Universidade do Minho
Escola de Engenharia

Telma Cristina Oliveira Fernandes

Analysis of Subpopulation Proteomics and stress investigation of *Pseudomonas putida* KT2440 as a step towards deciphering population heterogeneities under varying environmental conditions

Novembro de 2013

Universidade do Minho
Escola de Engenharia
Departamento de Informática

Telma Cristina Oliveira Fernandes

Analysis of Subpopulation Proteomics and stress investigation of *Pseudomonas putida* KT2440 as a step towards deciphering population heterogeneities under varying environmental conditions

Dissertação de Mestrado
Mestrado em Bioinformática

Trabalho realizado sob orientação de

Professor Doutor Ralf Takors
Professora Doutora Isabel Rocha

Novembro de 2013

Acknowledgements

This work is not just the result of an individual effort, but rather a set of efforts that enabled me to finish one of the most important stages of my life. So I need to express a few words of thanks and deep appreciation to all who contributed directly or indirectly to the realization of this thesis.

My supervisor, Ralf Takors, for accepting me at the Institute of Biotechnology of the University of Stuttgart to perform my master thesis.

I appreciate all advice given of my co-supervisor, Professor Isabel Rocha. I thank her for the wisdom given during these two years and for the corrections and suggestions for the thesis, which ensured a greater enrichment of the same.

To Sarah Lieder, my greatest thanks for the guidance provided, unconditional support, friendship, I also thank the comments and pertinent suggestions given me during the realization of this thesis.

Christiane Mack and Michael Löffler for the friendship, affection and help that they have always shown to me.

I am very grateful to all my family for the encouragement received over the years, especially my parents, Luísa and Jorge, my brother and best friend, Miguel, my grandparents, Tila and my cousins Andreia and Magda.

To my aunt Susy the assistance provided by the accommodation that made possible the realization of my thesis at the University of Stuttgart.

To my boyfriend, Luís Correia, for remaining always on my side and supporting me during my academic life.

To all my friends, in particular to Rita Pacheco, Adriana Valente, Sophia Torres, Victor Costa, João Ribeiro, José Pacheco, and Andreia Carvalho for appropriate demonstrations of friendship and encouragement.

Abstract

Cell behavior differs even under identical micro-environmental conditions, resulting in cell heterogeneity. This phenomenon can transform a well producing bacterial population which is exposed to production-like stress conditions into subpopulations with reduced or even stopped product formation properties.

As a step towards understanding population heterogeneity it is necessary to apply sophisticated techniques, such as a combination of fluorescence activated cell sorting (FACS) and mass spectrometry (MS) resulting in proteomic information on the subpopulation level rather than on the whole population level.

The proteome of *Pseudomonas putida* KT2440 subpopulations were analysed in standard conditions (μ=0.2 h^{-1}) using R and Bioconductor that will serve as basis for comparison of non-stressed to stressed conditions to find optimization targets to reduce cell-to-cell variability and improved productivity. Here, the differences in subpopulations at different growth rates (μ=0.1 h^{-1}, μ=0.2 h^{-1} and μ=0.7 h^{-1}) were investigated. The second part of the thesis included the investigation of tolerance levels of *P. putida* to biofuels precursors, for a successful establishment of the experimental set-up of stress conditions on *P. putida* cultures.

Flow cytometry analysis showed the existence of a two subpopulations at each growth rate: a one chromosome equivalent (C1n) and a two chromosome equivalent (C2n) subpopulation at μ=0.1 h^{-1} and μ=0.2 h^{-1} and at μ=0.7 h^{-1}it was possible to see a subpopulation with two chromosome equivalent (C2n) and an additional subpopulation with multiple chromosome equivalents (CXn).

The statistical analysis on proteomic data and the experiments on phase two showed that *P. putida* KT2440 is very stable metabolically and it can tolerate solvents better than other industrial host systems. Therefore it can be concluded that *P. putida* KT2440 could be a promising candidate for a microbial cell factory chassis for industrial processes involving solvent production or 2 phase cultivation systems.

Key-words: Cell heterogeneity; *Pseudomonas putida* KT2440; Flow cytometry; Proteomic data.

Resumo

O comportamento das células pode variar até com pequenas alterações no microambiente, podendo originar heterogeneidade celular. Este fenómeno pode converter populações bacterianas com elevada capacidade de produção em subpopulações onde a taxa de produtividade pode diminuir drasticamente ou até parar por completo.

Para esclarecer a heterogeneidade a nível bacteriano é necessário utilizar técnicas sofisticadas como a combinação da seleção de células ativadas por fluorescência (FACS) e espectrometria de massa (MS) originando dados de proteómica a nível das subpopulações bacterianas, e não apenas ao nível da população.

O proteoma das subpopulações de *Pseudomonas putida* KT2440 foi analisado em condições padrão (μ=0.2 h^{-1}) usando o R e o Bioconductor que servirá como base para a comparação dos resultados das condições padrão e as condições de stresse para encontrar novas metas de otimização para reduzir a variabilidade entre subpopulações e melhorar a produtividade a nível industrial. Desta forma, diferentes taxas de crescimento foram testadas (μ=0.1 h^{-1}, μ=0.2 h^{-1} and μ=0.7 h^{-1}) nas diferentes subpopulações. A segunda parte da tese inclui a investigação de níveis de tolerância de *P. putida* a precursores para a produção de biocombustíveis, para ajudar a estabelecer as condições de stresse.

A análise da citometria de fluxo mostrou a existência de duas subpopulações em cada taxa de crescimento testada: subpopulação com um cromossoma equivalente (C1n) e uma subpopulação com dois cromossomas equivalente (C2n) foram encontradas nas taxas de crescimento de μ=0.1 h^{-1} e μ=0.2 h^{-1}. Na taxa de crescimento de μ=0.7 h^{-1} foi possível ver uma subpopulação com dois cromossomas equivalente (C2n) e uma subpopulação adicional com vários cromossomos equivalente (CXn).

A análise estatística dos dados de proteómica bem como os resultados da fase dois mostraram que *P. putida* KT2440 é muito estável metabolicamente e pode tolerar solventes melhor que outras bactérias. Portanto, pode concluir-se que *P. putida* KT2440 pode ser um candidato promissor para processos industriais que envolvem a produção de solventes.

Palavras-chave: Heterogeneidade celular; *Pseudomonas putida* KT2440; Citometria de fluxo;Proteómica.

Table of Contents

List of Figures

List of Tables

1.Introduction

A well producing bacterial population which is exposed to stress conditions may convert to subpopulations with reduced or even stopped product formation properties. Since inefficiently producing subpopulations of *Pseudomonas putida* cultures may have significant impact on productivity of industrial cultures, it is important to quantify population heterogeneity and find strategies to attenuate it (Müller et al. 2010).

This master thesis addresses particularly the proteomic analysis of *P. putida* KT2440 continuous cultures to investigate the differences between the subpopulations at varying growth conditions. Results may lead a step towards targeted optimization methods with a new perspective on preventing the advent of subpopulation split-up resulting in a loss of productivity in industrial fermentation processes.

This thesis is organized in four chapters. This introductory chapter that includes a brief description of the state of the art about *P. putida* and its importance in Biotechnology, biofuels, and stress, cellular heterogeneity, the evolution of proteomic analysis, the motivation of this work and the research aims. Chapter 2 describes the material and methods used for the experimental part and the work flow used for the computational part. Results are presented and discussed in chapter 3 and the conclusions and future perspectives are presented in chapter 4.

1.1 *Pseudomonas putida*

P. putida is a rod-shaped, flagellated, gram-negative and saprophytic bacterium (Table 1). It is able to colonize different environments such as soil, fresh water and the surfaces of living organisms, e.g. plant roots and the surrounding environments (rhizosphere). It is one of the best studied species of the genus *Pseudomonas* with an extraordinary diversity of enzymes, metabolic degradation pathways and transporters for secondary metabolites. Furthermore, it is able to tolerate moderate to high solvent concentrations in its environment and has promoting effects on plant growth and protection from pathogens (Espinosa-Urgel et al. 2000). This wealth of interesting properties makes it applicable in various different biotechnological applications including bioremediation of contaminated areas (Kenneth N Timmis et al. 1994) and the

biocatalytic production of fine chemicals (A Schmid et al. 2001; Zeyer et al. 1985). Other applications are e.g. quality improvement of fossil fuels (Galán et al. 2000), the production of bioplastics, especially polyhydroxyalkanoic acids (PHA) (Olivera et al. 2001), and as agents of plant growth promotion and plant pest control (O'Sullivan and O'Gara 1992; Walsh et al. 2001).

P. putida is an organism that plays a role in various biotechnological fields, ranging from industrial production of chemicals (white biotechnology), medical biotechnology (red biotechnology) and also environmental biotechnology (green). Besides these different applications, *P. putida* is characterized by stress resistance, can be easily isolated, and is an excellent host for gene cloning and expression, making it well suitable for laboratory experiments and optimization processes. This bacterium holds a few distinctions, including being the first Gram-negative soil bacterium to be certified as a safety strain by the Recombinant DNA Advisory Committee in 1982 (Federal Register 1982).

Table 1 - Classification of *Pseudomonas putida.*

Domain	Bacteria
Phylum	Proteobacteria
Class	Gamma proteobacteria
Order	Pseudomonadales
Family	Pseudomonadaceae
Genus	*Pseudomonas*
Species	*Pseudomonas putida*

1.1.1 *Pseudomonas putida* KT2440

The species *Pseudomonas putida* includes several strains, e.g. *P. putida* KT2440. This strain was sequenced and annotated in 2002. It contains a single circular chromosome of 6,181,863 bp with an average G+C content of 61.6%, carrying 5420 open reading frames (ORFs) (K. E. Nelson et al. 2002) (Figure 1).

P. putida KT2440 is a plasmid-free derivative of a toluene-degrading strain, initially named *Pseudomonas arvilla* mt-2 (Kojima et al. 1967) and later designated as *Pseudomonas putida* mt-2 (Olivera et al. 2001; Williams and Murray 1974).

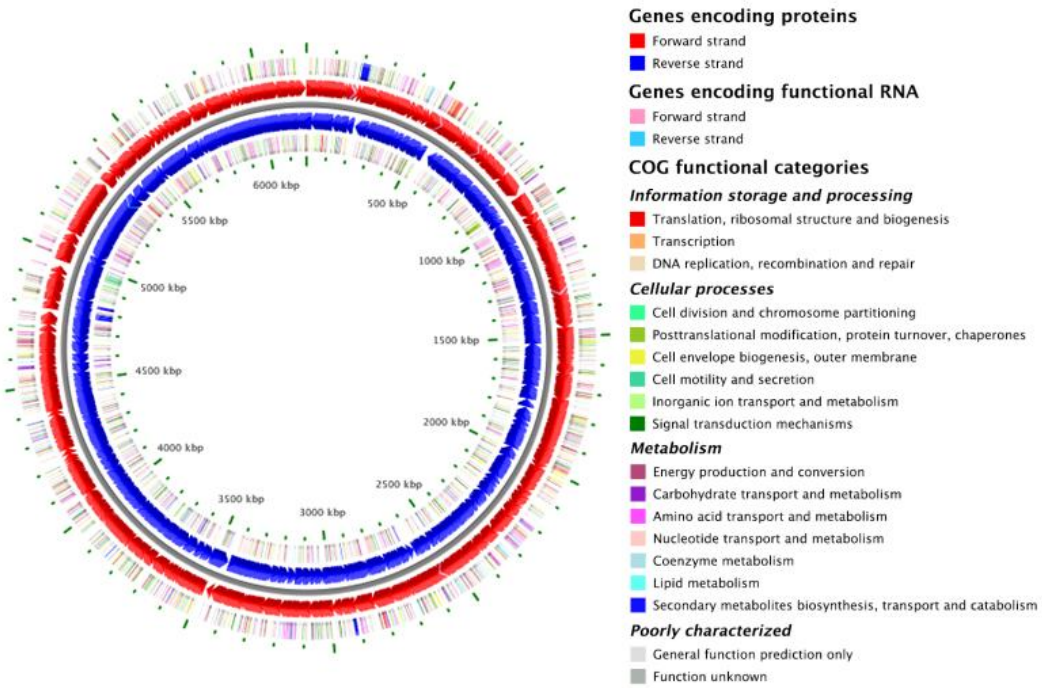

Figure 1 - Circular representation of the *Pseudomonas putida* KT2440 genome (Stothard et al. 2005).

P. putida KT2440 revealed a close evolutionary relationship with *P. aeruginosa,* one of the most dangerous opportunistic human (Stover et al. 2000) and plant pathogenic *Pseudomonas*. The genomic similarity between the two species accounts for 77.3 %. The main difference between *P. putida* KT2440 to *P. aeruginosa* is the lack of virulence genes. The similarity between *P. aeruginosa* and *P. putida* can be an advantage because discoveries made in *P. putida* might be able to be extrapolated to *P. aeruginosa* without having to manipulate this dangerous organism.

On a metabolic level *P. putida* is a versatile bacterium with a very complex metabolism including a large number and diversity of enzymes, transporters and proteins that protect this bacterium against toxic products and enables it to survive in different environments. Interesting pathways that exist in *P. putida* KT2440 are the catabolic degradation of aromatic compounds. *P. putida* is able to modify the structure of aromatic compounds, e.g. derived from plant material, into common intermediates of the central metabolic pathway that can be used as carbon and energy sources.

1.1.2 Stress and Biofuels

P. putida strains are known for being solvent-tolerant bacteria (Nicolaou et al. 2010) but the mechanisms and their interplay responsible for this tolerance are still to be understood.

Biofuels can be produced by microorganisms, but they can cause cellular stress leading to cellular heterogeneity harming the industrial production of these interesting compounds.

α-Pinene, (R)-(+)-Limonene and Bisabolene are considered next-generation biofuels because they can serve as precursors to produce fuel using suitable microorganisms. These solvents offer many advantages and might supplement existing petroleum fuels in the future. To reach this goal, a suitable industrial host is needed.

P. putida is one cell factory that can be a candidate as a biofuel-tolerant host strain. Therefore, in this thesis, first trial experiments for testing its tolerance towards α-Pinene, (R)-(+)-Limonene and Bisabolene are studied to pave the way to further investigate these stress conditions at single cell proteome level.

1.2 Cellular Heterogeneity

Since the publication of *P. putida* KT2440's genome (K. E. Nelson et al. 2002), and the genome-scale model reconstructions (Nogales et al. 2008; Puchałka et al. 2008; Sohn et al. 2010) our knowledge about this bacterium increased significantly. However, many mechanisms of adaptation, stress tolerance or recovery remain unknown and need to be studied.

It is now known that cell behavior differs even under identical micro-environmental conditions (Jahn et al. 2012) resulting in cell heterogeneity. This phenomenon can be caused by stress conditions, for example, in the bioreactors during fermentation processes. It can transform a well producing bacterial population in subpopulations with reduced or even stopped product formation properties (Figure 2). So far this diversity in cell cultures is associated with gene loss, mutations or variations in reactor environments (Fritzsch et al. 2012; Schweder et al. 1999) that can cause epigenetic modifications, variance in cell size owing to asymmetric cell division or different growth (Jahn et al. 2012). It is necessary to understand population heterogeneity because it has a significant impact on the productivity of industrial cultures.

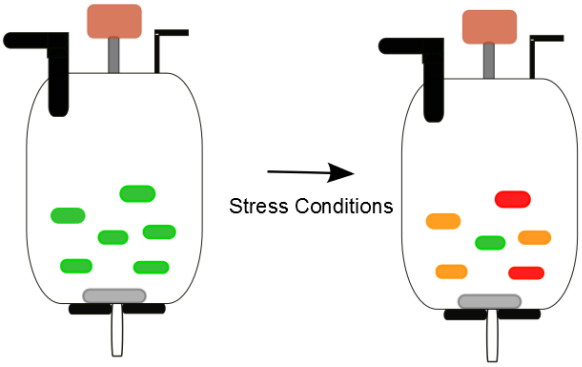

Figure 2 - Bioreactor with *P. putida* KT 2440 cultures after and before suffer stress conditions. The green rectangle represents producing *P. putida* KT2440 populations, the orange and red rectangle show *P. putida* KT2440 subpopulations with reduced and stopped product formation properties, respectively.

To obtain more information about population heterogeneity the use of sophisticated techniques is required, such as a combination of fluorescence activated cell sorting (FACS) and mass spectrometry (MS) resulting in proteomic information on subpopulation level rather than on whole population level.

The population elected for the proteomic analysis was *Pseudomonas putida* KT 2440 because of their biotechnological applicability, their versatile and non-pathogenic characteristics.

1.2.1 Identification of Subpopulations

Flow cytometry is a powerful technique to analyze different parameters of individual cells within heterogeneous populations. Thousands of cells pass through a laser beam per second and the light that emerges from each cell is captured.

The flow cytometer consists of a fluidic system, lasers, optics, electronic parts and a computer system. In the fluidic system, a central channel through which the sample is injected is enclosed by an outer sheath that contains a faster flowing fluid. Lasers are light sources for scatter and fluorescence. The optics receives the light and the electronics and the computer system convert the signals from the detectors into digital data (Figure 3).

When the cell passes through the laser, light is refracted and scattered. Light scatter is collected at two angles: forward scatter (FSC) and side scatter (SSC). Forward scatter measures scattered light in the direction of the laser path and measures the size of the cell and the side scatter measures scattered light at 90 degrees to the laser path and gives information about the granularity of the cell.

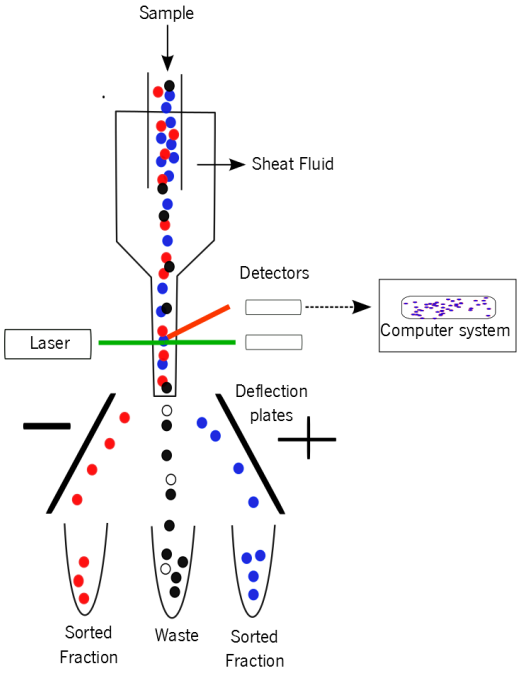

Figure 3 - Schematic diagram of a flow cytometer, showing the fluid sheath, laser, detectors, deflection plates and computer system.

One of the big advantages of this system is the fact that flow cytometers can detect very small particles (cells between 1 and 15 microns in diameter), which is a prerequisite for successful realization of prokaryotic population analysis.

To study cellular characteristics, flow cytometry is combined with fluorophore analysis and flow cytometry based cell sorting (FACS). A fluorophore, 4',6-diamidino-2-phenylindole (DAPI), is used to stain DNA. It is a fluorescent stain that binds strongly to A-T rich regions in DNA. DAPI can pass through an intact cell membrane. Therefore it can be used to stain both live and fixed cells. Previous staining of the samples with DAPI allows to measure DNA content of each single cell that passes through the fluorescence cytometer because during the flow cytometry process the DAPI are excited and emit a specific light that will be captured by a specific detector called channel 4 (FL 4). DNA content gives information on which cell in the whole populations is in which state of the cell cycle.

DNA content is then used as a parameter for fluorescence activated cell sorting to separate subpopulations, as the DNA molecule is very stable and it is easily labeled. The analysis of flow cytometry data will be carried out in this thesis using Bioconductor (R. C. Gentleman et al. 2004).

After separating *P. putida* KT 2440 in subpopulations, a detailed investigation is necessary of their characteristics and one appropriate approach is studying the proteomics of these cells.

1.3 Proteomics

Proteomics refer to the study of the proteome. The term proteome was defined as the entire set of proteins expressed by the genome. In detail, it is the set of proteins expressed at time of sampling under defined environmental conditions. Nowadays, proteomics is much more than this description. It is considered as all protein isoforms and modifications in a cell, and also the interactions between them, their structural description and their higher-order complexes (Tyers and Mann 2003). Proteins are an important key-element of life because they are the main components of the physiological pathways of the cells. Proteins represent about 50% of dry weight of a prokaryotic organism, catalyze biochemical reactions (enzymes), act as messengers and control elements that regulate cell reproduction, influence growth (growth factors), transport specific molecules and ions (hemoglobin), give organisms the ability to contract, to move or to change their shape, etc.

1.3.1 Proteomic Analysis

Proteomic analysis is the qualitative and quantitative investigation of the proteome. On a qualitative level, it allows to identify proteins, understand their distribution, posttranslational modifications, interactions between them and their structure and function. On a quantitative level it gives information about protein abundance and distribution, temporal changes in abundance due to synthesis and degradation or both and protein stoichiometry of reactions. The understanding of these mechanisms could be revolutionary on an industrial and medical level as it might allow optimization of the production of valuable compounds and can help to obtain information on disease development in an organism.

The method used to study the proteome of the different subpopulations in this project was mass spectrometry (MS).

MS was described by physicists in the late 1880s. It is a technique based on the formation of gas-phase ions from the analyte (positively or negatively charged), realized by the ion source component of the spectrometers. Ions can be isolated electrically (or magnetically) based on their mass-to-charge ratio (m/z) detected by the mass analyzer component. In a MS spectrum, the x- coordinate represents m/z values, whereas the y-axis indicates total ion counts

(El-aneed et al. 2009). The last part of the mass spectrometers is the ion detector to monitor the ions (Figure 4).

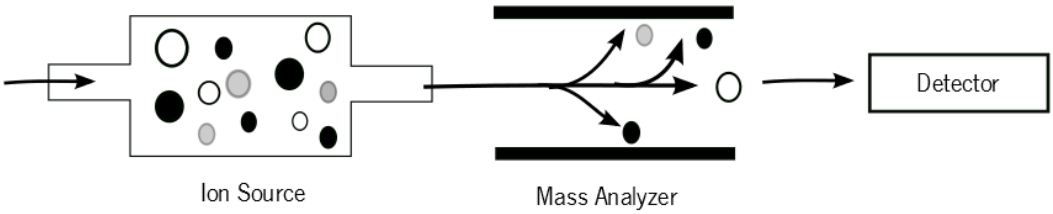

Figure 4 - Block diagram showing the mass spectrometers components (El-Aneed et al. 2009).

Mass-spectrometry-based proteomics is a high-throughput, sensitive and specific analysis for qualitative and quantitative applications that has become an important analytical tool in biology during the past 20 years. Driven by the genome sequencing projects including the human genome (Human Genome Project (Lander et al. 2001)), the success of protein MS was enabled by the development of efficient algorithms to extract data and the development of soft protein ionization methods.

Protein MS needs a high performance separation instrumentation to perform efficient separation of complex mixtures and prepare the molecules for the ionization source leading to high accuracy and sensitivity of the mass spectrometric analysis. The method chosen for the separation of complex peptides and proteins in this project was ultra-high performance liquid chromatography (UPLC).

UPLC is a variant of high-performance liquid chromatography (HPLC) using columns with particle size <2 μm (typically, 1.8 μm). It provides significantly better separation, peak capacity, speed, resolution and sensitivity for analytical determinations, particularly when coupled with mass spectrometers capable of high-speed acquisitions (Churchwell et al. 2005).

The UPLC device includes a sampler that brings the sample from the mobile phase into the stationary phase, pumps that deliver the desired flow and composition of the mobile phase through the column, the detector that generates a signal proportional to the amount of sample component emerging from the column, and the computer system, which controls the UPLC instrument and provides data analysis (Figure 5).

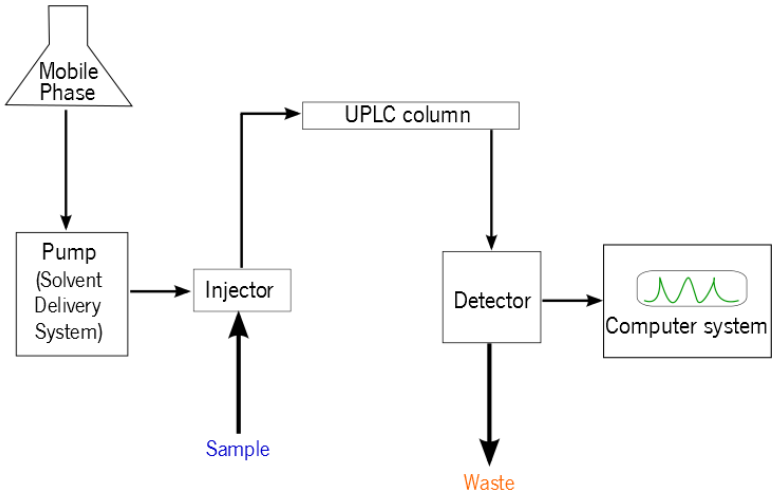

Figure 5 - Block diagram showing the components of an UPLC instrument.

Overall, the combination of fluorescence labeling, fluorescence activated cell sorting and UPLS+MS allows a separation of different subpopulations and getting the proteomic profiles of each subpopulation in order to give insights into molecular heterogeneity.

1.3.2 The MS Raw Data Analysis

After successful realization of MS experiments, algorithms need to be applied to extract information for analysis of mass spectrometry data in an efficient and robust way. Several thousand proteins in complex proteomes need to be identified and quantified with high-accuracy.

Known algorithms for the use in proteomic analysis are: ASAP Ratio (X.-J. Li et al. 2003), AYUMS (Saito et al. 2007), Census (S. K. Park et al. 2008), i-Tracker (Shadforth et al. 2005), jTraqX (Muth et al. 2010), MaxQuant (Cox and Mann 2008), MRmer (Martin et al. 2008), MSQuant (Mortensen et al. 2010), XPRESS (Han et al. 2006).

The algorithm used in this work with the cooperation of a partner UfZ Leipzig was MaxQuant, a quantitative proteomics software package designed for analyzing large mass spectrometric datasets. It is specifically aimed at high-resolution MS data. MaxQuant was developed at Max-Planck Institute for the NET framework and was written in the C# language. The interactive 3D data viewer was developed on the basis of DirectX. MaxQuant detects peaks, isotope clusters and stable amino acid isotope-labeled (SILAC) peptide pairs as three-dimensional objects in m/z, elution time and signal intensity space using correlation analysis and graph theory. It identifies and quantifies peptides and *in silico* composition of proteins. Each protein will be identified by at least two different peptides and the quantity is calculated by the integration of the peak. The software output is a table with identified proteins and peptides, which will be the basis for this thesis.

1.3.3 Proteomic Data Analysis

After the raw data analysis the list of all identified and quantified proteins was obtained from the cooperation partner UfZ Leipzig.

To build the basis for meaningful proteomic data analysis, various statistical function and methods must be applied, e.g. standard deviations, missing values imputation, statistical and significance analysis (*p*-value), gene ontology analysis etc.

One of the most used software in statistical computing and graphics is the R software (R Development Core Team 2012) and specifically the Bioconductor (R. C. Gentleman et al. 2004) for the analysis of high-throughput data.

Bioconductor is a free open source and open development software project built on the statistical freeware R software used for the analysis and comprehension of high-throughput gene expression data, but it can also be applied for another *omics* data as e.g. proteomics.

1.4 Research aims

This master thesis is integrated in the project: "Deciphering population dynamics as a Key for Process Optimization". The master thesis is divided in two phases: The first one is the proteomic analysis of *P. putida* KT 2440 cultured in standard conditions. Sarah Lieder established the standard conditions in the lab, and fermentation samples were analysed and sorted *via* FACS and the subpopulation proteome was analysed via UPLC+MS in the facilities of the cooperation partner UfZ Leipzig. The flow cytometry analysis and statistical analysis of the proteomic data is performed in this thesis and comprises the final part of the investigation of the standard conditions.

The second phase of the overall project is the investigation of different industrial relevant stress conditions. Therefore, as a second phase of this master project, tolerance levels of *P. putida* to α-Pinene, (R)-(+)-Limonene and Bisabolene are investigated, for a successful establishment of the experimental set-up of stress conditions on *P. putida* cultures. The overall project will contain fermentation scale solvent stress investigations on subpopulation level and the statistical analysis platform developed in the first part of the thesis will serve as tool for the analysis of the subpopulation proteomics under stressed conditions. The standard conditions analysed within this thesis will then serve as basis for comparison of non-stressed to stressed conditions to find optimization targets to reduce cell-to-cell variability and improved productivity. Figure 6 shows an overview of the thesis timeline and incorporation into the overall project.

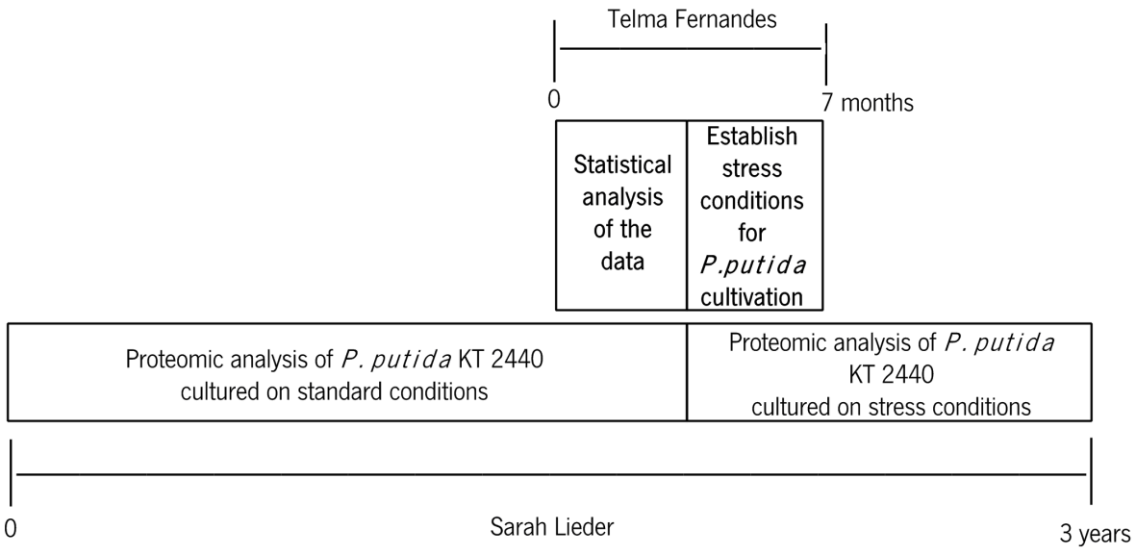

Figure 6 - Thesis timeline and incorporation into the overall project.

In summary, the main goals of this master thesis are a reliable data analysis of the *P. putida* subpopulation proteome to understand the differences in subpopulations at the same environmental condition comparing with the total population (Figure 7), but also at varying growth conditions. Additionally, it is tested if DNA content as a sorting criterion for subpopulation proteome analysis is giving insights into differences of subpopulations.

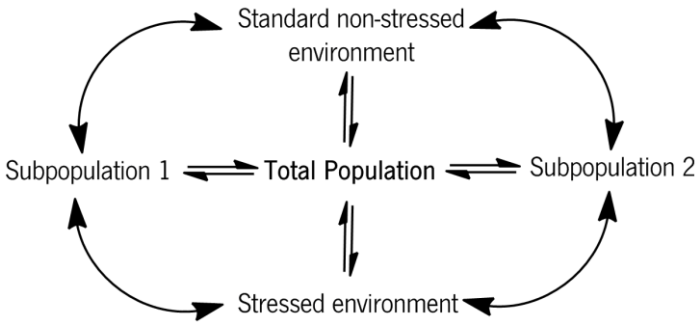

Figure 7 - Main goal of the overall project. Understand the differences between subpopulations at the same environmental conditions but also at varying growth conditions.

During the computational part of the thesis the MaxQuant software, with the cooperation of a partner, will be used to extract information of proteome data and the R software (R Development Core Team 2012) will be used for the statistical analysis to characterize the subpopulation split-up under various controlled steady-state conditions. A statistical analysis platform will be built for the analysis of standard condition data, but also for the analysis of the stress datasets, that will not be part of this project. Within the experimental part, called Shake Flask Experiments, preliminary experiments for the experimental set-up of stress investigation

experiments are carried out to find tolerance levels towards industrially interesting solvent compounds (Figure 8).

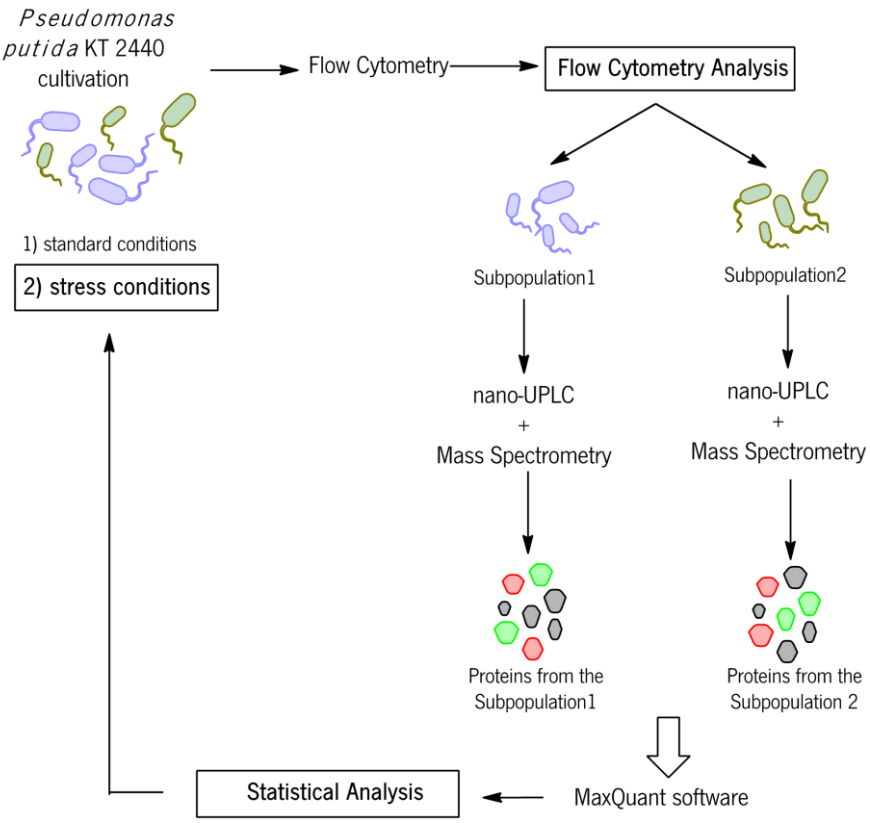

Figure 8 - Work methodology, showing the main steps from the overall project. The rectangles represent the work developed in this thesis. The first stage is the flow cytometry and statistical analysis of the proteomic data from the standard conditions and the second phase is the investigation of different relevant stress conditions. After set-up of the experiments for the stress conditions, these datasets will go through the same cycle of data analysis as the standard conditions.

2.Material and Methods

2.1. Flow Cytometry Analysis

Flow cytometry is one of the crucial tools available to investigate the behaviour of cell populations. R (R Development Core Team 2012) and Bioconductor (R. C. Gentleman et al. 2004) are used as a platform for analysis and visualization of the original complex multidimensional datasets originating from the flow cytometry output.

In this thesis the flow cytometry datasets of the standard continuous cultivation at different growth rates until wash out of the culture are analyzed. The intention is to find out if there are different subpopulations at varying growth rates and how these subpopulations evolve with increasing growth rate.

The packages chosen to perform the flow cytometry analysis and visualize this datasets were flowCore (Meur et al. 2013), flowViz (B Ellis et al. 2008) and latticeExtra (Deepayan Sarkar and Andrews 2012).

In a first step, the batch of FCS files was loaded into the R workspace to create a *flowSet*. A *flowSet* is a series of *flowFrames,* a meta-object that collects different types of information into a common identifier.

As a second step, the dataset is gated. Gating in flow cytometry is an important process for selecting populations of interest for further analysis and visualization. Noise, noncellular debris, cell debris, or clots of cells need to be removed using the *norm2filter* function. The gating was performed on forward scatter (FSC) and side scatter (SSC) levels to help ensure that the fluorescent measurements exhibit specificity for the target of interest.

The *xyplot* and *densityplot* were the functions used to perform the visualization.

The quantification of the subpopulations was performed using the Cyflogic v.1.2.1 (CyFlo Ltd, PL634, 20701 Turku, Finland).

2.2. Proteomic Data Analysis

Proteomic data analysis was performed on data sets of a continuous cultivation of *P. putida* KT2440 at standard conditions at various growth rates. Each sample was measured in

biological duplicates at three different growth rates: a very slow growth rate, μ=0.1 h^{-1}, a standard growth rate at which the continuous cultivations for stress conditions will be carried out in the future μ=0.2 h^{-1} and finally, the maximal growth rate μ=0.7 h^{-1}. Each sample was sorted into different subpopulations (for details see chapter 1.2.). Additionally the method needs to be checked by an internal standard sample. Here, the total population (TP) of the standard growth rate, both sorted (D02_TS) and not sorted (D02_TP) in the FACS serves as a standard for the method itself. Therefore, in total a number of 8 samples in duplicate were analyzed on proteomic level, generating 679 proteins displays.

The signal intensities of all peaks derived from each group in the mass spectrometer were exported to the MaxQuant software.

The proteomic data produced by MaxQuant were imported into R (version 2.15.1) (R Development Core Team 2012) to perform the statistical analysis. The workflow of statistical analysis is visualized in Figure 9.

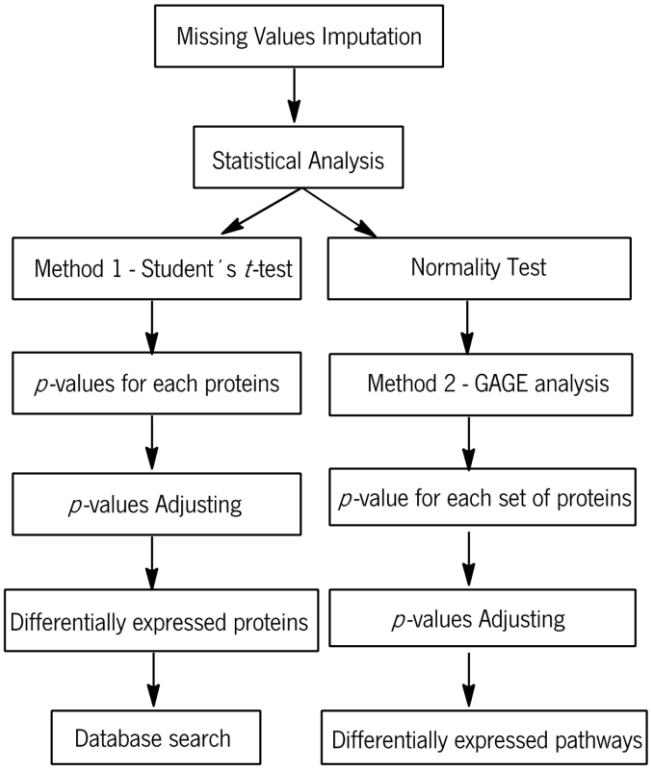

Figure 9 - Schematic workflow of statistical analysis.

2.2.1. Missing Values Imputation

Missing values are a frequent problem in large-scale profiling experiments, such as mass-spectrometry datasets. There are simple imputation methods, such as ignoring the missing data points or replacing them with zeros or average values (Aittokallio 2010). These methods may

ignore important data or do not take the correlation structure between the data into consideration resulting in harming the downstream analysis. To circumvent these consequences, the missing values imputation was carried out *via* a technique called *k*-nearest neighbors (KNN). KNN selects proteins with similar expression profiles to the missing protein to impute the missing value. It was proven to be a sensitive, fast and robust method for missing values estimation (Troyanskaya et al. 2001).

The imputation process of the KNN method is divided into two steps. In the first step, a set of proteins with a similar expression profile to the protein with a missing value is selected. The similarity was calculated via the Euclidean distance because it has been reported to be more accurate (Nguyen et al. 2004) and more sensitive to outliers (Troyanskaya et al. 2001). In the second step the missing values are predicted using a weighted average of the proteins selected in step 1.

KNN imputation was done using the Bioconductor in R computing language with the *impute* package (Trevor Hastie et al. 2005). Troyanskaya et al. (Troyanskaya et al. 2001) reported the best results for the number of neighbours (k) between 10 and 20. As a consequence of this, k=10 was chosen to perform the KNN method. Missing proteins were imputed by averaging the k nearest neighbour (non-missing) elements. In the event that 50 % neighbours were missing in a particular element, the overall sample mean for that block of proteins was used. In case more than 80 % of the entries were missing in any protein the program reported an error.

2.2.2. Testing the Normal distribution of the data

It is important to test the normality of the data before performing the statistical test to understand protein expression levels mainly because the proteomic data does not always follow a normal distribution (Albaum et al. 2011). This test was used only for the proteome datasets in method 2 (Proteins Set Analysis) because in method 1 (Student's t-test Analysis) the proteome datasets are only available in duplicates.

Therefore, the Shapiro-Wilk test (Shapiro and Wilk 1965) was used on the datasets in the R software (R Development Core Team 2012). A normally distributed population serves as the *null hypothesis* whereas the not-normally distribution serves as *alternative hypothesis.* A *p*-value < 0.05 was considered statistically significant. This data serves as input for further analyses.

2.2.3. Statistical Data Analysis - Method 1 (Student's t-test Analysis)

Significantly different protein expression between two different groups was determined using Student's *t*-tests for each protein. Unpaired analyses were used to compare the different samples with the control (D02_TS). To perform this statistical test, first a *null hypothesis* was stated, "protein not differentially expressed in relation to the standard sample" as well as the *alternative hypothesis*, "protein differentially expressed in relation to the standard sample". Based on the values measured and the normality assumption on the probability distribution of the data, either the null or the alternative hypothesis was maintained or rejected. A *p*-value < 0.05 was considered statistically significant.

2.2.4. Statistical Data Analysis - Method 2 (Proteins Set Analysis)

The proteome datasets are only available in duplicates due to experimental costs. Therefore, the normality of the proteomic dataset cannot be tested and the Student's *t*-test could be not the best test to estimate differentially expressed proteins. Therefore, an additional strategy was used for statistical analysis, a Gene Set Analysis (GSA). GSA is a powerful strategy to infer functional and mechanistic changes from high throughput microarray or sequencing data based on pathway knowledge.

GSA allows the identification of differentially expressed groups of proteins unlike the method described above that analyses individual differentially expressed proteins. The advantages of this approach are the encompassing of larger amount of biological information helping to interpret the data, the incorporation of biological pathway information, the facilitation to detect biologically important signals and more coherent results (Luo et al. 2009).

The GSA method chosen was a based parametric gene randomization procedure called Generally Applicable Gene-set Enrichment (GAGE). It can be used in datasets with different number of samples (Luo et al. 2009).

GAGE was implemented using the Bioconductor package *gage* (Luo et al. 2009) in the statistical computing language R.

The first step is the protein set preparation, in which the proteins are separated into two groups: the experimental sets and the set of pathways. The experimental data is the result of the MS analysis and the proteomic annotation. The set of pathways comes from extracted data from the database Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2012) (for details see chapter 2.2.6). Both of them are linked by the same ID system, the locus tag ID. The

second step is the differential expression test based on a one-on-one comparison between the two groups. For each pair of the experiment set the differential expression was calculated on a log based fold-change. The differential expressed set of proteins was calculated using two different methods to test the suitability of these methods for proteome data analysis: a parametric two-sample *t*-test and the non-parametric Mann–Whitney U test. In both statistical tests unpaired analyses were used to compare the different samples with the control sample (D02_TS).

Performing the GAGE analysis with a two sample *t*-tests, two assumptions on the log based fold changes of the target protein set and control sets are made: the data is approximately normally and independent.

In contrast, the Mann–Whitney U test tries to determine if two datasets differ significantly without the assumption of normality. The *null hypothesis* was defined as "the set of proteins (pathway) are not differentially expressed compared to the standard sample" and the alternative *hypothesis* as "the set of proteins (pathway) are differentially expressed in relation to the standard sample". The same *p*-value was used as in the Student's *t*-test ($p<0.05$).

2.2.5. Adjusting of *p*-values

To avoid a high number of false positives of differentially regulated proteins/pathways, the *p*-values were adjusted using an algorithm called "False Discovery Rate" (FDR) (Benjamini and Hochberg 1995). Take the *n* ordered *p*-values, $p_{(1)} \leq p_{(2)} \leq \ldots \leq p_{(n)}$", the FDR-adjusted *p*-values are given by equation 1.

$$\boldsymbol{p}_{(i)}(\boldsymbol{adjusted}) = \min_{k=i,\ldots,n} \left\{ \min\left(\frac{n}{k} \boldsymbol{p}_{(i)}, \mathbf{1} \right) \right\} (\boldsymbol{i} = \mathbf{1}, \ldots, \boldsymbol{n}) \qquad \text{Equation 1}$$

The adjustment procedures were implemented in the *p.adjust* method of the R-package "stats" (R Development Core Team 2012) and applied on the two methods described above. The result is a FDR analogue of the *p*-value, called *q*-value.

2.2.6. Proteomic Annotation

In order to facilitate the functional annotation and analysis of all proteins, DAVIDQuery (Day and Lisovich 2010), an R Package for retrieving data from DAVID (Database for Annotation, Visualization and Integrated Discovery) (Dennis et al. 2003) into R objects, was used. DAVID helps in visualization of high-throughput data and allows biological annotation (Dennis et al. 2003).

To get information about the locus tag, protein name, gene name, protein length, function, pathway and subcellular location etc., the protein ID given by the MS was converted into an Entrez gene ID and a Uniprot ID using the function *convertIDList* from DAVIDQuery (Day and Lisovich 2010). This procedure allowed access to information from the Pseudomonas Genome Database (Winsor et al. 2011) and the Universal Protein Research (Uniprot).

The search for information about the proteins differentially expressed in method 1 was directly carried out in different databases like KEGG (Kanehisa et al. 2012), Uniprot (Consortium 2012) and BIOCYC (Caspi et al. 2008).

For the method 2, extracted data from the database Kyoto Encyclopedia of Genes and Genomes (KEGG) (Kanehisa et al. 2012) was used to characterize the differentially expressed proteins. Differentially expressed proteins found using the locus tag ID were sorted into pathways for easier interpretation of the results.

2.2.7. Heat Map Plots

Visualization methods are very important for the interpretation of high-throughput data. Heat map plots were used to visualize the up or down regulated proteins from the differentially expressed proteins or pathways. The heat maps were drawn using the *heatmap.2* function in the *gplots* package (Gregory R. Warnes, 2013) in R (R Development Core Team 2012).

The rows represent the different proteins or pathways and the columns show the different samples and environmental conditions. The color coding is derived using the fold change values. Shades of blue represent up-regulation and shades of red represent down-regulation. The fold change is a measure used to compare two values of protein expression. Here, it is calculated by taking the logarithm of the ratio of the both expression values, the proteins (Ep) expression values of each sample and the expression values of the standard sample (Es) (Equation 2).

$$\boldsymbol{F = log_2 \frac{Ep}{Es}} \qquad \text{Equation 2}$$

2.3. Shake Flask Pulse Experiments

2.3.1. Organism and Media

The strain used in this study was *P. putida* KT2440 and it was grown in minimal medium (M12) (Table 1) supplemented with glucose (4.00 g/L), trace elements (Table 3) and phosphate buffer (2.00 g/L).

Stock solutions of trace elements, glucose and phosphate buffer were prepared and separately autoclaved to simplify media preparation.

The phosphate buffer stock solution was prepared dissolving 100 g of KH_2PO_4 in 1 L of distilled water. A pH of 6.7 was adjusted using KOH pills.

Table 2 - M12 solution composition and concentration in the final cultivation media.

Composition	Weight [g] per 1 L	Concentration (g/L)
$(NH_4)_2SO$	22	0.4
NaCl	0.2	0.02
$MgSO_4.7\ H_20$	4	0.4
$CaCl_2.2\ H_20$	0.4	0.04

Table 3 - Trace elements solution composition and concentration in the final cultivation media.

Composition	Weight [g] per 5 L	Concentration (g/L)
$ZnSO_4.7H_20$	1	0.002
$MnCl_2.4H_20$	0.5	0.001
$Na\text{-}Citrat.2H_20$	7.5	0.015
$CuSO_4.5H_20$	0.5	0.001
$NiCl_2.6\ H_20$	0.01	0.00002
$NaMoO_4.2.H_20$	0.015	0.00003
H_3BO_3	0.15	0.0003
$Fe(II)SO_4.7H_20$	5	0.01

2.3.2. *P. putida* KT 2440 cultivation

P. putida KT 2440 cultures were grown at 30° in 500 mL shake flasks in a horizontal rotary shaker at 180 rpm at a working volume of 50 ml. Cultures were grown for 4 hours, until they reached the exponential growth phase. At 4 hours of cultivation time, a pulse of a solvent with a specific concentration was given. Before the pulse, samples were taken every hour, after the pulse every 30 minutes, to monitor growth behaviour. The experiment was stopped after 7 hours and the biomass dry weight was determined (see section 2.3.3) The compounds used in this thesis were α-Pinene, (R)-(+) Limonene and Bisabolene at the concentrations of 0.1 %, 0.2 %, 0.3 %, 0.4 % and 0.5 % (v/v).

2.3.3. Biomass measurements

Biomass was monitored using two different methods.

The optical density can serve as a measure of the biomass concentration of the cultivation broth. This spectroscopic measurement is carried out at a wavelength of 600 nm in a Photometer Ultrospec 1100 pro (Amersham Biosciences) using polystyrol cuvettes (SARSTEDT). This measurement does not need a high volume of cultivation broth and therefore served as a biomass measurement to monitor the growth trajectory throughout the shake flask cultivation. To get a direct measurement of the biomass concentration, the biomass dry weight was measured at the end of each cultivation. Triplicates of aliquots of 10 mL of *P. putida* suspension were centrifuged and the pellet was dried at 95° C during 24 hours. The glass tubes with the dried pellets were cooled down in a vacuum desiccator for at least 2 hours, and then weighed. The weight difference of the empty glass tube to the glass tube with the dried pellet equals the biomass dry weight in 10 mL cultivation broth.

2.3.4. Statistical Analysis

The statistical analysis for the pulse experiments was performed using SigmaPlot 11.0 with the Holm-Sidak Multiple Comparison Test.

3.Results and Discussion

3.1. Flow Cytometry Analysis

Flow cytometry analysis was performed according to the procedure described in section 2.1. Samples for various growth rates from a chemostat experiment under standard conditions ranging from μ=0.1 h^{-1} to μ=0.8 h^{-1} were investigated.

Using a density plot to visualize the distribution of the events in the side scatter (SSC) and forward scatter (FSC) it is possible to realize that with increasing growth rate cells are getting bigger (Figure 10).

To get information on DNA content of the populations, the cells were stained with DAPI as mentioned in section 2.1. The fluorescence signal corresponds to the quantity of the DNA in each single cell of the population. To be able to identify subpopulations, a histogram view of the gated data was plotted (Figure 10).The quantification of the subpopulations was provided by Cyflogic v.1.2.1 (CyFlo Ltd, PL634, 20701 Turku, Finland) after plotting an histogram that display a single measurement parameter (fluorescence) on the x-axis and the number of events (cell count) on the y-axis (Table 4).

In the growth rates between μ=0.1 h^{-1} and μ=0.6 h^{-1} two different subpopulations of *P. putida* KT2440, that were called the subpopulation 1 (C1n) and subpopulation 2 (C2n) were found. The subpopulation 1 has one chromosome equivalent whereas the subpopulation 2 has two chromosome equivalents meaning that the subpopulation 1 is in the growth phase shortly after division and just about to start replicating and the subpopulation 2 is in the pre-division phase after finishing replication.

In the very fast growth rates (μ=0.7 h^{-1} and μ=0.8 h^{-1}) subpopulation 2 (C2n) and a subpopulation called CXn, where replication and division are uncoupled resulting in multiple chromosome containing cells can be found. The uncoupled cell cycle is found in some bacteria species growing at very fast growth rates.

Furthermore, the analysis of the DNA histogram (Figure 10) and Table 4 show that with increasing growth rate the subpopulation 2 increases, and subpopulation 1 decreases (μ=0.1 h^{-1}

to μ=0.6 h^{-1}) (Table 4). Subpopulation 1 vanishes at growth rates higher than μ=0.7 h^{-1} and the CXn subpopulation is evolving.

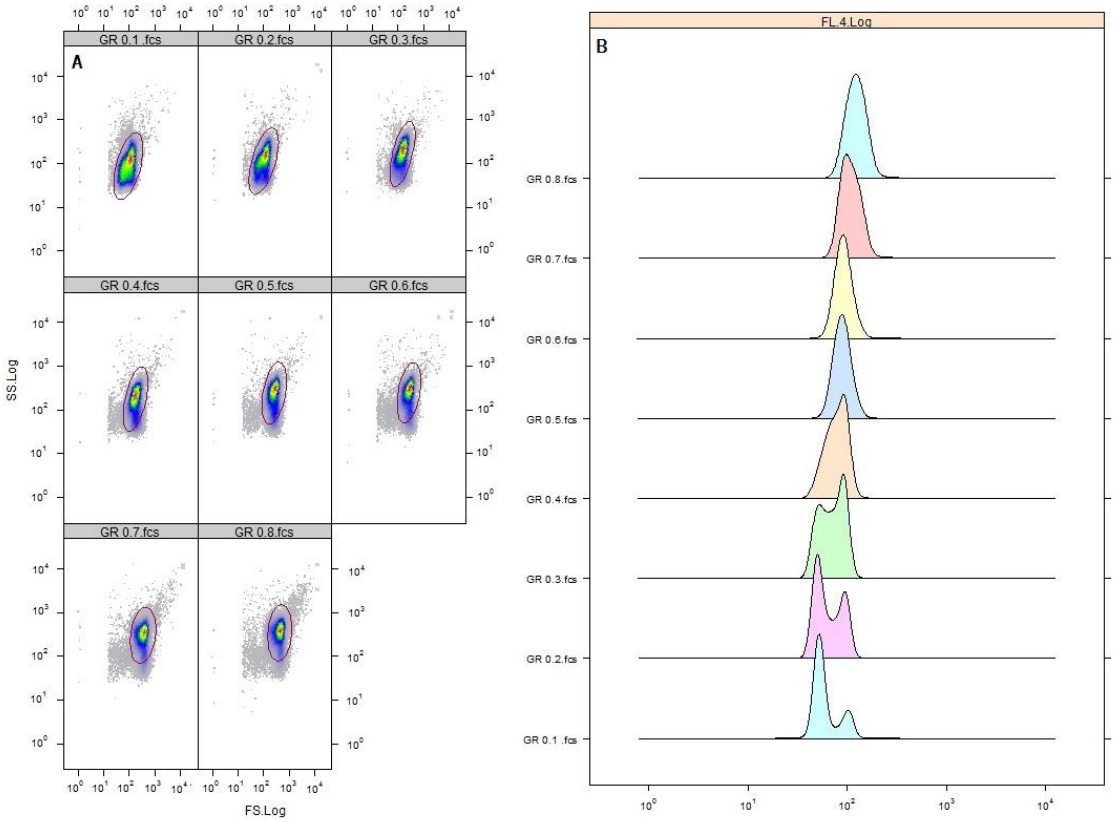

Figure 10 - (A) The plot represents the distribution of the events in the side scatter (SS.Log) and forward scatter (FS.Log) for the 8 samples. Inside the red ellipse are the cells of interest. The rest are debris according to the filter definition. (B) Flow cytometry characterization of *P. putida* KT 2440 subpopulations, the y-axis represents the samples and the x-axis represents the fluorescence signal captured for the channel 4 (FL.4.Log).

Table 4 - Quantification of different subpopulations (cell count) and their fluorescence (mean) in growth rate ranging μ =0.1 h^{-1} and μ =0.8 h^{-1}.

	Cell count (%)			Fluorescence FL-4 (mean)		
Samples	**C1n**	**C2n**	**CXn**	**C1n**	**C2n**	**CXn**
GR 0.1	77.4	22.6	-	34.5	61.5	-
GR 0.2	61.4	38.6	-	33.9	59.4	-
GR 0.3	50.0	50.0	-	32.9	56.2	-
GR 0.4	22.2	77.8	-	33.8	59.2	-
GR 0.5	17.9	82.1	-	43.8	58.7	-
GR 0.6	15.3	28.7	-	45.7	62.3	-
GR 0.7	-	41.8	58.2	-	58.4	81.3
GR 0.8	-	41.5	58.5	-	60.9	87.7

The reproducibility of the flow cytometry analysis was tested on duplicates of experimental datasets in this thesis. The results of the second experiment were consistent with the previous flow cytometry analysis (Figure 11, Table 5).

For complete analysis of the standard conditions and to get information about the proteomic content of different subpopulations at different growth rates, the very slow, μ=0.1 h^{-1}, and a very fast growth rate, μ=0.7 h^{-1}, were chosen for further investigation to get the proteomic composition of all 3 subpopulations found: C1, C2 and CX. As a reference for the slow and fast growth rates and to be able to compare the stress condition experiments that will be carried out in the future, the dilution rate μ=0.2 h^{-1} was also investigated.

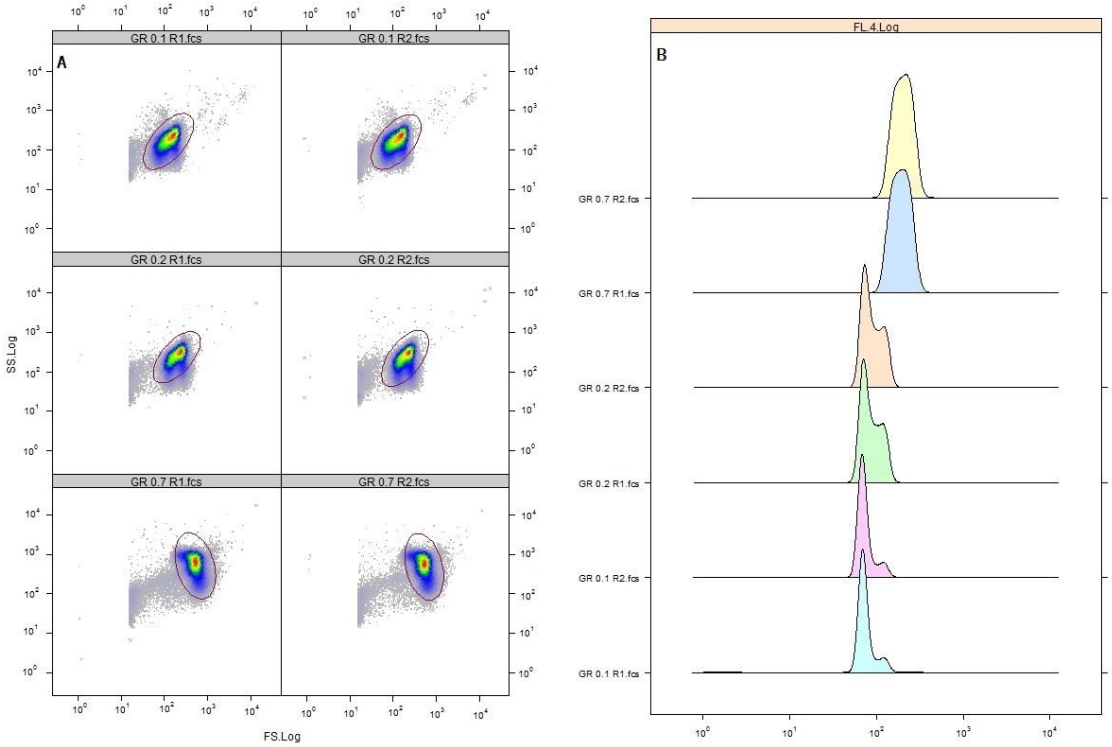

Figure 11 - (A) The plot represents the distribution of the events in the side scatter (SS.Log) and forward scatter (FS.Log) for the 6 samples. Inside the red ellipse are the cells of interest. The rest are debris according to the filter definition. (B) Flow cytometry characterization of *P. putida* KT 2440 subpopulations, the y-axis represents the samples and the x-axis represents the fluorescence signal captured for the channel 4 (FL.4.Log).

Table 5 - Quantification of different subpopulations (cell count) and their fluorescence (mean) in growth rate μ=0.1 h^{-1}, μ=0.2 h^{-1} and μ=0.7 h^{-1}.

	Cell count (%)			Fluorescence FL-4 (mean)		
Samples	**C1n**	**C2n**	**CXn**	**C1n**	**C2n**	**CXn**
GR 0.1 R1	82.7	17.3	-	44.6	76.7	-
GR 0.1 R2	81.3	18.7	-	44.1	77.5	-
GR 0.2 R1	66.7	33.3	-	45.2	78.8	-
GR 0.2 R2	65.7	34.3	-	46.7	81.3	-
GR 0.7 R1	1.91	42.3	55.8	45.1	85.5	142.7
GR 0.7 R2	1.28	41.7	56.9	47.6	90.1	153.4

3.2. Proteomic Data Analysis

3.2.1. Missing Values Imputation

31 % of missing values were found in the datasets of the 8 investigated samples (679 proteins). In the case of 190 proteins more than 50 % of the entries were missing. In these cases, as mentioned in section 2.2.1, the KNN method was performed to calculate the overall mean per sample.

3.2.2. Adjusting of *p*-values

The results of the analysis of differentially expressed proteins for the Student´s *t*-test and for the Protein Set Analysis method before and after the FDR adjustment can be found in the attachment.

3.2.3. KEGG analysis

The KEGG analysis (section 2.2.6) served as a preparation for the GAGE analysis. The KEGG file contains 2531 proteins, but only 1494 are unique proteins. On this file, just 407 were mapped and used for the output file for the protein set analysis. Having a total number of 679 proteins displays from the MS, the percentage of proteins analyzed in the GAGE method is approximately 60 %.

The table with the number of proteins distributed by 94 pathways that serves as a basis for further investigations in this thesis can be found in the attachment.

3.2.4. Normal distribution analysis

Testing the normal distribution is important to be able to choose a suitable statistical test for the application on ones dataset. For the test at least triplicate datasets need to be available. The proteomic data is only available in duplicates, as mentioned earlier. Therefore, this test was not used in the original data, but just for the data used for the GAGE analysis. The test was performed as explained in section 2.2.2.

For the normal distribution analysis 71 pathways were used from the original file. 23 pathways were removed, as they contained only one protein.

After the Shapiro-Wilk test was performed, 57 % of Non-Normal Distributed pathways were found in the 8 samples. This result confirms that proteomic data does not always follow a

normal distribution (Albaum et al. 2011). This result confirms the need for a normality test previous to proteomic statistical analysis.

3.2.5. Statistical Data Analysis – Method 1 (Student's t-test Analysis)

To understand the proteomic differences in between the subpopulations and in between different growth rates, it is crucial to understand the behavior of *P. putida* KT2440 as a population inside a bioreactor. The information of the statistical analysis of *P. putida* cultivated in a standard environment serves as a basis for further investigations on stress levels. The results will help to investigate *Pseudomonas'* ability to manage, to adapt and to recreate after being exposed to production-like stress conditions. The choice of a suitable statistical method to understand the proteomic data is crucial, as it can dramatically affect the results, in detail the selection of differentially expressed proteins. Two different methods were applied, as described in detail in section 2.2.3. In this chapter the results of the two methods are described and discussed, as well as the methods themselves are explored and evaluated.

The statistical data analysis was performed for a quantification of differences between the subpopulations at the different growth rates in comparison to the standard total population (D02_TS).

Using the Students *t*-test, at a slow growth rate of μ=0.1 h^{-1} (D01), 3 differentially expressed proteins were found in subpopulation 1 (D01_C1n) and 21 in subpopulation 2 (D01_C2n), respectively. It represents 0.4 % and 3.1 % of changed proteins in the subpopulation with one chromosome equivalent (D01_C1n) and in the subpopulation with two chromosome equivalents (D01_C2n), respectively when comparing to the standard sample (D02_TS).

At a growth rate of μ=0.2 h^{-1} (D02), 13 differentially expressed proteins for subpopulation 1 (D02_C1n), 16 for subpopulation 2 (D02_C2n) and 1 single protein for the control total population sample (D02_TP) were found. This corresponds to a difference of 1.9%, 2.4% and 0.1% between the samples D02_C1n, D02_C2n and D02_TP comparing with the standard sample (D02_TS), respectively.

At a growth rate of μ=0.7 h^{-1} (D07) 267 differentially expressed proteins were discovered in subpopulation 2 (D07_C2n) and 53 in the subpopulation with a multifold chromosome content X (D07_CXn). These results represent 39% of differentially expressed proteins of subpopulation 2 (D07_C2n) compared to the standard sample (D02_TS), and 7.8% of difference in protein expression between the multifold DNA content subpopulation (D07_CXn) and D02_TS.

In case of the third growth rate $\mu=0.7\ h^{-1}$, the different proteins are not directly looked up for detailed research in protein databases, as the search and interpretation of this high number of proteins would not lead to a targeted interpretation of the changes. Therefore, another test was performed called Generally Applicable Gene-set Enrichment (GAGE).

Discovering differentially expressed proteins is important to find optimization targets for e.g. improved stress tolerance or the production of an interesting compound. To draw conclusions for metabolic engineering strategies, it is not only important to find differentially expressed proteins, but also to quantify down or up-regulation. This information cannot be derived directly from the *t*-test analysis. Therefore, a heat map for visualization of fold-changes was constructed with all proteins differentially expressed for growth rates $\mu=0.1\ h^{-1}$ and $\mu=0.2\ h^{-1}$. Looking only at the fold change does not take the variance of the expression values measured into consideration. Therefore, it should be used only in combination with statistical methods (Tarca et al. 2008). For this reason, the Student *t*-test was applied in this thesis, before investigating fold-changes. In summary, proteins that were found to be differentially expressed in subpopulation 1 at a slow growth rate of $\mu=0.1\ h^{-1}$ were all down-regulated proteins. The same picture is found looking at subpopulation 2, given one up-regulated exception (protein 170). This trend continues in the other samples investigated ($\mu=0.2\ h^{-1}$, subpopulation 1 and 2). Most of the proteins found to be differentially expressed are down-regulated. In subpopulation 1, only three up-regulated proteins could be found (114, 328 and 589) whereas in subpopulation 2 five up-regulated proteins (285, 328, 362, 589 and 649) could be identified.

One protein was found to be down-regulated in the control sample (D02_TP). The visualization of these results can be seen in the heat map plot in Figure 12.

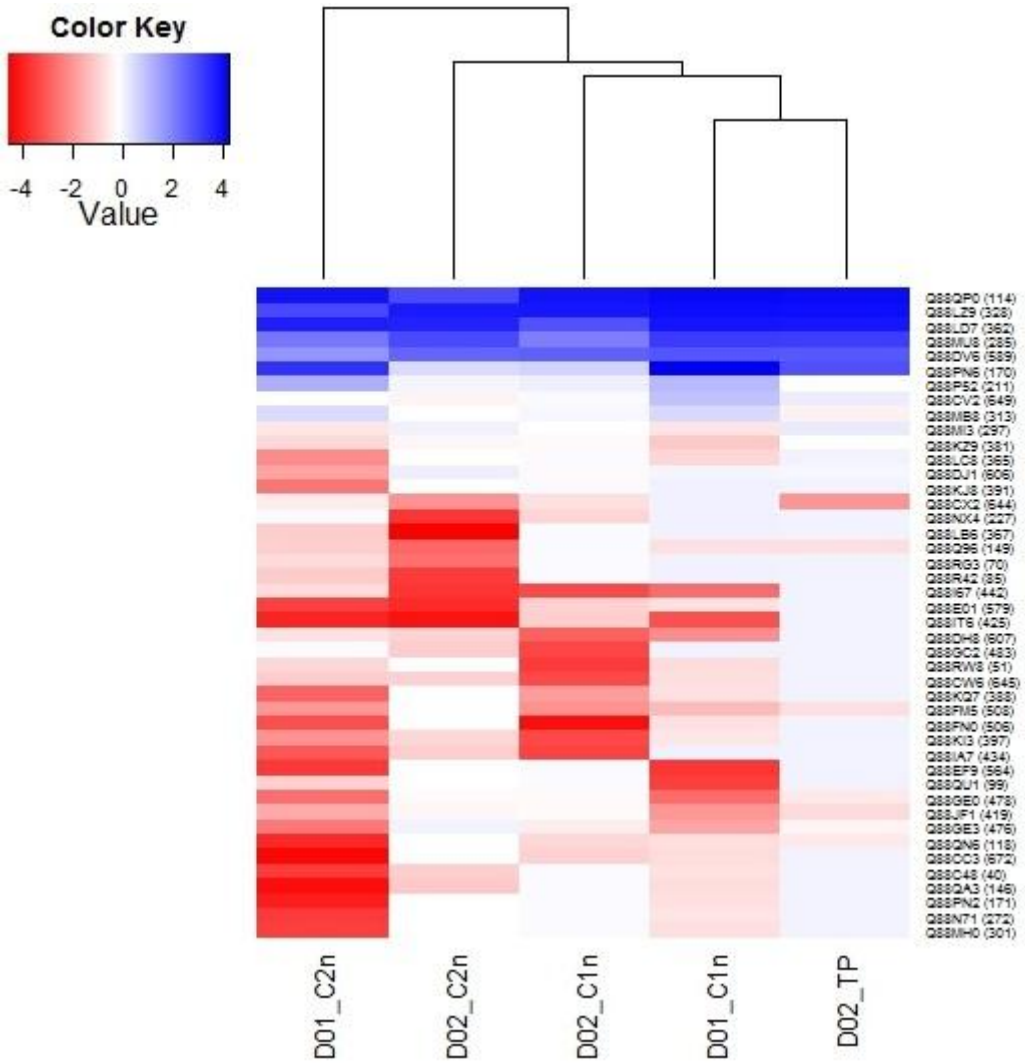

Figure 12 - Heat Map of up- and down-regulated differentially expressed proteins found in the samples D01_C1n, D01_C2n, D02_C1n, D02_C2n. D02_TP compared with the standard sample (D02_TS). Areas of red represent down-regulated proteins, while blue areas represent up-regulated proteins. Hierarchical clustering separates each sample. Proteins with Uniprot ID are indicated on the right of the figure.

Another important information that can be drawn out of this analysis is the proteins differentially expressed in common between the subpopulations of each growth rate. Common changes in proteome of both subpopulations compared to a different growth rate can be concluded to occur due to the change of the growth rate. One common protein at a growth rate of μ=0.1 h^{-1} and four common proteins at a growth rate of μ=0.2 h^{-1} could be detected. Detailed information on the proteins found, their biological function and the interpretation of these results can be found in the following paragraphs. The proteins found to be differentially expressed are summarized in the following set of tables and described and discussed in the following paragraphs.

At a growth rate of μ=0.1 h^{-1} the DNA topoisomerase was found to be down-regulated, (Table 6). DNA topoisomorase has the function to release the supercoiling and torsional tension of DNA introduced during the DNA replication and transcription by transiently cleaving and rejoining one strand of the DNA duplex. The decrease in protein quantity at a low growth rate fits the expectation. The cell adjusts the speed of replication and division to the slow growth rate. The other two proteins differentially expressed in this sample do not have an annotated function. Therefore, at this point, it is not possible to draw any conclusion about these proteins. In general it can be concluded that there is not a significant change in the proteome at a difference of growth rates from μ=0.1 h^{-1} to μ=0.2 h^{-1} between subpopulation 1 and the total standard population (D02_TS).

Table 6 - Information about the proteins differentially expressed in the D01_C1n sample.

ID	*q*-value	Uniprot	Entrez	Protein name	Length	Regulation
99	0,043	Q88QU1	1044108	Putative uncharacterized protein	522	Down regulated
381	0,043	Q88KZ9	1044994	DNA topoisomerase	869	Down regulated
564	0,003	Q88EF9	1041845	Nitroreductase family protein	185	Down regulated

In subpopulation 2 proteins were differentially down-regulated in nucleotide metabolism (40,146), amino-acid biosynthesis (171,211,297), ribosome pathway (118), glycolipid biosynthesis for membrane biogenesis (272, 301, 597) and the bacterial secretion system (434) (Table 7). These results can again be connected to the slow growth rate. Subpopulation 2 already finished replication of DNA in comparison to subpopulation 1. In comparison to the faster growing population at D=0.2 h^{-1} the amount of "building blocks" like amino-acids to integrate into proteins, nucleotides for the replication, ribosome enzymes for the translation machinery or lipids for the membranes are less needed.

Generally, the changes of the proteomic profile in subpopulation 2 at a growth rate of D=0.1 h^{-1} are bigger than in subpopulation 1 comparing to the standard population at a growth rate of D=0.2 h^{-1}.

Table 7 - Information about the proteins differentially expressed of the D01_C2n sample.

ID	*q*-value	Uniprot	Entrez	Protein name	Length	Regulation
40	0,010	Q88C48	1042138	Phosphoribosylaminoimidazole carboxylase; ATPase subunit	360	Down regulated
118	0,001	Q88QN6	1044199	30S ribosomal protein S10	103	Down regulated
146	0,003	Q88QA3	1044402	Adenine deaminase	315	Down regulated
170	0,042	Q88PN6	1044662	Cytochrome o ubiquinol oxidase; subunit I	672	Up-regulated
171	0,003	Q88PN2	1044669	Aminotransferase; class I	402	Down regulated
211	0,042	Q88P52	1044932	Arginine deiminase	417	Down regulated
272	0,007	Q88N71	1042966	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	303	Down regulated
297	0,005	Q88MI3	1045176	Aminotransferase; class I	398	Down regulated
301	0,004	Q88MH0	1044945	UDP-3-O-acylglucosamine N-acyltransferase	351	Down regulated
365	0,025	Q88LC8	1042987	P-47-related protein	410	Down regulated
388	0,004	Q88KQ7	1045106	Hydrolase; isochorismatase family	182	Down regulated
391	0,042	Q88KJ8	1045271	Putative uncharacterized protein	256	Down regulated
397	0,021	Q88KI3	1045292	Putative uncharacterized protein	259	Down regulated
434	0,031	Q88IA7	1043941	Putative uncharacterized protein	238	Down regulated
476	0,042	Q88GE3	1046188	Oxygen-independent Coproporphyrinogen III oxidase family protein	471	Down regulated
478	0,021	Q88GE0	1046194	Conserved domain protein	383	Down regulated
506	0,027	Q88FN0	1042130	Trehalose synthase; putative	1106	Down regulated
564	0,001	Q88EF9	1041845	Nitroreductase family protein	185	Down regulated
579	0,007	Q88E01	1045894	Methylmalonate semialdehyde dehydrogenase	508	Down regulated
606	0,019	Q88DJ1	1041846	SPFH domain/Band 7 family protein	284	Down regulated

672	0,001	Q88CC3	1042337	Aldehyde dehydrogenase family protein	496	Down regulated

Analysing the differentially expressed proteins of subpopulations at the growth rate of μ=0.2 h^{-1} gives information on the differences of proteome content between the two different subpopulations, but of course information on environmental influences cannot be drawn, as the growth rate of the standard population equals the growth rate of the investigated subpopulation. Interesting proteins down-regulated in subpopulation 1 (D02_C1n) were associated e.g. with the inhibition of the division (51) and the methylation of ribosome (442) (Table 8). It was found that in single chromosome content subpopulation, at standard growth rate, proteins that might suppress protein synthesis are down-regulated. A reason could be that in solely replicating cell population growth and replication machinery is prioritized. This can also be seen in ribosomal proteins important for translation process (114) and transcription factor (589) that are up-regulated. Additionally the protein 508 responsible for degradation of amino-acids is down-regulated. It might be interpreted in a way, that the cells need all available “building blocks” for new proteins and the protein responsible for amines and polyamines biosynthesis (645) in this stage purely focused on growth.

The differentially expressed proteins of subpopulation 1 (D02_C1n) are summarized in Table 8. Besides these proteins, in subpopulation 2 (D02_C2n), down-regulated proteins for amino-acid biosynthesis process (85, 227) were found additional to different up-regulated proteins related with transcription regulation (589) and amino-acid biosynthesis (285,649). This indicates changes in subpopulation with double chromosome content related to the upcoming division event in the cell. Replication of DNA is finished in these cells as well as growth in terms of volume gain. Therefore, different proteins regulating transcription and a reduced amino-acid biosynthesis process is fitting the expectation of the proteomic phenotype of this subpopulation.

Four proteins are commonly found in subpopulation 1 (D02_C1n) and subpopulation 2 (D02_C2n) (328, 442, 513, 589). As the standard population grows at the same growth rate, these proteins might be differentially expressed due to sampling methods or experimental procedures rather than biologically meaningful.

Table 8 - Information about the proteins differentially expressed of the D02_C1n sample.

ID	*q*.value	Uniprot	Entrez	Protein name	Length	Regulation
51	0,011	Q88RW8	1043472	tRNA uridine 5-carboxymethylaminomethyl modification enzyme MnmG	631	Down regulated
114	0,018	Q88QP0	1044192	30S ribosomal protein S12	123	Up-regulated
328	0,018	Q88LZ9	1043339	Phosphomannomutase	453	Up-regulated
397	0,004	Q88KI3	1045292	Putative uncharacterized protein	259	Down regulated
434	0,014	Q88IA7	1043941	Putative uncharacterized protein	238	Down regulated
442	0,018	Q88LV0	1043274	50S ribosomal protein L3 glutamine methyltransferase	302	Down regulated
483	0,018	Q88KY8	1045007	Soluble pyridine nucleotide transhydrogenase	464	Down regulated
506	0,018	Q88FN0	1042130	Trehalose synthase; putative	1106	Down regulated
508	0,049	Q88FM5	1042152	Isovaleryl-CoA dehydrogenase	424	Down regulated
513	0,031	Q88FI0	1042253	Isocitrate lyase	441	Down regulated
589	0,018	Q88DV6	1042024	N utilization substance protein A	493	Up-regulated
607	0,022	Q88DH8	1041837	Chaperone modulatory protein CbpM	101	Down regulated
645	0,002	Q88CW6	1043008	Choline dehydrogenase	565	Down regulated

Table 9 - Information about the proteins differentially expressed of the D02_C2n sample.

ID	*q*,value	Uniprot	Entrez	Protein name	Length	Regulation
70	0,040	Q88RG3	1043691	Toxin secretion ATP-binding protein	718	Down regulated
85	0,002	Q88R42	1043978	(EC 5.3.1.16) (Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase)	245	Down regulated
149	0,025	Q88Q96	1044413	PqiB family protein	766	Down regulated

227	0,002	Q88NX4	1044722	Ornithine carbamoyltransferase (OTCase) (EC 2.1.3.3)	306	Down regulated
285	0,040	Q88MU8	1045758	Homoserine dehydrogenase (EC 1.1.1.3)	434	Up-regulated
328	0,037	Q88LZ9	1043339	Phosphomannomutase	453	Up-regulated
362	0,040	Q88LD7	1043018	Putative uncharacterized protein	214	Up-regulated
367	0,004	Q88LB6	1042911	Membrane protein; putative	799	Down regulated
419	0,026	Q88JF1	1046177	5-methyltetrahydropteroyltriglutamate-homocysteine methyltransferase	344	Down regulated
425	0,021	Q88IT6	1042819	Delta-aminolevulinic acid dehydratase (EC 4.2.1.24)	324	Down regulated
442	0,021	Q88LV0	1043274	50S ribosomal protein L3 glutamine methyltransferase (L3 MTase) (EC 2.1.1.n15)	302	Down regulated
513	0,002	Q88FI0	1042253	Isocitrate lyase	441	Down regulated
579	0,004	Q88E01	1045894	Methylmalonate semialdehyde dehydrogenase	508	Down regulated
589	0,040	Q88DV6	1042024	N utilization substance protein A	493	Up-regulated
644	0,040	Q88CX2	1041641	Carboxyl-terminal protease	438	Down regulated
649	0,025	Q88CV2	1045947	3-dehydroquinate synthase (EC 4.2.3.4)	365	Up-regulated

The control sample (D02_TP) for assessing influences of the sorting procedure showed only one single differentially expressed protein, which did not appear in any other cases. This indicates that the sorting procedure does not interfere or deviate the proteome analysis of the subpopulations and it is a control for the reliability of the data. The protein found to be differentially expressed is the GTP pyrophosphokinase that it is a mediator of the stringent response that coordinates a variety of cellular activities in response to changes in nutritional abundance (Table - 10). It is possible that cells face some stress during the labelling and sorting

procedure, even though the metabolism is stopped *via* cold temperatures. As this protein was not found to be differentially expressed in any other sample and the *p*-value is very similar to *p*-value=0.05, it is still valid to conclude that the method and the procedures are valid for proteome measurements of subpopulations.

Table - 10 Information about the proteins differentially expressed of the D02_TP sample.

ID	*q*.value	Uniprot	Entrez	Protein name	Length	Regulation
313	0,049	Q88MB8	1044386	GTP pyrophosphokinase	746	Down regulated

3.2.6. Statistical Data Analysis – Method 2 (Proteins Set Analysis)

The Protein Set Analysis is based on the metabolic pathways derived from the KEGG analysis (section 3.2.3). Pathways with less than 2 proteins had to be removed for mathematical reasons, as at least two proteins need to be existent for application of the GAGE *t*-test or the Mann-Whitney U test analysis. These pathways were: D-Glutamine and D-glutamate metabolism, Glycerophospholipid metabolism, α-Linolenic acid metabolism, Dioxin degradation, Xylene degradation, Ethylbenzene degradation, Styrene degradation, Lipoic acid metabolism, Biosynthesis of ansamycins, Biosynthesis of vancomycin group antibiotics and the base excision repair metabolism. Removing these pathways resulted in 83 pathways containing 405 mapped proteins for the GAGE analysis.

Using the GAGE analysis performed in combination with the *t*-test, the ribosome pathway was found to be differentially expressed at the growth rate of μ=0.7 h^{-1} in both subpopulations (Figure 13). The calculated mean statistic is the mean of the individual statistics from multiple single array based gene set tests. Its absolute value measures the magnitude of the gene-set level changes. A positive mean t-statistics value indicates an up-regulation, and negative value a down-regulation.

The ribosome pathway is related with genetic information processing and translation. This result is in accordance to the expectation, as bacteria cultured at a very fast growth rate need additional ribosome machinery to perform the translation process fast.

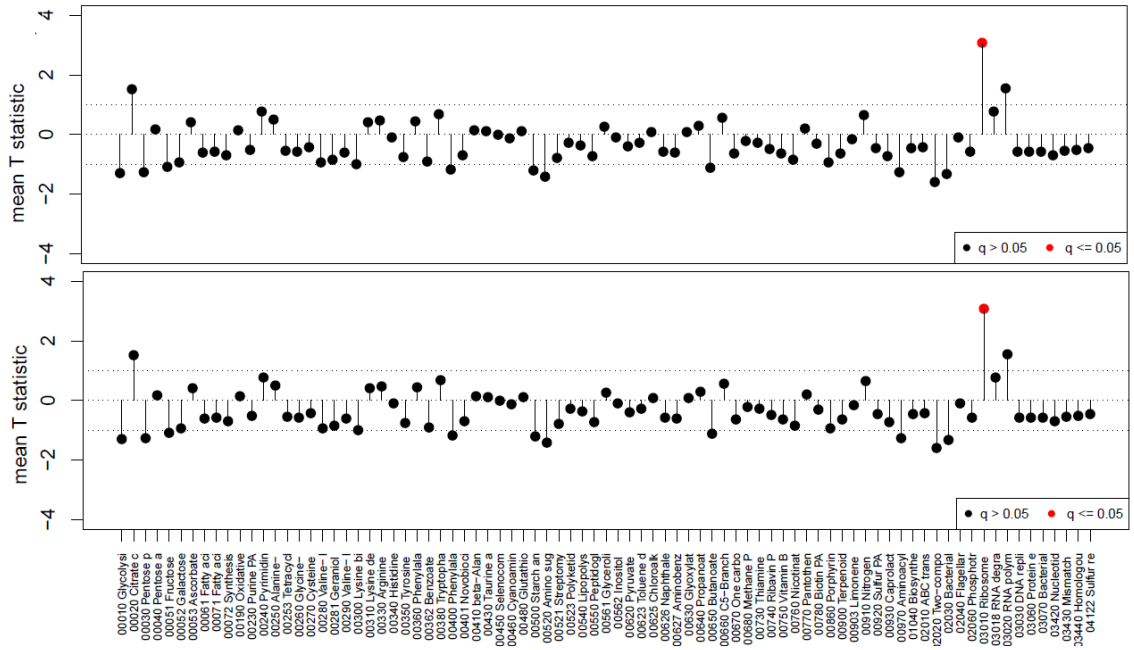

Figure 13 - GAGE *t*test analysis. The upper subfigure shows pathways differentially expressed in sample D07_ C2n and the bottom subfigure pathways that are differentially expressed in the sample D07_ CXn. The mean t statistics is the mean of the individual statistics from multiple single array based gene set tests. Red points show differentially expressed and black points commonly expressed pathways.

Combining the GAGE analysis with the Mann-Whitney U test, gives a partly different result compared to the *t*test statistics. The ribosome pathway is up-regulated in subpopulation 1 at the growth rate of μ=0.2 h^{-1} and both subpopulations at the faster growth rate, μ=0.7 h^{-1}, in contrast to the result achieved in combination with the *t*test. Looking into the *p*-values of the results, it becomes obvious that the difference in overexpression is much bigger at the growth rate of μ=0.7 h^{-1} in comparison to the growth rate of μ=0.2 h^{-1} (smaller *p*-values). This result is then again in accordance with the result of the *t*test analysis. The overexpression of the ribosome pathway in subpopulation 1 at the growth rate of μ=0.2h^{-1} might be interpreted as an artificial result or a higher sensitivity of the method.

In subpopulation 1 at the growth rate μ=0.2 h^{-1} besides the ribosome pathway, the RNA degradation were found to be up-regulated too (Figure 14).

RNA degradation is an important pathway for the expression of genetic information and the subpopulation 1 is in the process of starting replication and growth, whereas subpopulation 2 is in the state shortly before division. Therefore it can be assumed that the cells in subpopulation 1 need a higher amount of RNA degradation, as the RNA degradation pathway is responsible for adjusting the protein production in the cell to give metabolic changes that are required as cells grow and prepare for division (Arraiano et al. 2010).

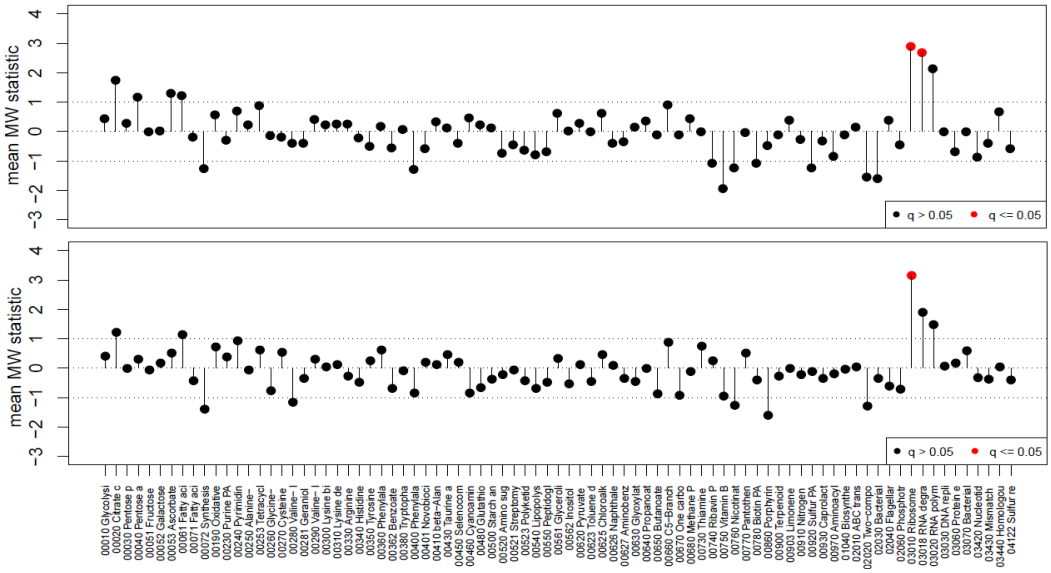

Figure 14 - GAGE Mann-Whitney U test analysis. The upper subfigure shows pathways differentially expressed in sample D02_ C1n and the bottom subfigure pathways that are differentially expressed in the sample D02_ TP. The mean MW statistics is the mean of the individual statistics from multiple single array based gene set tests. Red points show differentially expressed and black points commonly expressed pathways.

When looking into the control sample of the total population, one can find the ribosome pathway differentially expressed (Figure 14). As no pathway should be differentially expressed in this total population compared to the overall population after sorting, it is obvious, that the differentially expression of this pathway occurs biologically during the sorting procedure or is occurring as an artefact during the statistical analysis. Therefore, it will not be considered as biologically relevant in the discussion of differentially expressed pathways in the following paragraphs.

At the fast growth condition of μ=0.7 h^{-1}, the results are similar comparing with the GAGE analysis using the *t*-test. Subpopulation 2 and subpopulation X were found to show the ribosome pathway differentially expressed and up-regulated (Figure 15).

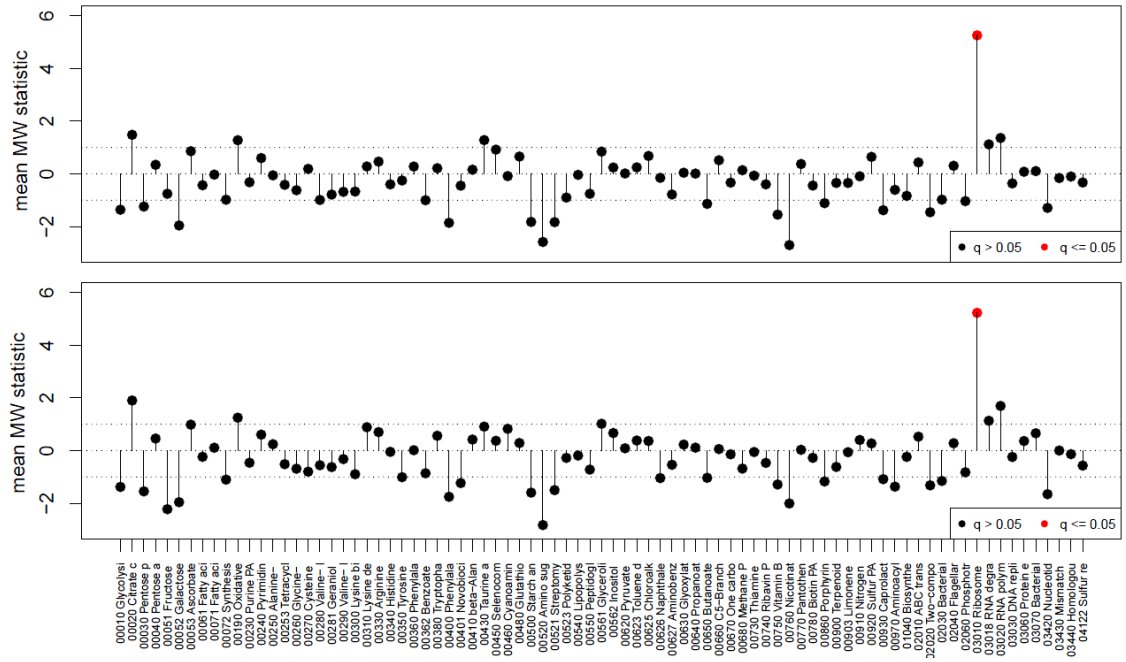

Figure 15 - GAGE Mann-Whitney U test analysis. The upper subfigure shows pathways differentially expressed in sample D07_ C2n and the bottom subfigure pathways that are differentially expressed in the sample D07_ CXn. The mean t statistics is the mean of the individual statistics from multiple single array based gene set tests. Red points show differentially expressed and black points commonly expressed pathways.

Several methods to perform statistical analysis on 'omics'-datasets, e.g. proteomic datasets, are known. Among all of the various methods, there is no perfect method to perform statistical analysis on proteomic data. Each method has special advantages but also specific limitations. Therefore, this thesis was also used to apply several methods in order to get a more complete and detailed insight into proteomic profiles, but also the influence and the power of the statistical methods applied.

The Student *t*-test analysis was chosen, as the *t*-test is a standard and powerful statistical tool commonly used in biology to find differences between samples.

The Student´s *t*-test has the advantage of using all the proteins found in the mass spectrometry analysis to perform the statistical analysis. Unfortunately, the direct search in the databases without clustering into groups or pathways is time consuming and the use and the value for interpretation might be discussable. In case of a high number of differentially expressed proteins, a direct search for each single protein in the database results in the same problem of interpretability and use of the analysis. One should be able to cluster and visualize datasets in a defined and reproducible standard operation procedure for giving deeper insights into systems understanding rather than piling up huge amount of information which are not interpretable.

The Student-*t*-test is a powerful technique to test differences between samples, but if the data does not follow the requirements for the *t*-test prerequisites or there are limited number of replicates some relevant information can be despised. The *t*-test requires that the observations within each group are normally distributed and the variances are equal in the two groups. These assumptions should be tested before performing statistical analysis.

The method of gene set analysis (GSA) was chosen because it became the first choice for extracting and explaining high-throughput data during the last years of experience in statistical handling of high throughput datasets in biological 'omics'-data interpretation (Khatri et al. 2012). Within the GSA methods, the GAGE test was the pathway method selected for this thesis, as it is the most frequently used GSA method and inferred statistically and biologically more relevant pathways (Luo et al. 2009).

Comparing the GAGE method directly to the Student *t*-test analysis the interpretation of the data is easier, as it clusters proteins into pathways. Therefore, the datasets are easier to overview. Especially in the case of samples at the high dilution rate μ=0.7 h^{-1} this method allowed interpretation of the dataset, whereas the Student *t*-test gave a high amount of differentially expressed single proteins, which could not be overviewed and therefore would be impossible to be interpreted biologically without sorting into pathways.

A disadvantage of combining the GAGE analysis with the *t*-test is that it cannot be assumed that proteomic data is following a Normal-distribution. Pathways might not be found to be differentially expressed because of their non-Gaussian distribution in this parametric test.

The GAGE analysis can also be performed in combination with a non-parametric test (Mann Whitney U test) for the statistical analysis despite having just duplicate values for each protein. The different results applying the parametric vs the non parametric test appear because of the different assessment of differentially expressed proteins/pathways in case of non-normally distributed datasets. As recommended by Albaum et al.(Albaum et al. 2011) one should first rely on the parametric test, but also take into consideration the non-parametric test secondly.

To be able to assess proteomic datasets in a better way, more biological replicates than duplicates of each protein would be helpful. However, FACS analysis in combination with proteome analysis is very time consuming, laborious and most of all very expensive.

Another problem of the proteome dataset analysis is the large amount of not annotated proteins. Many proteins cannot be taken into account in the study, as their function remains unknown. Here, the utmost importance of protein function research plays a role and future

investigations of protein function will improve the knowledge that can be drawn from proteomic datasets tremendously.

3.3. Shake Flask Pulse Experiments

Organic solvents are toxic to most microorganisms. Concerning processes in which organic solvents play a critical role, this leads to impasses at an industrial production level. Therefore, searching for solvent-tolerant microorganisms and exploring tolerance itself is important. The *Pseudomonas* genus has been studied to try to understand the mechanisms of organic solvent tolerance involving novel adaptations and some of them were described already as iterations at the level of the cell envelope structure. These suppress the effects of the solvents on membrane stability or limit their rate of diffusion into the cell (Heipieper et al. 1992). Furthermore, enhanced rates of phospholipids biosynthesis, speeding up repairing processes (Pinkart and D. C. White 1997) and active export systems that exclude solvents like toluene from the cell (Isken and De Bont 1996) were described.

Although these mechanisms have been studied in detail, to my knowledge, no studies have been made to test some interesting solvents as α-Pinene, (R)-(+) Limonene and Bisabolene. These compounds were found to be interesting building blocks for biofuel production (Peralta-Yahya et al. 2012). In this experimental part of the thesis the effects of these solvents on growth behavior and physiology of *P. putida* KT2440 were determined.

The pulse experiments with (R)-(+) Limonene showed that all concentrations tested (0.1, 0.2, 0.3, 0.5 % (v/v)) did not harm the growth behavior. In fact, the growth rate before the pulse with this compound in every concentration was smaller than the growth rate after the procedure (Figure 16). Only at a concentration of 0.1 % (v/v) the growth was practically the same. Table 11 summarizes the growth rates after and before pulse with the (R)-(+) Limonene tested concentrations.

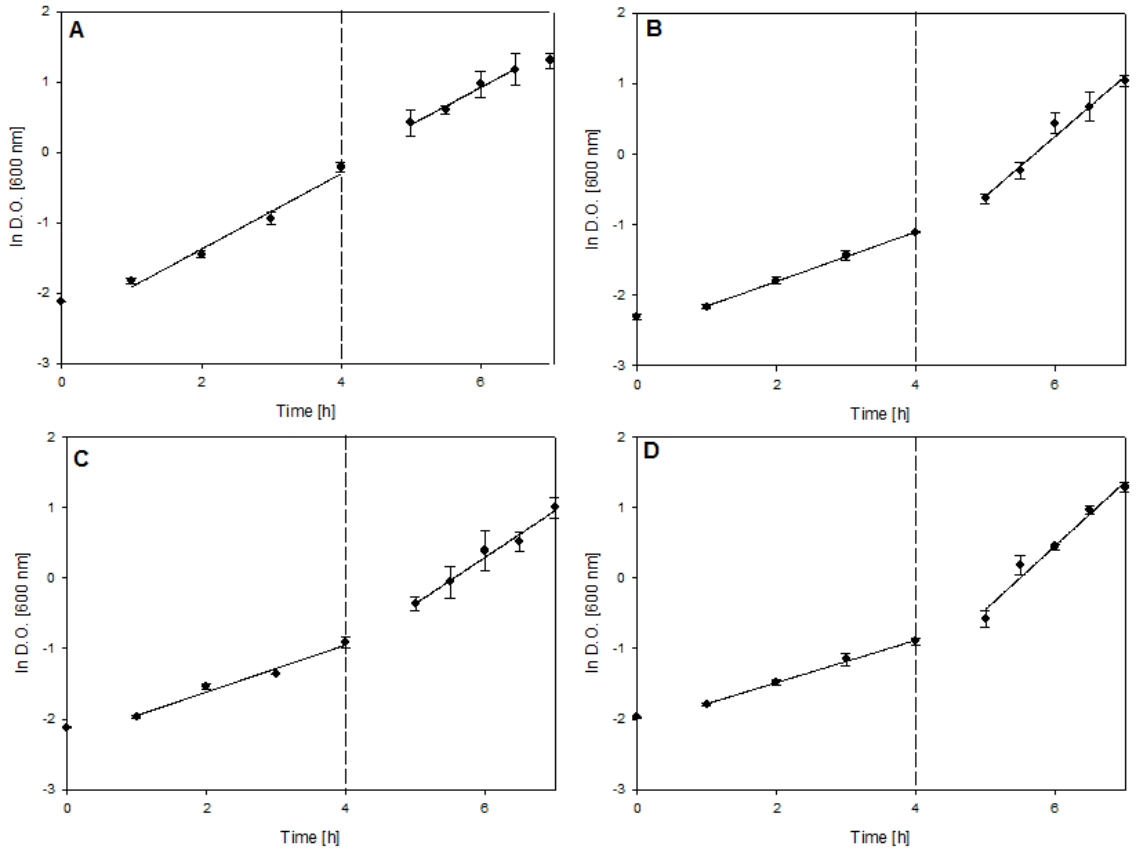

Figure 16 - Pulse experiments with (R)-(+) Limonene. Graphic determining the linear regression of the logarithmic optical density of the growth rates before and after each pulse. Plotted is the natural logarithm of the optical density at 600 nm against time. The dashed vertical line marks the time of the pulse. The slope of the regression line drawn in the first four hours and the slope of the regression line drawn after the hour 4 are the growth rates before and after the solvent pulse, respectively. (A) pulse with 0.1 % (v/v); (B) pulse with 0.2 % (v/v) ; (C) pulse with 0.3 % (v/v); (D) pulse with 0.5 % (v/v) of (R)-(+) Limonene.

Table 11 - Summary of the growth rates before and after the pulse procedure with (R)-(+)-Limonene.

(R)-(+)-Limonene	Growth rate μ(h^{-1})	
Concentration % (v/v)	Before Pulse	After Pulse
0.1	0.54	0.53
0.2	0.35	0.85
0.3	0.34	0.66
0.5	0.30	0.91

The increase in growth rate after the pulse might be due to the capability of *Pseudomonas* to cope very well with limonene in their surrounding and that they might even be able to metabolize this compound. When looking at the physiology and the shake flask cultivation visually, biofilm formation could be detected. Biofilm formation was found to interfere with the optical density measurements. Therefore, a strict conclusion on growth cannot be based on only optical density measurements and a second method for biomass build-up monitoring was used:

the biomass dry weight (section 2.3.3). For this method, for reproducible measurements, a high volume of V=30 mL liquid is needed. As the cultivation volume amounts 50 mL, only one biomass dry weight measurement at the end of the cultivation was conducted to compare the different cultivations among each other.

Looking at the biomass dry weight results, biomass was found significantly decreased compared to the standard sample (NO STRESS) in shaking flasks where the (R)-(+) Limonene was pulsed with 0.2 % (v/v), 0.3 % (v/v) and 0.5 % (v/v) (Figure 17). These differences are summarized in Table 12.

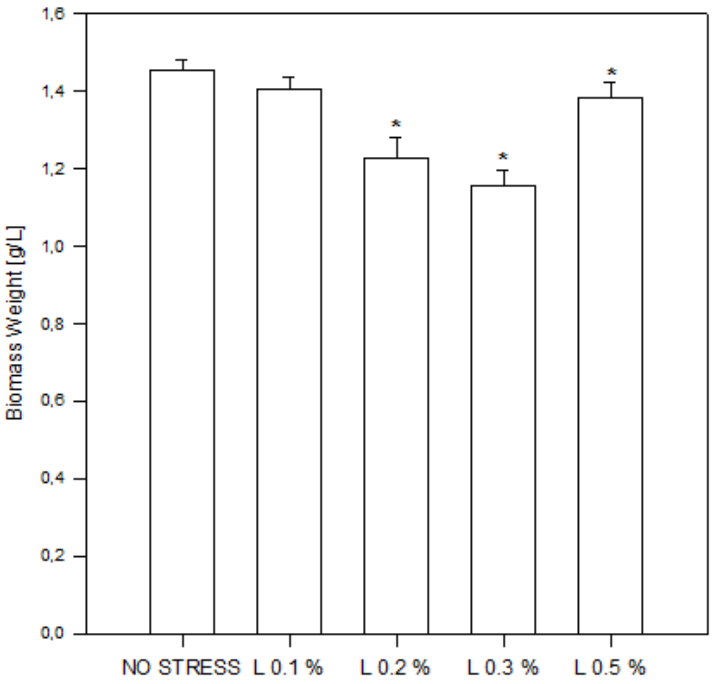

Figure 17 - Biomass dry weight. NO STRESS are the shaking flasks without a solvent pulse (standard sample). L 0.1 % represent the shaking flasks with 0.1 % (v/v) , L 0.2 % the shaking flasks with 0.2 % (v/v), L 0.3 % the shaking flasks with 0.3 % (v/v) and L 0.5 % the shaking flasks with 0.5 % (v/v) of (R)-(+) Limonene. * marks significantly different biomass dry weight results from the standard sample (NO STRESS).

Table 12 - Differences between biomass weight in each concentration tested of (R)-(+) Limonene and the standard sample (NO STRESS). The *p*-values from the statistical analysis can also be found in this table. ND means not significantly different compared to the standard sample (NO STRESS).

Concentration % (v/v)	*p*-value	Differences (%)
0.1	0.081	ND
0.2	<0.001	15.5
0.3	<0.001	20.6
0.5	0.010	4.40

At concentrations higher than 0.2% (v/v) of Limonene, biomass concentration was decreased. The decrease can be interpreted as a result of limited growth because of solvent

stress. Nevertheless, a general biomass growth could be detected. Therefore it also can be concluded, that these concentrations are not toxic but tolerated. Combining these results with the observed increased biofilm build-up with increasing limonene concentration, one can conclude, that growth is negatively influenced with increasing limonene concentration, but that at the same time, biofilm formation is increasing. That means that the cells tolerate limonene in general, but have to decrease their growth in order to be able to e.g. maintain tolerance mechanisms. Additionally, it might be possible that *Pseudomonas* is able to metabolize limonene, as the biofilm build-up is increasing with increasing limonene concentration, but biomass concentration is not decreasing dramatically at the highest investigated limonene concentration (0.5% v/v).

The pulse experiment with α-Pinene showed that all concentrations tested (0.1, 0.2, 0.3, 0.5 & (v/v)) made the growth rate increase after the pulse (Figure 19). The growth rates after and before pulse with α-Pinene are summarize in Table 13.

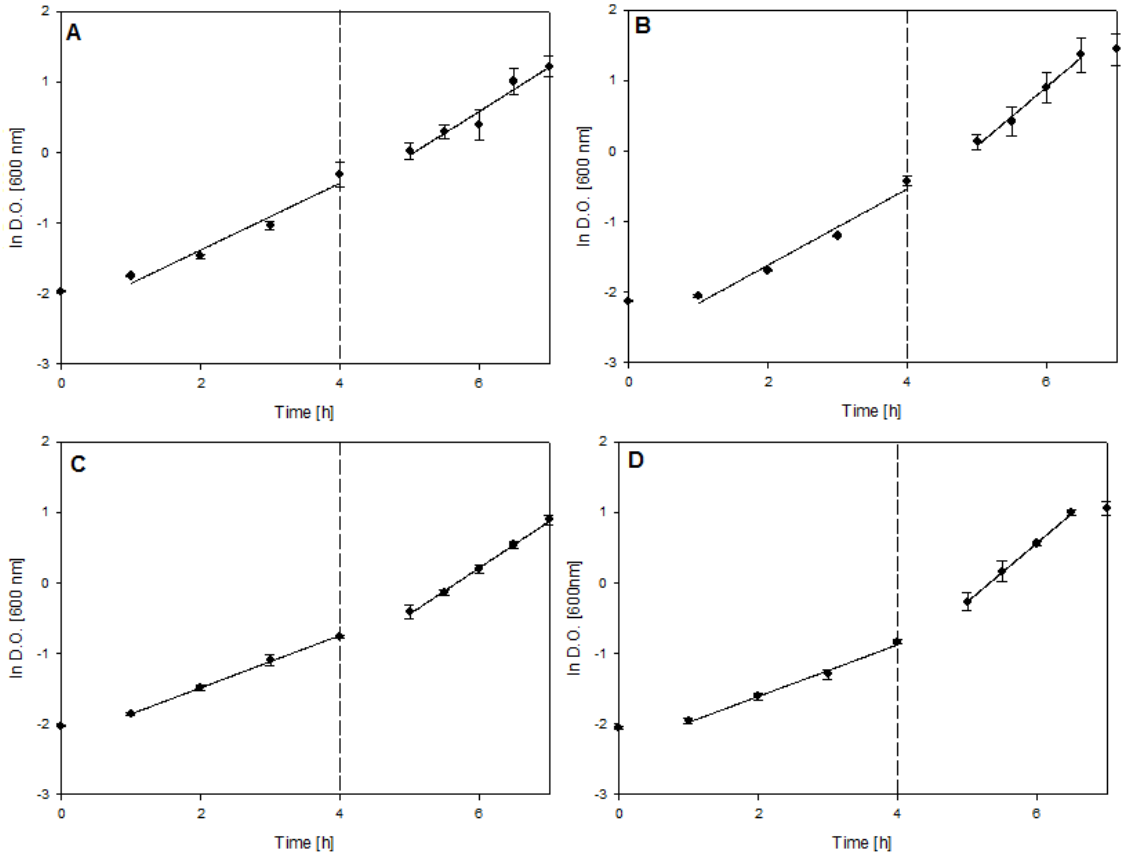

Figure 18 - Pulse experiments with α-Pinene. Graphic determining the linear regression of the logarithmic optical density of the growth rates before and after each pulse. Plotted is the natural logarithm of the optical density at 600 nm against time. The dashed vertical line marks the time of the pulse. The slope of the regression line drawn in the first four hours and the slope of the regression line drawn after the hour 4 are the growth rates before and after the solvent pulse, respectively. (A) pulse with 0.1 % (v/v); (B) pulse with 0.2 % (v/v) (C) pulse with 0.3 % (v/v); (D) pulse with 0.5 % (v/v) of α-Pinene .

Table 13 – Summarized growth rates before and after the pulse procedure with α-Pinene

α-Pinene	Growth rate (h^{-1})	
Concentration % (v/v)	Before Pulse	After Pulse
0.1	0.47	0.62
0.2	0.54	0.84
0.3	0.37	0.66
0.5	0.37	0.83

The biomass dry weight was found to be significantly decreased in all of the samples compared to the standard sample (NO STRESS). With increasing α-Pinene concentration, the biomass dry weight decreased (Figure 19). These differences are summarized in (Table 14).

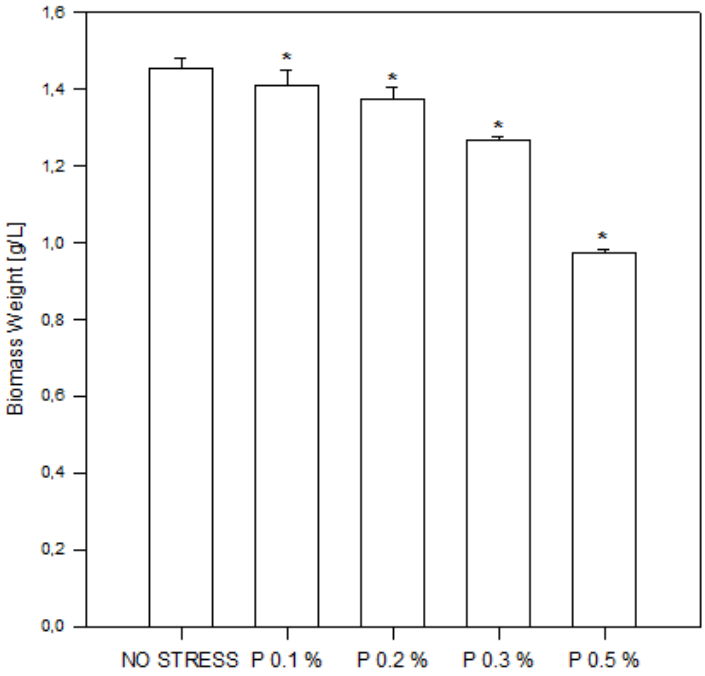

Figure 19 - Biomass dry weight. NO STRESS are the shaking flasks without a solvent pulse (standard sample). L 0.1 % represent the shaking flasks with 0.1 % (v/v) , L 0.2 % the shaking flasks with 0.2 % (v/v), L 0.3 % the shaking flasks with 0.3 % (v/v) and L 0.5 % the shaking flasks with 0.5 % (v/v) of α-Pinene. * marks significantly different biomass dry weight results from the standard sample (NO STRESS).

Table 14 - Differences between biomass weights in each concentration tested of α-Pinene and the standard sample (NO STRESS). The *p*-values from the statistical analysis can also be found in this table.

Concentration % (v/v)	*p*-value	Differences (%)
0.1	0.043	3.02
0.2	<0.001	5.61
0.3	<0.001	12.9
0.5	<0.010	33.1

It can be concluded that the cells can tolerate the tested concentrations, but that the stress is increasing with the increasing α-Pinene concentration, as biomass was decreased, while biofilm formation was increased.

When the experiments were conducted with Bisabolene, the optical density is increasing after the solvent pulsing in all tested concentrations (0.1, 0.2, 0.4, 0.5 % (v/v)) (Figure 20). These results are summarized on Table 15.

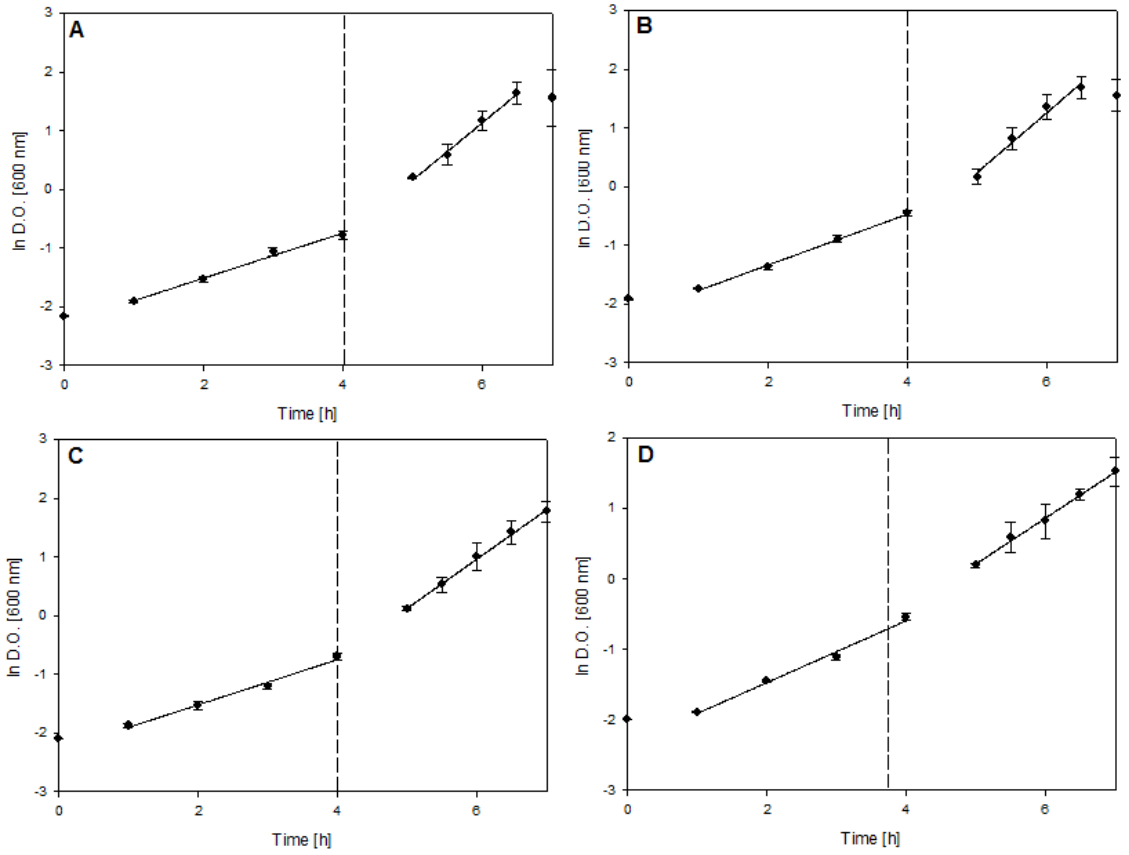

Figure 20 - Pulse experiments with Bisabolene. Graphic determining the linear regression of the logarithmic optical density of the growth rates before and after each pulse. Plotted is the natural logarithm of the optical density at 600 nm against time. The dashed vertical line marks the time of the pulse. The slope of the regression line drawn in the first four hours and the slope of the regression line drawn after the hour 4 are the growth rates before and after the solvent pulse respectively. (A) pulse with 0.1 % (v/v); (B) pulse with 0.2 % (v/v) (C) pulse with 0.3 % (v/v); (D) pulse with 0.5 % (v/v) of Bisabolene .

Table 15 - Summarized growth rates before and after the pulse procedure with Bisabolene.

Concentration % (v/v)	Before Pulse	After Pulse
0.1	0.38	0.98
0.2	0.43	1.02
0.3	0.38	0.84
0.5	0.44	0.65

The biomass dry weight was found to be significantly decreased in all of the samples compared to the standard sample (NO STRESS) (Figure 21). The difference between the samples and the *p*-values are summarized on Table 16.

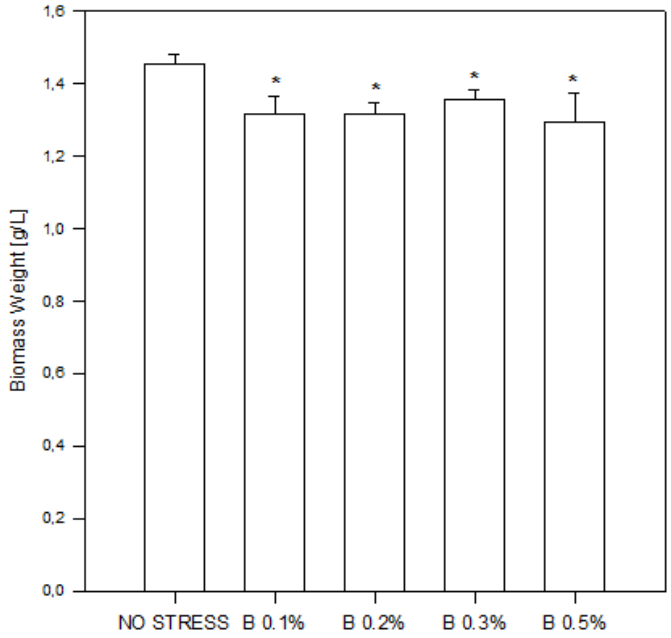

Figure 21 – Biomass dry weight. NO STRESS are the shaking flasks without a solvent pulse (standard sample). L 0.1 % represent the shaking flasks with 0.1 % (v/v) , L 0.2 % the shaking flasks with 0.2 % (v/v), L 0.3 % the shaking flasks with 0.3 % (v/v) and L 0.5 % the shaking flasks with 0.5 % (v/v) of Bisabolene. * marks significantly different biomass dry weight results from the standard sample (NO STRESS).

Table 16 - Differences between biomass weights in each concentration tested of Bisabolene and the standard sample (NO STRESS). The *p*-values from the statistical analysis can also be found in this table.

Concentration % (v/v)	*p*-value	Differences (%)
0.1	<0.001	9.60
0.2	<0.001	9.44
0.3	<0.003	6.60
0.5	<0.010	10.9

The decrease in biomass dry weight is generally smaller compared to the other solvents tested. This shows that *P. putida* KT 2440 can tolerate Bisabolene better. Nevertheless, biofilm formation was monitored. Combining this information, it might be possible that *P. putida* KT2440 is not only able to tolerate Bisabolene better compared to the other solvents tested, but also might be able to metabolize Bisabolene.

The results of the experimental part of this master thesis are that *P. putida* KT2440 can tolerate all the tested concentrations of solvents. α-Pinene limits growth the most, followed by a

small influence caused by (R)-(+)-Limonene. Bisabolene showed the least influence on growth. Biofilm formation was monitored in all cultivations, which compromised the validity of the growth rates calculated via optical density. Therefore, conclusions on influences on growth are solely based on biomass dry weight measurements. Even though growth was compromised, *P. putida* KT2440 tolerates the tested compounds (R)-(+) –Limonene, α-Pinene better than *Escherichia coli*. The tolerance of *E. coli* was tested with α-Pinene and Limonene and the maximum concentration tolerated was found to be 0.025 % for Limonene, 20 times less compared to the maximum concentration tested in this work (Dunlop et al. 2011). Looking at α-Pinene, the maximum concentration (0.5 % (v/v)) tested can also be tolerated by *E. coli,* but growth was compromised by 75 % (Dunlop et al. 2011).

Based on these results, it can be concluded, that *P. putida* KT2440 can tolerate solvents better than other industrial host systems, e.g. *E. coli* and therefore is a promising candidate for industrial processes involving solvent production or 2 phase cultivation systems. The metabolism of *P. putida* needs to be modified with metabolic engineering for the industrial application of the strain for production of these compounds, as a metabolization of products is contra productive. Nevertheless, we believe that cutting down metabolizing pathways might be an easier modification than implementation of various tolerance mechanisms in other hosts.

4.Conclusion

4.1 Flow Cytometry Analysis

Flow cytometry analysis showed the existence of a subpopulation with one chromosome equivalent (C1n) and a subpopulation with two chromosome equivalent (C2n) in the slow and normal grow rates (D=0.1 h^{-1} to D=0.6 h^{-1}). At very fast grow rates (D=0.7 h^{-1} to D=0.8 h^{-1}) it was possible to see a subpopulation with two chromosome equivalent (C2n) and an additional subpopulation with multiple chromosome equivalents (CXn). During the increase of growth rates subpopulation 1 decreases and subpopulation 2 increases until a D=0.6 h^{-1}. After that the subpopulation X increases and subpopulation 2 decreased. Subpopulation 1 having one chromosome equivalent can be interpreted as being in a growth stage shortly after division and just about to start replicating. Subpopulation 2 containing two chromosome equivalents is in the pre-division phase after finishing replication whereas the subpopulation X containing more than 2 chromosome equivalents is the result of a phenomenon called uncloupled cell cycle, where the division and replication happen simultaneously.

4.2 Proteomic Data Analysis

Proteomics experiments are complex and high-dimensional. Therefore, it is necessary to choose an appropriate statistical test to investigate if the data is independent, normally distributed, and has equal variance. However, to do that it is necessary to have multiple biological replicates. As experiments are time consuming and expensive, in conclusion statistical analysis and experiments have to be balanced to find a good compromise leading to good data quality.

The KNN imputation method is very important in proteomics MS resulting in more complete and coherent results as well as the use of adjustment of *p*-values methods, e.g. FDR. Without these methods the proteomic statistical analysis could be affected and might result in different proteomic profiles comparing with the reality.

The *t*-test is a powerful technique but in this thesis it cannot be concluded on its suitability because of the lack of replicates.

The GAGE analysis allows to have a good visualization of the data and permits the use of different statistical tests according to the data characteristics.

In the GAGE analysis either using the *t*-test or using Mann Whitney U test the ribosome pathway is over expressed because the bacteria cultured at a very fast growth rate need additional ribosome machinery to perform the translation process fast, as expected.

The uncharacterized proteins of *P. putida* KT2440 in the databases is a limiting step for the analysis of the proteomic data.

After the proteomics analysis there was no significant difference between subpopulation 1 and subpopulation 2 in growth rates μ=0.1 h^{-1} and μ=0.2 h^{-2}. On growth rate μ=0.7 h^{-1} the number of differentially expressed proteins is much bigger on subpopulation 2 (39%) comparing with subpopulation 1 (7.8%) meaning that subpopulation 2 needs more machinery to try to compensate the very fast growth rate.

In growth rates μ=0.1 h^{-1} and μ=0.2 h^{-2} mostly ribosomal, membrane and amino acid building block proteins were changed. That means that the strain is very stable metabolically, it does not shift during different growth rates, which is a good sign for a stable and robust metabolism and is one hint for the industrial suitability.

4.3 Shake Flask Experiments

P. putida KT2440 can tolerate α-Pinene and (R)-(+)-Limonene better than other industrial host systems, e.g. *E. coli* and therefore is a promising candidate for industrial processes involving solvent production or 2 phase cultivation systems. The tolerance of *E. coli* had been tested for α-Pinene and Limonene and the maximum concentration tolerated was found to be 0.025 % for Limonene, 20 times less compared to the maximum concentration tested in this work (Dunlop et al. 2011). Looking at α-Pinene, the maximum concentration (0.5 % (v/v)) tested can also be tolerated by *E. coli,* but growth was compromised by 75 % (Dunlop et al. 2011).

The metabolism of *P. putida* needs to be modified with metabolic engineering for the industrial application of the strain for production of these compounds, as a possible metabolization of products is contra productive. Nevertheless, we believe that cutting down metabolizing pathways might be an easier modification than implementation of various tolerance mechanisms in other hosts.

4.4 Future Work

For the future work it is necessary to establish the stress conditions in bioreactor scale according to the results of the shake flask pulse experiments. After this step the project will contain continuous fermentation of *P. putida* KT2440 in stress conditions and the statistical analysis platform developed in the first part of the thesis will serve as tool for the analysis of the subpopulation proteomics under stressed conditions.

The goal is to compare the proteomic results from the standard and stress conditions to find optimization targets to improve the production of interesting compounds in an industrial level.

References

Aittokallio, T. (2010). "Dealing with missing values in large-scale studies: microarray data imputation and beyond." *Briefings in bioinformatics*, 11(2), 253–64.

Albaum, S. P., Hahne, H., Otto, A., Haußmann, U., Becher, D., Poetsch, A., Goesmann, A., and Nattkemper, T. W. (2011). "A guide through the computational analysis of isotope-labeled mass spectrometry-based quantitative proteomics data: an application study." *Proteome science*, BioMed Central Ltd, 9(1), 30.

Arraiano, C. M., Andrade, J. M., Domingues, S., Guinote, I. B., Malecki, M., Matos, R. G., Moreira, R. N., Pobre, V., Reis, F. P., Saramago, M., Silva, I. J., and Viegas, S. C. (2010). "The critical role of RNA processing and degradation in the control of gene expression." *FEMS microbiology reviews*, 34(5), 883–923.

Benjamini, Y., and Hochberg, Y. (1995). "Controlling the False Discovery Rate: A Practical and Powerfull Approach to Multiple Testing." *Journal of the Royal Statistical Society*, 57(1), 289–300.

Caspi, R., Foerster, H., Fulcher, C. a, Kaipa, P., Krummenacker, M., Latendresse, M., Paley, S., Rhee, S. Y., Shearer, A. G., Tissier, C., Walk, T. C., Zhang, P., and Karp, P. D. (2008). "The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases." *Nucleic acids research*, 36(Database issue), D623–31.

Churchwell, M. I., Twaddle, N. C., Meeker, L. R., and Doerge, D. R. (2005). "Improving LC-MS sensitivity through increases in chromatographic performance: comparisons of UPLC-ES/MS/MS to HPLC-ES/MS/MS." *Journal of chromatography. B, Analytical technologies in the biomedical and life sciences*, 825(2), 134–43.

Consortium, T. U. (2012). "Reorganizing the protein space at the Universal Protein Resource (UniProt)." *Nucleic acids research*, 40(Database issue), D71–5.

Cox, J., and Mann, M. (2008). "MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification." *Nature biotechnology*, 26(12), 1367–72.

Day, R. S., and Lisovich, A. (2010). "The DAVIDQuery package in Bioconductor : Retrieving data from the DAVID Bioinformatics Resource into R." R package version 1.18.0.

Dennis, G., Sherman, B. T., Hosack, D. a, Yang, J., Gao, W., Lane, H. C., and Lempicki, R. a. (2003). "DAVID: Database for Annotation, Visualization, and Integrated Discovery." *Genome biology*, 4(5), P3.

Dunlop, M. J., Dossani, Z. Y., Szmidt, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., Hadi, M. Z., and Mukhopadhyay, A. (2011). "Engineering microbial biofuel tolerance and export using efflux pumps." *Molecular systems biology*, Nature Publishing Group, 7(487), 487.

El-aneed, A., Cohen, A., and Banoub, J. (2009). "Mass Spectrometry , Review of the Basics : Electrospray , MALDI , and Commonly Used Mass Analyzers." *Applied Spectroscopy Reviews*, 44, 210–230.

Ellis, B, Gentleman, R., Hahne, F., Sarkar, D, and Le Meur, N. (2008). "flowViz: Visualization for flow cytometry." *Bioinformatics (Oxford, England)*, R package version 1.22.0.

Espinosa-Urgel, M., Salido, A., and Ramos, J. (2000). "Genetic Analysis of Functions Involved in Adhesion of Pseudomonas putida to Seeds." *Journal of Bacteriology*, 182(9), 2363–2369.

"Federal Register." (1982). *Certified host-vector systems*, p17197.

Fritzsch, F. S. O., Dusny, C., Frick, O., and Schmid, Andreas. (2012). "Single-cell analysis in biotechnology, systems biology, and biocatalysis." *Annual review of chemical and biomolecular engineering*, 3, 129–55.

Galán, B., Díaz, E., and García, J. L. (2000). "Enhancing desulphurization by engineering a flavin reductase-encoding gene cassette in recombinant biocatalysts." *Environmental Microbiology*, 2(6), 687–94.

Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, Byron, Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J. Y. H., and Zhang, J. (2004). "Bioconductor: open software development for computational biology and bioinformatics." *Genome biology*, 5(10), R80.

Gregory R. Warnes, and Includes R source code and/or documentation contributed by: Ben Bolker, Lodewijk Bonebakker, Robert Gentleman, Wolfgang Huber Andy Liaw ,Thomas Lumley, Martin Maechler, Arni Magnusson, S. M. and M. S. (2013). "gplots: Various R programming tools for plotting data."

Han, D. K., Eng, J., Zhou, H., and Aebersold, R. (2006). "Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry." *Nature biotechnology*, 19(10), 946–951.

Hastie, Trevor, Tibshirani, Robert, Narasimhan, B., and Chu, G. (2005). "impute: impute: Imputation for microarray data." R package version 1.32.0.

Heipieper, H. J., Diefenbach, R., and Keweloh, H. (1992). "Conversion of cis unsaturated fatty acids to trans, a possible mechanism for the protection of phenol-degrading Pseudomonas putida P8 from substrate toxicity." *Applied and environmental microbiology*, 58(6), 1847–52.

Isken, S., and De Bont, J. a. (1996). “Active efflux of toluene in a solvent-resistant bacterium.” *Journal of bacteriology*, 178(20), 6056–8.

Jahn, M., Seifert, J., Von Bergen, M., Schmid, Andreas, Bühler, B., and Müller, S. (2012). “Subpopulation-proteomics in prokaryotic populations.” *Current Opinion in Biotechnology*, 1–9.

Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). “KEGG for integration and interpretation of large-scale molecular data sets.” *Nucleic acids research*, 40(Database issue), D109–14.

Khatri, P., Sirota, M., and Butte, A. J. (2012). “Ten Years of Pathway Analysis : Current Approaches and Outstanding Challenges.” *PLoS computational biology*, 8(2), 1–10.

Kojima, Y., Fujisama, H., Nakazawa, A., Nakazawa, T., Kanetsuna, F., Taniuchi, H., Nozaki, M., and Hayaishi, O. (1967). “Studies on Pyrocatechase.” *The Journal of Biological Chemistry*, 242(14), 3270–3278.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., et al. (2001). “Initial sequencing and analysis of the human genome.” *Nature*, 409(6822), 860–921.

Li, X.-J., Zhang, H., Ranish, J. a, and Aebersold, R. (2003). “Automated statistical analysis of protein abundance ratios from data generated by stable-isotope dilution and tandem mass spectrometry.” *Analytical chemistry*, 75(23), 6648–57.

Luo, W., Friedman, M. S., Shedden, K., Hankenson, K. D., and Woolf, P. J. (2009). “GAGE : generally applicable gene set enrichment for pathway analysis.” *BMC bioinformatics*, 10(161), 1–17.

Martin, D. B., Holzman, T., May, D., Peterson, A., Eastham, A., Eng, J., and McIntosh, M. (2008). “MRMer, an interactive open source and cross-platform system for data extraction and visualization of multiple reaction monitoring experiments.” *Molecular & cellular proteomics : MCP*, 7(11), 2270–8.

Martindale, W. (1996). *The Extra Pharmacopoeia*. (J. E. F. Reynold, ed.), London: Royal Pharmaceutical Society.

Meur, N. Le, Hahne, F., Ellis, B, and Haaland, P. (2013). “flowCore: data structures package for flow cytometry data.” R package version 1.24.2.

Mortensen, P., Gouw, J. W., Olsen, J. V, Ong, S., Rigbolt, K. T. G., Bunkenborg, J., Foster, L. J., Heck, A. J. R., Blagoev, B., Andersen, J. S., and Mann, M. (2010). “MSQuant , an Open Source Platform for Mass Spectrometry-Based Quantitative Proteomics research articles.” 393–403.

Müller, S., Harms, H., and Bley, T. (2010). “Origin and analysis of microbial population heterogeneity in bioprocesses.” *Current opinion in biotechnology*, 21, 100–13.

Muth, T., Keller, D., Puetz, S. M., Martens, L., Sickmann, A., and Boehm, A. M. (2010). "jTraqX: a free, platform independent tool for isobaric tag quantitation at the protein level." *Proteomics*, 10(6), 1223–5.

Nelson, K. E., Weinel, C., Paulsen, I. T., Dodson, R. J., Hilbert, H., Martins dos Santos, V a P, Fouts, D. E., Gill, S. R., Pop, M., Holmes, M., Brinkac, L., Beanan, M., DeBoy, R. T., Daugherty, S., Kolonay, J., Madupu, R., Nelson, W., White, O., Peterson, J., Khouri, H., Hance, I., Chris Lee, P., Holtzapple, E., Scanlan, D., Tran, K., Moazzez, A., Utterback, T., Rizzo, M., Lee, K., Kosack, D., Moestl, D., Wedler, H., Lauber, J., Stjepandic, D., Hoheisel, J., Straetz, M., Heim, S., Kiewitz, C., Eisen, J. a, Timmis, K N, Düsterhöft, A., Tümmler, B., and Fraser, C. M. (2002). "Complete genome sequence and comparative analysis of the metabolically versatile Pseudomonas putida KT2440." *Environmental Microbiology*, 4(12), 799–808.

Nguyen, D. V, Wang, N., and Carroll, R. J. (2004). "Evaluation of Missing Value Estimation for Microarray Data." *Journal of Data Science*, 2, 347–370.

Nicolaou, S. a, Gaida, S. M., and Papoutsakis, E. T. (2010). "A comparative view of metabolite and substrate stress and tolerance in microbial bioprocessing: From biofuels and chemicals, to biocatalysis and bioremediation." *Metabolic engineering*, Elsevier, 12(4), 307–31.

Nogales, J., Palsson, B. Ø., and Thiele, I. (2008). "A genome-scale metabolic reconstruction of Pseudomonas putida KT2440: iJN746 as a cell factory." *BMC systems biology*, 2, 79.

O'Sullivan, D. J., and O'Gara, F. (1992). "Traits of fluorescent Pseudomonas spp. involved in suppression of plant root pathogens." *Microbiological reviews*, 56(4), 662–76.

Olivera, E. R., Carnicero, D., Jodra, R., Miñambres, B., García, B., Abraham, G. a, Gallardo, A., Román, J. S., García, J. L., Naharro, G., and Luengo, J. M. (2001). "Genetically engineered Pseudomonas: a factory of new bioplastics with broad applications." *Environmental Microbiology*, 3(10), 612–8.

Park, S. K., Venable, J. D., Xu, T., and Iii, J. R. Y. (2008). "A quantitative analysis software tool for mass spectrometry – based proteomics." 5(4).

Peralta-Yahya, P. P., Zhang, F., Del Cardayre, S. B., and Keasling, J. D. (2012). "Microbial engineering for the production of advanced biofuels." *Nature*, 488(7411), 320–8.

Pinkart, H. C., and White, D. C. (1997). "Phospholipid biosynthesis and solvent tolerance in Pseudomonas putida strains." *Journal of bacteriology*, 179(13), 4219–26.

Puchałka, J., Oberhardt, M. a, Godinho, M., Bielecka, A., Regenhardt, D., Timmis, Kenneth N, Papin, J. a, and Martins dos Santos, Vítor a P. (2008). "Genome-scale reconstruction and analysis of the Pseudomonas putida KT2440 metabolic network facilitates applications in biotechnology." *PLoS computational biology*, 4(10), e1000210.

R Development Core Team. (2012). “R: A Language and Environment for Statistical Computing.” Vienna, Austria, ISBN 3–900051–07–0.

Saito, A., Nagasaki, M., Oyama, M., Kozuka-Hata, H., Semba, K., Sugano, S., Yamamoto, T., and Miyano, S. (2007). “AYUMS: an algorithm for completely automatic quantitation based on LC-MS/MS proteome data and its application to the analysis of signal transduction.” *BMC bioinformatics*, 8, 15.

Sarkar, Deepayan, and Andrews, F. (2012). “latticeExtra: Extra Graphical Utilities Based on Lattice.” R package version 0.6-24.

Schmid, A, Dordick, J. S., Hauer, B., Kiener, A., Wubbolts, M., and Witholt, B. (2001). “Industrial biocatalysis today and tomorrow.” *Nature*, 409(6817), 258–68.

Schweder, T., Krüger, E., Xu, B., Jürgen, B., Blomsten, G., Enfors, S. O., and Hecker, M. (1999). “Monitoring of genes that respond to process-related stress in large-scale bioprocesses.” *Biotechnology and bioengineering*, 65(2), 151–9.

Shadforth, I. P., Dunkley, T. P. J., Lilley, K. S., and Bessant, C. (2005). “i-Tracker: for quantitative proteomics using iTRAQ.” *BMC genomics*, 6, 145.

Shapiro, S., and Wilk, M. (1965). “An Analysis of Variance Test for Normality (Complete Samples).” *Biometrika*, 52(3/4), 591–611.

Sohn, S. B., Kim, T. Y., Park, J. M., and Lee, S. Y. (2010). “In silico genome-scale metabolic analysis of Pseudomonas putida KT2440 for polyhydroxyalkanoate synthesis, degradation of aromatics and anaerobic survival.” *Biotechnology journal*, 5(7), 739–50.

Stover, C. K., Pham, X. Q., Erwin, a L., Mizoguchi, S. D., Warrener, P., Hickey, M. J., Brinkman, F. S., Hufnagle, W. O., Kowalik, D. J., Lagrou, M., Garber, R. L., Goltry, L., Tolentino, E., Westbrock-Wadman, S., Yuan, Y., Brody, L. L., Coulter, S. N., Folger, K. R., Kas, A., Larbig, K., Lim, R., Smith, K., Spencer, D., Wong, G. K., Wu, Z., Paulsen, I. T., Reizer, J., Saier, M. H., Hancock, R. E., Lory, S., and Olson, M. V. (2000). “Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen.” *Nature*, 406(6799), 959–64.

Tarca, A., Romero, R., and Draghici, S. (2008). “Analysis of microarray experiments of gene expression profiling.” *Am J Obstet Gynecol*, 195(2), 373–388.

Timmis, Kenneth N, Steffan, R. J., and Unterman, R. (1994). “Designing Microorganisms for the Treatment of Toxic Wastes.” *Ann. Rev. Microbiol.*, 48, 525–57.

Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T, Tibshirani, R, Botstein, D., and Altman, R. B. (2001). “Missing value estimation methods for DNA microarrays.” *Bioinformatics (Oxford, England)*, 17(6), 520–5.

Tyers, M., and Mann, M. (2003). “From genomics to proteomics.” *Nature*, 422, 193–197.

Walsh, U. F., Morrissey, J. P., and O'Gara, F. (2001). "Pseudomonas for biocontrol of phytopathogens: from functional genomics to commercial exploitation." *Current Opinion in Biotechnology*, 12(3), 289–95.

Williams, P. A., and Murray, K. (1974). "Metabolism of Benzoate and the Methylbenzoates by Pseudomonas putida (arvilla) mt-2: Evidence for the Existence of a TOL Plasmid." *Journal of Bacteriology*, 120(1), 416–423.

Winsor, G. L., Lam, D. K. W., Fleming, L., Lo, R., Whiteside, M. D., Yu, N. Y., Hancock, R. E. W., and Brinkman, F. S. L. (2011). "Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes." *Nucleic acids research*, 39(Database issue), D596–600.

Witzig, T. E., Timm, M., Stenson, M., Svingen, P. A., and Kaufmann, S. H. (2000). "Induction of Apoptosis in Malignant B Cells by Phenylbutyrate or Phenylacetate in Combination with Chemotherapeutic Agents." 681–692.

Zambrano-Bigiarini, M. (2013). "hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series." R package version 0.3–6.

Zeyer, J., Lehrbach, P. R., and Timmis, K N. (1985). "Use of cloned genes of Pseudomonas TOL plasmid to effect biotransformation of benzoates to cis-dihydrodiols and catechols by Escherichia coli Cells." *Appl. Environ. Microbiol.*, 50(6), 1409–1413.

Attachments

Table 17 - Number of proteins for each pathway for the GAGE analysis.-

Pathways	Number of proteins
03010 Ribosome PATH:ppu03010	33
00020 Citrate cycle TCA cycle PATH:ppu00020	21
00240 Pyrimidine metabolism PATH:ppu00240	18
03018 RNA degradation PATH:ppu03018	11
03020 RNA polymerase PATH:ppu03020	3
00620 Pyruvate metabolism PATH:ppu00620	26
00230 Purine metabolism PATH:ppu00230	37
02010 ABC transporters PATH:ppu02010	16
00460 Cyanoamino acid metabolism PATH:ppu00460	2
00561 Glycerolipid metabolism PATH:ppu00561	3
00430 Taurine and hypotaurine metabolism PATH:ppu00430	2
00680 Methane metabolism PATH:ppu00680	16
01051 Biosynthesis of ansamycins	1
00061 Fatty acid biosynthesis PATH:ppu00061	7
00330 Arginine and proline metabolism PATH:ppu00330	26
00053 Ascorbate and aldarate metabolism PATH:ppu00053	2
00480 Glutathione metabolism PATH:ppu00480	8
00360 Phenylalanine metabolism PATH:ppu00360	5
00625 Chloroalkane and chloroalkene degradation PATH:ppu00625	5
00640 Propanoate metabolism PATH:ppu00640	23
00010 Glycolysis / Gluconeogenesis PATH:ppu00010	24
00410 beta-Alanine metabolism PATH:ppu00410	12
00380 Tryptophan metabolism PATH:ppu00380	11
00660 C5-Branched dibasic acid metabolism PATH:ppu00660	6
00562 Inositol phosphate metabolism PATH:ppu00562	2
00030 Pentose phosphate pathway PATH:ppu00030	15
00190 Oxidative phosphorylation PATH:ppu00190	24
00253 Tetracycline biosynthesis	3
00643 Styrene degradation PATH:ppu00643	1
00040 Pentose and glucuronate interconversions PATH:ppu00040	3
00592 alpha-Linolenic acid metabolism PATH:ppu00592	1
00642 Ethylbenzene degradation	1
00310 Lysine degradation PATH:ppu00310	12
03440 Homologous recombination PATH:ppu03440	6
00270 Cysteine and methionine metabolism PATH:ppu00270	14

00051 Fructose and mannose metabolism PATH:ppu00051	6
00910 Nitrogen metabolism PATH:ppu00910	13
00350 Tyrosine metabolism PATH:ppu00350	4
00770 Pantothenate and CoA biosynthesis PATH:ppu00770	9
03060 Protein export PATH:ppu03060	7
03030 DNA replication PATH:ppu03030	4
00730 Thiamine metabolism PATH:ppu00730	2
00290 Valine- leucine and isoleucine biosynthesis PATH:ppu00290	12
00970 Aminoacyl-tRNA biosynthesis PATH:ppu00970	23
00564 Glycerophospholipid metabolism PATH:ppu00564	1
00630 Glyoxylate and dicarboxylate metabolism PATH:ppu00630	9
00250 Alanine- aspartate and glutamate metabolism PATH:ppu00250	19
00401 Novobiocin biosynthesis PATH:ppu00401	3
00623 Toluene degradation PATH:ppu00623	3
00740 Riboflavin metabolism PATH:ppu00740	2
00300 Lysine biosynthesis PATH:ppu00300	9
00260 Glycine- serine and threonine metabolism PATH:ppu00260	19
03070 Bacterial secretion system PATH:ppu03070	11
00920 Sulfur metabolism PATH:ppu00920	7
00622 Xylene degradation PATH:ppu00622	1
00621 Dioxin degradation PATH:ppu00621	1
00450 Selenocompound metabolism PATH:ppu00450	5
00540 Lipopolysaccharide biosynthesis PATH:ppu00540	3
03430 Mismatch repair PATH:ppu03430	4
00626 Naphthalene degradation PATH:ppu00626	2
00280 Valine- leucine and isoleucine degradation PATH:ppu00280	20
02040 Flagellar assembly PATH:ppu02040	3
00071 Fatty acid metabolism PATH:ppu00071	14
00670 One carbon pool by folate PATH:ppu00670	6
00340 Histidine metabolism PATH:ppu00340	6
00500 Starch and sucrose metabolism PATH:ppu00500	10
02060 Phosphotransferase system PTS PATH:ppu02060	2
03410 Base excision repair PATH:ppu03410	1
00523 Polyketide sugar unit biosynthesis PATH:ppu00523	2
03420 Nucleotide excision repair PATH:ppu03420	2
00520 Amino sugar and nucleotide sugar metabolism PATH:ppu00520	10
01055 Biosynthesis of vancomycin group antibiotics	1
00650 Butanoate metabolism PATH:ppu00650	14
00281 Geraniol degradation PATH:ppu00281	7
02030 Bacterial chemotaxis PATH:ppu02030	18
00627 Aminobenzoate degradation PATH:ppu00627	3
00362 Benzoate degradation PATH:ppu00362	7
00903 Limonene and pinene degradation PATH:ppu00903	5
00521 Streptomycin biosynthesis PATH:ppu00521	5

00052 Galactose metabolism PATH:ppu00052	4
00930 Caprolactam degradation PATH:ppu00930	3
01040 Biosynthesis of unsaturated fatty acids PATH:ppu01040	2
00400 Phenylalanine- tyrosine and tryptophan biosynthesis PATH:ppu00400	10
00785 Lipoic acid metabolism PATH:ppu00785	1
00780 Biotin metabolism PATH:ppu00780	3
00550 Peptidoglycan biosynthesis PATH:ppu00550	5
00860 Porphyrin and chlorophyll metabolism PATH:ppu00860	6
00750 Vitamin B6 metabolism PATH:ppu00750	2
00072 Synthesis and degradation of ketone bodies PATH:ppu00072	3
00471 D-Glutamine and D-glutamate metabolism PATH:ppu00471	1
00760 Nicotinate and nicotinamide metabolism PATH:ppu00760	3
00900 Terpenoid backbone biosynthesis PATH:ppu00900	5
04122 Sulfur relay system PATH:ppu04122	3
02020 Two-component system PATH:ppu02020	30

Table 18 - Number of *p*-values before and after the *p-values* adjustment for each sample (method 1). Additionally, false positive findings of differentially expressed genes are quantified.

	Proteins differentially expressed		
Samples	Before FDR	After FDR	False Positives (%)
D01_C1n	105	3	97.1
D01_C2n	131	21	84.0
D02_C1n	68	13	80.9
D02_C2n	48	16	66.7
D02_TP	55	1	98.2
D07_C2n	343	267	22.2
D07_Cxn	322	52	83.9
Total	1072	373	65.2

Table 19 - Number of *p*-values before and after the *p-values* adjustment for each sample (method 2 using the *t*-test). Additionally, false positive findings of differentially expressed genes are quantified.

	Proteins differentially expressed		
Samples	Before FDR	After FDR	False Positives (%)
D01_C1n	1	0	100
D01_C2n	1	0	100
D02_C1n	0	0	0
D02_C2n	0	0	0
D02_TP	2	0	100
D07_C2n	3	1	66.7
D07_Cxn	3	1	66.7
Total	10	2	80.0

Table 20 - Number of *p*-values before and after the *p-values* adjustment for each sample (method 2 using the Mann–Whitney *U* test). Additionally, false positive findings of differentially expressed genes are quantified.

	Proteins differentially expressed		
Samples	Before FDR	After FDR	False Positives
D01_C1n	4	0	100
D01_C2n	5	0	100
D02_C1n	4	2	50
D02_C2n	2	0	100
D02_TP	3	1	66.7
D07_C2n	3	1	66.7
D07_Cxn	4	1	75.0
Total	25	5	80.0