein sequences were aligned to the nr-DB in order to detect possible homologies. This has been done so i) to search for enzymes possibly involved in fumarate and succinate processing, ii) to identify the cofactors used by the GPO (MPN051), iii) to shed light on the NOX isoform (MPN394), iv) to confirm that a reaction converting UTP into CTP does not exist in *M. pneumoniae*, and v) to search for proteins possibly catalyzing phospholipid production. All pBLAST results are shown in Appendix A, section A.1.

3.2.2. Experimental Procedures

Sample Preparation

M. pneumoniae M129 cells were grown in batch culture in suitable culture flasks. Generally, cells were grown as pre-culture for 96 hours, harvested and diluted into fresh growth medium and seeded into new culture flasks for experiments. Cells were grown for different time intervals, ranging from 24 to 96 hours. At indicated times, the growth medium was discarded and the cells were washed twice with ice cold PBS containing 0.05% glucose. After complete removal of the wash buffer, the culture flask was placed on a bed of dry ice and -80°C methanol was rapidly added for both quenching metabolism and lysing the cells. After cell scraping and collecting the sample, cell debris was spun down and the supernatant containing cellular metabolites were transferred to pre-cooled glass tubes containing internal standard as indicated below. Samples were immediately frozen in liquid nitrogen and lyophilized to dryness for 24h-72h.

Protein Concentration and Enzyme Assays

Protein content was determined using the commercially available BCA kit (Thermo), essentially following the manufacturers recommendations and as described by Yus et al. [2009]. The determination of extracellular glucose, lactic acid, acetic acid and ethanol was carried out using commercially available kits (BioVision #K606 and #K607, Megazyme K-ACETRM, K-ETOH) as described by Yus et al. [2009].

GC-MS Analysis

Different groups of compounds (free bases, free amino acids, fatty acids, and glycolysis products) were targeted specifically using tailored protocols as described in Maier et al., *under revision at MSB*. Depending on the case, growth medium, total cell content, cell pellet, or cytoplasm was analyzed as described in each protocol.

NMR Measurements

Dried extracts were redissolved, centrifuged, and 600 μl supernatant were transferred into NMR tubes as described in Maier et al., *under revision at MSB*. Spectra were acquired with an Avance 800 MHz NMR Spectrometer with triple resonance CryoProbe (Bruker

Biospin, Coventry, UK) following a procedure described by [Beckonert et al., 2007] and standard pulse sequences for 1D 1H-NOESY, 2D 1H-1H COSY, 1H-13C HSQC experiments. Metabolites were identified by comparing spectra from standard compounds and spectra available from online repositories (HMBD [Wishart et al., 2009] and BMRB http://bmrb.cerm.unifi.it).

LC-MS Measurements

Dried extracts were redissolved, centrifuged and supernatants were transferred to HPLC glass-vials as described in Maier et al., *under revision at MSB*. Samples were analyzed by UPLC-MS with a HILIC mode and a reversed phase mode separation [Spagou et al., 2011, Want et al., 2010]. In addition, samples were subjected to an ion-pairing mode HPLC-MS method [Liebeke et al., 2010] for the analysis of very polar metabolites like triphosphate nucleotides. LC-MS data was evaluated for predicted *M. pneumonia* metabolites (Appendix A, Table A.3), including possible ions for common adducts in ESI mass spectrometry (e.g. $[M + H]^+$, $[M + NH4]^+$, $[M + Na]^+$, or $[M + K]^+$ and in negative ion mode $[M\text{-}H]^-$) [Tong et al., 1999]. As LC-MS profiling data often contains a huge amount of uninformative "noise" [Jankevics et al., 2012] we applied a filtering to our peak list, considering only peaks above a specified abundance threshold and with ions not on the predicted metabolite list for further identification by database searches in Metlin, massbank.jp [Smith et al., 2005, Horai et al., 2010].

Transposon Screens

The 64 pools of an ordered collection of *M. pneumoniae* transposon mutants generated by "haystack mutagenesis" [Halbedel et al., 2006] were assorted into 10 groups. Then, genomic DNA extractions were performed using Illustrabacteria genomic KIT (GE). The disruptive insertions in genes *mpn133*, *mpn321*, *mpn392*, *mpn533* and *mpn595* were detected by PCR (Figure 3.10). Fragments corresponding to junctions between genes and the mini-transposon were amplified using the primer 3JpMT85 and the primers 5MPN133, 5MPN321, 5MPN392, 5MPN533 and 5MPN595, respectively (Appendix A, Table A.12). The position of the transposon insertion in the different genes was determined by DNA sequencing.

pH Experiment

To check the influence of the medium pH on growth performance, *M. pneumoniae* cells were grown in batch culture in 75 cm^2 culture flasks. Cells were grown in pre-culture for 96 hours in glucose containing medium, harvested by scraping and diluted into fresh growth medium. Medium pH was adjusted back to pH7.7 after four days of growth by titration with sterile 1 M NaOH. Samples from growth medium supplemented with 1% glucose (55.5 mM) were taken at indicated time points (Appendix A, Figure A.9). Glucose and lactic acid concentrations were determined with enzymatic assays as described (section 3.2.2).

3.3. Results

Applying the systems biology approach, we explored and characterized the metabolic network of *M. pneumoniae*. We build a predictive constraint-based metabolic model to validate the wiring diagram, to explore the energy metabolism, and to predict metabolic phenotypes for single and double mutants: *i*JW145 (Appendix A, Table A.1). The model building process can be divided into three main steps: i) model construction, ii) biomass definition and assignment of reversibilities, and iii) model refinement. It has been shown that automatic metabolic reconstructions are likely error-prone due to the one-dimensional annotation they are based on [Reed and Palsson, 2003]. Hence, we iteratively integrated different experimental data already during the model building process. To extract biologically relevant information, we accomplished different analyses consecutive to the integration of experimental data and during the model validation process. In parallel, we monitored cellular and extracellular metabolites *in vivo*, applying a combination of different technical approaches complementing each other. NMR, GC-MS, and LC-MS have been used to identify and in selected cases quantify metabolites, the concentration changes in external metabolites have been measured with metabolite assays, and the central carbon flux has been examined using ^{13}C-glucose tracer experiments. Model predictions and experimental results have been integrated with each other repeatedly, thus beneficially effecting experimental and computational design processes as well as the biological knowledge gain (Figure 3.1).

3.3.1. Model Construction

Using the modeling platform ToBiN we built a genome-scale constraint-based metabolic reconstruction of the minimal bacterium *M. pneumoniae* (*i*JW145_reconstruct; Appendix A, Figure A.1 and Table A.1) based on a curated wiring diagram for metabolism [Yus et al., 2009]. We adjoined transport reactions, which were not defined in the reaction list from Yus et al. [2009], for components that are exchanged with the environment and degradation reactions for proteins and RNA. Finally, we added source and sink reactions in order to account for the system boundaries, i.e. to enable the simulation of the exchange of metabolic compounds with the environment under steady state conditions.

To properly assign reaction reversibilities, we integrated consensus information obtained from the BRENDA enzyme database [Scheer et al., 2011] with reaction directions published for the metabolic network of *E. coli* [Fleming et al., 2009] and the metabolic map for *M. pneumoniae* [Yus et al., 2009] (Appendix A, Table A.2). Thus, we could define reversibilities of 65.1% of all model reactions. 8.5% of the reactions, for which contrary information was available, were set irreversible, among them all tRNA biosynthesis reactions and six reactions of the lipid metabolism. The reversibilities for transport reactions (12.7% of all reactions) and model specific reactions (source/sink, 21.8% of all reactions) have been set as required for the production of all metabolic components and to reproduce experimental findings. The NADH oxidase reaction (M017) has been changed to irreversible along with the functional re-annotation of the NADH oxidase (MPN394, see below). In case of new experimental evidence on reaction reversibilities,

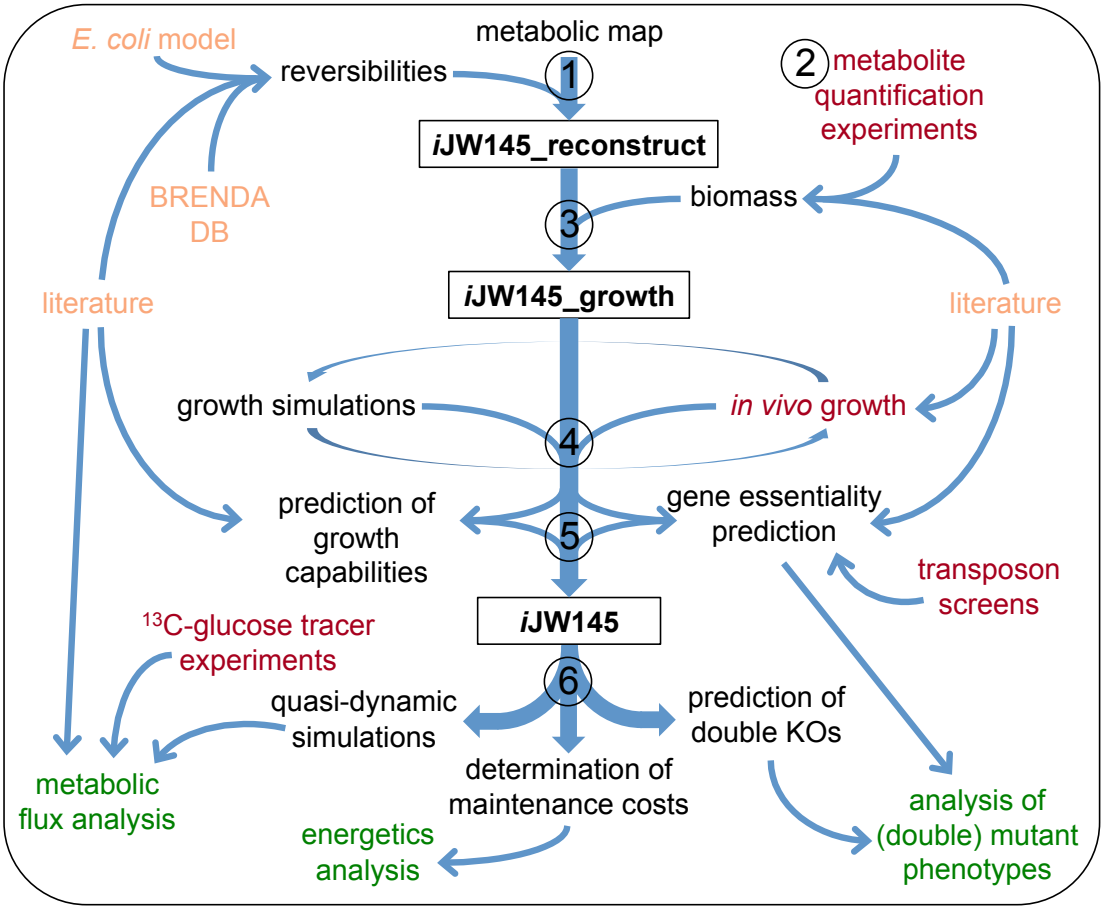

Figure 3.1.: Workflow describing the construction process of *i*JW145 and characterization of the *M. pneumoniae* metabolism: 1. Reconstruction of the metabolic network based on the metabolic map from [Yus et al., 2009]; 2. *In vivo* identification and quantification of the *M. pneumoniae* metabolome; 3. Definition of the biomass composition and integration of reaction reversibilities; 4. Model refinement based on an iterative cycle of *in silico* growth simulation and their comparison to *in vivo* findings and literature data; 5. Model evaluation based on the prediction of growth capabilities for different carbon sources and an *in silico* knock-out study to predict gene essentiality; 6. Model application for the prediction of i) double mutant phenotypes (KOs - knock-outs), ii) energy balancing for *M. pneumoniae*, and iii) the prediction of metabolic flux distributions that upon integration with literature and *in vivo* fluxes allowed to characterize the metabolic behavior of *M. pneumoniae* in batch culture growth. Model inputs are shown in red (this study) or orange (literature, databases), the different *in silico* analyses accomplished and the resulting model states are shown in black, and model outputs are shown in green.

the model can be adjusted easily.

To prove the completeness and the correct connectivity of the reconstructed metabolic network, we used FBA to individually maximize the production of each network component. Contrary to more complex organisms, the metabolic network of *M. pneumoniae* is mainly composed of linear pathway modules which are interconnected by only very few metabolites apart from ubiquitous cofactors, such as AMP, ADP, ATP, H^+, H2O, NAD^+, NADH, Pi, PPi (Appendix A, Figure A.1 and Table A.3). These results and the lack of rescue pathways known from other organisms, amongst them the Entner-Doudoroff-Pathway to bypass glycolysis, facilitate the analysis of inter-pathway crosstalk and limit the existence of multiple optima for FBA problems. Furthermore, the absence of most catabolic and anabolic routes in *M. pneumoniae* allows to relate external metabolite measurements directly to intracellular fluxes.

3.3.2. Metabolite Identification

Out of 216 reactants included in the *in silico* metabolome of *M. pneumoniae* we extracted 150 that can be verified experimentally, excluding tRNAs, protein-based reactants, and inorganic compounds not amenable to experimental analysis (Appendix A, Table A.3). Combining GC-MS, LC-MS, and NMR, we identified 86 different cellular metabolites in an organism-wide screen (Figure 3.2A and Appendix A, Table A.4). Those metabolites confirm 53% of the predicted metabolites (Figure 3.2B), surpassing experimental coverage of predicted metabolites in other prokaryotes [Soga et al., 2003, van der Werf et al., 2007, 2008, t'Kindt et al., 2010, Liebeke et al., 2011]. 83% of the metabolites experimentally detected in this study were identified in large-scale metabolic screens of other bacteria as well [Soga et al., 2003, van der Werf et al., 2007, Liebeke et al., 2011] (Figure 3.2C).

Experimentally determined metabolites map to all pathways of the metabolic reconstruction (Figure 3.3). High coverage was achieved for intermediates of the central carbon metabolism, the nucleotide metabolism, and the amino acid metabolism. Metabolites associated to alternative sugar metabolism and pentose phosphate pathway are observed less frequently. This finding agrees with the low abundances determined for the respective catalyzing proteins [Maier et al., 2011].

73 predicted metabolites were not detected. 49 of those can technically be identified using commercially available pure compounds as standards [Soga et al., 2003, van der Werf et al., 2007, Liebeke et al., 2011]. However, only 16 of them have been experimentally confirmed previously in *E. coli*, *B. subtilis* or *S. aureus* [Soga et al., 2003, van der Werf et al., 2007, Liebeke et al., 2011] suggesting general identification problems for all others. Due to the small size of *M. pneumoniae*, metabolites that are unstable, rapidly turned over, or of very low cellular abundance are likely to be present below the detection limit of the applied techniques. Furthermore, the 16 previously identified compounds are mainly intermediates of the pentose phosphate pathway and the CoA metabolism, two pathways that are supposed to have high activity only, if the associated cellular building blocks cannot be imported from the environment, which does not apply for the rich medium used during batch culture growth.

In addition to the metabolites contained in the metabolic model, we identified seven metabolites not predicted from the *in silico* metabolome, namely ADP-glucose, cholesterol, cytosine, ethanol, fumarate, succinate, and trans-4-hydroxyproline. ADP-glucose,

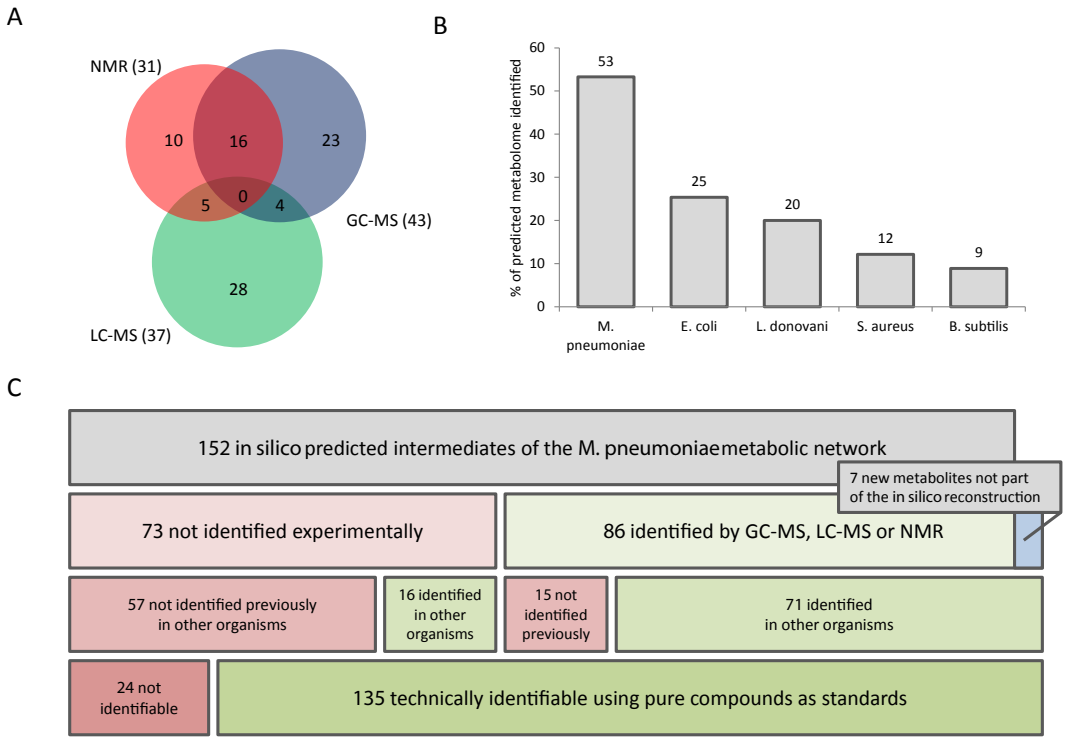

Figure 3.2.: Metabolomics in *M. pneumoniae*. A: Specificity and overlap of different technical approaches applied to detect *M. pneumoniae* metabolites. B: Integration of *in silico* predicted and experimentally confirmed metabolites in different organisms. C: Graphical representation of predicted, identified, and technically identifiable metabolites in *M. pneumoniae*.

cholesterol, and cytidine, despite being detected experimentally, are not included in the metabolic reconstruction for technical reasons. ADP-glucose is functionally redundant with UDP-glucose, which is included in the reconstruction and which also was identified experimentally. *M. pneumoniae* is not able to synthesize sterols, amongst others cholesterol, but directly imports them from the environment [Yus et al., 2009] and incorporates them into the plasma membrane [Johnson and Somerson, 1980]. Nevertheless, since the lipid composition of mycoplasmas has been shown to vary depending on the fatty acids provided with the growth medium [McElhaney and Tourtellotte, 1969, Pollack et al., 1970, Rottem, 1980], cholesterol and other sterols are not included explicitly in the model but merged into general artificial molecules (for details please refer to section 3.3.4). Experiments in defined growth medium showed that cytidine rather than

cytosine is directly imported by *M. pneumoniae* [Yus et al., 2009], suggesting that cytosine could eventually be produced from cytidine but not *vice versa*. Since no enzyme has been shown to catalyze this reaction and it cannot be excluded that cytosine arises due to technical challenges during sample preparation (see section 3.3.3), cytosine forms not part of the metabolic reconstruction so far.

Ethanol is a common end product of organic acid fermentation and for *M. pneumoniae* production from pyruvate via acetaldehyde has been suggested [Weiner et al., 2003]. However, neither acetaldehyde (this study) nor the enzyme catalyzing ethanol production (acetaldehyde dehydrogenase/alcohol dehydrogenase (ADH, MPN564) could be detected by mass spectrometry [Maier et al., 2011]. Furthermore, no ethanol secretion was observed *in vivo* (Appendix A, Figure A.2) [Yus et al., 2009], although it was detected in trace amounts in the growth medium using NMR.

Fumarate and succinate were identified in trace amounts by NMR and succinate but not fumarate was also identified as minor component of the growth medium. Using pBLAST alignments of *B. subtilis* enzymes involved in fumarate and succinate metabolism, we could not identify any *M. pneumoniae* enzyme showing significant sequence similarity (Appendix A, section A.1.1).
Post-translational modifications of proline residues lead to the synthesis of trans-4-hydroxyproline in eukaryotic cells. Bacteria, amongst them mycoplasmas, do not contain the enzyme required to modify proline accordingly. Nevertheless, *E. coli* is able to import hydrozyproline and incorporate it into proteins under suitable growth conditions [Buechter et al., 2003]. We confirmed the presence of trans-4-hydroxyproline in the growth medium using GC-MS, but it remains questionable if *M. pneumoniae* is able to incorporate this modified amino acid into proteins.

In summary, we achieved unprecedented coverage of the *in silico* predicted metabolic component space combining three complementary experimental approaches. The integration of the experimental findings with literature data and the metabolic reconstruction verifies accuracy and completeness of the latter.

3.3.3. Metabolite Quantification

The design of a stoichiometric model for metabolism provides the basis for understanding metabolic behavior of *M. pneumoniae*. However, since metabolism by definition is responsible for nutrient uptake and their subsequent processing into energy and cell building blocks, it is of upmost importance to not only identify but quantify such building blocks when aiming to reproduce metabolic behavior and to predict realistic metabolic flux distributions. To further characterize *M. pneumoniae* and better understand its cellular composition, we quantified metabolic key compounds, such as nucleobases, amino acids, and fatty acids. The processes leading to the quantification of those metabolites are outlined below.

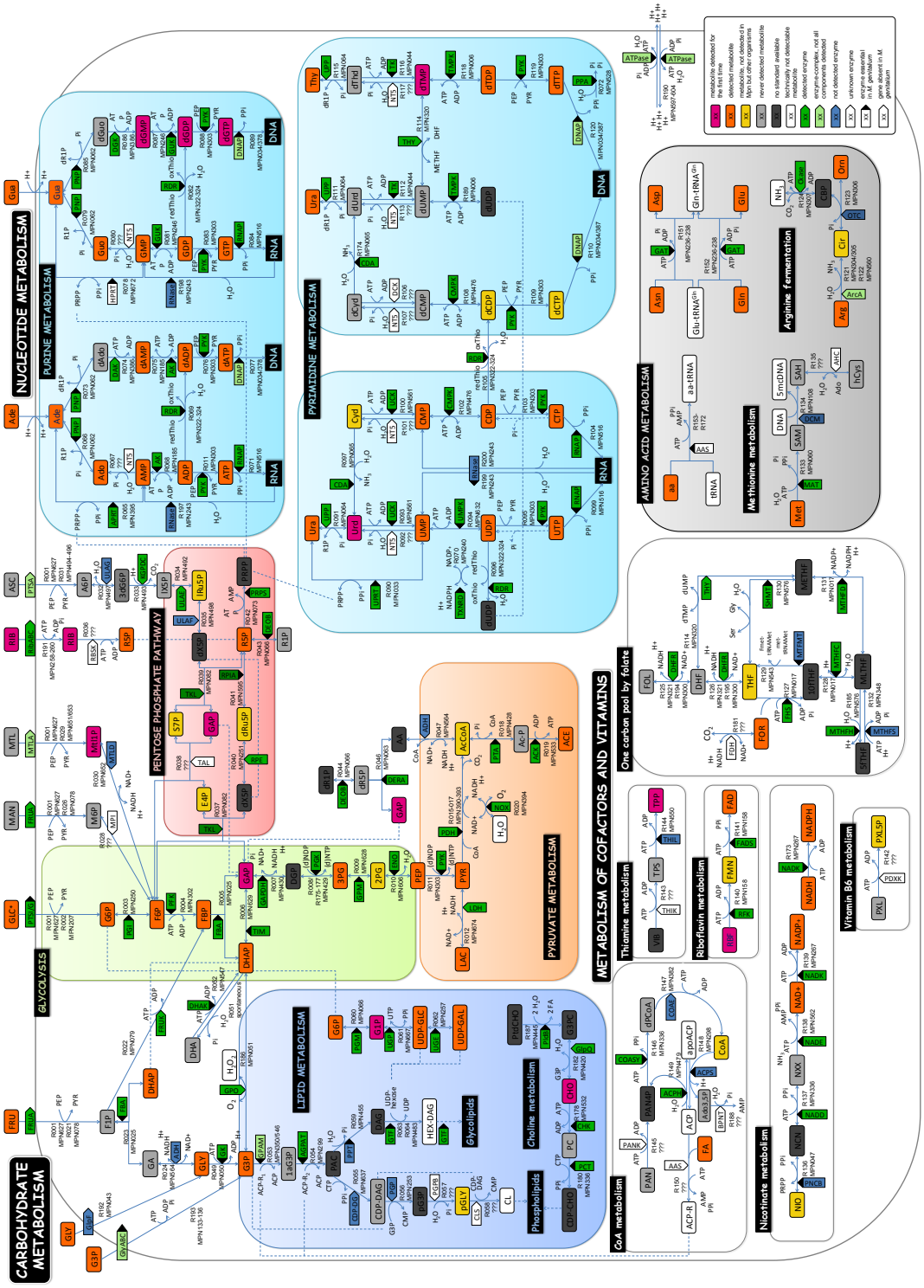

Figure 3.3.: The metabolic core network of *M. pneumoniae*: Metabolites and catalyzing enzymes have been colored according to results from genome-wide screens for metabolites (this work) and proteins [Maier et al., 2011] in *M. pneumoniae*. The color code is shown in the lower right corner of the figure, for information about the moonlighting enzymes please refer to Yus et al. [2009]. A larger version of this figure is provided for take out at the very end of this thesis.

Nucleobases

Nucleobases can be divided into two groups, purine and pyrimidine nucleobases. Adenine and guanine belong to the purine nucleobases, while cytosine, thymine, and uracil are pyrimidine nucleobases. Nucleobases are processed into nucleotides which i) are incorporated to different extent into DNA and RNA, ii) function as cellular energy carrier, mainly in form of ATP, iii) are incorporated into reaction cofactors, such as NAD^+/NADH or coenzyme A, and iv) have functions in signal transduction [Alberts et al., 2008].

We measured nucleobases as well as nucleosides in the *M. pneumoniae* cytosol and in the growth medium from samples taken at regular intervals during a four days growth experiment (Figure 3.4). Methodological constraints in sample preparation prior to GC-MS analysis result in substantial chemical conversion of nucleosides and nucleotides into their cognate bases for most nucleosides. Thereby, conversion rates ranged from about 3% for adenosine to about 50% for thymidine (Appendix A, Figure A.3). To this end, cumulative amounts of determined cytosolic nucleobases and nucleosides (NUBS) are assumed to represent the cellular inventory of free nucleobases, nucleosides, (labile) nucleotides and nucleobase containing coenzymes (Figure 3.5).

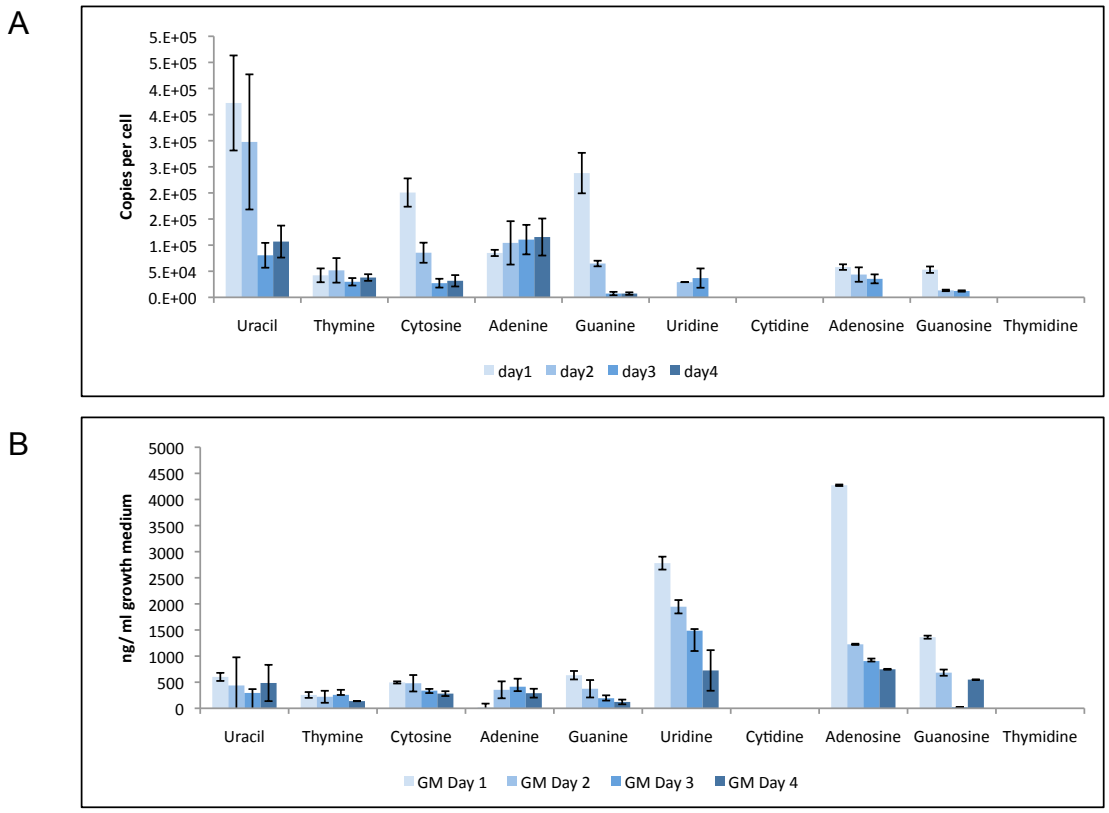

Figure 3.4.: Free bases and nucleosides have been measured along a four days growth experiment. A: in the *M. pneumoniae* cytosol; B: in the growth medium.

Despite their constant depletion from the growth medium along the four days growth course, we found that NUBS are not growth limiting based on their final concentrations after 96 hours of growth (Figure 3.4). Compared to the growth medium, NUBS are found on average 700 times enriched in the cytosol, suggesting a direct and active import by *M. pneumoniae*, although no transport protein has been identified so far. The observed extracellular changes are reflected by intracellular abundance changes for all NUBS but for adenine/adenosine, which was maintained at a constant concentration of 5 mM at all examined times. This finding indicates the existence of a separate mechanism responsible for maintaining the intracellular pools of ADP and ATP at constant levels independent of alternations in the growth conditions. Compared to *E. coli*, the determined intracellular concentrations in *M. pneumoniae* are on average three times lower [Bennett et al., 2009].

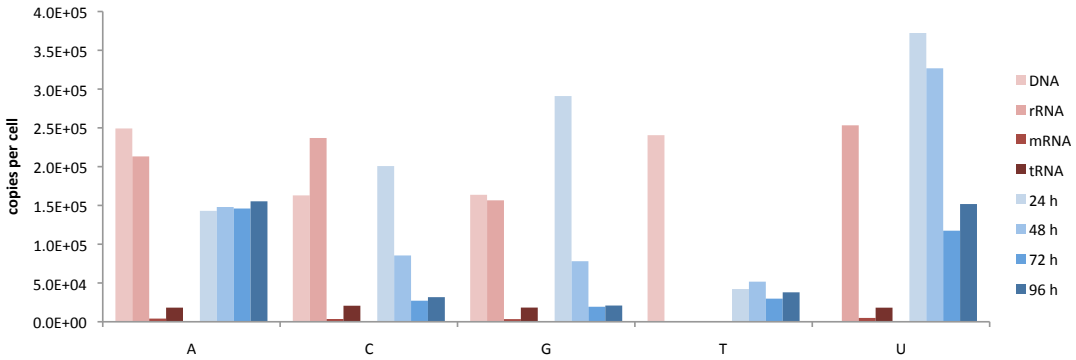

Figure 3.5.: NUBS (free + bound nucleobases) homeostasis: Bound nucleobases (red bars) and free NUBS determined for different days of a four days batch culture growth experiment (blue bars) in the *M. pneumoniae* cytosol.

To further analyze homeostasis of cellular nucleobase pools and their turnover, we quantified nucleobases bound in DNA, mRNA, and tRNA (Figure 3.5). To this end, we integrated available data on the genome sequence and ribosome, mRNA, and tRNA abundances [Yus et al., 2009, Maier et al., 2011]. We found that DNA and tRNA are the mayor cellular nucleobase sinks and that the amount of bound molecules exceeded the freely available NUBS for all nucleobases except uracil. We conclude, that the large intracellular pools of nucleobases determined in *M. pneumoniae* reflect their high functional diversity and the importance of especially adenine derivatives in living organisms.

Amino Acids

The initial reconstruction of the metabolic network [Yus et al., 2009], genetic evidence [Güell et al., 2009], and the proteome analysis [Maier et al., 2011] showed that *M. pneumoniae* lacks almost all anabolic pathways for amino acid synthesis and modification. Instead, *M. pneumoniae* takes up single amino acids and peptides from the environment [Yus et al., 2009]. We determined amino acid concentrations for the cytosol and for the

growth medium by GC-MS and integrated our results with quantitative information on the proteome composition (Figure 3.6A).

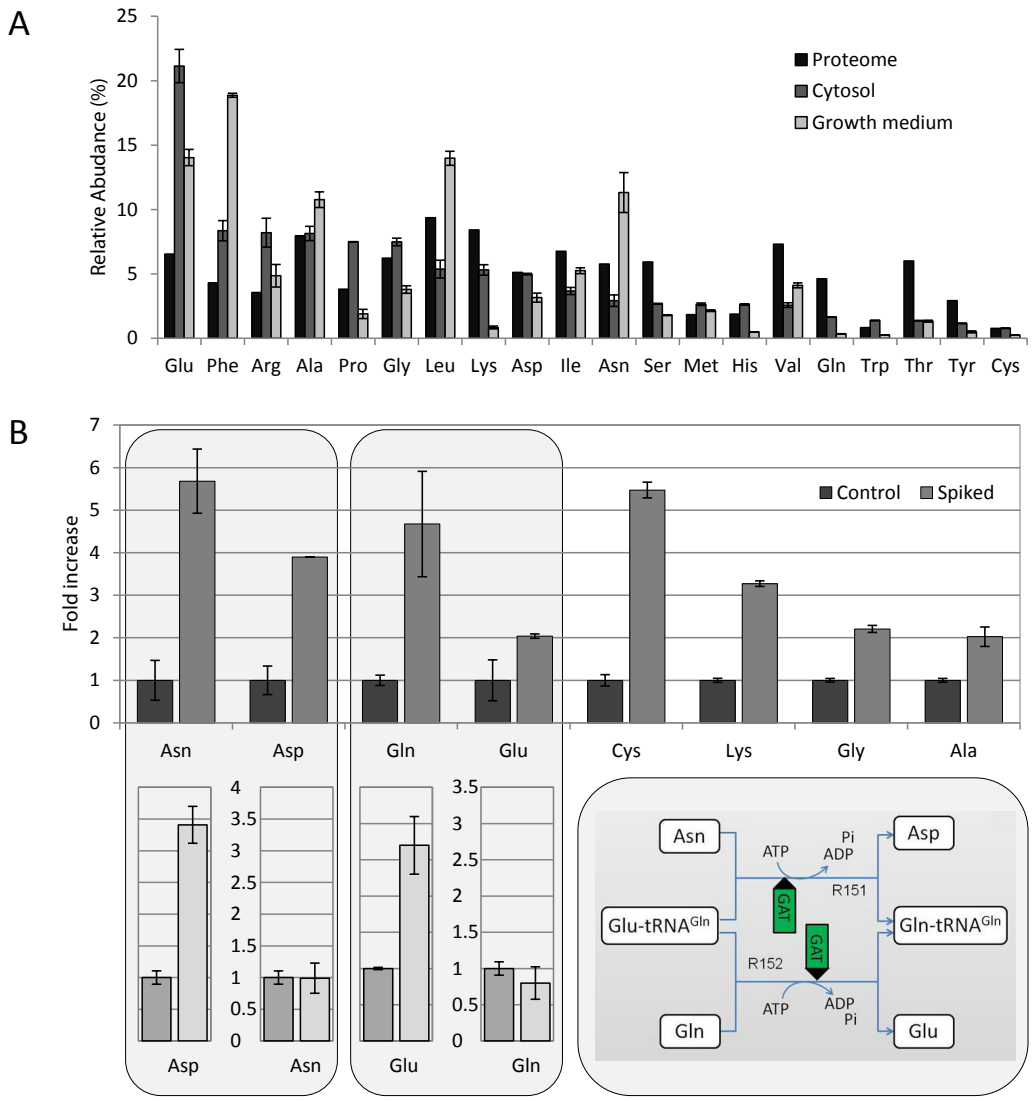

Figure 3.6.: Amino acid quantification. A: Relative amino acid abundances in the proteome (black), the cytosol (dark grey), and the growth medium (light grey). B: In the upper part, abundance changes of specific amino acids upon artificial enhancing of their concentrations in the growth medium (spiking). In the lower part, abundance changes for the labeled amino acids upon spiking of their respective vertically aligned amino acid in the upper part, proving the irreversibility of reactions M152 and M153.

Cytoplasmic amino acid concentrations range from 0.1 mM for cysteine to 3.3 mM for glutamic acid which accounts for 21.1% of the total amount of free amino acids.

As for nucleobases, we found intracellular amino acid concentrations to be significantly enriched compared to the surrounding growth medium (Appendix A, Figure A.4). Artificially increasing the concentrations of individual amino acids in the growth medium led to a concomitant increase of the respective intracellular concentrations (Figure 3.6B). Nevertheless, the amino acid pools of the growth medium are only partially reflected by cytosolic pools (Pearson correlation coefficient of relative amino acid abundances $r_p = 0.16$) and the quantitative distribution in the proteome ($r_p = 0.33$) [Maier et al., 2011] (Figure 3.6A). These findings suggest that *M. pneumoniae* actively imports amino acids from the growth medium but with different efficiencies.

Interestingly, an artificial increase of extracellular concentrations of asparagine and glutamine levels, resulted in an additional increase of intracellular concentrations of aspartate and glutamate, respectively, while inversely this effect was not observed (Figure 3.6B). The metabolic network of *M. pneumoniae* contains two reactions (M152 and M153) in which misacetylated Glu-tRNA(gln) is converted into Gln-tRNA(gln), thereby consuming ATP and glutamine or asparagine and releasing glutamate or aspartate, respectively. We thus confirm that the glutamyl-tRNA-amidotransferase (MPN236-MPN238), compensates for the lack of a Gln-tRNA-synthetase in *M. pneumoniae*, as previously shown also for other gram-positive bacteria [Curnow et al., 1997].

Fatty Acids

It has been shown that the membrane composition in mycoplasmas varies depending on the fatty acid composition of the growth medium [McElhaney and Tourtellotte, 1969, Pollack et al., 1970, Rottem, 1980]. To obtain insight into the uptake of fatty acids from the medium and their incorporation into the cell membrane, we quantified the fatty acids contained in the growth medium, present in the cytosol, and incorporated into membrane lipids (Figure 3.7A). While in the growth medium 80% of all fatty acids have C_{18} chains, in the cytosol as well as in the cellular membranes fatty acids with C_{16} chains (41% and 49% of all fatty acids, respectively) are dominant. In general, we found similar fatty acid profiles for the cytosol and the membranes, suggesting an actively regulated import of fatty acids but unregulated incorporation into membrane lipids.

Since no significant depletion in the growth medium could be measured between beginning and end of the growth course, we conclude that fatty acids are not growth limiting in *M. pneumoniae* (Appendix A, Figure A.5). While at early growth stages fatty acids with longer chains are found commonly accounting for 17% of all fatty acids, towards later growth stages they amount for only 3.5% of the total fatty acid pool (Figure 3.7B). Interestingly, cytoplasmic and membrane incorporated fatty acids are significantly enriched in saturated fatty acids (71% and 70%, respectively) compared to the growth medium (41%). This can be explained by their positive influence on membrane stability and integrity, since incorporation of unsaturated (*cis*) fatty acids disrupt the membrane structure [Lucy, 1972]. For *M. pneumoniae* membrane stability is of upmost importance due to the lack of a protective cell wall.

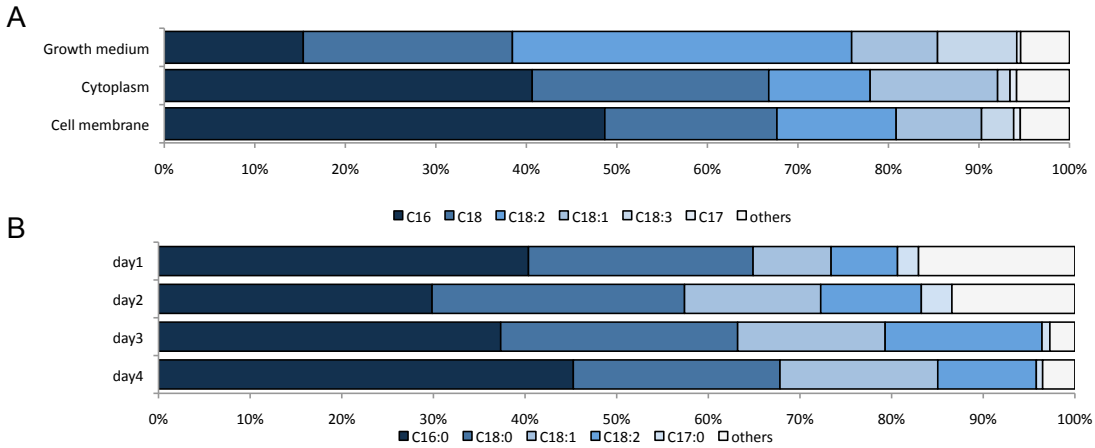

Figure 3.7.: Fatty acid composition analysis: Fatty acid chains of different length have been quantified. A: in the cytosol, the membranes, and the growth medium, B: in the cytosol at regular intervals of a four days batch culture growth experiment.

3.3.4. Defining the Biomass Composition of *M. pneumoniae*

To simulate cell growth and to allow the prediction of realistic metabolic flux distributions, an accurate quantitative representation of the biomass composition of an average *M. pneumoniae* cell, i.e. the macromolecules composing the cell, must be defined (Equation 1.4). Mycoplasmas have been shown to comprise 54-62% protein, 12-20% lipids, 3-8% carbohydrates, 8-17% RNA, and 4-7% DNA [Razin et al., 1963] and *M. pneumoniae* contains 10 fg of protein per cell [Yus et al., 2009]. Assuming that proteins account for 62% of the total cell mass, one *M. pneumoniae* cell has a total weight of 16.13 fg what allowed us to determine the different other biomass fractions (Table 3.1).

To determine a realistic mathematical representation of a *M. pneumoniae* cell, several assumptions and technical tricks necessary to allow *i*JW145 to simulate biomass production and growth have been made:

1. DNA and all RNA types have to be synthesized once de novo which is accomplished by building artificial DNA and RNA molecules of 100 bases length each, displaying the natural GC content of the respective molecules. mRNA turnover is considered by setting a minimal constraint on the degradation reaction (0.0028 $mmol^{*}g^{-1*}h^{-1}$) reflecting the minimum average mRNA half-life of about 1 minute [Maier et al., 2011]. In order to account for DNA degradation, DNA repair and measurement errors in the microarray data [Güell et al., 2009], the amount of DNA and RNA needed for biomass production has been determined to account for 5.2% and 6.5%, respectively, of the total biomass.

2. Protein production is modeled via building of artificial protein molecules of 345 amino acids length (average protein length according to Yus et al. [2009] reflect-

Table 3.1: Biomass Composition

biomass component	biomass fraction in % of total cell mass	biomass fraction in mmol/g of cells	biomass fraction in molecules/ cell	quantity determined by
DNA	5.00	1624	15774.70	sequence
RNA	6.50	2015	19572.67	sequence
protein	62.00	16049	155891.74	sequence
ACP	0.003	3	29.14	Maier et al. [2011]
glycolipid (Mpn)	10.00	63702	618768.49	GC/MS (fatty acid chains)
phosphatidic acid (Mpn)	10.00	148168	1439227.79	GC/MS (fatty acid chains)
glycine	0.07	9220	89558.34	GC/MS
L-alanine	0.09	9824	95425.29	GC/MS
L-arginine	0.07	3913	38008.87	GC/MS
L-asparagine	0.001	54	524.53	GC/MS
L-aspartate	0.12	9318	90510.26	GC/MS
L-cysteine	0.001	67	650.80	GC/MS
L-glutamate	0.27	18651	181166.23	GC/MS
L-glutamine	0.004	294	2855.76	GC/MS
L-histidine	0.04	2422	23526.06	GC/MS
L-isoleucine	0.02	1858	18047.66	GC/MS
L-leucine	0.28	21313	207023.53	GC/MS
L-lysine	0.03	1741	16911.18	GC/MS
L-methionine	0.02	1370	13307.48	GC/MS
L-phenylalanine	0.08	5122	49752.48	GC/MS
L-proline	0.08	6837	66411.10	GC/MS
L-serine	0.03	3202	31102.58	GC/MS
L-threonine	0.03	2489	24176.87	GC/MS
L-tryptophan	0.04	1864	18105.94	GC/MS
L-tyrosine	0.02	1366	13268.62	GC/MS
L-valine	0.03	2793	27129.77	GC/MS
adenosine	0.05	1981	19242.42	GC/MS
cytidine	0.01	503	4885.88	GC/MS
guanosine	0.05	1657	16095.25	GC/MS
thymidine	0.03	1242	12064.15	GC/MS
uridine	0.06	2541	24681.97	GC/MS
				E. coli
orthophosphate (Pi)	0.40	41537	403474.00	Amin and Peterkofsky [1995], Neidhardt [1996]
thiamin diphosphate	0.00	100	971.35	function
NADPH	0.01	100	971.35	function
NADP+	0.01	100	971.35	function
CoA	0.01	100	971.35	function
FAD	0.01	100	971.35	function
5fTHF	0.005	100	971.35	function
pyridoxal phosphate	0.002	100	971.35	function
S-adenosyl-L-met	0.004	100	971.35	function
CDP-CHO	0.0001	100	971.35	function
D-G6P	4.51	174748	1814216.37	100%biomass - rest

Table 3.1.: Biomass composition of an average *M. pneumoniae* cell: For the different biomass components information about their biomass fraction, the respective constraint set (note that this value has been adjusted to 1000000*biomass in order to avoid numerical problems of the solver), the equivalent number of molecules per cell, and the methods used to determine each constraint are provided.

ing the amino acid composition found in the quantified proteome [Maier et al., 2011]. An exception is the acyl-carrier-protein (ACP) that is modeled exactly, i.e. sequence-dependent, and also included in the biomass in the quantity it has been detected on the second day of growth [Maier et al., 2011] as it plays an important role in CoA metabolism. *M. pneumoniae* is known to take up peptides using the Opp transporter and amino acids using ABC systems to import and proton symport to export them [Yus et al., 2009]. Since nothing is known about the specificity and activity of the proteases located on the surface of the *M. pneumoniae* cells and the up-take of peptides of varying amino acid composition is more complicated to simulate, in the model we only consider ABC transporters for the import of single amino acids.

3. The exact lipid composition of *M. pneumoniae* is not known but fatty acids have been quantified based on the length of their carbon chains (section 3.3.3). Because of the varying lipid composition in mycoplasmas depending on the fatty acids provided with the medium [McElhaney and Tourtellotte, 1969, Pollack et al., 1970, Rottem, 1980] we made the following assumptions for lipids in our model:
   - Lipids provide 20% of the total cell mass (assumption based on [Razin et al., 1963]).
   - The "average" lipid has two fatty acid chains, one composed of 16 and the other of 18 carbon atoms. Fatty acids with carbon chains of 16 and 18 carbons length, respectively, have been found to be most abundant in *M. pneumoniae* contributing about 95% of the total fatty acids encountered (Figure 3.7). In addition, the defined medium designed for *M. pneumoniae* [Yus et al., 2009] only contains fatty acids with 16 and 18 carbon atoms length. Cholesterol, despite having been shown to be essential for *M. pneumoniae* growth [Rottem et al., 1971, Johnson and Somerson, 1980, Yus et al., 2009] and found to be abundant in the cytosol (section 3.3.3), has not been included explicitly, since no information about the specific up-take mechanism could be found.
   - Phosphatidic acid (PAC) and glycolipids provide half of the total lipid mass each.
   - Cardiolipin (CL) was not included in the biomass as the phospholipid branch (conversion of PAC into CL) is not essential in *M. genitalium* [Glass et al., 2006] and the functionality of this pathway has not been proven for *M. pneumoniae*. None of the intermediary metabolites nor CL have been detected (section 3.3.2), as neither have the two proteins assigned to this pathway [Maier et al., 2011].
   - For glycolipids we assumed the attachment of three galactose and three glucose molecules to each diacylglycerol molecule to account for the diversity of glycolipids.
   - The total amount of lipids needed to duplicate one cell was calculated based on the molecular masses of the "average" glycolipid and PAC, respectively.

4. Free amino acids and bases were determined by GC-MS experiments (section 3.3.3) and included in respective amounts in the biomass function.

5. Nucleobases have been considered (section 3.3.3) and included respectively.

6. Orthophosphate has been included based on the concentration measured in *E. coli* [Amin and Peterkofsky, 1995, Neidhardt, 1996].

7. Glucose 6-phosphate (G6P) has been quantitatively included in the biomass as model compound in order to account for all free organic metabolites (carbohydrates). To this end, we calculated the mass sum of all other defined biomass components, subtracted it from the total cell mass, and assigned the missing fraction to G6P (Table 3.1).

8. 5-formyltetrahydrofolate (5fTHF) was included in the biomass equation for its supposed inhibitory function on serine hydroxymethyltransferase (SHMT, mpn576), suggested from findings in plants [Goyer et al., 2005, Kruger et al., 1999].

9. Because ToBiN allows only integer values in reactions and the fraction of different compounds in biomass differs by several orders of magnitude, we rescaled the biomass mass units by a factor of 1,000,000 in order to avoid numerical problems of the solver.

Summing up, an average *M. pneumoniae* cell is composed of 62% protein, 5.2% DNA, 6.5% RNA, 1.8% diverse metabolites, as for example orthophosphate or free amino acids, 20% lipids and 4.5% other carbohydrates, represented by G6P. While the major building blocks have been included quantitatively, vitamins and cofactors proven essential have been included qualitatively. The assembly of the stoichiometric network (iJW145_reconstruct) and the assignment of reaction reversibilities together with the definition of the biomass composition enabled the model to simulate growth: *i*JW145_growth.

3.3.5. Model Refinement

To further validate the network structure of *i*JW145_growth and to avoid reconstruction errors [Reed et al., 2006, Henry et al., 2010], we simulated growth for different nutrient conditions *in silico*. To this end, constraint sets specifying the availability of growth limiting nutrients as well as accounting for essential cellular functions had to be defined. To not restrict the solution space of the model, but to keep a high predictive capacity [Edwards et al., 2002, Covert and Palsson, 2003, Price et al., 2004], we minimized the number of constraints to those indispensable for reproducing experimental findings.

Glucose is the main sugar source used for batch culture growth experiments of *M. pneumoniae in vivo*. Glucose consumption, organic acid secretion, and protein production have been measured during a four days batch culture growth experiment (Appendix A, Figure A.2). Based on those measurements and literature information [Yus et al., 2009, Maier et al., 2011], we defined initial flux constraints representing the nutrient

supply by defined and rich medium as described in Material and Methods, section 3.2.1. During four days of batch culture growth, a metabolic shift from mainly acetic acid fermentation towards mainly lactic acid fermentation can be observed (Appendix A, Figure A.2) [Yus et al., 2009]. In agreement, the abundance of the lactate dehydrogenase (LDH) increases from 203 copies per cell at early growth stages to above 1000 copies per cell after four days of growth [Maier et al., 2011]. To represent this shift in our model, we directly constrained the favorable acetic acid production.

Applying the defined constraint sets, the model was able to simulate growth under different conditions using biomass synthesis as objective function for the FBA. Initial simulations and flux distributions verified the network structure and confirmed a wide range of experimental data. Nevertheless, we also identified several conflicts between model predictions and available experimental results. These conflicts were resolved in an iterative process of model simulations and evaluation of the prediction results, thus guiding the correction of the wiring diagram (Figure 3.8). To this end, we accomplished

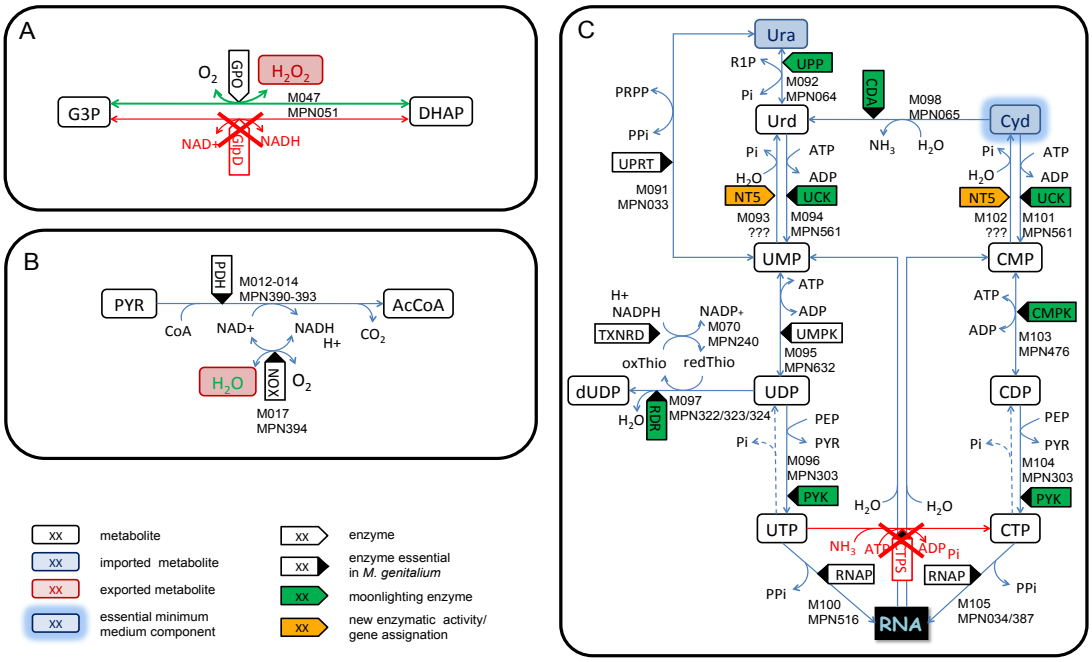

Figure 3.8.: Sub-networks of *M. pneumoniae* metabolism to which changes have been applied. A: pyrimidine metabolism to illustrate the removal of the reaction converting UTP into CTP; B: NOX reaction has been changed from producing H2O2 to producing H2O; C: changes applied to the reaction converting G3P into DHAP. Red crossed reactions have been removed from the network.

additional experiments and used sequence alignments, literature mining and repeated simulations with adjusted constraints, ultimately resulting in a refined metabolic network, *i*JW145. The processes leading to the correction of the wiring diagram are outlined below.

Redox state: For ATP generation, *M. pneumoniae* relies on glycolysis and the subsequent organic acid fermentation. Pyruvate dehydrogenation, leading to the production of acetic acid, causes an intracellular redox imbalance due to the accumulation of NADH. The model initially predicted a circular interconversion of DHAP and G3P to maintain the redox balance, thereby producing toxic H_2O_2. According to the wiring diagram from [Yus et al., 2009], this interconversion was catalyzed by a coupled glycerol phosphate dehydrogenase (GlpD)/glycerol phosphate oxidase (GPO) (MPN051). However, in ^{13}C-glucose tracer experiments we examined only very low conversion rates of DHAP into G3P (compare section 3.3.7). Additionally, sequence analysis (Appendix A, section A.1.2) and experimental results [Hames et al., 2009] independently characterized MPN051 as a GPO without GlpD similarity. In consequence, we deleted the GlpD reaction (R050 in [Yus et al., 2009]) from the model (Figure 3.8A).

Subsequently, *in silico* the NADH oxidase (NOX, MPN394) maintained the redox balance, thereby producing major quantities of H_2O_2 from oxygen. This hypothesis was not supported experimentally, since the GPO has been shown to be the major source for H_2O_2 production [Hames et al., 2009]. Nevertheless, the high abundance of the NOX (1763 copies per cell, [Maier et al., 2011]), suggest high cellular activity. Literature search revealed that two NOX isoforms, one producing H_2O_2 and the other H_2O, exist [Sakamoto et al., 1996]. According to sequence alignments using pBLAST (Appendix A, section A.1.3), the *M. pneumoniae* NOX was re-annotated as a H_2O-producing isoform containing even the characteristic FAD-binding fingerprint conserved through a variety of organisms (Appendix A, Figure A.6). The model has been adapted accordingly (Figure 3.8B), now predicting NOX-dependent maintenance of the cellular redox balance and the GPO being the major source of H_2O_2 production.

Pyrimidine requirements: In their minimal medium experiments [Yus et al., 2009] determined the nucleobase cytidine as essential and sufficient to synthesize all pyrimidine nucleotides. Contradicting this finding, but confirming literature data on other mycoplasmas [Maniloff et al., 1992, Pachkov et al., 2007], *i*JW145 predicted that also uracil could serve as precursor for the production of all pyrimidine nucleotides. One by one we silenced the reactions of the pyrimidine metabolism, thus identifying MPN256, a proposed CTP synthase, to be responsible for the sufficiency of uracil for *in silico* growth (Figure 3.8C). pBLAST results could not confirm this function assignment, showing that neither MPN256 nor any other *M. pneumoniae* protein has significant sequence similarity with CTP synthases from other organisms, such as *B. subtilis* or *E. coli* (Appendix A, section A.1.4). After removing the respective reaction (R098 in [Yus et al., 2009]) from the model (Figure 3.8C), only cytidine can be used as a precursor for the synthesis of all pyrimidine nucleotides.

Glycerol essentiality in minimal medium: Glycerol has been shown to be essential for *in vivo* growth of *M. pneumoniae* under minimal medium conditions [Yus et al., 2009]. Growth simulations initially did not support this finding due to the fact that glucose was used for the production of the lipid precursor G3P when silencing other

sugar sources. Yet, ^{13}C-glucose tracer experiments have shown that the interconversion rate of DHAP into G3P is very low (section 3.3.7, Figure 3.15B). Since the GPO is a membrane-bound protein releasing toxic H_2O_2 to the environment, one could speculate that the availability of intracellular H_2O_2 limits the production of G3P from DHAP. Limiting the GPO reaction (M048) accordingly, resulted in glycerol, G3P and fructose being the only sugars that can be used for lipid production in *M. pneumoniae*.

G3P was converted into phosphatidic acid, the primary lipid precursor, directly. Glycerol, when taken up, was phosphorylated beforehand. In case of fructose uptake, fructose 1-phosphate (F1P) was broken down into DHAP and glyceraldehyde (GA) which in turn is processed into glycerol. However, *in vivo* glycerol could not be substituted by fructose (G3P was not tested) [Yus et al., 2009]. A literature screen revealed that in *E. coli* and *Helix pomatia* the aldolase, the enzyme converting F1P into DHAP and GA, has only about 3.5% and 5% affinity towards F1P compared to its affinity towards FBP [Kochman et al., 1982, Szwergold et al., 1995]. Similar affinities in *M. pneumoniae* would lead to an extremely limited incorporation of carbon compounds into the lipid metabolism. Enzymes involved in fructose uptake and processing are very low abundant in *M. pneumoniae* when grown under rich medium conditions [Maier et al., 2011] and have been shown to be significantly over-expressed in the fructose-adapted strain [Yus et al., 2009]. In addition, glycerol has been shown to play a regulatory role in glucose uptake via the phosphorylation of the phosphocarrier protein HPr (MPN053) [Halbedel et al., 2006]. We constrained the conversion of DHAP into G3P in order to reproduce experimental findings and to allow the model to account for the regulatory role of glycerol.

Lipid metabolism: The lipid metabolism is probably the less studied metabolic pathway in *M. pneumoniae*. For the phospholipid branch, only two of the supposed four catalyzing enzymes are known and both could be disrupted in *M. genitalium* [Yus et al., 2009, Glass et al., 2006]. In *M. pneumoniae* the two respective genes, *mpn637* and *mpn253*, are expressed and show changes in their expression level under different conditions [Güell et al., 2009]. However, the corresponding proteins have not been detected [Maier et al., 2011] as neither has been cardiolipin (compare section 3.3.2). We accomplished alignments with the *E. coli* phosphatidylglycerol phosphatases (*pgpA* and *pgpB*) and cardiolipin synthase (*cls*) versus the *M. pneumoniae* proteome and the nr-DB using pBLAST. Neither *M. pneumoniae* nor any other mycoplasma species have proteins with significant sequence similarity. Without further experimental research it remains questionable if *M. pneumoniae* is able to synthesize cardiolipin. The proteins assigned to the phospholipid branch could have alternative activities as many other so-called moonlighting enzymes in *M. pneumoniae* [Yus et al., 2009] or just have not been eliminated yet during the genome reduction process. Due to lack of experimental proof for cardiolipin synthesis in *M. pneumoniae* we included phosphatidic acid instead of cardiolipin as a biomass component, despite the possibility that it might be processed further (Table 3.1).

In summary, the comparison of model predictions and experimental data guided the correction of the wiring diagram of *M. pneumoniae* (Figure 3.8). Furthermore, we cor-

rected the functional annotation of three *M. pneumoniae* proteins: MPN051 is a GPO with no GlpD activity, MPN394 is producing water not H_2O_2, and MPN256 has no CTP synthase activity.

3.3.6. Model Validation

Metabolic Capabilities When Grown on Alternative Sugar Sources

M. pneumoniae is able to process several alternative carbon sources despite from glucose: glycerol, G3P, fructose, ribose, mannose, mannitol, ascorbate, and G3PC [Yus et al., 2009, Schmidl et al., 2011]. To validate our metabolic reconstruction, we adjusted the constraint sets for rich and minimal medium compositions for growth on different carbon sources and predicted the qualitative *in silico* growth capabilities of *M. pneumoniae* (Table 3.2). We verified the prediction by comparing our results to experimentally determined growth capabilities from different studies [Yus et al., 2009, Schmidl et al., 2011].

Table 3.2: Growth on Alternative Sugars

	rich medium		**defined medium**
sugar	***in silico* growth**	***in vivo* growth [Yus et al., 2009]**	***in silico* growth**
glucose	✓	✓[1,2]	✓
fructose	✓	✓[1,2]	0
mannose	✓	✓[2]	✓
mannitol	✓	-2	✓
ribose	✓	(✓)2	✓
ascorbate	✓	(✓)2	✓
glycerol	✓	(✓)2	0
G3P	✓	not tested	0
G3PC	✓	(✓)1	0

Table 3.2.: Comparison of the *in silico* predictions to *in vitro* results for growth on alternative sugars. ✓- growth; (✓) - catabolic activity (growth not proven by protein measures); 0 - catabolic activity but no growth; not tested - growth on the respective sugar source has not been examined; 1 - this study, 2 - Yus et al. [2009].

In silico doubling times were comparable when the same amount of carbon was provided. In contrast, the *in vivo* doubling times differed significantly (Appendix A, Figure A.2) [Yus et al., 2009]. However, it has been also shown that several serial passages are necessary to adapt the wild type *M. pneumoniae* to growth on fructose and that in those adapted cells the enzymes involved in fructose uptake and processing are significantly overexpressed [Yus et al., 2009]. Comparing the abundances of the different sugar uptake proteins, we found that the glucose-specific protein (MPN207) has high copy numbers (∼385/cell), while all other known sugar transporters are about 14 - 100 times

less abundant (fructose and mannose: 3-5 copies/cell; ribose: 25 copies/cell; glycerol, G3P, mannitol and ascorbate: not detectable) [Maier et al., 2011].

For minimal medium conditions, the model predicted growth on glucose, ribose, mannose, mannitol, and ascorbate while no growth has been observed on fructose, glycerol, G3P, and phosphatidylcholine (Table 3.2). The inability to synthesize the pentose phosphate pathway precursor fructose 6-phosphate (F6P), indispensable for *de novo* nucleotide synthesis in absence of ribosylated nucleobases, has been identified as cause for this growth limitations.

Gene Essentiality Prediction and Analysis of Mutant Phenotypes

To further validate the model on a global scale, we accomplished an *in silico* knock-out study for 131 metabolic genes (genes coding for proteins involved in DNA degradation, ATPase function, and chaperone activity have not been included in knock-out study since the respective processes have not been modeled explicitly). To this end, we systematically silenced, i.e. limited to zero, all reactions catalyzed by the same gene product and applied FBA for growth under rich medium conditions. 73 genes (56% of enzymes included in the prediction) were predicted to be essential, since the respective knock-out led to either growth arrest or cell death when not all minimum constraints could be matched (Figure 3.9A and Appendix A, Table A.10). Conversely, 58 enzymes (44%) were predicted to be not essential, since their *in silico* knock-out resulted in objective values for the FBA larger than zero.

The prediction results were evaluated by comparing them to a genome-wide transposon mutagenesis study in the closely related bacterium *M. genitalium* [Glass et al., 2006]. Therefore, we assigned functional orthologs of *M. pneumoniae* and *M. genitalium* based on sequence alignments for the genes and the promotor regions and on an alignment of COG IDs (Appendix A, Table A.11). For 17 *M. pneumoniae* genes no ortholog in *M. genitalium* could be assigned. Those genes were considered to be not essential due to their absence in *M. genitalium* and the high similarity of the metabolic networks of both organisms.

Table 3.3: Statistics for the gene essentiality prediction

	comparison to Glass et al. [2006]	**comparison also to *M. pneumoniae* mutants**	**taking conditions into account**
TP (true positive)	72	72	72
TN (true negative)	41	48	53
FP (false positive)	1	1	1
FN (false negative)	17	10	5
ACC (prediction accuracy)	0.8626	0.9160	0.9542
SPC (prediction specificity)	0.9762	0.9796	0.9815
ACC in %	**86.26**	**91.60**	**95.42**
SPC in %	**97.62**	**97.96**	**98.15**

Table 3.3.

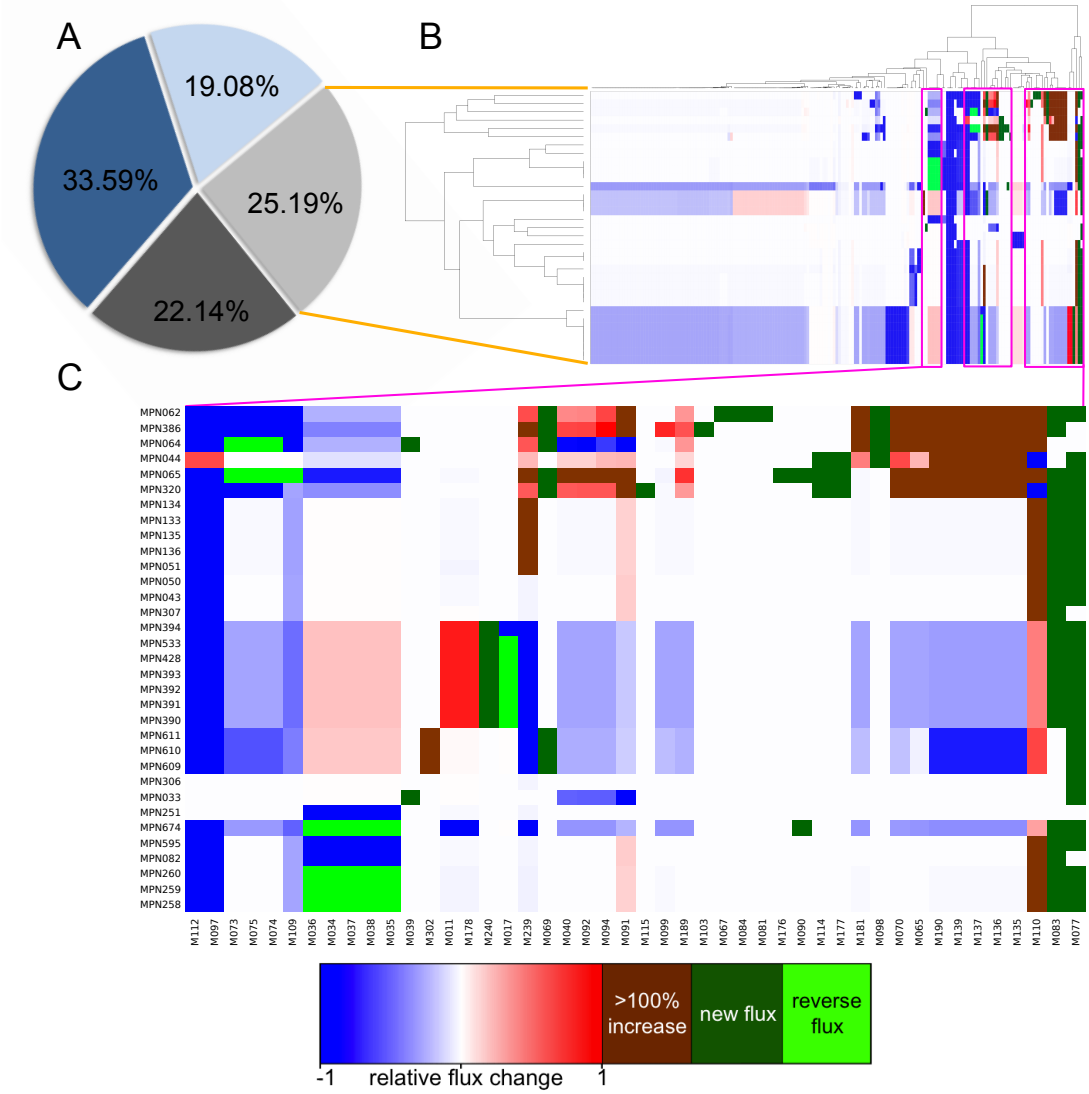

Figure 3.9.: *In silico* knock-out study. A: Knock-out effects on growth separated into lethal (dark blue), growth inhibiting (light blue), reduced fitness (light grey), and growth not affecting (dark grey) mutant phenotypes; B: relative flux changes for reduced fitness knock-out vs. wild type simulations (sink/source reactions are excluded): down-regulation (blue) → up-regulation up to 100% (red), inverse fluxes (light green), new fluxes (dark green) and flux changes >100% (brown); C: relative flux changes of reactions predicted to have inverse, new or highly up-regulated flux under at least one knock-out condition.

In a first, unbiased analysis using gene essentiality in *M. genitalium* and the complex assumption (see Material and Methods, section 3.2.1) as only criteria, we achieved 86% accuracy (correct predicted/total predicted) and 97% specificity (true negatives/(true negatives + false negatives)) with the *in silico* gene essentiality prediction (Table 3.3). In case of contradictions between model prediction and gene essentiality in *M. genitalium*, we screened a *M. pneumoniae* transposon library and could confirm the prediction of five non-essential genes (Figure 3.10). When further taking the simulated conditions (rich medium, growth on glucose) into account, the model predicts essentiality for metabolic genes with a final accuracy of more than 95% and specificity higher than 98% (Table 3.3). We conclude that the metabolic model *i*JW145 possesses high predictive power for metabolic phenotypes.

To obtain information on the impact of gene knock-outs on the metabolic behavior of *M. pneumoniae*, we quantitatively assayed the flux changes in *in silico* knock-outs producing reduced fitness phenotypes (Figure 3.9B). We performed unbiased clustering for genes and reactions, respectively, according to either their influence on reaction fluxes or their flux changes in reduced fitness phenotype producing knock-outs (Figure 3.9B). Genes of the acetate branch (pyruvate metabolism) are found to have similar influence on the network behavior and reactions of the same pathway often show similar changes in the same *in silico* knock-out. The many small clusters (composed nearly exclusively of

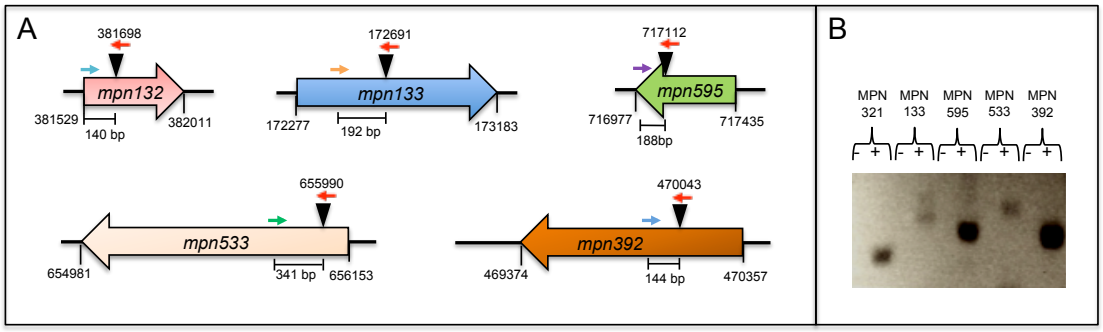

Figure 3.10.: Isolated knock-out mutants of *M. pneumoniae*. A: schematic representation of the disrupted genes (big colored arrows) with genome positions, insertion site (black triangle), and the length of the PCR fragment (indicated below the respective gene); B: Western Blot of positive screening results of a mutant library, confirming the mutants predicted *in silico*.

the components of just one metabolic complex) and single enzyme clusters reflect the high percentage of multifunctional enzymes, i.e. enzymes catalyzing more than one metabolic reaction, and transport proteins encoded by the non-essential genes (together 51.5%). When analyzing the qualitative flux changes, we found that most of the individual reaction fluxes (colored boxes in Figure 3.9B) are down-regulated (blue, ∼54%) or do not change at all (white, ∼34%) in response to *in silico* gene deletions resulting in reduced fitness phenotypes. However, we also identified several highly up-regulated reactions (red-brown, ∼10.4%) and a few flux direction changes (dark green, ∼0.4%). Only 88 new fluxes (light green, ∼1.1% of all fluxes) have been observed in a total number of

15 reactions in the 32 analyzed knock-outs, highlighting the impact of the reductive genome evolution and the resulting lack of metabolic rescue pathways on the metabolic capabilities of *M. pneumoniae*.

We analyzed the reactions with highly up-regulated, new or reverse fluxes under at least one *in silico* knock-out (Figure 3.9C). As expected, the LDH reaction (M011) gets significantly up-regulated in all knock-outs of genes related to the acetate branch of the pyruvate metabolism. Surprisingly, all other 28 selected reaction belong to either nucleotide metabolism (and associated glycolysis or cofactor metabolism reactions) or the pentose phosphate pathway. One could speculate that the non-essential genes in those two pathways have been preserved during the reductive genome evolution to allow *M. pneumoniae* to presumably maintain growth ability in case of non-constant supply of (some) nucleobases.

3.3.7. Model Application

In Silico Double Knock-out Prediction

Synthetic genetic array analysis has been applied to study network connectivity and the functional relation amongst genes of different metabolic pathways [Tong et al., 2001, 2004, Szappanos et al., 2011]. The analysis of sick and synthetic lethal interactions, i.e.

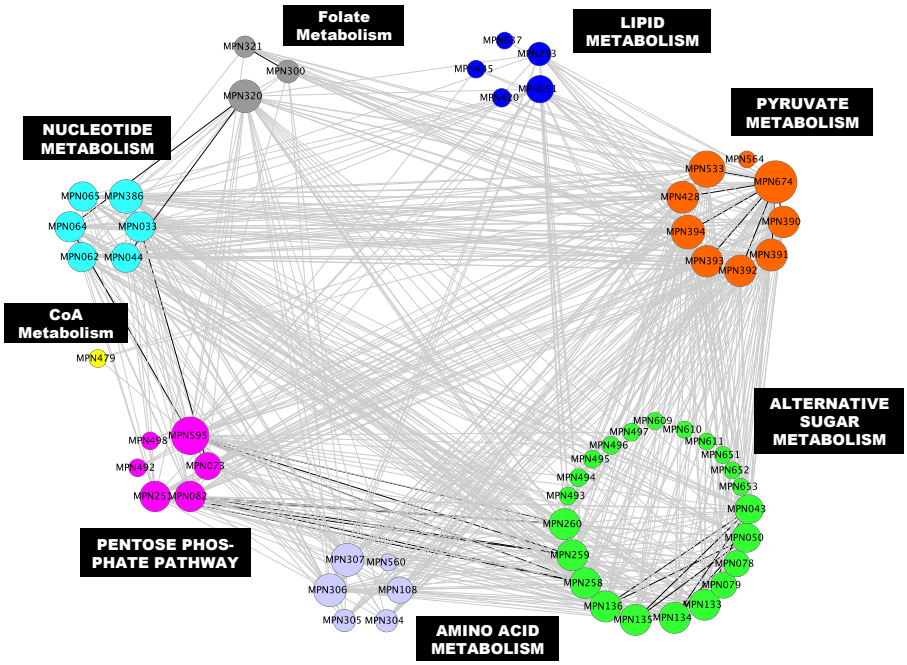

Figure 3.11.: Synthetic lethal (black) and sick (grey) interactions between *in silico* non-essential genes of *M. pneumoniae*; genes are sorted according to their metabolic pathways and sized based on the number of synthetic lethal and sick interactions.

double mutants that cause combined reduced fitness phenotypes or cell death, allows the identification of gene products impinging on the same biological process [Hartman et al., 2001]. Lacking appropriate experimental analysis tool for *M. pneumoniae* and encouraged by the high accuracy of the gene essentiality prediction, we performed an *in silico* interaction screen by predicting double mutant phenotypes for 58 genes assumed to be not essential from the single knock-out results (Appendix A, Table A.13). Analysis of *in silico* synthetic lethal and sick interactions showed that pyruvate metabolism genes have a global effect on the metabolic behavior in *M. pneumoniae* (Figure 3.11). Due to the limitations in acetic acid production, *mpn674*, encoding the LDH, has a particularly strong effect on the growth fitness, confirming the central role of pyruvate metabolism for ATP generation in *M. pneumoniae*. In addition, genes coding for proteins involved in sugar uptake and processing can have an impact on energy production but have much weaker influence than pyruvate metabolism genes.Nucleotide metabolism and pentose phosphate pathway genes, respectively, are enriched among genes producing synthetic lethal interactions. This supports the results of the single knock-out analysis, suggesting that remaining rescue routes in *M. pneumoniae* metabolism are contained mainly in these two pathways.

Cellular Energy Balance

During the mid exponential growth phase (at about 36 hours after inoculation), *in vivo M. pneumoniae* consumes ∼20000 glucose molecules per cell and second (Figure 3.13B). Applying FBA with biomass production as objective function, this leads to the production of ∼60000 ATP molecules per cell and second (Appendix A, Table A.14) and *in silico* doubling times of 2.3 to 3.8 hours (Figure 3.12A). In contrast, doubling times determined *in vivo* during the exponential phase in batch culture growth (19.7 to 59.7 hours, Figure 3.13A) and reported in previous microscope studies (∼8 hours) were much longer. These results suggested the existence of additional energy sinks not yet specified in the model.

To estimate the contribution of those maintenance tasks on energy homeostasis in *M. pneumoniae* we defined an unspecific energy consuming reaction. We then manually fitted the minimum constraint of this reaction for each simulated time point to allow reproduction of *in vivo* doubling times (Figure 3.12A). Interestingly, we find that *M. pneumoniae* uses more than 70% of its generated energy for yet unknown or not quantified functions (Figure 3.12B). Depending on the simulated growth time, only between 12% and 29% of the produced energy is used for the synthesis of biomass, for protein and mRNA turnover as well as for the detoxification functions defined in the model (Figure 3.12C and Appendix A, Table A.14). More precisely, at 36 hours of growth (optimal doubling rate determined), 9.8% of the total energy are used for protein production (assuming protein half-life of 23 hours [Maier et al., 2011]), while 8.4% is dedicated to RNA synthesis (assuming mRNA half-life of 1 min [Maier et al., 2011]). DNA synthesis consumes 0.05% of the available energy, lipid production about 0.4% and 4.9% are used up for the uptake and processing of reaction cofactors as well as the detoxification functions defined in the model (Figure 3.12C and Appendix A, Table A.14).

To further characterize the additional energy sinks in *M. pneumoniae*, we first classified them into GAM and NGAM tasks [Pirt, 1965, Varma and Palsson, 1994b]. To

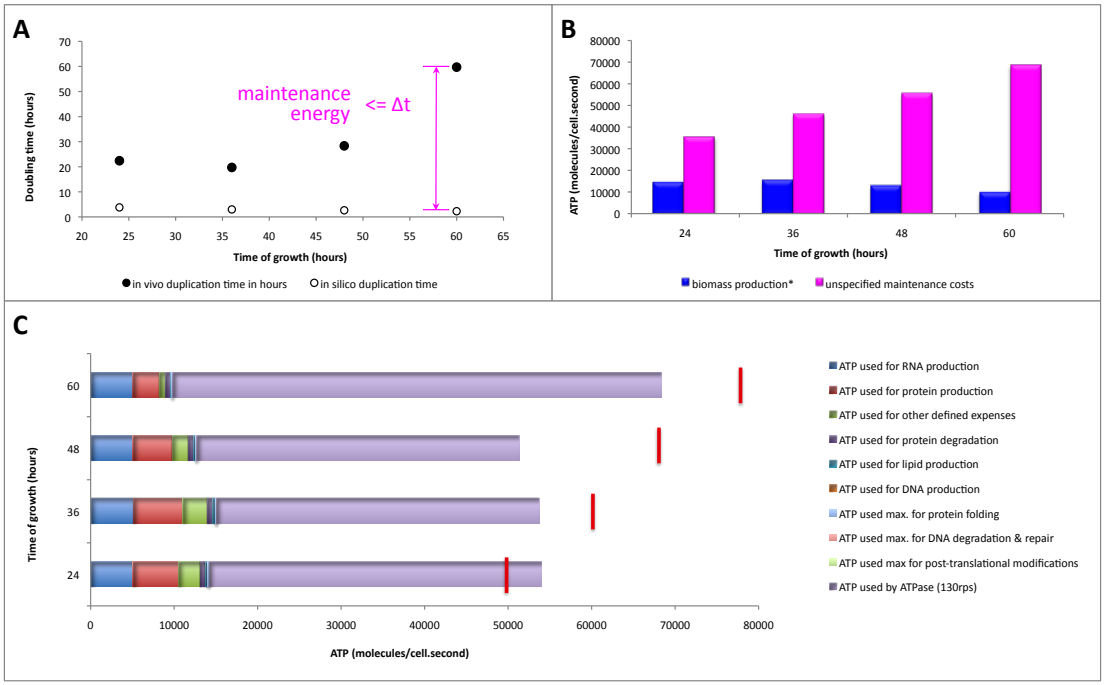

Figure 3.12.: Analysis of energetic expenditures at different times of a four days batch culture growth: A: Comparison of *in vivo* (white circles) and *in silico* (black circles) doubling times leading to the definition of the maintenance energy; B: Energetic expenditures used for biomass and functions specifically defined in the model (red) and for maintenance tasks (blue); C: ATP usage per cell and second for the different functions of a *M. pneumoniae* cell, f.l.t.r. synthesis of RNA, protein production, other specified functions, protein degradation, lipid production, synthesis of DNA, chaperone catalyzed protein folding, DNA maintenance, ATPase function. Red vertical bars indicate the total number of ATP molecules produced at each time point.

estimate the contribution of GAM to the total energy costs, we calculated upper boundaries for the ATP consumption by DNA repair (0.01%), post-translational modifications (0.03%), and chaperone-assisted protein folding (0.25%) based on available experimental data and literature (Appendix A, Table A.14) [van Noort et al., 2012, Drake et al., 1998, Naylor and Hartl, 2001, Maier et al., 2011].

DNA Maintenance Costs: DNA-based microbes have mutation rates <0.005 per genome per duplication [Drake et al., 1998]. In order to calculate the upper boundary for DNA maintenance costs we assumed a mutation rate of 0.05 for the *M. pneumoniae* genome and find that DNA maintenance can account for a maximum of about 0.01% of the total generated energy.

Post-translational Modification Costs: 93 phosphorylation sites and 720 acetylation

sites on 72 and 221 proteins, respectively, have been detected in *M. pneumoniae in vivo* [van Noort et al., 2012]. Integrating this information with the abundances of the respective proteins [Maier et al., 2011], we calculated the upper boundary for post-translational modifications.On one hand, we assumed that all phosphorylation and acetylation sites on all found proteins are modified once. This accounts for 0.01% of the total ATP expenses. Adding another 0.02% of the total ATP consumed at each time point in order to account for cyclic phosphorylation-dephosphorylation events [Shacter et al., 1984], we find that post-translational modifications account for 0.03% of the total energy. For lack of information about de-acetylation we did not consider it in further detail.
Chaperone-catalyzed Protein Folding Costs: We integrated the abundances determined for the two *M. pneumoniae* chaperones (GroEL, a 14mer, and DnaK) [Maier et al., 2011] with catalytic rates described for *E. coli* (GroEL: 20 sec & usage of 7 ATP per folding cycle, DnaK: 15 sec & usage of 1 ATP per folding cycle) [Naylor and Hartl, 2001]. Assuming constant activity of all chaperones we calculated the upper boundary for the ATP consumption for protein folding. Considering in addition protein turnover costs (1.1%) and the other expenses defined in the model total expenses on GAM account for a maximum of 6.9% of the total cellular energy.

Systematic literature screening identified proton translocation by the cellular ATPase as most significant quantifiable NGAM task [Kobayashi, 1985]. The ATPase is mainly involved in maintaining an optimal proton gradient across the cellular membrane to allow nutrient import. Towards later growth stages, when the acidification of the growth medium results in pH stress for the cells, the intracellular pH maintenance additionally challenges ATPase function [Kobayashi, 1985, Moreno et al., 1998]. We determined the amount of ATPases per *M. pneumoniae* cell based on the abundance of the β-subunit (MPN598) of which three copies are contained in the ATPase core [Maier et al., 2011]. Integrating mechanistic information on the ATP hydrolysis rate of the ATPase [Watanabe et al., 2008] with *in vivo* measurement data on ATPase rotation speed [Watanabe et al., 2008, Wu et al., 2010] and abundances of ATPase components (99 - 150 complexes/cell, [Maier et al., 2011]), we estimated the energy consumption of the *M. pneumoniae* ATPase. Assuming constant ATPase activity at maximum speed (130 rps), the ATPase uses a maximum of 38610 ATP per cell and second at 36 hours after inoculation *in silico*, accounting for about 57% of the total cellular energy, and an even higher fraction at later growth stages (Appendix A, Table A.14).

Summing up, we combined our metabolic model and experimental results to quantitatively analyze the global energy balance of *M. pneumoniae*. We can explain the consumption of 75% - 100% of the total energy produced, by considering all quantifiable ATP consuming processes (Figure 3.12C). Biomass production accounts for about 11 - 22%, GAM for about 2-7%, and NGAM for 57 - 80% of the total ATP generated during the exponential growth phase *in silico*. It is important to note that *M. pneumoniae* during the exponential growth phase of a batch culture experiment uses about 78% - 89% of the total generated energy not for the production of cell building blocks but intracellular homeostasis.

The definition of the maintenance energy sink and the subsequent determination of its

constraint for different growth times, completed the construction and refinement of our metabolic model for *M. pneumoniae*, *i*JW145. The final model is composed of 306 reactions connecting 216 metabolites and 145 catalyzing enzymes (Appendix A, Figure A.1 and Table A.1; http://mycoplasma.crg.es/pathways.php). It is able to reproduce experimentally determined doubling times as well as external metabolite measurements and to predict metabolic flux distributions for different selected time points during batch culture growth.

Analysis of Metabolic Flux Changes *In Silico*

For dynamic simulations, models based on ordinary differential equations (ODEs) are usually employed. However, these models have limited applicability for large-scale analysis mainly caused by overfitting, i.e. the failure to determine biologically useful parameter sets due to insufficient quantitative experimental data [Draper and Smith, 1998]. To overcome the limitations of the static modeling approach used and to gain information about the changes in metabolic flux distributions during the exponential growth phase, we designed an approach for quasi-dynamic simulations of a constraint-based model. To this end, we fitted a logarithmic function (Equation 3.3) to the determined maintenance costs in addition to the fitting of sigmoidal functions (Equations 3.1 & 3.2) to the concentration data for glucose, acetic acid, lactic acid, and protein (Figure 3.13 and Appendix A, Table A.8). Although these functions do not allow the extraction of kinetic parameters, they enabled us to calculate constraint sets for any given time point of the exponential growth phase, thereby also accounting for the experimental error.

Simulating growth with respectively determined constraint sets allows to predict metabolic flux distributions for different times of the exponential growth phase, thus providing information about changes in metabolic behavior during batch culture growth (Figure 3.14A and Appendix A, Table A.15). Analyzing the qualitative flux change between *in silico* flux distributions for 24, 36, 48, and 60 hours, we found that about 51.6% of all reactions show the same qualitative flux changes as biomass synthesis, i.e. the flux increases from 24 hour to 36 hours and decreases from 36 hours to 48 hours as well as from 48 hours to 60 hours (Figure 3.14B). Another 2.6% of the reactions show flux changes contrary to biomass synthesis, i.e. first increase (24 - 36 hours) and then decrease (36 - 48 - 60 hours). For 11.4% of the metabolic reactions the flux constantly increases during the exponential growth time while for 5.9% the flux constantly decreases. 1.6% show now flux change during the exponential growth phase (minimum constraint reactions), 0.7% show diverse changes, and 26.1% are not used under the simulated rich medium conditions.

Only in four pathways (for the assignment see Appendix A, Table A.1), namely in glycolysis, pyruvate metabolism, energy metabolism, and biomass production, all reactions are active under rich medium conditions *in silico* (Figure 3.14C). Interestingly, in amino acid metabolism, nucleotide metabolism, and pentose phosphate pathway between 25% and 35% of the reactions are not active. Lipid metabolism also contains 30% non-active reactions, which despite being considered a direct effect of the assumptions for the biomass composition, is in agreement with detected proteins [Maier et al., 2011] and

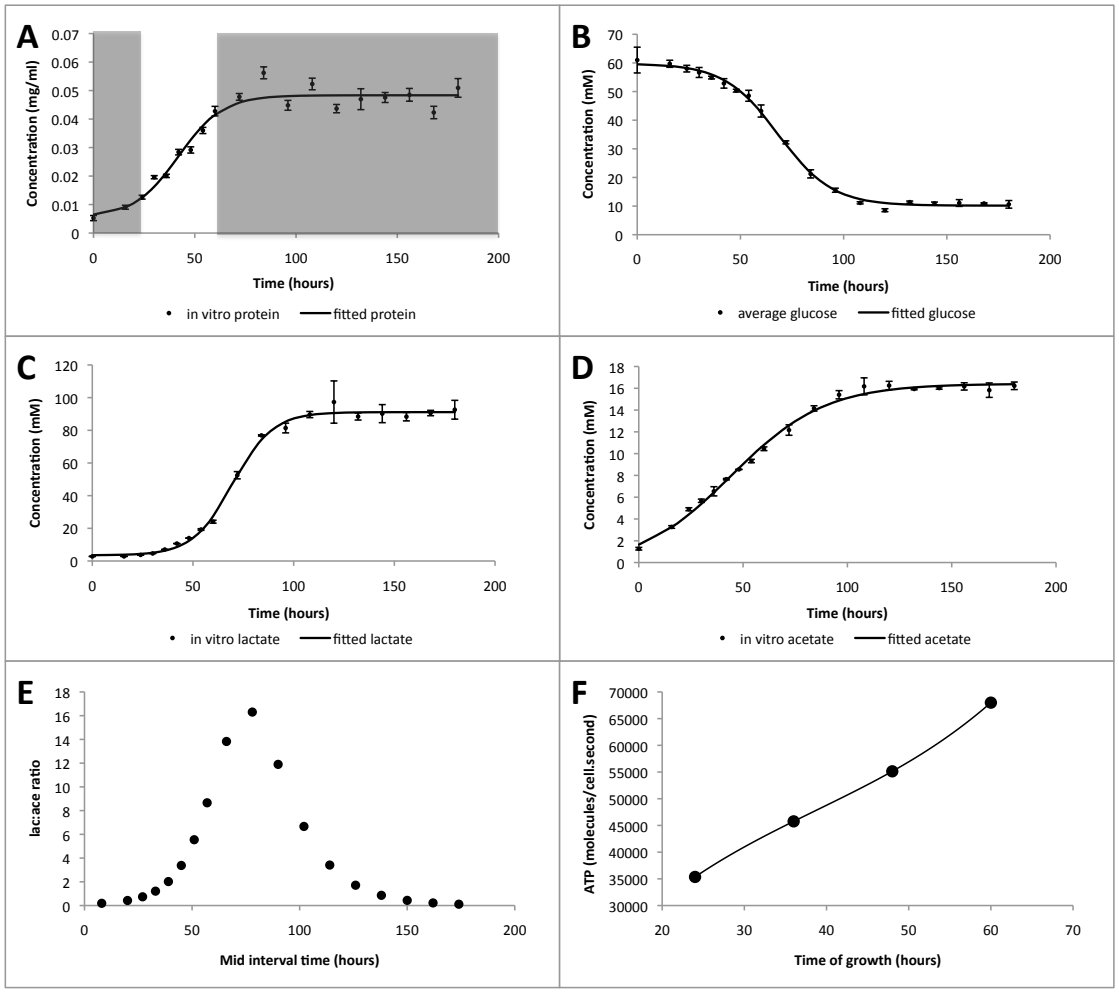

Figure 3.13.: Comparison of *in vivo* measurements of external metabolites and protein quantities with the nonlinearly fitted curves used to determine constraints to simulate different time points of growth *in silico*: A: protein, only the exponential growth phase (24-60 hours) for which *i*JW145 is defined has white background. B: glucose. C: lactic acid. D: acetic acid. E: lactate to acetate ratio (based on fittings). F: *in silico* maintenance costs.

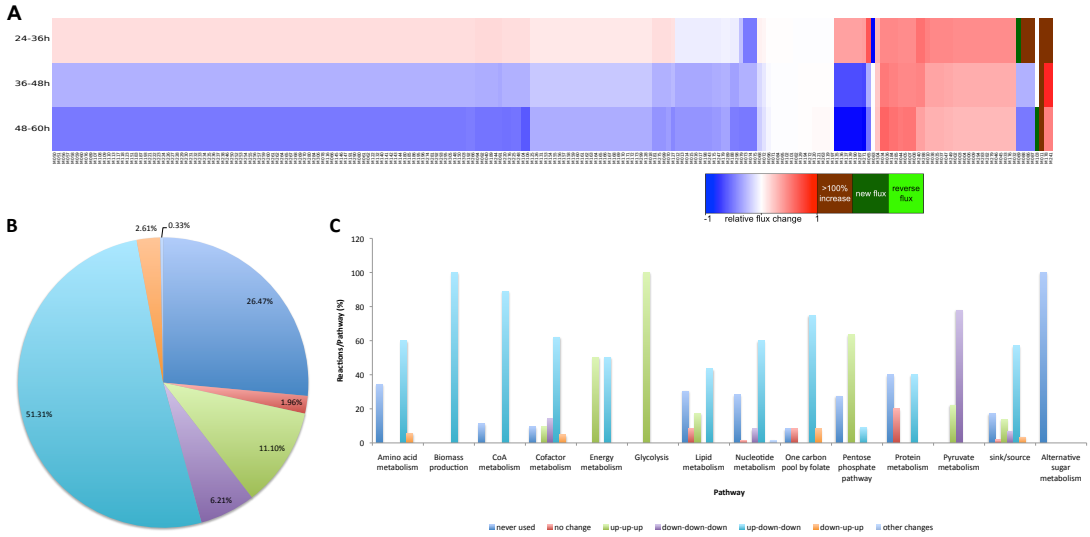

Figure 3.14.: A: qualitative flux changes between simulated time points (t = 24, 36, 48, 60 hours). B: Reaction classification according to their qualitative flux changes along the exponential growth phase. C: As in B, but distinguished for the different pathways.

metabolites (see section 3.3.2). Based on the simulated conditions, reactions involved in the processing of alternative sugars are not active. The constantly changing reactions, either belonging to glycolysis, to pyruvate metabolism, the energy producing arginine fermentation (amino acid metabolism) or being associated cofactor transport reactions (Figure 3.14C), comprise all catabolic reactions of *M. pneumoniae*. Their behavior can be attributed to the increase in maintenance costs during batch culture growth and the adaptation of the catabolic pathways as well as to the imposed acetate production constraints. The two reactions showing diverse changes belong to the nucleotide metabolism and a detailed analysis revealed that apparently the two routes producing deoxy-CDP and deoxy-GDP, respectively, are energetically equal and therefore can be used interchangeably.

We conclude, that the capability of *M. pneumoniae* to adapt to environmental stress observed *in vivo* [Güell et al., 2009] is probably mainly owed to those pathways still preserving unused metabolic routes, i.e. nucleotide metabolism and pentose phosphate pathway. In general, the high activity of metabolic reactions in *M. pneumoniae* reflects the reduced genome and the parasitic life, as the simple linear network structure is fine-tuned to take up as many cell building blocks and cofactors from the environment as possible.

The description and analysis of a cellular subsystem in isolation holds the risk to miss important regulatory influences from other subsystems of the cell. To gain insight into the relation between protein abundances and metabolic activity we integrated experi-

mentally determined protein quantification data [Maier et al., 2011] with *in silico* flux predictions for different time points of batch culture growth (Appendix A, Figure A.8). We aligned the qualitative trends of flux and protein abundance changes during the exponential growth phase as described in Material and Methods, section 3.2.1, finding that they match in about 86% of the cases (Appendix A, Figure A.8A). These findings are in agreement with a recent study analyzing the dynamic adaptation of *B. subtilis* to nutritional shifts [Nicolas et al., 2012]. Pyruvate metabolism (83.3% hit enzymes), amino acid metabolism (75% hits), nucleotide metabolism (76.5% hits), and folate metabolism (50% hits) are the only pathways for which not all enzyme abundance changes match the changes in fluxes of respectively catalyzed reactions (Appendix A, Figure A.8B), when taking into account the experimental error reported for protein quantification [Maier et al., 2011]. The integration of information about post-translational modifications [van Noort et al., 2012] did not lead to further conclusions about the influence of protein concentrations on metabolic regulation. No significant enrichment of modified proteins among the proteins showing abundance changes contradicting the predicted flux changes compared to the modified fraction of hit enzymes could be determined.

Analysis of the Central Carbon Metabolism *In Vivo*

To further analyze the central carbon metabolism, we measured cellular concentrations of key metabolites in sugar processing pathways. We found considerably small intracellular pools of glycolysis intermediates, ranging from 994 molecules per cell (0.03 mM) for GAP to 33400 molecules per cell (1.1 mM) for FBP (Figure 3.15A). Considering the nonlinear fitting to *in vivo* measurements of extracellular glucose, we determined that the glucose uptake rate during the exponential growth phase ranges from 8,000 to 35,000 molecules*cell^{-1*}second^{-1} (Figure 3.13 and Appendix A, Figure A.2).

To experimentally confirm the suggested fast turnover of intracellular pools of glycolytic intermediates, we accomplished ^{13}C-glucose tracer experiments. To this end, *M. pneumoniae* cell were pulse-fed with heavy isotope labeled glucose ($^{13}C_6H_{12}O_6$) and monitored the propagation of the labeled carbon atoms through glycolysis by GC-MS. Already 15 seconds after the supply of labeled glucose, the earliest time point for reproducible measurements, we found the intracellular pools of all detectable glycolytic intermediates labeled to high excess (Figure 3.15). Time-dependent incorporation of carbon 13 is well described by the curves shapes of one-phase exponential decay functions fitted to the fraction of labeled compound in percentage of the total intracellular pool of the respective compound (Appendix A, Figure A.7; R^2 values can be found in Appendix A, Table A.16).

The experimentally determined metabolic network [Yus et al., 2009] and the analysis of branching metabolites (see section 3.3.1) suggested an unusually low inter-pathway connectivity for *M. pneumoniae*. Analyzing the incorporation of labeled glucose into reporter compounds for lipid metabolism (G1P and G3P) and the pentose phosphate pathway (R5P), we explored the connectivity of glycolysis to other metabolic pathways *in vivo* (Figure 3.15B).

The conversion of G6P into G1P, a precursor for glycolipid synthesis, constitutes the

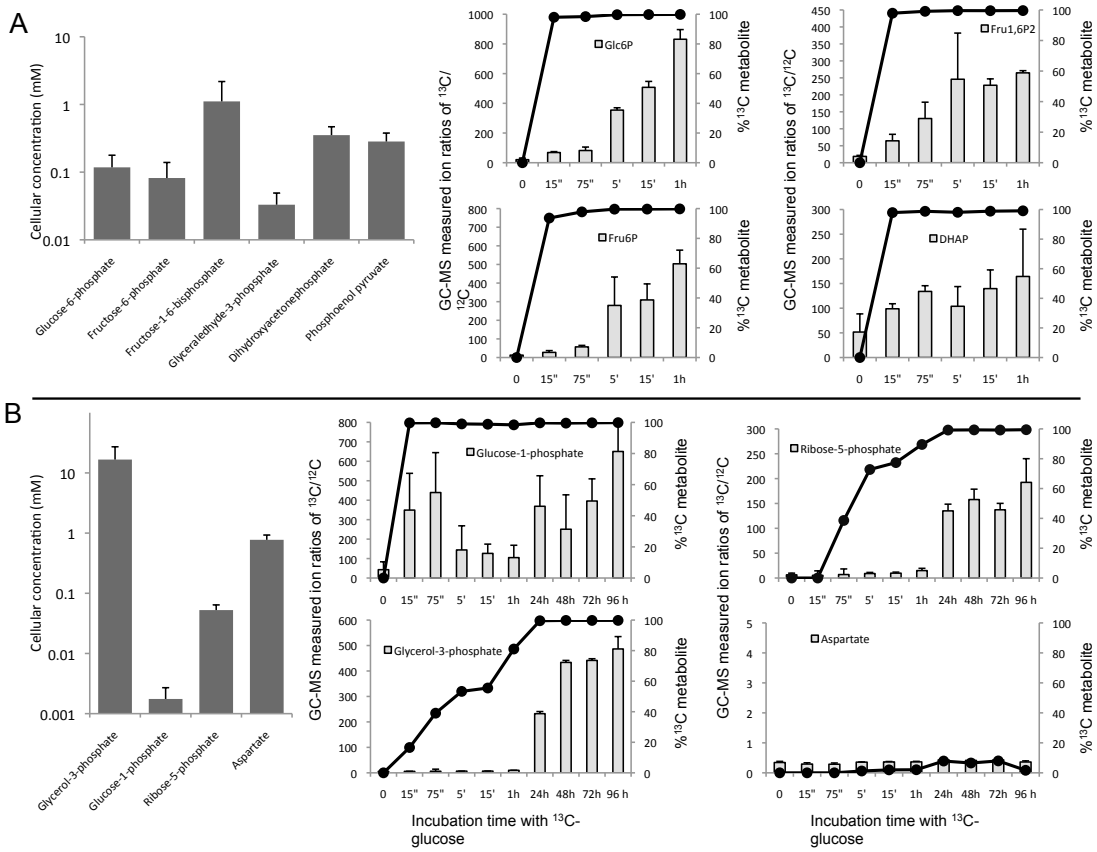

Figure 3.15.: Central carbon metabolism and flux analysis. A: Quantification of glycolytic intermediates and their turnover; light grey bars: $^{13}C/^{12}C$ ratios, black lines: % of labeled metabolite. B: Quantification and flux data for key metabolites in pathways branching off glycolysis.

first outgoing branch from glycolysis. As for glycolytic intermediates the saturation curve for G1P labeling follows a one-phase exponential function (Equation 3.4 and Appendix A, Figure A.7 and Table A.16), reflecting a fast and constant pool turnover. On first sight this contradict model simulations, which predict only very slow incorporation of carbon from G1P into lipids when compared to glycolysis speed (see section 3.3.7 and Appendix A, Table A.15). However, the phosphoglucomutase (MPN066) reversibly converts G1P and G6P, and this process is assumed to be close to equilibrium. In the second branch connecting glycolysis to lipid metabolism, DHAP is converted into G3P, which provides the polar head group for phospho- and glycolipids to which fatty acids are covalently attached (Appendix A, Figure A.1). In contrast to glycolytic intermediates, G3P is highly abundant (5^*10^5 molecules/cell, 16.7 mM) and conversion of DHAP into G3P found to be comparatively slow, reaching saturation not seconds but hours after incubation with ^{13}C-labeled glucose (Figure 3.15B). Accordingly, the saturation curve for G3P labeling, in contrast to G1P and glycolytic intermediate saturation curves, follows a two-phase exponential shape (Equation 3.5 and Appendix A, Figure A.7 and Table A.17).

Before the carbon enters lower glycolysis, it is possibly fed into the pentose phosphate pathway by transketolase (MPN082) and transaldolase (unknown MPN) catalyzed reactions (Appendix A, Figure A.1). R5P, a key intermediate of the pentose phosphate pathway, has a intracellular pool of about 1500 molecules per cell (Figure 3.15B). As for G3P, the incorporation of heavy labeled carbon into R5P is minimal compared to the conversion rates in glycolysis and the saturation curve is best fitted by a two-phase exponential decay function (Equation 3.5 and Appendix A, Figure A.7 and Table A.17). In addition, the synthesis of aspartic acid via oxaloacetate from pyruvate or malate has been suggested for *M. pneumoniae* [Manolukas et al., 1988]. Nevertheless, we could not observe any incorporation of labeled carbon into aspartic acid during four days of batch culture growth (Figure 3.15B). In agreement, none of the enzymes proposed to catalyze such conversion has been detected [Maier et al., 2011]. Thus, a link between central carbon metabolism and amino acid metabolism can be discarded for *M. pneumoniae*, confirming the metabolic network structure.

Integrating quantitative *in silico* fluxes with *in vivo* data on metabolite and enzyme abundances as well as on qualitative carbon flux, provides a comprehensive picture of metabolic activity in *M. pneumoniae* (Figure 3.16). Confirming the massive carbon flux shuttled through glycolysis observed *in silico* and *in vivo*, the glycolytic and pyruvate metabolism proteins belong to the most abundant enzymes encountered in *M. pneumoniae* ranging from hundreds to thousands of copies per cell [Maier et al., 2011].The glycolytic intermediates have been covered to high extent, missing only a few presumably transient intermediates. The in part contrary directions of *in vivo* and *in silico* fluxes connecting glycolysis and other metabolic pathways can be explained by the low abundances determined for the up-take systems of ribose, glycerol, and G3P *in vivo*. In the model, to qualitatively represent experimental findings on glycerol essentiality and the composition of the rich medium for *M. pneumoniae* [Chanock et al., 1962a, Yus et al., 2009], their sources have not been limited to zero as for the other alternative sugar sources. However, due to lacking quantitative information on their uptake by *M.*

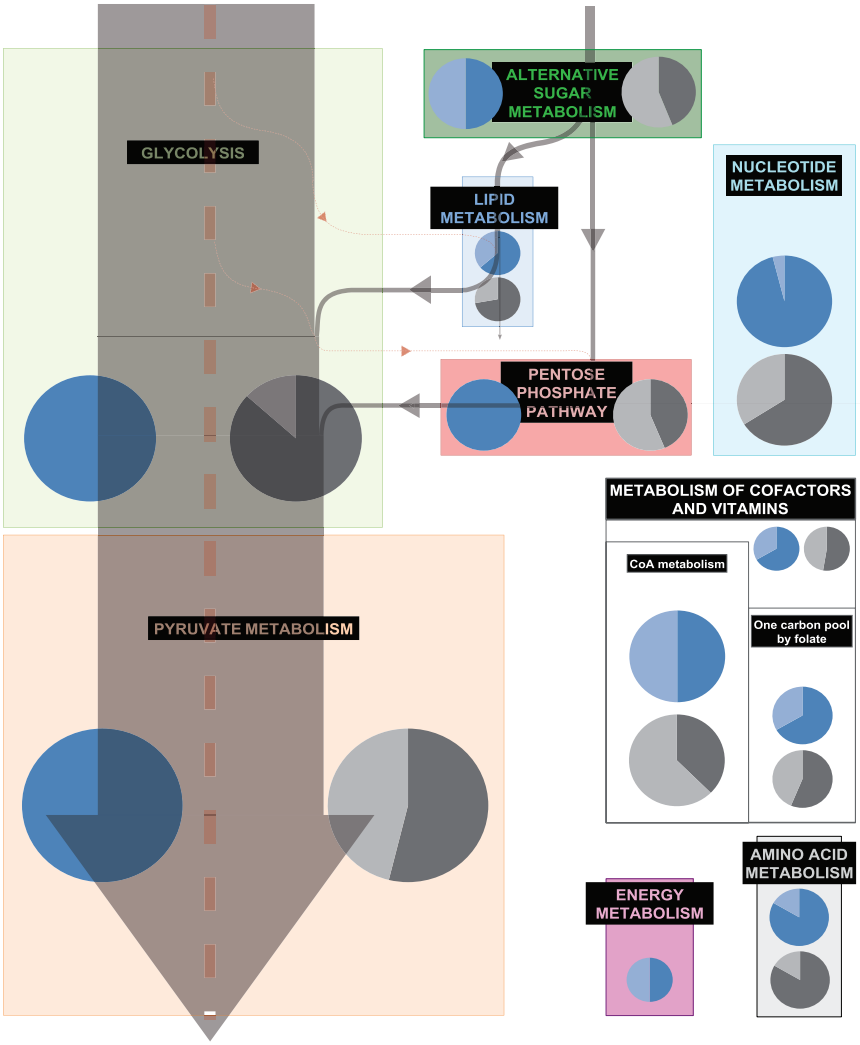

Figure 3.16.: Pathway-based integration of *in vivo* and *in silico* metabolomics data for *M. pneumoniae*. Pathway boxes are sized according to the average protein abundance for the respective pathway. Blue pie plots display detected (dark blue) vs. not detected (light blue) enzymes [Maier et al., 2011] and grey pie plots display detected (dark grey) vs. not detected (light grey) metabolites of the respective pathway. Grey arrows indicate quantitative *in silico* fluxes and the red dashed arrows display qualitative results from the ^{13}C-glucose tracer experiments.

pneumoniae, their usage is probably overestimated *in silico* and the model will have to be adjusted upon availability of new experimental evidence.

Determination of *In Vivo* Catalytic Rates in Glycolysis

Confirming the massive carbon flux shuttled through glycolysis *in silico* and *in vivo*, the enzymes involved in glycolysis and pyruvate metabolism belong to the most abundant enzymes encountered in *M. pneumoniae* ranging from hundreds to thousands of copies per cell [Maier et al., 2011]. Connecting this information to abundances of glycolytic intermediates and to the glycolytic speed determined *in silico* based on *in vivo* uptake rates for glucose, provides an overview about glycolytic activity (Figure 3.17A). Assuming that for reactions with large substrate pools, the effective reaction speed (v_{eff}) approaches the maximum velocity (V_{max}), we calculated *in vivo* catalytic rates for selected glycolytic enzymes (Figure 3.17B) by transforming Equation 1.10 accordingly:

$$k_{cat}(E) = \frac{v_{eff}}{[E]} \tag{3.6}$$

with $[E]$ being the abundance of enzyme E in molecules per cell.

The determined k_{cat} values varied during batch culture growth for all enzymes, usually peaking at 72 hours of growth (Figure 3.17B), coinciding with the maximal glucose consumption rate determined (section 3.3.5, Figure 3.13). It is important to note that *in vivo* k_{cat} values represent apparent turnover numbers and do not represent maximal reaction rates.Phosphofructokinase had the highest k_{cat} after 48 hours of growth ($293^{*}s^{-1}$). Turnover numbers for several glycolytic enzymes could not be determined, either when enzyme abundances far exceeded substrate pools (e.g. for GAPDH, M006) or when substrate metabolites could not be quantified (e.g. for enolase (ENO), M009). Apart from metabolite and enzyme abundances, several additional factors, amongst them allosteric and direct regulation of enzyme activity (regulatory feedbacks) or enzyme affinities and competition for common substrates, have an impact on the effective reaction rates. To gain further insight into those regulatory mechanisms for glycolysis in *M. pneumoniae*, the construction of a dynamic model based on quantitative data and time series following perturbations, ideally on the sub-second time scale, is indispensable. Currently, technical limitations in sample preparations prevent such analyses for *M. pneumoniae* or similar organisms, being 15 seconds the minimum time required to prepare metabolic samples for *M. pneumoniae*.

3.4. Discussion

I present in this chapter a comprehensive systems biology study of the metabolism of *M. pneumoniae in silico* and *in vivo*. To this end, we designed a predictive genome-scale metabolic model, *i*JW145 (comprising 306 reactions that connect 216 metabolites), and experimentally assessed the metabolic space. By integrating model predictions with experimental data obtained in *in vivo* and literature information, we were able to explore the metabolic network of an organism in unprecedented detail. We curated the wiring

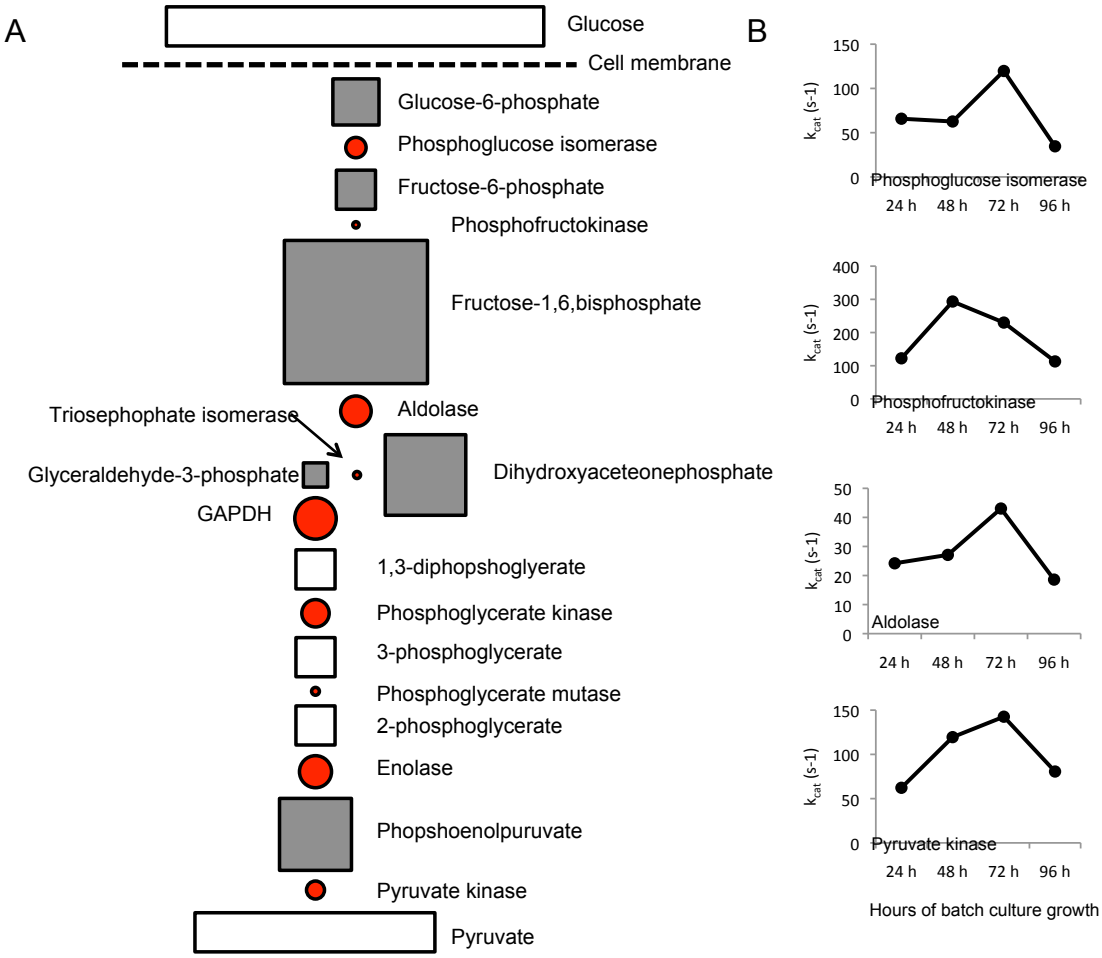

Figure 3.17.: A: Areas of circles (enzymes) and squares (metabolic intermediates) represent intracellular molar amounts at 96 hours of growth. Not detected or not quantified metabolic intermediates are represented by open squares. B: *In vivo* k_{cat} values for selected enzymes with large substrate pools along batch culture growth.

diagram of the metabolic network and the related functional annotations of metabolic key enzymes. Further, we identified and quantified *M. pneumoniae* metabolites, achieving extraordinary high coverage of the *in silico* predicted metabolome. Integrating additional literature data, we semi-quantitatively determined the biomass composition of an average *M. pneumoniae* cell, thus allowing *in silico* growth simulations. This curated network was validated by predicting growth capabilities on alternative carbon sources and gene essentiality based on an *in silico* knock-out study with remarkably high accuracy and specificity. Using the validated model, we theoretically analyzed the energy balancing of *M. pneumoniae* grown in batch culture, finding that *M. pneumoniae* dedicates most of its energy to cellular homeostasis. To analyze changes in the metabolic behavior during the exponential growth phase, we fitted equations to external metabolite measurements and *in silico* predicted maintenance costs. *In vivo*, we monitored the central carbon flux by heavy isotope labeled glucose tracer experiments. Finally, based on protein and metabolite quantification data as well as information on the overall speed of glycolysis *in silico*, we calculated *in vivo* catalytic rates for glycolytic key enzymes.

Metabolic reconstructions have been generated during the past years for various organisms, ranging from prokaryotes, the most prominent being *E. coli*, to different human tissues [Edwards and Palsson, 2000, Duarte et al., 2007, Gille et al., 2010, Rolfsson et al., 2011] (for a list of validated models see Feist et al. [2009], supplementary table 2). The original reconstruction for *E. coli* [Edwards and Palsson, 2000] has been constantly improved integrating new experimental data, for example mRNA expression data or thermodynamic information on reaction reversibilities [Covert et al., 2001, Shlomi et al., 2007, Feist et al., 2007, Fleming et al., 2009, Lewis et al., 2012]. The only mycoplasma for which a metabolic reconstruction was available so far, is *M. genitalium* [Suthers et al., 2009]. However, this reconstruction has been generated automatically from the annotated genome and such automated network generations have been shown to be error prone due to the one-dimensional annotation they are based on [Reed et al., 2006, Henry et al., 2010]. Taking advantage of the small size of *M. pneumoniae* and the large amount of available genome-scale data sets, we directly integrated different experimental data during the metabolic reconstruction process. This allowed us to not only obtain an accurate description of the metabolic network of *M. pneumoniae* but even to correct the network annotation and, subsequently, the functional annotation of related key metabolic enzymes.

Applying a cross-platform approach to analyze the metabolite space of *M. pneumoniae*, we were able to identify the majority of the *in silico* predicted metabolites and to quantify metabolic key components, such as nucleobases, amino acids, fatty acids, and glycolytic intermediates. Integrating determined pool sizes of free cytoplasmic metabolites with their corresponding bound forms and the respective abundances from the growth medium, permitted insight into regulatory mechanisms for the homeostasis of cellular building blocks. One of the most interesting findings of this analysis is probably the discovery that, as already observed for proteins [Maier et al., 2011], the relative abundances of metabolites of a certain pathway or functional group correlate moderately between different organisms [Papagianni et al., 2007, Bennett et al., 2009]. We conclude that different growth environments and the differences in network size and

used metabolic routes do influence on absolute metabolite concentrations but not on the relative ratios, suggesting that metabolism is regulated on the pathway-level rather than on the level of individual reactions. Regulating pathways as functional units presumably greatly simplifies metabolic adjustments. This is in agreement with recent studies on the evolution of flux distributions in bacteria showing that metabolic networks evolved balancing the maximization of biomass and energy production and the minimization of changes required to adapt to environmental changes Schuetz et al. [2012].

We semi-quantitatively determined the biomass composition of an average *M. pneumoniae* cell based on the quantitative metabolite data and literature information [Razin et al., 1963, Razin, 1978, Pollack et al., 1970, Rottem, 1980, Yus et al., 2009, Maier et al., 2011]. This quantitative definition of the *M. pneumoniae* cell composition provides the basis for *in silico* growth simulations for this model organism and possibly other closely related wall-less bacteria. The further integration of doubling times monitored during the exponential growth phase of *M. pneumoniae* grown in batch culture enabled us to quantitatively dissect the energy metabolism *in silico*. Most strikingly, we found that *M. pneumoniae*, at least under laboratory conditions, dedicates only a small part of its generated energy to the production of cellular building blocks. Instead, the quantification of alternative energy consuming processes identified the ATPase to be the major energy sink, using up 57-88% of the total produced ATP to maintain a favorable proton gradient across the membrane and a constant intracellular pH. This surprising finding can be explained by different factors. On the one hand, due to the small size of *M. pneumoniae*, membrane leaking and transport processes across the membrane have a higher impact on the cytoplasmic conditions (*M. pneumoniae* has a surface to volume ratio 2500 times higher than *E. coli*; Appendix A, section A.2 and Table A.18). On the other hand, the acidification of the growth medium through the continuous secretion of organic acids further complicate intracellular pH maintenance of *M. pneumoniae* when grown in batch culture. We verified this hypothesis experimentally and found that artificially re-buffering the pH of the growth medium back to pH7.7, grants metabolic activity until complete depletion of glucose from the growth medium and growth of *M. pneumoniae* colonies beyond 96 hours, when under normal batch culture growth conditions a growth arrest is observed (Appendix A, Figure A.9).

In contrast to the extraordinary high costs caused by the ATPase, the costs for chaperone-assisted protein folding and maintenance are unexpectedly low, especially taking into account their high abundance (chaperones comprise 10% of the total quantified proteome [Maier et al., 2011]). The energy used for GAM amounts to 2-6.9% of the total produced energy and to 2.5-10% of the NGAM costs. Contrarily, in *E. coli*, GAM costs were far exceeding NGAM expenses (∼2.5fold to ∼7fold) [Feist et al., 2007, Varma and Palsson, 1994a]. However, when artificially increasing *E. coli* doubling times to 20 hours, approximating the values observed for *M. pneumoniae*, GAM costs per time decrease amounting to only 8.5% of the NGAM costs and, thus, matching the values determined for *M. pneumoniae*. In summary, we could explain 75-100% of the energy demand of *M. pneumoniae* during the exponential phase of a four days batch culture growth experiment. The only known energy consuming processes not estimated by the presented analysis, are movement and attachment. Surface gliding has been shown to

be an energy consuming process in *M. mobile* [Jaffe et al., 2004] but details on the exact ATP consumption are lacking. Furthermore, the missing expenses could be accredited to i) the assumed mRNA and protein turnover rates (we used average values [Maier et al., 2011]), ii) experimental errors in absolute protein quantifications for single proteins (a twofold error has been reported [Maier et al., 2011]), and iii) the estimation of doubling times (protein quantities up to 36 hours after inoculation are below the detection limit and, thus, supposed to have a higher error: Appendix A, Figure A.2A).

Most bacteria follow a single objective function, the maximization of growth and progeny, in metabolic models mainly accomplished for by the production of biomass [Neidhardt, 1996, Buescher et al., 2012]. Recent studies revealed, that metabolic flux states in addition evolved to minimize the adjustment costs upon environmental changes [Schuetz et al., 2012]. Integrating our metabolic model with the accomplished energy calculations allowed the quantitative balancing of the energy metabolism in *M. pneumoniae*. The comparative analysis of energy expenses in *M. pneumoniae* and *E. coli* revealed fundamental differences, suggesting characteristic energy expense profiles for different bacteria. Three parameters have been identified to govern the composition of those energy profiles, namely the network topology, the cell size, the growth rate, and the environmental conditions. We conclude that the large fraction of energy dedicated to cellular homeostasis and the associated slow growth of *M. pneumoniae* in batch culture reflect the adaptation to the growth on human lung epithelial cells, a largely unchanging environment.

Gene essentiality has been predicted for other organisms, such as *E. coli*, based on *in silico* knock-out studies [Reed and Palsson, 2003, Feist et al., 2007]. We reach slightly higher accuracy and specificity with our prediction for *M. pneumoniae* than has been achieved for *E. coli* so far [Feist et al., 2007], thus proving the high accuracy of the metabolic reconstruction and the predictive capacity of the model. In consistency with the fact that *M. pneumoniae* has an exceptionally high fraction of essential metabolic genes (56.6% vs. 19% in *E. coli* [Baba et al., 2006, Joyce et al., 2006]), we found that rescue routes to buffer for loss of function causing gene deletions are only preserved in the nucleotide metabolism and the pentose phosphate pathway. In agreement with the results from the energetic analysis, this finding suggests that the reductive genome evolution process of *M. pneumoniae* eliminated most metabolic rescue pathways, while adapting to a parasitic life in the human lung.

Synthetic genetic array analysis has been used to automate the isolation and analysis of double mutants *in vivo* [Tong et al., 2001, 2004]. Predicting double mutant phenotypes *in silico*, allowed to unravel the combinatory effects caused by different gene deletions on the metabolic behavior and to further analyze the remaining adaptive capabilities of *M. pneumoniae*, confirming the single knock-out results. The applied *in silico* analysis provides an auspicious alternative to experimental approaches, especially for the analysis of organisms lacking appropriate *in vivo* analysis tools.

Since static modeling approaches, such as constraint-based modeling, by definition cannot provide information about network dynamics, models based on ODEs are usually employed for time-dependent simulations [Klipp et al., 2005]. However, dynamic models have limited applicability for large-scale networks, mainly due to overfitting [Draper and

Smith, 1998]. To overcome the limitations of the static approach while avoiding the problem of overfitting, we applied non-linear fittings to experimental data and *in silico* results. The resulting equations allow the calculation of growth constraints for every time point of the exponential growth phase of *M. pneumoniae* grown in batch culture and, subsequently, the analysis of changes in the metabolic behavior between those time points. Albeit not granting the extraction of kinetic parameters for the underlying biochemical reactions, this approach facilitates general information about network dynamics without requiring to determine the exact kinetics for the entire network.

Integrating the predicted flux distributions with ^{13}C-glucose tracing results as well as metabolite and protein quantification data, provided a qualitative picture of metabolic pathway activity in *M. pneumoniae*, as well as quantitative understanding of network connectivity. Furthermore, we could not only estimate the velocities of different metabolic routes, but calculate *in vivo* kinetic parameters for key glycolytic enzymes.

Presumably the most interesting general finding of the growth simulations is that oxygen consumption is tightly coupled to acetic acid production *in silico* under all simulated conditions. This prediction agrees with findings in *L. lactis*, for which the limited oxygen availability at later growth stages has been shown to prevent counterbalancing the redox imbalance associated with acetic acid production, equilibrated by releasing the lactate dehydrogenase from its supposed oxygen-dependent inhibition [Gottschalk, 1986, Neves et al., 2005]. Therefore, we propose that oxygen could have a similar regulatory role on pyruvate metabolism in *M. pneumoniae* leading this organism to switch from mainly acetic to mainly lactic acid fermentation during a four days batch culture growth experiment (this study, Yus et al. [2009]).

Summing up, the iterative integration of *in silico* results with *in vivo* data and other information granted understanding of the examined metabolic system beyond the static limitations of the mathematical approach. The presented metabolic model, *i*JW145, allows to predict metabolic behavior where it is not amenable to experimental analysis, for example due to lacking *M. pneumoniae* specific analysis tools. The extracted biological findings enlighten functional mechanisms of *M. pneumoniae* metabolism but also pinpoint open questions that remain to be investigated. In addition, the examinations with respect to the central energy producing carbon metabolism will facilitate the design of a dynamic model for glycolysis to further investigate underlying regulatory mechanisms controlling growth performance of this minimal bacteria.

4. Genome Re-annotation for *Mycoplasma pneumoniae*

In this chapter, the genome re-annotation of *M. pneumoniae* is presented. For publication the results will be integrated with a genome-wide essentiality study currently under development in our group. The manuscript to be submitted to MSB is still in preparation, but the combined project was already priced with the IOM Luis Denis Award (IOM - International Organization for Mycoplasmology) on the IOM congress 2012, 15-17/07, Toulouse, France.

I designed the theoretical peptide library for M. pneumoniae, analyzed the proteomics data, and conducted the bioinformatic analyses integrating in vivo and in silico results. Furthermore, I was involved in the project design and development and I participate in the manuscript writing as well as figure and table design.

4.1. Introduction

The interest in the minimal gene set sustaining life increased since the first genome sequences for minimal organisms became available. According to Koonin [2003], essential genes of an organism are defined as the minimum gene complement sustaining growth and cell division under the most favorable environmental conditions. In theoretical approaches, the number of essential genes has been predicted to be around 240 [Gil et al., 2004, Shuler et al., 2012]. Experimentally, gene essentiality has been studied in different organisms, including *M. genitalium* and *M. pulmonis* [Hutchison et al., 1999, Glass et al., 2006, Akerley et al., 2002, Jacobs et al., 2003, French et al., 2008, Langridge et al., 2009, Dybvig et al., 2010, Christen et al., 2011, Griffin et al., 2011]. However, the operon structure of the bacterial genome and the self-transposition of active transposons so far prevented the successful determination of the minimum essential genome sustaining life. In addition, the existence of antisense RNAs and other putative regulatory regions in bacterial genomes [Lluch-Senar et al., 2007, Yus et al., 2012] suggest that even in the simplest organism the genome is carefully regulated, leading to the essentiality of not only genes but also other non-coding regions. Finally, the high error rates reported for genome annotations [Casari et al., 1995, Brenner, 1999] pose another challenge within the aim to define the minimal genome able to sustain life of a cell.

Aiming to provide a thorough basis to decipher the essential functional genome of *M. pneumoniae*, we manually curated the genome annotation. To this end, we combined a peptide library obtained from translating all peptides encoded theoretically by the *M.*

pneumoniae genome with mRNA expression [Güell et al., 2009, 2011] as well as protein sequencing and MS results (this study) (Figure 4.1).

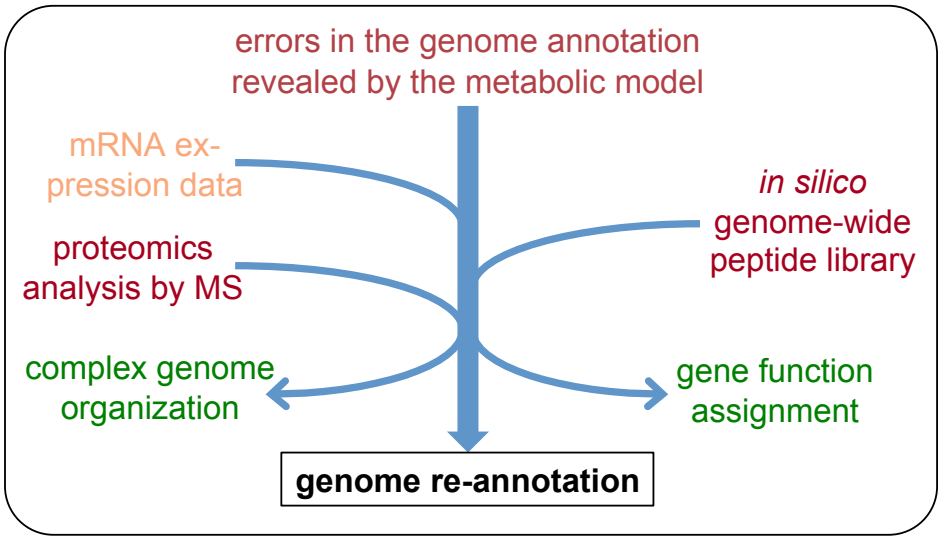

Figure 4.1.: Workflow of the *M. pneumoniae* genome analysis. 1: re-annotation of the genome based on integration of a theoretical peptide library, mRNA expression and protein MS data; 2: genome essentiality study based on HITS sequencing of a minitransposon mutant library and the definition of a distribution score (DSE) that allows to define essentiality for all genomic regions. Accomplished analyses integrated into the study have red fonts, literature data and the mutant library obtained previously by "haystack mutagenesis" [Halbedel and Stülke, 2007] have orange fonts, and results of the study have green fonts.

4.2. Material and Methods

4.2.1. Computational Procedures

Peptide Library Design

The three reading frames of the two strands of the genome of *M. pneumoniae* (reference genome: *M. pneumoniae* M129, NC_000912, NCBI) were translated *in silico* into peptides as described by the pseudocode in Appendix B, section B.1. The resulting peptide library contained all maximum length peptides encoded on the *M. pneumoniae* genome. Next, we applied a minimum length filter, discarding all predicted peptides shorter than 19 amino acids, resulting in a library containing a total number of 12,426 peptides (data not shown). This library has been used for the analysis of the MS results, to assure that determined unique peptides do not additionally match putative proteins shorter than 50 amino acids. The second minimum length filter, discarding all predicted peptides shorter than 50 amino acids except the two annotated proteins MPN188 (37aa) and MPN682 (48aa), resulted in a library of 3,748 peptides used for the result analysis (data

not shown). Annotated *M. pneumoniae* proteins have been matched to the peptides in the library and a list of 151 putatively longer proteins has been generated including those proteins that do not match to the entire sequence but only to the C-terminus of a maximum length peptide (Appendix B, Table B.1).

4.2.2. Experimental Procedures

Protein Identification and Quantification by Mass Spectrometry

M. pneumoniae M129 cells were grown in two T75 flasks (75 cm^2) in Hayflick media at 37°C. At 6 and 96 hours after inoculation, the attached cells were washed three times with PBS and lysated with 200 μl of urea 8M. 10 μl of the lysate were loaded in two NuPAGE 4-12% Bis-Tris and in one Tis-Glycine gels (Invitrogen). The different lanes were divided into 16 bands of approximately 15 mm. Proteins were identified and quantified by MS analysis from the bands and from unfractionated solution samples, using the designed peptide library for the result analysis.

4.3. Results

4.3.1. Transcriptome Analysis

Using transcriptomics data [Güell et al., 2009] and information about a recently discovered new class of short RNAs which precisely map the transcriptional start sites (TSSs) of *M. pneumoniae* genes (TSS RNAs) [Yus et al., 2012], we determined small RNAs (each identified by a *M. pneumoniae ID* short (MPNs) ID), TSSs, promoter sequences, and 5'-untranslated regions (5'-UTRs) for the *M. pneumoniae* genome (Appendix B, Table B.1). For 32 annotated genes, the experimentally detected TSS was located downstream of the annotated translational start codon (TSC), suggesting an erroneous annotated TSC. In all cases, the resulting transcript contained an alternative TSC in proximity to the TSS, giving rise to either a shorter version of the annotated protein or a completely different protein encoded in one of the two other reading frames (Figure 4.2A and Appendix B, Table B.1). According to multiple sequence alignments with ClustalW [Larkin et al., 2007], for six of those proteins exist orthologous short proteins in other organisms (Appendix B, tabletab:reannotated). Another 86 genes are found to contain multiple TSSs giving rise to different transcripts from which either size-differing versions of the annotated protein or completely different polypeptides can be translated (Figure 4.2A and Appendix B, Table B.1). Furthermore, for a total number of 152 genes we detected 5'-UTRs bigger than 40 nucleotides. Those large 5'-UTRs do possibly contain regulatory elements or even encode for small peptides, previously not considered [Rasmussen et al., 2009]. 25 out of the 152 large 5'-UTRs presumably extend the annotated ORF, 29 could encode for proteins of at least 50 amino acids length previously not annotated, and the remaining 98 could be involved in translational regulation (Appendix B, Table B.1). Of the 25 putative longer proteins, for 12 the additional amino acid sequence was common to orthologous proteins in other bacterial species (Appendix B, Table B.2). Translating

all peptides theoretically encoded by the detected transcripts predicted the putative existence of 115 previously not annotated protein coding genes. Finally, we identified 302

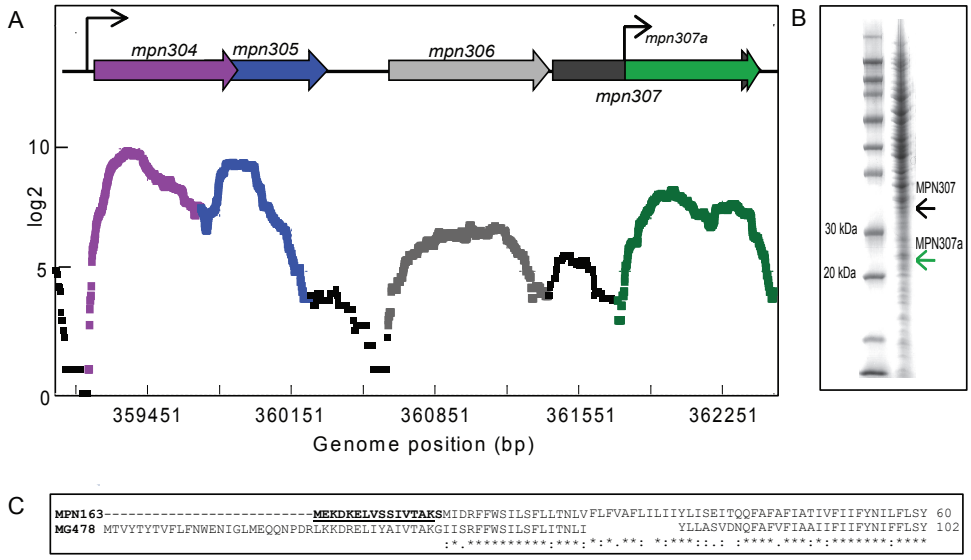

Figure 4.2.: Examples from the genome re-annotation highlighting the genomic complexity: A: mRNA expression profiles for genome positions 359000-362600, suggesting the expression of two different genes, *mpn307* and *mpn307a* from the annotated and an additional internal TSS of *mpn307*. B: Confirmation of the expression of two proteins of the expected size from *mpn307* and *mpn307a*. C: Alignment of MPN163 and MG478: *mpn163* is expressed with a different N-terminus, larger than previously annotated based on sequence similarity (bold and underlined amino acids indicated the additional protein part detected by MS).

new so-called short RNAs (MPNs IDs) of which about 33% are longer than 250 base pairs and could code for previously unknown proteins.

4.3.2. *In Silico* Translation of the genome

To account for gene products which are not detected experimentally due to low or time-dependent expression, we designed a comprehensive peptide library for *M. pneumoniae*. To this end, each of the three possible reading frames of both strands of the genome was scanned for the first TSC (TTG, GTG, ATG), the genome was translated until the next stop codon and scanned again for the next TSC and so on. Thus, for every frame only the longest possible version of each protein was considered for the peptide library. In addition, we applied a minimum length filter of 50 amino acids, discarding all shorter peptides for the following analysis.

All together, the peptide library contains 4.436 peptides, including the 689 annotated protein-coding genes, 690 ORFs coding for short RNAs, and 3.057 putative new ORFs (data not shown). For 151 genes, the annotated sequence matches only the C-terminal

part of a library peptide, suggesting the possible existence of a longer version or larger isoform of the same protein (Appendix B, Table B.1).

4.3.3. Integration of *In Silico* and *In Vivo* Data

Based on the integration of this peptide library with the information about genome positions of non-coding regions, as well as an analysis of the *M. pneumoniae* proteome, we re-annotated the genome (Appendix B, Table B.1). Therefore, to validate the predictions from the transcriptomics analysis and the peptide library, we determined the proteome of *M. pneumoniae* applying MS to non-fractionated total protein extracts obtained after 6 and 96 hours of growth in batch culture. In total, 516 of the 689 annotated proteins were identified (Appendix B, Table B.3), showing high overlap with a previous protein identification studies in *M. pneumoniae*: a) 93% overlap (557 proteins have been detected) [Jaffe et al., 2004] and b) 100% out of 414 quantified proteins by Maier et al. [2011]. All genes with significant expression level of full-length transcripts (average $log_2 > 11$ for tiling arrays and average $log_2 > 13$ for deep sequencing results) have been detected, while for proteins with lower expression level 71% have been identified, confirming high coverage of the putatively expressed proteome.

We fractionated protein extracts by high resolution SDS gel electrophoresis and analyzed the different fractions separately by MS (Figure 4.2B), to further evaluate the prediction of putative longer or shorter proteins. Proteins have been detected for 84 out of the 151 genes possibly coding for longer proteins (Appendix B, Table B.3). However, to confirm the existence of the larger isoform, unique peptides matching the additional peptide sequence have to be identified. This was the case for five proteins (MPN006, MPN148, MPN163, MPN388, and MPN664), confirming that these proteins indeed are expressed in a longer form than previously annotated (Figure 4.2C). In addition, for six out of 34 genes found to have internal TSSs, we could prove the expression of two proteins of different size: MPN310 (200 and 19 kDa), MPN130 (16,5 and 10 kDa), MPN410 (17,5 and 10 kDa), MPN073 (44 and 38 kDa), MPN196 (27 and 6,5 kDa), and MPN307 (33 and 20 kDa) (Figure 4.2A & B and Appendix B, Table B.3). The two isoforms for MPN310 had been previously described by Boonmee et al. [2009].

Analyzing the 32 genes with an annotated TSC outside the experimentally determined transcript, revealed that 20 code for hypothetical proteins, 12 of those showing high similarity to the functionally uncharacterized *mpn013* (Appendix B, Table B.2: underlined gene IDs). In all 20 cases, the TSS gives rise to an ORF lacking the same fragment of the putative protein. Unique peptides have been identified for 8 out of the 13 supposed homologs of *mpn013*, but no unique peptide corresponding to the fragment between TSC and TSS (56 aa length) could be determined. We conclude, that at least in *M. pneumoniae* this protein family is shorter than annotated.

We identified 33 new proteins of which 19 are derived from an internal TSS of a previously annotated gene (Appendix B, Table B.1). Three of the new proteins, MPN060a, MPN198a, and MPN394a, are only expressed during the stationary growth phase (Appendix B, Table B.3). Seven new proteins are homologs of other *M. pneumoniae* proteins, respectively located in a different region of the genome. Additionally, another eight new

proteins have homologs in other mycoplasma species or other *M. pneumoniae* strains (Appendix B, Table B.2), most of them coding for hypothetical proteins.

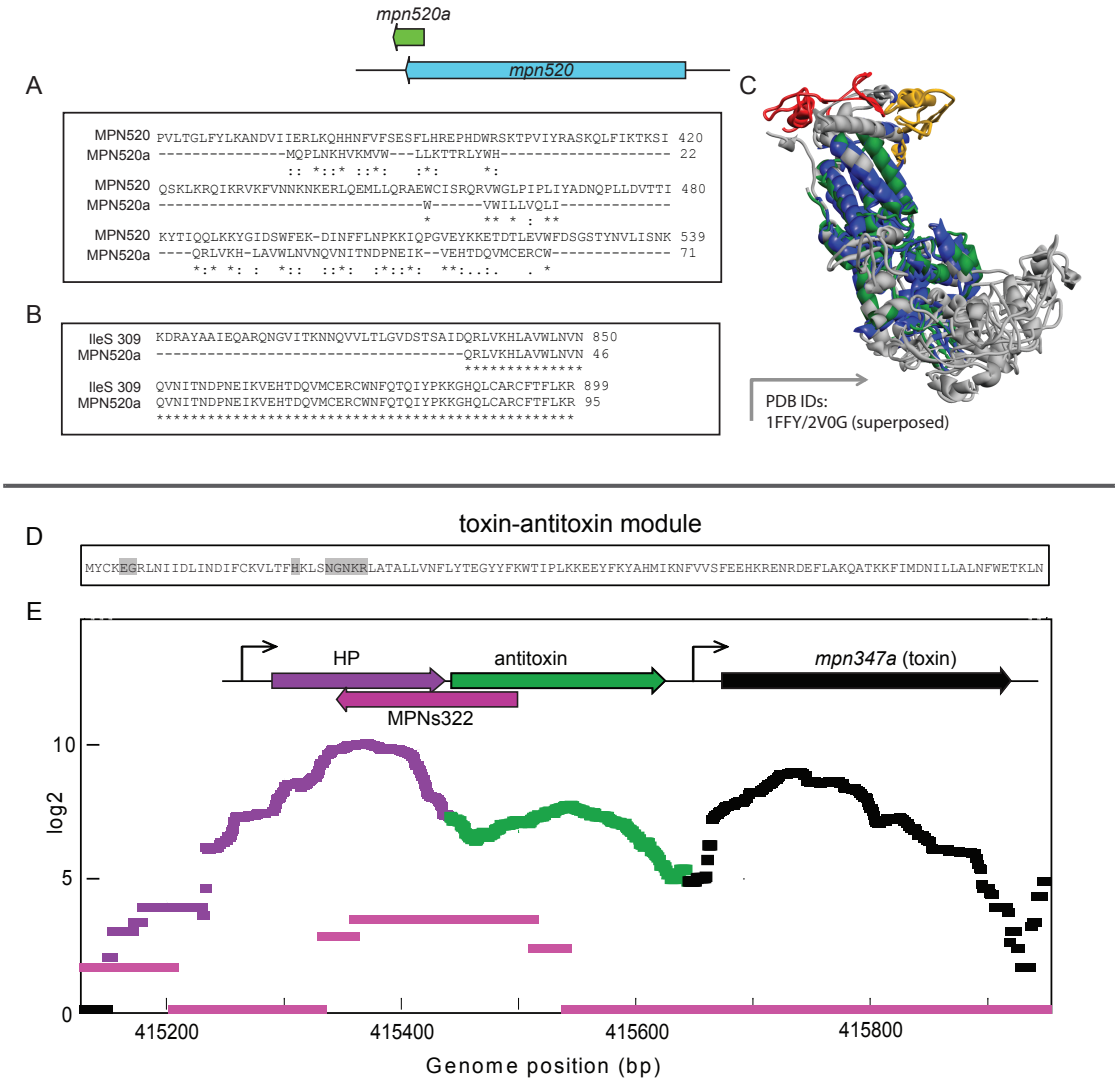

Figure 4.3.: Functionality analysis for new *M. pneumoniae* genes. A-C: *mpn520a*, A: Alignment of MPN520a to MPN520, showing low sequence similarity. B: Alignment of MPN520a to the isoleucyl-tRNA synthetase of *M. pneumoniae* 309 (IleS 309), showing a perfect match for the C-terminal of IleS. C: Superposed protein structures for two tRNA synthetases (IFFY/2V0G) [Silvian et al., 1999, Rock et al., 2007]; D-E: *mpn341a*, D: Conserved residues of the active center of toxin proteins detected in MPN347a. E: mRNA expression profiles showing expression i) of an antisense RNA, MPNs322, overlapping with two protein coding genes, which are expressed at the same time and ii) of the newly identified toxin (*mpn347a*).

For example, MPN520a, expressed from an ORF located in an intergenic region, is highly similar to the C-terminal region of isoleucyl-tRNA synthetase of the M. pneumoniae 309 strain, but not to the isoleucyl-tRNA synthetase of M. pneumoniae M129i (Figure 4.3). This finding suggests an additional function for this C-terminal region of the *M. pneumoniae* 309 isoleucyl-tRNA synthetase, that in *M. pneumoniae* 129 is accounted for by MPN520a. Interestingly, MPN347a, also expressed from an intergenic region, contains the same active site as the toxic death-on-curing (DOC) protein of phage P1 [Garcia-Pino et al., 2008] (Figure 4.3D). In *E. coli*, DOC is part of an operon containing two counteracting proteins known as toxin-antitoxin or addiction module [Buts et al., 2005]. In *M. pneumoniae*, upstream to MPN347a locates an ORF encoding the antitoxin (Figure 4.3E) but the protein, which has been shown to be very unstable and difficult to detect [Buts et al., 2005], has not been identified by MS (Appendix B, Table B.3). However, an antisense RNA (MPNs322) for the antitoxin has been identified by ultra-sequencing and tiling arrays [Güell et al., 2009] (Figure 4.3E). This antisense RNA could inhibit the translation of the antitoxin favoring toxin activity.

About two thirds of the new coding ORFs (21) are located in transgenic regions, suggesting that the same region of the genome can codify for two different proteins. *Mpn037a*, for instance, is a cis-antisense RNA of *mpn037* and contains two alternative TSSs putatively originating two size-differing versions of the same protein (11 and 13 kDa). In addition, *mpn037a* has high sequence similarity to *mpn139*, coding for a hypothetical protein of *M. pneumoniae*. We identified six unique peptides for MPN037a, proving that both strands of the same genomic regions can code for two different proteins expressed under the same conditions. This finding poses a formidable challenge on gene essentiality studies in bacteria, because genes could be wrongly considered essential due to the existence of an overlapping essential ORF on the opposing strand.

In summary, the integration of transcriptomics and proteomics data with the theoretical peptidome encoded in the *M. pneumoniae* genome allowed to identify 115 putative new ORFs (33 validated), to change the length of 44 proteins (12 longer and 32 shorter), to identify 53 proteins with putative isoforms of different length (6 validated), and to describe 126 new MPNs encoding ORFs.

4.4. Discussion

Despite the fact, that the genome of *M. pneumoniae* has been sequenced twice [Himmelreich et al., 1996, Dandekar et al., 2000], we were able to not only correct the annotation of 44 genes but to also identify and verify 33 new protein coding genes, thus correcting an error rate of about 10.7%. This rate even exceeds previous estimations for automatically derived genome annotations, amounting to 8% in *M. genitalium* [Brenner, 1999]. The integration of the theoretical peptide library with mRNA expression data [Güell et al., 2009, 2011] and proteome characterization, in addition, allowed to identify six genes from which different isoforms of the same protein are expressed from differing TSSs which could not have been detected based on the genome sequence alone. Finally, we discovered a total number of 126 new sRNAs.

Carefully analyzing the newly detected protein-coding genes enabled us to unravel the probable function of one of the newly annotated genes, showing high sequence similarity between the respective gene and the toxin of the toxin-antitoxin module in *E. coli*. Our results show that for an accurate genome annotation it is indispensable to consider experimental data, preferentially of different cellular entities, such as mRNAs and proteins, and obtained under different growth condition, as well as the theoretical coding capacities of the genome.

Interestingly, we found that the reduced genome of *M. pneumoniae* is not at all simply structured but in contrast highly compact. In addition to the high number of small RNAs, many of them antisense to protein coding ORFs, the genome even contains regions where both strands code for proteins found to be expressed under the applied growth conditions. These findings have the potential to revolutionize understanding of bacterial genomes, which are putatively structured and regulated in a far more complicated manner than previously assumed. For *M. pneumoniae*, we propose that the multitude of sense and antisense RNAs (MPNs) found to overlap with coding genes are involved in the regulation of mainly those overlapping genes and putatively others.

To understand life, even if only of a single-cellular organisms, one has to determine the functions and their related genes essential for survival of the cells. To study gene essentiality in bacteria, which are much smaller and simpler than eukaryotic cell, in general genome-wide mutagenesis studies are accomplished despite being not entirely accurate due to several technical reasons. However, all attempts to define the minimal essential genome of a cell has not been successful so far. One possible reason is the applied focus on protein and functional RNA encoding genes, thereby missing other, non-translated genomic regions. The detailed annotation of the *M. pneumoniae* genome is currently used to define essential genomic regions, independent of localization, coding capacities, and function. Upon finishing this *in vivo* essentiality study, the two project will be merged into a single manuscript for publication.

5. Summarizing Discussion and Concluding Remarks

Objectives and Summary

Understanding life of an entire cell is a highly ambitious goal - at the current state of technology and knowledge not amenable for eukaryotic cells. However, advances in computer technology and high throughput analysis methods bring this goal into reach for some of the most simple prokaryotes, mycoplasmas. The objective of this thesis was to contribute to a collaborative systems biology research project to comprehensively explore the minimal bacterium *M. pneumoniae* by conducting mathematical and computational analyses integrating high-throughput *in vivo* and *in silico* data. To this end, different sub-projects have been carried out, relating information on diverse aspects of *M. pneumoniae*, such as the cellular composition, metabolic behavior, the proteome organization, as well as genome structure and gene expression.

In science, the exchange of information is of upmost importance in order to produce further knowledge gain. To allow fast and easy exchange of information within the mycoplasma project and the scientific community, to obtain an overview about the available data and to classify it according to putative utility for the design of mathematical models, we developed and implemented a database for *M. pneumoniae*, **MyMpn**. Apart from data access, the database interface allows to further integrate new and unpublished data with the database content, providing advanced access to members of the mycoplasma project. The related data analysis revealed significant gaps in our understanding of the *M. pneumoniae* cell composition and the functional mechanisms governing metabolism, that in part could be addressed by a metabolic model.

As the main project of this thesis, we developed a predictive genome-scale constraint-based model for the *M. pneumoniae* metabolism, *i*JW145, based on a metabolic map previously designed in our group [Yus et al., 2009]. In addition, we qualitatively and quantitatively explored the metabolite space of *M. pneumoniae in vivo*, thereby assessing regulatory mechanisms for intracellular metabolite pools. This enabled us to semi-quantitatively define the biomass composition of an average *M. pneumoniae* cell, thus providing the basis for *in silico* growth simulations for *M. pneumoniae* and putatively other wall-less bacteria. Simulating growth with our metabolic model allowed us to correct the network annotation, to unravel the energy balancing of *M. pneumoniae* grown in batch culture, and to characterize the metabolic behavior during the exponential growth phase as well as in diverse mutant phenotypes. The iterative integration of *in silico* and *in vivo* results, as well as the collaborative effort put into experiment and model design, provided extraordinary insight into the metabolic behavior of *M. pneumoniae* which in addition to the reduced genome also reflects the high degree of adaptation

to parasitic life in the human lung.

Analyzing the metabolism of *M. pneumoniae*, amongst other findings, pointed out several errors in the genome annotation, both in the annotated ORFs and in the functional annotation of specific genes. Integrating information on the theoretical coding capacities of the genome sequence with mRNA as well as protein sequencing and MS data, enabled us to re-annotate the *M. pneumoniae* genome in unprecedented quality and detail. Our results prove the importance of the validation of automatically generated genome annotations by preferentially complementing genome-scale experimental data. Additionally, they suggest that even the most simple bacterial genomes are carefully organized resulting in complex regulatory mechanisms for gene expression.

The results from the different presented projects as well as from other ongoing studies will be incorporated into the **MyMpn** database, thus adding up to the scientific attempt of understanding *M. pneumoniae* in full quantitative detail.

State of the Art in *M. pneumoniae* Research

When I joined the mycoplasma project, several large-scale data sets had been produced, making available information on gene expression, protein-protein interactions, and the metabolic network of *M. pneumoniae* [Güell et al., 2009, Kühner et al., 2009, Yus et al., 2009]. Other studies on transcriptional regulation [Güell et al., 2011], proteomics [Maier et al., 2011], post-translational modifications [van Noort et al., 2012], and a new class of short RNAs (TSS RNAs) [Yus et al., 2012] were in preparation. However, a proper system to share and exchange the information amongst the different groups and to standardize their analysis was missing.

Our database provides the previously lacking central data storage for the mycoplasma community and significantly facilitates information exchange between the different research groups. Several of the analysis tools, as for example the genome browsers or the interactive metabolic map have been built by adapting available tools to *M. pneumoniae* and the necessities of the different scientists involved. The statistical analysis tools and the *Mycoplasma pneumoniae* genome browser can be applied to unpublished data since they allow to temporarily upload data that is not further stored or included in the database. Despite being still under development and thus not released to the public, the database is already used by the different groups of the mycoplasma project.

With respect to the understanding of *M. pneumoniae*, the database design process highlighted the requirement of a genome-scale metabolic model to further unravel underlying functional and regulatory mechanism. For the design of a dynamic model, describing all metabolic components (e.g. metabolites and catalyzing enzymes) and their biochemical relations in a time-dependent manner, the available experimental data, especially for metabolites, was still too sparse, suggesting the application of a static modeling approach, such as constraint-based modeling [Varma and Palsson, 1994b].

The metabolic network of *M. pneumoniae* had been reconstructed by integration of gene function assignments with growth curve measurements and subsequent manual gap filling based on database information on metabolic pathways [Yus et al., 2009]. To validate the reconstructed network and to further understand metabolism in *M. pneumoniae*,

the different metabolic components had to be identified, if possible quantified, and then placed into the context of their dynamic network.

Based on the available metabolic network reconstruction [Yus et al., 2009], we designed a genome-scale constraint-based model with predictive capacities: *i*JW145. The presented integration of this mathematical model with metabolite identification and quantification data as well as a broad literature screening allowed to draw a quantitative picture of the *M. pneumoniae* metabolism despite the application of a static modeling approach. Thereby, the combination of complementary experimental techniques facilitated the detection of a substantial fraction of the metabolome, reaching higher coverage of the *in silico* predicted metabolome than obtained for any other organism [Soga et al., 2003, van der Werf et al., 2007, 2008, t'Kindt et al., 2010, Liebeke et al., 2011]. The determination of the biomass composition of a cell wall-less bacteria, that significantly differs from that of higher prokaryotes, will serve as a basis for growth simulations in mycoplasmas and related organisms. By an in its complexity unprecedented iterative integration of *in silico* and *in vivo* results, we were able to unravel general principles governing the energy homeostasis in *M. pneumoniae.* In general, the metabolism of *M. pneumoniae* appears to be tightly adapted to growth in the constant environmental conditions it encounters in its natural habitat.

The extracted findings, such as the determined *in vivo* catalytic rates, are employed in the design of dynamic models for metabolic sub-systems as well as for the design of engineering tools for genetic manipulation of *M. pneumoniae* currently under development in our group. Besides, the correction of the functional annotation of metabolic key enzymes, revealed that important errors remain in the genome annotation of *M. pneumoniae* even after two annotation rounds [Himmelreich et al., 1996, Dandekar et al., 2000].

The determination of the minimal set of functions required to sustain life is unquestionable one of the important challenges in current biological research attracting the attention of many different research groups [Hutchison et al., 1999, Gerdes et al., 2003, Kobayashi et al., 2003, Glass et al., 2006, French et al., 2008, Christen et al., 2011]. *M. pneumoniae*, due to its reduced genome and the large amount of genome-scale data sets on transcriptomics and proteomics, is an ideal organism to tackle this question. Still, the high error rates reported for genome annotations in general [Casari et al., 1995, Brenner, 1999] and the deficits in the annotation of *M. pneumoniae* pinpointed by the metabolic analysis constituted a significant obstacle towards this aim.

We showed that it is indispensable for a high-quality annotation of a genome to complement the annotation obtained automatically based on sequence similarity with high-throughput *in vivo* data and the theoretical genomic coding capabilities. Our approach allowed us to correct the annotation of more than 10% of all protein-coding genes, including the 33 newly detected, and to annotate 126 new sRNA encoding ORFs.

This analysis reveals higher complexity of bacterial genomes than previously assumed, proving the existence of i) overlapping protein coding genes on opposing strands and ii) different protein isoforms expressed from the same gene (validated for six ORFs) in bacteria. In addition, this accurate genome annotation provides a thorough basis to assess

the minimal essential genome sustaining life of *M. pneumoniae*.

Advantages of the Collaborative Research Approach in Systems Biology
All projects presented in this thesis have been accomplished in collaboration with researchers from different natural sciences, each contributing by adding his or her field-specific knowledge and personal views. Thus, each of the involved scientists could profit from the expertise of the different colleagues, learning about highly diverse topics, the advantages and disadvantages of experimental and computational research, as well as a multitude of methodologies. This is of major importance, since one of the key aims of systems biology is to incorporate the perspectives of different scientific disciplines, thus providing the possibility to study a biological system as a whole and from different angles [Auyang, 1999, Kitano, 2002a, Oshry, 2007]. To this end, it is indispensable that researchers learn to communicate their work to people from different areas, not familiar with many of the field-specific terms and definitions, thereby establishing a common language. As a result, the systems biology approach, when applied at all levels of a research project, grants a knowledge gain far beyond the sum of the results of each single sub-project. Integration of different data sets with each other allows to complement for limitations in the applied techniques, both for the same type of data (i.e. transcriptomics, proteomics, metabolomics) and for the relation of different cellular levels.

As detailed in the different results chapters, the accomplished analyses and, especially, the iterative integration of the different experimental and computational results, enabled us to obtain unprecedented insight into a biological system, the minimal bacterium *M. pneumoniae*. The developed database will allow to further improve the collaboration amongst the different research groups of the mycoplasma project and possibly attract the interest of other scientists into this fascinating model organism for systems biology. Furthermore it provides an ideal starting point for the development of mathematical models for *M. pneumoniae*.

Mathematical modeling, especially if many different large-scale data sets for an organism are available, as in case of *M. pneumoniae*, provides a powerful tool to understand a system as one big entity, rather than as a collection of sparsely connected sub-systems. Recently, a whole-cell model has been presented for *M. genitalium*, describing the life cycle of the smaller relative of *M. pneumoniae* and deciphering previously unknown functions [Karr et al., 2012]. For *M. pneumoniae*, the metabolic model provided insight not only into the network structure but also into the regulation of metabolic activity under different conditions, especially with respect to energy usage. When compared to the model building process in *E. coli*, where the initial metabolic model [Edwards and Palsson, 2000] has been improved by several different researchers over the past decade [Covert et al., 2001, Shlomi et al., 2007, Fleming et al., 2009, Lewis et al., 2012], in case of *M. pneumoniae* this process was much faster, which can be attributed to the small size and the resulting comparatively high coverage of the predicted transcriptome, proteome, and metabolome by experimental analyses.

Despite being the probably best studied model organism for prokaryotes in the biological sciences, *E. coli* still is quite complex for a bacterial cell when compared to minimal organisms such as mycoplasmas (*M. pneumoniae* is about 10 times smaller). Taking into

account the small cell size, limiting the diversity and abundance of all cellular components, *M. pneumoniae* is of special interest for the study of essential cellular functions, common to all cells. We integrated data on genomics, transcriptomics, and proteomics with the theoretical peptidome encoded by the *M. pneumoniae* genome. This allowed us to correct the genome annotation of *M. pneumoniae*, thereby re-defining more than 10% of the protein coding ORFs, what even exceeds the high error rate of 8% estimated for *M. genitalium* [Brenner, 1999]. Our analysis reveals a high degree of genome packaging, which allows to maintain diverse regulatory mechanisms, such as antisense RNAs for example, contradicting the simplicity hitherto assumed for bacterial genomes.

The presented work improves our understanding of *M. pneumoniae* significantly. The drawn biological conclusions, for example with respect to energy homeostasis in bacteria or the unexpected structural complexity of their genomes, will have general impact on the research in related fields of biology. The metabolic model, the *in vivo* metabolome study, and the high-quality genome annotation, together with the genome-scale data sets on transcriptomics, proteomics, and metabolomics turn *M. pneumoniae* into one of the most promising model organisms for systems biology.

Future Prospects

Since the presented work only comprises a small part of a much bigger project, i.e. the aim to understand an organism in its entirety, a lot of future work remains to be done until even approximating this goal. First of all, the presented data and results have to be integrated into the **MyMpn** database for its release. In addition, a comparative analysis of *M. pneumoniae* and *M. genitalium* applying the data analysis and visualization tools provided by the database, will add up to our understanding of the two closely related organisms. Another study on mycoplasma evolution with a focus on metabolism is currently under development and aims to allow the reconstruction of an ancestral mycoplasma genome as well as of its metabolic network by including more distant species than *M. genitalium* in this analysis. This network would not only enable us to study the specific effects of metabolic sub-networks on the general network behavior, but also conclude on metabolic functions related to the different niches and virulence functions of various mycoplasma species. For our essentiality study, we are currently producing additional datasets aimed to allow i) further finetuning of the applied scoring system and, putatively, the sequential and functional (re-)annotation of additional genes.

Several mathematical models are currently under development, one for the transcription and translation of genes into proteins, one describing gene regulation by the few TFs of *M. pneumoniae*, and another one to analyze chromosome structuring and the influence on transcriptional regulation by chromosome packaging. Those models all attempt to shed light on the mechanisms by which minimal bacteria that lack many cellular functions known from more complex organisms are still able to show the differentiated response to environmental stresses observed *in vivo* [Güell et al., 2009, 2011]. Once finished, those models have to be connected to the metabolic model in order to study the connection between transcription, translation, and metabolic function. Furthermore, to deepen the knowledge obtained by the metabolic model and especially to understand regulatory mechanisms, for example the one governing the metabolic switch

from mainly acetic to mainly lactic acid fermentation *M. pneumoniae* undergoes in a four days batch culture growth experiment, dynamic models for metabolic sub-systems have to be designed. An ODE-based model for glycolysis, aimed to allow testing of the hypothesis that oxygen availability strongly influences the regulation of pyruvate processing, has been designed and will in the best case also provide information about kinetic parameters, which then can be compared to the estimates calculated from *in vivo* and *in silico* results. To this end, experimentally we plan to measure growth and external metabolites in an LDH^- strain in order to better understand the regulation of the mentioned metabolic switch. Despite the fact that *M. pneumoniae* will not be the first organism for which a whole-cell model is constructed with the recent one for *M. genitalium* [Karr et al., 2012], we are convinced that such a model would complement that for *M. genitalium* based on the differing foci of the mathematical models and the experimental analyses.

Based on the curation of the genome annotation, we are currently conducting a genome-wide essentiality study. A minitransposon mutant library obtained previously by "haystack mutagenesis" [Halbedel and Stülke, 2007] has been sequenced, determining the insertion positions in the genomes of the transposon mutants. While up to now, the criteria to define essentiality were only applicable to defined protein or functional RNA encoding ORFs, we are working on the establishment of a score allowing to decipher all essential genomic regions, regardless of their function, length, and transcription. This score, in combination with the high-quality genome annotation and the *in silico* gene essentiality prediction of the metabolic model, will be used to unravel the minimal essential genome of *M. pneumonaie* and insights into mycoplasma-specific functional mechanisms.

As a final conclusion from the presented work, we claim that mycoplasmas are ideal model organisms for systems biology due to their small size and the many biological principles, as for instance the existence of antisense RNAs in bacteria [Lluch-Senar et al., 2007], recently discovered. Compared to its closest relative, *M. genitalium*, apart from the high quality of the available data, for example the new genome annotation, the advantages in laboratory culturing of *M. pneumoniae* are putatively the major advantages. Nevertheless, research on both organisms will mutually profit from each other due to their close relationship and an integration of findings from the recently published whole-cell model for *M. genitalium* with our results should reveal further functional mechanisms of these minimal organisms.

A. Supplementary Material for Chapter 3

A.1. Sequence Alignments

We used NCBI pBLAST (Altschul et al, 1997) to determine sequence similarity for different *M. pneumoniae* proteins to those of other organisms. We used the nr-DB when searching for similarity to *M. pneumoniae* proteins and the *M. pneumoniae* proteome when using protein sequences from other organisms to check for similarities in *M. pneumoniae*. In case of significant hits (e-value < 1e-25) the alignment of the best hit is shown, in case of no significant hits the 5 first entries from the hit list are shown.

A.1.1. Putative Succinate Dehydrogenase Subunit

As an example for the negative results in the attempt to identify enzymes involved in succinate or fumarate processing in *M. pneumoniae* the alignment of *B. subtilis* sdhA protein sequence vs. the *M. pneumoniae* proteome is shown.

```
Query= succinate dehydrogenase flavoprotein subunit (sdhA) {Bacillus subtilis}
Length=586

                                                                    Score    E-
 Sequences producing significant alignments:                        (Bits)   Value
 ref|NP_110246.1| tRNA uridine 5-carboxymethylaminomethyl modi...   28.1     0.039
 ref|NP_110144.1| hypothetical protein MPN456 [Mycoplasma pneu...   27.3     0.087
 ref|NP_109928.1| thioredoxin reductase [Mycoplasma pneumoniae...   26.2     0.16
 ref|NP_109949.1| DNA topoisomerase I [Mycoplasma pneumoniae M...   26.2     0.18
 ref|NP_109936.1| Ser/Thr/Tyr protein kinase [Mycoplasma pneum...   24.6     0.54
```

A.1.2. Glycerol 3-phosphate Dehydrogenase/Oxidase (MPN051)

```
Query= glycerol 3-phosphate dehydrogenase (GlpD) {Bacillus subtilis}
Length=555
                                                                    Score    E-
 Sequences producing significant alignments:                        (Bits)   Value
 ref|NP_109739.1| glycerol-3-phospate dehydrogenase [Mycoplasm...   53.5     4e-10

ALIGNMENTS
>ref|NP_109739.1| glycerol-3-phospate dehydrogenase [Mycoplasma pneumoniae M129]
sp|P75063.1|Y051_MYCPN RecName:  Full=Uncharacterized protein MG039 homolog
gb|AAB95751.1| glycerol-3-phospate dehydrogenase [Mycoplasma pneumoniae M129]
Length=384
```

```
Score = 53.5 bits (127), Expect = 4e-10, Method:  Compositional matrix adjust.
Identities = 86/359 (24%), Positives = 143/359 (40%), Gaps = 72/359 (20%)

 Query    21  KTYDLFIIGGGITGAGTALDAASRGMKVALSEMQDFAAG-TSSRSTKLVHGGLRYLKQFE   79
              +T D+ I+GGG+ G  TA + +   +KV L E   + A  TS  ++ ++H G+
 Sbjct     2  ETRDVLIVGGGVIGCATAYELSQYKLKVTLVEKHHYLAQETSHANSGVIHTGI-------   54

 Query    80  VKMVAEVGKERAIVYENGPHVTTPEWMLLPFHKGGTFGSFTTSIGL---RVYDFLAGVKK  136
                            +  PH  T ++ +L   K     ++   +G    ++   +    +
 Sbjct    55  ---------------DPNPHKLTAKYNIL--GKKLWLNTYFKRLGFPRQKIRTLIVAFNE   97

 Query   137  SERRS-------------------MLSAKETLQKEPLVKKDGLKG----GGYYVEYRTDD  173
               ER                     MLS +ETL+ EP V  + + G    G + ++
 Sbjct    98  MEREQLEVLKQRGIANQINLEDIQMLSKEETLKLEPYVNPEIVAGLKIEGSWAIDPVLAS  157

 Query   174  ARLTIEVMKEAVKF--GAEPVNYSKVKELLYEKGKAVGVLIEDVLTKKEYKVYAKKIVNA  231
                L +   +  V+     E  N SK  +  Y       ++  +  T   +KV  KKI++A
 Sbjct   158  KCLALAAQQNKVQICTNTEVTNISKQVDGTY-------LVWTNNETTPSFKV--KKIIDA  208

 Query   232  TGPWVDQLREKDHSKNGKHLQHTKGIHLVFDQSVFPLKQAVYF-DTPDGRMVFAIPR-EG  289
               G + D L     + + +         +V +Q    L   V+   T  G+ V   P  +G
 Sbjct   209  AGHYADYLAHLAKADDFEQTTRRGQYVVVTNQGELHLNSMVFMVPTIHGKGVIVSPMLDG  268

 Query   290  KTYVGTT--DTVYKEALEHPRMTTEDRDYVIKSI-NYMFPELNITANDIESSWAGLRPL  345
                 VG T  D V KEA    R  T+D   ++  I  +M P LNI  N+   S+AG RP+
 Sbjct   269  NFLVGPTALDGVDKEAT---RYITKDAPCMLTKIGKHMVPSLNI--NNALISFAGSRPI  322
```

A.1.3. NADH Oxidase (NOX, MPN394)

a) H_2O-forming NOX of *S. mutants* vs. *M. pneumoniae* proteome

```
Query= NADH oxidase (H2O-forming) (NaoX) {Streptococcus mutans}
Length=457
                                                                  Score    E-
 Sequences producing significant alignments:                      (Bits)   Value
 lcl|58781 MPN394 NADH oxidase (nox) {Mycoplasma pneumoniae M129}  286      1e-81

ALIGNMENTS
>lcl|58781 MPN394 NADH oxidase (nox) {Mycoplasma pneumoniae M129}
Length=479


Score = 286 bits (731), Expect = 1e-81, Method:  Compositional matrix adjust.
Identities = 171/472 (36%), Positives = 281/472 (59%), Gaps = 20/472 (4%)

 Query     1  MSKIVIVGANHAGTAAINTILDNYGSENEVVVFDQNSNISFLGCGMALWIGKQISGPQGL   60
              M K++++G NHAGT+ I T+L     + +V  +D+N+NISFLGCG+AL +   +   + L
 Sbjct     1  MKKVIVIGVNHAGTSFIRTLLSK-SKDFQVNAYDRNTNISFLGCGIALAVSGVVKNTEDL   59

 Query    61  FYADKESLEAKGAKIYMESPVTAIDYDAKRVTA--LVNGQEHVESYEKLILATGSTPILP  118
              FY+  E L+A GA ++M   V  +D D K+V    L  G+E V+ Y++L++A+G+ PI
 Sbjct    60  FYSTPEELKAMGANVFMAHDVVGLDLDKKQVIVKDLATGKETVDHYDQLVVASGAWPICM  119

 Query   119  PIKGAAIK-----EGSRDFEATLKNLQFVKLYQNAEDVINKLQ-DKSQNLNRIAVVGAGY  172
               ++            +  +   +KNL   KLYQ+A  +I+  + DKS  +  +A+VG+GY
 Sbjct   120  NVENEVTHTQLQFNHTDKYCGNIKNLISCKLYQHALTLIDSFRHDKS--IKSVAIVGSGY  177
```

```
 Query  173  IGVELAEAFKRLGKEVILIDVVDTCLAGYYDQDLSEMMRQNLEDHGIELAFGETVKA--I  230
             IG+ELAEA  + GK+V +ID++D      +D++ +  + + ++  GI L  G  VK   +
 Sbjct  178  IGLELAEAAWQCGKQVTVIDMLDKPAGNNFDEEFTNELEKAMKKAGINLMMGSAVKGFIV  237

 Query  231  EGDGKVER-IVTDKASHDVDMVILAVGFRPNTAL--GNAKLKTFRNGAFLVDKK-QETSI  286
             + D  V + + TDK   D D+VI ++GFRPNT     + + +  RNG+  V++  Q  +
 Sbjct  238  DADKNVVKGVETDKGRVDADLVIQSIGFRPNTQFVPKDRQFEFNRNGSIKVNEYLQALNH  297

 Query  287  PDVYAIGDCATVYDNAINDTNYIALASNALRSGIVAGHNAAGHKLESL-GVQGSNGISIF  345
              +VY IG  A +YD A      I LA+NA++SG+VA  +  G K   L  + G+N + +F
 Sbjct  298  ENVYVIGGAAAIYDAASEQYENIDLATNAVKSGLVAAMHMIGSKAVKLESIVGTNALHVF  357

 Query  346  GLNMVSTGLTQEKAKRFGYNPEVTAFTDFQKASFIEHDNYPVTLKIVYDKDSRLVLGAQM  405
             GLN+ +TGLT+++AK  G++  V+   D  +  F+   +  V  K++YDK +  +LGAQ+
 Sbjct  358  GLNLAATGLTEKRAKMNGFDVGVSIVDDNDRPEFMGTFD-KVRFKLIYDKKTLRLLGAQL  416

 Query  406  AS-KEDMSMGIHMFSLAIQEKVTIERLALLDYFFLPHFNQPYNYMTKAALKA  456
              S   + S  I   +LA+Q+K+ I  L L+D +FLPH+N+P+N++  A L+A
 Sbjct  417  LSWNTNHSEIIFYIALAVQKKMLISELGLVDVYFLPHYNKPFNFVLAAVLQA  468
```

b) H_2O_2-forming NOX of *S. mutants* vs. *M. pneumoniae* proteome

```
Query= NADH oxidase (H2O2-forming) (NaoX) {Streptococcus mutans}
Length=457
                                                                  Score    E-
 Sequences producing significant alignments:                      (Bits)   Value
 lcl|2291 MPN394 NADH oxidase (nox) {Mycoplasma pneumoniae M129}  23.1     0.022

ALIGNMENTS
>lcl|2291 MPN394 NADH oxidase (nox) {Mycoplasma pneumoniae M129}
Length=479


Score = 23.1 bits (48), Expect = 0.022, Method:  Compositional matrix adjust.
Identities = 31/127 (24%), Positives = 53/127 (41%), Gaps = 16/127 (12%)

 Query  348  KKVAVIGGGNSGLEAAIDLAGLASHVYILEFLPELKADKILQDRAEALDN------ITIL  401
             K VA++G G  GLE A         V +++ L +   +   ++    L+       I ++
 Sbjct  168  KSVAIVGSGYIGLELAEAAWQCGKQVTVIDMLDKPAGNNFDEEFTNELEKAMKKAGINLM  227

 Query  402  TNVATKEII---GNDHVEGLRYSDRTTNEEYLLDLEGVFVQIGLVPSTDWL-KDSGLALN  457
                A K  I     + V+G+  +D+       +D + V   IG  P+T ++ KD     N
 Sbjct  228  MGSAVKGFIVDADKNVVKGVE-TDKGR-----VDADLVIQSIGFRPNTQFVPKDRQFEFN  281

 Query  458  EKGEIIV  464
               G I V
 Sbjct  282  RNGSIKV  288
```

A.1.4. Putative CTP Synthase

As an example for the negative results in the attempt of identifying a CTP synthase in *M. pneumoniae* the alignment of *M. gallisepticum* pyrG protein sequence vs. the *M. pneumoniae* proteome is shown.

```
Query= CTP synthase (pyrG) {Mycoplasma gallisepticum}
Length=540
```

```
                                                                    Score    E-
Sequences producing significant alignments:                         (Bits)   Value
lcl|57601 MPN001 DNA polymerase III beta subunit (dnaN) {Myco...   7.3      0.84
lcl|57603 MPN003 DNA gyrase subunit B (gyrB) {Mycoplasma pneu...    18.5     0.30
lcl|57604 MPN004 DNA gyrase subunit A (gyrA) {Mycoplasma pneu...    17.3     0.66
lcl|57605 MPN005 seryl-tRNA synthetase (serS) {Mycoplasma pne...    16.2     1.7
lcl|57606 MPN006 thymidylate kinase {Mycoplasma pneumoniae M129}    15.0     3.4
```

A.2. Comparative Calculations for *M. pneumoniae* and *E. coli*

To verify the surprising findings on energy consumption in *M. pneumoniae*, we compared the biophysical properties and the energy consuming processes in *M. pneumoniae* and *E. coli*. The numbers for biophysical properties of *M. pneumoniae* are taken from Yus et al. [2009] and the numbers for *E. coli* have been extracted from the Bionumbers database [Milo et al., 2010]. We calculated the surface to volume ratio for both organisms in order to identify a possible influence on cellular homeostasis tasks. Furthermore, we determined the relative fractions of energy dedicated to the synthesis of the major cellular building blocks in both organisms. The list of biophysical properties used and the calculation results can be seen in Table A.18.

A.3. Figures

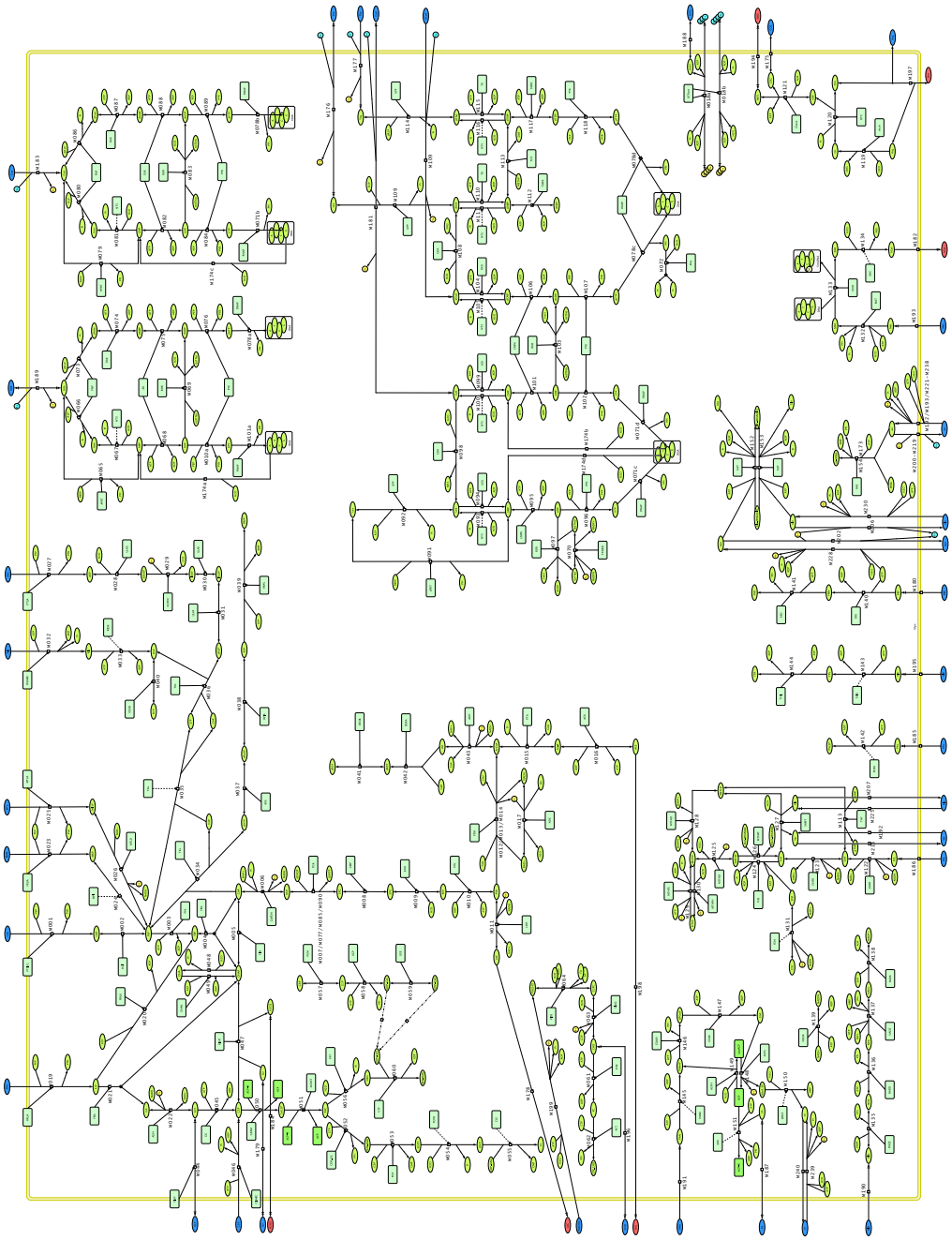

Figure A.1.: Reaction network implemented in the metabolic model for *M. pneumoniae*. Metabolites have oval shapes, while boxes display metabolic enzymes or enzyme complexes. Blue colored metabolites are imported from the environment, red colored metabolites are exported. The clickable version of the model map is found in the **MyMpn** database, menu section 'pathway maps', a larger version for take out at the very end of this thesis.

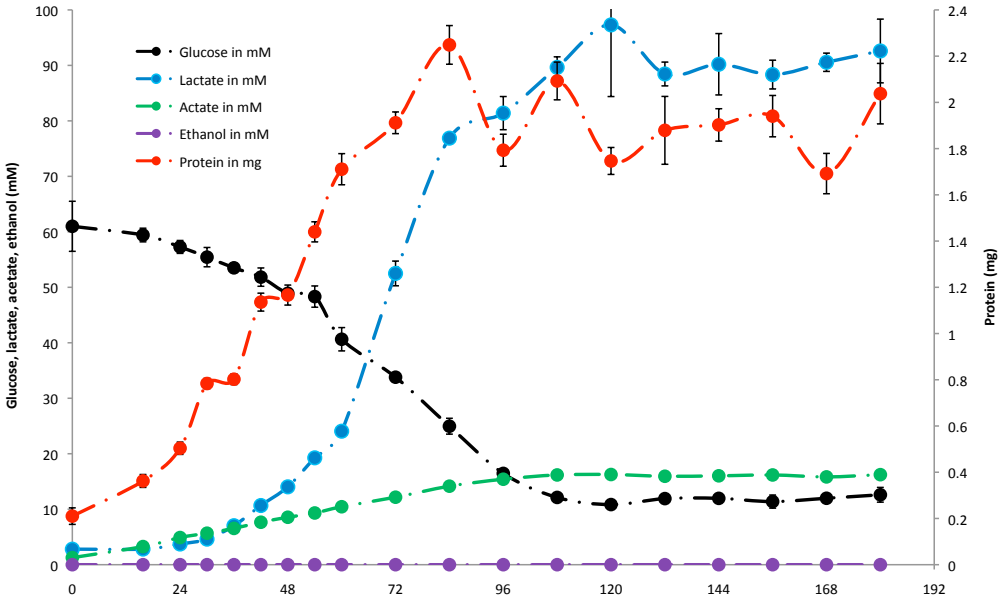

Figure A.2.: Metabolite assays: Extracellular glucose and organic acid concentrations (left y-axis), as well as protein concentration (right y-axis) during a four days batch culture growth experiment.

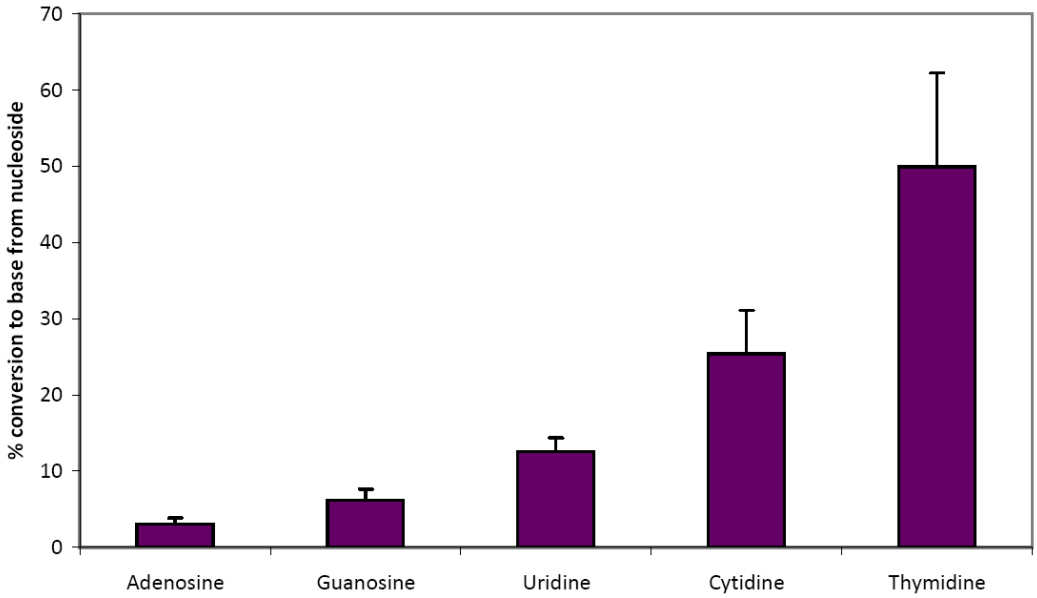

Figure A.3.: Conversion of nucleosides into nucleobases during sample preparation.

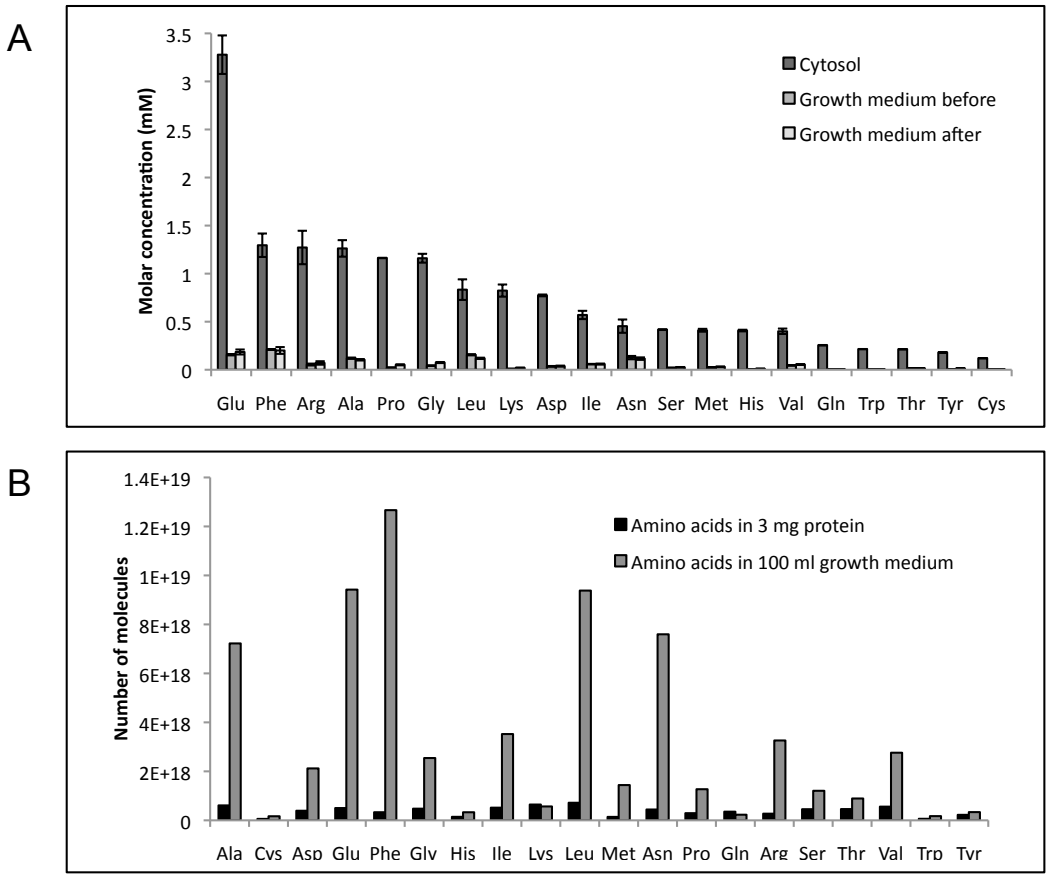

Figure A.4.: Amino acid quantifications. A: Amino acid concentrations in the cytosol are higher than in the growth medium suggesting active import. B: Amino acids are not growth limiting.

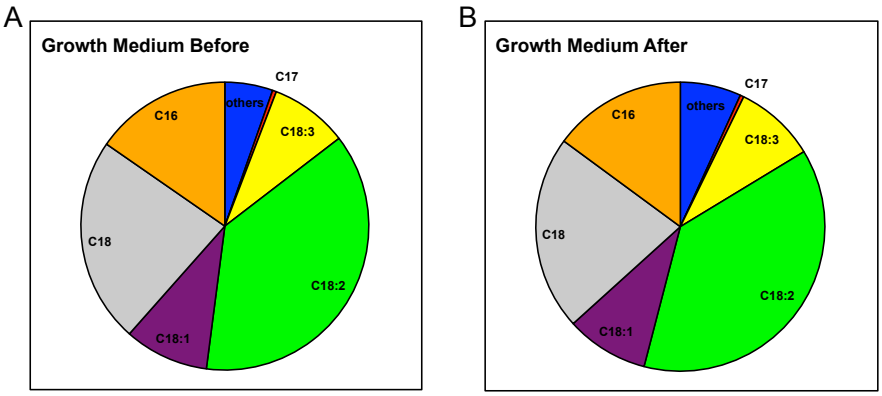

Figure A.5.: Fatty acid composition. A: at the beginning (0h) and B: at the end (96h) of a four days batch culture growth experiment.

```
MKKVIVIGVNHAGTSFIRT
M K++++G NHAGT+ I T
MSKIVIVGANHAGTAAINT
```

Figure A.6.: Conserved FAD-binding fingerprint of the H_2O-producing *M. pneumoniae* NOX.

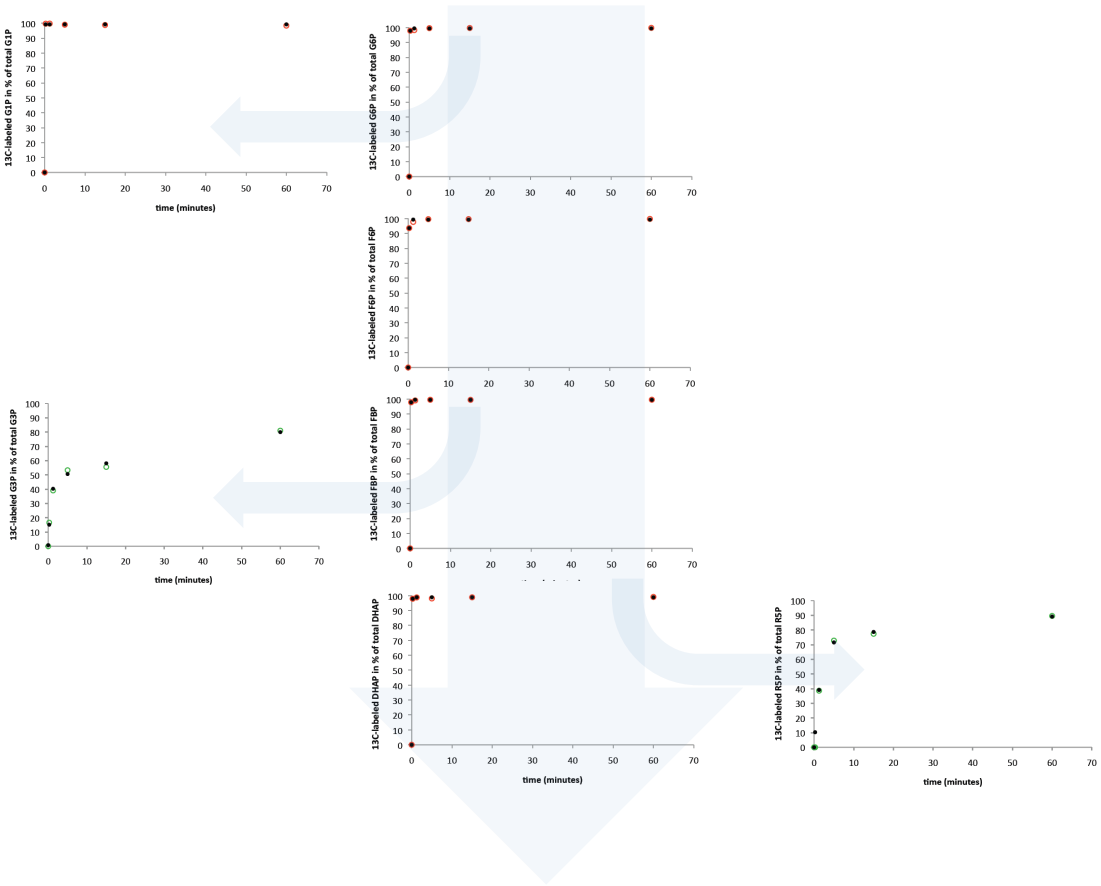

Figure A.7.: Analysis of heavy labeled carbon flux: Transparent blue arrows indicate glycolysis and corresponding outfluxes towards lipid metabolism (left) and the pentose phosphate pathway (right). Experimental data (black circles) is well described by single phase exponential decay functions for glycolytic compounds and for G1P (red rings). For G3P and R5P, two-phase exponential decay functions were fitted to the experimental data (green rings). See tables A.16 & A.17 for all fitting parameters.

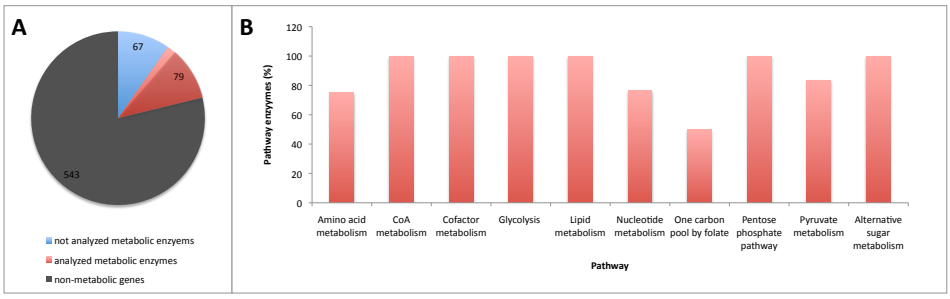

Figure A.8.: Alignment of qualitative changes of *in vivo* protein abundances and *in silico* fluxes during the exponential growth phase. A: Classification of *M. pneumoniae* enzymes; blue - metabolic enzymes without available protein data or only catalyzing metabolic reactions in complexes; red - metabolic enzymes for which the concentration changes have been assigned to flux changes, the dark red shadow indicates those enzymes that show concentration changes that qualitatively match the flux changes of their respective reactions (hit enzymes); grey - non-metabolic enzymes. B: Hit enzymes in percentage of the total number of enzymes involved in their respective pathways.

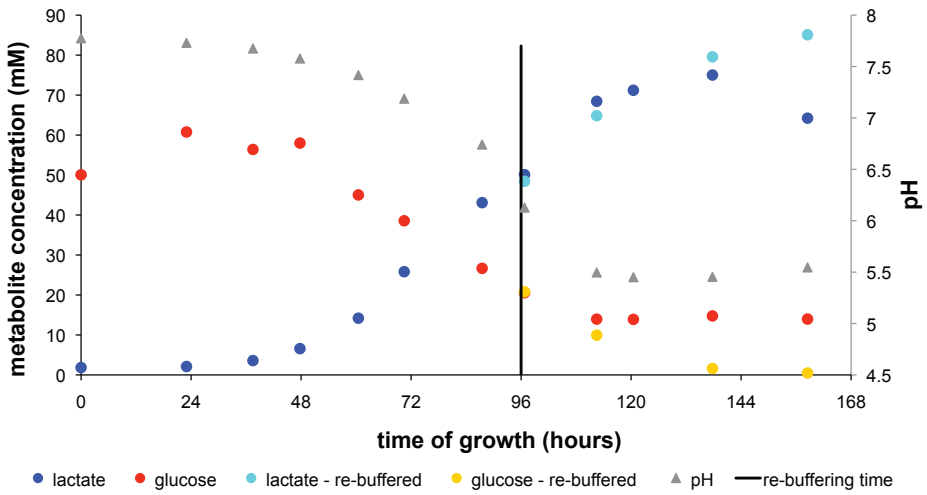

Figure A.9.: Glucose (red/orange) and lactic acid concentrations (blue/zyan) and the medium pH (grey triangles) for different time points of batch culture growth. At 96 hours the medium pH was re-buffered by NaOH titration to pH7.7 (vertical black line). In normal batch culture growth without re-buffering the medium pH cell growth comes to a halt at about 96 hours of growth (glucose - red, lactic acid - blue). In re-buffered medium *M. pneumoniae* was able to use up the remaining glucose (orange), to produce further lactic acid (zyan) and, subsequently, to grow.

A.4. Tables

Table A.1: Reaction List

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M001	R001/ R002	D-glucose[e] + phosphoenolpyruvate[c] ⟶ D-glucose 6-phosphate[c] + pyruvate[c]	MPN627/ MPN207	PTSI/G	2.7.3.9/ 2.7.1.69	glycolysis	I
M002	R003	[c]: D-glucose 6-phosphate ⟷ D-fructose 6-phosphate	MPN250	PGI	5.3.1.9	glycolysis	R
M003	R004	[c]: D-fructose 6-phosphate + ATP ⟶ ADP + D-fructose 1,6-bisphosphate + H+	MPN302	PFK	2.7.1.11	glycolysis	I
M004	R005	[c]: D-fructose 1,6-bisphosphate ⟷ dihydroxyacetone phosphate + D-glyceraldehyde 3-phosphate	MPN025	FBA	4.1.2.13	glycolysis	R
M005	R006	[c]: D-glyceraldehyde 3-phosphate ⟷ dihydroxyacetone phosphate	MPN629	TIM	5.3.1.1	glycolysis	R
M006	R007	[c]: NAD+ + orthophosphate + D-glyceraldehyde 3-phosphate ⟷ NADH + H+ + 1,3-bisphospho-D-glycerate	MPN430	GAPDH	1.2.1.12	glycolysis	R
M007	R008	[c]: 3-phospho-D-glycerate + ATP ⟷ ADP + 1,3-bisphospho-D-glycerate	MPN429	PGK	2.7.2.3	glycolysis	R
M008	R009	[c]: 2-phospho-D-glycerate ⟷ 3-phospho-D-glycerate	MPN628	PGM	5.4.2.1	glycolysis	R
M009	R010	[c]: 2-phospho-D-glycerate ⟷ phosphoenolpyruvate + H_2O	MPN606	ENO	4.2.1.11	glycolysis	R
M010	R011	[c]: ADP + phosphoenolpyruvate + H+ ⟶ ATP + pyruvate	MPN303	PYK	2.7.1.40	glycolysis	I
M011	R012	[c]: (S)-lactate + NAD+ ⟷ NADH + H+ + pyruvate	MPN674	LDH	1.1.1.27	pyruvate metabolism	R
M012	R015	[c]: lipoamide + H+ + pyruvate ⟶ S-acetyldihydrolipoamide + CO_2	MPN392/ MPN393	PDH-E1	1.2.4.1	pyruvate metabolism	I
M013	R016	[c]: acetyl-CoA + dihydrolipoamide ⟷ S-acetyldihydrolipoamide + CoA	MPN391	PDH-E2	2.3.1.12	pyruvate metabolism	R
M014	R017	[c]: dihydrolipoamide + NAD+ ⟷ lipoamide + NADH + H+	MPN390	PDH-E3	1.8.1.4	pyruvate metabolism	R
M015	R018	[c]: acetyl-CoA + orthophosphate ⟷ acetyl phosphate + CoA	MPN428	PTA	2.3.1.8	pyruvate metabolism	R
M016	R019	[c]: acetate + ATP ⟷ ADP + acetyl phosphate	MPN533	ACK	2.7.2.1	pyruvate metabolism	R
M017	R020	[c]: (2) NADH + oxygen + (2) H+ ⟷ (2) H_2O + (2) NAD+	MPN394	NOX	1.6.99.3	pyruvate metabolism	R
M018	R190	(4) H+[e] + ADP[c] + orthophosphate[c] ⟷ H_2O[c] + (3) H+[c] + ATP[c]	MPN597-MPN604	ATPase	3.6.3.14	Energy metabolism	R
M019	R021	D-fructose[e] + phosphoenolpyruvate[c] ⟶ D-fructose 1-phosphate[c] + pyruvate[c]	MPN078	FRUA	2.7.1.69	Alternative sugar metabolism	I
M020	R022	[c]: D-fructose 1-phosphate + ATP ⟶ ADP + D-fructose 1,6-bisphosphate + H+	MPN079	FRUK	2.7.1.56	Alternative sugar metabolism	I
M021	R023	[c]: D-fructose 1-phosphate ⟶ D-glyceraldehyde + dihydroxyacetone phosphate	MPN025	FBA	4.1.2.13	Alternative sugar metabolism	I
M022	R024	[c]: glycerol + NAD+ ⟷ D-glyceraldehyde + NADH + H+	MPN564	ADH	1.1.1.1	Alternative sugar metabolism	R
M023	R026	D-mannose[e] + phosphoenolpyruvate[c] ⟶ D-mannose 6-phosphate[c] + pyruvate[c]	MPN078	FRUA	2.7.1.69	Alternative sugar metabolism	I
M024	R028	[c]: D-mannose 6-phosphate ⟷ D-fructose 6-phosphate	?	MPI	5.3.1.8	Alternative sugar metabolism	R
M025	R029	mannitol[e] + phosphoenolpyruvate[c] ⟶ pyruvate[c] + D-mannitol 1-phosphate[c]	MPN651/ MPN653	MTLA	2.7.1.69	Alternative sugar metabolism	I
M026	R030	[c]: NAD+ + D-mannitol 1-phosphate ⟷ NADH + D-fructose 6-phosphate + H+	MPN652	MTLD	1.1.1.17	Alternative sugar metabolism	R
M027	R031	L-ascorbate[e] + phosphoenolpyruvate[c] + (2) H+[c] ⟶ L-ascorbate 6-phosphate[c] + pyruvate[c]	MPN494-MPN496	PTSA	2.7.1.69	Alternative sugar metabolism	I
M028	R032	[c]: H_2O + L-ascorbate 6-phosphate ⟷ 3-keto-L-gulonate 6-phosphate + (3) H+	MPN497	ULAG	3.1.1.-	Alternative sugar metabolism	R
M029	R033	[c]: 3-keto-L-gulonate 6-phosphate + H+ ⟶ CO_2 + L-xylulose 5-phosphate	MPN493	KGPDC	4.1.1.85	Alternative sugar metabolism	I
M030	R034	[c]: L-ribulose 5-phosphate ⟷ L-xylulose 5-phosphate	MPN492	ULAE	5.1.3.22	Alternative sugar metabolism	R

Continued on next page

Table A.1 – continued from previous page

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M031	R035	[c]: L-ribulose 5-phosphate ⟷ D-xylulose 5-phosphate	MPN498	ULAF	5.1.3.4	Alternative sugar metabolism	R
M032	R191	D-ribose[e] ⟶ D-ribose[c]	MPN258-MPN260	RibABC	3.6.3.17	pentose phosphate pathway	I
M033	R036	[c]: D-ribose + ATP ⟶ ADP + D-ribose 5-phosphate + H+	?	RBSK	2.7.1.15	pentose phosphate pathway	I
M034	R037	[c]: D-fructose 6-phosphate + D-glyceraldehyde 3-phosphate ⟷ D-xylulose 5-phosphate + D-erythrose 4-phosphate	MPN082	TKL	2.2.1.1	pentose phosphate pathway	R
M035	R038	[c]: sedoheptulose 7-phosphate + D-glyceraldehyde 3-phosphate ⟷ D-erythrose 4-phosphate + D-fructose 6-phosphate	?	TAL	2.2.1.2	pentose phosphate pathway	R
M036	R039	[c]: sedoheptulose 7-phosphate + D-glyceraldehyde 3-phosphate ⟷ D-xylulose 5-phosphate + D-ribose 5-phosphate	MPN082	TKL	2.2.1.1	pentose phosphate pathway	R
M037	R040	[c]: D-ribulose 5-phosphate ⟷ D-xylulose 5-phosphate	MPN251	RPE	5.1.3.1	pentose phosphate pathway	R
M038	R041	[c]: D-ribose 5-phosphate ⟷ D-ribulose 5-phosphate	MPN595	RPIA	5.3.1.6	pentose phosphate pathway	R
M039	R042	[c]: D-ribose 5-phosphate + ATP ⟶ AMP + 5-phospho-alpha-D-ribose 1-diphosphate + H+	MPN073	PRPS	2.7.6.1	pentose phosphate pathway	I
M040	R043	[c]: D-ribose 1-phosphate ⟷ D-ribose 5-phosphate + (2) H+	MPN066	DEOB	5.4.2.7	pentose phosphate pathway	R
M041	R044	[c]: 2-deoxy-D-ribose 1-phosphate ⟷ 2-deoxy-D-ribose 5-phosphate	MPN066	DEOB	5.4.2.7	pentose phosphate pathway	R
M042	R046	[c]: 2-deoxy-D-ribose 5-phosphate ⟶ acetaldehyde + D-glyceraldehyde 3-phosphate	MPN063	DERA	4.1.2.4	pentose phosphate pathway	I
M043	R047	[c]: acetaldehyde + NAD+ + CoA ⟷ acetyl-CoA + NADH + H+	MPN564	ADH	1.2.1.10	Alternative sugar metabolism	R
M044	R192	glycerol[c] ⟷ glycerol[e]	MPN043	GlpF	-	lipid metabolism	R
M045	R049	[c]: glycerol + ATP ⟶ ADP + sn-glycerol 3-phosphate + H+	MPN050	GK	2.7.1.30	lipid metabolism	I
M046	R193	sn-glycerol 3-phosphate[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + orthophosphate[c] + H+[c] + sn-glycerol 3-phosphate[c]	MPN133-MPN136	GlyABC	-	lipid metabolism	I
M047	R186	[c]: oxygen + sn-glycerol 3-phosphate ⟷ H_2O_2 + dihydroxyacetone phosphate	MPN051	GPO	1.1.3.21	lipid metabolism	R
M048	R051	[c]: H_2O + dihydroxyacetone phosphate ⟶ dihydroxyacetone + orthophosphate	-	-	-	lipid metabolism	I
M049	R052	[c]: dihydroxyacetone + ATP ⟶ ADP + dihydroxyacetone phosphate + H+	MPN547	DHAK	2.7.1.29	lipid metabolism	I
M050	R053	[c]: ACP-R (Mpn) + sn-glycerol 3-phosphate ⟶ 1-acyl-glycerol 3-phosphate (Mpn) + acyl carrier protein	MPN350/ MPN546	GPAM	2.3.1.15	lipid metabolism	I
M051	R054	[c]: ACP-R (Mpn) + 1-acyl-glycerol 3-phosphate (Mpn) ⟶ phosphatidic acid (Mpn) + acyl carrier protein	MPN299	AGPAT	2.3.1.51	lipid metabolism	I
M052	R055	[c]: phosphatidic acid (Mpn) + CTP + H+ ⟶ CDP-diacylglycerol (Mpn) + pyrophosphate	MPN637	CDP-DG	2.7.7.41	lipid metabolism	I
M053	R056	[c]: CDP-diacylglycerol (Mpn) + sn-glycerol 3-phosphate ⟶ CMP + phosphatidylglycerol 3-phosphate (Mpn) + H+	MPN253	PGP	2.7.8.5	lipid metabolism	I
M054	R057	[c]: H_2O + phosphatidylglycerol 3-phosphate (Mpn) ⟶ orthophosphate + phosphatidylglycerol (Mpn)	?	PGPB	3.1.3.27	lipid metabolism	I
M055	R058	[c]: CDP-diacylglycerol (Mpn) + phosphatidylglycerol (Mpn) ⟶ cardiolipin (Mpn) + CMP + H+	?	CLS	2.7.8.-	lipid metabolism	I
M056	R059	[c]: H_2O + phosphatidic acid (Mpn) ⟶ diacylglycerol (Mpn) + orthophosphate	MPN455	PPT	3.1.3.4	lipid metabolism	I
M057	R060	[c]: D-glucose 1-phosphate ⟷ D-glucose 6-phosphate	MPN066	PGM	5.4.2.2	lipid metabolism	R
M058	R061	[c]: D-glucose 1-phosphate + UTP + H+ ⟷ UDP-glucose + pyrophosphate	MPN667	UGP	2.7.7.9	lipid metabolism	R
M059	R062	[c]: UDP-glucose ⟷ UDP-galactose	MPN257	UGE	5.1.3.2	lipid metabolism	R
M060	R063/ R064	[c]: diacylglycerol (Mpn) + (3) UDP-galactose + (3) UDP-glucose ⟶ glycolipid (Mpn) + (6) H+ + (6) UDP	MPN483	GTF	-	lipid metabolism	I
M061	R178	[c]: choline + ATP ⟶ ADP + H+ + choline phosphate	MPN532	CHK	2.7.1.32	lipid metabolism	I
M062	R180	[c]: CTP + H+ + choline phosphate ⟶ CDP-choline + pyrophosphate	MPN336	PCT	2.7.7.15	lipid metabolism	I

Continued on next page

Table A.1 – continued from previous page

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M063	R182	[c]: sn-glycero-3-phosphocholine + H_2O ⟷ choline + sn-glycerol 3-phosphate + H+	MPN420	GlpQ	3.1.4.46	lipid metabolism	R
M064	R187	[c]: phosphatidylcholine + (2) H_2O ⟷ sn-glycero-3-phosphocholine + (2) fatty acid (Mpn)	MPN445	PldB	3.1.1.5	lipid metabolism	R
M065	R065	[c]: AMP + pyrophosphate ⟷ adenine + 5-phospho-alpha-D-ribose 1-diphosphate	MPN395	APRT	2.4.2.7	nucleotide metabolism	R
M066	R066	[c]: adenosine + orthophosphate + (2) H+ ⟷ D-ribose 1-phosphate + adenine	MPN062	PNP	2.4.2.1	nucleotide metabolism	R
M067	R067	[c]: H_2O + AMP ⟶ adenosine + orthophosphate	?	NT5	3.1.3.5	nucleotide metabolism	I
M068	R068	[c]: AMP + ATP ⟷ (2) ADP	MPN185	AK	2.7.4.3	nucleotide metabolism	R
M069	R069	[c]: ADP + reduced thioredoxin ⟶ dADP + H_2O + oxidized thioredoxin	MPN322-MPN324	RDR	1.17.4.1	nucleotide metabolism	I
M070	R070	[c]: oxidized thioredoxin + H+ + NADPH ⟷ NADP+ + reduced thioredoxin	MPN240	TXNRD	1.8.1.9	nucleotide metabolism	R
M071	R071/ R084/ R099/ R104	[c]: (28) GTP + (25) UTP + (40) H_2O + (18) CTP + (69) ATP ⟶ (40) ADP + RNA (Mpn) + (40) orthophosphate + (40) H+ + (100) pyrophosphate	MPN516	RNAP	-	nucleotide metabolism	I
M072	R072	[c]: H_2O + pyrophosphate ⟶ (2) orthophosphate + H+	MPN528	PPA	3.6.1.1	nucleotide metabolism	I
M073	R073	[c]: deoxyadenosine + orthophosphate ⟷ 2-deoxy-D-ribose 1-phosphate + adenine	MPN062	PNP	2.4.2.1	nucleotide metabolism	R
M074	R074	[c]: deoxyadenosine + ATP ⟷ ADP + H+ + dAMP	MPN386	DAK	2.7.1.76	nucleotide metabolism	R
M075	R075	[c]: dAMP + ATP ⟷ dADP + ADP	MPN185	AK	2.7.4.3	nucleotide metabolism	R
M076	R076	[c]: phosphoenolpyruvate + dADP + H+ ⟶ dATP + pyruvate	MPN303	PYK	2.7.1.40	nucleotide metabolism	I
M077	R175	[c]: 3-phospho-D-glycerate + dATP ⟷ dADP + 1,3-bisphospho-D-glycerate	MPN429	PGK	2.7.2.3	nucleotide metabolism	R
M078	R077/ R089/ R110/ R120	[c]: (20) dGTP + (140) H_2O + (30) dTTP + (20) dCTP + (140) ATP + (30) dATP ⟶ (140) ADP + DNA (Mpn) + (140) orthophosphate + (140) H+ + (100) pyrophosphate	MPN034/ MPN378	DNAP	-	nucleotide metabolism	I
M079	R078	[c]: GMP + pyrophosphate ⟷ 5-phospho-alpha-D-ribose 1-diphosphate + guanine	MPN672	HPRT	2.4.2.8	nucleotide metabolism	R
M080	R079	[c]: guanosine + orthophosphate + (2) H+ ⟷ D-ribose 1-phosphate + guanine	MPN062	PNP	2.4.2.1	nucleotide metabolism	R
M081	R080	[c]: GMP + H_2O ⟶ guanosine + orthophosphate	?	NT5	3.1.3.5	nucleotide metabolism	I
M082	R081	[c]: GMP + ATP ⟷ ADP + GDP	MPN246	GUK	2.7.4.8	nucleotide metabolism	R
M083	R082	[c]: reduced thioredoxin + GDP ⟶ H_2O + oxidized thioredoxin + dGDP	MPN322-MPN324	RDR	1.17.4.1	nucleotide metabolism	I
M084	R083	[c]: phosphoenolpyruvate + H+ + GDP ⟶ GTP + pyruvate	MPN303	PYK	2.7.1.40	nucleotide metabolism	I
M085	R176	[c]: GTP + 3-phospho-D-glycerate ⟷ GDP + 1,3-bisphospho-D-glycerate	MPN429	PGK	2.7.2.3	nucleotide metabolism	R
M086	R085	[c]: deoxyguanosine + orthophosphate ⟷ 2-deoxy-D-ribose 1-phosphate + guanine	MPN062	PNP	2.4.2.1	nucleotide metabolism	R
M087	R086	[c]: deoxyguanosine + ATP ⟷ ADP + H+ + dGMP	MPN386	DGK	2.7.1.113	nucleotide metabolism	R
M088	R087	[c]: ATP + dGMP ⟷ ADP + dGDP	MPN246	GUK	2.7.4.8	nucleotide metabolism	R
M089	R088	[c]: phosphoenolpyruvate + dGDP + H+ ⟶ dGTP + pyruvate	MPN303	PYK	2.7.1.40	nucleotide metabolism	I
M090	R177	[c]: 3-phospho-D-glycerate + dGTP ⟷ dGDP + 1,3-bisphospho-D-glycerate	MPN429	PGK	2.7.2.3	nucleotide metabolism	R
M091	R090	[c]: UMP + pyrophosphate ⟷ uracil + 5-phospho-alpha-D-ribose 1-diphosphate	MPN033	UPRT	2.4.2.9	nucleotide metabolism	R
M092	R091	[c]: orthophosphate + (2) H+ + uridine ⟷ uracil + D-ribose 1-phosphate	MPN064	UPP	2.4.2.3	nucleotide metabolism	R
M093	R092	[c]: H_2O + UMP ⟶ orthophosphate + uridine	?	NT5	3.1.3.5	nucleotide metabolism	I
M094	R093	[c]: ATP + uridine ⟶ ADP + UMP + H+	MPN561	UCK	2.7.1.48	nucleotide metabolism	I
M095	R094	[c]: UMP + ATP ⟷ ADP + UDP	MPN632	UMPK	2.7.4.22	nucleotide metabolism	R
M096	R095	[c]: phosphoenolpyruvate + H+ + UDP ⟶ UTP + pyruvate	MPN303	PYK	2.7.1.40	nucleotide metabolism	I
M097	R096	[c]: reduced thioredoxin + UDP ⟶ H_2O + oxidized thioredoxin + dUDP	MPN322-MPN324	RDR	1.17.4.1	nucleotide metabolism	I
M098	R097	[c]: cytidine + H_2O ⟶ NH_3 + uridine	MPN065	CDA	3.5.4.5	nucleotide metabolism	I
M099	R100	[c]: cytidine + ATP ⟶ ADP + CMP + H+	MPN561	UCK	2.7.1.48	nucleotide metabolism	I

Continued on next page

Table A.1 – continued from previous page

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M100	R101	[c]: H_2O + CMP ⟶ cytidine + orthophosphate	?	NT5	3.1.3.5	nucleotide metabolism	I
M101	R102	[c]: CMP + ATP ⟷ ADP + CDP	MPN476	CMPK	2.7.4.14	nucleotide metabolism	R
M102	R103	[c]: phosphoenolpyruvate + CDP + H+ ⟶ CTP + pyruvate	MPN303	PYK	2.7.1.40	nucleotide metabolism	I
M103	R105	[c]: CDP + reduced thioredoxin ⟶ H_2O + oxidized thioredoxin + dCDP	MPN322-MPN324	RDR	1.17.4.1	nucleotide metabolism	I
M104	R106	[c]: deoxycytidine + ATP ⟶ ADP + dCMP + H+	MPN386	DCK	2.7.1.74	nucleotide metabolism	I
M105	R107	[c]: H_2O + dCMP ⟶ deoxycytidine + orthophosphate	?	NT5	3.1.3.5	nucleotide metabolism	I
M106	R108	[c]: dCMP + ATP ⟷ ADP + dCDP	MPN476	CMPK	2.7.4.14	nucleotide metabolism	R
M107	R109	[c]: phosphoenolpyruvate + dCDP + H+ ⟶ dCTP + pyruvate	MPN303	PYK	2.7.1.40	nucleotide metabolism	I
M108	R174	[c]: H_2O + deoxycytidine ⟶ deoxyuridine + NH_3	MPN065	CDA	3.5.4.5	nucleotide metabolism	I
M109	R111	[c]: deoxyuridine + orthophosphate ⟷ uracil + 2-deoxy-D-ribose 1-phosphate	MPN064	UPP	2.4.2.3	nucleotide metabolism	R
M110	R112	[c]: deoxyuridine + ATP ⟶ ADP + dUMP + H+	MPN044	TK	2.7.1.21	nucleotide metabolism	I
M111	R113	[c]: H_2O + dUMP ⟶ deoxyuridine + orthophosphate	?	NT5	3.1.3.5	nucleotide metabolism	I
M112	R189	[c]: dUMP + ATP ⟷ ADP + dUDP	MPN006	TMPK	2.7.4.9	nucleotide metabolism	R
M113	R114	[c]: 5,10-methylenetetrahydrofolate + dUMP ⟷ dTMP + dihydrofolate	MPN320	THY	2.1.1.45	nucleotide metabolism	R
M114	R115	[c]: thymidine + orthophosphate ⟷ 2-deoxy-D-ribose 1-phosphate + thymine	MPN064	UPP	2.4.2.4	nucleotide metabolism	R
M115	R116	[c]: thymidine + ATP ⟶ ADP + dTMP + H+	MPN044	TK	2.7.1.21	nucleotide metabolism	I
M116	R117	[c]: H_2O + dTMP ⟶ thymidine + orthophosphate	?	NT5	3.1.3.5	nucleotide metabolism	I
M117	R118	[c]: dTMP + ATP ⟷ ADP + dTDP	MPN006	TMPK	2.7.4.9	nucleotide metabolism	R
M118	R119	[c]: phosphoenolpyruvate + dTDP + H+ ⟶ dTTP + pyruvate	MPN303	PYK	2.7.1.40	nucleotide metabolism	I
M119	R121/ R122	[c]: L-arginine + H_2O ⟶ NH_3 + L-citrulline + H+	MPN304/ MPN305; MPN560	ArcA	3.5.3.6	amino acid metabolism	I
M120	R123	[c]: carbamoyl phosphate + L-ornithine ⟷ orthophosphate + H+ + L-citrulline	MPN306	OTC	2.1.3.3	amino acid metabolism	R
M121	R124	[c]: CO_2 + NH_3 + ATP ⟷ ADP + carbamoyl phosphate + H+	MPN307	Ckase	2.7.2.2	amino acid metabolism	R
M122	R125/ R194	[c]: NAD+ + dihydrofolate ⟷ NADH + folic acid	MPN321; MPN300	DHFR	1.5.1.3	one carbon pool by folate	R
M123	R126/ R195	[c]: tetrahydrofolate + NAD+ ⟷ NADH + dihydrofolate + H+	MPN321; MPN300	DHFR	1.5.1.3	one carbon pool by folate	R
M124	R127	[c]: tetrahydrofolate + ATP + formate ⟶ ADP + orthophosphate + 10-formyltetrahydrofolate	MPN017	FHS	6.3.4.3	one carbon pool by folate	I
M125	R128	[c]: H_2O + 5,10-methenyltetrahydrofolate ⟷ 10-formyltetrahydrofolate	MPN017	MTHFC	3.5.4.9	one carbon pool by folate	R
M126	R129	[c]: L-methionyl-tRNA(Met) + 10-formyltetrahydrofolate ⟶ tetrahydrofolate + N-formylmethionyl-tRNA(Met)	MPN543	MTFMT	2.1.2.9	one carbon pool by folate	I
M127	R130	[c]: glycine + H_2O + 5,10-methylenetetrahydrofolate ⟷ tetrahydrofolate + L-serine	MPN576	SHMT	2.1.2.1	one carbon pool by folate	R
M128	R131	[c]: NADP+ + 5,10-methylenetetrahydrofolate ⟷ H+ + 5,10-methenyltetrahydrofolate + NADPH	MPN017	MTHFD	1.5.1.5	one carbon pool by folate	R
M129	R132	[c]: ATP + 5-formyltetrahydrofolate ⟶ ADP + orthophosphate + (2) H+ + 5,10-methenyltetrahydrofolate	MPN348	MTHFS	6.3.3.2	one carbon pool by folate	I
M130	R185	[c]: H_2O + H+ + 5,10-methenyltetrahydrofolate ⟶ 5-formyltetrahydrofolate	MPN576	MTHFH	?	one carbon pool by folate	I
M131	R181	[c]: NAD+ + formate ⟷ CO_2 + NADH	?	FDH	1.2.1.2	one carbon pool by folate	R
M132	R133	[c]: H_2O + ATP + L-methionine ⟶ orthophosphate + pyrophosphate + S-adenosyl-L-methionine	MPN060	MAT	2.5.1.6	amino acid metabolism	I
M133	R134	[c]: DNA (Mpn) + S-adenosyl-L-methionine ⟶ 5mcDNA (Mpn) + S-adenosyl-L-homocysteine	MPN108	DCM	2.1.1.37	amino acid metabolism	I
M134	R135	[c]: H_2O + S-adenosyl-L-homocysteine ⟶ L-homocysteine + adenosine	?	AHC	3.3.1.1	amino acid metabolism	I

Continued on next page

Table A.1 – continued from previous page

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M135	R136	[c]: nicotinate D-ribonucleotide + pyrophosphate ⟷ nicotinate + 5-phospho-alpha-D-ribose 1-diphosphate + H+	MPN047	PNCB	2.4.2.12	cofactor metabolism	R
M136	R137	[c]: nicotinate D-ribonucleotide + H+ + ATP ⟷ deamino-NAD+ + pyrophosphate	MPN336	NADD	2.7.7.18	cofactor metabolism	R
M137	R138	[c]: deamino-NAD+ + NH_3 + ATP ⟶ AMP + NAD+ + pyrophosphate	MPN562	NADE	6.3.1.5	cofactor metabolism	I
M138	R139	[c]: NAD+ + ATP ⟶ ADP + NADP+ + H+	MPN267	NADK	2.7.1.23	cofactor metabolism	I
M139	R173	[c]: NADH + ATP ⟶ ADP + H+ + NADPH	MPN267	NADK	2.7.1.23	cofactor metabolism	I
M140	R140	[c]: riboflavin + ATP ⟶ ADP + FMN + H+	MPN158	RFK	2.7.1.26	cofactor metabolism	I
M141	R141	[c]: FMN + H+ + ATP ⟶ pyrophosphate + FAD	MPN158	FAD	2.7.7.2	cofactor metabolism	I
M142	R142	[c]: pyridoxal + ATP ⟶ ADP + pyridoxal phosphate + H+	?	PDXK	2.7.1.35	cofactor metabolism	I
M143	R143	[c]: ATP + thiamin ⟶ ADP + H+ + thiamin monophosphate	?	THIK	2.7.1.89	cofactor metabolism	I
M144	R144	[c]: thiamin monophosphate + ATP ⟶ ADP + thiamin diphosphate	MPN550	THIL	2.7.4.16	cofactor metabolism	I
M145	R145	[c]: pantetheine + ATP ⟶ ADP + H+ + pantetheine 4-phosphate	?	PANK	2.7.1.33	CoA metabolism	I
M146	R146	[c]: H+ + ATP + pantetheine 4-phosphate ⟶ pyrophosphate + dephospho-CoA	MPN336	COASY	2.7.7.3	CoA metabolism	I
M147	R147	[c]: ATP + dephospho-CoA ⟶ ADP + CoA + H+	MPN382	COAE	2.7.1.24	CoA metabolism	I
M148	R148	[c]: apoprotein [acyl carrier protein] + CoA ⟶ adenosine 3',5'-bisphosphate + acyl carrier protein + H+	MPN298	ACPS	2.7.8.7	CoA metabolism	I
M149	R149	[c]: H_2O + acyl carrier protein ⟶ apoprotein [acyl carrier protein] + H+ + pantetheine 4-phosphate	MPN479	ACPH	3.1.4.14	CoA metabolism	I
M150	R188	[c]: H_2O + adenosine 3',5'-bisphosphate ⟶ AMP + orthophosphate	?	BPNT	3.1.3.7	CoA metabolism	I
M151	R150	[c]: fatty acid (Mpn) + acyl carrier protein + ATP ⟶ ACP-R (Mpn) + AMP + H+ + pyrophosphate	?	AAS	6.2.1.20	CoA metabolism	I
M152	R151	[c]: H_2O + L-glutamyl-tRNA(Gln) + L-asparagine + ATP ⟶ L-glutaminyl-tRNA(Gln) + ADP + orthophosphate + H+ + L-aspartate	MPN236-MPN238	GAT	6.3.5.7	amino acid metabolism	I
M153	R152	[c]: L-glutamine + H_2O + L-glutamyl-tRNA(Gln) + ATP ⟶ L-glutaminyl-tRNA(Gln) + ADP + L-glutamate + orthophosphate + H+	MPN236-MPN238	GAT	6.3.5.7	amino acid metabolism	I
M154	R153	[c]: tRNA(Met) + L-methionine + ATP ⟶ L-methionyl-tRNA(Met) + AMP + pyrophosphate	MPN023	METS	6.1.1.10	amino acid metabolism	I
M155	R154	[c]: tRNA(Ile) + L-isoleucine + ATP ⟶ L-isoleucyl-tRNA(Ile) + AMP + pyrophosphate	MPN520	ILES	6.1.1.5	amino acid metabolism	I
M156	R155	[c]: tRNA(Val) + L-valine + ATP ⟶ L-valyl-tRNA(Val) + AMP + pyrophosphate	MPN480	VALS	6.1.1.9	amino acid metabolism	I
M157	R156	[c]: tRNA(Leu) + L-leucine + ATP ⟶ AMP + L-leucyl-tRNA(Leu) + pyrophosphate	MPN384	LEUS	6.1.1.4	amino acid metabolism	I
M158	R157	[c]: tRNA(Cys) + ATP + L-cysteine ⟶ AMP + pyrophosphate + L-cysteinyl-tRNA(Cys)	MPN356	CYSS	6.1.1.16	amino acid metabolism	I
M159	R158	[c]: tRNA(Glu) + L-glutamate + ATP ⟶ L-glutamyl-tRNA(Glu) + AMP + pyrophosphate	MPN678	GLTX	6.1.1.17	amino acid metabolism	I
M160	R159	[c]: L-glutamate + ATP + tRNA(Gln) ⟶ L-glutamyl-tRNA(Gln) + AMP + pyrophosphate	MPN678	GLTX	6.1.1.17	amino acid metabolism	I
M161	R160	[c]: L-arginine + tRNA(Arg) + ATP ⟶ AMP + L-arginyl-tRNA(Arg) + pyrophosphate	MPN556	ARGS	6.1.1.19	amino acid metabolism	I
M162	R161	[c]: L-tyrosine + tRNA(Tyr) + ATP ⟶ AMP + L-tyrosyl-tRNA(Tyr) + pyrophosphate	MPN669	TYRS	6.1.1.1	amino acid metabolism	I
M163	R162	[c]: tRNA(Trp) + L-tryptophan + ATP ⟶ AMP + L-tryptophanyl-tRNA(Trp) + pyrophosphate	MPN265	TRPS	6.1.1.2	amino acid metabolism	I
M164	R163	[c]: tRNA(Ser) + L-serine + ATP ⟶ L-seryl-tRNA(Ser) + AMP + pyrophosphate	MPN005	SERS	6.1.1.11	amino acid metabolism	I
M165	R164	[c]: L-threonine + tRNA(Thr) + ATP ⟶ AMP + L-threonyl-tRNA(Thr) + pyrophosphate	MPN553	THRS	6.1.1.3	amino acid metabolism	I

Continued on next page

Table A.1 – continued from previous page

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M166	R165	[c]: tRNA(Pro) + ATP + L-proline ⟶ L-prolyl-tRNA(Pro) + AMP + pyrophosphate	MPN402	PROS	6.1.1.15	amino acid metabolism	I
M167	R166	[c]: tRNA(Asp) + L-aspartate + ATP ⟶ AMP + L-aspartyl-tRNA(Asp) + pyrophosphate	MPN046	ASPS	6.1.1.12	amino acid metabolism	I
M168	R167	[c]: tRNA(Asn) + L-asparagine + ATP ⟶ AMP + L-asparaginyl-tRNA(Asn) + pyrophosphate	MPN252	ASNS	6.1.1.22	amino acid metabolism	I
M169	R168	[c]: tRNA(Lys) + L-lysine + ATP ⟶ AMP + pyrophosphate + L-lysyl-tRNA(Lys)	MPN277	LYSS	6.1.1.6	amino acid metabolism	I
M170	R169	[c]: L-histidine + ATP + tRNA(His) ⟶ AMP + L-histidyl-tRNA(His) + pyrophosphate	MPN045	HISS	6.1.1.21	amino acid metabolism	I
M171	R170	[c]: tRNA(Phe) + L-phenylalanine + ATP ⟶ AMP + L-phenylalanyl-tRNA(Phe) + pyrophosphate	MPN105/ MPN106	PHES	6.1.1.20	amino acid metabolism	I
M172	R171	[c]: L-alanine + tRNA(Ala) + ATP ⟶ L-alanyl-tRNA(Ala) + AMP + pyrophosphate	MPN418	ALAS	6.1.1.7	amino acid metabolism	I
M173	R172	[c]: glycine + tRNA(Gly) + ATP ⟶ glycyl-tRNA(Gly) + AMP + pyrophosphate	MPN354	GLYS	6.1.1.14	amino acid metabolism	I
M174	-	[c]: (100) H_2O + RNA (Mpn) ⟶ (28) GMP + (29) AMP + (18) CMP + (25) UMP + (100) H+	-	-	-	nucleotide metabolism	I
M175	-	CO_2[e] ⟷ CO_2[c]	?	-	-	cofactor metabolism	R
M176	-	uracil[e] + H+[e] ⟷ uracil[c] + H+[c]	?	-	-	nucleotide metabolism	R
M177	-	thymine[e] + H+[e] ⟶ thymine[c] + H+[c]	?	-	-	nucleotide metabolism	I
M178	-	(S)-lactate[c] + H+[c] ⟶ (S)-lactate[e] + H+[e]	?	-	-	pyruvate metabolism	I
M179	-	oxygen[e] ⟷ oxygen[c]	?	-	-	cofactor metabolism	R
M180	-	riboflavin[e] ⟶ riboflavin[c]	?	-	-	cofactor metabolism	I
M181	-	cytidine[e] + H+[e] ⟷ cytidine[c] + H+[c]	?	-	-	nucleotide metabolism	R
M182	-	L-homocysteine[c] ⟶ L-homocysteine[e]	?	-	-	amino acid metabolism	I
M183	-	guanine[e] + H+[e] ⟶ guanine[c] + H+[c]	?	-	-	nucleotide metabolism	I
M184	-	H_2O_2[e] ⟷ H_2O_2[c]	?	-	-	cofactor metabolism	R
M185	-	pyridoxal[e] ⟶ pyridoxal[c]	?	-	-	cofactor metabolism	I
M186	-	folic acid[e] ⟶ folic acid[c]	?	-	-	one carbon pool by folate	I
M187	-	fatty acid (Mpn)[e] ⟷ fatty acid (Mpn)[c]	?	-	-	CoA metabolism	R
M188	-	H_2O[e] ⟷ H_2O[c]	?	-	-	cofactor metabolism	R
M189	-	adenine[c] + H+[c] ⟷ adenine[e] + H+[e]	?	-	-	nucleotide metabolism	R
M190	-	nicotinate[e] ⟶ nicotinate[c]	?	-	-	cofactor metabolism	I
M191	-	pantetheine[e] ⟶ pantetheine[c]	?	-	-	CoA metabolism	I
M192	-	L-serine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-serine[c] + orthophosphate[c] + H+[c]	?	-	-	one carbon pool by folate	I
M193	-	L-methionine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + orthophosphate[c] + L-methionine[c] + H+[c]	?	-	-	amino acid metabolism	I
M194	-	NH_3[c] ⟷ NH_3[e]	?	-	-	cofactor metabolism	R
M195	-	thiamin[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + orthophosphate[c] + H+[c] + thiamin[c]	?	-	-	cofactor metabolism	I
M196	-	H+[e] + choline[e] ⟷ H+[c] + choline[c]	?	-	-	lipid metabolism	R
M197	-	L-arginine[e] + L-ornithine[c] ⟷ L-ornithine[e] + L-arginine[c]	?	-	-	amino acid metabolism	R
M198	-	H+[c] + acetate[c] ⟶ H+[e] + acetate[e]	?	-	-	pyruvate metabolism	I
M199	-	phosphatidylcholine[e] + ATP[c] + H_2O[c] ⟶ phosphatidylcholine[c] + ADP[c] + orthophosphate[c] + H+[c]	?	-	-	lipid metabolism	I
M200	-	L-alanine[c] + H+[c] ⟶ L-alanine[e] + H+[e]	?	-	-	amino acid metabolism	I
M201	-	L-arginine[c] + H+[c] ⟶ L-arginine[e] + H+[e]	?	-	-	amino acid metabolism	I

Continued on next page

Table A.1 – continued from previous page

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M202	-	L-asparagine[c] + H+[c] ⟶ L-asparagine[e] + H+[e]	?	-	-	amino acid metabolism	I
M203	-	H+[c] + L-aspartate[c] ⟶ H+[e] + L-aspartate[e]	?	-	-	amino acid metabolism	I
M204	-	H+[c] + L-cysteine[c] ⟶ H+[e] + L-cysteine[e]	?	-	-	amino acid metabolism	I
M205	-	L-glutamate[c] + H+[c] ⟶ L-glutamate[e] + H+[e]	?	-	-	amino acid metabolism	I
M206	-	L-glutamine[c] + H+[c] ⟶ L-glutamine[e] + H+[e]	?	-	-	amino acid metabolism	I
M207	-	glycine[c] + H+[c] ⟶ glycine[e] + H+[e]	?	-	-	amino acid metabolism	I
M208	-	L-histidine[c] + H+[c] ⟶ L-histidine[e] + H+[e]	?	-	-	amino acid metabolism	I
M209	-	L-isoleucine[c] + H+[c] ⟶ L-isoleucine[e] + H+[e]	?	-	-	amino acid metabolism	I
M210	-	L-leucine[c] + H+[c] ⟶ H+[e] + L-leucine[e]	?	-	-	amino acid metabolism	I
M211	-	L-lysine[c] + H+[c] ⟶ L-lysine[e] + H+[e]	?	-	-	amino acid metabolism	I
M212	-	H+[c] + L-methionine[c] ⟶ H+[e] + L-methionine[e]	?	-	-	amino acid metabolism	I
M213	-	L-phenylalanine[c] + H+[c] ⟶ L-phenylalanine[e] + H+[e]	?	-	-	amino acid metabolism	I
M214	-	H+[c] + L-proline[c] ⟶ H+[e] + L-proline[e]	?	-	-	amino acid metabolism	I
M215	-	L-serine[c] + H+[c] ⟶ L-serine[e] + H+[e]	?	-	-	amino acid metabolism	I
M216	-	L-threonine[c] + H+[c] ⟶ L-threonine[e] + H+[e]	?	-	-	amino acid metabolism	I
M217	-	L-tryptophan[c] + H+[c] ⟶ L-tryptophan[e] + H+[e]	?	-	-	amino acid metabolism	I
M218	-	L-tyrosine[c] + H+[c] ⟶ L-tyrosine[e] + H+[e]	?	-	-	amino acid metabolism	I
M219	-	L-valine[c] + H+[c] ⟶ L-valine[e] + H+[e]	?	-	-	amino acid metabolism	I
M220	-	deoxycytidine[e] + H+[e] ⟶ deoxycytidine[c] + H+[c]	?	-	-	nucleotide metabolism	I
M221	-	L-arginine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-arginine[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M222	-	L-aspartate[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + orthophosphate[c] + H+[c] + L-aspartate[c]	?	-	-	amino acid metabolism	I
M223	-	L-cysteine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + orthophosphate[c] + H+[c] + L-cysteine[c]	?	-	-	amino acid metabolism	I
M224	-	L-glutamate[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + orthophosphate[c] + L-glutamate[c] + H+[c]	?	-	-	amino acid metabolism	I
M225	-	glycine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + glycine[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M226	-	L-isoleucine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-isoleucine[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M227	-	L-alanine[e] + H_2O[c] + ATP[c] ⟶ L-alanine[c] + ADP[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M228	-	L-asparagine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + orthophosphate[c] + L-asparagine[c] + H+[c]	?	-	-	amino acid metabolism	I
M229	-	L-leucine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + orthophosphate[c] + L-leucine[c] + H+[c]	?	-	-	amino acid metabolism	I
M230	-	L-glutamine[e] + H_2O[c] + ATP[c] ⟶ L-glutamine[c] + ADP[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M231	-	L-histidine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-histidine[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M232	-	L-lysine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-lysine[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M233	-	L-proline[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + orthophosphate[c] + H+[c] + L-proline[c]	?	-	-	amino acid metabolism	I
M234	-	L-phenylalanine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-phenylalanine[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M235	-	L-threonine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-threonine[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M236	-	L-tryptophan[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-tryptophan[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I

Continued on next page

Table A.1 – continued from previous page

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M237	-	L-tyrosine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-tyrosine[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M238	-	L-valine[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + L-valine[c] + orthophosphate[c] + H+[c]	?	-	-	amino acid metabolism	I
M239	-	orthophosphate[e] + H_2O[c] + ATP[c] ⟶ ADP[c] + (2) orthophosphate[c] + H+[c]	MPN609-MPN611	-	-	cofactor metabolism	I
M240	-	orthophosphate[c] ⟶ orthophosphate[e]	?	-	-	cofactor metabolism	I
M241	-	[e]: (S)-lactate ⟶	-	-	-	sink/source	I
M242	-	[c]: 5mcDNA (Mpn) ⟶	-	-	-	sink/source	I
M243	-	[e]: acetate ⟶	-	-	-	sink/source	I
M244	-	[e]: adenine ⟷	-	-	-	sink/source	R
M245	-	[c]: adenosine 3',5'-bisphosphate ⟶	-	-	-	sink/source	I
M246	-	[e]: biomass ⟶	-	-	-	sink/source	R
M247	-	[e]: CO_2 ⟷	-	-	-	sink/source	R
M248	-	[e]: glycine ⟷	-	-	-	sink/source	R
M249	-	[e]: H_2O_2 ⟶	-	-	-	sink/source	I
M250	-	[e]: L-alanine ⟷	-	-	-	sink/source	R
M251	-	[e]: L-arginine ⟷	-	-	-	sink/source	R
M252	-	[e]: L-asparagine ⟷	-	-	-	sink/source	R
M253	-	[e]: L-aspartate ⟷	-	-	-	sink/source	R
M254	-	[e]: L-cysteine ⟷	-	-	-	sink/source	R
M255	-	[e]: L-glutamate ⟷	-	-	-	sink/source	R
M256	-	[e]: L-glutamine ⟷	-	-	-	sink/source	R
M257	-	[e]: L-histidine ⟷	-	-	-	sink/source	R
M258	-	[e]: L-homocysteine ⟶	-	-	-	sink/source	I
M259	-	[e]: L-isoleucine ⟷	-	-	-	sink/source	R
M260	-	[e]: L-leucine ⟷	-	-	-	sink/source	R
M261	-	[e]: L-lysine ⟷	-	-	-	sink/source	R
M262	-	[e]: L-methionine ⟷	-	-	-	sink/source	R
M263	-	[e]: L-ornithine ⟶	-	-	-	sink/source	I
M264	-	[e]: L-phenylalanine ⟷	-	-	-	sink/source	R
M265	-	[e]: L-proline ⟷	-	-	-	sink/source	R
M266	-	[e]: L-serine ⟷	-	-	-	sink/source	R
M267	-	[e]: L-threonine ⟷	-	-	-	sink/source	R
M268	-	[e]: L-tryptophan ⟷	-	-	-	sink/source	R
M269	-	[e]: L-tyrosine ⟷	-	-	-	sink/source	R
M270	-	[e]: L-valine ⟷	-	-	-	sink/source	R
M271	-	[c]: NADP+ ⟶	-	-	-	sink/source	I
M272	-	[e]: NH_3 ⟷	-	-	-	sink/source	R
M273	-	[e]: ⟶ ascorbate	-	-	-	sink/source	I
M274	-	[e]: ⟷ choline	-	-	-	sink/source	R
M275	-	[e]: ⟶ cytidine	-	-	-	sink/source	I
M276	-	[e]: ⟶ deoxycytidine	-	-	-	sink/source	I
M277	-	[e]: ⟶ D-fructose	-	-	-	sink/source	I
M278	-	[e]: ⟶ D-mannose	-	-	-	sink/source	I
M279	-	[e]: ⟶ D-ribose	-	-	-	sink/source	I
M280	-	[e]: ⟷ fatty acid (Mpn)	-	-	-	sink/source	R
M281	-	[e]: ⟶ folic acid	-	-	-	sink/source	I
M282	-	[e]: ⟶ D-glucose	-	-	-	sink/source	I
M283	-	[e]: ⟶ glycerol	-	-	-	sink/source	I
M284	-	[e]: ⟶ guanine	-	-	-	sink/source	I
M285	-	[e]: ⟷ H+	-	-	-	sink/source	R
M286	-	[e]: ⟶ mannitol	-	-	-	sink/source	I
M287	-	[e]: ⟶ nicotinate	-	-	-	sink/source	I
M288	-	[e]: ⟷ orthophosphate	-	-	-	sink/source	R
M289	-	[e]: ⟶ oxygen	-	-	-	sink/source	I
M290	-	[e]: ⟶ pantetheine	-	-	-	sink/source	I
M291	-	[e]: ⟶ phosphatidylcholine	-	-	-	sink/source	I
M292	-	[e]: ⟶ pyridoxal	-	-	-	sink/source	I
M293	-	[e]: ⟶ riboflavin	-	-	-	sink/source	I
M294	-	[e]: ⟶ sn-glycerol 3-phosphate	-	-	-	sink/source	I
M295	-	[e]: ⟶ thiamin	-	-	-	sink/source	I
M296	-	[e]: ⟶ thymine	-	-	-	sink/source	I
M297	-	[e]: ⟶ uracil	-	-	-	sink/source	I
M298	-	[e]: ⟷ H_2O	-	-	-	sink/source	R

Continued on next page

Table A.1 – continued from previous page

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M299	-	[c]: (25) L-valyl-tRNA(Val) + (18) L-aspartyl-tRNA(Asp) + (13) L-arginyl-tRNA(Arg) + N-formylmethionyl-tRNA(Met) + (826) ATP + (32) L-lysyl-tRNA(Lys) + (17) L-glutaminyl-tRNA(Gln) + (22) L-isoleucyl-tRNA(Ile) + (482) H_2O + (21) L-seryl-tRNA(Ser) + (14) L-prolyl-tRNA(Pro) + (21) glycyl-tRNA(Gly) + (27) L-alanyl-tRNA(Ala) + (10) L-tyrosyl-tRNA(Tyr) + (21) L-threonyl-tRNA(Thr) + (20) L-asparaginyl-tRNA(Asn) + (5) L-methionyl-tRNA(Met) + (6) L-histidyl-tRNA(His) + (3) L-cysteinyl-tRNA(Cys) + (22) L-glutamyl-tRNA(Glu) + (3) L-tryptophanyl-tRNA(Trp) + (15) L-phenylalanyl-tRNA(Phe) + (29) L-leucyl-tRNA(Leu) ⟶ (3) tRNA(Trp) + (14) tRNA(Pro) + (10) tRNA(Tyr) + (29) tRNA(Leu) + (6) tRNA(Met) + (13) tRNA(Arg) + (22) tRNA(Glu) + (3) tRNA(Cys) + (27) tRNA(Ala) + (6) tRNA(His) + (32) tRNA(Lys) + (18) tRNA(Asp) + (25) tRNA(Val) + (826) H+ + (22) tRNA(Ile) + (21) tRNA(Ser) + (826) orthophosphate + (17) tRNA(Gln) + (826) ADP + (15) tRNA(Phe) + (21) tRNA(Gly) + (20) tRNA(Asn) + protein (Mpn) + (21) tRNA(Thr)	-	-	-	Protein metabolism	I
M300	-	[c]: H_2O + protein (Mpn) + ATP ⟶ ADP + orthophosphate + DnaK-folded protein (Mpn) + H+	MPN434/ MPN002/ MPN120	DnaK	-	Protein metabolism	I
M301	-	[c]: (7) H_2O + protein (Mpn) + (7) ATP ⟶ (7) ADP + GroEL-folded protein (Mpn) + (7) orthophosphate + (7) H+	MPN573/ MPN574	GroEL	-	Protein metabolism	I
M302	-	[c]: (1032) H_2O + protein (Mpn) + (688) ATP ⟶ (27) L-alanine + (10) L-tyrosine + (21) glycine + (3) L-tryptophan + (21) L-serine + (20) L-asparagine + (22) L-glutamate + (29) L-leucine + (688) H+ + (14) L-proline + (3) L-cysteine + (13) L-arginine + (32) L-lysine + (21) L-threonine + (15) L-phenylalanine + (688) orthophosphate + (18) L-aspartate + (6) L-methionine + (688) ADP + (17) L-glutamine + (22) L-isoleucine + (25) L-valine + (6) L-histidine	-	-	-	Protein metabolism	I
M303	-	[c]: (5) L-valyl-tRNA(Val) + (6) L-aspartyl-tRNA(Asp) + (2) L-arginyl-tRNA(Arg) + N-formylmethionyl-tRNA(Met) + (199) ATP + (11) L-lysyl-tRNA(Lys) + (4) L-glutaminyl-tRNA(Gln) + (8) L-isoleucyl-tRNA(Ile) + (199) H_2O + (5) L-seryl-tRNA(Ser) + L-prolyl-tRNA(Pro) + glycyl-tRNA(Gly) + (4) L-alanyl-tRNA(Ala) + L-threonyl-tRNA(Thr) + (3) L-asparaginyl-tRNA(Asn) + (3) L-methionyl-tRNA(Met) + L-histidyl-tRNA(His) + (11) L-glutamyl-tRNA(Glu) + (5) L-phenylalanyl-tRNA(Phe) + (12) L-leucyl-tRNA(Leu) ⟶ tRNA(Pro) + (12) tRNA(Leu) + (4) tRNA(Met) + (2) tRNA(Arg) + (11) tRNA(Glu) + (4) tRNA(Ala) + tRNA(His) + (11) tRNA(Lys) + (6) tRNA(Asp) + (5) tRNA(Val) + (282) H+ + (8) tRNA(Ile) + apoprotein [acyl carrier protein] + (5) tRNA(Ser) + (199) orthophosphate + (4) tRNA(Gln) + (199) ADP + (5) tRNA(Phe) + tRNA(Gly) + (3) tRNA(Asn) + tRNA(Thr)	-	-	-	Protein metabolism	I
M304	-	[c]: H_2O + ATP ⟶ ADP + orthophosphate + H+	-	-	-	Energy metabolism	I

Continued on next page

Table A.1 – continued from previous page

reaction ID	ID in 'YUS'	equation	gene ID	enzyme (short)	EC number	pathway	reversibility
M305	-	[c]: (1657) guanosine + (1366) L-tyrosine + (9220) glycine + (1864) L-tryptophan + (54) L-asparagine + (100) S-adenosyl-L-methionine + (100) pyridoxal phosphate + (67) L-cysteine + a100) thiamin diphosphate + (2489) L-threonine + (100) 5-formyltetrahydrofolate + (1242) thymidine + (2793) L-valine + (1681) DNA (Mpn) + (1981) adenosine + (100) CoA + (9824) L-alanine + (3202) L-serine + (2015) RNA (Mpn) + (63702) glycolipid (Mpn) + (503) cytidine + (167760) D-glucose 6-phosphate + (3) acyl carrier protein + (100) CDP-choline + (18651) L-glutamate + (21313) L-leucine + (6837) L-proline + (100) NADP+ + (3913) L-arginine + (1741) L-lysine + (5122) L-phenylalanine + (1370) L-methionine + (9318) L-aspartate + (100) FAD + (294) L-glutamine + (1858) L-isoleucine + (148168) phosphatidic acid (Mpn) + (16949) protein (Mpn) + (2422) L-histidine + (100) NADPH + (2541) uridine ⟶ (1000000) biomass	-	-	-	Biomass production	I
M306	-	biomass[c] + (25) H_2O[c] + (25) ATP[c] ⟶ biomass[e] + (25) ADP[c] + (25) orthophosphate[c] + (25) H+[c]	-	-	-	Biomass production	I

Table A.1.: Reactions included in the model: *reaction ID* (used throughout the main text and in Appendix A, figure A.1 to identify the model reactions), *ID in 'YUS'* (given for reactions also included in the metabolic network published by Yus et al. [2009]), *equation* (reaction equation), *gene ID* (the *M. pneumoniae* enzyme catalyzing the reaction), *enzyme (short)* (shortcut used throughout the text and in the figures), *EC number* , *pathway* (pathway a reactions is assigned to), and *reversibility* (reaction reversibility) are specified.

Table A.2 : Reaction Reversibilities

model ID	A	B	C	D	E	model ID	A	B	C	D	E	model ID	A	B	C	D	E
M001	I	I	I	-	I	M104	I	-	-	-	I	M207	I	I	I	-	-
M002	R	R	R	-	R	M105	I	I	R	-	I	M208	I	R	R	-	-
M003	I	I	R	-	R	M106	R	R	R	-	R	M209	I	R	I	-	-
M004	R	R	R	R	R	M107	I	-	-	-	I	M210	I	R	I	-	-
M005	R	R	R	-	R	M108	I	I	R	-	R	M211	I	I	I	-	-
M006	R	R	R	-	R	M109	R	R	R	-	R	M212	I	-	-	-	-
M007	R	R	R	-	R	M110	I	I	R	-	I	M213	I	R	I	-	-
M008	R	R	I	-	R	M111	I	I	R	-	I	M214	I	R	R	-	-
M009	R	R	I	-	R	M112	R	R	R	-	R	M215	I	R	I	-	-
M010	I	I	R	-	I	M113	R	I	R	-	I	M216	I	R	I	-	-
M011	R	-	-	-	R	M114	R	R	R	-	R	M217	I	R	I	-	-
M012	I	-	-	-	I	M115	I	I	R	-	I	M218	I	R	I	-	-
M013	R	-	-	-	R	M116	I	I	R	-	I	M219	I	R	I	-	-
M014	R	-	-	-	R	M117	R	R	R	R	R	M220	I	I	R	-	-
M015	R	R	R	-	R	M118	I	-	-	-	I	M221	I	-	-	-	-
M016	R	R	R	R	R	M119	I	-	-	-	I	M222	I	-	-	-	-
M017	R	-	-	-	I	M120	R	R	R	-	R	M223	I	-	-	-	-
M018	R	R	I	-	R	M121	R	R	I	-	R	M224	I	-	-	-	-
M019	I	I	I	-	I	M122	R	-	-	-	R	M225	I	-	-	-	-
M020	I	I	R	-	I	M123	R	-	-	-	R	M226	I	-	-	-	-
M021	I	-	-	R	R	M124	I	-	-	-	R	M227	I	-	-	-	-
M022	R	R	R	-	R	M125	R	R	R	-	R	M228	I	-	-	-	-
M023	I	I	I	-	I	M126	I	-	-	-	I	M229	I	-	-	-	-
M024	R	R	R	-	R	M127	R	R	R	-	R	M230	I	-	-	-	-
M025	I	I	I	-	I	M128	R	R	R	-	R	M231	I	-	-	-	-
M026	R	R	R	-	R	M129	I	-	-	-	I	M232	I	-	-	-	-
M027	I	-	-	-	I	M130	I	-	-	-	I	M233	I	-	-	-	-
M028	R	-	-	-	R	M131	R	-	-	-	R	M234	I	-	-	-	-
M029	I	I	I	-	I	M132	I	I	I	-	I	M235	I	-	-	-	-
M030	R	I	R	-	R	M133	I	-	-	-	I	M236	I	-	-	-	-
M031	R	R	R	-	R	M134	I	I	R	R	R	M237	I	-	-	-	-
M032	I	-	-	-	I	M135	R	-	-	R	I	M238	I	-	-	-	-

Continued on next page

Table A.2 – continued from previous page

model ID	A	B	C	D	E	model ID	A	B	C	D	E	model ID	A	B	C	D	E
M033	I	-	-	-	I	M136	R	R	R	R	R	M239	I	-	-	-	-
M034	R	R	R	-	R	M137	I	I	I	-	I	M240	I	-	-	-	-
M035	R	R	R	-	R	M138	I	I	R	-	I	M241	I	-	-	-	-
M036	R	-	-	-	R	M139	I	-	-	-	I	M242	I	-	-	-	-
M037	R	R	R	-	R	M140	I	I	I	-	I	M243	I	-	-	-	-
M038	R	-	-	-	R	M141	I	I	I	R	R	M244	R	-	-	-	-
M039	I	-	-	R	R	M142	I	I	R	-	I	M245	I	-	-	-	-
M040	R	-	-	-	R	M143	I	I	R	-	I	M246	R	-	-	-	-
M041	R	R	R	-	R	M144	I	I	R	-	I	M247	R	-	-	-	-
M042	I	I	R	-	R	M145	I	-	-	-	I	M248	R	-	-	-	-
M043	R	R	R	-	R	M146	I	I	R	R	R	M249	I	-	-	-	-
M044	R	R	R	-	R	M147	I	I	I	-	I	M250	R	-	-	-	-
M045	I	I	R	-	I	M148	I	I	I	R	R	M251	R	-	-	-	-
M046	I	I	R	-	I	M149	I	-	-	-	I	M252	R	-	-	-	-
M047	R	-	-	R	R	M150	I	I	R	-	R	M253	R	-	-	-	-
M048	I	-	-	-	I	M151	I	-	-	-	I	M254	R	-	-	-	-
M049	I	-	-	-	I	M152	I	-	-	-	I	M255	R	-	-	-	-
M050	I	-	-	-	I	M153	I	-	-	-	I	M256	R	-	-	-	-
M051	I	-	-	-	R	M154	I	-	-	-	R	M257	R	-	-	-	-
M052	I	-	-	-	R	M155	I	-	-	-	R	M258	I	-	-	-	-
M053	I	-	-	-	R	M156	I	-	-	-	R	M259	R	-	-	-	-
M054	I	-	-	-	R	M157	I	-	-	-	R	M260	R	-	-	-	-
M055	I	-	-	-	I	M158	I	-	-	-	R	M261	R	-	-	-	-
M056	I	-	-	-	I	M159	I	I	I	-	R	M262	R	-	-	-	-
M057	R	R	R	-	R	M160	I	-	-	-	R	M263	I	-	-	-	-
M058	R	I	R	-	R	M161	I	-	-	-	R	M264	R	-	-	-	-
M059	R	R	R	-	R	M162	I	-	-	-	R	M265	R	-	-	-	-
M060	I	-	-	-	I	M163	I	-	-	-	R	M266	R	-	-	-	-
M061	I	-	-	-	R	M164	I	-	-	-	R	M267	R	-	-	-	-
M062	I	-	-	-	R	M165	I	-	-	-	R	M268	R	-	-	-	-
M063	R	-	-	-	R	M166	I	-	-	-	R	M269	R	-	-	-	-
M064	R	-	-	-	R	M167	I	-	-	-	R	M270	R	-	-	-	-
M065	R	I	R	R	R	M168	I	-	-	-	R	M271	I	-	-	-	-
M066	R	R	R	-	R	M169	I	-	-	-	R	M272	R	-	-	-	-
M067	I	I	I	-	I	M170	I	-	-	-	R	M273	R	-	-	-	-
M068	R	R	R	-	R	M171	I	-	-	-	R	M274	I	-	-	-	-
M069	I	I	I	-	I	M172	I	-	-	-	R	M275	R	-	-	-	-
M070	R	I	R	-	I	M173	I	-	-	-	R	M276	I	-	-	-	-
M071	I	-	-	-	I	M174	I	-	-	-	-	M277	I	-	-	-	-
M072	I	I	I	-	I	M175	R	R	I	-	-	M278	I	-	-	-	-
M073	R	R	R	-	R	M176	R	R	R	-	-	M279	I	-	-	-	-
M074	R	-	-	R	I	M177	I	-	-	-	-	M280	I	-	-	-	-
M075	R	R	R	-	R	M178	I	-	-	-	-	M281	R	-	-	-	-
M076	I	-	-	-	I	M179	R	R	R	-	-	M282	I	-	-	-	-
M077	R	-	-	-	R	M180	I	-	-	-	-	M283	I	-	-	-	-
M078	I	-	-	-	I	M181	R	R	R	-	-	M284	I	-	-	-	-
M079	R	I	R	R	R	M182	I	-	-	-	-	M285	I	-	-	-	-
M080	R	R	R	-	R	M183	I	I	R	-	-	M286	R	-	-	-	-
M081	I	I	R	-	I	M184	R	-	-	-	-	M287	I	-	-	-	-
M082	R	R	R	-	R	M185	I	-	-	-	-	M288	I	-	-	-	-
M083	I	I	R	-	I	M186	I	-	-	-	-	M289	R	-	-	-	-
M084	I	-	-	-	I	M187	R	-	-	-	-	M290	I	-	-	-	-
M085	R	-	-	-	R	M188	R	R	R	-	-	M291	I	-	-	-	-
M086	R	R	R	-	R	M189	R	R	I	-	-	M292	I	-	-	-	-
M087	R	-	-	R	I	M190	I	-	-	-	-	M293	I	-	-	-	-
M088	R	R	R	-	R	M191	I	-	-	-	-	M294	I	-	-	-	-
M089	I	-	-	-	I	M192	I	-	-	-	-	M295	I	-	-	-	-
M090	R	-	-	-	R	M193	I	I	I	-	-	M296	I	-	-	-	-
M091	R	I	R	R	R	M194	R	-	-	-	-	M297	I	-	-	-	-
M092	R	R	R	-	R	M195	I	I	R	-	-	M298	I	-	-	-	-
M093	I	I	R	-	I	M196	R	I	R	-	-	M299	R	-	-	-	-
M094	I	-	-	-	I	M197	R	R	R	-	-	M300	I	-	-	-	-
M095	R	R	R	-	R	M198	I	-	-	-	-	M301	I	-	-	-	-
M096	I	-	-	-	I	M199	I	-	-	-	-	M302	I	-	-	-	-
M097	I	I	R	-	I	M200	I	I	I	-	-	M303	I	-	-	-	-
M098	I	I	R	-	R	M201	I	-	-	-	-	M304	I	-	-	-	-
M099	I	-	-	-	I	M202	I	R	I	-	-	M305	I	-	-	-	-
M100	I	I	R	-	I	M203	I	I	R	-	-	M306	I	-	-	-	-
M101	R	R	R	-	R	M204	I	-	-	-	-	M307	I	-	-	-	-
M102	I	-	-	-	I	M205	I	R	R	-	-						
M103	I	I	R	-	I	M206	I	-	-	-	-						

Table A.2.: Reaction reversibilities. A: in the presented model. B: according to quantitative reversibilities in *E.coli* [Fleming et al., 2009]. C: according to qualitative reversibilities in *E.coli* [Fleming et al., 2009]. D: according to BRENDA enzyme DB [Scheer et al., 2011]. E: in the initial reconstruction from Yus et al. [2009].

Table A.3 : Branching Metabolites

Metabolite	No of reac-tions	experi-mentally amenable	Metabolite	No of reac-tions	experi-mentally amenable
(S)-lactate	2	Y	L-ascorbate 6-phosphate	2	Y
1-acyl-glycerol 3-phosphate	2	Y	L-asparagine	6	Y
1,3-bisphospho-D-glycerate	5	Y	L-asparaginyl-tRNA(Asn)	3	N
10-formyltetrahydrofolate	3	Y	L-aspartate	6	Y
2-deoxy-D-ribose 1-phosphate	5	Y	L-aspartyl-tRNA(Asp)	3	N
2-deoxy-D-ribose 5-phosphate	2	Y	L-citrulline	2	Y
2-phospho-D-glycerate	2	Y	L-cysteine	5	Y
3-keto-L-gulonate 6-phosphate	2	Y	L-cysteinyl-tRNA(Cys)	2	N
3-phospho-D-glycerate	5	Y	L-glutamate	7	Y
5-formyltetrahydrofolate	3	Y	L-glutamine	5	Y
5-phospho-alpha-D-ribose 1-diphosphate	5	Y	L-glutaminyl-tRNA(Gln)	4	N
5,10-methenyltetrahydrofolate	4	Y	L-glutamyl-tRNA(Gln)	3	N
5,10-methylenetetrahydrofolate	3	Y	L-glutamyl-tRNA(Glu)	3	N
5mcDNA (Mpn)	1	N	L-histidine	5	Y
acetaldehyde	2	Y	L-histidyl-tRNA(His)	3	N
acetate	2	Y	L-homocysteine	2	Y
acetyl phosphate	2	Y	L-isoleucine	5	Y
acetyl-CoA	3	Y	L-isoleucyl-tRNA(Ile)	3	N
ACP-R (Mpn)	3	N	L-leucine	5	Y
acyl carrier protein	6	N	L-leucyl-tRNA(Leu)	3	N
adenine	4	Y	L-lysine	5	Y
adenosine	4	Y	L-lysyl-tRNA(Lys)	3	N
adenosine 3',5'-bisphosphate	2	Y	L-methionine	6	Y
ADP	73	Y	L-methionyl-tRNA(Met)	4	N
AMP	28	Y	L-ornithine	2	Y
apoprotein [acyl carrier protein]	3	N	L-phenylalanine	5	Y
ATP	99	Y	L-phenylalanyl-tRNA(Phe)	3	N
biomass	2	N	L-proline	5	Y
carbamoyl phosphate	2	Y	L-prolyl-tRNA(Pro)	3	N
cardiolipin	1	N	L-ribulose 5-phosphate	2	Y
CDP	3	Y	L-serine	6	Y
CDP-choline	1	Y	L-seryl-tRNA(Ser)	3	N
CDP-diacylglycerol (Mpn)	3	Y	L-threonine	5	Y
choline	3	Y	L-threonyl-tRNA(Thr)	3	N
choline phosphate	2	Y	L-tryptophan	5	Y
CMP	6	Y	L-tryptophanyl-tRNA(Trp)	2	N
CO2	5	N	L-tyrosine	5	Y
CoA	6	Y	L-tyrosyl-tRNA(Tyr)	2	N
CTP	4	Y	L-valine	5	Y
cytidine	5	Y	L-valyl-tRNA(Val)	3	N
D-erythrose 4-phosphate	2	Y	L-xylulose 5-phosphate	2	Y
D-fructose	1	Y	lipoamide	2	Y
D-fructose 1-phosphate	3	Y	mannitol	1	Y
D-fructose 1,6-bisphosphate	3	Y	N-formylmethionyl-tRNA(Met)	3	N
D-fructose 6-phosphate	6	Y	NAD+	12	Y
D-glucose	1	Y	NADH	11	Y
D-glucose 1-phosphate	2	Y	NADP+	4	Y
D-glucose 6-phosphate	4	Y	NADPH	4	Y
D-glyceraldehyde	2	Y	NH3	6	N
D-glyceraldehyde 3-phosphate	7	Y	nicotinate	2	Y
D-mannitol 1-phosphate	2	Y	nicotinate D-ribonucleotide	2	Y
D-mannose	1	Y	orthophosphate	62	N
D-mannose 6-phosphate	2	Y	oxidized thioredoxin	5	N
D-ribose	2	Y	oxygen	3	N
D-ribose 1-phosphate	4	Y	pantetheine	2	Y
D-ribose 5-phosphate	5	Y	pantetheine 4-phosphate	3	Y
D-ribulose 5-phosphate	2	Y	phosphatidic acid (Mpn)	4	Y
D-xylulose 5-phosphate	4	Y	phosphatidylcholine	2	Y
dADP	4	Y	phosphatidylglycerol (Mpn)	2	Y
dAMP	2	Y	phosphatidylglycerol 3-phosphate (Mpn)	2	Y
dATP	3	Y	phosphoenolpyruvate	14	Y
dCDP	3	Y	protein (Mpn)	5	N
dCMP	3	Y	pyridoxal	2	Y
dCTP	2	Y	pyridoxal phosphate	2	Y
deamino-NAD+	2	Y	pyrophosphate	36	N
deoxyadenosine	2	Y	pyruvate	15	Y
deoxycytidine	4	Y	reduced thioredoxin	5	N
deoxyguanosine	2	Y	riboflavin	2	Y
deoxyuridine	4	Y	RNA (Mpn)	3	N
dephospho-CoA	2	Y	S-acetyldihydrolipoamide	2	Y
dGDP	4	Y	S-adenosyl-L-homocysteine	2	Y
dGMP	2	Y	S-adenosyl-L-methionine	3	Y
dGTP	3	Y	sedoheptulose 7-phosphate	2	Y
diacylglycerol (Mpn)	2	Y	sn-glycero-3-phosphocholine	2	Y
dihydrofolate	3	Y	sn-glycerol 3-phosphate	6	Y
dihydrolipoamide	2	Y	tetrahydrofolate	4	Y
dihydroxyacetone phosphate	6	Y	thiamin	2	Y
DNA (Mpn)	3	N	thiamin diphosphate	2	Y
DnaK-folded protein (Mpn)	1	N	thiamin monophosphate	2	Y
dTDP	2	Y	thymidine	4	Y

Continued on next page

Table A.3 – continued from previous page

Metabolite	No of reactions	experimentally amenable	Metabolite	No of reactions	experimentally amenable
dTMP	4	Y	thymine	2	Y
dTTP	2	Y	tRNA(Ala)	3	N
dUDP	2	N	tRNA(Arg)	3	N
dUMP	4	N	tRNA(Asn)	3	N
FAD	2	Y	tRNA(Asp)	3	N
fatty acid (Mpn)	3	Y	tRNA(Cys)	2	N
FMN	2	Y	tRNA(Gln)	3	N
folic acid	2	Y	tRNA(Glu)	3	N
formate	2	Y	tRNA(Gly)	3	N
GDP	4	Y	tRNA(His)	3	N
glycerol	3	Y	tRNA(Ile)	3	N
glycerone	2	Y	tRNA(Leu)	3	N
glycine	6	Y	tRNA(Lys)	3	N
glycolipid (Mpn)	2	N	tRNA(Met)	3	N
glycyl-tRNA(Gly)	3	N	tRNA(Phe)	3	N
GMP	4	Y	tRNA(Pro)	3	N
GroEL-folded protein (Mpn)	1	N	tRNA(Ser)	3	N
GTP	3	Y	tRNA(Thr)	3	N
guanine	4	Y	tRNA(Trp)	2	N
guanosine	3	Y	tRNA(Tyr)	2	N
H+	133	N	tRNA(Val)	3	N
H2O	68	N	UDP	4	Y
H2O2	2	N	UDP-galactose	2	Y
L-alanine	5	Y	UDP-glucose	3	Y
L-alanyl-tRNA(Ala)	3	N	UMP	5	Y
L-arginine	7	Y	uracil	4	Y
L-arginyl-tRNA(Arg)	3	N	uridine	5	Y
L-ascorbate	1	Y	UTP	3	Y

Table A.3.: Number of metabolic reactions each metabolite participates in and if the metabolite is amenable to experimental verification (Y: yes, N: no).

Table A.4: Experimentally Identified Metabolites

metabolite name	A	B	C	metabolite name	A	B	C
D-fructose 1,6-bisphosphate	Y	Y		leucine		Y	Y
D-glucose 1-phosphate	Y	Y		lysine		Y	Y
D-glucose 6-phosphate	Y	Y		methionine		Y	Y
sn-glycerol 3-phosphate	Y	Y		phenylalanine		Y	Y
ADP	Y		Y	proline		Y	Y
ATP	Y		Y	serine		Y	Y
NAD+	Y		Y	threonine		Y	Y
UDP-D-galactose	Y		Y	tryptophane		Y	Y
UDP-D-glucose	Y		Y	tyrosine		Y	Y
3-phospho-D-glycerate	Y			Valine		Y	Y
ADP-glucose	Y			glycerol		Y	
AMP	Y			fructose		Y	
CDP	Y			adenine		Y	
CMP	Y			alanine		Y	
CTP	Y			asparagine		Y	
dADP	Y			cholesterol		Y	
dAMP	Y			cysteine		Y	
dATP	Y			cytosine		Y	
dGDP	Y			DHAP		Y	
dGMP	Y			fatty acids (various)		Y	
dGTP	Y			D-fructose 6-phosphate		Y	
D-mannitol 1-phosphate	Y			D-glyceraldehyde 3-phosphate		Y	
FAD	Y			glucose		Y	
GDP	Y			glutamine		Y	
GMP	Y			guanine		Y	
GTP	Y			phosphoenolpyruvate		Y	
NADH	Y			D-ribose		Y	
NADP+	Y			D-ribose 5-phosphate		Y	
NADPH	Y			thymine		Y	
riboflavin	Y			uracil		Y	
thiamine diphosphate	Y			adenosine		Y	
dTMP	Y			guanosine		Y	
dTDP	Y			uridine		Y	
dTTP	Y			(S)-lactate			Y
UDP	Y			acetate			Y
UMP	Y			choline			Y
UTP	Y			ethanol			Y
arginine		Y	Y	formate			Y
aspartate		Y	Y	fumarate			Y
glutamate		Y	Y	ornithine			Y
glycine		Y	Y	pyruvate			Y
histidine		Y	Y	succinate			Y
isoleucine		Y	Y	trans-4-hydroxy-L-proline			Y

Table A.4.: For each encountered metabolite the technique(s) that successfully identified it are indicated. A: LC-MS; B: GC-MS; C: NMR.

Table A.5: Retention Times for NUBS and cholesterol

Compound	RT (min)	Ions (m/z)
adenine	7.66	279, 264, 192
thymine	5.33	270, 255, 113
cytosine	6.03	254, 240, 170
guanine	8.76	367, 352, 280
uracil	4.94	256, 241, 147
adenosine	10.56	540, 322, 230
guanosine	11.02	643, 410, 324
thymidine	10	458, 353, 199
uridine	9.98	517, 348, 217
cytidine	10.75	516, 348, 223
cholesterol	12.07	458, 368, 329
nLeu (ISTD)	4.83	260,232, 158

Table A.5.: Retention times (RT) and characteristic ions used for monitoring bases, nucleosides, and cholesterol after per-trimethylsilyl derivatisation; underlined ions were used for quantification.

Table A.6: Retention Times foi Amino Acids

Amino acid	RT (min)	Ions (m/z)
arginine	6.89	630, 573, 286
alanine	3.65	260, 232, 158
glycine	3.77	246, 218, 144
valine	4.26	288, 260, 186
leucine	4.46	302, 274, 200
isoleucine	4.62	302, 274, 200
methionine	5.67	320, 292, 218
serine	5.77	390, 362, 288
threonin	5.9	404, 376, 303
phenylalanine	6.17	336, 302, 234
aspartate	6.43	418, 390, 316
cystein	6.62	406, 378, 304
glutamate	6.87	432, 330, 272
prolin	4.78	286, 258, 184
asparagine	6.99	417, 302, 315
lysin	7.26	431, 329, 300
glutamine	7.41	431, 357, 329
histidine	7.97	440, 338, 196
tyrosine	8.14	466, 364, 302
tryptophan	8.94	489, 302, 244
nLeu (ISTD)	4.7	302, 274, 200

Table A.6.: Retention times (RT) and characteristic ions for monitoring amino acids; underlined ions were used for quantification.

Table A.7: Amino Acid Quantification

amino acid	no of aa in proteome	no of aa in cytosol	enrichment factor	import/min for t_d = 20 hours	internal pool turnover (min)
leucine	2393036	25097.06	95.35	1994.20	12.59
glutamate	1672072	98709.28	16.94	1393.39	70.84
alanine	2032737	38012.46	53.48	1693.95	22.44
valine	1864124	12050.33	154.69	1553.44	7.76
isoleucin	1724690	17162.65	100.49	1437.24	11.94
serine	1515135	12566.03	120.57	1262.61	9.95
glycin	1588085	34936.66	45.46	1323.40	26.40
aspartate	1305481	23289.71	56.05	1087.90	21.41
threonine	1533247	6433.41	238.33	1277.71	5.04
phenylalanine	1102859	39005.54	28.27	919.05	42.44
arginine	907229	38302.25	23.69	756.02	50.66
prolin	974850	35012.34	27.84	812.38	43.10
asparagine	1473974	13655.65	107.94	1228.31	11.12
methionine	469166	12369.48	37.93	390.97	31.64
glutamine	1179691	7692.71	153.35	983.08	7.83
lysin	2149849	24800.01	86.69	1791.54	13.84
histidine	478238	12279.47	38.95	398.53	30.81
cystein	197940	3627.55	54.57	164.95	21.99
tyrosin	748700	5417.45	138.20	623.92	8.68
tryptophan	211534	6470.83	32.69	176.28	36.71

Table A.7.: Amino acids detected experimentally in the *M. pneumoniae* cytosol and the growth medium. In addition, the enrichment factor for amino acids in the cytosol, the import rate per minute when assuming exponential growth (d_t := doubling time) and the turnover time for the intracellular amino acid pools have been calculated.

Table A.8: Variables for Metabolite Fittings

variable name	a	b	c	d
glucose concentration	10.15	59.788	1.0797	111.62
acetic acid concentration	-0.40559	16.414	1.0455	44.465
lactic acid concentration	3.4644	91.127	1.1056	69.738
protein concentration	0.0055678	0.048346	1.0956	41.969
maintenance costs	0.4364	-50.722	2716.5	-6656

Table A.8.: Variable values for the applied fittings of mathematical equations (see section 3.2.1, Equations 3.1 - 3.3) to metabolite and protein concentration changes determined *in vivo*, as well as to the maintenance costs *in silico*.

Table A.9: Growth Constraints

	24h		36h		48h		60h		defined medium	
reaction ID	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**	**min**	**max**
M048	0.01	inf	0.01	inf	0.01	inf	0.01	inf	0.01	inf
M130	0.01	inf	0.01	inf	0.01	inf	0.01	inf	0.01	inf
M302	0.000349	inf	0.000349	inf	0.000349	inf	0.000349	inf	0.000349	inf
M174	0.007741	inf	0.007741	inf	0.007741	inf	0.007741	inf	0.007741	inf
M251	0	0.25	0	0.25	0	0.25	0	0.25	0	0.25
M273	0	0	0	0	0	0	0	0	0	0
M277	0	0	0	0	0	0	0	0	0	0
M278	0	0	0	0	0	0	0	0	0	0
M279	0	0.127709	0	0.184225	0	0.242428	0	0.309440	0	0
M282	0	5.108358	0	7.369004	0	9.697136	0	12.377609	0	7.369004
M283	0	0.127709	0	0.184225	0	0.242428	0	0.309440	0	0.184225
M286	0	0	0	0	0	0	0	0	0	0
M291	0	0	0	0	0	0	0	0	0	0
M294	0	0.127709	0	0.184225	0	0.242428	0	0.309440	0	0.184225
M243	0	7.4436	0	6.9311	0	4.6191	0	2.6777	0	6.9311
M304	13.105	inf	16.9625	inf	20.435	inf	25.2015	inf	-	-

Table A.9.: Final constraints as used for simulating growth of *M. pneumoniae* at different time points of the exponential growth phase under rich medium conditions and to simulate growth in minimal medium. For minimal medium simulations, glucose and acetate have been constrained as for the 36 hour time point under rich medium conditions (for details see section 3.2.1).

Table A.10: Quantitative *In Silico* Knock-out Results

gene ID	objective value	gene ID	objective value	gene ID	objective value
MPN005	-	MPN259	0.32959274	MPN480	-
MPN006	0	MPN260	0.32959274	MPN483	0
MPN017	-	MPN265	-	MPN492	0.33468981
MPN023	-	MPN267	0	MPN493	0.33468981
MPN025	-	MPN277	-	MPN494	0.33468981
MPN033	0.33392743	MPN298	0	MPN495	0.33468981
MPN034	0	MPN299	0	MPN496	0.33468981
MPN043	0.33163157	MPN300	0.33468981	MPN497	0.33468981
MPN044	0.33303073	MPN302	-	MPN498	0.33468981
MPN045	-	MPN303	-	MPN516	-
MPN046	-	MPN304	0.33468981	MPN520	-
MPN047	0	MPN305	0.33468981	MPN528	-
MPN050	0.33163157	MPN306	0.33171012	MPN532	0
MPN051	0.32782108	MPN307	0.33171012	MPN533	0.21934261
MPN060	0	MPN320	0.32997561	MPN543	-
MPN062	0.33043678	MPN321	0.33468981	MPN546	0
MPN064	0.32870802	MPN322	0	MPN547	-
MPN065	0.32511847	MPN323	0	MPN550	0
MPN066	0	MPN324	0	MPN553	-
MPN073	0.33468981	MPN336	0	MPN556	-
MPN078	0.33468981	MPN348	-	MPN560	0.33468981
MPN079	0.33468981	MPN350	0	MPN561	0
MPN082	0.33163294	MPN354	-	MPN562	0
MPN105	-	MPN356	-	MPN564	0.33468981
MPN106	-	MPN378	0	MPN576	0
MPN108	0.33468981	MPN382	0	MPN595	0.33163294
MPN133	0.32857332	MPN384	-	MPN606	-
MPN134	0.32857332	MPN386	0.32791069	MPN609	0.2512049
MPN135	0.32857332	MPN390	0.21934261	MPN610	0.2512049
MPN136	0.32857332	MPN391	0.21934261	MPN611	0.2512049
MPN158	0	MPN392	0.21934261	MPN627	-
MPN185	-	MPN393	0.21934261	MPN628	-
MPN207	-	MPN394	0.21942224	MPN629	-
MPN236	-	MPN395	0	MPN632	-
MPN237	-	MPN402	-	MPN637	0.33468981
MPN238	-	MPN418	-	MPN651	0.33468981
MPN240	0	MPN420	0.33468981	MPN652	0.33468981
MPN246	-	MPN428	0.21934261	MPN653	0.33468981
MPN250	-	MPN429	-	MPN667	0
MPN251	0.33163294	MPN430	-	MPN669	-
MPN252	-	MPN445	0.33468981	MPN672	0
MPN253	0.33468981	MPN455	0	MPN674	0.19892356
MPN257	0	MPN476	-	MPN678	-
MPN258	0.32959274	MPN479	0.33468981		

Table A.10.: Objective values (ov) of the FBA when simulating growth of *in silico* knock-outs of the listed genes. $ov > 0$ - growth; $ov = 0$ - no growth but catabolic activity; no ov (-) - FBA is infeasible, i.e. at least one minimum requirement defined in the model cannot be matched.

Table A.11: Functional Orthologs in *M. pneumoniae* and *M. genitalium*

Mpn gene	Mg gene	Mg alias	Mpn gene	Mg gene	Mg alias	Mpn gene	Mg gene	Mg alias
MPN001	MG001		MPN231	MG093		MPN461	MG323	
MPN002	MG002		MPN232	MG094		MPN462		
MPN003	MG003		MPN233	MG095		MPN463		
MPN004	MG004		MPN234			MPN464	MG192	
MPN005	MG005		MPN235	MG097		MPN465		
MPN006	MG006		MPN236	MG098		MPN466		
MPN007	MG007		MPN237	MG099		MPN467		
MPN008	MG008		MPN238	MG100		MPN468		
MPN009	MG009		MPN239	MG101		MPN469	MG515	MG323.1
MPN010			MPN240	MG102		MPN470	MG324	
MPN011			MPN241	MG103		MPN471	MG325	
MPN012			MPN242	MG476		MPN472	MG326	
MPN013			MPN243	MG104		MPN473	MG327	
MPN014	MG010		MPN244	MG105		MPN474	MG328	
MPN015	MG011		MPN245	MG106		MPN475	MG329	
MPN016	MG012		MPN246	MG107		MPN476	MG330	
MPN017	MG013		MPN247	MG108		MPN477	MG331	
MPN018	MG014		MPN248	MG109		MPN478	MG332	
MPN019	MG015		MPN249	MG110		MPN479	MG333	
MPN020	MG018		MPN250	MG111		MPN480	MG334	
MPN021	MG019		MPN251	MG112		MPN481	MG335	
MPN022	MG020		MPN252	MG113		MPN482	MG516	MG335.1
MPN023	MG021		MPN253	MG114		MPN483	MG517	MG335.2
MPN024	MG022		MPN254	MG115		MPN484		
MPN025	MG023		MPN255	MG116		MPN485		
MPN026	MG024		MPN256	MG117		MPN486		
MPN027			MPN257	MG118		MPN487	MG336	
MPN028	MG025		MPN258	MG119		MPN488	MG337	
MPN029	MG026		MPN259	MG120		MPN489	MG338	
MPN030	MG027		MPN260	MG121		MPN490	MG339	
MPN031	MG028		MPN261	MG122		MPN491		
MPN032	MG029		MPN262	MG123		MPN492		
MPN033	MG030		MPN263	MG124		MPN493		
MPN034	MG031		MPN264	MG125		MPN494		
MPN035			MPN265	MG126		MPN495		
MPN036	MG032		MPN266	MG127		MPN496		
MPN037			MPN267	MG128		MPN497		
MPN038			MPN268	MG129		MPN498		
MPN039			MPN269	MG130		MPN499		
MPN040			MPN270	MG131		MPN500	MG191	
MPN041			MPN271			MPN501		
MPN042			MPN272			MPN502		
MPN043	MG033		MPN273	MG132		MPN503		
MPN044	MG034		MPN274	MG133		MPN504		
MPN045	MG035		MPN275	MG134		MPN505		
MPN046	MG036		MPN276	MG135		MPN506		
MPN047	MG037		MPN277	MG136		MPN507		
MPN048			MPN278	MG137		MPN508	MG288	
MPN049			MPN279	MG138		MPN509		
MPN050	MG038		MPN280	MG139		MPN510		
MPN051	MG039		MPN281			MPN511		
MPN052	MG040		MPN282			MPN512		
MPN053	MG041		MPN283			MPN513		
MPN054			MPN284	MG260		MPN514		
MPN055	MG042		MPN285			MPN515	MG340	
MPN056	MG043		MPN286			MPN516	MG341	
MPN057	MG044		MPN287			MPN517	MG342	
MPN058	MG045		MPN288			MPN518	MG343	
MPN059	MG046		MPN289			MPN519	MG344	
MPN060	MG047		MPN290			MPN520	MG345	
MPN061	MG048		MPN291	MG208		MPN521	MG346	
MPN062	MG049		MPN292	MG209		MPN522	MG347	
MPN063	MG050		MPN293	MG210		MPN523	MG348	
MPN064	MG051		MPN294			MPN524		
MPN065	MG052		MPN295	MG480	MG210.1	MPN525	MG349	
MPN066	MG053		MPN296	MG481	MG210.2	MPN526	MG350	
MPN067	MG054		MPN297	MG211		MPN527	MG521	MG350.1
MPN068	MG055		MPN298	MG482	MG211.1	MPN528	MG351	
MPN069	MG473	MG055.1	MPN299	MG212		MPN528a	MG352	
MPN070	MG474	MG055.2	MPN300	MG213		MPN529	MG353	
MPN071	MG056		MPN301	MG214		MPN530	MG354	
MPN072	MG057		MPN302	MG215		MPN531	MG355	
MPN073	MG058		MPN303	MG216		MPN532	MG356	
MPN074	MG059		MPN304			MPN533	MG357	
MPN075	MG060		MPN305			MPN534		
MPN076	MG061		MPN306			MPN535	MG358	
MPN077	MG061		MPN307			MPN536	MG359	
MPN078	MG062		MPN308			MPN537	MG360	
MPN079	MG063		MPN309	MG217		MPN538	MG361	
MPN080	MG064		MPN310	MG218		MPN539	MG362	
MPN081	MG065		MPN311	MG491	MG218.1	MPN540	MG363	
MPN082	MG066		MPN312			MPN541	MG522	MG363.1

Continued on next page

Table A.11 – continued from previous page

Mpn gene	Mg gene	Mg alias	Mpn gene	Mg gene	Mg alias	Mpn gene	Mg gene	Mg alias
MPN083	MG067		MPN313	MG220		MPN542	MG364	
MPN084	MG068		MPN314	MG221		MPN543	MG365	
MPN085			MPN315	MG222		MPN544	MG366	
MPN086			MPN316	MG223		MPN545	MG367	
MPN087			MPN317	MG224		MPN546	MG368	
MPN088			MPN318	MG225		MPN547	MG369	
MPN089			MPN319	MG226		MPN548	MG370	
MPN090			MPN320	MG227		MPN549	MG371	
MPN091			MPN321	MG228		MPN550	MG372	
MPN092			MPN322	MG229		MPN551	MG373	
MPN093			MPN323	MG230		MPN552	MG374	
MPN094			MPN324	MG231		MPN553	MG375	
MPN095			MPN325	MG232		MPN554	MG376	
MPN096			MPN326	MG233		MPN555	MG377	
MPN097			MPN327	MG234		MPN556	MG378	
MPN098			MPN328	MG235		MPN557	MG379	
MPN099			MPN329	MG236		MPN558	MG380	
MPN100			MPN330	MG237		MPN559	MG381	
MPN101			MPN331	MG238		MPN560		
MPN102			MPN332	MG239		MPN561	MG382	
MPN103			MPN333			MPN562	MG383	
MPN104			MPN334			MPN563	MG384	
MPN105	MG194		MPN335			MPN564		
MPN106	MG195		MPN336	MG240		MPN565	MG524	MG384.1
MPN107			MPN337	MG241		MPN566	MG385	
MPN108			MPN338	MG242		MPN567	MG386	
MPN109			MPN339			MPN568	MG387	
MPN110			MPN340			MPN569	MG388	
MPN111			MPN341	MG244		MPN570	MG389	
MPN112			MPN342			MPN571	MG390	
MPN113			MPN343			MPN572	MG391	
MPN114			MPN344			MPN573	MG392	
MPN115	MG196		MPN345			MPN574	MG393	
MPN116	MG197		MPN346			MPN575		
MPN117	MG198		MPN347			MPN576	MG394	
MPN118	MG199		MPN348	MG245		MPN577		
MPN119	MG200		MPN349	MG246		MPN578		
MPN120	MG201		MPN350	MG247		MPN579		
MPN121	MG202		MPN351	MG248		MPN580		
MPN122	MG203		MPN352	MG249		MPN581		
MPN123	MG204		MPN353	MG250		MPN582		
MPN124	MG205		MPN354	MG251		MPN583		
MPN125	MG206		MPN355	MG252		MPN584		
MPN126	MG207		MPN356	MG253		MPN585		
MPN127			MPN357	MG254		MPN586		
MPN128			MPN358	MG255		MPN587		
MPN129			MPN359	MG256		MPN588	MG395	
MPN130			MPN360	MG257		MPN589		
MPN131			MPN361	MG258		MPN590		
MPN132			MPN362	MG259		MPN591		
MPN133	MG186		MPN363			MPN592	MG395	
MPN134	MG187		MPN364			MPN593		
MPN135	MG188		MPN365			MPN594		
MPN136	MG189		MPN366			MPN595	MG396	
MPN137			MPN367			MPN596	MG397	
MPN138			MPN368			MPN597	MG398	
MPN139			MPN369			MPN598	MG399	
MPN140	MG190		MPN370			MPN599	MG400	
MPN141	MG191		MPN371			MPN600	MG401	
MPN142	MG192		MPN372			MPN601	MG402	
MPN143			MPN373			MPN602	MG403	
MPN144			MPN374			MPN603	MG404	
MPN145			MPN375			MPN604	MG405	
MPN146			MPN376			MPN605	MG406	
MPN147			MPN377			MPN606	MG407	
MPN148			MPN378	MG261		MPN607	MG408	
MPN149			MPN379	MG262		MPN608	MG409	
MPN150			MPN380	MG498	MG262.1	MPN609	MG410	
MPN151			MPN381	MG263		MPN610	MG411	
MPN152			MPN382	MG264		MPN611	MG412	
MPN153	MG140		MPN383	MG265		MPN612	MG414	
MPN154	MG141		MPN384	MG266		MPN613	MG525	
MPN155	MG142		MPN385	MG267		MPN614	MG525	
MPN156	MG143		MPN386	MG268		MPN615		
MPN157	MG144		MPN387	MG269		MPN616	MG417	
MPN158	MG145		MPN388			MPN617	MG418	
MPN159	MG146		MPN389	MG270		MPN618	MG419	
MPN160	MG147		MPN390	MG271		MPN619	MG421	
MPN161	MG148		MPN391	MG272		MPN620	MG422	
MPN162	MG149		MPN392	MG273		MPN621	MG423	
MPN163	MG478	MG149.1	MPN393	MG274		MPN622	MG424	
MPN164	MG150		MPN394	MG275		MPN623	MG425	

Continued on next page

Table A.11 – continued from previous page

Mpn gene	Mg gene	Mg alias	Mpn gene	Mg gene	Mg alias	Mpn gene	Mg gene	Mg alias
MPN165	MG151		MPN395	MG276		MPN624	MG426	
MPN166	MG152		MPN396	MG277		MPN625	MG427	
MPN167	MG153		MPN397	MG278		MPN626	MG428	
MPN168	MG154		MPN398	MG279		MPN627	MG429	
MPN169	MG155		MPN399	MG280		MPN628	MG430	
MPN170	MG156		MPN400	MG281		MPN629	MG431	
MPN171	MG157		MPN401	MG282		MPN630	MG432	
MPN172	MG158		MPN402	MG283		MPN631	MG433	
MPN173	MG159		MPN403	MG284		MPN632	MG434	
MPN174	MG160		MPN404	MG285		MPN633		
MPN175	MG161		MPN405	MG286		MPN634		
MPN176	MG162		MPN406	MG287		MPN635		
MPN177	MG163		MPN407			MPN636	MG435	
MPN178	MG164		MPN408			MPN637	MG437	
MPN179	MG165		MPN409			MPN638	MG438	
MPN180	MG166		MPN410			MPN639		
MPN181	MG167		MPN411			MPN640		
MPN182	MG168		MPN412			MPN641		
MPN183	MG169		MPN413			MPN642		
MPN184	MG170		MPN414			MPN643	MG440	
MPN185	MG171		MPN415	MG289		MPN644	MG439	
MPN186	MG172		MPN416	MG290		MPN645		
MPN187	MG173		MPN417	MG291		MPN646		
MPN188			MPN418	MG292		MPN647		
MPN189	MG175		MPN419	MG505	MG291.1	MPN648	MG441	
MPN190	MG176		MPN420	MG293		MPN649		
MPN191	MG177		MPN421	MG294		MPN650		
MPN192	MG178		MPN422	MG295		MPN651		
MPN193	MG179		MPN423	MG296		MPN652		
MPN194	MG180		MPN424			MPN653		
MPN195	MG181		MPN425	MG297		MPN654		
MPN196	MG182		MPN426	MG298		MPN655		
MPN197	MG183		MPN427			MPN656	MG442	
MPN198	MG184		MPN428	MG299		MPN657	MG443	
MPN199	MG185		MPN429	MG300		MPN658	MG444	
MPN200			MPN430	MG301		MPN659	MG445	
MPN201			MPN431	MG302		MPN660	MG446	
MPN202			MPN432	MG303		MPN661	MG447	
MPN203			MPN433	MG304		MPN662	MG448	
MPN204			MPN434	MG305		MPN663		
MPN205			MPN435	MG306		MPN664	MG450	
MPN206			MPN436	MG307		MPN665	MG451	
MPN207	MG069		MPN437			MPN666	MG452	
MPN208	MG070		MPN438			MPN667	MG453	
MPN209	MG071		MPN439			MPN668	MG454	
MPN210	MG072		MPN440			MPN669	MG455	
MPN211	MG073		MPN441			MPN670	MG456	
MPN212	MG074		MPN442			MPN671	MG457	
MPN213	MG075		MPN443	MG308		MPN672	MG458	
MPN214	MG076		MPN444	MG309		MPN673	MG459	
MPN215	MG077		MPN445	MG310		MPN674	MG460	
MPN216	MG078		MPN446	MG311		MPN675		
MPN217	MG079		MPN447	MG312		MPN676		
MPN218	MG080		MPN448	MG313		MPN677	MG461	
MPN219	MG081		MPN449	MG314		MPN678	MG462	
MPN220	MG082		MPN450	MG315		MPN679	MG463	
MPN221	MG083		MPN451	MG316		MPN680	MG464	
MPN222	MG084		MPN452	MG317		MPN681	MG465	
MPN223	MG085		MPN453	MG318		MPN682	MG466	
MPN224	MG086		MPN454	MG319		MPN683	MG467	
MPN225	MG087		MPN455	MG320		MPN684	MG468	
MPN226	MG088		MPN456	MG321		MPN685	MG526	MG468.1
MPN227	MG089		MPN457			MPN686	MG469	
MPN228	MG090		MPN458			MPN687		
MPN229	MG091		MPN459			MPN688	MG470	
MPN230	MG092		MPN460	MG322				

Table A.11.: Functional orthologs of *M. pneumoniae* and *M. genitalium*.

Table A.12: Transposon Primer

primer name	sequence	ORF	PCR fragment (bp)	inser-tion-site	% of trun-cation
5MPN133	CCCAAGATTCTTCTTCTGCATCCAC	MPN321	140	381698	65
5MPN321	CGCCAGACCACGTTGCACCAAGATG	MPN133	192	172691	54
5MPN392	TGGTGTTCAAGTGCTTTAATCCCAC	MPN595	188	717112	27
5MPN533	TGTGTCAAAGACTGCCACATTCTTA	MPN533	341	655990	86
5MPN595	TCTCATAAGCAATGATCTTACCGAC	MPN392	144	470043	68
3JpMT85	GGTGGATCCGTACTAGTGTGTGTCCAAAG	reverse primer			

Table A.12.: Primer used for the identification of transposon insertion sites in *M. pneumoniae* genes predicted to be not essential by the model but not found to be disrupted in *M. genitalium* [Glass et al., 2006]. For each primer the sequence, the ORF they have been used for, the PCR fragment, the transposon insertion site, and the truncated fraction of the protein are listed.

Table A.13: *In Silico* Double Knock-outs

gene 1	pathway 1	gene 2	pathway 2	inter-action type	gene 1	pathway 1	gene 2	pathway 2	inter-action type
MPN033	NT m.	MPN043	sugar m.	sick	MPN108	aa m.	MPN394	pyruvate m.	sick
MPN033	NT m.	MPN044	NT m.	sick	MPN108	aa m.	MPN428	pyruvate m.	sick
MPN033	NT m.	MPN050	sugar m.	sick	MPN108	aa m.	MPN533	pyruvate m.	sick
MPN033	NT m.	MPN051	lipid m.	sick	MPN108	aa m.	MPN595	PPP	sick
MPN033	NT m.	MPN062	NT m.	sick	MPN108	aa m.	MPN674	pyruvate m.	sick
MPN033	NT m.	MPN064	NT m.	sick	MPN133	sugar m.	MPN134	sugar m.	sick
MPN033	NT m.	MPN065	NT m.	sick	MPN133	sugar m.	MPN135	sugar m.	sick
MPN033	NT m.	MPN073	PPP	SL	MPN133	sugar m.	MPN136	sugar m.	sick
MPN033	NT m.	MPN082	PPP	sick	MPN133	sugar m.	MPN251	PPP	sick
MPN033	NT m.	MPN133	sugar m.	sick	MPN133	sugar m.	MPN258	sugar m.	sick
MPN033	NT m.	MPN134	sugar m.	sick	MPN133	sugar m.	MPN259	sugar m.	sick
MPN033	NT m.	MPN135	sugar m.	sick	MPN133	sugar m.	MPN260	sugar m.	sick
MPN033	NT m.	MPN136	sugar m.	sick	MPN133	sugar m.	MPN306	aa m.	sick
MPN033	NT m.	MPN251	PPP	sick	MPN133	sugar m.	MPN307	aa m.	sick
MPN033	NT m.	MPN258	sugar m.	sick	MPN133	sugar m.	MPN320	folate m.	sick
MPN033	NT m.	MPN259	sugar m.	sick	MPN133	sugar m.	MPN386	NT m.	sick
MPN033	NT m.	MPN260	sugar m.	sick	MPN133	sugar m.	MPN390	pyruvate m.	sick
MPN033	NT m.	MPN306	aa m.	sick	MPN133	sugar m.	MPN391	pyruvate m.	sick
MPN033	NT m.	MPN307	aa m.	sick	MPN133	sugar m.	MPN392	pyruvate m.	sick
MPN033	NT m.	MPN320	folate m.	sick	MPN133	sugar m.	MPN393	pyruvate m.	sick
MPN033	NT m.	MPN386	NT m.	sick	MPN133	sugar m.	MPN394	pyruvate m.	sick
MPN033	NT m.	MPN390	pyruvate m.	sick	MPN133	sugar m.	MPN428	pyruvate m.	sick
MPN033	NT m.	MPN391	pyruvate m.	sick	MPN133	sugar m.	MPN533	pyruvate m.	sick
MPN033	NT m.	MPN392	pyruvate m.	sick	MPN133	sugar m.	MPN595	PPP	sick
MPN033	NT m.	MPN393	pyruvate m.	sick	MPN133	sugar m.	MPN674	pyruvate m.	sick
MPN033	NT m.	MPN394	pyruvate m.	sick	MPN134	sugar m.	MPN135	sugar m.	sick
MPN033	NT m.	MPN428	pyruvate m.	sick	MPN134	sugar m.	MPN136	sugar m.	sick
MPN033	NT m.	MPN533	pyruvate m.	sick	MPN134	sugar m.	MPN251	PPP	sick
MPN033	NT m.	MPN595	PPP	sick	MPN134	sugar m.	MPN258	sugar m.	sick
MPN033	NT m.	MPN674	pyruvate m.	sick	MPN134	sugar m.	MPN259	sugar m.	sick
MPN043	sugar m.	MPN044	NT m.	sick	MPN134	sugar m.	MPN260	sugar m.	sick
MPN043	sugar m.	MPN051	lipid m.	sick	MPN134	sugar m.	MPN306	aa m.	sick
MPN043	sugar m.	MPN062	NT m.	sick	MPN134	sugar m.	MPN307	aa m.	sick
MPN043	sugar m.	MPN064	NT m.	sick	MPN134	sugar m.	MPN320	folate m.	sick
MPN043	sugar m.	MPN065	NT m.	sick	MPN134	sugar m.	MPN386	NT m.	sick
MPN043	sugar m.	MPN082	PPP	sick	MPN134	sugar m.	MPN390	pyruvate m.	sick
MPN043	sugar m.	MPN133	sugar m.	SL	MPN134	sugar m.	MPN391	pyruvate m.	sick
MPN043	sugar m.	MPN134	sugar m.	SL	MPN134	sugar m.	MPN392	pyruvate m.	sick
MPN043	sugar m.	MPN135	sugar m.	SL	MPN134	sugar m.	MPN393	pyruvate m.	sick
MPN043	sugar m.	MPN136	sugar m.	SL	MPN134	sugar m.	MPN394	pyruvate m.	sick
MPN043	sugar m.	MPN251	PPP	sick	MPN134	sugar m.	MPN428	pyruvate m.	sick
MPN043	sugar m.	MPN258	sugar m.	sick	MPN134	sugar m.	MPN533	pyruvate m.	sick
MPN043	sugar m.	MPN259	sugar m.	sick	MPN134	sugar m.	MPN595	PPP	sick
MPN043	sugar m.	MPN260	sugar m.	sick	MPN134	sugar m.	MPN674	pyruvate m.	sick
MPN043	sugar m.	MPN306	aa m.	sick	MPN135	sugar m.	MPN136	sugar m.	sick
MPN043	sugar m.	MPN307	aa m.	sick	MPN135	sugar m.	MPN251	PPP	sick
MPN043	sugar m.	MPN320	folate m.	sick	MPN135	sugar m.	MPN258	sugar m.	sick
MPN043	sugar m.	MPN386	NT m.	sick	MPN135	sugar m.	MPN259	sugar m.	sick
MPN043	sugar m.	MPN390	pyruvate m.	sick	MPN135	sugar m.	MPN260	sugar m.	sick
MPN043	sugar m.	MPN391	pyruvate m.	sick	MPN135	sugar m.	MPN306	aa m.	sick
MPN043	sugar m.	MPN392	pyruvate m.	sick	MPN135	sugar m.	MPN307	aa m.	sick

Continued on next page

Table A.13 – continued from previous page

gene 1	pathway 1	gene 2	pathway 2	interaction type	gene 1	pathway 1	gene 2	pathway 2	interaction type
MPN043	sugar m.	MPN393	pyruvate m.	sick	MPN135	sugar m.	MPN320	folate m.	sick
MPN043	sugar m.	MPN394	pyruvate m.	sick	MPN135	sugar m.	MPN386	NT m.	sick
MPN043	sugar m.	MPN428	pyruvate m.	sick	MPN135	sugar m.	MPN390	pyruvate m.	sick
MPN043	sugar m.	MPN533	pyruvate m.	sick	MPN135	sugar m.	MPN391	pyruvate m.	sick
MPN043	sugar m.	MPN595	PPP	sick	MPN135	sugar m.	MPN392	pyruvate m.	sick
MPN043	sugar m.	MPN674	pyruvate m.	sick	MPN135	sugar m.	MPN393	pyruvate m.	sick
MPN044	NT m.	MPN050	sugar m.	sick	MPN135	sugar m.	MPN394	pyruvate m.	sick
MPN044	NT m.	MPN051	lipid m.	sick	MPN135	sugar m.	MPN428	pyruvate m.	sick
MPN044	NT m.	MPN062	NT m.	sick	MPN135	sugar m.	MPN533	pyruvate m.	sick
MPN044	NT m.	MPN064	NT m.	sick	MPN135	sugar m.	MPN595	PPP	sick
MPN044	NT m.	MPN065	NT m.	sick	MPN135	sugar m.	MPN674	pyruvate m.	sick
MPN044	NT m.	MPN082	PPP	sick	MPN136	sugar m.	MPN251	PPP	sick
MPN044	NT m.	MPN133	sugar m.	sick	MPN136	sugar m.	MPN258	sugar m.	sick
MPN044	NT m.	MPN134	sugar m.	sick	MPN136	sugar m.	MPN259	sugar m.	sick
MPN044	NT m.	MPN135	sugar m.	sick	MPN136	sugar m.	MPN260	sugar m.	sick
MPN044	NT m.	MPN136	sugar m.	sick	MPN136	sugar m.	MPN306	aa m.	sick
MPN044	NT m.	MPN251	PPP	sick	MPN136	sugar m.	MPN307	aa m.	sick
MPN044	NT m.	MPN258	sugar m.	sick	MPN136	sugar m.	MPN320	folate m.	sick
MPN044	NT m.	MPN259	sugar m.	sick	MPN136	sugar m.	MPN386	NT m.	sick
MPN044	NT m.	MPN260	sugar m.	sick	MPN136	sugar m.	MPN390	pyruvate m.	sick
MPN044	NT m.	MPN306	aa m.	sick	MPN136	sugar m.	MPN391	pyruvate m.	sick
MPN044	NT m.	MPN307	aa m.	sick	MPN136	sugar m.	MPN392	pyruvate m.	sick
MPN044	NT m.	MPN320	folate m.	SL	MPN136	sugar m.	MPN393	pyruvate m.	sick
MPN044	NT m.	MPN386	NT m.	sick	MPN136	sugar m.	MPN394	pyruvate m.	sick
MPN044	NT m.	MPN390	pyruvate m.	sick	MPN136	sugar m.	MPN428	pyruvate m.	sick
MPN044	NT m.	MPN391	pyruvate m.	sick	MPN136	sugar m.	MPN533	pyruvate m.	sick
MPN044	NT m.	MPN392	pyruvate m.	sick	MPN136	sugar m.	MPN595	PPP	sick
MPN044	NT m.	MPN393	pyruvate m.	sick	MPN136	sugar m.	MPN674	pyruvate m.	sick
MPN044	NT m.	MPN394	pyruvate m.	sick	MPN251	PPP	MPN258	sugar m.	SL
MPN044	NT m.	MPN428	pyruvate m.	sick	MPN251	PPP	MPN259	sugar m.	SL
MPN044	NT m.	MPN533	pyruvate m.	sick	MPN251	PPP	MPN260	sugar m.	SL
MPN044	NT m.	MPN595	PPP	sick	MPN251	PPP	MPN306	aa m.	sick
MPN044	NT m.	MPN674	pyruvate m.	sick	MPN251	PPP	MPN307	aa m.	sick
MPN050	sugar m.	MPN051	lipid m.	sick	MPN251	PPP	MPN320	folate m.	sick
MPN050	sugar m.	MPN062	NT m.	sick	MPN251	PPP	MPN386	NT m.	sick
MPN050	sugar m.	MPN064	NT m.	sick	MPN251	PPP	MPN390	pyruvate m.	sick
MPN050	sugar m.	MPN065	NT m.	sick	MPN251	PPP	MPN391	pyruvate m.	sick
MPN050	sugar m.	MPN082	PPP	sick	MPN251	PPP	MPN392	pyruvate m.	sick
MPN050	sugar m.	MPN133	sugar m.	SL	MPN251	PPP	MPN393	pyruvate m.	sick
MPN050	sugar m.	MPN134	sugar m.	SL	MPN251	PPP	MPN394	pyruvate m.	sick
MPN050	sugar m.	MPN135	sugar m.	SL	MPN251	PPP	MPN428	pyruvate m.	sick
MPN050	sugar m.	MPN136	sugar m.	SL	MPN251	PPP	MPN533	pyruvate m.	sick
MPN050	sugar m.	MPN251	PPP	sick	MPN251	PPP	MPN674	pyruvate m.	sick
MPN050	sugar m.	MPN258	sugar m.	sick	MPN253	lipid m.	MPN258	sugar m.	sick
MPN050	sugar m.	MPN259	sugar m.	sick	MPN253	lipid m.	MPN259	sugar m.	sick
MPN050	sugar m.	MPN260	sugar m.	sick	MPN253	lipid m.	MPN260	sugar m.	sick
MPN050	sugar m.	MPN306	aa m.	sick	MPN253	lipid m.	MPN306	aa m.	sick
MPN050	sugar m.	MPN307	aa m.	sick	MPN253	lipid m.	MPN307	aa m.	sick
MPN050	sugar m.	MPN320	folate m.	sick	MPN253	lipid m.	MPN320	folate m.	sick
MPN050	sugar m.	MPN386	NT m.	sick	MPN253	lipid m.	MPN386	NT m.	sick
MPN050	sugar m.	MPN390	pyruvate m.	sick	MPN253	lipid m.	MPN390	pyruvate m.	sick
MPN050	sugar m.	MPN391	pyruvate m.	sick	MPN253	lipid m.	MPN391	pyruvate m.	sick
MPN050	sugar m.	MPN392	pyruvate m.	sick	MPN253	lipid m.	MPN392	pyruvate m.	sick
MPN050	sugar m.	MPN393	pyruvate m.	sick	MPN253	lipid m.	MPN393	pyruvate m.	sick
MPN050	sugar m.	MPN394	pyruvate m.	sick	MPN253	lipid m.	MPN394	pyruvate m.	sick
MPN050	sugar m.	MPN428	pyruvate m.	sick	MPN253	lipid m.	MPN428	pyruvate m.	sick
MPN050	sugar m.	MPN533	pyruvate m.	sick	MPN253	lipid m.	MPN533	pyruvate m.	sick
MPN050	sugar m.	MPN595	PPP	sick	MPN253	lipid m.	MPN595	PPP	sick
MPN050	sugar m.	MPN674	pyruvate m.	sick	MPN253	lipid m.	MPN674	pyruvate m.	sick
MPN051	lipid m.	MPN062	NT m.	sick	MPN258	sugar m.	MPN306	aa m.	sick
MPN051	lipid m.	MPN064	NT m.	sick	MPN258	sugar m.	MPN307	aa m.	sick
MPN051	lipid m.	MPN065	NT m.	sick	MPN258	sugar m.	MPN320	folate m.	sick
MPN051	lipid m.	MPN082	PPP	sick	MPN258	sugar m.	MPN386	NT m.	sick
MPN051	lipid m.	MPN251	PPP	sick	MPN258	sugar m.	MPN390	pyruvate m.	sick
MPN051	lipid m.	MPN258	sugar m.	sick	MPN258	sugar m.	MPN391	pyruvate m.	sick
MPN051	lipid m.	MPN259	sugar m.	sick	MPN258	sugar m.	MPN392	pyruvate m.	sick
MPN051	lipid m.	MPN260	sugar m.	sick	MPN258	sugar m.	MPN393	pyruvate m.	sick
MPN051	lipid m.	MPN306	aa m.	sick	MPN258	sugar m.	MPN394	pyruvate m.	sick
MPN051	lipid m.	MPN307	aa m.	sick	MPN258	sugar m.	MPN428	pyruvate m.	sick
MPN051	lipid m.	MPN320	folate m.	sick	MPN258	sugar m.	MPN533	pyruvate m.	sick
MPN051	lipid m.	MPN386	NT m.	sick	MPN258	sugar m.	MPN595	PPP	SL
MPN051	lipid m.	MPN390	pyruvate m.	sick	MPN258	sugar m.	MPN674	pyruvate m.	sick
MPN051	lipid m.	MPN391	pyruvate m.	sick	MPN259	sugar m.	MPN306	aa m.	sick
MPN051	lipid m.	MPN392	pyruvate m.	sick	MPN259	sugar m.	MPN307	aa m.	sick
MPN051	lipid m.	MPN393	pyruvate m.	sick	MPN259	sugar m.	MPN320	folate m.	sick
MPN051	lipid m.	MPN394	pyruvate m.	sick	MPN259	sugar m.	MPN386	NT m.	sick
MPN051	lipid m.	MPN428	pyruvate m.	sick	MPN259	sugar m.	MPN390	pyruvate m.	sick
MPN051	lipid m.	MPN533	pyruvate m.	sick	MPN259	sugar m.	MPN391	pyruvate m.	sick
MPN051	lipid m.	MPN595	PPP	sick	MPN259	sugar m.	MPN392	pyruvate m.	sick
MPN051	lipid m.	MPN674	pyruvate m.	sick	MPN259	sugar m.	MPN393	pyruvate m.	sick

Continued on next page

Table A.13 – continued from previous page

gene 1	pathway 1	gene 2	pathway 2	interaction type	gene 1	pathway 1	gene 2	pathway 2	interaction type
MPN062	NT m.	MPN064	NT m.	sick	MPN259	sugar m.	MPN394	pyruvate m.	sick
MPN062	NT m.	MPN065	NT m.	sick	MPN259	sugar m.	MPN428	pyruvate m.	sick
MPN062	NT m.	MPN082	PPP	sick	MPN259	sugar m.	MPN533	pyruvate m.	sick
MPN062	NT m.	MPN133	sugar m.	sick	MPN259	sugar m.	MPN595	PPP	SL
MPN062	NT m.	MPN134	sugar m.	sick	MPN259	sugar m.	MPN674	pyruvate m.	sick
MPN062	NT m.	MPN135	sugar m.	sick	MPN260	sugar m.	MPN306	aa m.	sick
MPN062	NT m.	MPN136	sugar m.	sick	MPN260	sugar m.	MPN307	aa m.	sick
MPN062	NT m.	MPN251	PPP	sick	MPN260	sugar m.	MPN320	folate m.	sick
MPN062	NT m.	MPN258	sugar m.	sick	MPN260	sugar m.	MPN386	NT m.	sick
MPN062	NT m.	MPN259	sugar m.	sick	MPN260	sugar m.	MPN390	pyruvate m.	sick
MPN062	NT m.	MPN260	sugar m.	sick	MPN260	sugar m.	MPN391	pyruvate m.	sick
MPN062	NT m.	MPN306	aa m.	sick	MPN260	sugar m.	MPN392	pyruvate m.	sick
MPN062	NT m.	MPN307	aa m.	sick	MPN260	sugar m.	MPN393	pyruvate m.	sick
MPN062	NT m.	MPN320	folate m.	sick	MPN260	sugar m.	MPN394	pyruvate m.	sick
MPN062	NT m.	MPN386	NT m.	sick	MPN260	sugar m.	MPN428	pyruvate m.	sick
MPN062	NT m.	MPN390	pyruvate m.	sick	MPN260	sugar m.	MPN533	pyruvate m.	sick
MPN062	NT m.	MPN391	pyruvate m.	sick	MPN260	sugar m.	MPN595	PPP	SL
MPN062	NT m.	MPN392	pyruvate m.	sick	MPN260	sugar m.	MPN674	pyruvate m.	sick
MPN062	NT m.	MPN393	pyruvate m.	sick	MPN300	folate m.	MPN306	aa m.	sick
MPN062	NT m.	MPN394	pyruvate m.	sick	MPN300	folate m.	MPN307	aa m.	sick
MPN062	NT m.	MPN428	pyruvate m.	sick	MPN300	folate m.	MPN320	folate m.	sick
MPN062	NT m.	MPN533	pyruvate m.	sick	MPN300	folate m.	MPN321	folate m.	SL
MPN062	NT m.	MPN595	PPP	sick	MPN300	folate m.	MPN386	NT m.	sick
MPN062	NT m.	MPN674	pyruvate m.	sick	MPN300	folate m.	MPN390	pyruvate m.	sick
MPN064	NT m.	MPN065	NT m.	sick	MPN300	folate m.	MPN391	pyruvate m.	sick
MPN064	NT m.	MPN073	PPP	SL	MPN300	folate m.	MPN392	pyruvate m.	sick
MPN064	NT m.	MPN082	PPP	sick	MPN300	folate m.	MPN393	pyruvate m.	sick
MPN064	NT m.	MPN133	sugar m.	sick	MPN300	folate m.	MPN394	pyruvate m.	sick
MPN064	NT m.	MPN134	sugar m.	sick	MPN300	folate m.	MPN428	pyruvate m.	sick
MPN064	NT m.	MPN135	sugar m.	sick	MPN300	folate m.	MPN533	pyruvate m.	sick
MPN064	NT m.	MPN136	sugar m.	sick	MPN300	folate m.	MPN595	PPP	sick
MPN064	NT m.	MPN251	PPP	sick	MPN300	folate m.	MPN674	pyruvate m.	sick
MPN064	NT m.	MPN258	sugar m.	sick	MPN304	aa m.	MPN306	aa m.	sick
MPN064	NT m.	MPN259	sugar m.	sick	MPN304	aa m.	MPN307	aa m.	sick
MPN064	NT m.	MPN260	sugar m.	sick	MPN304	aa m.	MPN320	folate m.	sick
MPN064	NT m.	MPN306	aa m.	sick	MPN304	aa m.	MPN386	NT m.	sick
MPN064	NT m.	MPN307	aa m.	sick	MPN304	aa m.	MPN390	pyruvate m.	sick
MPN064	NT m.	MPN320	folate m.	SL	MPN304	aa m.	MPN391	pyruvate m.	sick
MPN064	NT m.	MPN386	NT m.	sick	MPN304	aa m.	MPN392	pyruvate m.	sick
MPN064	NT m.	MPN390	pyruvate m.	sick	MPN304	aa m.	MPN393	pyruvate m.	sick
MPN064	NT m.	MPN391	pyruvate m.	sick	MPN304	aa m.	MPN394	pyruvate m.	sick
MPN064	NT m.	MPN392	pyruvate m.	sick	MPN304	aa m.	MPN428	pyruvate m.	sick
MPN064	NT m.	MPN393	pyruvate m.	sick	MPN304	aa m.	MPN533	pyruvate m.	sick
MPN064	NT m.	MPN394	pyruvate m.	sick	MPN304	aa m.	MPN595	PPP	sick
MPN064	NT m.	MPN428	pyruvate m.	sick	MPN304	aa m.	MPN674	pyruvate m.	sick
MPN064	NT m.	MPN533	pyruvate m.	sick	MPN305	aa m.	MPN306	aa m.	sick
MPN064	NT m.	MPN595	PPP	sick	MPN305	aa m.	MPN307	aa m.	sick
MPN064	NT m.	MPN674	pyruvate m.	sick	MPN305	aa m.	MPN320	folate m.	sick
MPN065	NT m.	MPN082	PPP	sick	MPN305	aa m.	MPN386	NT m.	sick
MPN065	NT m.	MPN133	sugar m.	sick	MPN305	aa m.	MPN390	pyruvate m.	sick
MPN065	NT m.	MPN134	sugar m.	sick	MPN305	aa m.	MPN391	pyruvate m.	sick
MPN065	NT m.	MPN135	sugar m.	sick	MPN305	aa m.	MPN392	pyruvate m.	sick
MPN065	NT m.	MPN136	sugar m.	sick	MPN305	aa m.	MPN393	pyruvate m.	sick
MPN065	NT m.	MPN251	PPP	sick	MPN305	aa m.	MPN394	pyruvate m.	sick
MPN065	NT m.	MPN258	sugar m.	sick	MPN305	aa m.	MPN428	pyruvate m.	sick
MPN065	NT m.	MPN259	sugar m.	sick	MPN305	aa m.	MPN533	pyruvate m.	sick
MPN065	NT m.	MPN260	sugar m.	sick	MPN305	aa m.	MPN595	PPP	sick
MPN065	NT m.	MPN306	aa m.	sick	MPN305	aa m.	MPN674	pyruvate m.	sick
MPN065	NT m.	MPN307	aa m.	sick	MPN306	aa m.	MPN320	folate m.	sick
MPN065	NT m.	MPN386	NT m.	sick	MPN306	aa m.	MPN386	NT m.	sick
MPN065	NT m.	MPN390	pyruvate m.	sick	MPN306	aa m.	MPN390	pyruvate m.	sick
MPN065	NT m.	MPN391	pyruvate m.	sick	MPN306	aa m.	MPN391	pyruvate m.	sick
MPN065	NT m.	MPN392	pyruvate m.	sick	MPN306	aa m.	MPN392	pyruvate m.	sick
MPN065	NT m.	MPN393	pyruvate m.	sick	MPN306	aa m.	MPN393	pyruvate m.	sick
MPN065	NT m.	MPN394	pyruvate m.	sick	MPN306	aa m.	MPN394	pyruvate m.	sick
MPN065	NT m.	MPN428	pyruvate m.	sick	MPN306	aa m.	MPN428	pyruvate m.	sick
MPN065	NT m.	MPN533	pyruvate m.	sick	MPN306	aa m.	MPN533	pyruvate m.	sick
MPN065	NT m.	MPN595	PPP	sick	MPN306	aa m.	MPN595	PPP	sick
MPN065	NT m.	MPN674	pyruvate m.	sick	MPN306	aa m.	MPN674	pyruvate m.	sick
MPN073	PPP	MPN082	PPP	sick	MPN307	aa m.	MPN320	folate m.	sick
MPN073	PPP	MPN133	sugar m.	sick	MPN307	aa m.	MPN386	NT m.	sick
MPN073	PPP	MPN134	sugar m.	sick	MPN307	aa m.	MPN390	pyruvate m.	sick
MPN073	PPP	MPN135	sugar m.	sick	MPN307	aa m.	MPN391	pyruvate m.	sick
MPN073	PPP	MPN136	sugar m.	sick	MPN307	aa m.	MPN392	pyruvate m.	sick
MPN073	PPP	MPN251	PPP	sick	MPN307	aa m.	MPN393	pyruvate m.	sick
MPN073	PPP	MPN258	sugar m.	sick	MPN307	aa m.	MPN394	pyruvate m.	sick
MPN073	PPP	MPN259	sugar m.	sick	MPN307	aa m.	MPN428	pyruvate m.	sick
MPN073	PPP	MPN260	sugar m.	sick	MPN307	aa m.	MPN533	pyruvate m.	sick
MPN073	PPP	MPN306	aa m.	sick	MPN307	aa m.	MPN595	PPP	sick
MPN073	PPP	MPN307	aa m.	sick	MPN307	aa m.	MPN674	pyruvate m.	sick

Continued on next page

Table A.13 – continued from previous page

gene 1	pathway 1	gene 2	pathway 2	interaction type	gene 1	pathway 1	gene 2	pathway 2	interaction type
MPN073	PPP	MPN320	folate m.	sick	MPN320	folate m.	MPN386	NT m.	sick
MPN073	PPP	MPN386	NT m.	sick	MPN320	folate m.	MPN390	pyruvate m.	sick
MPN073	PPP	MPN390	pyruvate m.	sick	MPN320	folate m.	MPN391	pyruvate m.	sick
MPN073	PPP	MPN391	pyruvate m.	sick	MPN320	folate m.	MPN392	pyruvate m.	sick
MPN073	PPP	MPN392	pyruvate m.	sick	MPN320	folate m.	MPN393	pyruvate m.	sick
MPN073	PPP	MPN393	pyruvate m.	sick	MPN320	folate m.	MPN394	pyruvate m.	sick
MPN073	PPP	MPN394	pyruvate m.	sick	MPN320	folate m.	MPN428	pyruvate m.	sick
MPN073	PPP	MPN428	pyruvate m.	sick	MPN320	folate m.	MPN533	pyruvate m.	sick
MPN073	PPP	MPN533	pyruvate m.	sick	MPN320	folate m.	MPN595	PPP	sick
MPN073	PPP	MPN595	PPP	sick	MPN320	folate m.	MPN674	pyruvate m.	sick
MPN073	PPP	MPN674	pyruvate m.	sick	MPN321	folate m.	MPN386	NT m.	sick
MPN078	sugar m.	MPN082	PPP	sick	MPN321	folate m.	MPN390	pyruvate m.	sick
MPN078	sugar m.	MPN133	sugar m.	sick	MPN321	folate m.	MPN391	pyruvate m.	sick
MPN078	sugar m.	MPN134	sugar m.	sick	MPN321	folate m.	MPN392	pyruvate m.	sick
MPN078	sugar m.	MPN135	sugar m.	sick	MPN321	folate m.	MPN393	pyruvate m.	sick
MPN078	sugar m.	MPN136	sugar m.	sick	MPN321	folate m.	MPN394	pyruvate m.	sick
MPN078	sugar m.	MPN251	PPP	sick	MPN321	folate m.	MPN428	pyruvate m.	sick
MPN078	sugar m.	MPN258	sugar m.	sick	MPN321	folate m.	MPN533	pyruvate m.	sick
MPN078	sugar m.	MPN259	sugar m.	sick	MPN321	folate m.	MPN595	PPP	sick
MPN078	sugar m.	MPN260	sugar m.	sick	MPN321	folate m.	MPN674	pyruvate m.	sick
MPN078	sugar m.	MPN306	aa m.	sick	MPN386	NT m.	MPN390	pyruvate m.	sick
MPN078	sugar m.	MPN307	aa m.	sick	MPN386	NT m.	MPN391	pyruvate m.	sick
MPN078	sugar m.	MPN320	folate m.	sick	MPN386	NT m.	MPN392	pyruvate m.	sick
MPN078	sugar m.	MPN386	NT m.	sick	MPN386	NT m.	MPN393	pyruvate m.	sick
MPN078	sugar m.	MPN390	pyruvate m.	sick	MPN386	NT m.	MPN394	pyruvate m.	sick
MPN078	sugar m.	MPN391	pyruvate m.	sick	MPN386	NT m.	MPN428	pyruvate m.	sick
MPN078	sugar m.	MPN392	pyruvate m.	sick	MPN386	NT m.	MPN533	pyruvate m.	sick
MPN078	sugar m.	MPN393	pyruvate m.	sick	MPN386	NT m.	MPN595	PPP	sick
MPN078	sugar m.	MPN394	pyruvate m.	sick	MPN386	NT m.	MPN674	pyruvate m.	sick
MPN078	sugar m.	MPN428	pyruvate m.	sick	MPN390	pyruvate m.	MPN394	pyruvate m.	sick
MPN078	sugar m.	MPN533	pyruvate m.	sick	MPN390	pyruvate m.	MPN595	PPP	sick
MPN078	sugar m.	MPN595	PPP	sick	MPN390	pyruvate m.	MPN674	pyruvate m.	SL
MPN078	sugar m.	MPN674	pyruvate m.	sick	MPN391	pyruvate m.	MPN394	pyruvate m.	sick
MPN079	sugar m.	MPN082	PPP	sick	MPN391	pyruvate m.	MPN420	lipid m.	sick
MPN079	sugar m.	MPN133	sugar m.	sick	MPN391	pyruvate m.	MPN595	PPP	sick
MPN079	sugar m.	MPN134	sugar m.	sick	MPN391	pyruvate m.	MPN674	pyruvate m.	SL
MPN079	sugar m.	MPN135	sugar m.	sick	MPN392	pyruvate m.	MPN394	pyruvate m.	sick
MPN079	sugar m.	MPN136	sugar m.	sick	MPN392	pyruvate m.	MPN595	PPP	sick
MPN079	sugar m.	MPN251	PPP	sick	MPN392	pyruvate m.	MPN674	pyruvate m.	SL
MPN079	sugar m.	MPN258	sugar m.	sick	MPN393	pyruvate m.	MPN394	pyruvate m.	sick
MPN079	sugar m.	MPN259	sugar m.	sick	MPN393	pyruvate m.	MPN595	PPP	sick
MPN079	sugar m.	MPN260	sugar m.	sick	MPN393	pyruvate m.	MPN674	pyruvate m.	SL
MPN079	sugar m.	MPN306	aa m.	sick	MPN394	pyruvate m.	MPN428	pyruvate m.	sick
MPN079	sugar m.	MPN307	aa m.	sick	MPN394	pyruvate m.	MPN533	pyruvate m.	sick
MPN079	sugar m.	MPN320	folate m.	sick	MPN394	pyruvate m.	MPN595	PPP	sick
MPN079	sugar m.	MPN386	NT m.	sick	MPN394	pyruvate m.	MPN674	pyruvate m.	SL
MPN079	sugar m.	MPN390	pyruvate m.	sick	MPN420	lipid m.	MPN428	pyruvate m.	sick
MPN079	sugar m.	MPN391	pyruvate m.	sick	MPN420	lipid m.	MPN533	pyruvate m.	sick
MPN079	sugar m.	MPN392	pyruvate m.	sick	MPN420	lipid m.	MPN595	PPP	sick
MPN079	sugar m.	MPN393	pyruvate m.	sick	MPN420	lipid m.	MPN674	pyruvate m.	sick
MPN079	sugar m.	MPN394	pyruvate m.	sick	MPN428	pyruvate m.	MPN595	PPP	sick
MPN079	sugar m.	MPN428	pyruvate m.	sick	MPN428	pyruvate m.	MPN674	pyruvate m.	SL
MPN079	sugar m.	MPN533	pyruvate m.	sick	MPN445	lipid m.	MPN533	pyruvate m.	sick
MPN079	sugar m.	MPN595	PPP	sick	MPN445	lipid m.	MPN595	PPP	sick
MPN079	sugar m.	MPN674	pyruvate m.	sick	MPN445	lipid m.	MPN674	pyruvate m.	sick
MPN082	PPP	MPN133	sugar m.	sick	MPN479	CoA m.	MPN533	pyruvate m.	sick
MPN082	PPP	MPN134	sugar m.	sick	MPN479	CoA m.	MPN595	PPP	sick
MPN082	PPP	MPN135	sugar m.	sick	MPN479	CoA m.	MPN674	pyruvate m.	sick
MPN082	PPP	MPN136	sugar m.	sick	MPN492	PPP	MPN533	pyruvate m.	sick
MPN082	PPP	MPN258	sugar m.	SL	MPN492	PPP	MPN595	PPP	sick
MPN082	PPP	MPN259	sugar m.	SL	MPN492	PPP	MPN674	pyruvate m.	sick
MPN082	PPP	MPN260	sugar m.	SL	MPN493	sugar m.	MPN533	pyruvate m.	sick
MPN082	PPP	MPN306	aa m.	sick	MPN493	sugar m.	MPN595	PPP	sick
MPN082	PPP	MPN307	aa m.	sick	MPN493	sugar m.	MPN674	pyruvate m.	sick
MPN082	PPP	MPN320	folate m.	sick	MPN494	sugar m.	MPN533	pyruvate m.	sick
MPN082	PPP	MPN386	NT m.	sick	MPN494	sugar m.	MPN595	PPP	sick
MPN082	PPP	MPN390	pyruvate m.	sick	MPN494	sugar m.	MPN674	pyruvate m.	sick
MPN082	PPP	MPN391	pyruvate m.	sick	MPN495	sugar m.	MPN533	pyruvate m.	sick
MPN082	PPP	MPN392	pyruvate m.	sick	MPN495	sugar m.	MPN595	PPP	sick
MPN082	PPP	MPN393	pyruvate m.	sick	MPN495	sugar m.	MPN674	pyruvate m.	sick
MPN082	PPP	MPN394	pyruvate m.	sick	MPN496	sugar m.	MPN533	pyruvate m.	sick
MPN082	PPP	MPN428	pyruvate m.	sick	MPN496	sugar m.	MPN595	PPP	sick
MPN082	PPP	MPN533	pyruvate m.	sick	MPN496	sugar m.	MPN674	pyruvate m.	sick
MPN082	PPP	MPN674	pyruvate m.	sick	MPN497	sugar m.	MPN533	pyruvate m.	sick
MPN108	aa m.	MPN082	PPP	sick	MPN497	sugar m.	MPN595	PPP	sick
MPN108	aa m.	MPN133	sugar m.	sick	MPN497	sugar m.	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN134	sugar m.	sick	MPN498	PPP	MPN533	pyruvate m.	sick
MPN108	aa m.	MPN135	sugar m.	sick	MPN498	PPP	MPN595	PPP	sick
MPN108	aa m.	MPN136	sugar m.	sick	MPN498	PPP	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN251	PPP	sick	MPN533	pyruvate m.	MPN595	PPP	sick

Continued on next page

Table A.13 – continued from previous page

gene 1	pathway 1	gene 2	pathway 2	interaction type	gene 1	pathway 1	gene 2	pathway 2	interaction type
MPN108	aa m.	MPN258	sugar m.	sick	MPN533	pyruvate m.	MPN674	pyruvate m.	SL
MPN108	aa m.	MPN259	sugar m.	sick	MPN560	aa m.	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN260	sugar m.	sick	MPN564	pyruvate m.	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN306	aa m.	sick	MPN595	PPP	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN307	aa m.	sick	MPN609	sugar m.	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN320	folate m.	sick	MPN610	sugar m.	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN386	NT m.	sick	MPN611	sugar m.	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN390	pyruvate m.	sick	MPN637	lipid m.	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN391	pyruvate m.	sick	MPN651	sugar m.	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN392	pyruvate m.	sick	MPN652	sugar m.	MPN674	pyruvate m.	sick
MPN108	aa m.	MPN393	pyruvate m.	sick	MPN653	sugar m.	MPN674	pyruvate m.	sick

Table A.13.: Synthetic lethal and sick interactions: Pairs of genes with their respective pathway affiliations and the type of interaction when simulating double knock-outs *in silico* are shown; metabolism is abbreviated by 'm.'.

Table A.14: Energetic Expenditures in *M. pneumoniae*

time of growth in hours	24	36	48	60	source
total ATP produced	49359	60702	67683	77666	*in silico* result
ATP used for DNA production	27	31	22	10	*in silico* result
ATP used for RNA production	5071	5079	5059	5034	*in silico* result
ATP used for protein production	5464	5961	4694	3168	*in silico* result
ATP used for protein degradation	648	648	648	648	*in silico* result (based on half-life [Maier et al., 2011])
ATP used for lipid production	227	258	180	85	*in silico* result
ATP used for other defined functions	2564	2962	1947	726	*in silico* result
ATP used max. for protein folding	174	161	161	175	upper boundary calculation
ATP used max. for DNA degradation & repair	5	6	4	2	upper boundary calculation
ATP used max for post-translational modifications	14	17	18	20	upper boundary calculation
ATP used max by ATPase (130rps)	39780	38610	38610	58500	upper boundary calculation
rest ATP	-4616	6970	16341	9297	total ATP - defined expenses

Table A.14.: The ATP produced and used for various cellular functions at different time points of the exponential growth phase in batch culture growth, as well as the source for the respective amount are shown. *In silico* results have been extracted from the respective predicted flux distributions, upper boundary calculations are described in detail in section 3.3.7.

Table A.15: *In Silico* Fluxes of *M. pneumoniae*

reaction ID	24h	36h	48h	60h
M001	5.10836	7.369	9.69714	12.3776
M002	5.0835	7.34078	9.67749	12.3683
M003	5.16243	7.45648	9.8343	12.5723
M004	5.16243	7.45648	9.8343	12.5723
M005	-5.40839	-7.81419	-10.3117	-13.1877
M006	10.6103	15.3285	20.2244	25.862
M007	-10.6103	-15.3285	-20.2244	-25.862
M008	-10.6103	-15.3285	-20.2244	-25.862
M009	10.6103	15.3285	20.2244	25.862

Continued on next page

Table A.15 – continued from previous page

reaction ID	24h	36h	48h	60h
M010	4.9216	7.37505	9.95335	12.9233
M011	-3.16668	-8.39742	-15.6053	-23.1843
M012	7.4436	6.9311	4.6191	2.6777
M013	-7.4436	-6.9311	-4.6191	-2.6777
M014	7.4436	6.9311	4.6191	2.6777
M015	7.4436	6.9311	4.6191	2.6777
M016	-7.4436	-6.9311	-4.6191	-2.6777
M017	7.44291	6.93027	4.61863	2.67759
M018	0.0922169	0.100645	0.0791637	0.0533939
M019	0	0	0	0
M020	0	0	0	0
M021	0	0	0	0
M022	0	0	0	0
M023	0	0	0	0
M024	0	0	0	0
M025	0	0	0	0
M026	0	0	0	0
M027	0	0	0	0
M028	0	0	0	0
M029	0	0	0	0
M030	0	0	0	0
M031	0	0	0	0
M032	0.127709	0.184225	0.242428	0.30944
M033	0.127709	0.184225	0.242428	0.30944
M034	-0.0394621	-0.0578489	-0.0784016	-0.102028
M035	0.0394621	0.0578489	0.0784016	0.102028
M036	-0.0394621	-0.0578489	-0.0784016	-0.102028
M037	0.0789242	0.115698	0.156803	0.204056
M038	0.0789242	0.115698	0.156803	0.204056
M039	0	0	0	0
M040	-0.00932266	-0.0106783	-0.00722312	-0.00335526
M041	0	0	0	0
M042	0	0	0	0
M043	0	0	0	0
M044	-0.127709	-0.184225	-0.242428	-0.30944
M045	0.127709	0.184225	0.242428	0.30944
M046	0.127709	0.184225	0.242428	0.30944
M047	0.245959	0.357712	0.477379	0.615333
M048	0.01	0.01	0.01	0.01
M049	0.01	0.01	0.01	0.01
M050	0.00945854	0.0107382	0.00747673	0.00354742
M051	0.00945854	0.0107382	0.00747673	0.00354742
M052	0	0	0	0
M053	0	0	0	0
M054	0	0	0	0
M055	0	0	0	0
M056	0.00284386	0.0032286	0.00224799	0.00106659
M057	-0.0170631	-0.0193716	-0.013488	-0.00639952
M058	0.0170631	0.0193716	0.013488	0.00639952
M059	0.00853157	0.00968581	0.00674398	0.00319976
M060	0.00284386	0.0032286	0.00224799	0.00106659
M061	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M062	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M063	0	0	0	0
M064	0	0	0	0
M065	-0.00281929	-0.00468143	-0.00315359	-0.00146034
M066	-8.84E-05	-0.000100403	-6.99E-05	-3.32E-05
M067	0	0	0	0
M068	0.613989	0.651928	0.558451	0.446129
M069	0	0.00143352	0.000998121	0.00047357

Continued on next page

Table A.15 – continued from previous page

reaction ID	24h	36h	48h	60h
M070	0.00126269	0.00143352	0.000998121	0.000621014
M071	0.00783096	0.00784313	0.00781211	0.00777474
M072	1.1943	1.23496	1.13134	1.00678
M073	-0.00217501	-0.00103576	-0.000721171	-0.000342168
M074	0.00217501	0.00103576	0.000721171	0.000342168
M075	0.00217501	0.00103576	0.000721171	0.000342168
M076	0.00217501	0.00246927	0.00171929	0.000815738
M077	0	0	0	0
M078	7.25E-05	8.23E-05	5.73E-05	2.72E-05
M079	-0.00378145	-0.00285953	-0.00199102	-0.000944662
M080	-7.40E-05	-8.40E-05	-5.85E-05	-2.77E-05
M081	0	0	0	0
M082	0.220529	0.219608	0.218739	0.217693
M083	0.00126269	0	0	0
M084	0.219267	0.219608	0.218739	0.217693
M085	0	0	0	0
M086	-0.000187323	-0.00164618	-0.00114619	-0.000543825
M087	0.000187323	0.00164618	0.00114619	0.000543825
M088	0.000187323	0.00164618	0.00114619	0.000543825
M089	0.00145001	0.00164618	0.00114619	0.000543825
M090	0	0	0	0
M091	0.00679791	0.00781201	0.00522737	0.00240835
M092	-0.00916025	-0.010494	-0.00709473	-0.00329435
M093	0	0	0	0
M094	0.00904681	0.0103652	0.00700506	0.0032518
M095	0.195774	0.196078	0.195303	0.194368
M096	0.212837	0.21545	0.208791	0.200768
M097	0	0	0	0
M098	0	0	0	0
M099	0.00162367	0.00184334	0.00128347	0.000756401
M100	0	0	0	0
M101	0.140962	0.141181	0.140621	0.140094
M102	0.140962	0.141181	0.140621	0.139947
M103	0	0	0	0.000147444
M104	0.00145001	0.00164618	0.00114619	0.000396382
M105	0	0	0	0
M106	0.00145001	0.00164618	0.00114619	0.000396382
M107	0.00145001	0.00164618	0.00114619	0.000543825
M108	0.00459279	0.00521416	0.00363049	0.00172253
M109	0.00236234	0.00268194	0.00186737	0.000885993
M110	0.00223046	0.00253222	0.00176312	0.000836533
M111	0	0	0	0
M112	0	0	0	0
M113	0.00223046	0.00253222	0.00176312	0.000836533
M114	0	0	0	0
M115	0	0	0	0
M116	5.54E-05	6.29E-05	4.38E-05	2.08E-05
M117	0.00217501	0.00246927	0.00171929	0.000815738
M118	0.00217501	0.00246927	0.00171929	0.000815738
M119	0.240511	0.239227	0.242499	0.246441
M120	-0.240511	-0.239227	-0.242499	-0.246441
M121	-0.240511	-0.239227	-0.242499	-0.246441
M122	-4.46E-06	-5.07E-06	-3.53E-06	-1.67E-06
M123	-0.00223492	-0.00253729	-0.00176665	-0.000838208
M124	0	0	0	0
M125	0.00106551	0.00116246	0.000915358	0.000617665
M126	0.00106551	0.00116246	0.000915363	0.000617665
M127	-0.00330044	-0.00369975	-0.00268201	-0.00145587
M128	0.00106998	0.00116753	0.000918892	0.000619339
M129	0.00999554	0.00999493	0.00999647	0.00999833

Continued on next page

Table A.15 – continued from previous page

reaction ID	24h	36h	48h	60h
M130	0.01	0.01	0.01	0.01
M131	0.00106538	0.00116231	0.000915257	0.000617615
M132	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M133	0	0	0	0
M134	0	0	0	0
M135	-0.000197175	-0.000271053	-8.28E-05	-3.35E-06
M136	0.000197175	0.000271053	8.28E-05	3.35E-06
M137	0.000197175	0.000271053	8.28E-05	3.35E-06
M138	0	0	0	0
M139	0.000197175	0.000271053	8.28E-05	3.35E-06
M140	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M141	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M142	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M143	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M144	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M145	4.60E-06	5.22E-06	3.63E-06	1.72E-06
M146	4.60E-06	5.22E-06	3.63E-06	1.72E-06
M147	4.60E-06	5.22E-06	3.63E-06	1.72E-06
M148	1.34E-07	1.52E-07	1.06E-07	5.02E-08
M149	0	0	0	0
M150	1.34E-07	1.52E-07	1.06E-07	5.02E-08
M151	0.0189171	0.0214764	0.0149535	0.00709484
M152	0	0	0	0
M153	0.018112	0.0197599	0.0155598	0.0104997
M154	0.0063928	0.00697447	0.00549196	0.00370589
M155	0.0234394	0.025572	0.0201365	0.0135879
M156	0.0266351	0.0290585	0.022882	0.0154406
M157	0.0308976	0.0337088	0.0265437	0.0179114
M158	0.00319613	0.00348693	0.00274577	0.00185284
M159	0.0234398	0.0255725	0.0201368	0.0135881
M160	0.018112	0.0197599	0.0155598	0.0104997
M161	0.0138502	0.0151103	0.0118986	0.00802909
M162	0.0106538	0.0116231	0.00915257	0.00617615
M163	0.00319613	0.00348693	0.00274577	0.00185284
M164	0.0223736	0.0244093	0.0192209	0.0129702
M165	0.0223731	0.0244087	0.0192205	0.01297
M166	0.0149154	0.0162725	0.0128137	0.00864666
M167	0.0191776	0.0209225	0.0164753	0.0111174
M168	0.0213079	0.0232467	0.0183055	0.0123524
M169	0.0340935	0.0371956	0.0292894	0.0197642
M170	0.0063924	0.00697401	0.00549165	0.00370574
M171	0.0159813	0.0174354	0.0137294	0.00926447
M172	0.0287657	0.031383	0.0247124	0.0166758
M173	0.0223731	0.0244087	0.0192205	0.01297
M174	0.007741	0.007741	0.007741	0.007741
M175	-7.68518	-7.17149	-4.86251	-2.92476
M176	0	0	0	0
M177	0	0	0	0
M178	3.16668	8.39742	15.6053	23.1843
M179	7.68887	7.28799	5.09601	3.29292
M180	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M181	0.00164613	0.00186883	0.00130122	0.000764823
M182	0	0	0	0
M183	0.00404275	0.00458969	0.00319569	0.00151623
M184	-0.245959	-0.357712	-0.477379	-0.615333
M185	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M186	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M187	0.0189171	0.0214764	0.0149535	0.00709484
M188	-7.53968	-7.05254	-4.47367	-2.21489
M189	0.00508274	0.00581759	0.00394467	0.00183568

Continued on next page

Table A.15 – continued from previous page

reaction ID	24h	36h	48h	60h
M190	0.000197175	0.000271053	8.28E-05	3.35E-06
M191	4.60E-06	5.22E-06	3.63E-06	1.72E-06
M192	0.0184901	0.0209444	0.014689	0.00715274
M193	0.00436503	0.00495557	0.00345044	0.0016371
M194	0.485417	0.483397	0.488546	0.494601
M195	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M196	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M197	0.240511	0.239227	0.242499	0.246441
M198	7.4436	6.9311	4.6191	2.6777
M199	0	0	0	0
M200	0	0	0	0
M201	0	0	0	0
M202	0	0	0	0
M203	0	0	0	0
M204	0	0	0	0
M205	0	0	0	0
M206	0	0	0	0
M207	0	0	0	0
M208	0	0	0	0
M209	0	0	0	0
M210	0	0	0	0
M211	0	0	0	0
M212	0	0	0	0
M213	0	0	0	0
M214	0	0	0	0
M215	0	0	0	0
M216	0	0	0	0
M217	0	0	0	0
M218	0	0	0	0
M219	0	0	0	0
M220	0.0060428	0.00686034	0.00477668	0.00211891
M221	0.00948916	0.010773	0.00750094	0.00355891
M222	0.0133134	0.0151146	0.0105239	0.00499318
M223	0.00215242	0.00244363	0.00170144	0.000807266
M224	0.0165966	0.018842	0.0131192	0.00622456
M225	0.0121573	0.0138493	0.00953696	0.00434156
M226	0.0158465	0.0179904	0.0125263	0.00594323
M227	0.019784	0.0224606	0.0156387	0.00741998
M228	0.0143324	0.0162714	0.0113294	0.00537535
M229	0.0217309	0.0246709	0.0171777	0.00815018
M230	0.0121938	0.0138435	0.00963887	0.00457327
M231	0.00440712	0.00500337	0.00348372	0.00165289
M232	0.0230065	0.026119	0.018186	0.00862857
M233	0.010336	0.0117344	0.00817038	0.00387653
M234	0.0109765	0.0124615	0.00867663	0.00411673
M235	0.0151573	0.0172079	0.0119814	0.00568473
M236	0.00223265	0.0025347	0.00176485	0.000837354
M237	0.00722576	0.00820333	0.00571177	0.00271002
M238	0.0180373	0.0204776	0.014258	0.00676488
M239	0	0	0	0
M240	0.0945652	0.146456	0.216448	0.297221
M241	3.16668	8.39742	15.6053	23.1843
M242	0	0	0	0
M243	7.4436	6.9311	4.6191	2.6777
M244	-0.00508274	-0.00581759	-0.00394467	-0.00183568
M245	0	0	0	0
M246	0.0446431	0.0506829	0.0352892	0.0167434
M247	7.68518	7.17149	4.86251	2.92476
M248	-0.0121573	-0.0138493	-0.00953696	-0.00434156
M249	0.245959	0.357712	0.477379	0.615333

Continued on next page

Table A.15 – continued from previous page

reaction ID	24h	36h	48h	60h
M250	-0.019784	-0.0224606	-0.0156387	-0.00741998
M251	-0.25	-0.25	-0.25	-0.25
M252	-0.0143324	-0.0162714	-0.0113294	-0.00537535
M253	-0.0133134	-0.0151146	-0.0105239	-0.00499318
M254	-0.00215242	-0.00244363	-0.00170144	-0.000807266
M255	-0.0165966	-0.018842	-0.0131192	-0.00622456
M256	-0.0121938	-0.0138435	-0.00963887	-0.00457327
M257	-0.00440712	-0.00500337	-0.00348372	-0.00165289
M258	0	0	0	0
M259	-0.0158465	-0.0179904	-0.0125263	-0.00594323
M260	-0.0217309	-0.0246709	-0.0171777	-0.00815018
M261	-0.0230065	-0.026119	-0.018186	-0.00862857
M262	-0.00436503	-0.00495557	-0.00345044	-0.0016371
M263	0.240511	0.239227	0.242499	0.246441
M264	-0.0109765	-0.0124615	-0.00867663	-0.00411673
M265	-0.010336	-0.0117344	-0.00817038	-0.00387653
M266	-0.0184901	-0.0209444	-0.014689	-0.00715274
M267	-0.0151573	-0.0172079	-0.0119814	-0.00568473
M268	-0.00223265	-0.0025347	-0.00176485	-0.000837354
M269	-0.00722576	-0.00820333	-0.00571177	-0.00271002
M270	-0.0180373	-0.0204776	-0.014258	-0.00676488
M271	0.000188246	0.000260917	7.57E-05	0
M272	0.485417	0.483397	0.488546	0.494601
M273	0	0	0	0
M274	-4.46E-06	-5.07E-06	-3.53E-06	-1.67E-06
M275	0.00164613	0.00186883	0.00130122	0.000764823
M276	0.0060428	0.00686034	0.00477668	0.00211891
M277	0	0	0	0
M278	0	0	0	0
M279	0.127709	0.184225	0.242428	0.30944
M280	0.0189171	0.0214764	0.0149535	0.00709484
M281	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M282	5.10836	7.369	9.69714	12.3776
M283	0.127709	0.184225	0.242428	0.30944
M284	0.00404275	0.00458969	0.00319569	0.00151623
M285	-10.2246	-14.9068	-19.8945	-25.6422
M286	0	0	0	0
M287	0.000197175	0.000271053	8.28E-05	3.35E-06
M288	0.0945652	0.146456	0.216448	0.297221
M289	7.68887	7.28799	5.09601	3.29292
M290	4.60E-06	5.22E-06	3.63E-06	1.72E-06
M291	0	0	0	0
M292	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M293	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M294	0.127709	0.184225	0.242428	0.30944
M295	4.46E-06	5.07E-06	3.53E-06	1.67E-06
M296	0	0	0	0
M297	0	0	0	0
M298	-7.53968	-7.05254	-4.47367	-2.21489
M299	0.00106538	0.00116231	0.000915257	0.000617615
M300	0	0	0	0
M301	0	0	0	0
M302	0.0003489	0.0003489	0.0003489	0.0003489
M303	1.34E-07	1.52E-07	1.06E-07	5.02E-08
M304	13.105	16.9618	20.435	25.2015
M305	4.46E-08	5.07E-08	3.53E-08	1.67E-08
M306	0.0446431	0.0506829	0.0352892	0.0167434

Table A.15.: *In silico* fluxes predicted for different time points of the exponential growth phase of *M. pneumoniae* batch culture growth by the final model *i*JW145.

Table A.16: Fitting Parameters for One-phase Exponential Decay Functions

	G3P	G1P	R5P	G6P	F6P	FBP	DHAP
span	-77.1	-99.23	-88.62	-99.52	-99.36	-99.53	-98.9
K	0.3716	31510000	0.3764	16.63	11.42	16.37	18.49
plateau	77.1	99.23	88.62	99.52	99.36	99.53	98.9
$\mathbf{R^2}$	0.8173	0.9998	0.9679	0.9998	0.9996	1	0.9999

Table A.16.: Paramters for fitting a one-phase exponential decay function (section 3.2.1, Equation 3.4) to the incorporation rate of heavy labeled isotop incorporation into intracellular pools of glycolytic intermediates and key metabolites of pathway interconnecting branches.

Table A.17: Fitting Parameters for Two-phase Exponential Decay Functions

	G3P	G1P	R5P	G6P	F6P	FBP	DHAP
$\mathbf{Y_0}$	0.7979	0	0	0	0	1.911E-09	1.204E-08
$\mathbf{part_1}$	0.4616	0.3242	0.7369	0.9782	0.9188	0.2902	0.3242
$\mathbf{span_1}$	-45.58858536	-32.170366	-73.17417	-97.67327	-91.659488	-28.883606	-32.12822
$\mathbf{K_1}$	1.478	2693	0.6003	707600000	28480000	16.43	18.38
$\mathbf{span_2}$	99.56	99.23	99.3	99.85	99.76	99.53	99.1
$\mathbf{K_2}$	-53.17351464	-67.059634	-26.12583	-2.17673	-8.100512	-70.646394	-66.97177999
plateau	0.01663	2693	0.01574	0.404	0.6142	16.35	17.46
$\mathbf{R^2}$	0.9973	0.9998	0.9995	1	1	1	0.9998

Table A.17.: Paramters for fitting a two-phase exponential decay function (section 3.2.1, Equation 3.5) to the incorporation rate of heavy labeled isotope into intracellular pools of glycolytic intermediates and key metabolites of pathway interconnecting branches.

Table A.18: Comparative Calculations for *M. pneumoniae* and *E. coli*

property	***Mpn***	***E.coli***	**unit**
volume	6.70E-17	1.10E-06	liter
proteins	130000	2350000	molecules/cell
mRNA	230	4352	molecules/cell
average mRNA length	1035	705	bases/cell
protein:mRNA	565.22	540	proteins/mRNA
mRNA half-life	1	3.69	min
bases in mRNA	238050	3068056	molecules/cell
surface (assuming ideal sphere form)	7.97751E-11	0.000515323	m^2
suface:volume	1190672.714	468.4750222	m^2/liter
ATP to de novo synthesize mRNA	1047420	13499444	molecules/replication
ATP to account for mRNA half-life	338164308	175602529	molecules/replication
total ATP used for mRNA synthesis during 1 cell replication	339211728	189101974	molecules
ribosomes	190	20100	molecules/cell
average rRNA length	4523	4566	nucleotides
bases in rRNA	859370	91776600	nucleotides/cell
ATPs to replicate rRNA of 1 cell	3781228	403817040	molecules/replication
tRNA per cell	3300	58000	molecules
average tRNA length	80	80	nucleotides
bases in tRNA	264000	4640000	nucleotides/cell
ATP to replicate tRNA	1161600	20416000	molecules/replication
bases in DNA	1632788	9279350	nucleotides/cell
ATP for DNA replication	8817055.2	50108490	molecules/replication
average protein length	345	235	amino acids
peptide bonds in the proteome	44720000	549900000	$cell^{-1}$
ATP for de novo protein synthesis	196768000	2419560000	molecules/replication
average protein half-life	23	-	hours
ATP to account for protein synthesis due to turnover	92068758	-	molecules/replication
ATP for protein degradation	76723965	-	molecules/replication
ATP to account for protein turnover	168792723	-	molecules/replication
ATP used for DNA, RNA & protein replication	718532335	3083003504	molecules/replication
ATP used for DNA, RNA & protein replication	18481	1284585	molecules/second
ATP used for protein replication	50.88	78.48	% of total ATP used for protein, DNA & RNA production

Table A.18.: In the upper part of the table, the information extracted from Yus et al. [2009] and the Bionumbers database [Milo et al., 2010] and used for the calculations are listed, the second part contains calculations on the surface to volume ratio for both organisms and on the dedication of energy to the production of major biomass building blocks.

B. Supplementary Material for Chapter 4

B.1. Pseudocode

Script for Translating Nucleotide Sequences Into All Putatively Encoded Peptides:

```
var1 = 0
pos = 0
while (NT sequence != "") do {
  if (var1 = 0) do {
    if codon(pos) is TSC do {
      protein = "M"
      determine start position from pos
      var1 = 1
    }
  }
  else do {
    if (codon(pos) != stop codon) do {
      add amino acid(codon) to protein
    }
    else do {
      print protein sequence to the list of peptides
      var1 = 0
    }
  }
  cut off the first 3 NTs from NT sequence
  pos = pos + 3
  if length(NT sequence < 3) do {
    NT sequence = ""
  }
}
```

Thereby, *pos* defines the position of the sequence, at which the script is momentaneously working (this can be either an internal position of the sequence or a genome position), *NT sequence* is the DNA sequence that shall be translated into peptides, *codon*(*pos*) stands for the DNA codon starting at position *pos*, *amino acid*(*codon*) is the amino acid encoded by *codon*, and *protein* contains the translated amino acid sequence.

B.2. Tables

Table B.1: Genome Re-annotation

gene ID (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN001	+	655 597		692	692	1834	alternative TSS → 2 transcripts		
MPN002	+		2720	1838	1838	2767			
MPN003	+	2800		2869	2869	4821			yes
MPN004	+			4821	4821	7340			
MPN005	+	7225		7312	7312	8574			yes
MPN006	+			8579	8549	9211	putative longer	longer	yes
MPN007	+		10000	9184	9184	9945			
MPN008	+		11310	9947	9947	11275			
MPN009	+		11960	11275	11275	12060			
MPN010	+	12342/ 12372	12750	12257	12392	12652	alternative TSS → 2 transcripts or shorter	shorter	
MPN011	-		12865	13533	13533	12838			
MPN012	-	14310	13450	14265	14265	13558			yes
MPN013	+	15030/ 15132	15800	14992	15088	15765	alternative TSS → 2 transcripts or shorter	shorter	
MPN014	+	15890/ 15880	160125	15867	15939	16505	alternative TSS → 2 transcripts or shorter	shorter	
MPN015	-		16170	17339	17339	16482			
MPN016	-	18345	17493	18205	18205	17339			yes
MPN017	-	19032	18345	18989	18989	18180			yes
MPN018	+	19314		19325	19325	21196			
MPN019	+		23120	21108	21108	23012			
MPN020	+	22876		23022	23022	26114			yes
MPN021	+	26140		26160	26160	27332			
MPN022	+		28550	27316	27316	28245			
MPN023	+		29803	28245	28245	29783			
MPN024	+	29803		29804	29804	30244			
MPN025	+		31200	30244	30244	31110			
MPN026	+	31085	32900	31111	31111	32199			
MPN027	-	33036	31925	33026	33026	32202			
MPN028	+	33046	33960	33059	33059	33958			
MPN029	+	33800		33979	33979	34551	putative NEW		yes
MPN030	+		35000	34469	34469	34975			
MPN031	+		36000	34975	34975	35586			
MPN032	-		35800	36136	36136	35810			
MPN033	-			36760	36760	36140			
MPN034	-			41131	41131	36800			
MPN035	+	41322	43400	41409	41409	43409			yes
MPN035a	-	42050			42013	41696	alternative TSS → NEW	NEW	
MPN036	+	43500	45640	43581	43581	45602			yes
MPN037	+			45770	45770	46213			
MPN037a	-	46145/ 46025	45765		46068	45736	alternative TSS → NEW	NEW	
MPN038	-	46750	46480	46792	46711	46442	putative shorter	shorter	yes
MPN039	+	47103	48050	47194	47194	48210			yes
MPN040	-			48416	48416	48105			
MPN041	+	48550		48670	48670	49230			yes
MPN042	+	48961		49292	49292	51310	putative NEW		yes
MPN043	+	51580	52600	51634	51634	52428			yes
MPN044	-	53069	52400	53050	53050	52475			
MPN045	+	53073	54400	53077	53077	54321			
MPN046	+	54156	56000	54293	54293	55966			yes
MPN047	+		57350	55942	55942	57297			
MPN047a	-	56979			56759	56502	alternative TSS → NEW	NEW	
MPN048	+		59488	57886	57886	59442			
MPN048a	+	57000			57029	57229	alternative TSS → NEW	NEW	
MPN049	+	59480	61495	59619	59619	61517			yes
MPN050	-		61800	63503	63503	61977			
MPN051	-	64655	63200	64648	64648	63494			
MPN052	+	64805		64831	64831	66804			
MPN053	+		67250	66822	66822	67088			
MPN054	+			68627	68627	68998			
MPN055	+	70382	72200	70404	70404	72086			
MPN056	+		73600	72088	72088	72948			
MPN057	+			72941	72941	73801			
MPN058	+		75203	73786	73786	75243			
MPN059	+	75203	76100	75243	75243	76202			
MPN060	+		77100	76186	76186	77337			
MPN060a	+	77198	77420		77280	77594	putative NEW	NEW	
MPN061	-	78982	77575	0	78977	77625			
MPN062	+	79008		79033	79033	79749			
MPN063	+			79753	79753	80427			
MPN064	+			80414	80414	81679			
MPN065	+		82000	81690	81690	82091			
MPN066	+		83850	82081	82081	83745			

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN067	-		83175	84675	84675	83713			
MPN068	-			85060	85060	84683			
MPN069	-	85350/ 85220	84975	85212	85212	85066	alternative TSS → 2 transcripts		yes
MPN070	-	85630	85290	85611	85611	85228			
MPN071	-	86485	85760	86428	86428	85598			yes
MPN072	-		86485	86946	86946	86410			
MPN073	-	88480/ 88005	86920	88091	88091/ 87960	86925	alternative TSS → 2 transcripts or NEW	2	yes
MPN074	-	88485		88341	88341	87898			yes
MPN075	+			88348	88348	89247			
MPN076	-	91057/ 89520	89250	90993	90993	89299	alternative TSS → longer		yes
MPN077	-	92905	91200	92839	92884	91199	putative longer	longer	yes
MPN078	+	93165		93191	93191	95275			
MPN079	+		96171	95268	95268	96170			
MPN080	+			96407	96407	100567			
MPN081	+		101980	100567	100567	101964			
MPN082	+	101980	103900	101985	101985	103931			
MPN083	+	103930		103941	103941	105542			
MPN084	+		107250	105592	105592	107166			
MPN085	+	107273	108690	107273	107273	108595			
MPN086	+	108780/ 108856	109200	108857	108857	109174	alternative TSS → 2 transcripts		yes
MPN087	+			109346	109346	109798			
MPN088	+	110230/ 110394	111350	110813	110813	111118	alternative TSS → 2 transcripts or NEW		yes
MPN089	+	111470	112600	111610	111610	112617			yes
MPN090	+	112710		112772	112772	113761			yes
MPN091	+	113646		113838	113838	114254			yes
MPN091a	-	114140			114043	113870	alternative TSS → NEW	NEW	
MPN092	+			114572	114572	115093			
MPN093	+		116000	114948	114948	115853			
MPN094	+	116435		116287	116455	116709	putative shorter	shorter	yes
MPN095	+	123250		123290	123290	124054			
MPN096	+		124998	124054	124054	124848			
MPN097	+	124998	126950	125024	125024	126649			
MPN098	+	126625	127400	126947	126947	127390			yes
MPN099	+			128076	128076	129119			
MPN100	+	129608	130007	129458	129626	130009	putative shorter	shorter	yes
MPN101	+			130466	130466	131752			
MPN101a	+	131776			131848	132513	putative NEW	NEW	
MPN102	+			132568	132568	133386			
MPN103	-			134570	134573	134055			
MPN104	+	134175/ 134212	134900	134583	134583	134897	alternative TSS → longer		yes
MPN104a	+	134985	135350		135088	135360	putative NEW	NEW	
MPN105	+	135718	137100	135731	135731	136756			
MPN106	+		139000	136759	136759	139176			
MPN107	+		140200	139289	139289	140044			
MPN108	+		141600	140046	140046	141260			
MPN109	+	141936	142650	141940	141940	142437			
MPN110	+		143180	142361	142361	144517			
MPN111	+	144431/ 144542	147138	145021	145021	146289	alternative TSS → 2 transcripts or NEW		yes
MPN112	+	147128	147750	147149	147149	147541			
MPN113	+			148090	148090	148761			
MPN114	+		150525	148751	148751	150553			
MPN115	+	150556/ 150588		150643	150643	151248	alternative TSS → 2 transcripts		yes
MPN116	+			151254	151254	151433			
MPN117	+		151790	151444	151444	151827			
MPN118	-	152520	151740	152512	152512	151802			
MPN119	+	152539		152568	152568	155300			
MPN120	+			155323	155323	155976			
MPN121	+			155976	155976	156341			
MPN122	+		158200	156467	156467	158374			
MPN123	+		160900	158374	158374	160743			
MPN124	-	161790	160080	161790	161785	160730			
MPN125	+		163600	161817	161817	163577			
MPN126	+	163604	164250	163613	163613	164092			
MPN127	+	164598/ 164623	165240	164484	164628	165026	alternative TSS → 2 transcripts or shorter or NEW	shorter	yes
MPN127a	+	165786			165867	166262	putative NEW	NEW	
MPN128	+	166310		166483	166483	166932	putative longer or NEW		yes
MPN129	+	167600/ 167795		167632	167632	168081	alternative TSS → 2 transcripts		
MPN130	+	169030/ 169190	169510	169042	169042/ 169210	169464	alternative TSS → 2 transcripts		yes

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN131	+	170040		170068	170068	170733			
MPN132	+			171049	171049	171819			
MPN132a	+	—			171494	171805	putative NEW	NEW	
MPN133	+			172277	172277	173182			
MPN134	+			173206	173206	174966			
MPN135	+			174959	174959	175948			
MPN136	+		177000	175920	175920	176879			
MPN137	-	178230/ 178060	177430	178143	178143 169210	177457	alternative TSS → 2 transcripts		yes
MPN138	-	178960/ 178850	178410	178890	178892 169210	178392	alternative TSS → 2 transcripts		yes
MPN139	-	179670/ 179465	179100	179620	179620 169210	179129	alternative TSS → 2 transcripts		yes
MPN140	+	179860	180355	179871	179871	180845			
MPN141	+			180858	180858	185741			
MPN142	+		189400	185747	185747	189403			
MPN143	+		190300	189520	189520	190047			
MPN144	+	190531		190621	190621	191862			yes
MPN145	+	192320		192214	192349	192753	putative shorter	shorter	yes
MPN146	+	192998	193900	193071	193071	193868			yes
MPN147	+		195000	193875	193875	195332			
MPN148	+	195450/ 195480	196400	195875	195539	196297	alternative TSS → longer	longer	yes
MPN149	+			196850	196850	198154			
MPN150	+			197982	197982	198656			
MPN151	+	198910/ 199020	199540	199139	199139	199540	alternative TSS → 2 transcripts		yes
MPN152	+	199800	202358	199823	199823	202207			
MPN152a	-	199982	198875		199929	199768	alternative TSS → NEW	NEW	
MPN153	+	202358	205724	202383	202383	205724			
MPN153a	-	205470			205468	205250	alternative TSS → NEW	NEW	
MPN154	+	205722	207500	205801	205801	207423			yes
MPN154a	-	206820			206273	206031	alternative TSS → NEW	NEW	
MPN155	+	207518		207704	207704	209557	putative NEW		yes
MPN155a	+	207500			207505	207717	putative NEW	NEW	
MPN156	+			209560	209560	209910			
MPN157	+	209912	211200	209926	209926	211134			
MPN158	+		211980	211147	211147	211956			
MPN159	+	211965		211982	211982	213256			
MPN160	+		214528	213243	213243	214376			
MPN161	+	214528	215850	214530	214530	215867			
MPN162	+	215970	216950	215989	215989	216951			
MPN163	+	217121/ 217540	217540	217198	217150	217536	alternative TSS → longer	longer	yes
MPN164	+	217550	217800	217630	217630	217956			yes
MPN165	+			217970	217970	218833			
MPN166	+			218833	218833	219471			
MPN167	+			219474	219474	220187			
MPN168	+			220187	220187	221050			
MPN169	+			221050	221050	221313			
MPN170	+			221240	221240	221794			
MPN171	+			221796	221796	222617			
MPN172	+			222617	222617	223036			
MPN173	+			223036	223036	223371			
MPN174	+			223371	223371	223628			
MPN175	+			223632	223632	224000			
MPN176	+			224000	224000	224335			
MPN177	+			224338	224338	224880			
MPN178	+			224882	224882	225067			
MPN179	+			225061	225061	225489			
MPN180	+			225496	225496	226050			
MPN181	+			226051	226051	226401			
MPN182	+	226232		226405	226405	227064			yes
MPN183	+			227068	227068	227523			
MPN184	+	227408		227523	227523	228956			yes
MPN185	+		229680	228950	228950	229597			
MPN186	+			229597	229597	230343			
MPN187	+		230579	230343	230343	230579			
MPN188	+	230585		230590	230590	230703			
MPN189	+			230703	230703	231077			
MPN190	+		231500	231077	231077	231442			
MPN191	+	231400		231448	231448	232431			yes
MPN192	+			232434	232434	232808			
MPN193	+			232792	232792	233616			
MPN194	+	233600		233604	233604	234515			
MPN195	+			234508	234508	235812			
MPN196	+	233600/ 236310		235805	235805/ 236318	236536	alternative TSS → 2 transcripts or NEW		yes
MPN197	+		238300	236529	236529	238364			

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN198	+	238246/ 238804	239373	238364	238364	239323	alternative TSS → 2 transcripts		yes
MPN198a	+	240170			240189	240398	alternative TSS → NEW	NEW	
MPN199	+		241952	239417	239417	241699			
MPN199a	+	241420			241575	241853	alternative TSS → NEW	NEW	
MPN200	+	241958	244379	241983	241983	244379			
MPN201	+		245950	244853	244853	245569			
MPN202	+			246049	246049	246990			
MPN203	+		247340	246909	246909	247292			
MPN204	+	247650/ 247740	248150	247655	247655	248101	alternative TSS → 2 transcripts		
MPN205	+	248351	249600	248562	248562	249878			yes
MPN206	-			250270	250274	249933			
MPN207	+		253152	250326	250326	253148			
MPN207a	+	249998			250000	250293	putative NEW	NEW	
MPN208	+	253750		253758	253758	254642			
MPN208a	-	253325			253317	253117	putative NEW	NEW	
MPN209	+		257203	254635	254635	257253			
MPN210	+	257205/ 257232		257240	257240	259666	alternative TSS → 2 transcripts		
MPN211	+		261600	259641	259641	261614			
MPN212	+	261600		261617	261617	262024			
MPN213	+		265300	262129	262129	265221			
MPN214	-	265662/ 265628	265200	265630	265634	265218	alternative TSS → 2 transcripts		
MPN215	+	265875		265910	265865	267079		longer	no (35 bp)
MPN216	+			267072	267072	268202			
MPN216a	-	—			267844	267638	putative NEW	NEW	
MPN217	+		269400	268204	268204	269475			
MPN218	+		272197	269438	269438	271993			
MPN219	+	272000		272049	272049	272462			yes
MPN220	+		273150	272462	272462	273142			
MPN221	+			273142	273142	273708			
MPN222	+			273701	273701	274570			
MPN223	+			274571	274571	275509			
MPN224	+	275422	276600	275499	275499	276668			yes
MPN225	+	276678		276697	276697	277116			
MPN226	+			277170	277170	277637			
MPN227	+		279726	277660	277660	279726			
MPN228	+	279755	280500	279783	279783	280430			
MPN229	+			280430	280430	280930			
MPN230	+			280920	280920	281234			
MPN231	+		281850	281237	281237	281686			
MPN232	+			281673	281673	283094			
MPN233	+		284500	283098	283098	284462			
MPN234	+			285190	285190	286443			
MPN235	+	286310	287190	286468	286468	287190			yes
MPN236	+	287260/ 287320	289150	287388	287388	28882[7]	alternative TSS → 2 transcripts		yes
MPN237	+			288830	288830	290266			
MPN238	+		291652	290256	290256	291692			
MPN239	+	291652		291667	291667	292335			
MPN240	+		293375	292328	292328	293275			
MPN241	+		294075	293265	293265	294107			
MPN242	+	294125		294140	294140	294370			
MPN243	+	294375	296550	294381	294381	296561			
MPN244	+		297450	296561	296561	297169			
MPN245	-	297732	297080	297790	297719	297138	putative shorter	shorter	yes
MPN246	+	297580/ 297745	298800	297610	297760	298329	alternative TSS → 2 transcripts or shorter	shorter	
MPN247	+		300050	298329	298329	299108			
MPN248	+			299099	299099	300268			
MPN249	+		301089	300255	300255	301091			
MPN250	+	301100	302680	301113	301113	302405			
MPN251	+			302383	302383	303030			
MPN252	+		304400	303037	303037	304404			
MPN253	+		305050	304361	304361	305044			
MPN254	+	305047		305066	305066	305539			
MPN255	+		306300	305565	305565	306320			
MPN256	+		307300	306313	306313	306984			
MPN257	+		307970	306965	306965	307981			
MPN258	+	307970		307996	307996	309714			
MPN259	+			309731	309731	311284			
MPN260	+			311284	311284	312219			
MPN261	+		314550	312256	312256	314391			
MPN262	+	313950	315844	314381	314381	315808	putative NEW		yes
MPN263	+	315855	316220	315874	315874	316182			
MPN264	+		317000	316194	316194	317039			
MPN265	-	318128	316860	318087	318087	317047	putative NEW		yes

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN266	+	318110		318144	318144	318581			
MPN267	+			318581	318581	319360			
MPN268	+	319350	319600	319360	319360	319713			
MPN269	+			319703	319703	321184			
MPN270	+		321500	321187	321187	321474			
MPN271	+	321715	322880	321736	321736	322491			
MPN272	+	324185	325000	324197	324197	324478			
MPN273	-		324400	324889	324889	324455			
MPN274	-	325807		325658	325658	324858			yes
MPN275	+	325620/ 325600	325920	325680	325680	325982	alternative TSS → 2 transcripts		yes
MPN276	-	327065/ 326820	325910	326811	326811	325954	alternative TSS → longer		yes
MPN277	+	326832	328375	326856	326856	328325			
MPN278	+			328285	328285	329484			
MPN279	+		331293	329484	329484	331229			
MPN280	+	331290	333069	331360	331360	333069			yes
MPN281	+	333255	334400	333280	333280	334413			
MPN282	+		335580	334768	334768	335268			
MPN283	+	336620/ 336511	337000	336479	336626	336826	alternative TSS → 2 transcripts or shorter	shorter	
MPN284	+	337743	340200	337770	337770	340154			
MPN285	+	340673	341400	340613	341207	341533	putative shorter	shorter	yes
MPN286	+		343220	341636	341636	343033			
MPN287	+	343620	344158	343764	343764	344120	putative longer		yes
MPN288	+	344690	346660	344716	344716	347079			
MPN289	+	346937		347169	347169	347732	putative NEW		yes
MPN290	+		348250	347871	347871	348308			
MPN291	-		348300	348891	348891	348301			
MPN292	-			349804	349804	348875			
MPN293	-	350458		350358	350358	349804			yes
MPN294	+	350392		350397	350397	351017			
MPN295	+			351066	351066	351728			
MPN296	+			351728	351728	351910			
MPN297	+			351935	351935	352384			
MPN298	+		352400	352387	352387	352746			
MPN299	+		353600	352734	352734	353534			
MPN300	+			353535	353535	355055			
MPN301	+		355675	355039	355039	355665			
MPN302	+	355675		355685	355685	356671			
MPN303	+		358180	356677	356677	358203			
MPN304	+	359160/ 359168	360200	359183	359183	359899	alternative TSS → 2 transcripts		
MPN305	+			359800	359800	360396			
MPN306	+			360644	360644	361465			
MPN307	+	359160/ 361750	361750	361468	361468/ 361783	362397	alternative TSS → 2 transcripts or NEW		yes
MPN308	+		364120	362400	362400	364097			
MPN309	+	364188	365780	364231	364231	365448			yes
MPN310	+	365356/ 370344	366000	365468	365468/ 3704124	370924	alternative TSS → 2 transcripts		yes
MPN311	+			370939	370939	372012			
MPN312	+		373000	372002	372002	372658			
MPN313	-			373013	373013	372741			
MPN314	+	373373		373414	373414	373839			yes
MPN315	+			373829	373829	374755			
MPN316	+			374758	374758	376017			
MPN317	+		377250	376023	376023	377165			
MPN318	+			377208	377208	378683			
MPN319	+	378420	380175	378658	378658	380169	putative longer or NEW		yes
MPN320	+	380638		380541	380664	381527	putative shorter	shorter	yes
MPN321	+			381529	381529	382011			
MPN322	+			382018	382018	383037			
MPN323	+			383040	383040	383501			
MPN324	+		385680	383525	383525	385690			
MPN325	+	385760	386120	385837	385837	386139			yes
MPN326	+			386132	386132	386434			
MPN327	+		387050	386427	386427	386741			
MPN328	+	386670		386734	386734	387594			yes
MPN329	+			387548	387548	388024			
MPN330	+		388500	388014	388014	388898			
MPN331	+	388910		388923	388923	390257			
MPN332	+	390282	392800	390328	390328	392715	putative longer		yes
MPN333	+			392910	392910	395162			
MPN334	+			395165	395165	396145			
MPN335	+		396380	396132	396132	398357			
MPN335a	-	397770			397616	397269	alternative TSS → NEW	NEW	
MPN336	-	399405	398150	399390	399392	398358			
MPN337	+	399440		399477	399477	401342			

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN338	+			401345	401345	403243			
MPN339	+		404000	403243	403243	403917			
MPN340	+			403921	403921	405510			
MPN341	+		408150	405573	405573	407720			
MPN342	+	407804	409750	407861	407861	409492			yes
MPN343	+	409438		409871	409871	410863	putative NEW		yes
MPN344	+	411098/ 411168	411850	411086	411383	411736	alternative TSS → 2 transcripts or shorter	shorter	
MPN345	+	412681		412681	412681	413301			
MPN346	+	413310	413750	413431	413431	413778			yes
MPN347	+			413985	413985	415115			
MPN347a	+	415550			415697	416032	putative NEW	NEW	
MPN348	+	415140		416070	416070	416564	putative NEW		yes
MPN349	+		417800	416554	416554	417399			
MPN350	-	418115	417190	418110	418110	417391			
MPN351	-		418160	418760	418757	418116			
MPN352	-	420268	418740	420240	420240	418741			
MPN353	-		420725	422160	422155	420293			
MPN354	-	423493	422150	423490	423488	422139			
MPN355	-	424240	423575	424240	424238	423510			
MPN356	-	425460	424230	425550	425440	424241	putative shorter	shorter	yes
MPN357	-			427511	427511	425535			
MPN358	+	427670	429300	427679	427679	429283			
MPN359	+		430335	429513	429513	430289			
MPN360	+	430310	430600	430334	430334	430627			
MPN361	+		431600	430628	430628	431707			
MPN362	+		433050	431707	431707	433068			
MPN363	+	433180	433540	433200	433200	433508			
MPN364	+			433547	433547	435580			
MPN365	+	435610		435924	435924	436730			yes
MPN366	+			436814	436814	437632			
MPN367	+			437563	437563	438531			
MPN367a	+	438480			438525	438827	alternative TSS → NEW	NEW	
MPN368	+	439370		439220	439388	439762	putative shorter	shorter	yes
MPN369	-	440670	439950	440655	440655	439894			
MPN370	+		442600	441296	441296	443509			
MPN371	-	444266	444085	444190	444187	443552			yes
MPN372	+	444330	446500	444341	444341	446116			
MPN373	-	446606	446000	446740	446606	446127	putative shorter	shorter	yes
MPN374	-	448632		448459	448459	447770			yes
MPN375	-	449345	448600	449194	449194	448805			yes
MPN376	-	452995	449500	452995	452995	449573			
MPN377	+	453392	453700	453422	453422	453646			
MPN378	+			453650	453650	456268			
MPN379	+	456054	457035	456272	456272	457147	putative NEW		yes
MPN380	+	457040	457900	457147	457147	457980			yes
MPN381	+	458048	459000	458126	458126	459004			yes
MPN382	+		459500	458998	458998	459600			
MPN383	-		459450	460378	460378	459530			
MPN384	-			462762	462762	460381			
MPN385	-			463107	463107	462763			
MPN386	-	463855	463100	463831	463831	463142			
MPN387	-		464000	465055	465055	463979			
MPN388	-			465434	465449	465048	putative longer	longer	yes
MPN389	-		465360	466446	466446	465427			
MPN390	-			467846	467846	466473			
MPN391	-	469068	467900	469055	469055	467847			
MPN392	-		469365	470357	470357	469374			
MPN393	-			471455	471455	470379			
MPN393a	-	472340			472203	472030	alternative TSS → NEW	NEW	
MPN394	-	472927		472915	472915	471476			
MPN395	-	473530	473000	473516	473408	472983	putative shorter	shorter	
MPN396	-	476487	473600	476466	476466	473551			
MPN397	+	476485/ 476500	478620	476470	476509	478671	alternative TSS → 2 transcripts or shorter	shorter	
MPN398	-		478450	479324	479324	478668			
MPN399	-		479250	480187	480187	479324			
MPN400	-	482110	480528	481997	481997	480249			yes
MPN401	-	483269	482800	483247	483247	482765			
MPN402	+		484950	483529	483529	484980			
MPN403	+			485125	485125	485493			
MPN404	+		486450	485474	485474	486514			
MPN405	+		487120	486521	486521	487114			
MPN406	+	487135	487390	487149	487149	487403			
MPN407	-	490223	487500	490147	490147	487508			yes
MPN408	+		492650	490375	490375	492657			
MPN409	+		494300	492730	492730	494331			
MPN410	+	494680 494860	495150	494694	494694/ 494862	495140	alternative TSS → 2 transcripts		yes

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN411	-	496060		496040	496040	495282			
MPN412	+	496525		496634	496634	497470			yes
MPN412a	+	497073			497073	497474	alternative TSS → NEW	NEW	
MPN413	+			497547	497547	497966			
MPN414	+		499550	498076	498076	499557			
MPN415	+	499800	501050	499808	499808	500950			
MPN416	+			500943	500943	501677			
MPN417	+		503250	501670	501670	503298			
MPN418	-		503150	503721	503721	503299			
MPN419	-		503925	506423	506423	503721			
MPN420	-		506450	507200	507200	506475			
MPN421	-	508682/ 508642		508620	508620	507193	alternative TSS → 2 transcripts		yes
MPN422	-		508682	509863	509863	508751			
MPN423	-	510255		510253	510253	509864			
MPN424	-		510255	510561	510561	510253			
MPN425	-			511588	511588	510542			
MPN426	-	514575		514542	514542	511594			
MPN427	+	514480/ 514590	515700	514727	514727	515599	alternative TSS → 2 transcripts		yes
MPN428	-		515600	516567	516567	515605			
MPN429	-			517834	517834	516605			
MPN430	-	518890/ 518862	517780	518850	518850	517837	alternative TSS → 2 transcripts		
MPN431	-		518900	519834	519834	518881			
MPN432	-	521685	519750	520972	520972	519824	putative NEW		yes
MPN433	-	521685	520800	521679	521679	520867			
MPN434	-	523635	521825	523624	523624	521837			
MPN435	-		523635	524867	524867	523680			
MPN436	-	528761/ 527402		528611	528611	524877	alternative TSS → 2 transcripts		yes
MPN437	-			530638	530638	528920			
MPN438	-			531893	531893	530856			
MPN439	-	532715		532662	532662	531949			yes
MPN440	-			534998	534998	532818			
MPN441	+	535050		535468	535468	535776	putative longer or NEW		yes
MPN442	-		535900	536541	536541	536089			
MPN443	-			537755	537755	536526			
MPN444	-	541760	537750	541739	541739	537762			
MPN445	+	541770	542560	541734	541794	542603	putative shorter	shorter	
MPN446	-	543283	542650	543281	543281	542664			
MPN447	-	546393		546360	546360	543304			
MPN448	-		546400	547281	547281	546424			
MPN449	-		547100	548678	548678	547332			
MPN450	-	549604		549630	549589	548690	putative shorter	shorter	
MPN451	-		549590	550622	550622	549513			
MPN452	-		550750	552680	552680	550662			
MPN453	-		552750	553518	553518	552694			
MPN454	-	554180/ 554138	553600	554120	554120	553539	alternative TSS → 2 transcripts		yes
MPN455	-	555085	554258	555072	555072	554209			
MPN456	-	558430	555400	558415	558415	555398			
MPN457	-			559787	559787	558798			
MPN458	-		559550	560202	560202	559729			
MPN459	-	561945	560100	561898	561898	560123			yes
MPN460	-	564365	562650	564355	564355	562658			
MPN461	+	564380		564390	564390	565085			
MPN462	+			565237	565237	566049			
MPN463	+			566124	566124	566597			
MPN464	+		568200	566891	566891	568228			
MPN464a	+	568540/ 568540	569190	568625/ 568793	568625/ 568793	569155	putative NEW (with alternative TSS)	NEW	
MPN465	-			569244	569244	568645			
MPN466	-			569710	569710	569288			
MPN467	-	570060		570045	570045	569740			
MPN468	+			570782	570782	571408			
MPN469	-		572300	572985	573063	572254	putative longer	longer	yes
MPN470	-		572990	574090	574090	573026			
MPN471	-	574255		574251	574251	574090			
MPN472	-	575170	574300	575151	575151	574270			
MPN473	-		575200	575963	575963	575157			
MPN474	-	577082	575900	579050	576942	575950	putative shorter	shorter	
MPN474a	-	579392			579389	579105	putative NEW	NEW	
MPN475	-		579400	580757	580757	579408			
MPN476	-		580320	581412	581412	580759			
MPN477	-	582398		582068	582068	581439	putative NEW		yes
MPN478	-		582100	582778	582778	582071			
MPN479	-	583393	582670	583375	583375	582782			
MPN480	-		583500	585919	585919	583403			

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN481	-			586484	586484	585903			
MPN482	-	586885/ 586910	586450	586686	586713	586492	alternative TSS → longer	longer	yes
MPN483	+	586760	587800	586772	586772	587797			
MPN484	-	588658	588300	588613	588613	588302			yes
MPN485	-	590042	590920	589980	589980	589030	putative longer		yes
MPN486	+			589922	589922	590365			
MPN487	+	590500		590569	590569	591795			yes
MPN488	+		592200	591798	591798	592220			
MPN488a	-	592578			592479	592201	alternative TSS → NEW	NEW	
MPN489	-	596335	592000	596300	596300	592398			
MPN490	-	597842/ 597790	596820	597826	597826	596816	alternative TSS → 2 transcripts		
MPN491	+	597836/ 597892	599400	597900	597900	599324	alternative TSS → 2 transcripts		yes
MPN492	-		599500	600380	600380	599463			
MPN493	-			600972	600972	600316			
MPN494	-		601050	601453	601453	600974			
MPN495	-	601882/ 601755		601742	601742	601455	alternative TSS → 2 transcripts		yes
MPN496	-		601885	603727	603727	601745			
MPN497	-	604801	604200	604586	604586	604155	putative NEW		yes
MPN498	+	604897		604897	604897	605625			
MPN499	+			605579	605579	606070			
MPN500	+			606218	606218	607801			
MPN501	+	608320		608167	608335	608757	putative shorter	shorter	yes
MPN502	+			609202	609202	610470			
MPN503	+			611025	611025	612170			
MPN504	+	612880/ 613000	613200	612740	612887	613120	alternative TSS → 2 transcripts or shorter	shorter	
MPN505	-	613960	612800	613949	613949	613188			
MPN506	+	614842		614865	614865	617246			
MPN506a	-	616120			615980	615744	alternative TSS → NEW	NEW	
MPN507	+	617350	617600	617366	617366	618457			
MPN508	-	619746/ 619782	618500	620010	619735	618485	alternative TSS → 2 transcripts or shorter	shorter	
MPN509	-	622592	620570	621844	621844	620561	putative longer or NEW		yes
MPN510	-		623450	624250	624250	622874			
MPN511	-		624778	625644	625644	624862			
MPN512	-	626858	625710	626139	626139	625675	putative longer or NEW		yes
MPN513	-			627369	627369	626917			
MPN514	-	628288	627920	628248	628248	627886			
MPN515	-			632282	632282	628410			
MPN516	-			636461	636461	632286			
MPN517	-	637070	636580	637054	637054	636554			
MPN518	-	639210	638170	639193	639193	638147			
MPN519	-		639218	640014	640014	639196			
MPN520	-		639950	642717	642717	640132			
MPN520a	-	640400			640304	640017	alternative TSS → NEW	NEW	
MPN521	+	642750		642760	642760	643260			
MPN522	+		643900	643251	643251	643883			
MPN523	-		644068	644955	644955	644038			
MPN524	-	645912	645520	646050	645904	645545	putative shorter	shorter	yes
MPN525	+	646630		646644	646644	647885	putative NEW		yes
MPN526	+		648900	647866	647866	648852			
MPN527	-	649819	649000	649784	649784	649107			
MPN528	+	649838	650450	649839	649839	650393			
MPN528a	+			650401	650402	650920			
MPN529	-	651315	650940	651246	651246	650917			yes
MPN530	+	651338	652000	651395	651395	651805			yes
MPN531	-	653975	651900	653949	653949	651802			
MPN532	-		654150	654973	654973	654125			
MPN533	-	656172	654750	656153	656153	654981			
MPN534	+			656494	656494	656916			
MPN535	+			656922	656922	657542			
MPN536	+		658400	657517	657517	658440			
MPN537	-	659740/ 659668	658470	659664	659664	658426	alternative TSS → 2 transcripts		yes
MPN538	+	659695		659782	659782	660267	putative NEW		yes
MPN539	+			660305	660305	660673			
MPN540	+		660825	660680	660680	660853			
MPN541	-	661175/ 661155	660838	661126	661126	660863	alternative TSS → 2 transcripts or NEW		yes
MPN542	+	661170	661950	661187	661187	661843			
MPN543	+			661848	661848	662783			
MPN544	+	662582/ 662582	664700	662783	662783	664777	alternative TSS → 2 transcripts		yes
MPN545	-		664780	665609	665609	664761			
MPN546	-	666592	665650	666585	666585	665599			

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN547	-	668307/ 667096	666630	668261	668261	666585	alternative TSS → 2 transcripts		yes
MPN548	-		668420	669296	669296	668316			
MPN549	-	670442	669750	670260	670260	669283			yes
MPN550	-		670575	671416	671416	670253			
MPN551	-			672248	672248	671403			
MPN552	-			673064	673064	672255			
MPN553	-		673000	674751	674751	673057			
MPN554	-			675065	675065	674751			
MPN555	-	675658	675000	675648	675648	675067			
MPN556	-			677304	677304	675691			
MPN557	+	677330		677332	677332	679170			
MPN558	+		680000	679163	679163	679738			
MPN559	+		681000	679692	679692	680399			
MPN560	-	681802	680482	681793	681793	680477			
MPN561	+		682500	681847	681847	682488			
MPN562	-		682450	683231	683231	682485			
MPN563	-		684000	684525	684525	683224			
MPN564	-	685816	684800	685784	685784	684729			
MPN565	-		685822	686456	686456	685998			
MPN566	-		686620	687339	687339	686626			
MPN567	-	690500/ 687703	687230	690453	690453	687343	alternative TSS → 2 transcripts		yes
MPN568	-		690600	691428	691428	690553			
MPN569	-			691742	691742	691416			
MPN570	-			692253	692253	691864			
MPN571	-			694237	694237	692255			
MPN572	-		694340	695678	695678	694341			
MPN573	-			697323	697323	695692			
MPN574	-	697680		697676	697676	697326			
MPN575	-	698225	697900	698250	698193	697864	putative shorter	shorter	
MPN576	-	699913	698700	699905	699905	698685			
MPN577	-	701322	700000	701107	701107	700067	putative longer		yes
MPN578	+			701545	701545	701847			
MPN578a	+	701700			701714	702049	alternative TSS → NEW	NEW	
MPN579	-	702258	701790	702192	702192	701863			yes
MPN580	-		702400	702756	702756	702334			
MPN581	-			703621	703621	702824			
MPN582	-	705130	703742	705116	705116	703797			
MPN582a	-	705530			705505	705191	putative NEW	NEW	
MPN583	-	706680		706231	706231	705554	putative NEW		yes
MPN584	-		706680	707142	707142	706735			
MPN585	-	708122		708112	708112	707204			
MPN586	-			709229	709229	708186			
MPN587	-		708920	709557	709557	709105			
MPN588	-			711210	711210	709615			
MPN589	-		712375	711896	711896	711423			
MPN590	-			712690	712690	712037			
MPN591	-		712800	713908	713908	712847			
MPN592	-	716007		715996	715996	714431			
MPN593	-	716500	716070	716465	716465	716097			
MPN594	-	717059		717059	717059	716691			
MPN595	-	717470/ 717448		717435	717435	716977	alternative TSS → 2 transcripts		
MPN596	-	718749/ 718338		719144	718649	717435	alternative TSS → 2 transcripts or shorter	shorter	
MPN597	-		719200	719545	719545	719144			
MPN598	-			720975	720975	719548			
MPN599	-			721814	721814	720975			
MPN600	-			723370	723370	721814			
MPN601	-			723909	723909	723373			
MPN602	-			724525	724525	723902			
MPN603	-	724942		724845	724845	724528			yes
MPN604	-			725729	725729	724848			
MPN605	-			726168	726168	725695			
MPN606	-	727568	726180	727544	727544	726174			
MPN607	+	727575/ 727602	728050	727599	727599	728072	alternative TSS → 2 transcripts or NEW		yes
MPN608	-		727900	728746	728746	728069			
MPN609	-			729738	729738	728749			
MPN610	-			731678	731678	729723			
MPN611	-	732812	731700	732840	732796	731678	putative shorter	shorter	
MPN612	-			736070	736070	733077			
MPN613	-			737219	737219	736185			
MPN614	-			738159	738159	737155			
MPN615	-			738994	738994	738245			
MPN616	-		739400	739822	739822	739424			
MPN617	-	740287	739850	740268	740268	739828			
MPN618	-		740287	742344	742344	740299			

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPN619	-	745407	742360	745199	745199	742353	putative NEW		yes
MPN620	-	747710		747696	747696	745177			
MPN621	-		747500	749427	749427	747742			
MPN622	-	749708	749405	749692	749692	749432			
MPN623	+	749726/ 749782		749850	749850	751202	alternative TSS → 2 transcripts		yes
MPN624	+	751215	751400	751224	751224	751421			
MPN625	+	751458	751895	751465	751465	751890			
MPN626	+	752070	752360	752025	752121	752543	putative shorter	shorter	yes
MPN627	-		752420	754258	754258	752540			
MPN628	-			755789	755789	754263			
MPN629	-	756522		756516	756516	755782			
MPN630	-	757832	756600	757783	757783	756569			yes
MPN631	+	757845	758820	757870	757870	758766			
MPN632	+		759450	758768	758747	759475	putative longer	longer	yes
MPN633	+	759515	760300	759578	759578	760321			yes
MPN634	+		761115	760403	760403	760948			
MPN635	+		762620	761504	761504	762562			
MPN636	+	762620	763250	762639	762639	763193			
MPN637	+		764400	763193	763193	764380			
MPN638	+	764400	765600	764400	764400	765527			
MPN639	-	766400	765522	766387	766387	765524			
MPN640	-		766400	767297	767297	766395			
MPN641	-			768127	768127	767297			
MPN642	-		768000	768969	768969	768130			
MPN643	-			769877	769877	768969			
MPN644	-			770649	770649	769798			
MPN645	-			771503	771503	770652			
MPN646	-		772400	772336	772336	771503			
MPN647	-			773231	773231	772359			
MPN648	-	773840		773752	773752	773342			yes
MPN649	-			775049	775049	774639			
MPN650	-	775425	774920	775339	775339	775034			yes
MPN651	+	776242	776365	776337	776337	777476	putative longer		yes
MPN652	+	777470	778510	777463	777481	778557	putative shorter	shorter	
MPN653	+		779025	778538	778538	778969			
MPN654	-	779731	779240	779730	779731	779342			
MPN655	+	780185	780600	780008	780338	780622	putative shorter	shorter	yes
MPN656	-		781243	781910	781910	781095			
MPN657	-		781850	783084	783084	781879			
MPN658	-		783000	783459	783459	783100			
MPN659	-			784070	784138	783443	putative longer	longer	yes
MPN660	-	784455	784000	784407	784407	784141			yes
MPN661	-		784500	786138	786138	784489			
MPN662	-			786594	786594	786139			
MPN663	-		786672	787262	787262	786558			
MPN664	-			787980	788141	787269	putative longer	longer	yes
MPN665	-	789400/ 789350	788150	789325	789325	788141	alternative TSS → 2 transcripts		yes
MPN666	+	789390	790180	789410	789410	790165			
MPN667	+		791000	790153	790153	791028			
MPN668	-	791462	790850	791447	791447	791025			
MPN669	+	791370/ 791465	792650	791473	791473	792672	alternative TSS → 2 transcripts		yes
MPN670	-	793766	792720	793749	793749	792712			
MPN670a	-	792860			792833	792213	alternative TSS → NEW	NEW	
MPN671	-	796468	794180	796431	796431	794302			
MPN671a	-	794340			794318	793737	alternative TSS → NEW	NEW	
MPN672	-		796400	797140	797181	796612	putative longer	longer	yes
MPN673	-	797662		797648	797648	797139			
MPN674	+	797805	798800	797819	797819	798757			
MPN675	-	799516	799160	799467	799467	799162			yes
MPN676	+	799780		799858	799858	800178	putative longer		yes
MPN677	-		800675	802012	802012	800735			
MPN678	-			803447	803447	801993			
MPN679	-		803732	804225	804225	803434			
MPN680	-			805375	805375	804218			
MPN681	-			805712	805712	805356			
MPN682	-	805863	805680	805848	805848	805702			
MPN683	-		805863	806890	806890	805871			
MPN684	-			812540	812540	806892			
MPN685	-		812200	813400	813400	812546			
MPN686	-		813520	814787	814787	813468			
MPN687	-			815539	815539	814787			
MPN688	-	816360	815200	816338	816338	815526			
MPNr01	+	118215/ 118248	119800		118312	119825	alternative TSS → 2 transcripts		
MPNr02	+	120055	119818						
MPNr03	+		123130						

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPNs001	-	463950	463875				non-coding RNA		
MPNs002	-	469317	463875				non-coding RNA		
MPNs003	-	571800	571478				putative NEW (no stop)		
MPNs004	+	571820	572220				non-coding RNA		
MPNs005	+	13390	14200				putative NEW		
MPNs006	+	15895	16000				putative NEW (no stop)		
MPNs007	+	57400	57920				putative NEW (no stop)		
MPNs008	+	61723/ 61723	62040				alternative TSS → 2 transcripts		
MPNs010	+	88180	88340				non-coding RNA		
MPNs011	+	88738	88840				putative NEW (no stop)		
MPNs012	+	91205	92600				putative NEW		
MPNs013	+	96173	96450				putative NEW		
MPNs014	+	99078	99175				putative NEW (no stop)		
MPNs015	+	108600	108680				non-coding RNA		
MPNs016	+	109300	109420				putative NEW (no stop)		
MPNs018	+	161700	162030				putative NEW (no stop)		
MPNs019	+	167025	167500				putative NEW		
MPNs021	+	172232	172400				putative NEW (no stop)		
MPNs022	+	229682	229950				putative NEW (no stop)		
MPNs024	+	238805	239015				putative NEW		
MPNs025	+	239373	239800				putative NEW (no stop)		
MPNs026	+	240997	241150				putative NEW (no stop)		
MPNs027	+	253203	253483				non-coding RNA		
MPNs028	+	265330	265450				putative NEW (no stop)		
MPNs029	+	265662	265770				non-coding RNA		
MPNs030	+	276675	276775				putative NEW (no stop)		
MPNs031	+	284500	284700				putative NEW (no stop)		
MPNs032	+	323470	323700				putative NEW		
MPNs033	+	336518	336700				putative NEW (no stop)		
MPNs034	+	348825	350250				non-coding RNA		
MPNs036	+	398612	398775				non-coding RNA		
MPNs037	+	411950	412210				putative NEW		
MPNs038	+	432403	432680				putative NEW (no stop)		
MPNs039	+	449750	450170				putative NEW		
MPNs040	+	450690	450750				non-coding RNA		
MPNs041	+	451000	452890				putative NEW		
MPNs042	+	457035	457200				putative NEW (no stop)		
MPNs043	+	459957	460408				putative NEW		
MPNs044	+	460418	460520				non-coding RNA		
MPNs045	+	460835	461290				putative NEW		
MPNs046	+	461290	461400				non-coding RNA		
MPNs047	+	468454	469000				putative NEW		
MPNs048	+	470460	470550				putative NEW (no stop)		
MPNs049	+	471462	471580				non-coding RNA		
MPNs050	+	480900	481180				non-coding RNA		
MPNs051	+	482020	482100				putative NEW (no stop)		
MPNs052	+	482240	482350				non-coding RNA		
MPNs053	+	483320	483640				putative NEW (no stop)		
MPNs054	+	485160	485380				putative NEW (no stop)		
MPNs055	+	490325	490650				putative NEW (no stop)		
MPNs056	+	505640	505890				non-coding RNA		
MPNs057	+	507805	508120				putative NEW		
MPNs058	+	509125	509220				non-coding RNA		
MPNs059	+	509880	509940				non-coding RNA		
MPNs060	+	511840	511930				non-coding RNA		
MPNs061	+	517155	517200				non-coding RNA		
MPNs062	+	518870	519080				non-coding RNA		
MPNs063	+	530410	530570				non-coding RNA		
MPNs064	+	531550	531700				putative NEW (no stop)		
MPNs065	+	532795	532960				non-coding RNA		
MPNs066	+	533810	533890				non-coding RNA		
MPNs067	+	538250	538420				non-coding RNA		
MPNs068	+	540657	540800				putative NEW (no stop)		
MPNs069	+	541487	541700				putative NEW (no stop)		
MPNs070	+	543310	543450				putative NEW (no stop)		
MPNs071	+	544638	544700				non-coding RNA		
MPNs072	+	544790	544900				non-coding RNA		
MPNs073	+	545910	546065				non-coding RNA		
MPNs074	+	547620	547800				non-coding RNA		
MPNs075	+	549062	549270				putative NEW (no stop)		
MPNs076	+	550540	550750				putative NEW (no stop)		
MPNs077	+	556300	556430				putative NEW (no stop)		
MPNs078	+	556612	556885				non-coding RNA		
MPNs079	+	557780	557860				non-coding RNA		
MPNs080	+	560140	560300				non-coding RNA		
MPNs081	+	562660	562740				non-coding RNA		
MPNs083	+	569330	569580				putative NEW		
MPNs084	+	570160	570310				putative NEW (no stop)		

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPNs085	+	571560	571650				non-coding RNA		
MPNs086	+	572810	573000				putative NEW (no stop)		
MPNs087	+	576650	576800				putative NEW (no stop)		
MPNs088	+	596490	596720				non-coding RNA		
MPNs089	+	596760	597600				putative NEW		
MPNs090	+	599595	599900				putative NEW		
MPNs091	+	618643	618900				non-coding RNA		
MPNs092	+	623980	624210				non-coding RNA		
MPNs093	+	637870	638138				putative NEW (no stop)		
MPNs094	+	641225	641380				non-coding RNA		
MPNs095	+	645860	645970				non-coding RNA		
MPNs096	+	656095	656460				putative NEW		
MPNs097	+	664700	664820				non-coding RNA		
MPNs098	+	665160	665220				putative NEW (no stop)		
MPNs099	+	681820	682030				putative NEW (no stop)		
MPNs100	+	683647	684500				putative NEW		
MPNs101	+	684870	685090				non-coding RNA		
MPNs102	+	687560	687660				putative NEW (no stop)		
MPNs103	+	690525	690650				non-coding RNA		
MPNs104	+	702946	703380				non-coding RNA		
MPNs105	+	703758	703950				non-coding RNA		
MPNs106	+	708540	709300				putative NEW		
MPNs107	+	711420	711620				non-coding RNA		
MPNs108	+	716570	716800				non-coding RNA		
MPNs109	+	733795	734120				putative NEW (no stop)		
MPNs110	+	739840	740075				putative NEW (no stop)		
MPNs111	+	752360	753150				putative NEW		
MPNs112	+	757437	757550				non-coding RNA		
MPNs113	+	761115	761400				non-coding RNA		
MPNs114	+	770075	770250				non-coding RNA		
MPNs115	+	776010	776360				putative NEW		
MPNs116	+	778610	778730				putative NEW (no stop)		
MPNs117	+	780180	780310				putative NEW (no stop)		
MPNs118	+	780800	781560				putative NEW		
MPNs119	+	782321	782410				non-coding RNA		
MPNs120	+	783890	784000				non-coding RNA		
MPNs121	+	785298	785420				non-coding RNA		
MPNs122	+	786500	786850				putative NEW		
MPNs123	+	787600	787700				non-coding RNA		
MPNs124	+	797290	797800				putative NEW		
MPNs125	+	800647	800780				putative NEW (no stop)		
MPNs126	+	807600	807780				putative NEW (no stop)		
MPNs127	+	813618	813700				non-coding RNA		
MPNs128	+	701170	701260				non-coding RNA		
MPNs200	-	180	45				putative NEW (no stop)		
MPNs201	-	575	395				non-coding RNA		
MPNs202	-	15690	15150				non-coding RNA		
MPNs203	-	16038	15922				non-coding RNA		
MPNs204	-	17493	17300				putative NEW (no stop)		
MPNs205	-	19340	19295				non-coding RNA		
MPNs206	-	19408	19370				non-coding RNA		
MPNs208	-	23165	22750				putative NEW		
MPNs209	-	27205	26525				putative NEW		
MPNs210	-	28960	28505				putative NEW		
MPNs211	-	30400	30025				putative NEW (no stop)		
MPNs212	-	30725	30670				putative NEW (no stop)		
MPNs213	-	33795	33505				non-coding RNA		
MPNs214	-	41195/ 41148	41040				alternative TSS → 2 transcripts		
MPNs215	-	45720	45100				non-coding RNA		
MPNs217	-	48820	48720				non-coding RNA		
MPNs218	-	55493	55375				putative NEW (no stop)		
MPNs219	-	56440	56260				non-coding RNA		
MPNs220	-	57462	57240				non-coding RNA		
MPNs221	-	65550	64890				putative NEW (no stop)		
MPNs222	-	66790	66600				non-coding RNA		
MPNs223	-	67905	67200				putative NEW		
MPNs224	-	68400	68220				non-coding RNA		
MPNs225	-	68970	68820				non-coding RNA		
MPNs226	-	82028	81720				putative NEW		
MPNs227	-	69100	69038				non-coding RNA		
MPNs228	-	82625	82375				non-coding RNA		
MPNs229	-	69293	69150				putative NEW (no stop)		
MPNs230	-	72525	71625				putative NEW (from MPNs)		
MPNs231	-	93233	93185				non-coding RNA		
MPNs232	-	73020	72825				non-coding RNA		
MPNs233	-	96150	96005				non-coding RNA		
MPNs234	-	106760	106600				putative NEW (no stop)		
MPNs235	-	108240	108030				non-coding RNA		

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPNs236	-	110280	110080				non-coding RNA		
MPNs237	-	111750/ 111710	111630				non-coding RNA		
MPNs238	-	113410	113320				non-coding RNA		
MPNs239	-	114880	114730				putative NEW (no stop)		
MPNs240	-	118107	117280				putative NEW		
MPNs241	-	119835	119380				putative NEW (no stop)		
MPNs242	-	122760	122075				putative NEW		
MPNs243	-	123910	123790				putative NEW (no stop)		
MPNs244	-	124845	124450				putative NEW (no stop)		
MPNs245	-	125320	125250				non-coding RNA		
MPNs246	-	126560	126170				putative NEW		
MPNs247	-	127960	127270				putative NEW		
MPNs249	-	132940	132575				putative NEW (no stop)		
MPNs250	-	136742	136550				putative NEW (no stop)		
MPNs251	-	141660	141600				non-coding RNA		
MPNs254	-	142560	142350				putative NEW (no stop)		
MPNs255	-	145910	145530				putative NEW (no stop)		
MPNs256	-	146935	146810				non-coding RNA		
MPNs257	-	149842	149660				putative NEW (no stop)		
MPNs259	-	154492	154380				non-coding RNA		
MPNs261	-	162090	161850				non-coding RNA		
MPNs262	-	163560	163160				putative NEW (no stop)		
MPNs263	-	164360	164240				putative NEW (no stop)		
MPNs264	-	167200	167080				putative NEW (no stop)		
MPNs265	-	168637	168550				putative NEW (no stop)		
MPNs267	-	170685	170350				putative NEW		
MPNs268	-	173242	173050				non-coding RNA		
MPNs269	-	190268	190000				putative NEW (no stop)		
MPNs270	-	195320	195220				putative NEW (no stop)		
MPNs271	-	197915	197880				putative NEW (no stop)		
MPNs273	-	207660	207540				non-coding RNA		
MPNs274	-	213968	213240				putative NEW		
MPNs275	-	218972	218400				putative NEW		
MPNs276	-	223180	223140				putative NEW (no stop)		
MPNs278	-	224865	224620				non-coding RNA		
MPNs279	-	229660	229520				non-coding RNA		
MPNs280	-	234380	234050				putative NEW		
MPNs281	-	243000	242800				non-coding RNA		
MPNs282	-	251685	251510				non-coding RNA		
MPNs283	-	260383	260230				non-coding RNA		
MPNs284	-	271983	271250				putative NEW		
MPNs285	-	277120	277000				non-coding RNA		
MPNs286	-	291345	291190				non-coding RNA		
MPNs287	-	292580	292375				non-coding RNA		
MPNs288	-	296050	295940				non-coding RNA		
MPNs289	-	298665	298560				non-coding RNA		
MPNs290	-	301008	300900				putative NEW (no stop)		
MPNs291	-	302938	302878				putative NEW (no stop)		
MPNs292	-	305025	304900				non-coding RNA		
MPNs293	-	306882	306500				putative NEW		
MPNs294	-	309578	309440				putative NEW (no stop)		
MPNs295	-	313570	313050				putative NEW		
MPNs296	-	313855	313700				putative NEW (no stop)		
MPNs297	-	320762	320675				non-coding RNA		
MPNs298	-	323140	322670				non-coding RNA		
MPNs299	-	323973	323882				putative NEW (no stop)		
MPNs300	-	328352/ 328238	328238				alternative TSS → 2 transcripts		
MPNs301	-	332805	332320				putative NEW		
MPNs302	-	337968	337775				putative NEW (no stop)		
MPNs303	-	340880	340600				putative NEW (no stop)		
MPNs304	-	355225	355142				putative NEW (no stop)		
MPNs305	-	357220	357070				non-coding RNA		
MPNs306	-	360400	360250				putative NEW (no stop)		
MPNs307	-	361440	361318				putative NEW (no stop)		
MPNs308	-	361730	361600				putative NEW (no stop)		
MPNs309	-	363205	362675				putative NEW		
MPNs310	-	364705	364340				putative NEW (no stop)		
MPNs311	-	365492	365280				non-coding RNA		
MPNs312	-	369520	369280				putative NEW		
MPNs313	-	373355	373225				non-coding RNA		
MPNs314	-	378383	378210				non-coding RNA		
MPNs315	-	385795/ 385682	385155				alternative TSS → 2 transcripts		
MPNs316	-	388900	388250				putative NEW		
MPNs317	-	395860	395630				putative NEW (no stop)		
MPNs318	-	401278	401120				non-coding RNA		
MPNs319	-	403180	403038				non-coding RNA		

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPNs320	-	406625	406350				non-coding RNA		
MPNs321	-	407738	407622				non-coding RNA		
MPNs322	-	415460	415300				non-coding RNA		
MPNs323	-	409390	409295				non-coding RNA		
MPNs324	-	416880	416683				non-coding RNA		
MPNs326	-	431605	431400				putative NEW (no stop)		
MPNs327	-	443075	441770				putative NEW		
MPNs328	-	447072	446962				putative NEW (no stop)		
MPNs329	-	457902	457600				putative NEW (no stop)		
MPNs330	-	473207	473000				putative NEW (no stop)		
MPNs331	-	477250	477180				non-coding RNA		
MPNs333	-	491444	491280				non-coding RNA		
MPNs334	-	491740	491520				non-coding RNA		
MPNs335	-	493524	492700				putative NEW		
MPNs336	-	499965	499850				putative NEW (no stop)		
MPNs337	-	501498	501300				non-coding RNA		
MPNs339	-	563350	562850				putative NEW		
MPNs340	-	591265	591120				putative NEW (no stop)		
MPNs341	-	591705	591500				non-coding RNA		
MPNs342	-	596738	596653				non-coding RNA		
MPNs343	-	607514	605500				putative NEW		
MPNs344	-	620510	620220				putative NEW (no stop)		
MPNs345	-	624778	624640				putative NEW (no stop)		
MPNs346	-	625128	624920				putative NEW (no stop)		
MPNs347	-	627540	627388				non-coding RNA		
MPNs348	-	636558	636420				putative NEW (no stop)		
MPNs349	-	642723	642575				putative NEW (no stop)		
MPNs350	-	643850	642900				putative NEW		
MPNs351	-	644000	643920				non-coding RNA		
MPNs352	-	651280	650925				putative NEW (no stop)		
MPNs353	-	657224	656675				putative NEW		
MPNs354	-	663380	662700				putative NEW		
MPNs355	-	664437	664200				non-coding RNA		
MPNs356	-	666596	666400				putative NEW (no stop)		
MPNs357	-	670575	670480				putative NEW (no stop)		
MPNs358	-	677306					non-coding RNA		
MPNs359	-	678640	678550				putative NEW (no stop)		
MPNs360	-	691910	691820				putative NEW (no stop)		
MPNs361	-	703742	703270				putative NEW (no stop)		
MPNs362	-	711220	710925				putative NEW (no stop)		
MPNs363	-	718750	718500				putative NEW (no stop)		
MPNs364	-	738930	738320				putative NEW (no stop)		
MPNs365	-	742355	741810				putative NEW (no stop)		
MPNs366	-	749770	749750				putative NEW (no stop)		
MPNs367	-	759226	758925				putative NEW		
MPNs368	-	765522	765395				non-coding RNA		
MPNs369	-	776443	776220				putative NEW		
MPNs370	-	778310	778160				non-coding RNA		
MPNs371	-	781243	781120				non-coding RNA		
MPNs372	-	786672	786330				putative NEW (no stop)		
MPNs373	-	789578	789540				putative NEW (no stop)		
MPNs374	-	791993	791943				non-coding RNA		
MPNs375	-	800080/ 800042	799800				alternative TSS → 2 transcripts		
MPNs376	-	800460	800400				non-coding RNA		
MPNs377	-	800535	800495				non-coding RNA		
MPNs378	-	803732	803500				putative NEW (no stop)		
MPNs379	-	813458	813000				putative NEW (no stop)		
MPNs380	-	813522	813485				non-coding RNA		
MPNs381	+	596	633				non-coding RNA		
MPNs382	+	134175	134430				putative NEW (no stop)		
MPNs383	-	180395	180030				putative NEW (no stop)		
MPNt01	-		19078		19149	19073			
MPNt02	-	19256			19231	19154			
MPNt03	-	93150	93050		93140	93051			
MPNt04	-	156432	156350		156430	156350			
MPNt05	+	358232			358243	358320			
MPNt06	+				358372	358449			
MPNt07	+				358452	358529			
MPNt08	+				358539	358615			
MPNt09	+				358655	358746			
MPNt10	+				358758	358835			
MPNt11	+		358900		358836	358910			
MPNt12	+		358978		358911	358987			
MPNt13	+	373075	373158		373083	373160			
MPNt14	-	427650	427580		427650	427580			
MPNt15	+	429258/ 429280	429359		429284	429359	alternative TSS → 2 transcripts		
MPNt16	+	429420	429500		429432	429506			

Continued on next page

Table B.1 – continued from previous page

ORF (re-anno-tated)	str.	TSS1	TTS	TSC NCBI	TSC re-anno-tated	stop	prediction from transcripts or theoretical transcriptome	re-anno-tation	5'-UTR >40bp
MPNt17	-	453340	453260		453336	453260			
MPNt18	-		482280		482360	482290			
MPNt19	-				482450	482360			
MPNt20	-				482520	482440			
MPNt21	-				482640	482560			
MPNt22	-	482745			482730	482640			
MPNt23	+	487392	487454		487394	487469			
MPNt24	+	555135			555142	555233			
MPNt25	+		555330		555234	555323			
MPNt26	+	558618	558700		558635	558724			
MPNt27	-		562060		562140	562050			
MPNt28	-				562220	562140			
MPNt29	-				562310	562230			
MPNt30	-				562390	562320			
MPNt31	-				562470	562400			
MPNt32	-				562470	562390			
MPNt33	-	562640			562630	562550			
MPNt34	+	643898	643995		643913	643989			
MPNt35	-	645078			645060	644980			
MPNt36	+	648900			648912	648993			
MPNt37	-	680482	680410		680480	680410			

Table B.1.: For each gene ID the strand (str.), a transcription start site (TSS1) and transcription termination site (TTS) if detected by Güell et al. [2009, 2011], the annotated transcription start codon (TSC) at NCBI, the re-annotated TSC, the stop codon, the prediction for re-annotation from the theoretical peptide library, the type of re-annotation if applied, and if the respective ORF contains a 5'-UTR larger than 40 base pairs.

Table B.2: Newly and Re-annotated ORFs

ORF	str.	TSC	stop	(re-)annotated	ClustalW results
MPN006	+	8549	9211	longer	longer in *M. pulmonis*, *H. influenzae*
MPN010	+	12392	12652	shorter	—
MPN013	+	15088	15765	shorter	—
MPN014	+	15939	16505	shorter	homolog in other species: not shorter
MPN035a	-	41998	41696	new ORF	HP (*Phytophthora sojae*)
MPN037a	-	46068/ 46011	45736	new ORF, two isoforms	MPN139
MPN038	-	46711	46442	shorter	shorter in *M. pneumoniae* FH
MPN047a	-	56759	56502	new ORF	pentapeptide repeat protein (*Vibrio parahaemolyticus* 16)
MPN048a	+	57029	57229	new ORF	transcriptional regulator, GntR (*B. cereus*)
MPN060a	+	77313	77594	new ORF	HP (*M. pneumoniae* FH)
MPN073	-	88091/87959	86925	two isoforms	shorter in *M. genitalium*, *M. pulmonis*
MPN077	-	92884	91199	longer	longer in *M. genitalium*
MPN091a	-	114043	113870	new ORF	—
MPN094	+	116455	116709	shorter	—
MPN100	+	129626	130009	shorter	—
MPN101a	+	131848	132513	new ORF	MPN149
MPN104a	+	135094	135360	new ORF	HP (*M. pneumoniae* FH)
MPN127	+	164628	165026	shorter	—
MPN127a	+	165867	166262	new ORF	Adhesin P1 (*M. pneumoniae*)
MPN130	+	169042/169210	169464	two isoforms	MPN130
MPN132a	+	171494	171805	new ORF	MPN170
MPN145	+	192349	192753	shorter	—
MPN148	+	195539	196297	longer	—
MPN152a	-	199929	199768	new ORF	NsdA (*Streptomyces qingfengmyceticus*)
MPN153a	-	205468	205250	new ORF	ApbE family lipoprotein (*Lysinibacillus fusiformis* ZC1)
MPN154a	-	206273	206031	new ORF	similar to 6 PDZ domain containing proteins (*Gallus gallus*)
MPN155a	+	207505	207717	new ORF	CHP (*M. pneumoniae* FH) MG477
MPN163	+	217150	217536	longer	longer in *M. genitalium*
MPN196	+	235805/236318	236536	two isoforms	MPN196
MPN198a	+	240189	240398	new ORF	—
MPN199a	+	241575	241853	new ORF	unknown protein (*M. genitalium*)
MPN207a	+	250000	250293	new ORF	CHP (*M. pneumoniae* FH)
MPN208a	-	253317	253117	new ORF	conserved hypothetical protein (*Ajellomyces capsulatus* H88)
MPN215	+	265865	267079	longer	longer in *M. genitalium*, *B. subtilis*, *Staphylococcus aureus*, *E. coli*
MPN216a	-	267844	267638	new ORF	predicted protein (*Hordeum vulgare*)
MPN245	-	297719	297138	shorter	shorter in other species
MPN246	+	297760	298329	shorter	shorter in other species

Continued on next page

Table B.2 – continued from previous page

ORF	str.	TSC	stop	(re-)annotated	ClustalW results
MPN283	+	336626	336826	shorter	—
MPN285	+	341207	341533	shorter	—
MPN307	+	361468/361783	362397	two isoforms	MPN307
MPN310	+	365468/370424	370924	two isoforms	other experimental evidences*
MPN320	+	380664	381527	shorter	homolog in other species: not shorter
MPN335a	-	397616	397269	new ORF	HP (*Coprobacillus*)
MPN344	+	411383	411736	shorter	—
MPN347a	+	415697	416032	new ORF	death-on-curing family protein (*M. pneumoniae* FH)
MPN356	-	425440	424241	shorter	shorter in other species
MPN367a	+	438615	438827	new ORF	MPN150
MPN368	+	439388	439762	shorter	—
MPN373	-	446606	446127	shorter	—
MPN388	-	465449	465048	longer	—
MPN393a	-	472203	472030	new ORF	—
MPN395	-	473408	472983	shorter	shorter in other species
MPN397	+	476509	478671	shorter	shorter in other species
MPN410	+	494694/494862	495140	two isoforms	—
MPN412a	+	497073	497474	new ORF	hypothetical protein VCJ_001109 (Vibrio)
MPN445	+	541794	542603	shorter	shorter in other species
MPN450	-	549589	548690	shorter	shorter in *M. genitalium*
MPN464a	+	568625/ 568793	569155	new ORF, altern. TSS & TSC	MPN100
MPN469	-	573063	572254	longer	longer in *M. genitalium*
MPN474	-	576942	575950	shorter	shorter in *M. genitalium*
MPN474a	-	579389	579105	new ORF	CHP (*M. pneumoniae* FH) MG328
MPN482	-	586713	586492	longer	longer in *M. genitalium*, *M. pulmonis*
MPN488a	-	592377	592201	new ORF	MG338
MPN501	+	608335	608757	shorter	—
MPN504	+	612887	613120	shorter	—
MPN506a	-	615980	615744	new ORF	2,4-dienoyl-CoA reductase (NADPH) precursor related protein (*Leeuwenhoekiella blandensis*)
MPN508	-	619735	618485	shorter	
MPN520a	-	640304	640017	new ORF	isoleucyl-tRNA synthetase (*M. pneumoniae* 309)
MPN524	-	645904	645545	shorter	—
MPN575	-	698193	697864	shorter	—
MPN578a	+	701714	702049	new ORF	cyclopropane fatty acid synthase B (*P. putida*)
MPN582a	-	705505	705191	new ORF	MPN584
MPN596	-	718649	717435	shorter	homolog in *M. genitalium*: not shorter
MPN611	-	732796	731678	shorter	shorter in *M. genitalium*
MPN626	+	752121	752543	shorter	homolog in *M. genitalium*: not shorter
MPN632	+	758747	759475	longer	longer in *M. genitalium*, *B. subtilis*
MPN652	+	777481	778557	shorter	shorter in *M. pulmonis*, *B. subtilis*, *E. coli*, *S. pneumoniae*
MPN655	+	780338	780622	shorter	shorter in *M. pneumoniae* FH and 309 strains
MPN659	-	784138	783443	longer	longer in *M. genitalium*, *M. pulmonis*, *B. subtilis*, extitC. crescentus, *E. coli*, *S. pneumoniae*, *H. influenzae*
MPN664	-	788141	787269	longer	longer in *M. genitalium*
MPN670a	-	792833	792213	new ORF	gamma-glutamyltransferase 7-like (*Acyrthosiphon pisum*)
MPN671a	-	794318	793737	new ORF	HP (*M. pneumoniae* FH)
MPN672	-	797181	796612	longer	longer in *M. pulmonis*

Table B.2.: ClustalW results for newly and re-annotated *M. pneumoniae* ORFs; 'altern.' abbreviates alternative.

Table B.3: Molecular Weight of *M. pneumoniae* Proteins

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN001	398.66	393.22	L	id00013	45	210	-
MPN002	13.01	12.78	L	id00032	32.5	210	45
MPN003	92.56	96.48	L	id00049	60	210	-
MPN004	95.03	93.07	L	id00081	70	210	-
MPN005	83.1	90.77	J	id00114	45	70	210
MPN006	26.39	25.22	F	id00133	22.5	-	-
MPN007	10.91	9.61	L	id00143	27.5	-	-
MPN008	8.3	7.62	J	id00153	45	-	-
MPN009	7.06	5.75	L	id00174	27.5	-	-
MPN010	0	0	N	-	-	-	-
MPN011	12.47	16.89	M	id00207	22.5	10	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN012	27.64	7.5	M	id00216	18.75	10	-
MPN013	12.07	13.11	N	id00227	16.25	32.5	-
MPN014	0	0	L	-	-	-	-
MPN015	26.85	16.33	J	id00254	32.5	45	-
MPN016	2.3	22.41	J	id00266	32.5	-	-
MPN017	15.91	10.2	H	id00275	27.5	-	-
MPN018	31.31	30.94	I	id00281	55	70	210
MPN019	24.25	31.1	I	id00306	55	210	70
MPN020	58.9	53.8	L	id00334	95	210	-
MPN021	81.6	95.96	O	id00375	45	-	-
MPN022	56.17	79.06	O	id00390	27.5	45	-
MPN023	15.23	16.71	J	id00406	45	-	-
MPN024	149.87	133.65	K	id00433	16.25	-	-
MPN025	611.7	556.28	G	id00443	27.5	45	-
MPN026	17.26	12.52	J	id00456	45	260	95
MPN027	12.86	9.18	J	id00493	27.5	-	-
MPN028	0	0	M	-	-	-	-
MPN029	211.63	261.13	J	id00513	27.5	-	-
MPN030	10.48	12.82	J	id00524	16.25	-	-
MPN031	13.66	11.69	A	id00534	22.5	-	-
MPN032	0	0	O	id00555	10	-	-
MPN033	26.3	26.93	F	id00563	22.5	-	-
MPN034	50.33	47.55	L	id00638	210	95	-
MPN035	0	0	M	-	-	-	-
MPN035a	-	-	-	-	-	-	-
MPN036	13.84	8.26	M	id00674	60	18.75	10
MPN037	0	0	M	-	-	-	-
MPN037a	-	-	-	id00715	10	-	-
MPN038	0	0	N	-	-	-	-
MPN039	0	0	M	-	-	-	-
MPN040	0	0	M	-	-	-	-
MPN041	0	0	M	-	-	-	-
MPN042	0	0	M	-	-	-	-
MPN043	46.86	66.38	G	id00796	18.75	-	-
MPN044	112.97	119.44	F	id00818	22.5	45	-
MPN045	54.1	52.5	J	id00819	45	70	210
MPN046	17.12	19.65	J	id00838	55	70	210
MPN047	10.3	8	H	id00859	45	-	-
MPN047a	-	-	-	id00873	12.5	-	-
MPN048	0	0	M	-	-	-	-
MPN048a	-	-	-	id00876	18.75	-	-
MPN049	0	0	M	-	-	-	-
MPN050	134.86	146.26	C	id00971	55	210	-
MPN051	188.63	161.29	C	id00989	45	70	210
MPN052	343.63	400.29	S	id00993	27.5	10	-
MPN053	472.99	603.33	G	id01022	10	18.75	-
MPN054	0	0	M	-	-	-	-
MPN055	60.72	50.85	E	id01069	60	210	-
MPN056	0	0	E	id01096	22.5	-	-
MPN057	0	0	E	id01109	22.5	-	-
MPN058	20.06	22.92	R	id01119	45	12.5	-
MPN059	27.51	25.79	O	id01142	32.5	45	-
MPN060	37.42	33.56	H	id01157	45	210	70
MPN060a	-	-	-	id01176	12.5	-	-
MPN061	101.01	94.8	U	id01193	45	70	210
MPN062	481.66	533.49	F	id01194	22.5	-	-
MPN063	230.35	246.97	F	id01207	22.5	10	-
MPN064	183.21	231.49	F	id01214	45	70	210

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN065	18.31	16.68	F	id01230	12.5	18.75	-
MPN066	76.59	90.09	G	id01237	55	70	210
MPN067	176.58	193.15	K	id01282	45	32.5	95
MPN068	13.16	14.33	U	id01288	16.25	-	-
MPN069	15.66	6.61	J	id01291	10	-	-
MPN070	7.96	4.26	M	id01296	16.25	-	-
MPN071	15.04	12.54	R	id01315	27.5	-	-
MPN072	12.56	9.08	J	id01319	22.5	-	-
MPN073	128.56	149.77	F	id01336	32.5	45	70
MPN074	3.35	0	J	id01340	18.75	-	-
MPN075	7.29	5.11	M	id01341	32.5	10	-
MPN076	240.08	233.95	G	id01388	45	210	70
MPN077	87.14	108.58	R	id01412	260	95	60
MPN078	15.65	19.67	G	id01418	55	70	210
MPN079	13.68	12.78	G	id01451	32.5	45	-
MPN080	12.62	24.05	L	id01464	18.75	27.5	55
MPN081	10.78	12.68	E	id01511	45	-	-
MPN082	290.1	311.53	G	id01539	60	210	-
MPN083	12.54	9.72	S	id01566	55	18.75	10
MPN084	10.59	17.2	S	id01595	55	27.5	-
MPN085	0	0.41	M	-	-	-	-
MPN086	0	0	M	-	-	-	-
MPN087	0	0	M	-	-	-	-
MPN088	0	0	S	-	-	-	-
MPN089	0	0	V	-	-	-	-
MPN090	21.88	4.82	M	id01695	27.5	10	-
MPN091	0	0	S	-	-	-	-
MPN091a	-	-	-	id01712	-	-	-
MPN092	0	0	M	-	-	-	-
MPN093	34.81	44.01	M	-	-	-	-
MPN094	50.45	18.38	N	id01748	10	-	-
MPN095	0.12	9.24	E	-	-	-	-
MPN096	0	0	E	-	-	-	-
MPN097	0	0.06	M	-	-	-	-
MPN098	12.93	15.14	M	-	-	-	-
MPN099	88.51	70.22	M	id01956	210	-	-
MPN100	18.7	14.08	N	-	-	-	-
MPN101	227.16	207.06	M	-	-	-	-
MPN101a	-	-	-	-	-	-	-
MPN102	0	8.8	M	-	-	-	-
MPN103	0	0	S	-	-	-	-
MPN104	23.12	16.84	N	id02053	16.25	-	-
MPN104a	-	-	-	id02064	12.5	-	-
MPN105	60.24	54.03	J	id02077	32.5	45	260
MPN106	68.06	65.77	J	id02091	70	210	-
MPN107	0	0	L	-	-	-	-
MPN108	0	0	L	-	-	-	-
MPN109	42.02	41.08	V	id02159	18.75	-	-
MPN110	0	0	V	-	-	-	-
MPN111	0	0	V	-	-	-	-
MPN112	0	0	U	-	-	-	-
MPN113	0	0	U	-	-	-	-
MPN114	0	0	I	-	-	-	-
MPN115	134.42	124.74	J	id02306	22.5	-	-
MPN116	16.7	14.47	J	id02314	12.5	-	-
MPN117	31.35	33.28	J	id02319	18.75	-	-
MPN118	15.68	12.62	L	id02337	22.5	-	-
MPN119	103.49	100.82	O	id02338	210	95	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN120	261.04	222.59	O	id02379	22.5	-	-
MPN121	113.89	110.15	S	id02386	12.5	18.75	-
MPN122	12.34	9.27	L	id02395	60	260	95
MPN123	8.98	7.56	L	id02422	70	210	-
MPN124	111.36	62.67	K	id02467	45	210	-
MPN125	9.7	8.02	L	id02468	60	260	-
MPN126	29.3	31.13	R	id02498	18.75	-	-
MPN127	2.31	0.88	N	id02514	55	-	-
MPN127a	-	-	-	-	-	-	-
MPN128	85.32	89.03	M	-	-	-	-
MPN129	0	0	M	-	-	-	-
MPN130	68.9	25.12	N	id02580	16.25	10	-
MPN131	0	0	M	-	-	-	-
MPN132	87.68	48.39	M	-	-	-	-
MPN132a	-	-	-	id02622	10	-	-
MPN133	2.61	1.48	L	id02636	18.75	27.5	12.5
MPN134	139.47	122.39	G	id02653	60	260	-
MPN135	11.56	7.77	G	id02678	27.5	-	-
MPN136	10	18.35	G	id02694	27.5	18.75	-
MPN137	0	1.84	N	-	-	-	-
MPN138	0.48	0	N	-	-	-	-
MPN139	8.4	16.44	N	id02747	12.5	-	-
MPN140	186.67	160.18	J	id02752	32.5	45	210
MPN141	312.74	269.96	M	id02773	210	95	-
MPN142	121.02	130.37	M	id02838	70	32.5	210
MPN143	0	0	S	-	-	-	-
MPN144	189.94	174.57	M	-	-	-	-
MPN145	0	0	N	-	-	-	-
MPN146	0	0	M	-	-	-	-
MPN147	0	0	M	-	-	-	-
MPN148	8.34	4.23	M	id02983	27.5	-	-
MPN149	68.78	7.85	M	-	-	-	-
MPN150	0	0	M	-	-	-	-
MPN151	16.33	16.44	N	id03032	16.25	-	-
MPN152	27.86	32.6	M	id03047	10	-	-
MPN152a	-	-	-	-	-	-	-
MPN153	105.96	104.9	L	id03087	210	95	-
MPN153a	-	-	-	id03131	260	-	-
MPN154	170.56	149.49	K	id03137	60	210	-
MPN154a	-	-	-	-	-	-	-
MPN155	89.83	79.65	J	id03163	60	210	-
MPN155a	-	-	-	-	-	-	-
MPN156	16.47	18.33	J	id03189	12.5	-	-
MPN157	48.25	48.52	M	id03196	32.5	45	-
MPN158	12.41	11.25	H	id03211	27.5	-	-
MPN159	11.94	19.9	V	id03223	70	45	-
MPN160	8.58	8.97	M	-	-	-	-
MPN161	67.44	55.24	S	id03259	45	210	-
MPN162	69.6	30.57	M	id03281	10	18.75	-
MPN163	31.59	22.79	M	id03301	22.5	-	-
MPN164	92.89	80.44	J	id03308	12.5	-	-
MPN165	73.11	101.18	J	id03313	32.5	45	-
MPN166	115.14	120.72	J	id03329	22.5	-	-
MPN167	161.85	159.79	J	id03336	27.5	45	-
MPN168	76.59	102.63	J	id03341	32.5	45	-
MPN169	40.57	61.8	J	id03357	10	-	-
MPN170	51.61	44.49	J	id03359	18.75	-	-
MPN171	87.46	93.11	J	id03367	27.5	45	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN172	39.76	48.19	J	id03383	16.25	-	-
MPN173	35.5	30.34	J	id03391	12.5	-	-
MPN174	44.54	56.94	J	id03394	12.5	18.75	-
MPN175	24.95	31.33	J	id03398	16.25	-	-
MPN176	83.92	92.34	J	id03403	16.25	-	-
MPN177	101.93	121.09	J	id03409	18.75	-	-
MPN178	52.46	52.58	J	id03417	10	-	-
MPN179	41.29	48.38	J	id03422	18.75	-	-
MPN180	96.38	103.69	J	id03427	22.5	-	-
MPN181	60.7	68.48	J	id03434	12.5	-	-
MPN182	86.57	91.08	J	id03441	27.5	45	-
MPN183	56.95	44.74	J	id03452	18.75	-	-
MPN184	23.92	20.14	U	id03458	32.5	55	210
MPN185	66.28	62.01	F	id03477	22.5	-	-
MPN186	20.34	11.19	J	id03490	27.5	-	-
MPN187	33.88	33.83	J	id03504	10	-	-
MPN188	0.42	16.01	J	-	-	-	-
MPN189	72.76	67.03	J	id03510	16.25	-	-
MPN190	66.45	54.19	J	id03518	12.5	-	-
MPN191	329.64	325.62	K	id03522	32.5	45	210
MPN192	44.69	33.43	J	id03539	16.25	-	-
MPN193	40.55	32.98	P	id03545	27.5	45	-
MPN194	47.93	41.8	P	id03558	32.5	45	-
MPN195	23.75	25.95	P	id03572	32.5	45	-
MPN196	5.82	3.27	J	id03592	27.5	10	-
MPN197	143.46	146.9	E	id03609	60	210	-
MPN198	4.61	2.36	V	id03635	32.5	-	-
MPN198a	-	-	-	-	-	-	-
MPN199	2.6	3.37	M	id03651	27.5	-	-
MPN199a	-	-	-	id03681	27.5	-	-
MPN200	31.63	29.7	M	id03686	18.75	27.5	10
MPN201	2.73	0	V	id03725	-	-	-
MPN202	105.83	130.17	M	-	-	-	-
MPN203	0	0	M	-	-	-	-
MPN204	61.69	22.85	N	id03769	10	-	-
MPN205	181.26	175.6	M	-	-	-	-
MPN206	0	0	S	-	-	-	-
MPN207	430.31	442.76	G	id03807	70	210	45
MPN207a	-	-	-	id03803	12.5	-	-
MPN208	117.17	106.58	J	id03848	32.5	45	-
MPN208a	-	-	-	id03840	27.5	-	-
MPN209	13.59	10.57	P	id03861	70	260	-
MPN210	135.46	138.07	U	id03897	70	210	-
MPN211	41.78	37.4	L	id03931	60	260	-
MPN212	-	-	S	id03955	22.5	-	-
MPN213	19.36	54.02	M	id03965	95	210	-
MPN214	11.15	35.23	M	id04011	16.25	-	-
MPN215	40.78	43.69	E	id04015	32.5	45	-
MPN216	60.6	59.48	E	id04032	32.5	45	260
MPN216a	-	-	-	id04047	-	-	-
MPN217	50.59	48.73	E	id04052	45	210	-
MPN218	72.58	94.7	E	id04070	70	210	-
MPN219	88.59	87.66	J	id04113	16.25	-	-
MPN220	115.04	126.97	J	id04118	27.5	-	-
MPN221	32.16	35.19	J	id04129	22.5	-	-
MPN222	7.26	5.77	J	-	-	-	-
MPN223	16.72	16.38	T	id04151	32.5	45	-
MPN224	15.34	7.89	U	id04163	32.5	-	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN225	78.34	69.37	J	id04178	18.75	-	-
MPN226	102.23	96.53	J	id04187	18.75	10	-
MPN227	231.47	243.73	J	id04194	70	210	-
MPN228	138.62	142.07	J	id04222	27.5	45	-
MPN229	103.69	90.2	L	id04235	18.75	-	-
MPN230	119.29	115.06	J	id04243	16.25	-	-
MPN231	40.78	50.13	J	id04249	18.75	-	-
MPN232	30.9	29.81	L	id04254	55	70	-
MPN233	73.12	40.81	M	id04274	16.25	27.5	-
MPN234	3.64	0	M	-	-	-	-
MPN235	3.78	0.85	L	id04322	27.5	-	-
MPN236	24.36	23.73	J	id04337	45	12.5	-
MPN237	69.43	73.56	J	id04359	45	210	-
MPN238	44.84	42.3	J	id04386	45	-	-
MPN239	92.46	103.53	K	id04405	22.5	-	-
MPN240	84.72	83.49	O	id04421	32.5	45	-
MPN241	260.06	3.37	R	-	-	-	-
MPN242	0	0	U	-	-	-	-
MPN243	93.09	98.33	K	id04457	70	210	-
MPN244	33.97	35.31	L	id04492	18.75	-	-
MPN245	97.97	94.81	J	id04515	22.5	-	-
MPN246	145.6	134.84	F	id04510	18.75	-	-
MPN247	63.33	51.16	T	id04520	27.5	-	-
MPN248	12.53	11.14	T	id04531	-	-	-
MPN249	0	0	J	-	-	-	-
MPN250	432.42	463.01	G	id04560	45	70	260
MPN251	20.71	22.89	G	id04578	22.5	-	-
MPN252	68.1	77.15	J	id04587	45	70	210
MPN253	0	0	I	id04602	18.75	-	-
MPN254	52.97	93.44	L	id04614	16.25	-	-
MPN255	111.66	129.88	I	id04625	27.5	-	-
MPN256	86.95	93.12	R	id04633	27.5	-	-
MPN257	36.52	38.62	I	id04646	32.5	45	70
MPN258	145.85	158.03	G	id04660	55	70	210
MPN259	51.44	67.86	G	id04687	45	210	70
MPN260	91.42	63.85	G	id04710	22.5	-	-
MPN261	133.99	132.88	L	id04728	70	210	-
MPN262	34.51	32.63	M	id04765	45	-	-
MPN263	956.35	832.69	O	id04786	10	18.75	-
MPN264	41.59	26.59	R	id04794	27.5	-	-
MPN265	46.48	59.88	J	id04831	32.5	45	-
MPN266	128.95	123.69	K	id04833	16.25	-	-
MPN267	278.98	266.07	H	id04837	27.5	45	-
MPN268	59.03	60.5	G	id04855	12.5	18.75	-
MPN269	25.57	27.25	R	id04861	55	70	210
MPN270	0	0	M	-	-	-	-
MPN271	18.87	7.71	M	id04891	27.5	-	-
MPN272	54.37	39.83	S	id04936	10	-	-
MPN273	275.02	253.17	F	id04946	16.25	-	-
MPN274	0	0	P	-	-	-	-
MPN275	57.02	43.2	L	id04957	10	-	-
MPN276	24.81	29.88	S	id04978	27.5	45	60
MPN277	91.89	84.58	J	id04977	55	70	210
MPN278	14.33	10.5	M	id05004	45	-	-
MPN279	16.09	21.36	J	id05023	60	210	-
MPN280	115.22	106.77	J	id05051	55	210	-
MPN281	29.7	26.83	M	id05076	10	22.5	-
MPN282	0	0	M	-	-	-	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN283	0	0	N	-	-	-	-
MPN284	98.78	118.54	M	id05146	16.25	60	27.5
MPN285	4.41	0	V	-	-	-	-
MPN286	136.58	71.29	M	-	-	-	-
MPN287	0	6.41	N	id05241	12.5	18.75	-
MPN288	56.52	66.25	M	id05259	27.5	18.75	12.5
MPN289	0	0	V	-	-	-	-
MPN290	0	0	V	-	-	-	-
MPN291	18.41	18.41	O	id05326	22.5	-	-
MPN292	9.58	9.31	J	id05335	32.5	45	-
MPN293	0	0.22	U	id05340	18.75	-	-
MPN294	351.29	311.65	O	id05341	22.5	-	-
MPN295	404.69	387.09	K	id05356	27.5	-	-
MPN296	27.64	24.46	J	id05366	10	18.75	-
MPN297	112.43	118.12	D	id05369	16.25	-	-
MPN298	0	3.43	I	id05376	12.5	-	-
MPN299	62.4	62.36	I	id05382	27.5	45	-
MPN300	10.98	8.31	D	id05392	55	210	-
MPN301	17.63	17.91	D	id05415	22.5	-	-
MPN302	224.48	243.62	G	id05426	27.5	45	-
MPN303	320.57	359.48	G	id05444	55	70	210
MPN304	5.7	6.98	C	id05492	27.5	-	-
MPN305	0	0	C	-	-	-	-
MPN306	0	0	C	-	-	-	-
MPN307	6.56	17.02	C	id05530	32.5	22.5	-
MPN308	7.42	16.55	E	id05544	-	-	-
MPN309	69.12	68.92	S	id05570	55	95	-
MPN310	146.38	148.75	M	id05589	210	22.5	-
MPN311	81.74	93.59	D	id05659	45	-	-
MPN312	21.38	22.81	S	id05674	32.5	-	-
MPN313	0	0	S	-	-	-	-
MPN314	977.59	809.35	D	id05694	16.25	-	-
MPN315	217.05	171.58	D	id05699	32.5	45	-
MPN316	11.72	6.74	D	id05714	45	-	-
MPN317	17.99	6.74	D	id05732	-	-	-
MPN318	15.28	14.19	E	id05748	45	32.5	60
MPN319	7.16	10.34	E	-	-	-	-
MPN320	480.14	400.85	F	id05805	32.5	45	-
MPN321	212.48	221.03	FH	id05824	18.75	-	-
MPN322	827.73	921.4	F	id05833	32.5	45	70
MPN323	492.23	664.84	F	id05849	18.75	-	-
MPN324	737.03	778.53	F	id05859	70	210	-
MPN325	51.96	79.44	J	id05896	12.5	18.75	-
MPN326	26.78	16.34	J	id05902	10	-	-
MPN327	38.11	54.74	J	id05906	12.5	-	-
MPN328	53.84	46.92	L	id05915	32.5	45	-
MPN329	11.98	12.22	K	id05925	18.75	-	-
MPN330	11.34	12.17	S	id05934	32.5	-	-
MPN331	215.57	226.96	O	id05949	55	210	-
MPN332	82.05	76.23	O	id05968	70	210	-
MPN333	0	0	R	-	-	-	-
MPN334	0	0	R	-	-	-	-
MPN335	0	0	R	-	-	-	-
MPN335a	-	-	-	-	-	-	-
MPN336	35.16	34.87	H	id06098	32.5	45	210
MPN337	30.54	25.21	A	id06100	55	70	210
MPN338	38.35	47.1	A	id06125	60	210	-
MPN339	2.74	2.24	A	id06153	22.5	-	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN340	8.63	9.39	L	id06165	55	-	-
MPN341	2.1	1.51	L	id06195	-	-	-
MPN342	19.53	9.16	V	id06224	55	260	-
MPN343	2.73	0	V	-	-	-	-
MPN344	10.48	25.22	N	id06266	12.5	18.75	-
MPN345	0	0	V	-	-	-	-
MPN346	0	0	V	-	-	-	-
MPN347	0	0	V	-	-	-	-
MPN347a	-	-	-	-	-	-	-
MPN348	18.32	13.53	H	id06339	18.75	-	-
MPN349	56.83	51.56	R	id06346	27.5	45	-
MPN350	0.55	0	I	id06366	22.5	-	-
MPN351	0	0	V	-	-	-	-
MPN352	111.19	107.74	K	id06391	55	70	210
MPN353	20.46	19.27	L	id06420	60	210	-
MPN354	58.84	64.23	J	id06442	45	70	210
MPN355	13.45	15.06	J	id06452	27.5	-	-
MPN356	10.16	7.43	J	id06477	45	-	-
MPN357	21.3	21.45	L	id06502	60	210	-
MPN358	46.13	41.61	A	id06503	45	210	70
MPN359	31.35	35.65	A	id06529	22.5	-	-
MPN360	40.8	39.76	J	id06542	12.5	-	-
MPN361	24.47	23.52	J	id06546	45	-	-
MPN362	7.43	7.1	J	id06562	45	-	-
MPN363	9.78	10.65	M	-	-	-	-
MPN364	19.58	16.92	M	-	-	-	-
MPN365	2.73	0	V	-	-	-	-
MPN366	59.77	42.27	M	-	-	-	-
MPN367	59.32	50.76	M	-	-	-	-
MPN367a	-	-	-	-	-	-	-
MPN368	14.71	5.33	N	id06679	12.5	-	-
MPN369	40.03	0	M	-	-	-	-
MPN370	219.28	198.7	M	-	-	-	-
MPN371	0	0	S	-	-	-	-
MPN372	159.46	156.56	V	id06748	60	210	-
MPN373	0	0	M	-	-	-	-
MPN374	0	0	M	-	-	-	-
MPN375	0	0	M	-	-	-	-
MPN376	131.22	227.52	A	id06896	32.5	95	210
MPN377	576.82	525.35	A	id06907	10	18.75	-
MPN378	35.75	35.21	L	id06912	70	210	-
MPN379	29	27.62	L	id06950	32.5	-	-
MPN380	31.76	34.06	L	id06960	27.5	-	-
MPN381	84.01	91.31	R	id06976	32.5	45	-
MPN382	0	2.17	H	id06995	22.5	-	-
MPN383	102.58	71.48	R	id07023	27.5	45	-
MPN384	109.06	82.6	J	id07056	70	210	-
MPN385	0	9.9	A	-	-	-	-
MPN386	333.18	345.29	F	id07070	22.5	-	-
MPN387	60.56	68.99	S	id07090	45	-	-
MPN388	54.26	24.1	S	id07098	12.5	-	-
MPN389	521.01	539.55	O	id07110	45	32.5	210
MPN390	736.37	651.63	C	id07133	45	210	-
MPN391	740.46	755.66	C	id07155	45	210	-
MPN392	2017.58	2077.91	C	id07174	32.5	45	70
MPN393	2228.58	2269.69	C	id07190	45	210	32.5
MPN393a	-	-	-	-	-	-	-
MPN394	1271.12	1095.2	C	id07211	45	210	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN395	50.48	49.25	F	id07224	18.75	-	-
MPN396	48.32	46	U	id07270	70	260	-
MPN397	44.1	50.36	TF	id07271	70	210	-
MPN398	60.31	65.49	A	id07313	18.75	-	-
MPN399	29.33	21.93	A	id07322	16.25	27.5	-
MPN400	56.31	23.52	A	id07348	45	-	-
MPN401	319.8	321.92	K	id07373	18.75	-	-
MPN402	39.71	33.3	J	id07379	45	70	260
MPN403	0	0	S	-	-	-	-
MPN404	0	0	M	-	-	-	-
MPN405	0	0	A	-	-	-	-
MPN406	5.65	15.49	I	id07440	10	-	-
MPN407	8.55	14.24	I	id07490	10	18.75	45
MPN408	37.81	31.54	M	id07497	27.5	16.25	-
MPN409	18.46	10.48	M	-	-	-	-
MPN410	53.73	16.14	N	id07561	10	18.75	-
MPN411	0.3	0	M	-	-	-	-
MPN412	120.78	130.65	M	-	-	-	-
MPN412a	-	-	-	id07601	10	18.75	27.5
MPN413	0	0	S	-	-	-	-
MPN414	63.69	49.01	M	-	-	-	-
MPN415	15.97	16.97	P	id07642	18.75	55	-
MPN416	4.19	3.02	P	id07662	27.5	-	-
MPN417	0	0	P	-	-	-	-
MPN418	46.11	41.23	J	-	-	-	-
MPN419	71.78	76.66	L	id07745	70	210	-
MPN420	68.7	64.47	I	id07756	27.5	-	-
MPN421	28.13	16.77	U	id07776	32.5	-	-
MPN422	25.12	24.98	J	id07797	45	32.5	95
MPN423	20.07	23.42	S	id07804	12.5	-	-
MPN424	0	1.42	K	id07807	12.5	-	-
MPN425	55.25	52.9	U	id07821	32.5	55	-
MPN426	47.32	44.47	D	id07858	95	210	-
MPN427	101.12	108.26	R	id07862	27.5	45	-
MPN428	552.01	515.91	G	id07899	32.5	45	-
MPN429	604.15	513.68	G	id07923	45	70	210
MPN430	2503.62	2644.96	IG	id07935	32.5	45	70
MPN431	5.72	7.15	P	id07951	27.5	-	-
MPN432	7.58	8.41	P	id07968	32.5	-	-
MPN433	9.44	46.29	P	id07978	27.5	-	-
MPN434	2197.63	2293.61	O	id08000	60	210	-
MPN435	0	12.61	A	id08018	32.5	-	-
MPN436	32.82	26.93	M	id08071	45	12.5	22.5
MPN437	0	0	M	-	-	-	-
MPN438	0	0	S	-	-	-	-
MPN439	0	0	M	-	-	-	-
MPN440	15.49	0.89	M	-	-	-	-
MPN441	0	0	S	-	-	-	-
MPN442	0	0	M	-	-	-	-
MPN443	6.47	21.7	J	id08212	45	-	-
MPN444	26.7	12.11	M	id08268	10	27.5	18.75
MPN445	46.84	44.45	I	id08265	27.5	-	-
MPN446	114.09	96.29	J	id08288	22.5	-	-
MPN447	168.04	200.39	M	id08329	210	95	-
MPN448	0	0	H	-	-	-	-
MPN449	26.01	8.72	A	id08362	45	95	260
MPN450	14.75	12.53	L	id08373	32.5	-	-
MPN451	0	0	L	-	-	-	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN452	241.71	209.54	M	id08410	95	210	-
MPN453	46.95	46.92	M	id08418	27.5	60	-
MPN454	77.64	95.78	S	id08427	18.75	12.5	-
MPN455	57.26	10.82	I	id08439	27.5	45	-
MPN456	82.75	84.05	E	id08490	18.75	10	45
MPN457	53.92	64.11	S	-	-	-	-
MPN458	55.34	65.75	S	-	-	-	-
MPN459	35.3	40.55	E	id08552	10	-	-
MPN460	11.12	0	P	id08590	-	-	-
MPN461	40.6	42.79	P	id08591	27.5	18.75	-
MPN462	53.82	96.87	M	-	-	-	-
MPN463	0	0	S	-	-	-	-
MPN464	44.2	43.01	M	-	-	-	-
MPN464a	-	-	-	id08650	12.5	-	-
MPN465	0	0	M	-	-	-	-
MPN466	5.48	5.29	M	-	-	-	-
MPN467	40.03	0	M	-	-	-	-
MPN468	136.58	71.29	M	-	-	-	-
MPN469	13.49	8.72	A	id08722	22.5	-	-
MPN470	431	438.32	E	id08741	45	32.5	70
MPN471	20.56	11.32	J	id08743	10	-	-
MPN472	37.85	31.95	I	id08754	27.5	45	-
MPN473	7.9	5.67	I	id08766	27.5	-	-
MPN474	496.82	461.72	N	id08809	210	95	-
MPN474a	-	-	-	id08811	210	95	-
MPN475	13.47	16.51	T	id08827	45	95	-
MPN476	36.2	41.25	F	id08830	22.5	-	-
MPN477	9.01	7.98	A	id08842	22.5	-	-
MPN478	50.29	46.79	F	id08852	27.5	-	-
MPN479	466.06	451.84	I	id08864	18.75	-	-
MPN480	25.16	30.15	J	id08906	70	210	-
MPN481	15.22	14.51	T	id08917	22.5	-	-
MPN482	3.04	2.77	A	id08921	10	-	-
MPN483	20.24	22.68	I	id08922	32.5	-	-
MPN484	22.59	9.4	N	id08952	10	-	-
MPN485	10.95	7.3	S	-	-	-	-
MPN486	0	0	S	-	-	-	-
MPN487	26.97	26.58	O	id08982	45	210	-
MPN488	15.46	23.17	O	id09001	10	-	-
MPN488a	-	-	-	id09007	27.5	-	-
MPN489	26.19	12.6	M	id09062	27.5	12.5	18.75
MPN490	3.93	16.33	L	id09086	32.5	-	-
MPN491	66.07	79.2	L	id09089	27.5	18.75	12.5
MPN492	10.63	12.85	G	id09129	32.5	-	-
MPN493	13.72	11.01	G	id09138	22.5	70	-
MPN494	35.37	26.95	G	id09145	18.75	-	-
MPN495	35.54	30.26	G	id09150	10	-	-
MPN496	13.88	1.86	G	id09178	45	-	-
MPN497	0	0	G	-	-	-	-
MPN498	24.77	26.94	G	id09202	27.5	-	-
MPN499	13.24	10.77	S	id09212	18.75	-	-
MPN500	63.2	69.53	M	-	-	-	-
MPN501	44.56	16.53	N	id09244	12.5	-	-
MPN502	195.92	174.84	M	id09262	210	-	-
MPN503	2.77	5.88	M	-	-	-	-
MPN504	4.54	13.3	N	id09316	10	-	-
MPN505	18.99	9.44	M	id09333	10	-	-
MPN506	32.43	51.42	M	id09348	18.75	27.5	10

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN506a	-	-	-	-	-	-	-
MPN507	13.98	1.45	V	-	-	-	-
MPN508	0	0	U	id09427	10	-	-
MPN509	9.06	2.51	S	id09463	60	-	-
MPN510	0	0	S	-	-	-	-
MPN511	0	0	S	-	-	-	-
MPN512	7.36	0	M	-	-	-	-
MPN513	0	0	V	-	-	-	-
MPN514	0	0	S	-	-	-	-
MPN515	220.99	231.33	K	id09608	210	95	-
MPN516	285.85	330.15	K	id09681	210	95	-
MPN517	253.68	279.25	H	id09691	18.75	-	-
MPN518	162.23	172.32	T	id09718	32.5	45	210
MPN519	15.13	14.92	I	id09731	27.5	-	-
MPN520	66.46	68.52	J	id09776	70	210	-
MPN520a	-	-	-	-	-	-	-
MPN521	22.06	17.4	J	id09778	18.75	-	-
MPN522	13.14	8.74	J	id09788	22.5	-	-
MPN523	50.63	21.04	M	id09823	27.5	18.75	-
MPN524	70.82	19.63	N	id09841	12.5	18.75	-
MPN525	0	0	L	-	-	-	-
MPN526	32.77	33.72	K	id09868	32.5	45	70
MPN527	0	0	A	-	-	-	-
MPN528	193.49	181.37	C	id09903	22.5	-	-
MPN528a	0	0	L	-	-	-	-
MPN529	67.64	54.51	D	id09928	12.5	18.75	-
MPN530	511.99	495.83	S	id09930	12.5	18.75	-
MPN531	164.31	234.46	O	id09961	70	210	-
MPN532	39.94	53.46	I	id09982	32.5	-	-
MPN533	917.76	1001.36	C	id09999	45	70	210
MPN534	0	0	S	-	-	-	-
MPN535	0	0	L	-	-	-	-
MPN536	0	0	L	-	-	-	-
MPN537	0	0	L	-	-	-	-
MPN538	49.02	55.41	J	id10063	18.75	-	-
MPN539	232.77	256.42	J	id10071	12.5	18.75	-
MPN540	37.36	25.55	J	id10078	10	-	-
MPN541	90.78	59.29	J	id10086	12.5	-	-
MPN542	9.86	9.85	A	id10088	27.5	-	-
MPN543	8.17	5.79	J	id10099	32.5	-	-
MPN544	6.97	6.16	S	id10113	70	-	-
MPN545	41.25	36.72	K	id10155	32.5	-	-
MPN546	38.77	35.99	I	id10168	32.5	45	-
MPN547	91.54	81.97	C	id10196	55	70	210
MPN548	9.63	5.95	J	id10210	-	-	-
MPN549	37.02	45.25	L	id10225	32.5	45	-
MPN550	130.3	8.78	H	id10242	-	-	-
MPN551	7.76	7.64	L	id10257	27.5	-	-
MPN552	11.25	10.55	S	id10275	27.5	-	-
MPN553	63.45	74.17	J	id10301	55	210	-
MPN554	13.49	20.92	D	id10307	12.5	-	-
MPN555	1073.93	1055.13	O	id10317	22.5	-	-
MPN556	146.7	139.02	J	id10339	55	70	210
MPN557	28.24	27.59	D	id10341	60	210	-
MPN558	23.47	21.99	DJ	id10362	18.75	-	-
MPN559	3.04	5.85	S	id10369	22.5	-	-
MPN560	83	91.45	C	id10398	45	70	210
MPN561	18.29	24.75	F	id10399	22.5	-	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN562	40.65	38.71	H	id10423	27.5	-	-
MPN563	23.55	22.93	J	id10445	45	95	260
MPN564	17.33	38.32	C	id10468	18.75	-	-
MPN565	0	0	A	id10479	18.75	-	-
MPN566	48.54	42.94	I	id10487	27.5	-	-
MPN567	54.4	31.95	M	id10537	210	95	-
MPN568	14.93	14.5	D	id10552	32.5	-	-
MPN569	10.07	4.96	O	id10559	16.25	-	-
MPN570	0	0	A	-	-	-	-
MPN571	0	0	U	-	-	-	-
MPN572	1146.1	1066.33	O	id10620	45	70	210
MPN573	1284.93	1359.75	O	id10636	55	210	-
MPN574	978.79	1103.42	O	id10642	12.5	-	-
MPN575	0	0	A	-	-	-	-
MPN576	288.85	305.24	E	id10669	45	70	210
MPN577	0	0	S	-	-	-	-
MPN578	0	0	S	-	-	-	-
MPN578a	-	-	-	-	-	-	-
MPN579	0	0	A	-	-	-	-
MPN580	0	0	O	-	-	-	-
MPN581	0	0	O	-	-	-	-
MPN582	0	0	O	id10758	10	-	-
MPN582a	-	-	-	-	-	-	-
MPN583	0	0	S	-	-	-	-
MPN584	0	0	M	-	-	-	-
MPN585	0	0	M	id10809	10	-	-
MPN586	0	0	O	-	-	-	-
MPN587	0	0	M	-	-	-	-
MPN588	0	0.77	M	id10855	12.5	-	-
MPN589	0	0	S	-	-	-	-
MPN590	14.83	23.46	M	-	-	-	-
MPN591	29.13	32.71	M	id10891	10	27.5	18.75
MPN592	20.74	25.2	O	id10918	55	22.5	10
MPN593	13.33	0	M	-	-	-	-
MPN594	0	0	M	-	-	-	-
MPN595	43.95	33.83	G	id10936	16.25	-	-
MPN596	6.05	5.43	D	id10956	55	-	-
MPN597	28.45	28.36	C	id10962	16.25	-	-
MPN598	316.92	360.82	C	id10986	45	210	-
MPN599	28.92	37.48	C	id11002	32.5	-	-
MPN600	205.87	200.18	C	id11028	55	210	-
MPN601	19.82	30.42	C	id11038	18.75	-	-
MPN602	55.88	76.12	C	id11048	18.75	-	-
MPN603	0	13.19	C	id11054	10	-	-
MPN604	26.67	13.88	C	id11071	22.5	55	-
MPN605	0	0	A	id11075	18.75	-	-
MPN606	1183.11	1137.37	G	id11095	45	210	-
MPN607	41.35	40.07	OV	id11097	18.75	-	-
MPN608	29.73	26.48	P	id11111	27.5	-	-
MPN609	29.49	32.27	P	id11125	32.5	45	260
MPN610	13.08	8.37	P	id11148	45	260	-
MPN611	80.45	79.36	P	id11171	45	32.5	18.75
MPN612	0	0	S	-	-	-	-
MPN613	0	0	S	-	-	-	-
MPN614	0	0	S	-	-	-	-
MPN615	2.49	0.66	V	-	-	-	-
MPN616	108.07	100.06	J	id11274	16.25	-	-
MPN617	19.99	23.01	J	id11283	18.75	-	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN618	86.28	59.85	L	id11316	70	210	-
MPN619	36.35	35.53	L	id11361	95	210	-
MPN620	15.29	16.82	S	id11389	70	210	-
MPN621	82.05	83.58	J	id11413	55	70	210
MPN622	13.52	16.57	J	id11416	10	-	-
MPN623	32.14	10.04	KJ	id11420	45	260	95
MPN624	30.41	18.04	J	id11447	10	-	-
MPN625	330.04	350.61	OV	id11453	16.25	-	-
MPN626	0	0	K	-	-	-	-
MPN627	186.61	199.76	G	id11500	55	210	-
MPN628	94.91	117.79	G	id11524	55	70	260
MPN629	127.05	126.62	G	id11538	27.5	-	-
MPN630	13.73	14.34	M	id11562	32.5	55	-
MPN631	195.74	195.05	J	id11564	32.5	45	-
MPN632	13.05	13.14	F	id11578	22.5	-	-
MPN633	0	0	O	-	-	-	-
MPN634	0	0	TF	-	-	-	-
MPN635	0	0	S	-	-	-	-
MPN636	279.69	265.26	J	id11646	22.5	-	-
MPN637	0	4.06	I	-	-	-	-
MPN638	321.78	277.22	V	id11672	45	70	210
MPN639	11.59	5.3	M	id11692	27.5	10	-
MPN640	0	0	M	-	-	-	-
MPN641	8.06	1.81	M	id11718	27.5	12.5	-
MPN642	8.98	4.65	M	id11731	27.5	-	-
MPN643	6.79	4.16	M	id11737	27.5	-	-
MPN644	0	0	M	-	-	-	-
MPN645	0	0	M	id11756	-	-	-
MPN646	0	0	M	id11768	27.5	-	-
MPN647	0	2.83	M	id11780	27.5	-	-
MPN648	0	0	S	-	-	-	-
MPN649	0	0	M	-	-	-	-
MPN650	0	0	M	-	-	-	-
MPN651	0	0	G	-	-	-	-
MPN652	9.5	18.21	G	id11845	-	-	-
MPN653	27.43	18.89	G	id11853	16.25	-	-
MPN654	29.7	0	M	-	-	-	-
MPN655	7.66	26.61	N	id11877	27.5	12.5	18.75
MPN656	7.17	0.58	J	id11900	27.5	-	-
MPN657	0	1.66	A	-	-	-	-
MPN658	89.49	109.31	J	id11925	16.25	10	-
MPN659	14.33	5.61	J	id11936	22.5	-	-
MPN660	61.11	77.69	J	id11941	12.5	-	-
MPN661	16.13	18.97	V	id11968	45	95	-
MPN662	162.33	143.14	EV	id11977	16.25	-	-
MPN663	30.4	30.12	J	id11986	27.5	-	-
MPN664	57.99	49.23	I	id12002	27.5	45	-
MPN665	2630.78	2477.83	J	id12024	45	210	-
MPN666	19.15	15.68	R	id12025	22.5	-	-
MPN667	37.14	35.55	M	id12043	27.5	-	-
MPN668	412.08	381.82	O	id12065	12.5	-	-
MPN669	31.93	32.86	J	id12066	45	-	-
MPN670	73.44	80.23	S	id12096	32.5	45	210
MPN670a	-	-	-	-	-	-	-
MPN671	197.78	188.49	O	id12132	60	210	-
MPN671a	-	-	-	id12104	18.75	-	-
MPN672	8.37	35.15	F	id12145	18.75	-	-
MPN673	164.43	155.87	I	id12151	18.75	-	-

Continued on next page

Table B.3 – continued from previous page

MPN	protein copies/cell at 6h	protein copies/cell at 96h	COG category	MS ID	MW1 (kDa)	MW2 (kDa)	MW3 (kDa)
MPN674	960.99	995.72	C	id12153	32.5	45	-
MPN675	5.76	0	N	id12181	10	-	-
MPN676	0	0	S	-	-	-	-
MPN677	40.29	46.54	R	id12223	45	-	-
MPN678	89.52	95.2	J	id12248	45	70	210
MPN679	12.27	6.93	J	id12259	27.5	-	-
MPN680	13.48	12.91	U	id12272	32.5	45	-
MPN681	0	0	J	-	-	-	-
MPN682	10.31	25.94	J	-	-	-	-
MPN683	39.88	20.73	P	id12291	32.5	45	-
MPN684	60.98	26.78	P	id12369	95	210	27.5
MPN685	104.82	78.4	P	id12385	27.5	45	-
MPN686	45.6	32.35	LD	id12405	45	260	95
MPN687	24.95	17.87	S	id12414	45	-	-
MPN688	201.07	199.15	D	id12426	27.5	45	-

Table B.3.: Applying MS to fractionated protein extracts allowed to identify genes from which proteins of different sizes are expressed; the functions encoded by the diffeernt COG classes can be found in Appendix B, Table B.4.

Table B.4: COG Categories

COG category	function
A	membrane Proteins of unknown function
C	energy production and conversion, coenzyme metabolism
D	cell division and chromosome partitioning
E	amino acid transport and metabolism
F	nucleotide transport and metabolism; coenzyme metabolism
G	carbohydrate transport and metabolism
H	coenzyme metabolism
I	lipid metabolism
J	translation, ribosomal structure and biogenesis
K	transcription
L	DNA replication, recombination and repair
M	cell envelope biogenesis, outer membrane
N	cell motility and secretion
O	post-translational modification, protein turnover, chaperones
P	inorganic ion transport and metabolism
R	general function prediction only
S	function unknown
T	signal transduction mechanisms
U	intracellular trafficking, secretion and vesicular transport
V	defense mechanisms

Table B.4.: Abbreviations for COG categories and the associated functions.

Bibliography

B. J. Akerley, E. J. Rubin, V. L. Novick, K. Amaya, N. Judson, and J. J. Mekalanos. A genome-scale analysis for identification of genes required for growth or survival of haemophilus influenzae. *Proc Natl Acad Sci U S A*, 99(2):966–971, Jan 2002.

B. Alberts, A. Johnson, and J. Lewis. *Molecular Biology of the Cell.* Garland Science, 2008.

U. Alon. How to choose a good scientific problem. *Mol Cell*, 35(6):726–728, Sep 2009.

R. B. Altman. Building successful biological databases. *Brief Bioinform*, 5(1):4–5, Mar 2004.

S. F. Altschul, T. L. Madden, A. A. Schäffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. Gapped blast and psi-blast: a new generation of protein database search programs. *Nucleic Acids Res*, 25(17):3389–3402, Sep 1997.

N. Amin and A. Peterkofsky. A dual mechanism for regulating camp levels in escherichia coli. *J Biol Chem*, 270(20):11803–11805, May 1995.

N. L. Anderson and N. G. Anderson. Proteome and proteomics: new technologies, new concepts, and new words. *Electrophoresis*, 19(11):1853–1861, Aug 1998.

J. Andreev, Z. Borovsky, I. Rosenshine, and S. Rottem. Invasion of hela cells by mycoplasma penetrans and the induction of tyrosine phosphorylation of a 145-kda host cell protein. *FEMS Microbiol Lett*, 132(3):189–194, Oct 1995.

B. P. Anton and E. A. Raleigh. Transposon-mediated linker insertion scanning mutagenesis of the escherichia coli mcra endonuclease. *J Bacteriol*, 186(17):5699–5707, Sep 2004.

L. Arvestad, A.-C. Berglund, J. Lagergren, and B. Sennblad. Bayesian gene/species tree reconciliation and orthology analysis using mcmc. *Bioinformatics*, 19 Suppl 1:i7–15, 2003.

Atta-Ur-Rahman. *Nuclear Magnetic Resonance: Basic Principles.* Springer Verlag, Berlin, 1986.

S. Y. Auyang. *Foundations of Complex-system Theories: in Economics, Evolutionary Biology, and Statistical Physics.* Cambridge University Press, 1999.

V. L. Avila. *Biology: Investigating life on earth.* Jones and Bartlett, Boston, 1995.

G. J. E. Baart and D. E. Martens. Genome-scale metabolic models: reconstruction and analysis. *Methods Mol Biol*, 799:107–126, 2012.

T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori. Construction of escherichia coli k-12 in-frame, single-gene knockout mutants: the keio collection. *Mol Syst Biol*, 2:2006.0008, 2006.

M. Babu, J. J. Díaz-Mejía, J. Vlasblom, A. Gagarinova, S. Phanse, C. Graham, F. Yousif, H. Ding, X. Xiong, A. Nazarians-Armavil, M. Alamgir, M. Ali, O. Pogoutse, A. Pe'er, R. Arnold, M. Michaut, J. Parkinson, A. Golshani, C. Whitfield, S. J. Wodak, G. Moreno-Hagelsieb, J. F. Greenblatt, and A. Emili. Genetic interaction maps in escherichia coli reveal

functional crosstalk among cell envelope biogenesis pathways. *PLoS Genet*, 7(11):e1002377, Nov 2011.

O. Beckonert, H. C. Keun, T. M. D. Ebbels, J. Bundy, E. Holmes, J. C. Lindon, and J. K. Nicholson. Metabolic profiling, metabolomic and metabonomic procedures for nmr spectroscopy of urine, plasma, serum and tissue extracts. *Nat Protoc*, 2(11):2692–2703, 2007.

B. D. Bennett, E. H. Kimball, M. Gao, R. Osterhout, S. J. Van Dien, and J. D. Rabinowitz. Absolute metabolite concentrations and implied enzyme active site occupancy in escherichia coli. *Nat Chem Biol*, 5(8):593–599, Aug 2009.

H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne. The protein data bank. *Nucleic Acids Res*, 28(1):235–242, Jan 2000.

W. P. Blackstock and M. P. Weir. Proteomics: quantitative and physical mapping of cellular proteins. *Trends Biotechnol*, 17(3):121–127, Mar 1999.

H. G. Bohn. *Aristotle's History of Animals (Translated by Cresswell, R.).* Adamant Media Corporation, elibron classics replica edition edition, 1862.

A. Boonmee, T. Ruppert, and R. Herrmann. The gene mpn310 (hmw2) from mycoplasma pneumoniae encodes two proteins, hmw2 and hmw2-s, which differ in size but use the same reading frame. *FEMS Microbiol Lett*, 290(2):174–181, Jan 2009.

G. E. P. Box and N. R. Draper. *Empirical Model-Building and Response Surfaces.* Wiley, New York, 1987.

S. E. Brenner. Errors in genome annotation. *Trends Genet*, 15(4):132–133, Apr 1999.

A. J. Brown. Enzyme action. *J. Chem. Society*, 81:373–386, 1902.

D. D. Buechter, D. N. Paolella, B. S. Leslie, M. S. Brown, K. A. Mehos, and E. A. Gruskin. Co-translational incorporation of trans-4-hydroxyproline into recombinant proteins in bacteria. *J Biol Chem*, 278(1):645–650, Jan 2003.

J. M. Buescher, W. Liebermeister, M. Jules, M. Uhr, J. Muntel, E. Botella, B. Hessling, R. J. Kleijn, L. Le Chat, F. Lecointe, U. Mäder, P. Nicolas, S. Piersma, F. Rügheimer, D. Becher, P. Bessieres, E. Bidnenko, E. L. Denham, E. Dervyn, K. M. Devine, G. Doherty, S. Drulhe, L. Felicori, M. J. Fogg, A. Goelzer, A. Hansen, C. R. Harwood, M. Hecker, S. Hubner, C. Hultschig, H. Jarmer, E. Klipp, A. Leduc, P. Lewis, F. Molina, P. Noirot, S. Peres, N. Pigeonneau, S. Pohl, S. Rasmussen, B. Rinn, M. Schaffer, J. Schnidder, B. Schwikowski, J. M. Van Dijl, P. Veiga, S. Walsh, A. J. Wilkinson, J. Stelling, S. Aymerich, and U. Sauer. Global network reorganization during dynamic adaptations of bacillus subtilis metabolism. *Science*, 335(6072):1099–1103, Mar 2012.

L. Buts, J. Lah, M.-H. Dao-Thi, L. Wyns, and R. Loris. Toxin-antitoxin modules as bacterial metabolic stress managers. *Trends Biochem Sci*, 30(12):672–679, Dec 2005.

M. A. Canonaco, C. O. Gualerzi, and C. L. Pon. Alternative occupancy of a dual ribosomal binding site by mrna affected by translation initiation factors. *Eur J Biochem*, 182(3):501–506, Jul 1989.

G. Casari, M. A. Andrade, P. Bork, J. Boyle, A. Daruvar, C. Ouzounis, R. Schneider, J. Tamames, A. Valencia, and C. Sander. Challenging times for bioinformatics. *Nature*, 376(6542):647–648, Aug 1995.

E. C. Y. Chan, K. K. Pasikanti, and J. K. Nicholson. Global urinary metabolic profiling procedures using gas chromatography-mass spectrometry. *Nat Protoc*, 6(10):1483–1499, Oct 2011.

B. S. W. Chang, K. Jönsson, M. A. Kazmi, M. J. Donoghue, and T. P. Sakmar. Recreating a functional ancestral archosaur visual pigment. *Mol Biol Evol*, 19(9):1483–1489, Sep 2002.

R. M. Chanock, L. Hayflick, and M. F. Barile. Growth on artificial medium of an agent associated with atypical pneumonia and its identification as a pplo. *Proc Natl Acad Sci U S A*, 48:41–49, Jan 1962a.

R. M. Chanock, W. D. James, H. H. Fox, H. C. Turner, M. A. Mufson, and L. Hayflick. Growth of eaton pplo in broth and preparation of complement fixing antigen. *Proc Soc Exp Biol Med*, 110:884–889, 1962b.

E. Chiner, J. Signes-Costa, A. L. Andreu, and L. Andreu. Mycoplasma pneumoniae pneumonia: an uncommon cause of adult respiratory distress syndrome. *An Med Interna*, 20(11):597–598, Nov 2003.

B. Christen, E. Abeliuk, J. M. Collier, V. S. Kalogeraki, B. Passarelli, J. A. Coller, M. J. Fero, H. H. McAdams, and L. Shapiro. The essential genome of a bacterium. *Mol Syst Biol*, 7:528, 2011.

C. Churcher. *Beginning Database Design: From Novice to Professional.* Springer, 2007.

F. D. Ciccarelli, T. Doerks, C. von Mering, C. J. Creevey, B. Snel, and P. Bork. Toward automatic reconstruction of a highly resolved tree of life. *Science*, 311(5765):1283–1287, Mar 2006.

A. Cornish-Bowden and M. L. Cárdenas. From genome to cellular phenotype–a role for metabolic flux analysis? *Nat Biotechnol*, 18(3):267–268, Mar 2000.

G. Corso, M. Esposito, and M. Gallo. Transformation of arginine into ornithine during the preparation of its tert-butyldimethylsilyl derivative for analysis by gas chromatography/mass spectrometry. *Biological Mass Spectrometry*, 22:698–702, 1993.

M. W. Covert and B. O. Palsson. Constraints-based models: regulation of gene expression reduces the steady-state solution space. *J Theor Biol*, 221(3):309–325, Apr 2003.

M. W. Covert, C. H. Schilling, I. Famili, J. S. Edwards, I. I. Goryanin, E. Selkov, and B. O. Palsson. Metabolic modeling of microbial strains in silico. *Trends Biochem Sci*, 26(3):179–186, Mar 2001.

D. J. Creek, A. Jankevics, R. Breitling, D. G. Watson, M. P. Barrett, and K. E. V. Burgess. Toward global metabolomics analysis with hydrophilic interaction liquid chromatography-mass spectrometry: improved metabolite identification by retention time prediction. *Anal Chem*, 83(22):8703–8710, Nov 2011.

A. W. Curnow, K. Hong, R. Yuan, S. Kim, O. Martins, W. Winkler, T. M. Henkin, and D. Söll. Glu-trnagln amidotransferase: a novel heterotrimeric enzyme required for correct decoding of glutamine codons during translation. *Proc Natl Acad Sci U S A*, 94(22):11819–11826, Oct 1997.

T. Dandekar, M. Huynen, J. T. Regula, B. Ueberle, C. U. Zimmermann, M. A. Andrade, T. Doerks, L. Sánchez-Pulido, B. Snel, M. Suyama, Y. P. Yuan, R. Herrmann, and P. Bork. Reannotating the mycoplasma pneumoniae genome sequence: adding value, function and reading frames. *Nucleic Acids Res*, 28(17):3278–3288, Sep 2000.

C. Darwin. *On the Origin of Species by Means of Natural Selection: The Preservation of Favored Races in the Struggle of Life*. Murray, London, 1859.

J. W. Drake, B. Charlesworth, D. Charlesworth, and J. F. Crow. Rates of spontaneous mutation. *Genetics*, 148(4):1667–1686, Apr 1998.

N. Draper and H. Smith. *Applied Regression Analysis*. Wiley, 1998.

N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas, and B. Ã. Palsson. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A*, 104(6):1777–1782, Feb 2007.

W. B. Dunn, D. I. Broadhurst, H. J. Atherton, R. Goodacre, and J. L. Griffin. Systems level studies of mammalian metabolomes: the roles of mass spectrometry and nuclear magnetic resonance spectroscopy. *Chem Soc Rev*, 40(1):387–426, Jan 2011.

K. Dybvig, P. Lao, D. S. Jordan, and W. L. Simmons. Fewer essential genes in mycoplasmas than previous studies suggest. *FEMS Microbiol Lett*, Jul 2010.

M. D. Eaton, G. Meiklejohn, W. van Herick, and M. Corey. Studies on the etiology of primary atypical pneumonia: Ii. properties of the virus isolated and propagated in chick embryos. *J Exp Med*, 82(5):317–328, Oct 1945.

J. S. Edwards and B. O. Palsson. Metabolic flux balance analysis and the in silico analysis of escherichia coli k-12 gene deletions. *BMC Bioinformatics*, 1:1, 2000.

J. S. Edwards, M. W. Covert, and B. O. Palsson. Metabolic modelling of microbes: the flux-balance approach. *Environ Microbiol*, 4(3):133–140, Mar 2002.

J. A. Eisen and C. M. Fraser. Phylogenomics: intersection of evolution and genomics. *Science*, 300(5626):1706–1707, Jun 2003.

A. M. Feist, C. S. Henry, J. L. Reed, M. Krummenacker, A. R. Joyce, P. D. Karp, L. J. Broadbelt, V. Hatzimanikatis, and B. O. Palsson. A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. *Mol Syst Biol*, 3:121, 2007.

A. M. Feist, M. J. Herrgård, I. Thiele, J. L. Reed, and B. O. Palsson. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol*, 7(2):129–143, Feb 2009.

D. A. Fell and J. R. Small. Fat synthesis in adipose tissue. an examination of stoichiometric constraints. *Biochem J*, 238(3):781–786, Sep 1986.

O. Fiehn. Metabolomics–the link between genotypes and phenotypes. *Plant Mol Biol*, 48(1-2): 155–171, Jan 2002.

W. M. Fitch. Distinguishing homologous from analogous proteins. *Syst Zool*, 19(2):99–113, Jun 1970.

R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, and J. M. Merrick. Whole-genome random sequencing and assembly of haemophilus influenzae rd. *Science*, 269(5223):496–512, Jul 1995.

R. M. T. Fleming and I. Thiele. von bertalanffy 1.0: a cobra toolbox extension to thermodynamically constrain metabolic models. *Bioinformatics*, 27(1):142–143, Jan 2011.

R. M. T. Fleming, I. Thiele, and H. P. Nasheuer. Quantitative assignment of reaction directionality in constraint-based models of metabolism: application to escherichia coli. *Biophys Chem*, 145(2-3):47–56, Dec 2009.

L. A. Flórez, S. F. Roppel, A. G. Schmeisky, C. R. Lammers, and J. Stülke. A community-curated consensual annotation that is continuously updated: the bacillus subtilis centred wiki subtiwiki. *Database (Oxford)*, 2009:bap012, 2009.

L. A. Flórez, C. R. Lammers, R. Michna, and J. Stülke. Cellpublisher: a web platform for the intuitive visualization and sharing of metabolic, signalling and regulatory pathways. *Bioinformatics*, 26(23):2997–2999, Dec 2010.

A. C. Forster and G. M. Church. Towards synthesis of a minimal cell. *Mol Syst Biol*, 2:45, 2006.

C. M. Fraser, J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, R. D. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. F. Tomb, B. A. Dougherty, K. F. Bott, P. C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison, 3rd, and J. C. Venter. The minimal gene complement of mycoplasma genitalium. *Science*, 270(5235):397–403, Oct 1995.

C. T. French, P. Lao, A. E. Loraine, B. T. Matthews, H. Yu, and K. Dybvig. Large-scale transposon mutagenesis of mycoplasma pulmonis. *Mol Microbiol*, 69(1):67–76, Jul 2008.

A. Friboulet and D. Thomas. Systems biology-an interdisciplinary approach. *Biosens Bioelectron*, 20(12):2404–2407, Jun 2005.

T. Gabaldón and M. A. Huynen. Shaping the mitochondrial proteome. *Biochim Biophys Acta*, 1659(2-3):212–220, Dec 2004.

T. Gabaldón and M. A. Huynen. Lineage-specific gene loss following mitochondrial endosymbiosis and its potential for function prediction in eukaryotes. *Bioinformatics*, 21 Suppl 2:ii144–ii150, Sep 2005.

T. Gabaldón and M. A. Huynen. *Reconstruction of ancestral proteomes. In Ancestral Sequence Reconstruction, edited by Liberles, D. A.* Oxford University Press, 2007.

L. A. Gallagher, E. Ramage, M. A. Jacobs, R. Kaul, M. Brittnacher, and C. Manoil. A comprehensive transposon mutant library of francisella novicida, a bioweapon surrogate. *Proc Natl Acad Sci U S A*, 104(3):1009–1014, Jan 2007.

A. Garcia-Pino, M. Christensen-Dalsgaard, L. Wyns, M. Yarmolinsky, R. D. Magnuson, K. Gerdes, and R. Loris. Doc of prophage p1 is inhibited by its antitoxin partner phd through fold complementation. *J Biol Chem*, 283(45):30821–30827, Nov 2008.

E. A. Gaucher, J. M. Thomson, M. F. Burgan, and S. A. Benner. Inferring the palaeoenvironment of ancient bacteria on the basis of resurrected proteins. *Nature*, 425(6955):285–288, Sep 2003.

F. M. Geier, E. J. Want, A. M. Leroi, and J. G. Bundy. Cross-platform comparison of caenorhabditis elegans tissue extraction strategies for comprehensive metabolome coverage. *Anal Chem*, 83(10):3730–3736, May 2011.

S. Y. Gerdes, M. D. Scholle, J. W. Campbell, G. Balázsi, E. Ravasz, M. D. Daugherty, A. L. Somera, N. C. Kyrpides, I. Anderson, M. S. Gelfand, A. Bhattacharya, V. Kapatral, M. D'Souza, M. V. Baev, Y. Grechkin, F. Mseeh, M. Y. Fonstein, R. Overbeek, A.-L. Barabási,

Z. N. Oltvai, and A. L. Osterman. Experimental determination and system level analysis of essential genes in escherichia coli mg1655. *J Bacteriol*, 185(19):5673–5684, Oct 2003.

S. Ghaemmaghami, W.-K. Huh, K. Bower, R. W. Howson, A. Belle, N. Dephoure, E. K. O'Shea, and J. S. Weissman. Global analysis of protein expression in yeast. *Nature*, 425(6959):737–741, Oct 2003.

F. M. Ghanem, A. C. Ridpath, W. E. Moore, and L. V. Moore. Identification of clostridium botulinum, clostridium argentinense, and related organisms by cellular fatty acid analysis. *J Clin Microbiol*, 29(6):1114–1124, Jun 1991.

D. G. Gibson, G. A. Benders, C. Andrews-Pfannkoch, E. A. Denisova, H. Baden-Tillson, J. Zaveri, T. B. Stockwell, A. Brownley, D. W. Thomas, M. A. Algire, C. Merryman, L. Young, V. N. Noskov, J. I. Glass, J. C. Venter, C. A. Hutchison, 3rd, and H. O. Smith. Complete chemical synthesis, assembly, and cloning of a mycoplasma genitalium genome. *Science*, 319 (5867):1215–1220, Feb 2008.

D. G. Gibson, J. I. Glass, C. Lartigue, V. N. Noskov, R.-Y. Chuang, M. A. Algire, G. A. Benders, M. G. Montague, L. Ma, M. M. Moodie, C. Merryman, S. Vashee, R. Krishnakumar, N. Assad-Garcia, C. Andrews-Pfannkoch, E. A. Denisova, L. Young, Z.-Q. Qi, T. H. Segall-Shapiro, C. H. Calvey, P. P. Parmar, C. A. Hutchison, 3rd, H. O. Smith, and J. C. Venter. Creation of a bacterial cell controlled by a chemically synthesized genome. *Science*, 329(5987): 52–56, Jul 2010.

R. Gil, F. J. Silva, J. Peretó, and A. Moya. Determination of the core of a minimal bacterial gene set. *Microbiol Mol Biol Rev*, 68(3):518–537, Sep 2004.

C. Gille, C. Bölling, A. Hoppe, S. Bulik, S. Hoffmann, K. Hübner, A. Karlstädt, R. Ganeshan, M. König, K. Rother, M. Weidlich, J. Behre, and H.-G. Holzhütter. Hepatonet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol Syst Biol*, 6:411, Sep 2010.

J. I. Glass, N. Assad-Garcia, N. Alperovich, S. Yooseph, M. R. Lewis, M. Maruf, C. A. Hutchison, 3rd, H. O. Smith, and J. C. Venter. Essential genes of a minimal bacterium. *Proc Natl Acad Sci U S A*, 103(2):425–430, Jan 2006.

R. Goodacre, S. Vaidyanathan, W. B. Dunn, G. G. Harrigan, and D. B. Kell. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Biotechnol*, 22(5): 245–252, May 2004.

G. Gottschalk. *Bacterial Metabolism*. Springer, 1st edition, 1986.

A. Goyer, E. Collakova, R. DÃaz de la Garza, E. P. Quinlivan, J. Williamson, J. F. Gregory, 3rd, Y. Shachar-Hill, and A. D. Hanson. 5-formyltetrahydrofolate is an inhibitory but well tolerated metabolite in arabidopsis leaves. *Journal of Biological Chemistry*, 280(28):26137–26142, Jul 2005.

J. E. Griffin, J. D. Gawronski, M. A. Dejesus, T. R. Ioerger, B. J. Akerley, and C. M. Sassetti. High-resolution phenotypic profiling defines genes essential for mycobacterial growth and cholesterol catabolism. *PLoS Pathog*, 7(9):e1002251, Sep 2011.

M. Güell, V. van Noort, E. Yus, W.-H. Chen, J. Leigh-Bell, K. Michalodimitrakis, T. Yamada, M. Arumugam, T. Doerks, S. Kühner, M. Rode, M. Suyama, S. Schmidt, A.-C. Gavin, P. Bork, and L. Serrano. Transcriptome complexity in a genome-reduced bacterium. *Science*, 326(5957):

1268–1271, Nov 2009.

M. Güell, E. Yus, M. Lluch-Senar, and L. Serrano. Bacterial transcriptomics: what is beyond the rna horiz-ome? *Nat Rev Microbiol*, 9(9):658–669, Sep 2011.

N. Gupta, S. Tanner, N. Jaitly, J. N. Adkins, M. Lipton, R. Edwards, M. Romine, A. Osterman, V. Bafna, R. D. Smith, and P. A. Pevzner. Whole proteome analysis of post-translational modifications: applications of mass-spectrometry for proteogenomic annotation. *Genome Res*, 17(9):1362–1377, Sep 2007.

S. Halbedel and J. Stülke. Tools for the genetic analysis of mycoplasma. *Int J Med Microbiol*, 297(1):37–44, Feb 2007.

S. Halbedel, J. Busse, S. R. Schmidl, and J. Stülke. Regulatory protein phosphorylation in mycoplasma pneumoniae. a pp2c-type phosphatase serves to dephosphorylate hpr(ser-p). *J Biol Chem*, 281(36):26253–26259, Sep 2006.

C. Hames. *Glycerolmetabolismus und Pathogenitaet von Mycoplasma pneumoniae.* PhD thesis, Georg-August-Universitaet Goettingen, 2008.

C. Hames, S. Halbedel, M. Hoppert, J. Frey, and J. Stülke. Glycerol metabolism is important for cytotoxicity of mycoplasma pneumoniae. *J Bacteriol*, 191(3):747–753, Feb 2009.

J. L. Hartman, 4th, B. Garvik, and L. Hartwell. Principles for the buffering of genetic variation. *Science*, 291(5506):1001–1004, Feb 2001.

L. M. Harwood and C. J. Moody. *Experimental Organic Chemistry: Principles and Practice.* Blackwell Scientific Publications, 1989.

C. S. Henry, M. DeJongh, A. A. Best, P. M. Frybarger, B. Linsay, and R. L. Stevens. High-throughput generation, optimization and analysis of genome-scale metabolic models. *Nat Biotechnol*, 28(9):977–982, Sep 2010.

R. Hershberg, S. Altuvia, and H. Margalit. A survey of small rna-encoding genes in *Escherichia coli*. *Nucleic Acids Res*, 31(7):1813–1820, Apr 2003.

R. Himmelreich, H. Hilbert, H. Plagens, E. Pirkl, B. C. Li, and R. Herrmann. Complete sequence analysis of the genome of the bacterium mycoplasma pneumoniae. *Nucleic Acids Res*, 24(22): 4420–4449, Nov 1996.

M. H. N. Hoefnagel, A. van der Burgt, D. E. Martens, J. Hugenholtz, and Snoep. Time dependent responses of glycolytic intermediates in a detailed glycolytic model of lactococcus lactis during glucose run-out experiments. *Mol Biol Rep*, 29(1-2):157–161, 2002.

H. Horai, M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito, and T. Nishioka. Massbank: a public repository for sharing mass spectral data for life sciences. *J Mass Spectrom*, 45(7):703–714, Jul 2010.

C. A. Hutchison, 3rd, S. N. Peterson, S. R. Gill, R. T. Cline, O. White, C. M. Fraser, H. O. Smith, and J. C. Venter. Global transposon mutagenesis and a minimal mycoplasma genome. *Science*, 286(5447):2165–2169, Dec 1999.

M. A. Huynen and P. Bork. Measuring genome evolution. *Proc Natl Acad Sci U S A*, 95(11): 5849–5856, May 1998.

I. Irnov, C. M. Sharma, J. Vogel, and W. C. Winkler. Identification of regulatory rnas in bacillus subtilis. *Nucleic Acids Res*, 38(19):6637–6651, Oct 2010.

M. A. Jacobs, A. Alwood, I. Thaipisuttikul, D. Spencer, E. Haugen, S. Ernst, O. Will, R. Kaul, C. Raymond, R. Levy, L. Chun-Rong, D. Guenthner, D. Bovee, M. V. Olson, and C. Manoil. Comprehensive transposon mutant library of pseudomonas aeruginosa. *Proc Natl Acad Sci U S A*, 100(24):14339–14344, Nov 2003.

J. D. Jaffe, M. Miyata, and Berg. Energetics of gliding motility in mycoplasma mobile. *J Bacteriol*, 186(13):4254–4261, Jul 2004.

A. Jankevics, M. E. Merlo, M. de Vries, R. J. Vonk, E. Takano, and R. Breitling. Separating the wheat from the chaff: a prioritisation pipeline for the analysis of metabolomics datasets. *Metabolomics*, 8(Suppl 1):29–36, Jun 2012.

J. K. Johnson and N. L. Somerson. Cholesterol as a limiting factor in the growth of mycoplasma pneumoniae. *Appl Environ Microbiol*, 40(2):391–399, Aug 1980.

A. Joshi and B. O. Palsson. Metabolic dynamics in the human red cell. part i–a comprehensive kinetic model. *J Theor Biol*, 141(4):515–528, Dec 1989.

A. R. Joyce, J. L. Reed, A. White, R. Edwards, A. Osterman, T. Baba, H. Mori, S. A. Lesely, B. O. Palsson, and S. Agarwalla. Experimental and computational assessment of conditionally essential genes in escherichia coli. *J Bacteriol*, 188(23):8259–8271, Dec 2006.

B. K. Kaiser, M. C. Clifton, B. W. Shen, and B. L. Stoddard. The structure of a bacterial duf199/whia protein: domestication of an invasive endonuclease. *Structure*, 17(10):1368–1376, Oct 2009.

M. Kanehisa and S. Goto. Kegg: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, 28(1):27–30, Jan 2000.

J. R. Karr, J. C. Sanghvi, D. N. Macklin, M. V. Gutschow, J. M. Jacobs, B. Bolival, Jr, N. Assad-Garcia, J. I. Glass, and M. W. Covert. A whole-cell computational model predicts phenotype from genotype. *Cell*, 150(2):389–401, Jul 2012.

K. J. Kauffman, P. Prakash, and Edwards. Advances in flux balance analysis. *Curr Opin Biotechnol*, 14(5):491–496, Oct 2003.

I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muñiz Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D. Karp. Ecocyc: a comprehensive database of escherichia coli biology. *Nucleic Acids Res*, 39(Database issue):D583–D590, Jan 2011.

H. Kitano. *Foundations of Systems Biology*. MIT Press, Cambridge, Massachusetts, 2001.

H. Kitano. Systems biology: a brief overview. *Science*, 295(5560):1662–1664, Mar 2002a.

H. Kitano. Computational systems biology. *Nature*, 420(6912):206–210, Nov 2002b.

H. Kitano. Introductions to systems biology. *Tanpakushitsu Kakusan Koso*, 48(7):789–793, Jun 2003.

H. Kitano, A. Funahashi, Y. Matsuoka, and K. Oda. Using process diagrams for the graphical representation of biological networks. *Nat Biotechnol*, 23(8):961–966, Aug 2005.

E. Klipp, R. Herwig, A. Kowald, C. Wierling, and Lehrach. *Systems Biology in Practice*. Wiley-VCH, 2005.

H. Kobayashi. A proton-translocating atpase regulates ph of the bacterial cytoplasm. *J Biol Chem*, 260(1):72–76, Jan 1985.

K. Kobayashi, S. D. Ehrlich, A. Albertini, G. Amati, K. K. Andersen, M. Arnaud, K. Asai, S. Ashikaga, S. Aymerich, P. Bessieres, F. Boland, S. C. Brignell, S. Bron, K. Bunai, J. Chapuis, L. C. Christiansen, A. Danchin, M. Débarbouille, E. Dervyn, E. Deuerling, K. Devine, S. K. Devine, O. Dreesen, J. Errington, S. Fillinger, S. J. Foster, Y. Fujita, A. Galizzi, R. Gardan, C. Eschevins, T. Fukushima, K. Haga, C. R. Harwood, M. Hecker, D. Hosoya, M. F. Hullo, H. Kakeshita, D. Karamata, Y. Kasahara, F. Kawamura, K. Koga, P. Koski, R. Kuwana, D. Imamura, M. Ishimaru, S. Ishikawa, I. Ishio, D. Le Coq, A. Masson, C. MauÃ«l, R. Meima, R. P. Mellado, A. Moir, S. Moriya, E. Nagakawa, H. Nanamiya, S. Nakai, P. Nygaard, M. Ogura, T. Ohanan, M. O'Reilly, M. O'Rourke, Z. Pragai, H. M. Pooley, G. Rapoport, J. P. Rawlins, L. A. Rivas, C. Rivolta, A. Sadaie, Y. Sadaie, M. Sarvas, T. Sato, H. H. Saxild, E. Scanlan, W. Schumann, J. F. M. L. Seegers, J. Sekiguchi, A. Sekowska, S. J. Séror, M. Simon, P. Stragier, R. Studer, H. Takamatsu, T. Tanaka, M. Takeuchi, H. B. Thomaides, V. Vagner, J. M. van Dijl, K. Watabe, A. Wipat, H. Yamamoto, M. Yamamoto, Y. Yamamoto, K. Yamane, K. Yata, K. Yoshida, H. Yoshikawa, U. Zuber, and N. Ogasawara. Essential bacillus subtilis genes. *Proc Natl Acad Sci U S A*, 100(8):4678–4683, Apr 2003.

M. Kochman, P. A. Hargrave, and J. Buczylko. Fructose-bisphosphate aldolase from helix pomatia. *Methods Enzymol*, 90 Pt E:259–263, 1982.

E. V. Koonin. Comparative genomics, minimal gene-sets and the last universal common ancestor. *Nat Rev Microbiol*, 1(2):127–136, Nov 2003.

N. J. Kruger, S. A. Hill, and R. Ratcliffe. *Regulation of primary metabolic pathways in plants*. Springer Verlag, 1999.

S. Kühner, V. van Noort, M. J. Betts, A. Leo-Macias, C. Batisse, M. Rode, T. Yamada, T. Maier, S. Bader, P. Beltran-Alvarez, D. Castaño Diez, W.-H. Chen, D. Devos, M. Güell, T. Norambuena, I. Racke, V. Rybin, A. Schmidt, E. Yus, R. Aebersold, R. Herrmann, B. Böttcher, A. S. Frangakis, R. B. Russell, L. Serrano, P. Bork, and A.-C. Gavin. Proteome organization in a genome-reduced bacterium. *Science*, 326(5957):1235–1240, Nov 2009.

M. J. Lambeth and M. J. Kushmerick. A computational model for glycogenolysis in skeletal muscle. *Ann Biomed Eng*, 30(6):808–827, Jun 2002.

G. C. Langridge, M.-D. Phan, D. J. Turner, T. T. Perkins, L. Parts, J. Haase, I. Charles, D. J. Maskell, S. E. Peters, G. Dougan, J. Wain, J. Parkhill, and A. K. Turner. Simultaneous assay of every salmonella typhi gene using one million transposon mutants. *Genome Res*, 19(12): 2308–2316, Dec 2009.

M. A. Larkin, G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. Clustal w and clustal x version 2.0. *Bioinformatics*, 23(21):2947–2948, Nov 2007.

C. Lartigue, J. I. Glass, N. Alperovich, R. Pieper, P. P. Parmar, C. A. Hutchison, 3rd, H. O. Smith, and J. C. Venter. Genome transplantation in bacteria: changing one species to another.

Science, 317(5838):632–638, Aug 2007.

N. E. Lewis, H. Nagarajan, and B. O. Palsson. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol*, Feb 2012.

C. Li, M. Donizelli, N. Rodriguez, H. Dharuri, L. Endler, V. Chelliah, L. Li, E. He, A. Henry, M. I. Stefan, J. L. Snoep, M. Hucka, N. Le Novère, and C. Laibe. Biomodels database: An enhanced, curated and annotated resource for published quantitative kinetic models. *BMC Systems Biology*, 4:92, Jun 2010.

L. M. Liberman, R. Sozzani, and P. N. Benfey. Integrative systems biology: an attempt to describe a simple weed. *Curr Opin Plant Biol*, 15(2):162–167, Apr 2012.

M. Liebeke, H. Meyer, S. Donat, K. Ohlsen, and M. Lalk. A metabolomic view of staphylococcus aureus and its ser/thr kinase and phosphatase deletion mutants: involvement in cell wall biosynthesis. *Chem Biol*, 17(8):820–830, Aug 2010.

M. Liebeke, K. Dörries, D. Zühlke, J. Bernhardt, S. Fuchs, J. Pané-Farré, S. Engelmann, U. Völker, R. Bode, T. Dandekar, U. Lindequist, M. Hecker, and M. Lalk. A metabolomics and proteomics study of the adaptation of staphylococcus aureus to glucose starvation. *Mol Biosyst*, 7(4):1241–1253, Apr 2011.

M. Lluch-Senar, M. Vallmitjana, E. Querol, and J. Piñol. A new promoterless reporter vector reveals antisense transcription in mycoplasma genitalium. *Microbiology*, 153(Pt 8):2743–2752, Aug 2007.

M. Lluch-Senar, E. Querol, and J. Piñol. Cell division in a minimal bacterium in the absence of ftsz. *Mol Microbiol*, 78(2):278–289, Oct 2010.

M. Lluch-Senar, J. Delgado, J. Wodke, and L. Serrano. Deciphering the minimal functional genomes. *MSB*, manuscript in preparation.

J. A. Lucy. Functional and structural aspects of biological membranes: a suggested structural role for vitamin e in the control of membrane permeability and stability. *Ann N Y Acad Sci*, 203:4–11, Dec 1972.

U. Mäder, A. G. Schmeisky, L. A. Flórez, and J. Stülke. Subtiwiki–a comprehensive community resource for the model organism bacillus subtilis. *Nucleic Acids Res*, 40(Database issue): D1278–D1287, Jan 2012.

R. Mahadevan, J. S. Edwards, and F. J. Doyle, 3rd. Dynamic flux balance analysis of diauxic growth in escherichia coli. *Biophys J*, 83(3):1331–1340, Sep 2002.

T. Maier, A. Schmidt, M. Güell, S. Kühner, A.-C. Gavin, R. Aebersold, and L. Serrano. Quantification of mrna and protein and integration with protein turnover in a bacterium. *Mol Syst Biol*, 7:511, 2011.

J. Maniloff, R. N. McElhany, L. R. Finch, and J. B. Baseman. *Mycoplasmas. Molecular Biology and Pathogenesis. - Source of nucleotides.* American Society for Microbiology, 1992.

J. T. Manolukas, M. F. Barile, D. K. Chandler, and J. D. Pollack. Presence of anaplerotic reactions and transamination, and the absence of the tricarboxylic acid cycle in mollicutes. *J Gen Microbiol*, 134(3):791–800, Mar 1988.

B. McClintock. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci U S A*, 36

(6):344–355, Jun 1950.

R. N. McElhaney and M. E. Tourtellotte. Mycoplasma membrane lipids: variations in fatty acid composition. *Science*, 164(3878):433–434, Apr 1969.

S. McKnight. Gene switching by metabolic enzymes–how did you get on the invitation list? *Cell*, 114(2):150–152, Jul 2003.

M. Merzbacher, C. Detsch, W. Hillen, and J. Stülke. Mycoplasma pneumoniae hpr kinase/phosphorylase. *Eur J Biochem*, 271(2):367–374, Jan 2004.

R. Milo, P. Jorgensen, U. Moran, G. Weber, and M. Springer. Bionumbers–the database of key numbers in molecular and cell biology. *Nucleic Acids Res*, 38(Database issue):D750–D753, Jan 2010.

J. Monod. From enzymatic adaptation to allosteric transitions. *Science*, 154(3748):475–483, Oct 1966.

S. N. Moreno, L. Zhong, H. G. Lu, W. D. Souza, and M. Benchimol. Vacuolar-type h+-atpase regulates cytoplasmic ph in toxoplasma gondii tachyzoites. *Biochem J*, 330 (Pt 2):853–860, Mar 1998.

H. J. Morowitz and Tourtellotte. The smallest living cells. *Sci Am*, 206:117–126, Mar 1962.

D. Nagrath, M. Avila-Elchiver, F. Berthiaume, A. W. Tilles, A. Messac, and Yarmush. Integrated energy and flux balance based multiobjective framework for large-scale metabolic networks. *Ann Biomed Eng*, 35(6):863–885, Jun 2007.

D. J. Naylor and F. U. Hartl. Contribution of molecular chaperones to protein folding in the cytoplasm of prokaryotic and eukaryotic cells. *Biochem Soc Symp*, pages 45–68, 2001.

F. C. Neidhardt. *Escherichia coli and Salmonella: Cellular and Molecular Biology.* ASM press, 1996.

A. R. Neves, A. Ramos, M. C. Nunes, M. Kleerebezem, J. Hugenholtz, W. M. de Vos, J. Almeida, and H. Santos. In vivo nuclear magnetic resonance studies of glycolytic kinetics in lactococcus lactis. *Biotechnol Bioeng*, 64(2):200–212, Jul 1999.

A. R. Neves, W. A. Pool, J. Kok, O. P. Kuipers, and Santos. Overview on sugar metabolism and its control in lactococcus lactis - the input from in vivo nmr. *FEMS Microbiol Rev*, 29 (3):531–554, Aug 2005.

J. K. Nicholson and J. C. Lindon. Systems biology: Metabonomics. *Nature*, 455(7216):1054–1056, Oct 2008.

P. Nicolas, U. Mäder, E. Dervyn, T. Rochat, A. Leduc, N. Pigeonneau, E. Bidnenko, E. Marchadier, M. Hoebeke, S. Aymerich, D. Becher, P. Bisicchia, E. Botella, O. Delumeau, G. Doherty, E. L. Denham, M. J. Fogg, V. Fromion, A. Goelzer, A. Hansen, E. Härtig, C. R. Harwood, G. Homuth, H. Jarmer, M. Jules, E. Klipp, L. Le Chat, F. Lecointe, P. Lewis, W. Liebermeister, A. March, R. A. T. Mars, P. Nannapaneni, D. Noone, S. Pohl, B. Rinn, F. Rügheimer, P. K. Sappa, F. Samson, M. Schaffer, B. Schwikowski, L. Steil, J. Stülke, T. Wiegert, K. M. Devine, A. J. Wilkinson, J. M. van Dijl, M. Hecker, U. Völker, P. BessiÃ¨res, and P. Noirot. Condition-dependent transcriptome reveals high-level regulatory architecture in bacillus subtilis. *Science*, 335(6072):1103–1106, Mar 2012.

D. Noble. *The Music of Life: Biology beyond genes.* Oxford University Press, 2008.

M. A. Oberhardt, J. Puchalka, K. E. Fryer, V. A. P. Martins dos Santos, and Papin. Genome-scale metabolic network analysis of the opportunistic pathogen pseudomonas aeruginosa pao1. *J Bacteriol*, 190(8):2790–2803, Apr 2008.

M. A. Oberhardt, B. O. Palsson, and J. A. Papin. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol*, 5:320, 2009.

J. D. Orth, I. Thiele, and B. Ã. Palsson. What is flux balance analysis? *Nat Biotechnol*, 28(3): 245–248, Mar 2010.

B. Oshry. *Seeing Systems: Unlocking the Mysteries of Organizational Life.* Mcgraw-Hill Professionals, 2007.

M. Pachkov, T. Dandekar, J. Korbel, P. Bork, and Schuster. Use of pathway analysis and genome context methods for functional genomics of mycoplasma pneumoniae nucleotide metabolism. *Gene*, 396(2):215–225, Jul 2007.

M. Papagianni, N. Avramidis, and G. Filiousis. Glycolysis and the regulation of glucose transport in lactococcus lactis spp. lactis in batch and fed-batch culture. *Microb Cell Fact*, 6:16, 2007.

D. L. Pavia, G. M. Lampman, G. S. Kriz, and R. G. Engel. *Introduction to Organic Laboratory Techniques: A Microscale Approach.* David Harris, 2005.

S. J. Pirt. The maintenance energy of bacteria in growing cultures. *Proc R Soc Lond B Biol Sci*, 163(991):224–231, Oct 1965.

J. D. Pollack, N. L. Somerson, and L. Senterfit. Isolation, characterization, and immunogenicity of mycoplasma pneumoniae membranes. *Infect Immun*, 2(3):326–339, Sep 1970.

J. D. Pollack, N. L. Somerson, and Senterfit. Chemical composition and serology of mycoplasma pneumoniae lipids. *J Infect Dis*, 127:Suppl:S32–Suppl:S35, Mar 1973.

J. D. Pollack, M. V. Williams, and McElhaney. The comparative metabolism of the mollicutes (mycoplasmas): the utility for taxonomic classification and the relationship of putative gene annotation and phylogeny to enzymatic function in the smallest free-living cells. *Crit Rev Microbiol*, 23(4):269–354, 1997.

N. D. Price, J. L. Reed, and B. O. Palsson. Genome-scale models of microbial cells: evaluating the consequences of constraints. *Nat Rev Microbiol*, 2(11):886–897, Nov 2004.

J. Puchalka, M. A. Oberhardt, M. Godinho, A. Bielecka, D. Regenhardt, K. N. Timmis, J. A. Papin, and V. A. P. Martins dos Santos. Genome-scale reconstruction and analysis of the pseudomonas putida kt2440 metabolic network facilitates applications in biotechnology. *PLoS Comput Biol*, 4(10):e1000210, Oct 2008.

T. Pupko, I. Pe'er, M. Hasegawa, D. Graur, and N. Friedman. A branch-and-bound algorithm for the inference of ancestral amino-acid sequences when the replacement rate varies among sites: Application to the evolution of five gene families. *Bioinformatics*, 18(8):1116–1123, Aug 2002.

A. Ramos, A. R. Neves, R. Ventura, C. Maycock, P. López, and H. Santos. Effect of pyruvate kinase overproduction on glucose metabolism of lactococcus lactis. *Microbiology*, 150(Pt 4): 1103–1111, Apr 2004.

S. Rasmussen, H. B. Nielsen, and H. Jarmer. The transcriptionally active regions in the genome of bacillus subtilis. *Mol Microbiol*, 73(6):1043–1057, Sep 2009.

S. Razin. The mycoplasmas. *Microbiol Rev*, 42(2):414–470, Jun 1978.

S. Razin, M. Aragaman, and J. Avigan. Chemical composition of mycoplasma cells and membranes. *J Gen Microbiol*, 33:477–487, Dec 1963.

S. Razin, D. Yogev, and Y. Naot. Molecular biology and pathogenicity of mycoplasmas. *Microbiol Mol Biol Rev*, 62(4):1094–1156, Dec 1998.

J. L. Reed and B. O. Palsson. Thirteen years of building constraint-based in silico models of escherichia coli. *J Bacteriol*, 185(9):2692–2699, May 2003.

J. L. Reed, I. Famili, I. Thiele, and B. O. Palsson. Towards multidimensional genome annotation. *Nat Rev Genet*, 7(2):130–141, Feb 2006.

M. Ridley, editor. *Evolution*. Oxford Readers, 2003.

M. Rizzi, U. Theobald, E. Querfurth, T. Rohrhirsch, M. Baltes, and Reuss. In vivo investigations of glucose transport in saccharomyces cerevisiae. *Biotechnol Bioeng*, 49(3):316–327, Feb 1996.

M. Rizzi, M. Baltes, U. Theobald, and M. Reuss. In vivo analysis of metabolic dynamics in saccharomyces cerevisiae: Ii. mathematical model. *Biotechnol Bioeng*, 55(4):592–608, Aug 1997.

F. L. Rock, W. Mao, A. Yaremchuk, M. Tukalo, T. Crépin, H. Zhou, Y.-K. Zhang, V. Hernandez, T. Akama, S. J. Baker, J. J. Plattner, L. Shapiro, S. A. Martinis, S. J. Benkovic, S. Cusack, and M. R. K. Alley. An antifungal agent inhibits an aminoacyl-trna synthetase by trapping trna in the editing site. *Science*, 316(5832):1759–1761, Jun 2007.

O. Rolfsson, B. O. Palsson, and I. Thiele. The human metabolic reconstruction recon 1 directs hypotheses of novel human metabolic functions. *BMC Syst Biol*, 5:155, 2011.

K. Römer. Programming paradigms and middleware for sensor networks. *GI/ITG Fachgespräch Sensornetze, Karlsruhe*, Jahrgang 3(3):33–35, Feb 2004.

Rottem. Membrane lipids of mycoplasmas. *Biochim Biophys Acta*, 604(1):65–90, May 1980.

S. Rottem, E. A. Pfendt, and L. Hayflick. Sterol requirements of t-strain mycoplasmas. *J Bacteriol*, 105(1):323–330, Jan 1971.

M. Sakamoto, T. Uchimera, and Komagata. Comparison of h2o-forming nadh oxidase from leuconostoc mesenteroides subsp. mesenteroides nric 1541t and h2o2-forming nadh oxidase from sporolactobacillus inulinus nric 1133t. *Journal of Fermentation and Bioengineering*, 82: 531 – 537, 1996.

J. M. Savinell and B. O. Palsson. Optimal selection of metabolic fluxes for in vivo measurement. ii. application to escherichia coli and hybridoma cell metabolism. *J Theor Biol*, 155(2):215–242, Mar 1992a.

J. M. Savinell and B. O. Palsson. Optimal selection of metabolic fluxes for in vivo measurement. i. development of mathematical methods. *J Theor Biol*, 155(2):201–214, Mar 1992b.

A. Scalbert, L. Brennan, O. Fiehn, T. Hankemeier, B. S. Kristal, B. van Ommen, E. Pujos-Guillot, E. Verheij, D. Wishart, and S. Wopereis. Mass-spectrometry-based metabolomics:

limitations and recommendations for future progress with particular focus on nutrition research. *Metabolomics*, 5(4):435–458, Dec 2009.

M. Scheer, A. Grote, A. Chang, I. Schomburg, C. Munaretto, M. Rother, C. Söhngen, M. Stelzer, J. Thiele, and D. Schomburg. Brenda, the enzyme information system in 2011. *Nucleic Acids Res*, 39(Database issue):D670–D676, Jan 2011.

J. Schellenberger, R. Que, R. M. T. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian, J. Kang, D. R. Hyduke, and B. O. Palsson. Quantitative prediction of cellular metabolism with constraint-based models: the cobra toolbox v2.0. *Nat Protoc*, 6(9):1290–1307, Sep 2011.

S. R. Schmidl, A. Otto, M. Lluch-Senar, J. Piñol, J. Busse, D. Becher, and J. Stülke. A trigger enzyme in mycoplasma pneumoniae: impact of the glycerophosphodiesterase glpq on virulence and gene expression. *PLoS Pathog*, 7(9):e1002263, Sep 2011.

R. Schuetz, N. Zamboni, M. Zampieri, M. Heinemann, and U. Sauer. Multidimensional optimality of microbial metabolism. *Science*, 336(6081):601–604, May 2012.

S. Seto, A. Uenoyama, and Miyata. Identification of a 521-kilodalton protein (gli521) involved in force generation or force transmission for mycoplasma mobile gliding. *J Bacteriol*, 187(10): 3502–3510, May 2005.

A. Seybert, R. Herrmann, and A. S. Frangakis. Structural analysis of mycoplasma pneumoniae by cryo-electron tomography. *J Struct Biol*, 156(2):342–354, Nov 2006.

E. Shacter, P. B. Chock, and E. R. Stadtman. Energy consumption in a cyclic phosphorylation/dephosphorylation cascade. *J Biol Chem*, 259(19):12260–12264, Oct 1984.

T. Shlomi, Y. Eisenberg, R. Sharan, and E. Ruppin. A genome-scale computational study of the interplay between transcriptional regulation and metabolism. *Mol Syst Biol*, 3:101, 2007.

M. L. Shuler, P. Foley, and J. Atlas. Modeling a minimal cell. *Methods Mol Biol*, 881:573–610, 2012.

L. F. Silvian, J. Wang, and T. A. Steitz. Insights into editing from an ile-trna synthetase structure with trnaile and mupirocin. *Science*, 285(5430):1074–1077, Aug 1999.

C. A. Smith, G. O'Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan, and G. Siuzdak. Metlin: a metabolite mass spectral database. *Ther Drug Monit*, 27(6):747–751, Dec 2005.

L. R. Snyder, J. J. Kirkland, and J. W. Dolan. *Introduction to Modern Liquid Chromatography, Third Edition*. Wiley, 2010.

T. G. Sobolevsky, A. I. Revelsky, B. Miller, V. Oriedo, E. S. Chernetsova, and I. Revelsky. Comparison of sitylation and esterificationacylation procedures in gc-ms analysis of amino acids. *Journal of Separation Science*, 26:1474–1478, 2003.

T. Soga, Y. Ohashi, Y. Ueno, H. Naraoka, M. Tomita, and T. Nishioka. Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. *J Proteome Res*, 2(5):488–494, 2003.

R. Sorek and P. Cossart. Prokaryotic transcriptomics: a new view on regulation, physiology and pathogenicity. *Nat Rev Genet*, 11(1):9–16, Jan 2010.

K. Spagou, I. D. Wilson, P. Masson, G. Theodoridis, N. Raikos, M. Coen, E. Holmes, J. C. Lindon, R. S. Plumb, J. K. Nicholson, and E. J. Want. Hilic-uplc-ms for exploratory urinary metabolic profiling in toxicological studies. *Anal Chem*, 83(1):382–390, Jan 2011.

O. D. Sparkman. *Mass Spec Desk Reference*. Global View Publishing, 2000.

J. Stelling. Mathematical models in microbial systems biology. *Curr Opin Microbiol*, 7(5): 513–518, Oct 2004.

C. E. V. Storm and E. L. L. Sonnhammer. Automated ortholog inference from phylogenetic trees and calculation of orthology reliability. *Bioinformatics*, 18(1):92–99, Jan 2002.

P. F. Suthers, M. S. Dasika, V. S. Kumar, G. Denisov, J. I. Glass, and C. D. Maranas. A genome-scale metabolic reconstruction of mycoplasma genitalium, ips189. *PLoS Comput Biol*, 5(2):e1000285, Feb 2009.

B. Szappanos, K. Kovács, B. Szamecz, F. Honti, M. Costanzo, A. Baryshnikova, G. Gelius-Dietrich, M. J. Lercher, M. Jelasity, C. L. Myers, B. J. Andrews, C. Boone, S. G. Oliver, C. Pál, and B. Papp. An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nat Genet*, 43(7):656–662, Jul 2011.

B. S. Szwergold, K. Ugurbil, and T. R. Brown. Properties of fructose-1,6-bisphosphate aldolase from escherichia coli: an nmr analysis. *Arch Biochem Biophys*, 317(1):244–252, Feb 1995.

Y. Taniguchi, P. J. Choi, G.-W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, and X. S. Xie. Quantifying e. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991):533–538, Jul 2010.

R. L. Tatusov, E. V. Koonin, and D. J. Lipman. A genomic perspective on protein families. *Science*, 278(5338):631–637, Oct 1997.

T. A. Tatusova, I. Karsch-Mizrachi, and J. A. Ostell. Complete genomes in www entrez: data representation and analysis. *Bioinformatics*, 15(7-8):536–543, 1999.

B. Teusink, J. Passarge, C. A. Reijenga, E. Esgalhado, C. C. van der Weijden, M. Schepper, M. C. Walsh, B. M. Bakker, K. van Dam, H. V. Westerhoff, and J. L. Snoep. Can yeast glycolysis be understood in terms of in vitro kinetics of the constituent enzymes? testing biochemistry. *Eur J Biochem*, 267(17):5313–5329, Sep 2000.

L. Thiele, K. Miettinen, P. J. Korhonen, and J. Molina. A preference-based evolutionary algorithm for multi-objective optimization. *Evol Comput*, 17(3):411–436, 2009.

S. G. Thorleifsson and I. Thiele. rbionet: A cobra toolbox extension for reconstructing high-quality biochemical networks. *Bioinformatics*, 27(14):2009–2010, Jul 2011.

R. t'Kindt, R. A. Scheltema, A. Jankevics, K. Brunker, S. Rijal, J.-C. Dujardin, R. Breitling, D. G. Watson, G. H. Coombs, and S. Decuypere. Metabolomics to unveil and understand phenotypic diversity between pathogen populations. *PLoS Negl Trop Dis*, 4(11):e904, 2010.

A. H. Tong, M. Evangelista, A. B. Parsons, H. Xu, G. D. Bader, N. Pagé, M. Robinson, S. Raghibizadeh, C. W. Hogue, H. Bussey, B. Andrews, M. Tyers, and C. Boone. Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science*, 294(5550):2364–2368, Dec 2001.

A. H. Y. Tong, G. Lesage, G. D. Bader, H. Ding, H. Xu, X. Xin, J. Young, G. F. Berriz,

R. L. Brost, M. Chang, Y. Chen, X. Cheng, G. Chua, H. Friesen, D. S. Goldberg, J. Haynes, C. Humphries, G. He, S. Hussein, L. Ke, N. Krogan, Z. Li, J. N. Levinson, H. Lu, P. Ménard, C. Munyana, A. B. Parsons, O. Ryan, R. Tonikian, T. Roberts, A.-M. Sdicu, J. Shapiro, B. Sheikh, B. Suter, S. L. Wong, L. V. Zhang, H. Zhu, C. G. Burd, S. Munro, C. Sander, J. Rine, J. Greenblatt, M. Peter, A. Bretscher, G. Bell, F. P. Roth, G. W. Brown, B. Andrews, H. Bussey, and C. Boone. Global mapping of the yeast genetic interaction network. *Science*, 303(5659):808–813, Feb 2004.

H. Tong, D. Bell, and K. Tabei. Automated data massaging, interpretation, and e-mailing modules for high throughput open access mass spectrometry. *Society for Mass Spectrometry*, 0305:1174–1187, 1999.

A. Uenoyama and M. Miyata. Identification of a 123-kilodalton protein (gli123) involved in machinery for gliding motility of mycoplasma mobile. *J Bacteriol*, 187(16):5578–5584, Aug 2005.

A. Uenoyama, A. Kusumoto, and M. Miyata. Identification of a 349-kilodalton protein (gli349) responsible for cytadherence and glass binding during gliding of mycoplasma mobile. *J Bacteriol*, 186(5):1537–1545, Mar 2004.

unpublished results. Proteomics data for mycoplasma pneumoniae. Lluch-Senar, M. and Yus., E.

M. J. van der Werf, K. M. Overkamp, B. Muilwijk, L. Coulier, and T. Hankemeier. Microbial metabolomics: toward a platform with full metabolome coverage. *Anal Biochem*, 370(1):17–25, Nov 2007.

M. J. van der Werf, K. M. Overkamp, B. Muilwijk, M. M. Koek, B. J. C. van der Werff-van der Vat, R. H. Jellema, L. Coulier, and T. Hankemeier. Comprehensive analysis of the metabolome of pseudomonas putida s12 grown on different carbon sources. *Mol Biosyst*, 4(4):315–327, Apr 2008.

W. M. van Gulik. Fast sampling for quantitative microbial metabolomics. *Curr Opin Biotechnol*, 21(1):27–34, Feb 2010.

V. van Noort, J. Seebacher, S. Bader, S. Mohammed, I. Vonkova, M. J. Betts, S. Kühner, R. Kumar, T. Maier, M. O'Flaherty, V. Rybin, A. Schmeisky, E. Yus, J. Stülke, L. Serrano, R. B. Russell, A. J. R. Heck, P. Bork, and A.-C. Gavin. Cross-talk between phosphorylation and lysine acetylation in a genome-reduced bacterium. *Mol Syst Biol*, 8:571, 2012.

T. van Opijnen and A. Camilli. Genome-wide fitness and genetic interactions determined by tn-seq, a high-throughput massively parallel sequencing method for microorganisms. *Curr Protoc Microbiol*, Chapter 1:Unit1E.3, Nov 2010.

T. van Opijnen, K. L. Bodi, and A. Camilli. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nat Methods*, 6(10):767–772, Oct 2009.

A. Varma and B. O. Palsson. Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type escherichia coli w3110. *Appl Environ Microbiol*, 60(10):3724–3731, Oct 1994a.

A. Varma and B. O. Palsson. Metabolic flux balancing: Basic concepts, scientific and practical use. *Nature Bio/Technology*, 12, 1994b.

V. E. Velculescu, L. Zhang, B. Vogelstein, and K. W. Kinzler. Serial analysis of gene expression. *Science*, 270(5235):484–487, Oct 1995.

K. B. Waites and Talkington. Mycoplasma pneumoniae and its role as a human pathogen. *Clin Microbiol Rev*, 17(4):697–728, table of contents, Oct 2004.

E. J. Want, I. D. Wilson, H. Gika, G. Theodoridis, R. S. Plumb, J. Shockcor, E. Holmes, and J. K. Nicholson. Global metabolic profiling procedures for urine using uplc-ms. *Nat Protoc*, 5 (6):1005–1018, Jun 2010.

V. C. Wasinger, S. J. Cordwell, A. Cerpa-Poljak, J. X. Yan, A. A. Gooley, M. R. Wilkins, M. W. Duncan, R. Harris, K. L. Williams, and I. Humphery-Smith. Progress with gene-product mapping of the mollicutes: Mycoplasma genitalium. *Electrophoresis*, 16(7):1090–1094, Jul 1995.

R. Watanabe, R. Iino, K. Shimabukuro, M. Yoshida, and H. Noji. Temperature-sensitive reaction intermediate of f1-atpase. *EMBO Rep*, 9(1):84–90, Jan 2008.

J. D. Watson and F. H. Crick. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, 171(4356):737–738, Apr 1953a.

J. D. Watson and F. H. Crick. Genetical implications of the structure of deoxyribonucleic acid. *Nature*, 171(4361):964–967, May 1953b.

J. Weiner, 3rd, C.-U. Zimmerman, H. W. H. Göhlmann, and R. Herrmann. Transcription profiles of the bacterium mycoplasma pneumoniae grown at different temperatures. *Nucleic Acids Res*, 31(21):6306–6320, Nov 2003.

H. V. Westerhoff, C. Winder, H. Messiha, E. Simeonidis, M. Adamczyk, M. Verma, F. J. Bruggeman, and W. Dunn. Systems biology: the elements and principles of life. *FEBS Lett*, 583(24): 3882–3890, Dec 2009.

M. R. Wilkins, C. Pasquali, R. D. Appel, K. Ou, O. Golaz, J. C. Sanchez, J. X. Yan, A. A. Gooley, G. Hughes, I. Humphery-Smith, K. L. Williams, and D. F. Hochstrasser. From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. *Biotechnology (N Y)*, 14(1):61–65, Jan 1996.

D. S. Wishart, C. Knox, A. C. Guo, R. Eisner, N. Young, B. Gautam, D. D. Hau, N. Psychogios, E. Dong, S. Bouatra, R. Mandal, I. Sinelnikov, J. Xia, L. Jia, J. A. Cruz, E. Lim, C. A. Sobsey, S. Shrivastava, P. Huang, P. Liu, L. Fang, J. Peng, R. Fradette, D. Cheng, D. Tzur, M. Clements, A. Lewis, A. De Souza, A. Zuniga, M. Dawe, Y. Xiong, D. Clive, R. Greiner, A. Nazyrova, R. Shaykhutdinov, L. Li, H. J. Vogel, and I. Forsythe. Hmdb: a knowledgebase for the human metabolome. *Nucleic Acids Res*, 37(Database issue):D603–D610, Jan 2009.

S. M. S. Wong, J. D. Gawronski, D. Lapointe, and B. J. Akerley. High-throughput insertion tracking by deep sequencing for the analysis of bacterial pathogens. *Methods Mol Biol*, 733: 209–222, 2011.

D. Wu, P. Hugenholtz, K. Mavromatis, R. Pukall, E. Dalin, N. N. Ivanova, V. Kunin, L. Goodwin, M. Wu, B. J. Tindall, S. D. Hooper, A. Pati, A. Lykidis, S. Spring, I. J. Anderson, P. D'haeseleer, A. Zemla, M. Singer, A. Lapidus, M. Nolan, A. Copeland, C. Han, F. Chen, J.-F. Cheng, S. Lucas, C. Kerfeld, E. Lang, S. Gronow, P. Chain, D. Bruce, E. M. Rubin, N. C. Kyrpides, H.-P. Klenk, and J. A. Eisen. A phylogeny-driven genomic encyclopaedia of bacteria and archaea. *Nature*, 462(7276):1056–1060, Dec 2009.

W.-X. Wu, Y. Zhan, T.-J. Zhao, Y.-R. Han, and Y.-F. Chen. Stochastic four-state mechanochemical model of f1-atpase. *Commun. Theor. Phys.*, 54:630–634, 2010.

R. Yasuda, H. Noji, K. Kinosita, Jr, and M. Yoshida. F1-atpase is a highly efficient molecular motor that rotates with discrete 120 degree steps. *Cell*, 93(7):1117–1124, Jun 1998.

E. Yus, T. Maier, K. Michalodimitrakis, V. van Noort, T. Yamada, W.-H. Chen, J. A. H. Wodke, M. Güell, S. Martínez, R. Bourgeois, S. Kühner, E. Raineri, I. Letunic, O. V. Kalinina, M. Rode, R. Herrmann, R. Gutiérrez-Gallego, R. B. Russell, A.-C. Gavin, P. Bork, and L. Serrano. Impact of genome reduction on bacterial metabolism and its regulation. *Science*, 326(5957):1263–1268, Nov 2009.

E. Yus, M. Güell, A. P. Vivancos, W.-H. Chen, M. Lluch-Senar, J. Delgado, A.-C. Gavin, P. Bork, and L. Serrano. Transcription start site associated rnas in bacteria. *Mol Syst Biol*, 8:585, 2012.

P. Zipori, L. Bosch, and J. van Duin. Translation of ms2 rna in vitro in the absence of initiation factor if-3. *Eur J Biochem*, 92(1):235–241, Dec 1978.

Acknowledgments

First of all, I want to thank my two supervisors, **Edda Klipp** and **Luis Serrano**, who enabled me to accomplish this special interdisciplinary thesis and who supported me during all this time, scientifically and humanly.

Next, i am indebted to **Konstantinos Michalodimitrakis** and **Tobias Maier** who not only tought me experimental basics and chemistry but supervised my all day work and my mental health during the hard times in Barcelona.

Another big thank goes to **Clemens Kühn**, **Maria Lluch-Senar**, **Javier Delgado**, **Almer van der Slot**, and **Julia Riedl** for fruitful discussions and for their great support for my personal well-being.

Also, I am very grateful towards all my other colleagues, that were working together with me in the mycoplasma project: **Jacek Puchalka**, **Miguel Godinho**, **Andreu Alibes**, **Toni Hermoso**, **Luca Cozzuto**, **Guglielmo Roma**, **Eva Yus**, **Bernhard Petzold**, **Sira Martínez**, **Tony Ferrar**, **Marie Trussart**, **Vitor Martins dos Santos**, **Hinnerk Eilers**, **Jörg Stülke**, **Francesco Mancuso**, **Ricardo Gutíerrez-Gallego**, **Josep Marcos de Aguila**, and members of the different facilities at the CRG and the EMBL Heidelberg.

Furthermore, i thank all not yet mentioned past and present members of the 'Theoretical Biophysics Group' at the HU Berlin and the 'Design of Biological Systems Group' at the CRG in Barcelona with whom I had the pleasure to work, to pass lunch hours together, and to share quite some out-of-work activities.

As of working with computers that do not always react on nice persuasive talking, I owe the successful finishing of the thesis to **Ivo Mainz**, system administrator in the TBP group, and to the TIC department of the CRG, naming amongs all of them **Ismael De Mingo Benitez**. Further great support has been provided by the program secretaries, **Blanka Wysocka**, **Eva Navarette**, and **Reyes Perza**, by **Nicola Collu** from the purchase department, and by the human resources members in Barcelona and by **Sabine Wagnitz** in Berlin. You made my life much more easy.

In addition, I want to thank the people reviewing this thesis and thus granting the final quality. Unless already mentioned, these were **Thomas Spiesser** and the last-minute reader.

This thesis was funded by a fellowship from **Obra Social, Fundacíon "laCaixa"**.

As many of you may know by personal or indirect experience, the time of a PhD thesis is not always ease and fun and each PhD student presumably thinks about quitting it all quite a few times. However, I did not quit but finished and this was possible thanks to many people, sharing parts of their lives with me (friends, flatmates, sports mates, ...). Representative for all of them I want to mention some that stayed true throughout all the years and over the distance: **Lisa**, **Stefan**, **Tobse**, **Juri**, **Felix**, **Ingo**, **Anne-Katrin**, **Carlos**, **Maria**, **Iván**, **Brito**, **Rebe**, as well as the one who left: **Christian**.

Last but not least, I want to thank my family who supported me on all possible levels. First, those given to me by birth, especially my parents **Andrea and Horst**, my sister **Birgitta** and my brother **Tino**, my grandparents **Johanna and Joachim**, and all the others who obstained from my presence during so many important events of the real life. Second, those chosen freely, especially **Anke + family**, **Alex + family**, and **Cristóbal** & **Jose Miguel** who gave me a second home. I love you all!

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur, Quellen und Hilfsmittel angefertigt habe.

Berlin, den 08.08.2012 Judith Andrea Heidrun Wodke