TAMPERE UNIVERSITY OF TECHNOLOGY

AARO SALMINEN

SAMPLING OF INTRACELLULAR METABOLITES OF *ACINETOBACTER BAYLYI* SP. ADP1 FOR METABOLOMIC ANALYSIS

Master of Science Thesis

Examiners: PhD Ville Santala and
PhD Alain Perret
Examiner and topic approved in
the Faculty of Science and Environmental Engineering council
meeting on November 9th, 2011

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master's Degree Programme in Biotechnology
SALMINEN, AARO: Sampling of Intracellular Metabolites of *Acinetobacter baylyi* sp. ADP1 for Metabolomic Analysis
Master of Science Thesis, 58 pages, 6 Appendix pages
November 2011
Major: Environmental biotechnology
Examiner: PhD Ville Santala and PhD Alain Perret
Keywords: Metabolomics, *Acinetobacter*, mass spectrometry, metabolite extraction

Acinetobacter baylyi sp. ADP1 is a nutritionally versatile bacterial strain capable of natural transformation and homologous recombination. It has a unique metabolism, as it is capable of using a vast collection of compounds as a sole carbon source. It is non-pathogenic and easy to work with in the laboratory. The great potential of ADP1 lies in its usage as a laboratory model organism, as well as in bioremediation. The strain is well characterized, as full genomic sequence and metabolic model exist already. However, no metabolomic analyses have been performed so far.

Metabolomics has recently developed into an important field in biotechnology with many potential new applications. As a new area of interest, it is only in early stages of development, but nevertheless many powerful techniques already exist. However, lots remain to be discovered. As every bacterial species can be very different, the methods of metabolomics need to be designed for every study independently. This is why it is of great importance to optimize the metabolite extraction protocol specifically for each project.

The purpose of this work was to optimize a new extraction protocol of intracellular metabolites from ADP1, for the purpose of metabolomic analysis with high resolution mass spectrometry. An initial protocol, set up and Genoscope, and based on literature and some previous testing, was used as a basis for the optimization. From this protocol, the culture conditions, quenching procedure, sample storage and data analysis were reassessed. In the end, the final protocol was tested with the cells cultivated in two different carbon sources. The final protocol improved the quality of acquired data significantly, providing a better efficiency in quenching of the metabolism and metabolite extraction, leading to more detected metabolites in the final data. In addition, the extraction process was made more organized and faster to perform.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Biotekniikan koulutusohjelma
SALMINEN, AARO: Solunsisäisten pienmolekyylien eristäminen *Acinetobacter baylyi* ADP1 –bakteerista metabolomiseen analyysiin
Diplomityö, 58 sivua, 6 liitesivua
Marraskuu 2011
Pääaine: Ympäristöbiotekniikka
Tarkastajat: FT Ville Santala ja FT Alain Perret
Avainsanat: Metabolomiikka, *Acinetobacter*, massaspektrometria

Acinetobacter baylyi ADP1 on monipuolinen bakteerikanta, joka kykenee luonnolliseen transformaatioon ja homologiseen rekombinaatioon. Se kykenee käyttämään erilaisia hiililähteitä monista eri kemiallisista luokista, mikä tekee sen aineenvaihdunnasta ainutlaatuisen. ADP1 ei ole patogeeninen, ja sen kanssa on helppo työskennellä laboratorio-olosuhteissa. ADP1 soveltuu hyvin laboratorioiden malliorganismiksi, sekä bioremediaation käytännön sovelluksiin. Se on pitkälle karakterisoitu kanta, sillä sen koko genomi on jo kartoitettu ja siitä on olemassa laskennallinen metaboliamalli. Metabolomisia analyysejä siitä ei kuitenkaan ole toistaiseksi tehty.

Metabolomiikka on viimeaikoina kehittynyt tärkeäksi osaksi biotekniikkaa, sillä sen tuottamaa tietoa voidaan hyödyntää moniin käytännön sovelluksiin. Metabolomiikka ei ole uutena tieteenalana kehittynyt vielä kovinkaan pitkälle, vaikkakin monia tehokkaita tekniikoita on jo kehitetty. Paljon tutkimusta tarvitaan kuitenkin vielä menetelmien kehittämisessä. Koska jokainen bakteerilaji on erilainen, metabolomiikan menetelmät pitää suunnitella jokaista tutkimusta varten erikseen. Tästä syystä uutta metabolomiikkaprojektia aloitettaessa on tärkeää optimoida metaboliittien eristysmenetelmä erityisesti kyseisen tutkimuksen tarpeisiin.

Tämän työn tarkoituksena oli optimoida uusi solunsisäisten pienmolekyylien eristysmenetelmä ADP1-bakteerista korkean resoluution massaspektrometrillä suoritettavaa analyysiä varten. Pohjana optimoinnille käytettiin kirjallisuuteen ja aiempiin alustaviin testeihin perustuvaa menetelmää. Tästä protokollasta optimoitiin bakteerien viljelyolosuhteet, metabolian sammutus, valmiiden näytteiden säilytys sekä tulosten analysointivaiheen parametrit. Lopuksi valmista menetelmää testattiin käyttämällä kahdella eri hiililähteellä kasvatettuja ADP1-soluja. Uusi menetelmä paransi saatujen tulosten laatua selvästi sekä näytteiden energialatauksen että niissä havaittujen pienmolekyylien kannalta. Eristysprosessia parannettiin lisäksi järjestelmällisemmäksi sekä nopeammaksi toteuttaa.

RESUME

UNIVERSITE DE TECHNOLOGIE DE TAMPERE
Programme de master en biotechnologie
SALMINEN, AARO: Préparation d'échantillon des métabolites intracellulaires d'*Acinetobacter baylyi* ADP1 pour l'analyse métabolomique
Mémoire de master en sciences, 58 pages, 6 pages d'annexes
Novembre 2011
Matière principale: Biotechnologie de l'environnement
Examinateur: Dr. Ville Santala et Dr. Alain Perret
Mots-clefs: Métabolomique, Acinetobacter, spectrométrie de masse

Acinetobacter baylyi ADP1 est une souche bactérienne polyvalente, capable de transformation naturelle et de recombinaison homologue. Son métabolisme est unique, puisqu'elle est capable d'utiliser une grande variété de composés comme sa seule source de carbone. Elle n'est pas pathogène et elle est facile à utiliser dans le travail de laboratoire. L'ADP1 a un grand potentiel comme un organisme modèle de laboratoire, et aussi dans la bioremédiation. La souche est bien caractérisée, puisque la séquence génomique complète et un modèle metabolique existent déjà. Toutefois, aucune analyse métabolomique n'a été exécutée jusqu'à présent.

La métabolomique a pris de l'importance récemment en biotechnologie avec un grand nombre de nouvelles applications potentielles. Etant un nouveau domaine d'intérêt, la métabolomique n'en est qu'à ses débuts. Cependant, un grand nombre de techniques puissantes existent déjà. Toutefois, il y a encore beaucoup à découvrir. Puisque chaque souche bacterienne est trés différente, les méthodes de la métabolomique doivent être conçues séparément pour chaque étude. Il est donc trés important d'optimiser le protocole d'extraction des metabolites spécifiquement pour chaque projet.

Le but de cette étude était d'optimiser un nouveau protocole d'extraction des métabolites intracellulaires d'ADP1, pour une analyse métabolomique avec spectrométrie de masse de haute resolution. Un protocole initial, fondé sur la littérature et les essais précédents, a été utilisé comme une base d'optimisation. Les conditions de culture bactérienne, procédure de quenching et le stockage des échantillons de ce protocole ont été validés. Enfin, le protocole final a été testé avec deux sources differentes de carbone. Le protocole final améliorait la qualité des données de façon significative, en apportant une meilleure charge d'énergie et plus de métabolites detectées dans les données finales. De plus, le processus d'extraction est devenu plus organisé et plus rapide à exécuter.

PREFACE

This thesis is based on the work conducted during an internship at Genoscope laboratories (funded and supervised by Commissariat à l'énergie atomique et aux énergies alternatives, French atomic energy commission, CEA) in the city of Evry, France. It is a commencement to a project to analyze the metabolomics of *Acinetobacter baylyi* sp. ADP1. The metabolomics project is part of a vast project to characterize the named bacterial strain.

I would like to thank the supervisors of this thesis, Dr. Alain Perret (Genoscope) and Dr. Ville Santala (Tampere University of Technology) for their invaluable help and guidance during the process. I would like to give special thanks to Christophe Lechaplais for help during the laboratory experiments and Lucille Stuani and Sabine Tricot for using the LC-MS devices and performing those analyses. I also appreciate the help of Véronique de Berardinis and Marcel Salanoubat to arrange me the internship at the Genoscope laboratories.

In addition, I wish to extend thanks to the lovely Pauline Curtet for hosting and supporting me during the internship, and to my family and friends for their support and encouragement during the work process.

11th of November 2011

_____

AARO SALMINEN

CONTENTS

TERMS AND DEFINITIONS

ACN	Acetonitrile
ADP1	*Acinetobacter baylyi* sp. ADP1
AEC	Adenylate energy charge
APCI	Atmospheric pressure chemical ionization
C18	Hydrophobic interaction column based on a 18-carbon hydrocarbon silica
centWave	A peak detection algorithm for XCMS
CoA	Coenzyme A
dsDNA	Double stranded DNA
ESI	Electrospray ionization
FTMS	Fourier transform mass spectrometry
GC	Gas chromatography
GC-MS	Gas chromatography coupled with mass spectrometry
HMDB	Human Metabolome Database
HPLC	High-performance liquid chromatography
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid chromatography
LC-MS	Liquid chromatography coupled with mass spectrometry
MA	Medium acinetobacter
MALDI	Matrix-assisted laser desorption ionization
MAS	Medium acinetobacter with added succinate
matchedFilter	A peak detection algorithm for XCMS
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
m/z	Mass-to-charge ratio
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600 nanometers
ORFeome	Library of open reading frames
PTFE	Polytetrafluoroethylene (Teflon)
TOF	Time-of-flight
UHPLC	Ultra high-performance liquid chromatography
XCMS	A data analysis software for LC/MS data
ZIC-HILIC	Zwitterionic hydrophilic interaction column for liquid chromatography
ZIC-pHILIC	Zwitterionic polymer-based hydrophilic interaction column for liquid chromatography

1 INTRODUCTION

Genome sequences of more than a thousand bacteria have already been released in public databases, providing great amounts of information which may lead to better understanding of prokaryotic physiology. However, a significant part of predicted genes still have no real functional annotation. Thus, many metabolic pathways are incomplete, making the prediction of metabolic properties and metabolic network modelling of sequenced organisms difficult. Consequently, the sequencing and research center of Genoscope has started a systematic analysis of the strictly aerobic soil bacterium *Acinetobacter baylyi* sp. ADP1 as an alternative to E. coli to complete the knowledge of prokaryotic aerobic metabolism.

ADP1 is a strain of bacteria that has great potential in biotechnology as a vehicle of bioremediation. It is capable of natural transformation and homologous recombination, and has a unique metabolism. This is why it is able to live on a wide array of carbon sources due to its diverse degradation pathways, and can also be used in degradation of environmental pollutants. (Abdel-El-Haleem 2003) Because of the ease of transformation and heterologous gene expression, ADP1 could, in the future, be a new model organism for biotechnology, when it has been sufficiently characterized. (de Berardinis 2009) At the Genoscope laboratories, the project of characterization of ADP1 has already produced the complete sequence of the genome (Barbe 2004), a computational metabolic model (Durot 2008) and various other tools for working with ADP1. The next step in the project is to do metabolomic analyses on ADP1, in order to gain further information about the metabolism.

Metabolomics is a relatively new field in bacterial studies. It aims to study the complete metabolome of the bacteria in certain conditions, identifying and linking the compounds present in the cell. Metabolomic data can be used for example to characterize pathogens, design new drugs, or for metabolic engineering. (Dunn 2005) One of the most basic tools of metabolomics is mass spectrometry, often coupled with liquid chromatography. This powerful tool is able to detect the metabolites with sufficient mass accuracy to identify single compounds from a complex biological sample. (Werner 2008)

For this kind of analysis, the bacterial samples need to be prepared carefully. In the ideal case no compounds are lost and no bias is introduced to the data. This is impossible due to the vast amount of different compounds, with different chemical properties, in the cells. Consequently, the study needs to be more focused, concentrating on a smaller group of compounds, in the case of this study, on the small polar molecules. The ideal sample preparation protocol is also highly dependent on the bacterial strain used. Due to these facts, the sampling protocol needs to be optimized for every purpose and every microbial species separately. (Dettmer 2007)

To be able to start the metabolomic analyses at Genoscope, a reliable sampling protocol is needed. The goal of this thesis was to optimize a sampling protocol for intracellular metabolites of ADP1 for analysis with high resolution mass spectrometry. Starting from an initial protocol adapted from literature and some initial testing, the culture conditions, quenching of the metabolism and storage conditions were optimized for the purpose of the project. Also the data analysis parameters were optimized for these kinds of samples. In the end, the final protocol was tested with two different carbon sources, in order to both test the acquired protocol and try to assess the impact of a different carbon source to the metabolism of ADP1.

2 THEORETICAL BACKGROUND

This chapter explains the major concepts necessary to understand this thesis. First, the utilized bacterial species, *Acinetobacter baylyi* sp. ADP1 is introduced. In the following chapters, the aim and motivation of metabolomics is discussed, and the sample preparation methods are introduced.

2.1 Acinetobacter baylyi sp. ADP1

This subchapter introduces the bacterial species used in the experimental part of the thesis, *Acinetobacter baylyi* sp. ADP1. First, the bacterial species is introduced in general terms, after which the current knowledge about the species is reviewed, and some tools related to it are introduced.

2.1.1 General Information

The members of the *Acinetobacter* genus can be found in nature in both soil and waters, as well as in living organisms. They are widespread, Gram-negative bacteria that usually appear as paired cocci. They are non-motile and strictly aerobic and as versatile chemoheterotrophs can use many different compounds as carbon sources. The *Acinetobacter* genus is closely related to the *Pseudomonas* one. (Gerischer 2008) The *Acinetobacter* genus has also recently been of great interest in the field of environmental biotechnology, since it has the ability to degrade several pollutants and produce interesting new compounds, such as polysaccharides (Abdel-El-Haleem 2003).

Possibly the most interesting strain of the genus is the *Acinetobacter baumanii* due to its clinical importance. It is a well-known pathogen with great resistance to antibiotics, often known for infecting hospital patients. However, *Acinetobacter baylyi* sp. ADP1 (later referred to simply as ADP1) is easier to work with in the laboratory, and is thus used as a model organism. This is due to the fact that it is not pathogenic. (Gerischer 2008) Also, advanced genetic tools exist for ADP1, which make the laboratory work easier (de Berardinis 2009).

ADP1 originates from the strain BD4 of *Acinetobacter*, which was isolated from soil (Taylor 1961), and mutated by radiation into the present strain of ADP1. The main difference between ADP1 and BD4 is the absence of cell capsule in ADP1, which makes it easier to work with. The strain has the ability of natural transformation and homologous recombination, which makes it interesting for many different genetic applications. ADP1 also grows fast on several carbon sources, with the regeneration time of less than one hour. (Barbe 2004)

The natural transformation of ADP1 happens in three phases. In the beginning, the non-competent cell acquires competence through a complicated cascade of intra- and extracellular signals. In the second phase, the extracellular dsDNA binds to the DNA uptake machinery. After binding, one of the strands of the dsDNA is hydrolyzed into nucleotides, while the other one is taken into the cell. The process is finished by homologous recombination, which integrates the up taken DNA into the host chromosome, if homologous regions exist. (Palmen 1997) The process is visualized in Figure 2.1.

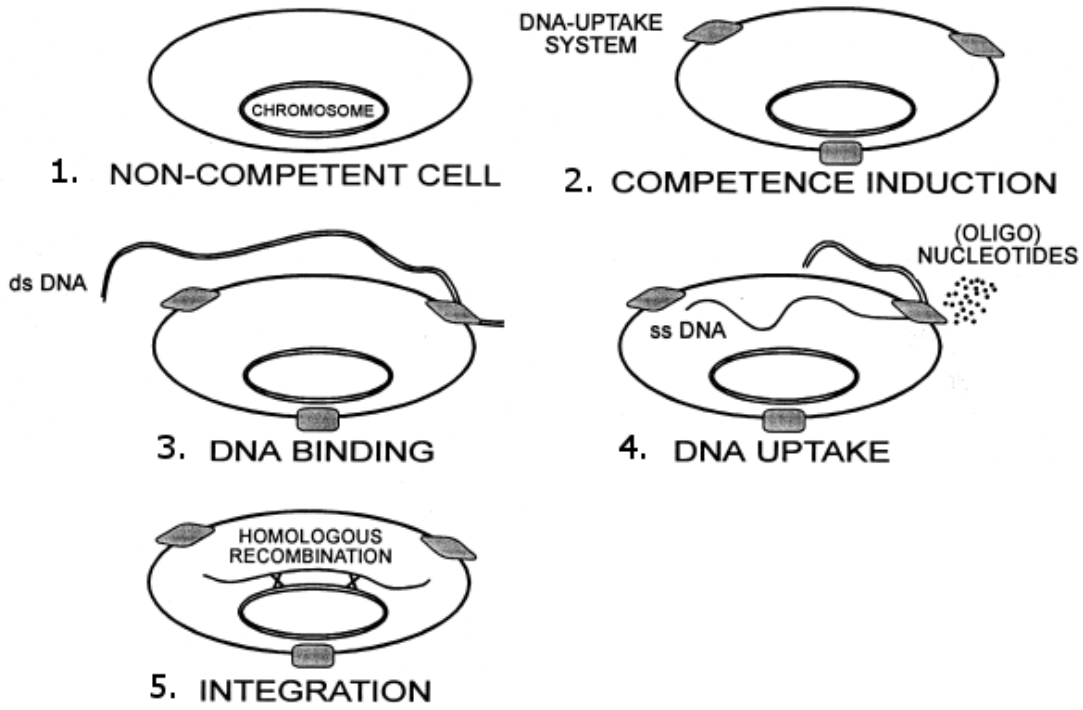


Figure 2.1. *The natural transformation process of ADP1. After the induction of competence (2), the dsDNA binds to the receptors on cell surface (3). The DNA is then taken into the cell (4), and integrated into the genome by homologous recombination (5). Figure edited from (Palmen 1997).*

The natural transformation can be very useful in constructing new metabolic systems in ADP1 host. Such experiments have already been done, and will be done more in the future. (Young 2005) A methodology for the transformation of ADP1 has been published, and can be used in everyday laboratory work. The method is simple and easy to use due to the natural transformation capabilities of the species. (Metzgar 2004)

In addition to the natural transformation and homologous recombination, also the metabolism of ADP1 is special. As mentioned, ADP1 can use a wide selection of compounds, such as the aliphatic succinate and the cyclic quinate, as single carbon sources, and is able to degrade many widely spread environmental pollutants, such as certain aromatic compounds. This is made possible by the β-ketoadipate pathway (Young 2005). The aromatic compounds are first metabolized into either catechol or protocatechuate, and then directed through the pathway into succinyl-CoA and to the Krebs

cycle. The degradation pathway is summarized in Figure 2.2. All of the genes used in these pathways are clustered into five unique islands in ADP1 genome. (de Berardinis 2009)

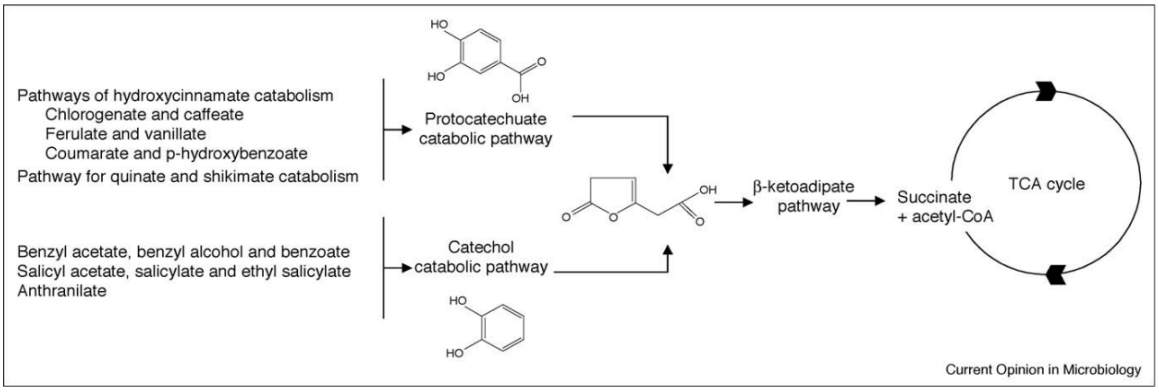


Figure 2.2. *The degradation pathway of aromatic compounds in ADP1. (de Berardinis 2009 Unlike stated in this original figure, the pathway leads to succinyl-CoA, instead of succinate, in the citric acid cycle.*

The degradation pathway of quinate is well characterized both for the regulation and the enzymatic activity, which is why it is often used as a carbon source to test new methods for ADP1 in the Genoscope laboratories. Quinate is transformed first into dehydroquinate, and further to dehydroshikimate. The dehydroshikimate is then converted into protocatechuate, which is metabolized into succinyl-CoA and acetyl-CoA through the β-ketoadipate pathway. These compounds can then enter the central metabolism in the citric acid cycle. The complete pathway of quinate degradation with all the intermediate metabolites is illustrated in Figure 2.3. All the genes required to the degradation are located in a single transcriptional unit within the ADP1 genome. (Gerischer 2008)

These specialized pathways are useful in biotechnology when the degradation of aromatic environmental contaminants is needed. In ADP1, some of the contaminants can be assimilated through the central metabolism, and thus fully metabolized. For example, alkanes with up to 12 carbons can be used by ADP1 as a carbon source (Geissdorfer 1999). *Acinetobacter* species are also important in the degradation and turnover of phytochemicals in the environment. These properties make the ADP1 strain interesting also in the environmental biotechnology point of view. (de Berardinis 2009)

Many possible applications exist for ADP1, and for *Acinetobacter* in general. In addition to bioremediation, ADP1 has been used in production of biosurfactants and bioemulsifiers. Also different kinds of biosensors and enzymatic applications have been shown to be possible applications for the *Acinetobacter*. In addition, ADP1 has shown potential in lipid production because of its tendency to accumulate wax esters and triacylglycerols. (Gerischer 2008) This is why ADP1 has recently been used also for triacylglycerol production in bioenergetics research (Santala 2011).

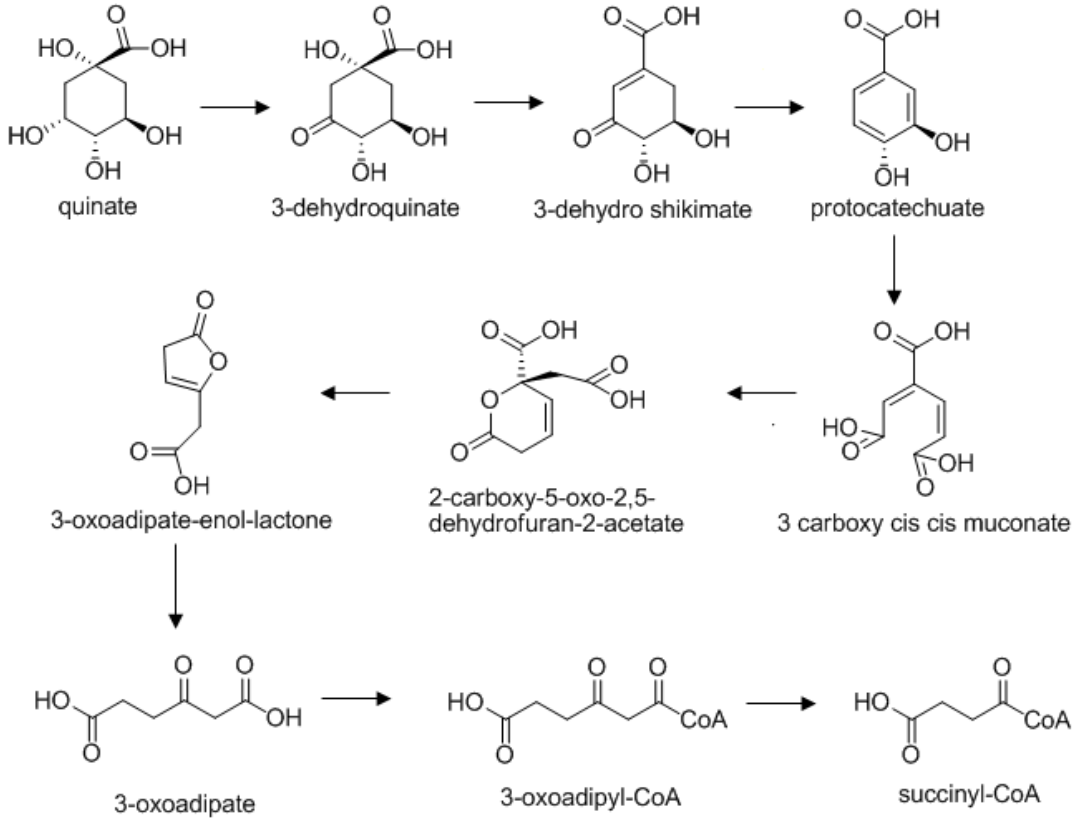


Figure 2.3. *The complete degradation pathway of quinate in ADP1.*

The main drawback of ADP1 in metabolomic studies is that the cells are very difficult to break, probably due to the remaining capsule from its ancestor BD4. As noticed at the Genoscope laboratories in previous unpublished tests, the methods commonly used to cell disruption, such as mechanical stress and sonication often fail to break the cells efficiently and cause loss of metabolites in the analysis. The cells can be broken, however, with sudden consecutive heat and cold shocks performed in liquid nitrogen and a warm water bath. This works better, and most cells can be seen broken under a microscope after the cell disruption.

2.1.2 Current Knowledge

In addition to metabolomic studies, ADP1 is also suitable for genetic research because of the many tools and great amount of knowledge that already exist for the strain. The strain has been mostly characterized at the Genoscope laboratories during the last years. The first one of the available tools to be completed was the full genomic sequence of ADP1. The whole circular chromosome of the bacterial strain was sequenced and annotated in 2004 by Barbe *et al*. A schematic presentation of the organization of the genes is presented in Figure 2.4.

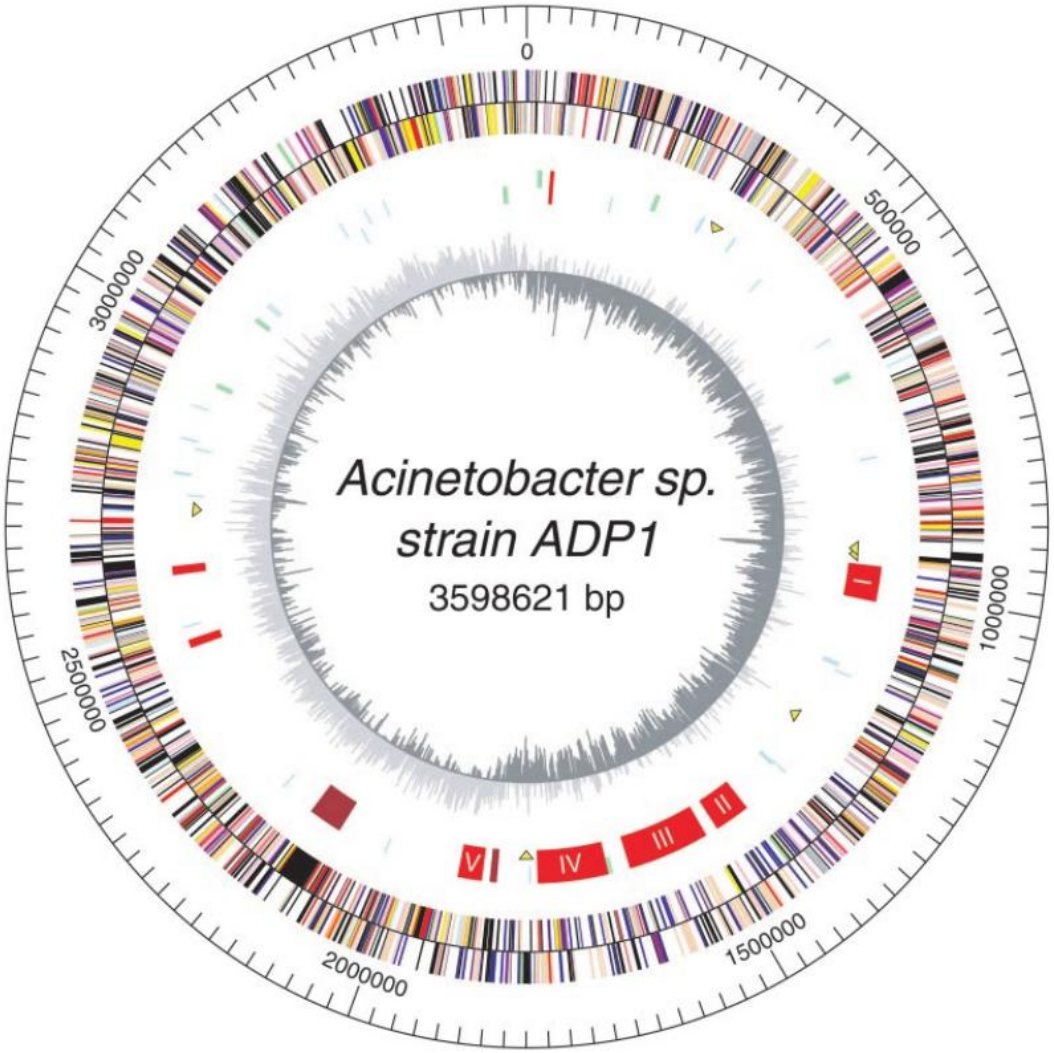


Figure 2.4. *The genome of ADP1, as presented by Barbe et al. (2004). The outermost circles represent the predicted coding regions transcribed in clockwise (outer circle) and counterclockwise (inner circle). The five catabolic islands are numbered I to V inside the circles. The GC bias is presented in the innermost circles. (Barbe 2004)*

The chromosome was approximately 3.6 million base pairs long, and had an average GC content of 40.3 percent. This was surprisingly far from the GC content of the closely related *Pseudomonas* genome (62 %). In the sequence analysis it was confirmed that the genes of the major catabolic pathways are located in five so called 'islands of catabolic diversity'. This arrangement is unique within known bacterial genomes. According to the sequence, the metabolism of ADP1 seems to be specialized to the degradation of diverse aromatic compounds. This is the case in many aerobic soil bacteria, due to many aromatic compounds released to the soil by plants. (Barbe 2004) One curious feature of the genome is the fact that no corresponding gene was found to pyruvate kinase, a gene involved in glycolysis in most organisms. Thus, ADP1 cannot directly phosphorylate glucose, although it can grow with glucose as the sole carbon source. This is not surprising, since all the genes of the Entner–Doudoroff pathway, an alternate pathway

converting glucose to pyruvate, were identified in the genome. However, the pathway of gluconeogenesis and the pentose phosphate shunt were complete. (Barbe 2004)

As perhaps the most important tool in metabolic research, a gene deletion mutant library has also been created by de Berardinis *et al.* (2008). In the process of constructing the library, 2594 single-gene deletion mutants were obtained, as opposed to 499 (16% of the genes) mutants that could not be cultured for one reason or another. These 499 genes are potentially essential genes for ADP1 growth on minimal medium. Of these genes, 56 were of only putative function and 46 of unknown function. 104 mutants were not obtained due to technical reasons. This kind of a library is a powerful tool for studying gene functions, as it can be used to screen the essentiality of a wide array of genes to a certain process, for example the utilization of a carbon source. It is also used for generating new data for systems biology approaches. (de Berardinis 2008)

The library was compared with available data regarding the metabolism of ADP1, as well as with similar libraries of other organisms. The data generated several new hypotheses about different pathways in ADP1, questioning the validity of the annotation-based previous knowledge. The new hypotheses included the annotation of genes involved in the biosynthesis pathways of methionine and ubiquinone. According to these hypotheses, in ADP1 these pathways differ significantly from those described in other bacteria, such as *E. coli*. This conclusion was derived from the gene essentiality pattern, which was unique in ADP1. When compared to other organisms, the gene essentiality data in general was 88% consistent with the data of *E. coli*, and 80% consistent with that of *Pseudomonas aeruginosa*. (de Berardinis 2008)

Also a constraint-based global metabolic model has been created for ADP1 in 2008 to simulate the metabolism of ADP1 in different nutritional conditions. The model includes the genomic data from the sequencing and annotation project, metabolic information from various publications and results of additional experiments. The model encompasses altogether 875 reactions, 701 metabolites and 774 genes, and almost every metabolic route identified in ADP1. The model was constructed in several steps of tests and model corrections. (Durot 2008)

The model was refined by comparing the predictions produced by the model to observations in cultures. These refinements were done in several cycles. In the final model, the predictions were consistent in 91% of wild-type growth experiments, 94% of gene essentiality experiments and 94% of phenotypes of grown mutants. The final model is called iAbaylyi^{v4} and is available to public use online with the help of a web interface developed specifically for the model (http://www.genoscope.cns.fr/cycsim/). (Durot 2008)

In addition to these published tools, other yet unpublished tools exist to examine ADP1. A vast collection of transcriptomics data of ADP1 grown in the presence of different carbon sources has been collected, and can be used at the Genoscope laboratories. The data has been acquired by direct RNA sequencing, as opposed to the traditional microarray-based analysis methods. This provides much more reliable data, as the method is more sensitive than microarrays. (van Vliet 2010) The complete data includes

the expression data of more than 3000 genes in different nutritional conditions. This data can be used to examine which genes are expressed with greater intensity in the different conditions, and thus new functions and metabolic pathways may be discovered and can be analyzed. (Unpublished data)

Also a complete library of open reading frames (ORFeome) of ADP1 exists at Genoscope. The ORFeome was created with an emphasis on soluble proteins with unknown or putative function, and that are related to metabolism. In total, 2157 genes (70% of all ADP1 genes) were cloned, including approximately 1300 enzymes and putative enzymes, and 800 proteins with unknown function. To help the analysis of the library, the genes have been classified by their putative function. (Unpublished data) This data can be easily used to large screenings of enzymatic functions, enzyme characterization or metabolomics analyses. This approach is reciprocal to the conventional methods of enzyme function studies. In ORFeome studies, specific activities are searched for purified recombinant proteins without known functions. (Saito 2006)

2.2 Metabolomics

In this subchapter, the field of metabolomics is introduced, regarding its purpose and different ways of usage. In the beginning of the subchapter, backgrounds of the field are reviewed in general terms. After that, the basic techniques of metabolomics are introduced, with particular attention to the techniques used in the practical part of this thesis. Next the techniques and software of data analysis are discussed. In the end, the recent developments in the field are looked at through various examples, to demonstrate the potential of metabolomics.

2.2.1 Background

Metabolomics is a branch of biological science studying metabolites, their presence and their relationships. The field of metabolomics developed in the 1990s, alongside with transcriptomics and proteomics, as a third level of functional genomics. Metabolomic approaches have been used for a longer time for diagnostics (Horning 1971), but only in recent times have they been used for larger-scale analyses of complex biological samples. (Villas-Bôas 2005) Due to its relatively late commencement, metabolomics is still a new and fast developing field with lots of challenges ahead. The most important challenge is the lack of an analysis platform that would be able to analyze on one run the whole set of metabolites present in the cell. (Dettmer 2007)

To achieve the full analysis of the metabolome, in certain plants up to 200,000 different compounds from many different physico-chemical classes need to be identified, quantified and mapped (Dettmer 2007). In bacteria, however, this number of different compounds is significantly lower. The metabolites include compounds all the way from hydrophobic lipids to hydrophilic sugars in concentrations ranging from pM to mM. This is why the use of two or more different methods is often required to gather all information. For example, a hydrophobic and a hydrophilic column could be used in the

chromatographic step to have a wider set of data for analysis. However, even then the complete set of information may be impossible to gather. (Werner 2008)

In microbial metabolomics, the two main approaches are called fingerprinting and footprinting. Fingerprinting involves the metabolites that are present inside the bacteria at a certain moment of time. To be able to analyze this, a reliable extraction method with a fast quenching of the metabolism is needed. If the quenching is too slow, the cells have time to use their pools of certain metabolites when stressed, and the final sample does not represent the wanted conditions anymore. After the quenching, the cells need to be disrupted and the metabolites collected carefully. (Dettmer 2007) In footprinting the focus is on the extracellular metabolites, secreted by the cells. In this approach, the cells are discarded and the extracellular medium is taken into further metabolomic analysis. These two approaches can result in vastly different results, as cells tend to keep certain molecules, such as phosphorylated compounds, inside the cell, while secreting signaling molecules and excess metabolites. (Dettmer 2007)

The analysis of the samples in both of these approaches is performed with a method powerful enough to detect single metabolites in the complex sample. The mass resolution of the final data needs to be good enough to allow identification of the compounds with the help of metabolite databases. (Dunn 2005) The resulting data is analyzed computationally with the help of specifically designed algorithms. The algorithms can identify masses and compare them to values found in the databases. However, in most cases, a vast amount of false data will interfere with the database searches. (Werner 2008)

The resulting metabolomic data can be used to a variety of different purposes, both in microbes and multicellular organisms. The most obvious use is to map the metabolism of an organism to assess different stress responses or to gain the needed knowledge to be able to produce certain compounds in the organism through metabolic engineering. This information leads to better understanding of the organism, and thus the organism can be used for example in bioremediation, through bioremediation applications, or in synthetic biology, through the means of metabolic engineering. The metabolic information of certain pathogens can also help to design new drugs and to help in diagnostics. (Dunn 2005)

Metabolomics can also be very useful for systems biology. Metabolic information is vital when building new computational models of bacterial organisms. When combined with genomic, transcriptomics and proteomic data, powerful tools can be created to model the organisms *in silico*. In fact, metabolomics has a huge potential in biotechnology, and the field is expected to grow as more and more scientists become aware of the potential of metabolomics. (Dunn 2005)

2.2.2 Basic Techniques

The analysis of metabolomic samples is challenging due to the wide variety of compounds from different classes and sizes. The most common strategies to analyze these samples are mass spectrometry coupled with either liquid chromatography (LC-MS) or gas chromatography (GC-MS), and nuclear magnetic resonance (NMR) spectroscopy.

These techniques develop all the time, constantly improving the separation and mass accuracy. (Villas-Bôas 2005)

When using mass spectrometric methods to detect the compounds, the separation of compounds, based on different variables such as affinity to a column, can be done by chromatographic methods. In gas chromatography (GC), the mobile phase carrying the metabolites is gaseous. This is why it is limited to volatile compounds and compounds that can be made volatile through derivatization. This is a major inconvenience in metabolomics, as the derivatization makes the quering of the compounds from metabolite databases impossible. (Spagou 2010)

This is the reason why liquid chromatography (LC) is often preferred in metabolomics. In LC, the solutes are carried and eluted by a liquid mobile phase. Nowadays, high-performance liquid chromatography (HPLC) with solid columns of porous stationary phase for trapping the metabolites is a standard tool in metabolomics. The function of the stationary phase of the column is often based either on the affinity of the solutes to the stationary phase, size of the particles or ion exchange. (Dunn 2005) In this study, also an ultra high-performance liquid chromatographic (UHPLC) system, which works with higher pressures than standard HPLC, is used. The two columns used in this study are a C18 column (BeH C18 2 x 150 mm 1,7µm) and a hydrophilic interaction (HILIC) column (polymeric beads ZIC-pHILIC 4.6 x 150 mm 5µm).

The C18 column is a chromatographic column based on hydrophobic interactions between the solutes and the porous column. It is based on 18-carbon-atom hydrocarbon silica which works as a stationary phase, as the solvent elutes the compounds according to their hydrophobicity. The column is based on reverse-phase chromatography. (Dettmer 2007) The column used in this study is one designed specifically for UHPLC.

The problem of hydrophobic interaction chromatography, however, is the polarity of many important metabolites in the cell. For example all amino acids and phosphorylated compounds, such as ATP, are very hydrophilic and are thus not retained at all in the C18 column. These compounds are very important to identify, and thus other columns are often preferred in metabolomics. The inability to retain these compounds can result in heavy ion suppression in the ionization phase. In addition, it leads to insufficient capability to identify compounds of the same exact mass, such as threonine and homoserine, from each other. (Dettmer 2007)

The HILIC column is based on hydrophilic interactions between the column and the metabolites. The compounds are eluted in a mobile phase of mostly water, and retained by a zwitterionic hydrophilic stationary phase. This causes the solutes to be separated in order of increasing hydrophilicity. With this method, the hydrophilic metabolites are retained well in the column, and can be more easily identified, due to a well defined retention time. However, if hydrophobic compounds such as lipids are of special interest, HILIC columns are not a good choice, since those compounds are not retained. (Spagou 2010)

The ZIC-pHILIC column used in this study differs from standard hydrophilic interaction columns by its polymer composition. A polymer is added to the silica-based

composition of the standard ZIC-HILIC, stabilizing the structure and allowing the utilization of the column in a wider range of pH, which can help the selectivity of many compounds. The best separation occurs when the mobile phase pH is far from the pKa of the phosphate groups and hydroxyl groups, in order to avoid protonation and deprotonation. The separation of phosphorylated compounds such as ATP has been shown to be significantly better on the ZIC-pHILIC, compared to the conventional ZIC-HILIC. In a recent study, the column was able to separate all biologically important triphosphates (ATP, GTP, CTP, UTP, dATP, dGTP, dCTP and dTTP) from each other in a satisfactory way. (Johnsen 2011)

As the ions exit from the HPLC column, they are directed towards the mass detector. At this stage, the separated compounds of interest need to be ionized for the mass detection and go from a liquid phase to a gaseous one. The ionization allows the analysis of large and non-volatile compounds in the mass spectrometry. The most common ionization methods include electrospray ionization (ESI), the technique used in this study, and matrix-assisted laser desorption ionization (MALDI). (Cech 2001)

In ESI, the solvent passing through the chromatographic column is channeled through a capillary charged with a high either positive or negative voltage. The positive or negative value of the voltage determines if the ionization is performed in the so called “positive ion” or “negative ion” mode. Some compounds can be better seen in positive ion mode, while others are more visible in negative ion mode. As both analyses cannot be performed at the same time, often the samples are run in both modes consecutively. (Cech 2001)

In the process of ionization, the voltage of the capillary generates a charge separation on the surface of the liquid and the liquid bulges from the end of the capillary. This phenomenon is known as the Taylor Cone. As more solution passes through the capillary to the cone, droplets with excess charge detach from the tip of the cone, with the help of increased temperature and nitrogen gas, and move towards the mass spectrometer entrance. The high voltage disperses the liquid stream, forming a mist of highly charged droplets that undergo desolvation during their passage across the source of the mass spectrometer. As the size of the droplets reduces, a point is reached at which the repulsive charges on the surface of the droplets are sufficient to overcome the cohesive force of surface tension. This leads to a Coulombic explosion that produces a number of smaller droplets. After a series of such explosions, free ions are produced and transferred into the mass spectrometer through a small window in the metal plate connected to the power supply, and their mass-charge ratio can be analyzed. (Cech 2001) The ESI method is further illustrated in Figure 2.5.

In the process of ESI, however, not all molecules are ionized correctly. Some molecule can, and will, produce adducts with other ions. This happens before the charge separation step, especially for molecules that have no acidic or basic groups. In positive ion mode, adducts are most often formed with sodium, lithium and ammonium ions. In negative ion mode, on the other hand, chloride adducts are most common. Adduct formation generates problems for the data analysis afterwards, as it changes the masses of

the ions, and thus the exact masses of the detected peaks cannot be directly compared to database values. The problem can be faced with a custom database of commonly occurring adducts of the compounds of interest. (Cech 2001)

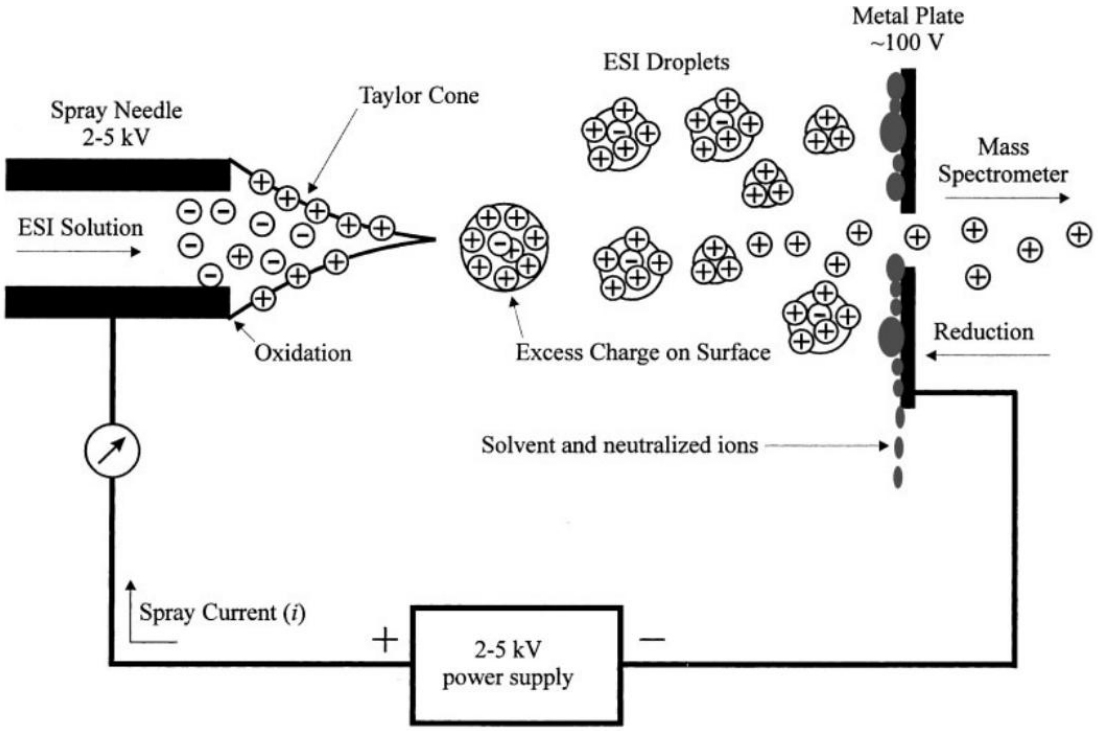


Figure 2.5. *The process of electrospray ionization (ESI) as a schematic drawing. The molecules are ionized in the capillary (left), and passed as droplets to the gas phase. The single ions then proceed to the mass detector. (Cech 2001)*

In addition to ion adducts, other factors producing difficulties in the ionization step include ion clustering and fragmentation. In the study by Tautenhahn et al. (2008) it was noticed that ion clusters such as $[2M+H]^+$ and $[3M+Na]^+$ can form in the ESI process. In addition, up to eight in-source fragments were detected for each compound in their standard mixture. Also isotopic peaks were detected for some compounds. These problems hinder the data analysis, as all of the compounds cannot be detected anymore with their own masses, and thus cannot be queried from metabolite databases. (Tautenhahn 2008)

Another problem in the ionization phase is ion suppression. Ion suppression may happen when non-volatile or less volatile compounds change the droplet solution properties. This makes the droplet formation and evaporation more difficult. When ion suppression happens, ions cannot detach from the droplet or even be formed, and thus they do not reach the mass detector and are not detected. The compounds causing ion suppression include salts, ion pairing agents and even some metabolites. It has been shown that molecules of high mass and very polar compounds may induce ion suppression. This is why phosphorylated compounds are very susceptible to ion suppression when not separated properly. (Annesley 2003)

This problem is most evident when a diverse set of compounds arrive to the ionization step at the same time. The separation of the compounds with chromatography is of utmost importance and helps the problem to a certain extent, but does not solve it completely. For example, as the C18 column does not retain the polar compounds at all, they arrive to ionization at the same time as the buffer solution and salts, and strong suppression occurs. In every mass spectroscopy experiment aiming to quantification of polar metabolites, experiments to determine ion suppression should be conducted in order to have reliable quantification data. Also, the choice of column is important in these cases. (Cech 2001)

In addition to ESI, other methods of ionization exist. In MALDI, a laser is used for ionization of the molecules. The laser pulse is directed to a matrix of crystallized molecules, and the molecules are ionized. Ions then detach from the matrix to the gas phase, and move at high speed to the mass detector. MALDI is often used for larger molecules such as proteins and polymers, and was thus not fully suitable for this study. (Hommerson 2011)

In atmospheric pressure chemical ionization (APCI), the solution containing the molecules is superheated and sprayed with nitrogen at high speed. This creates an aerosol cloud, which is exposed to an electrical discharge for ionization. This method, however, is more violent towards the molecules of interest and results in more fragmentation of the compounds. This is why the ESI method was preferred in this study despite of its ion suppression problems. (Hommerson 2011)

After the ionization, the accurate masses of the formed ions need to be detected. This is often done by mass spectrometry (MS), which can be performed using several different kinds of equipment. Different kinds of MS strategies include time-of-flight (TOF) mass spectrometry, quadrupole mass analysis, ion traps, including the more advanced Orbitrap device, and Fourier transform mass spectrometry (FTMS). (Werner 2008) In this study, an Orbitrap mass detection device was used, due to its great mass accuracy and resolving power, and relatively affordable price.

A layout of an Orbitrap device is shown in Figure 2.6. There is a wide variety of Orbitrap devices on the market, which have a few principles in common. The Orbitrap mass analyzer is often preceded by an external linear ion trap feeding the ions to the analyzer. The ion trap stores great populations of ions and injects them to the mass analyzer in short pulses. Thus, each population of ions of a single mass-to-charge ratio (m/z) is injected to the Orbitrap in a sub-microsecond pulse. The ions are pushed to an axial oscillation by a strong electrical field when they enter the mass analyzer. The ions also rotate around the Orbitrap axis, which keeps them from collapsing into the central electrode. Thus, the ions move on two axes around the Orbitrap electrode. The ions form a ring around the electrode, and the whole ring oscillates harmonically with a period proportional to $(m/z)^{1/2}$. The signal from the mass analyzer is then detected, and processed with Fourier transform into an m/z spectrum, from which the exact masses can be derived. (Makarov 2010; Perry 2008)

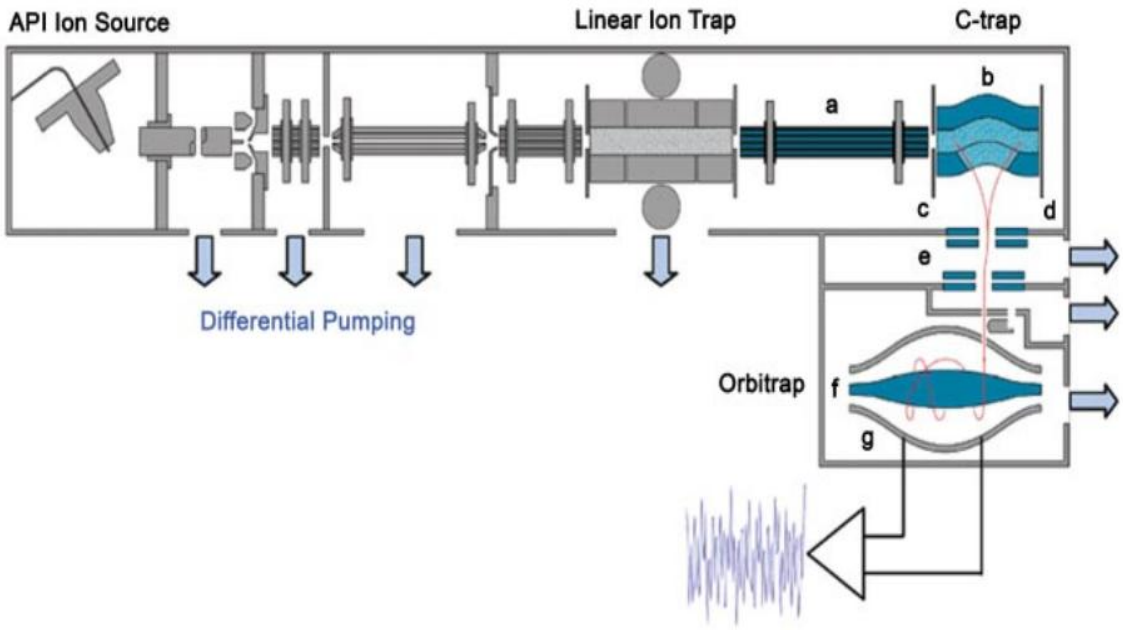


Figure 2.6. *A schematic drawing of the LTQ Orbitrap mass detector. The ions pass from the ion source, through the linear ion trap, into the C-Trap. From there they are moved into the Orbitrap mass detector where the exact masses are detected based on their movement around the electrode. (Perry 2008)*

The strength of Orbitrap mass detection is its mass accuracy and resolution. When the mass accuracy is sufficiently high, only a few elemental compositions for a mass peak are proposed, which can facilitate the identification of the compound. Depending on the resolution, the mass accuracy of Orbitrap can reach 3 ppm, which is sufficient for this kind of identification. (Makarov 2010) However, the accuracy can be even below 1 ppm when background ions are used in the process (Scheltema 2008). The resolution of Orbitrap depends on the scan time. If the scan time is set at 0.4 seconds, the resolution is 30,000, which is also sufficient for identification. With a resolution of 100,000 the scan time could be as high as 2 seconds. The scan time is important, as the whole mass spectrum has to be scanned in significantly less time than the width of a chromatographic peak. For example, for a short peak of 5 seconds, a 2-second scan time is not enough to identify the peak as only two scan points are acquired during the peak. This is why lower resolutions are sometimes preferred. (Werner 2008)

In TOF mass analysis, the ions are accelerated with the help of an electric field, and their velocity is measured. This is directly dependent on the m/z ratio of the ion, from which the mass can be again derived. The TOF devices can reach a mass accuracy of 1-5 ppm, and a resolution of 10,000. The problem of TOF, however, is the low dynamic range, usually at two or three orders of magnitude. (Werner 2008) This may not be enough for the purposes of the present study, and thus the Orbitrap method was preferred.

In addition to Orbitrap, FTMS also provides a mass accuracy and resolution high enough for metabolite identification. The mass accuracy of FTMS is in the same range

or slightly better than for Orbitrap, and the resolution is significantly higher. The FTMS technology is based on a constant magnetic field that accelerates the ions in a large cyclotron, and the frequency of the ions is measured. This is proportional to the m/z ratio of the ions, and the exact mass can be processed through Fourier transformation. FTMS is a still developing technology, and the devices are still very expensive. (Ohta 2010) The mass accuracies and the resolutions of the different MS devices in usual analysis can be compared in Table 2.1.

Table 2.1. *The mass accuracies and resolutions of the commonly used MS detection devices. The mass resolutions are expressed as m/Δm.(Werner 2008)*

Device	Mass resolution	Mass accuracy
Ion traps	500	0.3 to 1 u
TOF devices	800-20,000	1-5 ppm
Orbitrap	Up to 100,000	<3 ppm
FTMS devices	Up to 1,000,000	<1 ppm

To improve the quality of the acquired data, tandem mass spectrometry (MS/MS) is sometimes used. In MS/MS, the compounds are fragmented either in the ESI phase or afterwards, after the first MS analysis. In the second run of MS, the fragments are analyzed again. With the combined information of both MS analyses, it results in more information on the structure of the molecules, and compounds of interest can thus be identified even without a list of standard compounds. Different types of fragmentation can be used to acquire different kind of information on the structures. (Griffiths 2007)

In addition to MS techniques, NMR spectroscopy is also widely used in metabolomic research. NMR is based on measuring the resonance of the molecules in a magnetic field, and to thus determine the exact mass. The nuclei of the atoms absorb electromagnetic radiation in a characteristic frequency and can thus be identified. NMR does not need to be coupled to chromatographic separation, which makes it less discriminant than MS. Also, sample preparation is easier for NMR, as almost any kind of sample can be directly injected to NRM analysis. (Villas-Bôas 2005)

NMR spectroscopy, however, has some major drawbacks when compared to MS techniques. When examining complex biological samples, the sensitivity of NMR is not as good as that of the best MS techniques. Also, the chemical analysis of the molecules in MS, when combined with chromatographic separation, may be more complete, and specific molecule identification easier. In addition, the high-throughput utilization of MS is significantly more advanced than in NMR. (Villas-Bôas 2005)

2.2.3 Data Analysis

The data sets produced by mass spectrometry, when dealing with small molecules, are very vast and complex, and thus very difficult to read and interpret manually without designated software. The data consists of two dimensions. The first dimension is chromatographic data, presented as a function of the retention time of the specific molecules

in the column. The other part of the data consists of the data produced by the mass analysis. It presents intensities of detected compounds as function of their corresponding masses. The intensities can be presented in centroid or full profile mode. In centroid mode the masses are presented as discrete peaks, which makes the data easier to read. In full profile mode the peaks have a width, and are easier to differentiate from background noise.

The data requires heavy preprocessing before proper computational analysis can be performed. The optimal situation would be having software to automatically detect MS peaks and the corresponding chromatographic retention times from the data and match them with masses and retention times of specific compounds used as standards. This is what programs like the open-source software XCMS can be utilized for. The blind approach used by XCMS can also provide information that cannot be expected. (Smith 2006) However, big difficulties remain in the process.

The main difficulties are created by the presence of great numbers of false positives in the data, resulting from different origins. Most of the false positives originate from fragments of compounds, adducts and cluster ions that the metabolites give birth to in the process of ionization. Also high chemical noise can create false peaks and hide actual ones. (Werner 2008). Also different contaminants of various sources can pollute the data, and need to be ignored in the data processing phase. Many contaminants are known to be often present in MS data sets, and can thus be ignored in the data analysis phase. (Keller 2008)

Another problem is the variation of retention times from one chromatographic method to another, and even the shift of retention times when using a single method. To deal with the variation from one injection to another, the samples are often injected to the MS devices multiple times, to be able to have, on average, the correct retention time. The software used for data analysis often has a function for retention time correction for these cases. In order to know the actual retention time for the exact method used, standard solutions are often used, including the compounds of interest. Even the standards need to be injected several times however, as the retention times might vary also in their case.

Several different software products are available for the analysis of the LC/MS data. Many MS device manufacturers provide their own analysis software which are compatible with their own hardware. Such software include Sieve by Thermo (Thermo Fisher Scientific, Waltham, Massachusetts, USA) and MarkerLynx by Waters (Waters Corporation, Milford, Massachusetts; USA), among others. These software solutions, however, are closed-source and often not as flexible as the free, open-source options. In addition, they often are either expensive or only available when purchasing the related hardware.

However, there are also freely available software solutions. MetAlign is a free but closed-source data analysis application used mainly in plant research. The open-source alternatives include XCMS and MZmine. MZmine, originally initiated at the VTT Technical Research Centre in Finland, was originally published in 2005 (Katajamaa

2005) and has been developed since then. It is a widely used user-friendly toolkit with a good user interface, and includes also some statistic tools.

XCMS, the automatic data analysis software mainly used in this study, is a collection of many tools to analyze the LC/MS data. It includes tools for retention time alignment, feature detection and matching, and data filtration. XCMS is an R programming language based toolkit with no graphical user interface. (Smith 2006) XCMS has been shown to be slightly more efficient in peak detection than MZmine when the right tools are used (Tautenhahn 2008).

XCMS has a matched filtration algorithm to filter noise and detect features in the data. This tool is called matchedFilter, and it is available with XCMS by default. In this technique, a filter is used whose coefficients represent the expected shape of the signal, in the case of LC/MS data, Gaussian peaks. The peaks are then detected from the filtered data. The modifiable parameters for the filtration step are explained in Table 2.2. However, matchedFilter is originally dedicated to low resolution mass spectrometry, and could be thus not optimal for use with the Orbitrap device.

Table 2.2. *The modifiable parameters of the matchedFilter algorithm and their explanations. (Smith 2009)*

Parameter	Explanation
fwhm	Full width of the Gaussian peak at half maximum intensity
max	Maximum number of peaks per chromatogram
snthresh	Signal to noise ratio cutoff
step	Step size used for profile generation
steps	Number of steps to merge before filtration
mzdiff	Minimum m/z difference for peaks with overlapping retention times
sleep	Length of pause (s) between peak finding cycles
nSlaves	Number of MPI slaves in parallel peak detection

The other algorithm for filtration and feature detection, mainly used in the data analysis of this thesis, is called centWave. It is shown to be faster and more efficient than matchedFilter or MZmine in peak detection when working with high resolution MS data. In a comparison study, centWave detected 2634 peaks from an example data set, whereas matchedFilter could detect only 1568 of them. The data was similar LC-MS data as used in this thesis. The algorithm was also the fastest of the three to process the full data set. (Tautenhahn 2008) However, this algorithm is probably not optimal for Orbitrap data either, as it is developed for slightly lower resolution devices, such as TOF mass spectrometers. The algorithm gathers partial mass traces and then uses continuous wavelet transformation and Gauss-fitting for filtration and detection. The main parameters and the explanations of the meaning of each parameter are introduced in Table 2.3.

Table 2.3. *The modifiable parameters of the centWave algorithm, and their explanations. (Smith 2009)*

Parameter	Explanation
ppm	Tolerated m/z deviation between injections
peakwidth	Chromatographic peak width range (sec)
snthresh	Signal to noise ratio cutoff
noise	Noise level
prefilter	Minimum number and intensity of mass peaks
mzdiff	Minimum m/z difference for peaks with overlapping retention times
sleep	Length of pause (s) between peak finding cycles
nSlaves	Number of MPI slaves in parallel peak detection

After the filtration and peak detection, the peaks are aligned through different injections. Peaks are matched with the XCMS default algorithm that generates an average chromatogram and then matches peaks across different samples. After this, the possible retention time shift between samples is corrected, after which the peaks are matched again to generate an overlay of different chromatograms with corrected retention times. The last step of the data processing strategy is to fill missing peaks, with the help of the baseline, in the samples. (Smith 2006) The typical workflow of the analysis of LC/MS data with XCMS is reviewed in Figure 2.7.

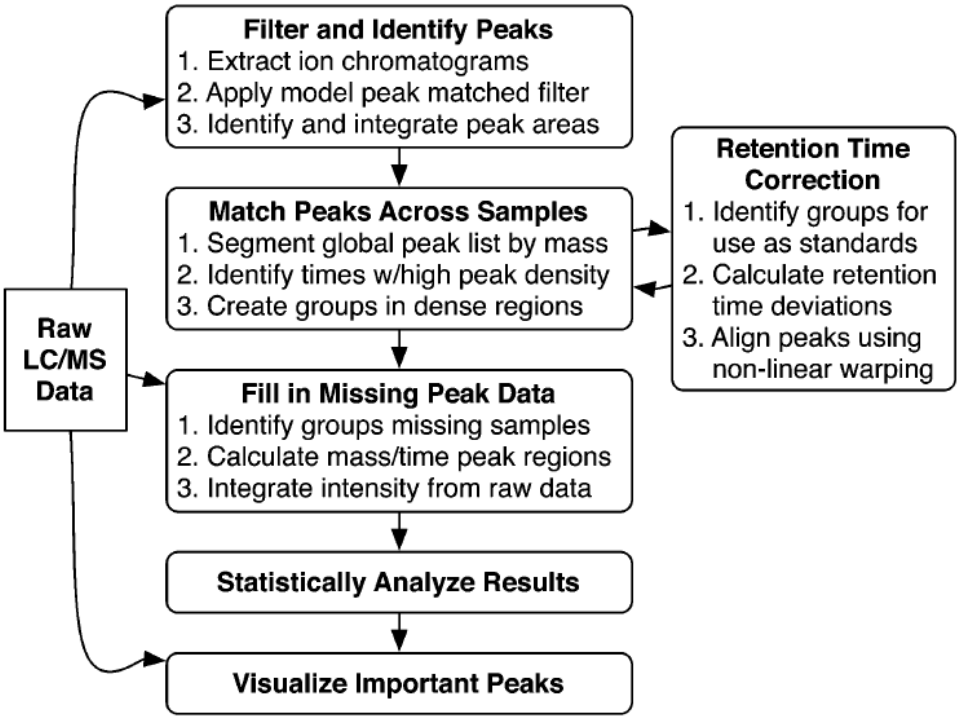


Figure 2.7. *The workflow of a typical LC/MS data analysis process with XCMS. Figure from (Smith 2006).*

The result of the data processing is a list of the detected peaks with the corresponding retention times and m/z values. For both variables, the average values are presented, as well as the minimum and maximum values in the data set. In addition to these data, the final data contains statistical results for validation purposes, and the sufaces of the peaks. This list includes some known contaminants, which have to be filtered out. To achieve this, a search of known contaminants in the samples is run, and the contaminants are marked to the final list of results. The contaminants list is based on previously published lists of known contaminants in mass spectrometry (Keller 2008). These contaminants can result from both sample handling and instruments, and are impossible to avoid completely. The potential contaminants include, among others, keratins, peptide fragments, adducts and solvents. (Keller 2008)

In the end of the analysis, extracted chromatograms corresponding to the peaks can be created. These can be viewed to visualize single peaks in the mass detection data. The exact masses of the detected peaks are also compared to online databases in order to ease the search of interesting compounds. As databases, the Kyoto Encyclopedia of Genes and Genomes (KEGG) compound database (http://www.genome.jp/) and the Human Metabolome Database (HMDB) (http://www.hmdb.ca/) are used. In the end, the data is arranged and exported into a format that can be viewed in Microsoft Excel, where all the data can be easily scrutinized.

The blind XCMS approach is not perfect yet. It can be seen when working with LC/MS data that peaks visible in the raw data are sometimes not detected by XCMS peak detection algorithms. This could be due to the low peak intensity, non-Gaussian peak shape or possible other factors. Also, the high number of false positives in the final data can be a problem, creating lists of thousands of detected peaks. This can hinder the detection of compounds of interest, as they are more difficult to filter from the vast data sets. (Tautenhahn 2008)

Another way to analyze LC/MS data is to view directly the data created by the detection device (the so called raw data). This can be done by such software as Xcalibur (Thermo Fisher Scientific, Waltham, Massachusetts, USA). In this way, the compounds of interest need to be known, as the data is too complicated to analyze with a blind approach by this method. When using standard lists and with a limited number of compounds of interest, however, it is possible to analyze the data also this way. In Xcalibur, by searching compounds by mass, the chromatographic and mass peaks can be viewed, and the compound can be identified and the peak areas can be integrated.

The computational data analysis cannot control all sources of error. There are some major biases, created by the experimental procedures, associated with this kind of data. The final number of bacteria used in the extraction, if not controlled, can create major bias in the amounts of different compounds in the final mix. Thus, samples made without controlling the number of bacteria might not be directly comparable. Also different extraction methods may introduce a bias in the final data sets. The problem of different bacterial biomasses can be encountered by normalization of the data. The first way to normalize data in bacterial extractions is to use the final OD_{600} value of the bacterial

culture if it is available. (Alfassi 2004) This, however, does not include the bias introduced by different solvents. Different extraction solvents could produce a bias by favoring certain kinds of metabolites to be extracted. For example, when water or other hydrophilic solvent is used as an extraction solvent, the most hydrophilic metabolites will aggregate and may be filtered off.

There are, however, at least three methods to normalize data mathematically, without knowing the final amount of bacteria. In the first approach, the current of all detected ions is divided by the highest current in the data. The second method divides the intensity of each compound with the sum of all detected ions in order to be able to compare different samples. The third method of normalization uses the sum of the squares of all peaks to normalization. The usage of this kind of normalization enables the quantitative comparison of different samples. (Alfassi 2004)

One important aspect of analysis of data from bacterial extractions is to determine if the quenching of the metabolism has been sufficient or not. This can tell if the samples actually represent the metabolic state in the cells in stable conditions, or have the cells suffered stress before quenching. The most common method for this is calculating the adenylate energy charge (AEC) in the samples. This is done using the following equation:

$$\mathrm{AEC} = \frac{[\mathrm{ATP}] + 0.5[\mathrm{ADP}]}{[\mathrm{ATP}] + [\mathrm{ADP}] + [\mathrm{AMP}]} \quad \text{(Chapman 1971)}$$

A high value of AEC (>80%) indicates that the quenching was done efficiently and reflects a cell population in a steady state. A low AEC indicates either too slow quenching causing additional stress to the bacteria, or stationary growth phase of the bacteria during quenching. (Chapman 1971) The concentrations of the compounds can be derived from the peak areas in the biological samples, with the help of a calibration curve acquired in the exact same conditions as the samples. If the data is not normalized, the concentrations themselves are not comparable, but the relationships are. Thus the AEC represents well the energy charge of the bacteria after quenching.

2.2.4 Recent advancements

Although metabolomics is a new field of study, very specialized techniques have already been emerging to study the metabolism of different kinds of cells. More specialized and more accurate techniques are developed constantly. The new techniques aim to extend metabolomics to directions it traditionally cannot reach. These directions include metabolomics of single cells, and new pathway analysis methods utilizing a more precise quantitative method than before.

One example of such a technique is the tracer-based metabolomics. It is a systems biology approach to metabolomics, which is targeted and based on metabolic network models. (Paul Lee 2010) In tracer-based metabolomics, a certain metabolite is labeled

with ^{13}C, and added to the growth medium of the cells. The cell metabolome is then screened, and the distribution of the marker in various metabolites is mapped. With the help of known pathways and models the relations between these compounds can then be determined. With the distribution of ^{13}C the quantitative relationships between the precursors and products can then be determined. With this data and the constraint-based metabolic models, phenotypes may then be compared quantitatively. (Paul Lee 2010)

The data produced is complementary with traditional metabolic profiling. Together these two sets of data describe completely the relationship of the inner cell metabolism and the nutrients derived from the environment. The data differs from that of fluxomics in the sense that there is an emphasis on the final products of the pathways. This technique could be very useful in biomedical research in the future, as they are further developed and moved to a larger scale. (Paul Lee 2010)

Another new promising initiative is so called single cell metabolomics. It aims at revealing phenomena in single cells that are hidden in conventional metabolomic profiles under the population average. So far a functional technique to do this has not been found, but scientists are working on it on several fronts. If successful, this approach could provide relevant data for systems biology, as well as provide a more detailed analysis on laboratory cultures, which can be surprisingly heterogeneous. (Heinemann 2011)

The problem of single cell metabolomics, however, is the small quantities of metabolites. These days mass spectrometry is able to detect compounds in the required attomole range, but quantification is still impossible. Also the transfer of one cell to the analysis device without losing content is problematic. This problem has been faced with micropipette sampling of the metabolites, followed by nano-ESI ionization. Alternatively, microfluidic chips could be used, followed by MALDI ionization. These techniques still have several problems to deal with, but the scientists are optimistic. (Heinemann 2011)

One recently developed approach is also the integration of metabolomics data to other omics data, in tries to create more comprehensive models. For example, transcriptomics and metabolomics have been analyzed in parallel. This may shed new light on the relationships between the genes and the end products, the metabolites. When comparing these data in different conditions, correlations can be made between the abundance of a certain metabolite and the activity of a gene. (Zhang 2010)

It has been also tried to integrate more than two layers of data into one analysis platform. In a study with *E. coli*, genomic, proteomic, metabolomic and fluxomic data were used together. The response to different kinds of perturbations was measured, and the different levels of data were correlated with each other. Surprisingly small impact was found on all levels of the data caused by the perturbations. Even metabolite concentrations were stable. These kinds of experiments can be useful in order to know the relationships between the different levels of regulation, as well as to test and refine the constraint-based metabolic models of the organisms. (Ishii 2007)

2.3 Sampling of Intracellular Metabolites

The sampling techniques used in microbial metabolomics are very important for the final results of the analysis. The goal of sampling is to extract the metabolites of interest from the biological matrix and to prepare them for the analysis. The interfering components of the matrix need to be removed in the process. In the case of bacterial intracellular metabolites (fingerprinting) and complete metabolic analysis, only salts and macromolecules are aimed to be removed as matrix components. (Dettmer 2007)

The bacterial cells need to be grown in the desired medium and conditions until the good OD_{600} value is reached. In most cases, the cells should be in logarithmic growth phase when starting the extraction, to get the best results. In the logarithmic phase, the AEC of the cells is the highest and the metabolite composition reflects the one of freely growing cells. Also, no excess stocks of metabolites and stress induced by too big population is present yet in the logarithmic phase. This is why, in the logarithmic phase, the metabolite composition reflects best the composition in normal, non-stressed cells. (Chapman 1971)

The cells can be grown in liquid medium until the end, as it is done in most cases, either in bioreactors (Link 2008), in flasks (Wellerdiek 2009) or on plates (Ewald 2009). In these cases the final OD_{600} is easy to measure and verify the good time to start the extraction. Another method is to do the cell growth on filter membranes on solid medium. This method was introduced by Brauer *et al.* (2006). The advantage of this method is that there is no extracellular medium to remove in the sampling process. This prevents metabolite loss if cell leakage occurs and makes the process faster to perform. Also, the quenching is easier to do fast, as the filter can simply be submerged into the quenching solvent. However, in the filter culture the measurement of the final OD_{600} value can be more complicated, and cannot be controlled continuously as in liquid culture.

This method of quenching is more efficient than the commonly used fast filtration (e.g. Shin 2010) and centrifugation (e.g. Marcinowska 2011), when cells are grown in a liquid culture. From liquid culture the cells need to be collected with one of the aforementioned methods, which causes additional stress to the cells. This may lower the AEC and thus lower the quality of the samples. In addition, metabolites may be lost if cell leakage occurs. In filter culture, these problems do not occur, as the cells are not in a liquid medium and can be quenched instantaneously.

When the cells are at the desired concentration, the cell metabolism needs to be quenched. This means freezing the metabolism to prevent loss of any metabolites from a certain moment of time on. The quenching needs to be done as fast as possible, in order to acquire the optimal result. If the quenching is too slow, the cells are affected by the stress caused by it, and have time to consume their pools of some metabolites, such as ATP and ADP. (Chapman 1971) This is why the efficiency of the quenching is measured commonly by the AEC of the final sample. In a successful quenching the pools of ATP and ADP remain in the state of the living cells, and the AEC value is high.

The quenching is often performed in extreme conditions, in the presence of a solvent to stop the metabolism quickly. Commonly used solvents include acetonitrile, methanol, chloroform, liquid nitrogen, glycerol and different mixes of the five. (Link 2008; Rabinowitz 2007/2) In the literature, acetonitrile has been found as the best solution for quenching in *E. coli* by some studies (Rabinowitz 2007/2). Methanol quenching has been seen to lose some important metabolites, such as triphosphates (Bolten 2007). Chloroform has not been seen to be good for most metabolites. In addition, the addition of formic acid to the quenching solvent, in order to decrease the pH, has been seen to increase the yield of phosphorylated compounds in extractions done on *E. coli.* (Rabinowitz 2007/2)

The best quenching solvent and conditions, however, are highly dependent on the microorganism used. In *E. coli,* acetonitrile is often used as the solvent (Rabinowitz 2007/2), whereas studies performed with *Corynebacterium glutamicum* and *Saccharophagus degradans* prefer methanol quenching (Wellerdiek 2009; Shin 2010). In yeast studies, boiling ethanol and a mixture of chloroform and methanol have been shown to give the best results (Canelas 2009). This is why it is important to determine the best quenching solvent individually for each microorganism. The wrong choice of quenching solvent and quenching conditions could result in highly biased data, and undermine the whole study.

After quenching, the metabolites need to be extracted from the cells. This requires cell disruption. Before the cell disruption, the cells are often washed to avoid the extracellular medium contaminating the final sample. This is why it is important that the cells remain intact during the process of quenching. Otherwise the metabolites can leak out of the cells, and will be washed away in the washing step. This can cause severe underestimation of the concentrations of metabolites in the cells. (Bolten 2007) With the filter culture, however, the washing step is not necessary, as no extracellular medium is present.

The cells can be broken in many different ways. Popular methods include bead mills, freeze-thaw cycles and ultrasound. In bead milling, the cells are loaded into a vessel containing small glass beads and agitated. The collisions with the beads break the cells, and the intracellular metabolites are released to the supernatant. In freeze-thaw extraction, the cells are broken by subsequent freezing and thawing with a high temperature gradient. Often, liquid nitrogen and a warm water bath are used in order to maximize the temperature difference. The cycle is repeated several times. In ultrasound, the cells are sonicated for 5-10 minutes in an ultrasonic bath, which breaks the cells efficiently. (Marcinowska 2011)

No universal extraction protocol exists that would work for all metabolites, bacterial species and analytical techniques. Studies exist, where the extraction protocol or certain parts of it are taken directly from literature and not tested and optimized for the used conditions (Marcinowska 2011). This can lead to major bias in the results, as the ideal solvents and conditions vary significantly from one species to another. For example, cold methanol quenching might be good in *E. coli,* but when used for *Lactococcus lac-*

tis, it causes severe cell leakage. This is why, in every new metabolomic study, the extraction protocol needs to be optimized and tested carefully before the analyses. (Bolten 2007)

In ADP1, no publications exist for a complete, optimized extraction protocol. The closest organisms to have an optimized protocol are probably *Pseudomonas aeruginosa* (Marcinowska 2011) and *Pseudomonas putida S112* (van der Werf 2008). The protocol of *P. aeruginosa* is for preparation of metabolomic samples for GC-MS from liquid cultures of *P. aeruginosa*. In this protocol, fast filtration and freezing of the cells is used for quenching. However, the quenching step was not considered important for the study, and was not optimized, but taken directly from literature. Bead mill was used for cell disruption. (Marcinowska 2011)

In the analyses of *P. putida S112* the analysis was performed both with GC-MS and LC-MS. The quenching and extraction method was based on literature, and not optimized for the study. (van der Werf 2008) The quenching was done in -45°C in a methanol quenching solvent. The method is originally developed for *Lactobacillus plantarum*, and thus could be not optimal for *P. putida*. (Pieterse 2006) The extraction was performed with the help of chloroform. (Rujiter 1996)

However, even these protocols are far from what is needed for the metabolomic analysis of ADP1 with LC-MS, due to the difference in methods and bacterial species. For example, as ADP1 is an exceptionally difficult bacterium to break, bead milling would not work for its disruption as well as for *P. aeruginosa*. In addition, as the quenching is not optimized for the species in these studies, the protocol cannot be used directly for ADP1. This is why a completely new extraction protocol is desperately needed for reliable analysis of the ADP1 metabolome.

3 RESEARCH METHODS AND MATERIAL

This chapter introduces the materials and methods used in the practical part of this thesis. In the first subchapters, the bacteria and culture conditions, mass spectrometry techniques and parameters and data analysis tools are revealed, and the initial extraction protocol is reviewed. In the fourth part, the actual experiments of sample preparation protocol optimization are described.

3.1 Culture Conditions and Initial Protocol

Bacteria and culture medium

The *Acinetobacter baylyi* sp. ADP1 bacteria, used in all the experiments, were derived from the same Genoscope laboratory stocks that have been used in several previously published research projects (de Berardinis 2008, Barbe 2004). The strain is a wild type ADP1.

The minimal culture medium (*Milieu Acinetobacter*, MA) for all the experiments was the same one used previously in Genoscope projects (de Berardinis 2008). The composition of the basic medium is as follows:

- 31 mM Na_2HPO_4
- 25 mM KH_2PO_4
- 18 mM NH_4Cl
- 41 mM nitrilotriacetic acid
- 2 mM $MgSO_4$
- 0.45 mM $CaCl_2$
- 3 mM $FeCl_3$
- 1 mM $MnCl_2$
- 1 mM $ZnCl_2$
- 0.3 mM ($CrCl_3$, H_3BO_3, $CoCl_2$, $CuCl_2$, $NiCl_2$, Na_2MoO_2, Na_2SeO_3).

This minimal defined medium composition is used in order to know the concentrations of different ions exactly. In for example LB medium, the composition and the concentrations of different nutrients may vary from one batch to another. Also, probably not all components present in the medium are known. Thus the conditions are not exactly constant from one culture to another. In addition, 25 mM succinate is added to the medium as a carbon source. With added succinate, the culture medium is referred to as MAS. All the compounds needed were from Sigma-Aldrich. For solid media, the same MAS composition is used with added agar to solidify the plate. When other carbon sources than succinate are used, they are added in the same concentration.

The initial culture and sample preparation protocol

The initial protocol of bacterial culture, before the optimization experiments, had been adapted from publication about experiments on other organisms (Rabinowitz 2007/1; Brauer 2006) and some initial testing done previously in the Genoscope laboratories (unpublished data). No published data exists on these kinds of experiments on ADP1 or any other member of the *Acinetobacter* genus, and thus no optimized protocol has been available.

The culture protocol was divided into two parts, an initial liquid culture and a subsequent solid medium culture on a filter membrane. This way of culture was first used by Yuan et al. in 2006, in a study performed with *E. coli* (Yuan 2006). This method was decided to be used in order to be able to quench the cell metabolism as quickly as possible after removal from the incubator, without causing excessive stress to the bacteria in between. In addition, the prevention of cell leakage was also important. (Rabinowitz 2007/1) The initial protocol is illustrated in Figure 3.1.

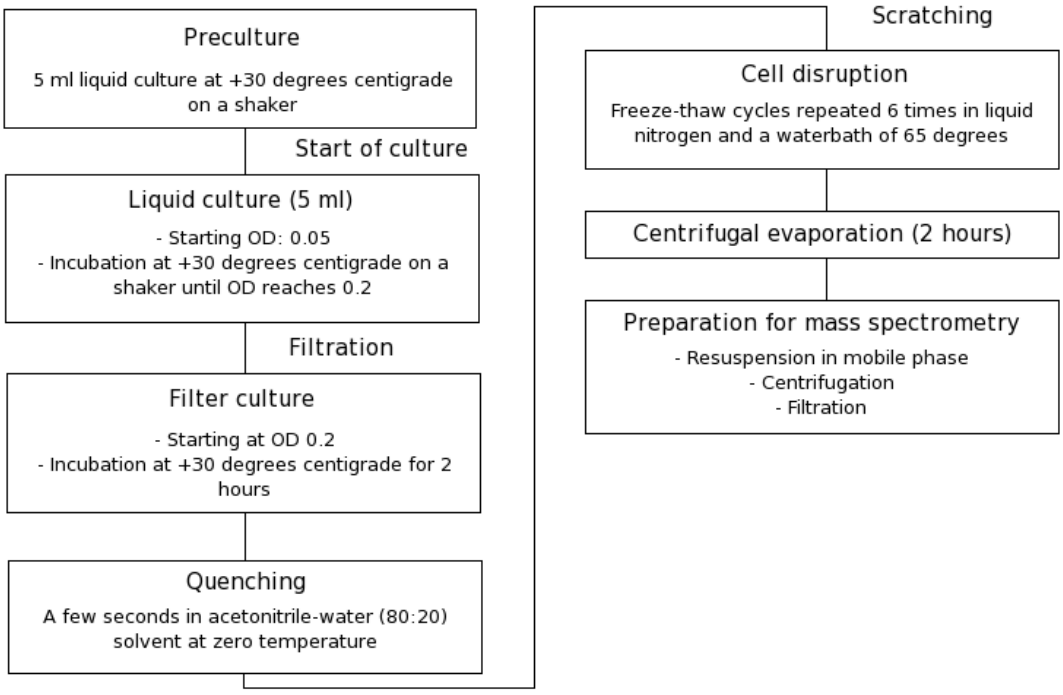


Figure 3.1. *The initial sampling protocol for ADP1.*

The protocol starts with overnight culture of the ADP1 cells in the MAS medium. The culture is performed as a liquid culture of 5 ml, in a shaker (Infors AG CH-4103 Incubator, Infors AG, Bottmingen-Basel, Switzerland) at 30 °C. The next morning, an aliquot is taken from the confluent culture, and its optical density at 600 nm (OD_{600}) is measured (Device: SAFAS UVmc2, SAFAS Monaco, Monaco). According to this measurement, the culture is diluted into two new 5 ml liquid cultures in MAS medium so that the OD_{600} in the beginning of the culture is 0.05. One culture serves as the actual

culture for extraction, while the other one is an OD control. Also one blank control with no bacterial inoculum is put to the incubator at the same time.

The new culture is incubated in the same shaker-incubator for 2 hours, until the OD_{600} reaches approximately the value of 0.2. At this time, each of the tubes is taken out of the incubator and the liquid cultures are filtered on hydrophilic PTFE filter membranes (Omnipore® Membrane Filters, 0.45 µm JH, Merck Millipore, Billerica, Massachusetts, USA). Once the culture is filtered, the filter paper is placed on a Petri dish with solid MAS medium, bacteria faced upwards. This way the bacteria can grow on the membrane, while the nutrients diffuse through it from the solid medium. The dish is placed in an incubator (Memmert Model 600, Memmert GmbH, Schwabach, Germany) at 30 degrees for a 2-hour period. After this, the culture has supposedly reached its logarithmic growth phase, and the metabolism of the bacteria can be quenched.

The quenching is done in acetonitrile-water (80:20) solution on a ethanol-ice bath at a temperature slightly below zero. The solvent had been chosen in previous tests at the laboratory, by looking at the cell behavior through a microscope. Acetonitrile was chosen over methanol because it prevented cells from clustering and gave better results in cell disruption. The temperature is not controlled, and could vary from one extraction to another. However, very low temperatures could not be used as the freezing point of the acetonitrile solution was only slightly below zero.

After the incubation, the filter culture and the blank control are removed, one at a time, from the incubator, and the filters are submerged face down in 3 ml of the cold quenching solution. After a few seconds, the filter is turned and the bacteria are carefully scratched off the filter membrane into the solution with tweezers. Once the cells are in the solution, 2 ml of the quenching solution is added and the total amount of 5 ml is transferred into 2.5 ml cryotubes (Sarstedt AG & Co, Nümbrecht, Germany). The same procedure is done to the negative control dish. Finally the OD control culture is also removed from the incubator and the cells are recovered as before into 5 ml of MAS medium. The OD_{600} of the suspension is then measured to know the final OD_{600} of the cultures.

The four cryotubes, two from the blank control and two from the culture, are then processed through the metabolite extraction phase. ADP1 are exceptionally strong bacteria, and do not break as easily as *E. coli*, for example. In previous studies at Genoscope laboratories, it has been discovered with the help of microscopy that bead mills and sonication fail to break the cells efficiently. (Unpublished data) This is why, for ADP1, freeze-thaw cycles are the best way to break cells, as they produce more stress to the cells and are able to break them.

The freeze-thaw cycles are executed with the help of liquid nitrogen and a hot water bath (Grant GP200-S26, Grant Instruments, Ltd, Cambridgeshire, United Kingdom) at 65 ºC. The tubes are submerged first into liquid nitrogen to freeze the solution rapidly. Right after this, the tubes are transferred into the water bath (65ºC) in order to unfreeze them quickly. This cycle is repeated six times for efficient disruption of the cells. After disruption, the sample is lyophilized in a centrifugal evaporator (Savant SpeedVac®

Systems AES2010, GMI, Ramsey, Minnesota, USA), and the pellet is re-suspended into the initial mobile phase of the used chromatographic method. The cell extracts can be stored dry after the evaporation at -20 ºC until analysis. The solution is then centrifuged, filtered through sterile 0.22 µm filters and transferred into vials for the mass spectrometric analysis.

3.2 Mass Spectrometry

In the analytical part of this study, two different chromatographic columns were used, coupled to a HESI-LTQ Orbitrap mass detector (Thermo Fisher Scientific, Waltham, Massachusetts, USA). These columns were the hydrophobic C18 column (Waters Acquity UPLC BEH C18 1.7 µm, Waters Corporation, Milford, Massachusetts, USA) and the hydrophilic ZIC-pHILIC column (Merck SeQuant polymeric beads 5 µm, Merck SeQuant AB, Umeå, Sweden). Both of the columns were 150 mm long. The thickness of the C18 column was 2.1 mm, while the HILIC was 4.6 mm thick. Both of the columns had UHPLC (the C18 column) or HPLC (the HILIC column) conditions especially optimized for them.

The samples, dissolved in the initial mobile phase of the method and filtered, were always kept at 4ºC until the injections. Each sample was always injected 4 times (2 times for standard mixtures) consecutively to deal with retention time shifts, and for statistical reasons. Each sample was injected in both positive and negative ion modes to facilitate the detection of as many metabolites as possible.

When using the hydrophobic C18 column, the column temperature was kept at 50 ºC. The flow through the column was 400 µl/min, and a single injection volume of 10 µl was used. As the polar mobile phase, either deionized water (negative ion mode) or 1.0% formic acid with 10 mM ammonium acetate at pH 3.5 (positive ion mode), was used. Methanol was used as the eluting solvent. The gradient method used can be seen in Table 3.1.

Table 3.1. *UHPLC method used for the C18 column. Solvent A is deionized water in negative ion mode, and 1.0% formic acid with 10 mM ammonium acetate (pH 3.5) in positive mode. Solvent B is 100% methanol.*

t (min)	Solvent A (%)	Solvent B (%)
0	100	0
1	100	0
10	0	100
18	0	100
20	100	0
25	100	0

When more emphasis was wanted to hydrophilic compounds, the ZIC-pHILIC column was used. In the dedicated method, the column temperature was 30 ºC and the flow of mobile phase through the column was 500 µl/min. Injection volume was kept again at

10 µl. Acetonitrile was used as the organic solvent, whereas water with 10 mM ammonium carbonate at pH 9.1 was used as the eluting solvent. The exact method used can be seen in Table 3.2.

Table 3.2. *HPLC method used for the ZIC-pHILIC column. Solvent A is water with 10 mM ammonium carbonate at pH 9.1, Solution B is acetonitrile.*

t (min)	Solution A (%)	Solution B (%)
0	20	80
2	20	80
22	80	20
30	80	20
35	20	80
40	20	80

In the Orbitrap mass detection, the same settings were used for all samples, regardless of the column. The range of the m/z ratio was set at 50-1000, and the mass resolution was set at 30,000. This means a scan time of 0.4 s. The resolution was kept relatively low to be able to perform more scans per minute, which helps to define chromatographic peaks more accurately. This is useful also for space saving purposes for the data files.

3.3 Data Analysis

For the analysis of the mass spectrometry data, the XCMS and Xcalibur software products were used. In the XCMS script, with the help of a previously acquired set of data, the parameters were optimized so that as many metabolites as possible could be identified. The parameters were optimized with the help of standard mixes and extractions performed previously at Genoscope with the help of the initial protocol. The optimization was done in an iterative fashion, repeating the analysis with slightly different parameters every time. In the process, the efficiency of the centWave and matchedFilter algorithms was compared with different parameters. The Xcalibur software was used to confirm from the RAW data if the detected compounds were actually present in the data set or not, and in some cases to identify the metabolites present in the standard injections.

The standard mixtures were created to know the retention times of each compound, in order to identify the compounds in the sample data. 122 potential metabolites were chosen to the mixtures, according to predictions of computational metabolic models of ADP1 (Durot 2008) and compounds detected in *E. coli* in previous publications (Bajad 2006). Some compounds were also chosen based on results from earlier test injections, suggested by XCMS analysis. The standard compounds were divided into four mixtures in order to avoid having compounds of the same mass in one mixture. This would prevent the identification of individual retention times. The final mixtures included the compounds presented in Table 3.3. All compounds had the concentration of 50 µM in

the final mixtures. The standard mixtures were run in both C18 and HILIC columns, with LTQ-Orbitrap as the mass detector. The retention times were detected from the data set using Xcalibur and recorded. The detected retention times for all compounds in both of the used columns in positive and negative ion mode are shown in Appendix 1. All compounds were acquired from Sigma-Aldrich.

Table 3.3. *A list of the 122 potential metabolites included in the standard mixtures, in order by their molecular masses. The retention times of each compound can be seen in Appendix I.*

Isoprene	Phenylalanine
Butanal	Phosphoenolpyruvate
Propanoate	Pyridoxamine
Diaminopropane	Pyridoxine
Glycine	Capric acid
Pyruvate	Aconitate
Butanoic acid	Shikimate
Alanine	Acetylornithine
Lactate	Arginine
Acetoacetate	Citrulline
γ-Aminobutyrate	Glucosamine
Hydroxypyruvate	Hydroxyphenylpyruvate
Serine	myo-Inositol
Glyceric acid	Tyrosine
Catechol	3-Phosphoglycerate
Isocytosine	Citrate
Cytosine	Tryptophan
Uracil	Pantothenic acid
Proline	Cystathionine
Fumarate	Deoxycytidine
3-Methyl-2-Oxobutyrate	Deoxyguanosine
Betaine	Deoxyuridine
Valine	Thymidine
Succinate	Cytidine
Threonine	Uridine
Homoserine	Biotin
Mercaptopyruvate	Pyridoxamine-5-phosphate
Cysteine	Deoxyadenosine
Nicotinamide	Deoxyinosine
Nicotinate	Glucosamine-6-phosphate
Taurine	Glucose-6-phosphate
Thymine	Thiamine
4-Methyloxovalerate	Adenosine
2-Ketohexanoate	Inosine
(Iso)Leucine	Guanosine
Oxaloacetate	Xanthosine

Ornithine	dCMP
Aspartate	Reduced Glutathione
Malate	CMP
Deoxyribose	UMP
Adenine	Cyclic AMP
Hypoxanthine	AMP
p-Aminobenzoate	dGMP
Anthranilic acid	IMP
Trigonelline	GMP
Tyramine	Riboflavin
Histidinol	S-Adenosyl-L-methionine
α-Ketoglutarate	CDP
Glutamine	ADP
Lysine	dGDP
Glutamate	Folate
Methionine	CTP
Guanine	ATP
Xanthine	dGTP
2,3-Dihydroxybenzoic acid	GTP
Protocatechuate	Oxidized Glutathione
Histidine	NAD
Orotic acid	NADP
Allantoin	CoA
Phenylpyruvate	FAD
7-Methylguanine	Acetyl-CoA

3.4 Protocol Optimization Experiments

This subchapter describes the experiments performed in order to optimize the sample preparation protocol for ADP1 to be used in the metabolomics project. First the experiments to optimize the culture conditions are discussed, followed by the descriptions of the experiments to optimize the quenching and sample storage. In the end the experiments performed to test the final protocol are described.

3.4.1 Culture Conditions

Optimization of cell recovery from filter

The first step of the experiment was to determine whether mild sonication in a water bath or scratching was the best way to recover cells from a filter membrane in the end of the culture. Also the reproducibility of the cell recovery was tested. For this step, the ADP1 cells were cultured overnight in a 5 ml liquid culture on a shaker at 30 degrees. In the next morning, the culture was diluted to an OD_{600} of 0.05, into a 25 ml liquid culture. The culture was incubated approximately 2 hours on the shaker, until it had

reached an OD_{600} of 0.2. At this point, 5 ml aliquots of the culture were filtered on four filter papers.

The filter cultures were incubated again for two hours in 30 degrees, after which the cells were recovered to filtered Tris-HCl 0.1 M buffer (pH 6.8) with two different methods. From two filters, the cells were scratched off with tweezers, as in the original protocol. From the other two, cells were recovered by a 5-minute sonication, while the filter was submerged in 5 ml of the buffer solution. The sonication was performed in cold water (+4 ºC) with Bransonic® Tabletop Ultrasonic Cleaner model 2510 (Branson Ultrasonic Corporation, Danbury, CT, USA). After the cell recovery, OD_{600} was measured from all samples.

After the resuspension, the cells were filtrated onto four new filter papers, and the cells were again recovered from the filter membrane into a cell suspension in the same buffer solution. This time, the cells that had been sonicated before were scratched, and vice versa. The final OD_{600} was measured from each sample. This was performed in order to assess the reproducibility of the cell recovery. The first recovery provided the initial amount of cells, and as they were filtrated and recovered again, the efficiency and reproducibility of the recovery could be determined. A lower OD_{600} value than on the first recovery would indicate that the recovery process is losing cell mass. In case of a similar OD_{600} value, the recovery method would be efficient.

In the second part of the experiment, it was determined whether the sonication method worked as well in MAS medium as in acetonitrile-water quenching solution. These solutions were chosen in order to compare if it is possible to do the sonication cell recovery in the same solvent as the quenching (in this case acetonitrile-water), or if they rather need to be transferred back to the MAS medium after quenching. The incubation times were similar to the first part of the experiment, so the results are also comparable to those performed in Tris-HCl buffer before. By comparing the OD_{600} values of the recoveries in three different solutions, it would be possible to determine which one would be optimal to be used in the final protocol.

The cells were again cultured overnight in a liquid culture on a shaker. As in the first step, the cells were diluted to an OD_{600} of 0.05 and incubated for two hours until the OD_{600} had reached 0.2. After the incubation, the cells were filtered and grown on the filter until the cell recovery. In the cell recovery step, two of the filters were sonicated submerged in 5 ml MAS medium, and two in 5 ml acetonitrile-water (80:20). In the end, the OD_{600} of each sample was measured against the corresponding blanks.

In the third part of the experiment, it was confirmed that the five-minute sonication does not break the cells. In this small experiment, one plate of ADP1 cells grown on a filter membrane was obtained in the same way as in the previous experiments. The plate was incubated 2 hours at 30°C, in order to have the usual cell density in the final culture. The cells were then scratched off the membrane, into Tris-HCl 0.1 M buffer (pH 6.8) with tweezers. At this point the complete absorbance spectrum (400-600 nm) of the cell suspension was measured.

After this, the cells were sonicated for five minutes on a glass plate in order to simulate the sonication process of the cell recovery. The same device and settings was used as previously. After the sonication, the spectrum was measured again and the two spectra were compared. If the sonication did not break cells, the two spectra would show the same absorbance throughout the wavelength scale. In case cell disruption would happen, however, the spectra would differ from one another significantly, as the number of intact cells would be decreased.

Growth kinetics in filter culture

The growth kinetics of ADP1 on filter culture could differ significantly from those in the liquid medium. This is why it was important to determine the kinetics in order to know when the culture is in the logarithmic growth phase to quench the metabolism and start the extraction at the right time. The growth kinetics were determined for both succinate and quinate media. The kinetics between these two conditions are different because of the different molecular structures, and thus the different degradation pathways for the substrates.

This was done by starting a liquid culture (60 ml) of ADP1 in both succinate and quinate-containing media from OD_{600} of 0.05, as is done in the extraction protocol. When reaching an OD_{600} value of approximately 0.2, 5 ml aliquots of the liquid culture were filtered on 15 filter membranes, and the filters were transferred onto solid MAS or MAQ culture plates to grow at 30 ºC. After this, one plate was taken from the incubator every 30 minutes, the cells were recovered with sonication into Tris-HCl 0.1 M buffer (pH 6.8), and the OD_{600} value was measured. After the measurements, a growth curve was drawn, and an ideal point for the quenching of the metabolome was decided.

The 30-minute interval between the OD_{600} measurements was considered sufficient, since the regeneration time of ADP1 is roughly one hour. In addition, it had been discovered before that the cell recovery works well in Tris-HCl buffer, so this was used as the solvent. Otherwise, the liquid culture, filtration and OD_{600} measurement were done as in the initial protocol, except sonication was used for cell recovery.

3.4.2 Quenching

Quenching of the bacterial metabolism needed to be optimized, because in the data acquired beforehand by the initial protocol, the adenylate energy charge was low, approximately 50 percent, suggesting inefficient quenching. Also, it needed to be verified if other solvents than the acetonitrile-water would be more efficient in general metabolite recovery, or faster to work with. In the experiments to determine the ideal quenching conditions, five sets of samples were prepared. The sample preparation was performed according to the initial protocol described earlier, the only change being the changing of the quenching solvent and temperature. Also, samples were sonicated for cell recovery from the filter.

The first condition used was the one in the initial protocol: acetonitrile-water (80:20) was used at the temperature close to 0 ºC. The temperature is kept low due to the high

freezing point of the solvent. This was performed in order to be able to compare the other results to the original method. In the second sample, the same solvent was used with added formic acid. Formic acid has been described in the literature as a good agent to prevent the loss of phosphorylated compounds. This is necessary to calculate the adenylate energy charge of the bacteria, and thus to evaluate the efficiency of quenching. A concentration of 0.1 M of formic acid was used, as often seen in literature. (Rabinowitz 2007/2)

The third samples were prepared with a mixture of acetonitrile and methanol (80:20) as the solvent, quenching at -30 degrees. This would combine the potentially best aspects of each solvent: acetonitrile as the main solvent and 20% methanol to bring the melting point of the solution down, in order to be able to work in a lower temperature. Acetonitrile and methanol have both been described before as potential quenching solvents (Rabinowitz 2007/2).

In the fourth sample, the solvent was changed to 100% methanol. Methanol has been also described as a potential quenching solvent in some microbial species (Rabinowitz 2007/2), and in addition, with pure methanol it was possible to work in dry ice temperature, -80 ºC. Even though methanol has been said to be a bad quenching solvent in many different bacteria due to its tendency to lose high amounts of metabolites in the process (Bolten 2007), it was decided to test it due to the vast differences between different bacterial species. Even if methanol does not work well in one species, it can be the optimal in another one. The fifth samples were prepared with another solvent, a mixture of chloroform and methanol (1:2). The working temperature was set at -30 ºC. (elaborated from Rabinowitz 2007/2) All the extracts were prepared within two weeks, and always injected right after the preparation. All the samples were run with the ZIC-pHILIC column and the Orbitrap mass detector, in both positive and negative ion mode, and the metabolite compositions were determined using Xcalibur and XCMS.

Special attention was given to the visibility of the peaks of AMP, ADP and ATP, to be able to calculate the adenylate energy charge. In order to compute the energy charge, calibration of the peak areas of AMP, ADP and ATP in relation to the concentrations was performed. Mixes of all three compounds at four concentrations (5, 10, 25 and 50 µM) were prepared and analyzed in the same conditions as the samples. Calibration curves of the concentrations as function of the peak area was computed. Using the standard curve, the concentrations of ATP, ADP and AMP in each sample were computed, and the energy charge was calculated for each condition using the standard AEC equation.

3.4.3 Sample storage

The stability of the extracted samples after preparation has been rarely discussed in literature before, even if it is of great importance. The samples are often prepared one by one, stored, and analyzed all together. The samples are often frozen dry after the centrifugal evaporation (Link 2008; Jozefczuk 2010), or in a liquid phase (Taymaz-Nikerel 2009; Winder 2008) and analyzed days or weeks later. There could be, however, degra-

dation of some compounds over time and in the melting and thawing process. In the work of this thesis, it was decided to assess if the samples are stable in storage at -20 ºC after preparation.

In the first experiment, three extractions were prepared in the same conditions, according to the initial extraction protocol. One of the samples was analyzed right after the preparation, one was stored dry for two weeks and one was stored dry for four weeks before analysis at -20 ºC. The analysis was performed in negative ion mode on the C18 column and the Orbitrap mass detection device. The resulting data was analyzed with Xcalibur and the centWave algorithm in XCMS.

In the second experiment, it was tried if one-week storage would be possible for the samples. It was also compared if it is better to store the samples dry after centrifugal evaporation, or in liquid phase after the cell disruption step. Two samples were prepared, using the initial extraction protocol, simultaneously from the same liquid culture. The samples were stored at -20 ºC in the two above-mentioned conditions for 7 days, and analyzed on the ZIC-pHILIC column and the Orbitrap in both positive- and negative-ion mode. Data analysis was performed as above.

3.5 Testing of the final protocol

The efficiency and reproducibility of the acquired final protocol was tested by making two extractions of ADP1 in the presence of two different carbon sources. The reproducibility could be determined by comparing the results of the succinate extractions to the ones obtained in the quenching experiments earlier. Quinate was chosen as the second carbon source, as the degradation pathway of quinate is well known (Figure 2.3). It would also be interesting to assess the modifications of the metabolism of the bacteria, introduced by the different carbon source.

The cells were grown, from the beginning to the end, in the MA medium with either succinate or quinate as the sole carbon source. Both of the carbon sources were added in concentration of 25 mM. The growth and extraction was done following the final optimized protocol. The final samples were run on the ZIC-pHILIC column and the Orbitrap mass detector. The results were analyzed using the Xcalibur software. The results of the cells grown on succinate were compared to the results obtained in the quenching experiments. If a similar AEC and number of metabolited would be detected, the reproducibility of the protocol could be confirmed.

The two different conditions were analyzed in look for major differences between the samples, and possible previously undetected pathways. When analyzing the sample grown on quinate, a special attention was given to the intermediate compounds of the degradation pathway of quinate (Figure 2.3). The detection of these compounds would both give further confirmation to the degradation pathway and give validity to the protocol.

4 RESULTS

This chapter presents the results of the experimental part of the thesis. The results are presented in the order of the previous chapter: first the experiments dealing with the culture protocol are displayed, after which the results of the quenching of the metabolism and the sample storage are presented. In the end, the results of the protocol testing are shown.

4.1 Data analysis parameters

The parameters of data analysis for both centWave and matchedFilter algorithms in the XCMS software were optimized with the help of standard mixtures and extraction data. The final optimized parameters of both algorithms are shown in Table 4.1. In the process, centWave was found to be more efficient in identifying the wanted peaks. The centWave algorithm also produced less false positive results, which leads to more efficient analysis of the final data.

Table 4.1. *The optimized parameters for centWave and matchedFilter algorithms in this study.*

matchedFilter		**centWave**	
Parameter	**Value**	**Parameter**	**Value**
fwhm	30	ppm	20
max	5	peakwidth	5-25
snthresh	100	snthresh	3
step	0.1	noise	200
steps	2	prefilter	3, 250
mzdiff	0.001	mzdiff	-0.0005
sleep	0	sleep	0
nSlaves	0	nSlaves	0

4.2 Culture conditions

Optimization of cell recovery from filter

The first part of the experiment was performed without problems. The OD_{600} of the culture at the time of filtration was 0.20 for all samples. Table 4.2 displays the results of the OD_{600} measurements performed after the first cell recovery, as well as the end of the study, for all samples in Tris buffer.

Table 4.2. *The results of the OD_{600} measurements in the first part of the cell recovery experiment.*

Sample	OD (First recovery)	OD (Second recovery)	Method
1	0,73	0,27	Sonication + Scratching
2	0,83	0,29	
3	0,49	0,20	Scratching + Sonication
4	0,53	0,25	

On the first cell recovery, cell recovery by sonication resulted in significantly higher OD_{600} values than the scratching method. This suggests that the mild sonication is a more efficient method for cell recovery. The duplicate recoveries produced similar OD_{600} values, suggesting a well reproducible recovery. On the second recovery, the values of the duplication recoveries again were almost the same, which further suggests that the recovery is reproducible. On the other hand, a significant part of the cell mass had disappeared after the first recoveries. This could be due to adhesion to the glass plates where the solution was kept, or cell aggregation. Dried cell mass was detected on the glass plates the following day.

In the second part of the experiment, the OD_{600} at the time of filtration was again 0.20. The results of the final OD_{600} measurements are presented in the Table 4.3. The cell recovery by sonication seemed to work better in acetonitrile-water solution than in MA medium. The final OD_{600} values were more than two-fold greater when acetonitrile was used in the solvent. In addition, the cells seemed to be heavily aggregated after the sonication in the MA medium. In addition, the values of sonication in acetonitrile-water correspond to the earlier results with Tris-HCl buffer, which suggests that the results in two conditions are comparable.

Table 4.3. *The results of the OD_{600} measurements of the second part of the cell recovery experiment. ACN stands for acetonitrile, MA stands for medium acinetobacter.*

Sample	OD	Solvent
1	0,328	MA medium
2	0,338	
3	0,852	ACN-water (80:20)
4	0,792	

In the third part of the experiment, the two spectra before and after sonication looked identical, suggesting no major cell disruption due to the five-minute sonication. This confirmed our hypothesis that the mild sonication in a water bath is not harmful to the cells, and it can be safely used in cell recovery from the filter without a major risk of cell leakage. The two spectra acquired can be seen in Figure 4.1.

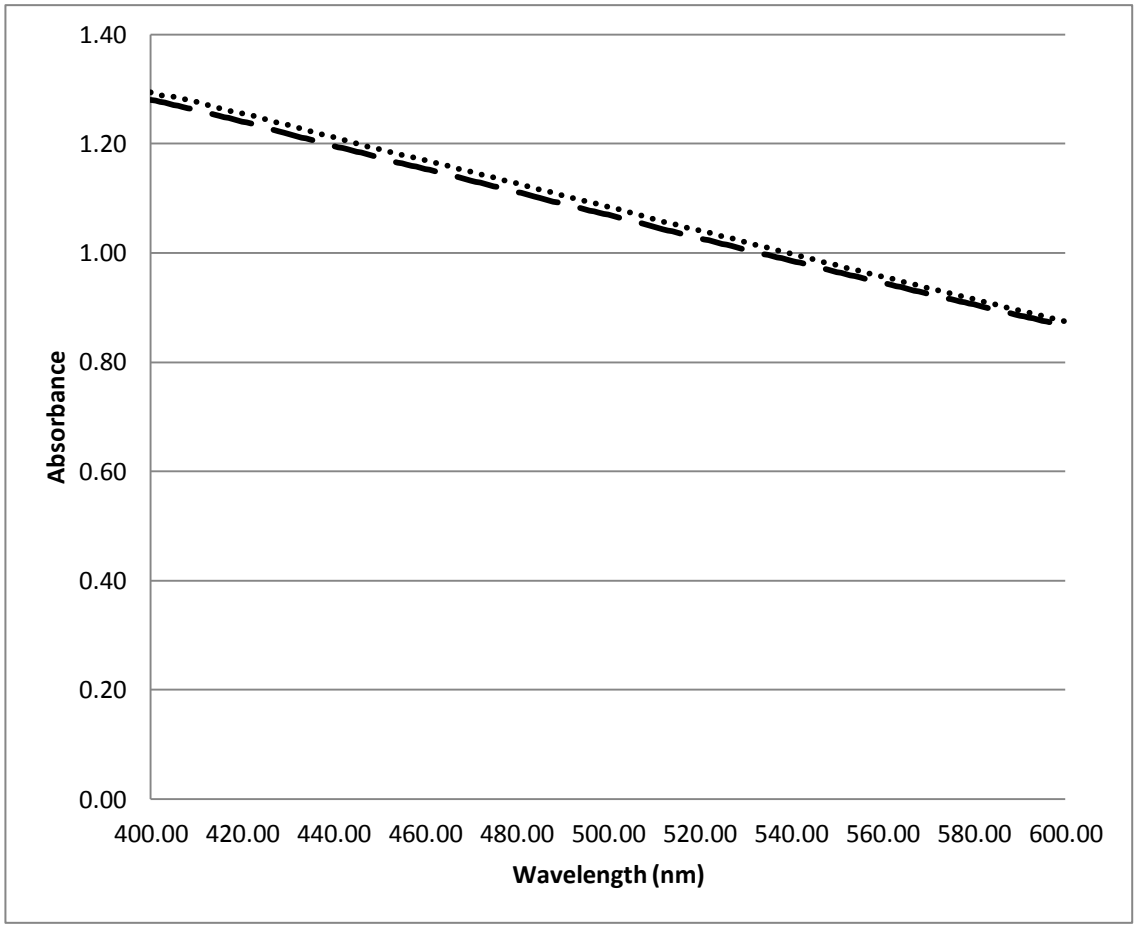


Figure 4.1. *The absorption spectra of the sample of ADP1 before and after mild sonication in a water bath. The dotted line represents the spectrum before sonication, while the dashed line represents the spectrum after sonication.*

Growth kinetics in filter culture

The experiment was performed without problems, and the OD_{600} measurements were done every 30 minutes, as planned. The growth curve of ADP1 on a succinate substrate on filter membranes can be seen in Figure 4.2. The fitting of a growth curve was difficult with traditional tools, since the culture is already at a logarithmic growth phase at time point 0 (filtration). This is why the points are presented without a traditional fitted growth curve.

The overall looks of the growth progression make sense, since the filtration is done at early logarithmic phase. Thus there is no clear lag phase in the beginning of the filter culture. The growth proceeds in a logarithmic way until about two hours, and then settles onto a plateau at approximately the OD_{600} value of 1. This is similar to the growth behavior of ADP1 in liquid succinate medium. This kind of behavior was expected, as it was suggested by previous experience at Genoscope on working with the filter cultures.

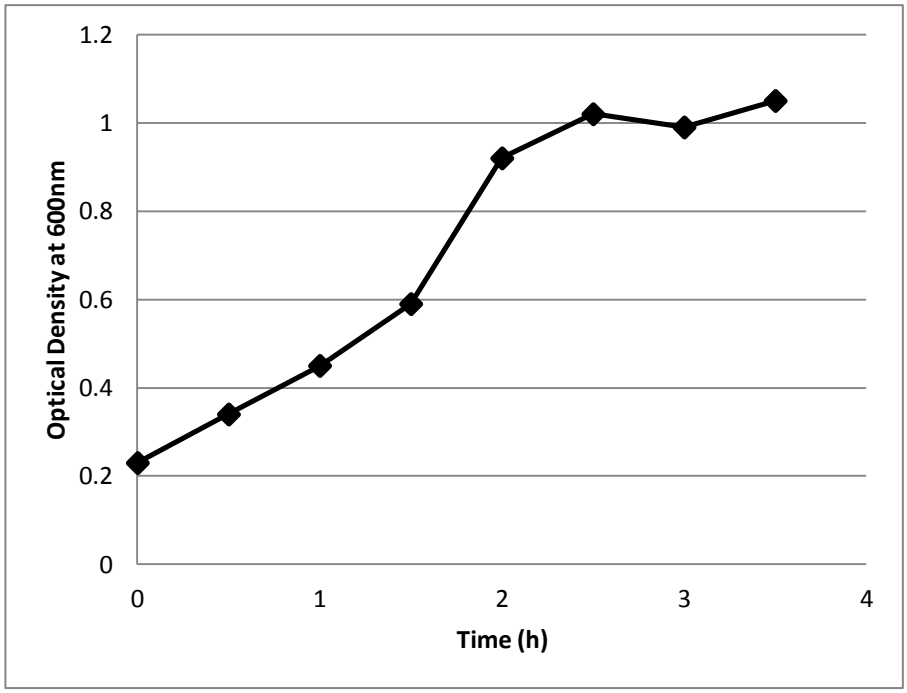


Figure 4.2. *Growth kinetics data of ADP1 grown on PTFE filter membranes with succinate as the carbon source.*

The growth kinetics for the quinate medium is shown in Figure 4.3. In these conditions, the cells took more time to reach the growth plateau than in the succinate medium. The plateau is also on a significantly higher OD_{600} level on quinate. The plateau was reached at OD_{600} value of 3. This could be explained by the higher amount of carbon atoms in the substrate.

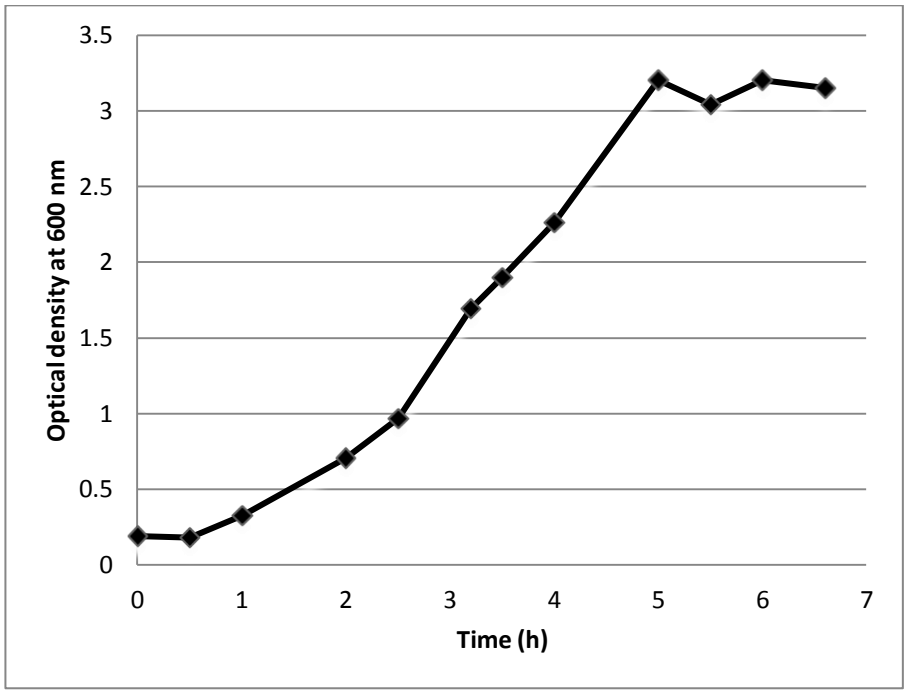


Figure 4.3. *Growth kinetics data of ADP1 grown on PTFE filter membranes with quinate as the carbon source.*

4.3 Quenching

The first differences between the solvents occurred already during the extraction work. The solvents with no water were significantly faster to work with due to the quicker evaporation of the solvent in the centrifugal evaporation. The evaporation took up to one hour shorter time with the methanol-based solvents than with the original water-based solvent. The condition involving chloroform, however, was noticed not to be a good solution with the current tube materials. The chloroform in the solvent broke a part of the tubes during the cell disruption step, and part of the cell extract was lost.

From the standard list used, approximately 60 compounds were identified in the best samples. Table 4.4 summarizes the amount of compounds identified in each sample. The total amount of metabolites is close to the same (60 metabolites on average) in all samples, except for the chloroform and acetonitrile-water-formic acid extractions, where it is slightly lower. In the extraction with formic acid, AMP ADP and ATP either were undetectable or had bad peak shape and low intensity. This was not expected, as in the literature formic acid has been shown to help the detection of phosphorylated compounds (Rabinowitz 2007/2). In the original acetonitrile-water protocol, more phosphorylated compounds were detected, compared to the other solvents. The complete qualitative results of each sample are shown in Appendix 2.

Table 4.4. *Number of standard compounds identified in each sample in the quenching solvent experiment. Samples A and B are duplicates, performed in a parallel way. ACN means acetonitrile, FA formic acid and MeOH methanol.*

Condition	Sample A	Sample B
ACN-water	61	59
ACN-water + FA	53	48
ACN-MeOH	60	63
MeOH	62	57
Chloroform-MeOH	55	55

The calibration curve of AMP, ADP and ATP is presented in Figure 4.4. The curve gives a good idea of the concentrations, but is probably not highly accurate due to strong tailing of the peaks in the calibration mixes. The tailing occurred already in small concentrations and was breaking the peak shape heavily in the higher concentrations. Integration of the peaks was done regardless of the tailing. The origin of the tailing is unknown. No tailing of this kind was detected in the biological samples. However, the calibration curve was considered accurate enough for the AEC analyses performed for this study.

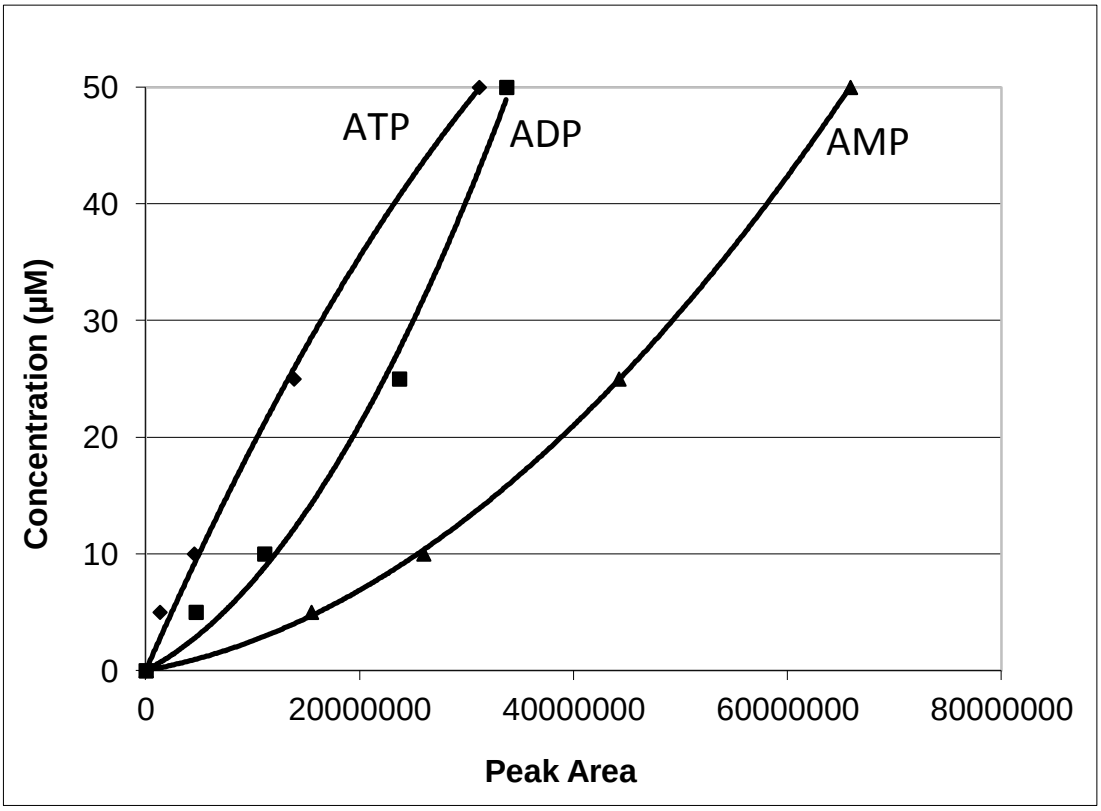


Figure 4.4. *Calibration curve of AMP, ADP and ATP from a LTQ-Orbitrap mass.*

The adenylate energy charge calculated with the help of the calibration curve revealed even more significant differences between the samples than the number of detected metabolites. The last three conditions had the best energy charges, whereas the original solvent used had a much lower charge. The energy charge of the original quenching solvent corresponds to the results obtained previously at Genoscope. In the sample with formic acid, the energy charge was not calculable since ATP, ADP and AMP were either not detected or had bad peaks in the data. Clearly the best energy charge was in the last condition, involving chloroform. Table 4.5 displays the computed energy charges with different solvents in each sample.

Table 4.5. *The adenylate energy charges in each sample in the quenching solvent experiment. A and B are duplicates, performed in a parallel way. ACN means acetonitrile, FA formic acid and MeOH methanol.*

Solvent	Sample A	Sample B
ACN-water	42,7%	38,5%
ACN-water + FA	N/A	N/A
ACN-MeOH	86,2%	85,8%
MeOH	83,8%	74,4%
Chloroform-MeOH	93,6%	94,5%

4.4 Sample storage

In the first sample storage experiment, it was noticed during the experiments that the prepared samples were not very stable when frozen and thawed again. The peak shape was considerably worse, and the intensity was systematically lower in stored samples, compared to those analyzed right after the preparation. In literature, similar observations can not be found. The results were systematical and concerned almost all metabolites found in the extractions. The degradation can be seen in the number of metabolites detected: in the sample analyzed right away, 27 metabolites were detected, whereas in the ones stored for 2 and 4 weeks, 25 and 12 metabolites were visible, respectively.

As an example of metabolite degradation, Figure 4.5 shows the chromatographic peaks of AMP found in the three samples.

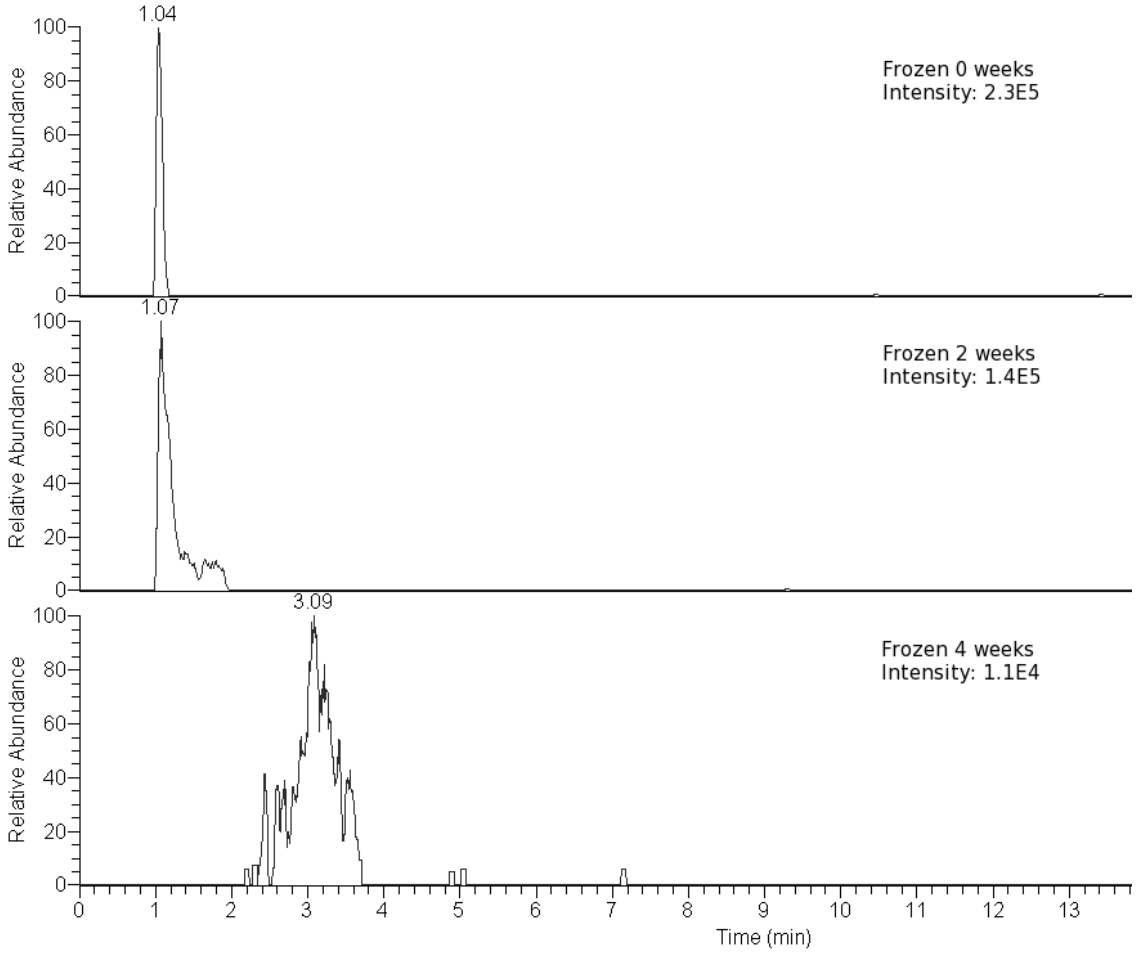


Figure 4.5. *Chromatographic peaks of AMP in the three samples, analyzed right away (uppermost panel) and 2 and 4 weeks (second and third panels, respectively) after sample preparation with the C18 column. Picture taken from Xcalibur. retention times are shown above the peaks, maximum intensities on the right side.*

The peak shape is significantly worse in the samples analyzed after storage. In the first sample, the peak is almost perfectly Gaussian, as it should be. In the sample analyzed after two weeks, the peak has somewhat retained its shape, but some tailing can be no-

ticed, suggesting minor chemical changes in the compound. The tailing of the peak also makes the quantification of the compound difficult, although it is still possible. Also the maximum intensity of the peak had decreased from 2.3E5 to 1.4E5, which could still be explained by sample-to-sample variation. In the last sample the peak shape is broken, and the retention time seems to be changed. The compound is not identifiable anymore in the data. The maximum intensity was only 1.1E4.

In the second experiment, it was noticed that after one week, the metabolite degradation is not yet very evident. In the dry-stored samples, 55 metabolites were detected (34 in the negative-ion mode, 34 in the positive-ion mode). The number is lower than in the previous experiments due to the C18 column used. The sample stored in the liquid phase contained 56 compounds from the standard list (32 in negative and 35 in positive-ion mode). The peaks were mostly Gaussian, and no significant peak tailing was noticed. However, in the case of ATP, a small tail was detected in the sample stored in the liquid phase.

4.5 Testing of the final protocol

The extractions were performed according to the new optimized protocol without problems in the workflow. The whole extraction process was completed in approximately 6 hours. The samples were analyzed about 7 days after the preparation due to technical reasons, but this did not seem to affect the results, as they were similar to the ones in previous experiments. The cells were growing slightly faster on quinate than succinate, which resulted in slightly higher final OD_{600} for the quinate samples. This may result in more detected metabolites in the data.

The quenching worked well for both of the samples. The AEC was 80.5% in the sample grown on succinate, and 85.7% for the one grown on quinate. These values correspond to the ones acquired in the quenching optimization experiments. The number of metabolites identified from the old standard list (Table 3.3) was also good. 73 metabolites were detected in the succinate sample, whereas 78 metabolites were present in the quinate one. From the standards added for this analysis, 3 metabolites were discovered in the succinate samples, and 2 in the quinate samples.

Differences were noticed in the metabolite compositions of the two samples. As expected, the cells grown on quinate contained protocatechuate which was not present in the succinate sample. Protocatechuate is an important metabolite in the degradation pathway of quinate. Other compounds that were identified with the help of standards in the quinate sample but not with succinate included p-aminobenzoate, UMP, CTP and GTP. No compounds on the standard list were exclusively present in the succinate samples.

From the quinate degradation pathway, out of 10 compounds, 5-7 were detected. 3-dehydroshikimate was not detected in the fingerprints, as were not 3-oxoadipyl-CoA and succinyl-CoA. Large size could explain the absence of the CoA derivatives in the data. 3-carboxymuconate and 2-carboxy-5-oxo-2,5-dehydrofuran-2-acetate have the

same exact mass, which was detected. It could not be determined from this data if only one of the compounds was present, or both. The other five compounds were detected with the corresponding masses. However, standard exists only for protocatechuate, which is thus the only one detected with certainty. The others were detected solely based on the mass, without knowing the correct retention times. None of the metabolites of the quinate degradation pathway were detected in the sample grown on succinate.

5 DISCUSSION

Culture conditions

For the cell recovery from the filter, the sonication was a significantly better method to resuspend the cells than the initially used scratching method. The OD_{600} values, and thus the number of cells after the first recoveries, were higher when the sonication method was used. In addition, the sonication worked better in the acetonitrile-water (80:20) solution than in the MA medium. No significant cell disruption was caused by the sonication.

If, however, the MA medium is preferred as a solvent, the sonication time should be increased to allow full detachment of cells from the filter and from each other. With the 5-minute sonication performed, all the cells were obviously not detached and some of them were heavily aggregated. Longer sonication times were not tested in MA medium, because the quenching solvent was decided to be used in the final protocol. It is faster to use the quenching solvent, as the cells do not need to be transferred between solvents. It is also possible that the difference in the OD_{600} value was due to the solvent. The spectrophotometer is known not to work as well with some solvents as others.

The reproducibility of the cell recovery was good in this experiment. No significant differences can be seen between the duplicate experiments, even after two cell recoveries. In the light of these results, the duplicate extractions in other experiments can be considered to represent the same condition. Based on this, the OD_{600} control plates are a valid method to measure the final cell density of the sample. However, in the second cell recoveries, significant amount of the cells was lost. This is probably due to aggregation of the cells on the glass surface.

The sonication is also a more reproducible method of cell recovery than scratching by tweezers in the sense that it is done completely by a sonication device, without human intervention. In initial tests it was noticed that when scratching the cells, significant differences were noticed from one scratcher to another, and even between the different plates of the same scratcher. In sonication, the ‘human factor’ is not present, and the results will be more comparable in the future, even if the performer of the extraction changes.

Because of the aggregation, it is recommended to keep the suspension of cells on the glass plate for as short time as possible. The cells seem to attach to the glass plate surface easily, and are not easy to fully resuspend afterwards. In addition, the aggregation is difficult to notice with naked eye before the cells have dried. Possibly due to this attachment, the second recovery in this experiment was not fully successful. Thus, cell suspension should be transferred into a cryotube in order to continue the extraction as quickly as possible after the cell recovery.

In the growth kinetics experiment, it was noticed that the bacteria grow on the filter membrane in a very similar way as in liquid culture: the exponential growth phase is

lasting about two hours after the filtration, during it the regeneration time of the bacteria is approximately one hour and the cells reach plateau at the OD_{600} value of 1 when using succinate as the carbon source. In other words, the growth on filter membrane follows roughly the same kinetics as in liquid culture. This suggests that the filter culture works well, and the metabolites can diffuse through the filter at a sufficient rate for the growth of the bacteria.

The growth kinetics data confirm that the original incubation time of two hours for the filter culture is good when working with succinate as the carbon source. After two hours, the culture is in the end of the logarithmic growth phase, at an OD_{600} value of 0.8. This is the optimal time for metabolite extraction, since the number of bacteria is then the highest possible with no stress caused by the limits of growth. (Chapman 1971) This is why, it is recommended to keep the incubation time at two hours, and to avoid exceeding it significantly. If the OD_{600} value at quenching is too close to the plateau, the extraction cannot be considered valid anymore.

In the quinate growth medium, however, the incubation time can be extended up to 4 hours in order to have more cells in the extraction. The plateau is reached later in these conditions, and thus it is possible to have the cells in the logarithmic growth phase for a longer time. However, if it is wanted that the OD_{600} value is the same in simultaneous succinate and quinate samples, the same 2-hour incubation time can be used as for the succinate medium. The growth rate in the beginning of the growth was similar in both samples.

The third part of the experiment showed that the sonication is safe to perform, and does not disrupt the cells. This means that with these settings the sonication can be safely used for cell recovery, only detaching the cells from the membrane. This is due to ADP1 being extraordinarily strong bacteria to break. Thus, the sonication method with the quenching solvent as a solvent is recommended to be used in the final protocol for cell recovery from the filter membrane.

Quenching

The preferred quenching condition, considering the prevailing laboratory conditions and equipment, was the acetonitrile-methanol (80:20) solvent at the working temperature of -30 °C. This condition was fast and reasonably easy to work with and produced a significantly better adenylate energy charge than the original protocol. In addition it produced the highest number of metabolites detected. The solvent did not cause any problems to the workflow.

The quenching solvent used in the original protocol resulted in a too low energy charge, suggesting insufficient quenching efficiency. The condition resulted in a greater amount of detected phosphorylated compounds, but can not be used in the final protocol due to the energy charge. The detection of less phosphorylated compounds can also result from slightly dirtier ion source in the other experiments. The original solvent with added formic acid did not prove to improve the stability of phosphorylated compounds in ADP1 extracts, as suggested in earlier literature (Rabinowitz 2007/2), vice versa. In

addition, both of these solvents were one hour slower to work with, which is an important quality when the extractions will be done in bigger scale.

The 100% methanol solvent was good in terms of workflow, causing no problems. Also the energy charge and number of detected metabolites were nearly as good as for the preferred acetonitrile-methanol solvent. Acetonitrile-methanol was, however preferred due to the slightly better results and fact that 100% methanol in the protocol requires more safety precautions. The chloroform-methanol solvent had the best energy charge of the experiment. However, the solvent caused severe problems to the workflow, breaking tubes during the cell disruption step. In addition, chloroform can be more hazardous to health than acetonitrile and methanol. Consequently, as the energy charge of the acetonitrile-methanol solvent was considered adequate, it was preferred as the final quenching solvent.

Normalization of the data was not performed, which is a minor drawback for these results. The quantitative analysis would have brought a new level of information to the dataset, and the comparison of the different methods could have been done more precisely, comparing the conditions in groups of different classes of compounds. A method for normalization needs to be set up in the future. The simplest way would be to measure the final OD_{600} of the bacteria from a separate plate in the end of the culture, and normalize the results according to that. However, a mathematical normalization method would be more conclusive.

Sample Storage

It was noticed that the storage of samples can potentially cause chemical changes to the metabolites, and cause the chromatographic peaks not to be identifiable anymore. Already two weeks of storage dry at -20 ºC was seen to cause these effects, elevated after 4 weeks of storage. One week of storage both in dry and liquid form, however, resulted in no significant changes in the sample. Thus, it is recommended that in the future the samples are analyzed right after the preparation, or at latest one week later, to get best results. Longer storage may cause severe bias to the data, and undermine the research.

In previous literature it can be seen that other researchers have stored their extraction samples dry for up to two weeks at -20 ºC without paying attention to the possible degradation (e.g. Jozefczuk 2010). Thus obtained results could be very biased or insufficient without assessment of the impact of storage on the sample composition. It can be seen that in metabolomics it is important not to blindly trust the previous studied performed in different laboratories, but to test the applicability of the method to your own conditions yourself.

Final Protocol

The final protocol obtained in the study can be viewed on Figure 5.1. The biggest difference to the original protocol is the changing of the quenching solvent and temperature. This change improved the protocol the most by making the final extract more representative of the state inside the cell in normal conditions. It also made possible the

reduction of the centrifugal evaporation time to approximately one hour, due to the lower boiling point of the methanol, compared to water. This saves one hour in the time to complete the protocol. The protocol was made more efficient also in other ways, related to the workflow, during the optimization. These changes included mainly questions related to organization and productivity. The time used for a single extraction could be reduced by up to 3 hours. This is an important point when the protocol is used in larger scale in the future.

In addition, cell recovery from filters after quenching was changed from scratching to sonication. In addition to recovering more cells into the final solution, sonication also is less dependent on the human factor (strength and duration of scratching), which makes the results more comparable in the long run. In the final protocol the samples are recommended to be analyzed right after the preparation or at latest during the next seven days, in order to avoid degradation of certain compounds during storage. The incubation time of the filter culture was also assessed, and it was confirmed that two hours is the optimal incubation time when grown on succinate medium. For data analysis, the XCMS algorithm centWave is recommended with the parameter values presented in Table 4.1.

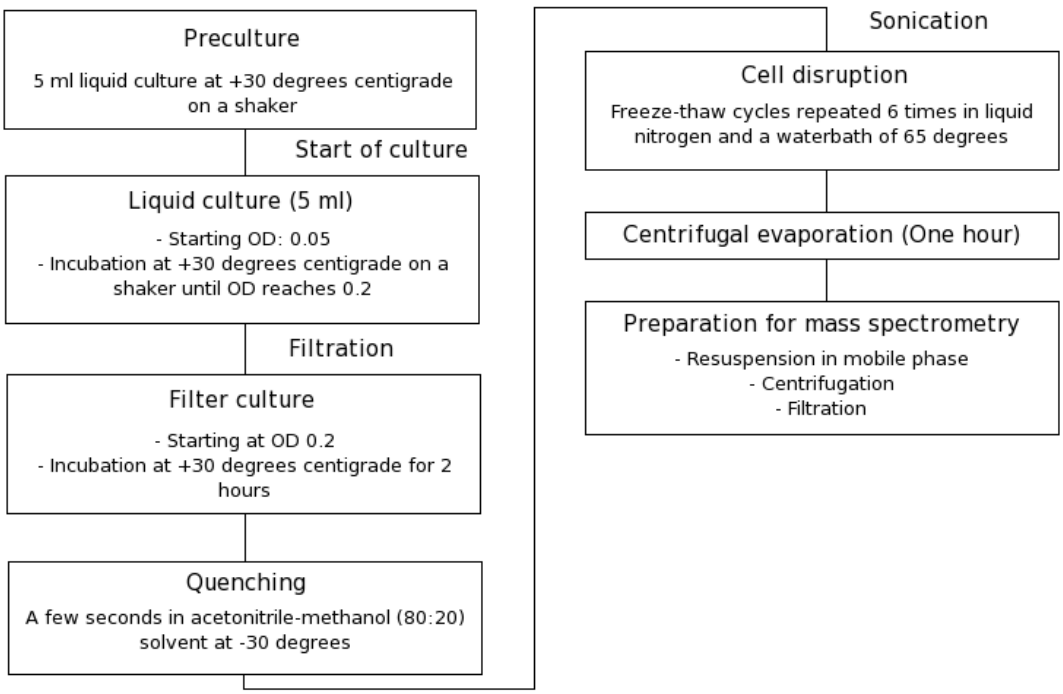


Figure 5.1. *The final optimized extraction protocol for ADP1 intracellular metabolites.*

Testing of the protocol

The AECs of both of the samples prepared, as well as the number of metabolites identified, suggest that the new protocol works very well. The AECs are in the same level as in the quenching experiment, which shows that the quenching in the new conditions is reproducible. The number of metabolites is even significantly higher than previously. This may be due to the slightly higher final OD_{600} of the bacteria, as compared to the

previous experiments. The cell density can also explain the detection of CTP and GTP in the quinate sample, and not in the succinate one. Both of them have been detected on succinate before. Another explanation for the increased number of metabolites could be the new clean ion source that was installed before the injections. It could provide more accurate detection in the new samples.

The three other compounds detected exclusively in the quinate sample, protocatechuate, p-aminobenzoate and UMP, have never been seen in succinate samples. Protocatechuate was expected as a degradation product of quinate. P-aminobenzoate could be also related to these pathways, as it is an aromatic compound like quinate. However, no known pathway is connecting them. No logical explanation could be found for the presence of UMP.

The quinate degradation pathway was detected in the samples grown on quinate nearly completely. Only three compounds of the pathway were surely not detected. This suggests that the pathway is valid, and that the extraction protocol is sufficiently efficient to detect these compounds. In addition, these compounds were detected only in the quinate samples, which was expected. This shows that the compounds were closely related to quinate metabolism. Thus, even without the standard solutions, it can be concluded with near certainty that these are the suggested quinate degradation intermediates. The compounds that were not detected could be explained by low concentrations or compartmentation of the compounds outside the cell.

In a similar study with *E. coli*, 141 metabolites were identified in an LC-MS/MS study (Bajad 2006). The maximum of 78 detected metabolites in the present study is much lower compared to these results. Especially a far greater amount of nucleotides and other phosphorylated compounds were detected in the *E. coli* study. In addition, 69 of the detected compounds were successfully quantified. The better results can be explained by the more powerful MS/MS method used in the Bajad *et aI.* study. Also, the standard list used in the experiment included 164 metabolites, as compared to the 122 metabolites in this study. This is why it could be interesting to further expand the amount of standard compounds with the help of the metabolic model of ADP1 (Durot 2008). This has already been started at Genoscope.

In addition, to further enhance the results of ADP1 metabolomics, MS/MS studies and isotopic studies are planned. These analyses could greatly improve the precision of the analysis, and possibly many new metabolites could be detected from similar ADP1 metabolomic samples. The methods related to these analyses, however, still require further development at the Genoscope laboratories, and were thus not conducted within the timeframe of this study.

Another method to improve the results presented here would be to use a thinner HILIC column in the LC-MS analyses. In a recent study, a ZIC-pHILIC column of the inner diameter of 2.1 mm, instead of the 4.6 mm used in this study, showed much better results. When comparing these two methods, the thinner column resulted in almost 50% more annotations. (t'Kindt 2010) This could be interesting also in the ADP1 studies, and the method is currently planned at Genoscope.

In additional notes, the cleanliness of the ion source of the Orbitrap device was seen to be of great importance for the results of the analysis. The samples were run both with a dirty and a clean ion source, and significant differences were found in the results. The peak areas of certain phosphorylated compounds were increased up to 1000-fold after the clean ion source was installed. Also the number of detected metabolites was increased significantly. On this basis, it is recommended that the ion source is cleaned regularly before the injections of biological samples.

In the future it could be interesting to compare the data of cells grown on two different carbon sources with a blind approach of XCMS. With this method, unexpected compounds appearing only in one of the data sets could be detected, and new, previously unknown pathways could be discovered. In addition to succinate and quinate, also other carbon sources could be tested. This strategy could provide results that cannot be expected beforehand, thanks to the blind approach.

It would be interesting in the future projects at Genoscope to try to integrate the data of the metabolomics and transcriptomics to a single set of data, in the case of different carbon sources. These kinds of integrations have been done before in literature with different bacterial strains (Zhang 2010), and would be interesting to do in the case of ADP1 in order to gain further knowledge about the consequences of changing the carbon source to the bacteria.

An interesting approach to the ADP1 characterization would also be using the knock-out mutant library to compare the metabolism of a mutant to that of the wild type correspondent. When a mutant where an unknown gene is knocked out is chosen to the study, the metabolomic profile could reveal the function of the deleted gene. The cell could be depleted of the product of the enzyme coded by the deleted gene, or the substrate of the enzyme could accumulate. Afterwards, this hypothetical annotation could be tested with help of the ORFeome. Thus it could be possible to further annotate the genome of ADP1.

6 CONCLUSIONS

The optimized protocol for extraction of intracellular metabolites from ADP1 was created successfully. The changes made to the protocol improved the quality of the data significantly, and the test extractions showed that the results are well reproducible. This makes it possible for the Genoscope laboratories to continue their project of ADP1 characterization also in the area of metabolomics. In addition to better results, the protocol is up to three hours faster to perform than in the beginning of the project, which makes it easier to develop ADP1 extractions in a larger scale.

The test extractions provided the identification of almost 80 metabolites in ADP1, which is a great improvement to the earlier results. This number can be improved in the future by a more extensive list of standard compounds, MS/MS methods, as well as with isotope studies. In addition, new interesting possibilities, such as exploration of new metabolic pathways and new annotation of ADP1 genes, have opened because of the new protocol.

To further develop the method in the future, a way to normalize the data needs to be set up. This is essential when quantifications of the metabolites need to be performed. The quantification is essential when comparisons of different samples are performed. Also, to further reduce the time needed for the extraction, the extraction step needs to be optimized. Less than six freeze-thaw cycles could be enough to break the cells. These experiments were unfortunately not possible to perform within the timeframe of the internship.

REFERENCES

Abdel-El-Haleem, D. 2003. Acinetobacter: environmental and biotechnological applications. African Journal of Biotechnology 2, 4, pp. 71-74.

Alfassi, Z.B. 2004. On the normalization of a mass spectrum for comparison of two spectra. Journal of the American Society for Mass Spectrometry 15, 3, pp. 385-387.

Annesley, T.M. 2003. Ion Suppression in Mass Spectrometry. Clinical Chemistry, 49, 7, pp. 1041-1044.

Bajad, S.U., Lu, W., Kimball, E.H., Yuan, J., Peterson, C. & Rabinowitz, J.D. 2006. Separation and quantitation of water soluble cellular metabolites by hydrophilic interaction chromatography-tandem mass spectrometry. Journal of Chromatography A 1125, 1, pp. 76-88.

Barbe, V., Vallenet, D., Fonknechten, N., Kreimeyer, A., Oztas, S., Labarre, L., Cruveiller, S., Robert, C., Duprat, S., Wincker, P., Ornston, L.N., Weissenbach, J., Marlière, P., Cohen, G.N. & Médigue, C. 2004. Unique features revealed by the genome sequence of Acinetobacter sp. ADP1, a versatile and naturally transformation competent bacterium. Nucleic Acids Research 32, 19, pp. 5766-5779.

Bolten, C.J., Kiefer, P., Letisse, F., Portais, J.C. & Wittmann, C. 2007. Sampling for metabolome analysis of microorganisms. Analytical Chemistry 79, 10, pp. 3843-3849.

Brauer, M.J., Yuan, J., Bennett, B.D., Lu, W., Kimball, E., Botstein, D. & Rabinowitz, J.D. 2006. Conservation of the metabolomic response to starvation across two divergent microbes. Proceedings of the National Academy of Sciences of the United States of America 103, 51, pp. 19302-19307.

Canelas, A.B., ten Pierick, A., Ras, C., Seifar, R.M., van Dam, J.C., van Gulik, W.M. & Heijnen, J.J. 2009. Quantitative evaluation of intracellular metabolite extraction techniques for yeast metabolomics. Analytical Chemistry 81, 17, pp. 7379-7389.

Cech, N.B. & Enke, C.G. 2001. Practical implications of some recent studies in electrospray ionization fundamentals. Mass Spectrometry Reviews, 20, 6, pp. 362-387.

Chapman, A.G., Fall, L. & Atkinson, D.E. 1971. Adenylate energy charge in Escherichia coli during growth and starvation. Journal of Bacteriology 108, 3, pp. 1072-1086.

de Berardinis, V., Durot, M., Weissenbach, J. & Salanoubat, M. 2009. Acinetobacter baylyi ADP1 as a model for metabolic system biology. Current Opinion in Microbiology 12, 5, pp. 568-76.

de Berardinis, V., Vallenet, D., Castelli, V., Besnard, M., Pinet, A., Cruaud, C., Samair, S., Lechaplais, C., Gyapay, G., Richez, C., Durot, M., Kreimeyer, A., Le Fèvre, F., Schächter, V., Pezo, V., Döring, V., Scarpelli, C., Médigue, C., Cohen, G.N., Marlière, P., Salanoubat, M. & Weissenbach, 2008. J. A complete collection of single-gene deletion mutants of Acinetobacter baylyi ADP1. Molecular Systems Biology 4, 174.

Dettmer, K., Aronov, P.A. & Hammock, B.D. 2007. Mass spectrometry-based metabolomics. Mass Spectrometry Reviews 26, 1, pp. 51-78.

Dunn, W.B., Bailey, N.J. & Johnson, H.E. 2005. Measuring the metabolome: current analytical technologies. The Analyst 130, 5, pp. 606-625.

Durot, M., Le Fèvre, F., de Berardinis, V., Kreimeyer, A., Vallenet, D., Combe, C., Smidtas, S., Salanoubat, M., Weissenbach, J. & Schachter, V. 2008. Iterative reconstruction of a global metabolic model of Acinetobacter baylyi ADP1 using high-throughput growth phenotype and gene essentiality data. BMC Systems Biology 2, 85.

Ewald, J.C., Heux, S. & Zamboni, N. 2009. High-throughput quantitative metabolomics: workflow for cultivation, quenching, and analysis of yeast in a multiwell format. Analytical Chemistry 81, 9, pp. 3623-3629.

Geissdorfer, W., Kok, R.G., Ratajczak, A., Hellingwerf, K.J. & Hillen, W. 1999. The genes rubA and rub for alkane degradation in Acinetobacter sp. strain ADP1 are in an operon with estB, encoding an esterase, and oxyR. Journal of Bacteriology 181, 4292-4298.

Gerischer U. 2008. Acinetobacter Molecular Biology. Norfolk, UK, Caister Academic Press, pp. 100-105.

Griffiths, W.J., Karu, K., Hornshaw, M., Woffendin, G. & Wang, Y. 2007. Metabolomics and metabolite profiling: past heroes and future developments. European Journal of Mass Spectrometry 13, 1, pp. 45-50.

Heinemann, M. & Zenobi, R. 2011. Single cell metabolomics. Current Opinion in Biotechnology 22, 1, pp. 26-31.

Hommerson, P., Khan, A.M., de Jong, G.J. & Somsen, G.W. 2011. Ionization techniques in capillary electrophoresis-mass spectrometry: Principles, design, and application. Mass Spectrometry Reviews, Epub 2011 Apr 1.

Horning, E.C. & Horning, M.G. 1971. Human metabolic profiles obtained by GC and GC/MS. Journal of Chromatographic Science 9, pp. 129–140.

Ishii, N., Nakahigashi, K., Baba, T., Robert, M., Soga, T., Kanai, A., Hirasawa, T., Naba, M., Hirai, K., Hoque, A., Ho, P.Y., Kakazu, Y., Sugawara, K., Igarashi, S., Harada, S., Masuda, T., Sugiyama, N., Togashi, T., Hasegawa, M., Takai, Y., Yugi, K., Arakawa, K., Iwata, N., Toya, Y., Nakayama, Y., Nishioka, T., Shimizu, K., Mori, H. & Tomita, M. 2007. Multiple high-throughput analyses monitor the response of E. coli to perturbations. Science 316, 5824, pp. 593-597.

Johnsen, E., Wilson, S.R., Odsbu, I., Krapp, A., Malerod, H., Skarstad, K. & Lundanes, E. 2011. Hydrophilic interaction chromatography (HILIC) of nucleoside triphosphates with temperature as a separation parameter. Journal of Chromatography A, Epub 2011 Jan 27.

Jozefczuk, S., Klie, S., Catchpole, G., Szymanski, J., Cuadros-Inostroza, A., Steinhauser, D., Selbig, J. & Willmitzer, L. 2010. Metabolomic and transcriptomic stress response of Escherichia coli. Molecular Systems Biology 11, 6, 364.

Katajamaa, M. & Oresic, M. 2005. Processing methods for differential analysis of LC/MS profile data. BMC Bioinformatics 6, 179.

Keller, B.O., Sui, J., Young, A.B. & Whittal, R.M. 2008. Interferences and contaminants encountered in modern mass spectrometry. Analytica Chimica Acta 627, 1, pp.71-81.

Link, H., Anselment, B. & Weuster-Botz, D. 2008. Leakage of adenylates during cold methanol/glycerol quenching of Escherichia coli. Metabolomics 4, pp. 240–247

Makarov, A. & Scigelova, M. 2010. Coupling liquid chromatography to Orbitrap mass spectrometry. Journal of Chromatography A 1217, 25, pp. 3938-3945.

Marcinowska, R., Trygg, J., Wolf-Watz, H., Mortiz, T. & Surowiec, I. 2011. Optimization of a sample preparation method for the metabolomic analysis of clinically relevant bacteria. Journal of Microbiological Methods, Epub 2011 Jul 13.

Metzgar, D., Bacher, J.M., Pezo, V., Reader, J., Döring, V., Schimmel, P., Marlière, P. & de Crécy-Lagard, V. 2004. Acinetobacter sp. ADP1: an ideal model organism for

genetic analysis and genome engineering. Nucleic Acids Research 32, 19, pp.5780-5790.

Ohta, D., Kanaya, S. & Suzuki, H. 2010. Application of Fourier-transform ion cyclotron resonance mass spectrometry to metabolic profiling and metabolite identification. Current Opinion in Biotechnology 21, 1, pp. 35-44.

Palmen, R. & Hellingwerf, K.J. 1997. Uptake and processing of DNA by *Acinetobacter calcoaceticus*--a review. Gene 192, 1, pp. 179-190.

Paul Lee, W.N., Wahjudi, P.N., Xu, J. & Go, V.L. 2010. Tracer-based metabolomics: concepts and practices. Clinical Biochemistry 43, 16-17, pp. 1269-1277.

Perry, R.H., Cooks, R.G. & Noll, R.J. 2008. Orbitrap mass spectrometry: instrumentation, ion motion and applications. Mass Spectrometry Reviews 27, 6, pp. 661-699.

Pieterse, B., Jellema, R.H. & van der Werf, M.J. 2006. Quenching of microbial samples for increased reliability of microarray data. Journal of Microbiological Methods 64, 2, pp. 207-216.

Rabinowitz, J.D. 2007. Cellular metabolomics of Escherichia coli. Expert Review of Proteomics 4, 2, pp. 187-198.

Rabinowitz, J.D. & Kimball, E. 2007. Acidic acetonitrile for cellular metabolome extraction from Escherichia coli. Analytical Chemistry 79, 16, pp. 6167-6173.

Rujiter, G.J.G. & Visser, J. 1996. Determination of intermediary metabolites in *Aspergillus Niger*. Journal of Microbiological Methods 25, pp. 295-302.

Saito, N., Robert, M., Kitamura, S., Baran, R., Soga, T., Mori, H., Nishioka, T. & Tomita, M. 2006. Metabolomics approach for enzyme discovery. Journal of Proteome Research 5, 8, pp. 1979-1987.

Santala, S., Efimova, E., Kivinen, V., Larjo, A., Aho, T., Karp, M. & Santala, V. 2011. Improved Triacylglycerol Production in Acinetobacter baylyi ADP1 by Metabolic Engineering. Microbial Cell Factories 10, 36.

Scheltema, R.A., Kamleh, A., Wildridge, D., Ebikeme, C., Watson, D.G., Barrett, M.P., Jansen, R.C. & Breitling, R. 2008. Increasing the mass accuracy of high-resolution LC-MS data using background ions: a case study on the LTQ-Orbitrap. Proteomics 8, 22, pp. 4647-4656.

Shin, M.H., Lee do, Y., Liu, K.H., Fiehn, O. & Kim, K.H. 2010. Evaluation of sampling and extraction methodologies for the global metabolic profiling of Saccharophagus degradans. Analytical Chemistry, 82, 15, pp. 6660-6666.

Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. & Siuzdak, G. 2006. XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification. Analytical Chemistry 78, pp. 779-787.

Smith, C.A. 2009. XCMS User Guide. [WWW] (Cited on June 17th, 2011, available at bioconductor.org/packages/2.6/bioc/manuals/xcms/man/xcms.pdf).

Spagou, K., Tsoukali, H., Raikos, N., Gika, H., Wilson, I.D. & Theodoridis, G. 2010. Hydrophilic interaction chromatography coupled to MS for metabonomic/metabolomic studies. Journal of Separation Science 33, 6-7, pp.716-727.

Tautenhahn, R., Böttcher, C. & Neumann S. 2008. Highly sensitive feature detection for high resolution LC/MS. BMC Bioinformatics 9, 504.

Taylor, W.H. & Juni, E. 1961. Pathways for biosynthesis of a bacterial capsular polysaccharide. I. Characterization of the organism and polysaccharide. Journal of Bacteriology 81, pp. 688-693.

Taymaz-Nikerel, H., de Mey, M., Ras, C., ten Pierick, A., Seifar, R.M., van Dam, J.C., Heijnen, J.J. & van Gulik, W.M. 2009. Development and application of a differential method for reliable metabolome analysis in Escherichia coli. Analytical Biochemistry 386, 1, pp. 9-19.

t'Kindt, R., Jankevics, A., Scheltema, R.A., Zheng, L., Watson, D.G., Dujardin, J.C., Breitling, R., Coombs, G.H. & Decuypere, S. 2010. Towards an unbiased metabolic profiling of protozoan parasites: optimisation of a Leishmania sampling protocol for HILIC-orbitrap analysis. Analytical and Bioanalytical Chemistry 398, 5, pp. 2059-2069.

van der Werf, M.J., Overkamp, K.M., Muilwijk, B., Koek, M.M., van der Werff-van der Vat, B.J., Jellema, R.H., Coulier, L. & Hankemeier, T. 2008. Comprehensive analysis of the metabolome of Pseudomonas putida S12 grown on different carbon sources. Molecular Biosystems 4, 4, pp. 315-327.

van Vliet, A.H. 2010. Next generation sequencing of microbial transcriptomes: challenges and opportunities. FEMS Microbiology Letters 302, 1, pp. 1-7.

Villas-Bôas, S.G., Mas, S., Åkesson, M., Smedsgaard, J. & Nielsen, J. 2005. Mass spectrometry in metabolome analysis. Mass Spectrometry Reviews 24, 5, pp. 613-646.

Wellerdiek, M., Winterhoff, D., Reule, W., Brandner, J. & Oldiges, M. 2009. Metabolic quenching of Corynebacterium glutamicum: efficiency of methods and impact of cold shock. Bioprocess and Biosystems Engineering 32, 5, pp. 581-592.

Werner, E., Heilier, J.F., Ducruix, C., Ezan, E., Junot, C. & Tabet, J.C. 2008. Mass spectrometry for the identification of the discriminating signals from metabolomics: current status and future trends. Journal of chromatography B, Analytical technologies in the biomedical and life sciences 871, 2, pp. 143-63.

Winder, C.L., Dunn, W.B., Schuler, S., Broadhurst, D., Jarvis, R., Stephens, G.M. & Goodacre, R. 2008. Global metabolic profiling of Escherichia coli cultures: an evaluation of methods for quenching and extraction of intracellular metabolites. Analytical Chemistry 80, 8, pp. 2939-48.

Young, D.M., Parke, D. & Ornston, L.N. 2005. Opportunities for genetic investigation afforded by Acinetobacter baylyi, a nutritionally versatile bacterial species that is highly competent for natural transformation. Annual Review of Microbiology 59, pp. 519-51.

Yuan, J., Fowler, W.U., Kimball, E., Lu, W. & Rabinowitz, J.D. 2006. Kinetic flux profiling of nitrogen assimilation in Escherichia coli. Nature Chemical Biology 2, 10, pp. 529-530.

Zhang, W., Li, F. & Nie, L. 2010. Integrating multiple 'omics' analysis for microbial biology: application and methodologies. Microbiology 156, 2, pp. 287-301.

APPENDIX 1: DETECTED RETENTION TIMES OF ALL COMPOUNDS IN THE STANDARD LIST IN BOTH COLUMNS

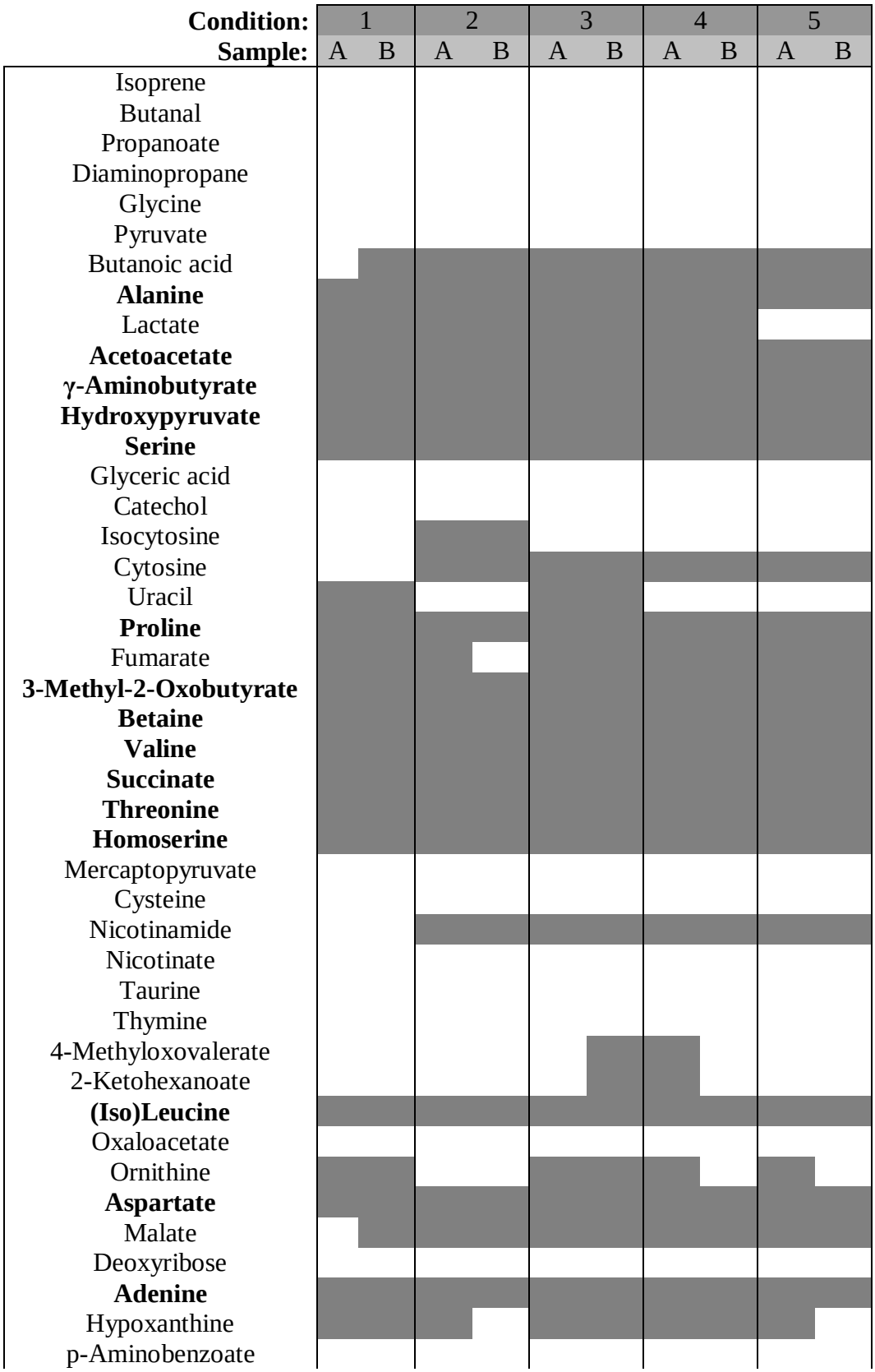
Column:	**C18**		**HILIC**	
Ionization mode:	Positive	Negative	Positive	Negative
Isoprene				
Butanal				
Propanoate	1.22		4.47	
Diaminopropane		0.89		
Glycine		0.96	11.28	
Pyruvate	1.02		4.24	
Butanoic acid	1.83		3.58	
Alanine	0.96	0.97	10.63	10.60
Lactate	0.99		5.16	
Acetoacetate	1.06		10.77	
γ-Aminobutyrate	0.97	1.02	10.03	11.17
Hydroxypyruvate	0.90		6.89	
Serine	0.94	0.96	11.30	11.34
Glyceric acid	0.94		7.51	
Catechol	4.60		3.78	
Isocytosine	1.38	1.21	7.20	7.19
Cytosine	1.18	1.66	7.74	7.64
Uracil	1.35	1.39	5.36	
Proline	1.07	1.10	9.10	9.09
Fumarate	0.90		11.16	
3-Methyl-2-Oxobutyrate	1.40		7.75	5.09
Betaine		1.00	7.80	7.79
Valine	1.21	1.25	8.91	8.88
Succinate	1.12		10.50	
Threonine	0.99	1.00	10.80	10.84
Homoserine	0.96	0.97	10.78	10.80
Mercaptopyruvate	1.07		11.54	
Cysteine		1.17		11.94
Nicotinamide	3.38	2.66	4.36	4.34
Nicotinate	1.74	1.56	3.87	3.86
Taurine	0.95	0.97	10.31	10.33
Thymine	2.62	4.14	4.31	4.29
4-Methyloxovalerate	2.63		2.72	
2-Ketohexanoate	2.67		2.69	
(Iso)Leucine	1.80	1.94	7.95	7.94
Oxaloacetate	0.91		14.15	
Ornithine		0.91	18.84	18.85
Aspartate	0.92	0.96	10.35	10.38
Malate	0.92	1.43	11.19	
Deoxyribose				

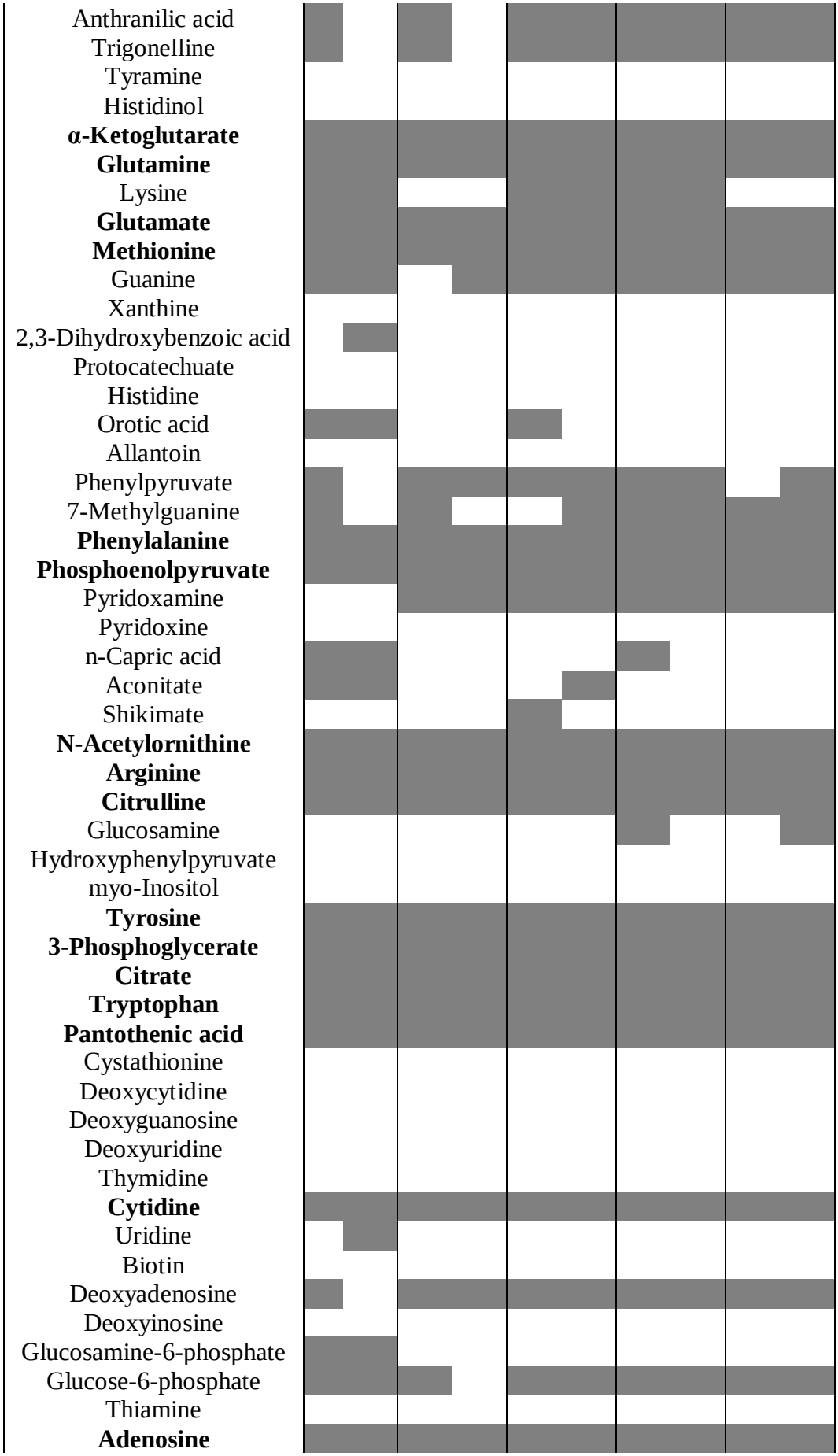
Adenine	2.74	1.67	6.27	6.26
Hypoxanthine	1.62	1.58	6.77	7.67
p-Aminobenzoate	1.72	4.29	6.19	
Anthranilic acid	3.13	6.40	4.10	8.24
Trigonelline		1.05	8.23	8.23
Tyramine				
Histidinol				
α-Ketoglutarate	0.97		16.44	
Glutamine		1.35		10.68
Lysine		0.92	20.54	20.55
Glutamate	0.93	0.98	10.07	10.07
Methionine	1.40	1.44	8.08	8.09
Guanine	1.59	3.60	8.67	8.69
Xanthine	1.60		7.44	
2,3-Dihydroxybenzoic acid	1.49		4.18	
Protocatechuate	1.89		8.94	
Histidine	1.18	0.95	12.84	
Orotic acid	1.04		5.87	
Allantoin	1.02	1.02	9.62	
Phenylpyruvate	3.81	1.64	2.62	9.29
7-Methylguanine	3.34	2.51	6.50	6.49
Phenylalanine	3.82	3.88	6.88	6.85
Phosphoenolpyruvate	0.91	0.99	12.62	
Pyridoxamine	1.91			8.74
Pyridoxine	3.23	2.02	4.84	4.82
n-Capric acid	10.92	9.96	2.38	
Aconitate	0.90		12.69	7.40
Shikimate	0.95		9.47	
N-Acetylornithine	1.05		11.05	11.08
Arginine		0.96	22.02	22.08
Citrulline	1.00	1.08	11.52	11.54
Glucosamine		0.96		11.33
Hydroxyphenylpyruvate	1.91	10.86	11.41	
myo-Inositol	0.99		12.28	
Tyrosine	1.57	1.64	9.28	9.29
3-Phosphoglycerate	0.85		12.24	
Citrate	0.92	1.11	13.08	
Tryptophan	4.52	4.68	8.32	8.27
Pantothenic acid	1.93	4.51	4.50	4.51
Cystathionine	0.95	0.97	12.81	12.42
Deoxycytidine	2.03	1.66	7.17	7.16
Deoxyguanosine				
Deoxyuridine	2.65	2.53	5.06	
Thymidine	4.16	4.13	4.22	
Cytidine	1.55	1.42	8.37	8.38
Uridine	1.96	1.88	6.52	

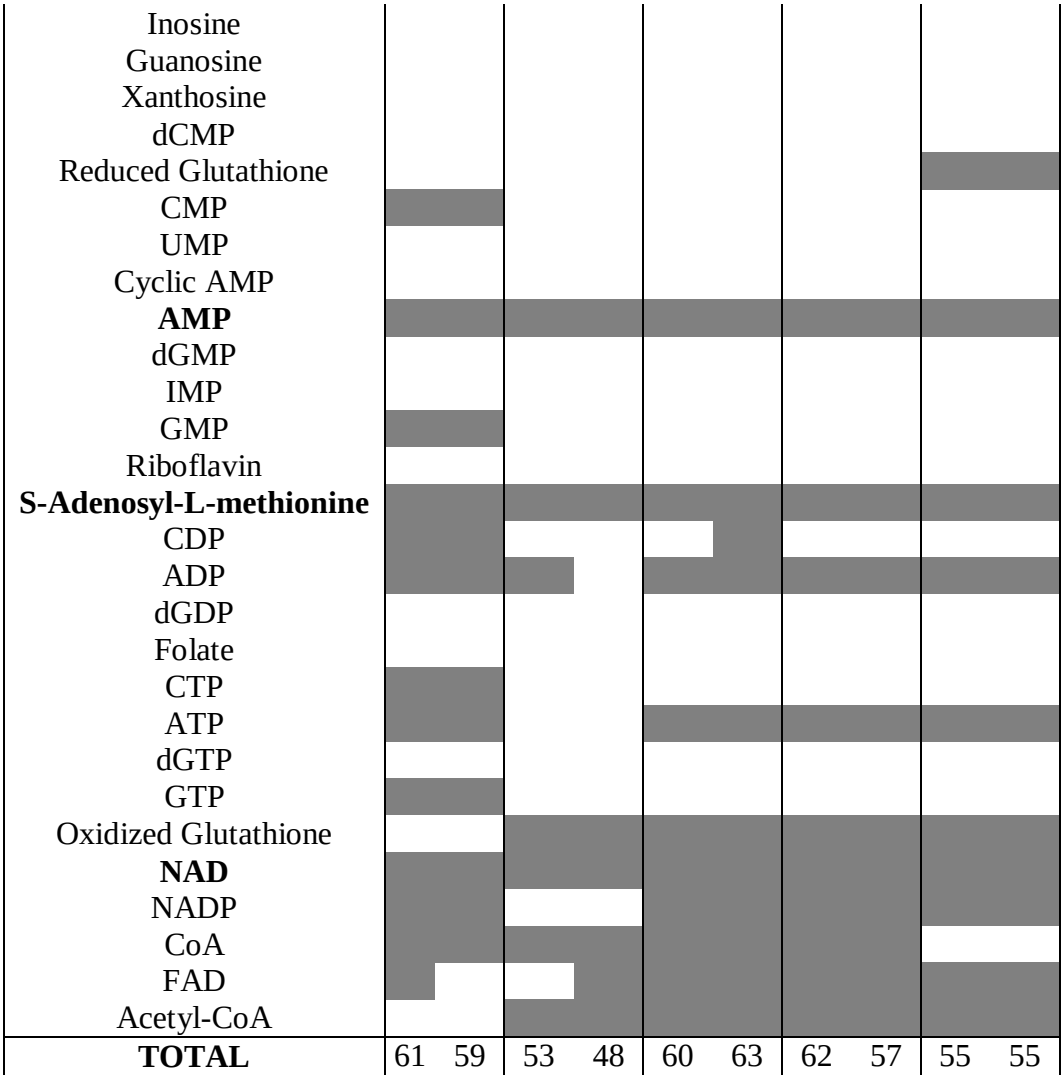
Biotin	3.09	6.10	4.74	4.65
Deoxyadenosine	4.50	4.16	5.02	5.05
Deoxyinosine	3.40		6.29	6.28
Glucosamine-6-phosphate	0.87		11.51	11.56
Glucose-6-phosphate	0.92	0.91	11.91	11.93
Thiamine		1.49		20.78
Adenosine	4.36	4.03	5.84	5.81
Inosine	3.13	2.97	7.49	7.49
Guanosine	3.31	3.11	8.96	8.96
Xanthosine	2.17	3.71	7.94	7.94
dCMP	1.03		10.26	10.28
Reduced Glutathione		1.00	13.28	13.31
CMP	0.94		11.08	11.09
UMP	0.98		10.50	10.52
Cyclic AMP	4.17		5.38	5.43
AMP	1.26	1.46	9.38	9.43
dGMP	1.22		10.88	10.88
IMP	1.15	1.24	10.80	10.82
GMP	1.11	1.20	11.71	11.73
Riboflavin	6.31	6.34	5.43	5.41
S-Adenosyl-L-methionine		1.18		12.83
CDP	0.85		12.22	12.24
ADP	1.05	1.11	10.77	10.78
dGDP	1.00		12.07	12.11
Folate	3.21	5.31	11.78	11.79
CTP	0.84		13.23	13.26
ATP	0.99	1.04	11.84	11.90
dGTP	0.88		13.08	13.12
GTP	0.86		13.90	13.91
Oxidized Glutathione	0.98	1.49	12.12	12.18
NAD	1.32		9.96	9.82
NADP	0.96		12.08	12.03
CoA	3.78		10.02	13.21
FAD	4.07	5.36	7.81	7.77
Acetyl-CoA	5.05		8.98	8.95

If retention time is not marked, the compound was not detected in the corresponding column and ionization mode. Retention times are marked in minutes.

APPENDIX 2: QUALITATIVE RESULTS OF THE QUENCHING SOLVENT EXPERIMENTS




Anthranilic acid
Trigonelline
Tyramine
Histidinol
α-Ketoglutarate
Glutamine
Lysine
Glutamate
Methionine
Guanine
Xanthine
2,3-Dihydroxybenzoic acid
Protocatechuate
Histidine
Orotic acid
Allantoin
Phenylpyruvate
7-Methylguanine
Phenylalanine
Phosphoenolpyruvate
Pyridoxamine
Pyridoxine
n-Capric acid
Aconitate
Shikimate
N-Acetylornithine
Arginine
Citrulline
Glucosamine
Hydroxyphenylpyruvate
myo-Inositol
Tyrosine
3-Phosphoglycerate
Citrate
Tryptophan
Pantothenic acid
Cystathionine
Deoxycytidine
Deoxyguanosine
Deoxyuridine
Thymidine
Cytidine
Uridine
Biotin
Deoxyadenosine
Deoxyinosine
Glucosamine-6-phosphate
Glucose-6-phosphate
Thiamine
Adenosine



The different solvents used: 1) Acetonitrile-water (80:20) 2) Acetonitrile-water (80:20) with 0.1 M formic acid 3) Acetonitrile-methanol (80:20) 4) 100% methanol 5) Chloroform-methanol (1:2).

Grey area means the compound was found in the sample prepared in the named condition, blank area means it was not detected.

The bolded metabolites were found universally in all prepared samples.