

TECHNISCHE UNIVERSITÄT MÜNCHEN

Lehrstuhl für Analytische Lebensmittelchemie

Host-Pathogen metabolomics of *Pseudomonas aeruginosa* infection models

Michael Witting

Vollständiger Abdruck der von der Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr. E. Grill

Prüfer der Dissertation:

1. Priv.-Doz. Dr. P. Schmitt-Kopplin
2. Univ.-Prof. Dr. M. Rychlik
3. Univ.-Prof. Dr. T. Rattei (Universität Wien / Österreich)

Die Dissertation wurde am 19.12.2012 bei der Technischen Universität eingereicht und durch die Fakultät Wissenschaftszentrum Weihenstephan für Ernährung, Landnutzung und Umwelt am 03.04.2013 angenommen.

„The whole is more than the sum of its parts“

Aristotle



To Fanny.

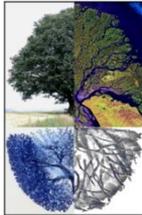
To my family.

To my friends and colleagues.



Acknowledgements

Research Unit Analytical BioGeoChemistry:



My special thanks go to Philippe Schmitt-Kopplin giving me the opportunity to work in his group on this interesting topic. I want to thank Jenny, Alesia and Juliano sharing the office with me and all the craziness in our room!!! Thank you Constanze showing me everything during my first month and all the discussion on host-pathogen interactions. Thanks to Alex for helping me with optimization of UHPLC-MS methods. Very special thanks go to Marianna, with all the help on statistical analysis of the data. And thank you to all the others, for the nice discussions, parties, BBQ's and the science we shared.

Department of Genome Oriented Bioinformatics



I have to thank Brigitte Wägele for her patience in teaching me the basics of programming and bioinformatics, Leonie Corry for taking care on all the bureaucracy and Thomas Rattei for all the support.

Laboratoire d'Ingénierie des Systèmes

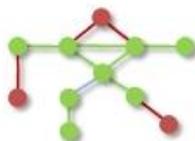
Macromoléculaires



From the CNRS I want to thank Romé Voulhoux and Thibault Sana on their input concerning *Pseudomonas aeruginosa*. Very special thanks go to Steve Garvis! Thank you for growing all that cells, worms and bacteria. Without you, this work would have been impossible!!!

This dissertation was prepared from 01.12.2009 to 30.11.2012 at Research Unit Analytical BioGeoChemistry at the Helmholtz Zentrum München, National Research Center for Environment and Health in the Helmholtz Association in Neuherberg.

The work was funded by the ERA-NET Pathomics project.



Pathomics



ERA-NET
PathoGenoMics



List of Abbreviations

ACN	Acetonitrile
APCI	Atmospheric pressure chemical ionization
API	Application programming interface
APPI	Atmospheric pressure photo ionization
BRENDA	Braunschweig Enzyme Database
CCUG	Culture Collection University of Göteborg
CE-MS	Capillary electrophoresis mass spectrometry
CF	Cystic fibrosis
CHCl ₃	Chloroform
Da	Dalton
DCC	Dicyclohexancarbondiimide
DI-MS	Direct infusion-MS
DNA	Deoxyribonucleic acid
EC number	Enzyme Commission number
EMSA	Electrophoretic mobility shift analysis
ESI	Electrospray ionization
FAD / FADH ₂	Flavin adenine dinucleotide
GC	Gas Chromatography
GCRMA	Gene chip robust multi array averaging
HILIC	Hydrophilic interaction liquid chromatography
HMDB	Human Metabolome Database
HPLC	High performance liquid chromatography
ICR-FT/MS	Ion Cyclotron Resonance Fourier Transform Mass Spectrometry
iPrOH	iso-Propanol
KEGG	Kyoto Encyclopedia of Genes and Genomes
LC	Liquid Chromatography
MALDI	Matrix assisted laser desorption ionization
MAS-NMR	Magic angle spinning-NMR
MCS	Maximum common substructure
MeOH	Methanol
MHz	Megahertz
MS	Mass Spectrometry
MTBE	Methyl-tert-butyl-ether
NADH / NAD ⁺	Nicotinamide adenine dinucleotide
NADPH / NADP ⁺	Nicotinamide adenine dinucleotide phosphate
NGM	Nematode growth medium
NGS	Next generation sequencing
NHS	N-Hydroxysuccinimide
NMR	Nuclear magnetic resonance
PATRIC	Pathosystems Resource Integration Center
PC	Phosphatidylcholine
PCA	Principal component analysis
PDA	Photo diode array
PE	Phosphatidylethanolamine
PLS	Partial least squares regression
ppm	Parts per million
PS	Phosphatidylserine
QQQ-MS	Triple quadrupole-MS
QS	Quorum sensing
QSRR	Quantitative structure retention relationship

<i>Q-ToF-MS</i>	<i>Quadrupole-Time of Flight-MS</i>
<i>RNA</i>	<i>Ribonucleic acid</i>
<i>RP</i>	<i>Reversed phase</i>
<i>SAGE</i>	<i>Serial analysis of gene expression</i>
<i>SPE</i>	<i>Solid phase extraction</i>
<i>T3SS</i>	<i>Type three secretion system</i>
<i>TCA</i>	<i>tricarboxylic acid</i>
<i>TG</i>	<i>Triacylglycerol</i>
<i>UHPLC-MS</i>	<i>Ultrahigh performance liquid chromatography – mass spectrometry</i>
<i>UHPLC-UHR-ToF-MS</i>	<i>Ultrahigh performance liquid chromatography Ultrahigh resolution Time of Flight Mass Spectrometry</i>





Summary

The gram negative bacterium *Pseudomonas aeruginosa* represents a major burden in public health. It causes a wide range of infections, from urogenital tract to blood stream infections. Broad metabolic capabilities, several antibiotic resistances and virulence factors complicate treatment of *P. aeruginosa* infection. Therefore novel diagnostic methods and drug targets for improved therapy are needed. The ERA-Net Pathomics project aims to find such novel diagnostic strategies and targets by combining different approaches. This work was part of metabolomics studies on different *P. aeruginosa* infection models.

Metabolomics, the global investigation on a organisms or systems metabolism, is from all “omics”-levels closest to the observed phenotype and allows to elucidate direct effects of infection on a hosts metabolism.

Different non-targeted metabolomics methods, including ICR-FT/MS and UHPLC-UHR-ToF-MS have been developed and applied two different infection models. On the one hand *Caenorhabditis elegans*, a convenient model for studying host-pathogen interactions was used. Within this infection model two different strains of this nematode, *daf-2* and *fer-15* have been challenged with fully virulent or attenuated *P. aeruginosa* and *Salmonella enterica*. Analysis of obtained data revealed specific metabolic response for each condition. Especially energy metabolism was heavily influenced by infection. On the contrary, HeLa cells infected with *P. aeruginosa* showed a specific profile in the cells lipid signature. Lastly, to study the influence of the bacterial virulence factor ExoY, cells expressing this protein were subjected to metabolome analysis. Similar to the *C. elegans* model a big influence on the central carbon metabolism was found.

Integration of data from all models showed that *P. aeruginosa* infection greatly influences the host’s metabolism. Results from the used non-targeted metabolomics platforms provided novel hypothesis, which have to be proven in more sophisticated setups. In conclusion, metabolomics holds great opportunities in infection research as diagnostic tool or for elucidation of drug targets.



Zusammenfassung

Das gram-negative Bakterium *Pseudomonas aeruginosa* repräsentiert ein großes Problem für die allgemeine Gesundheit und versucht eine breite Palette von Infektionen, von urogenitalen bis zu Infektionen der Blutbahn. Ein breites, metabolisches Potential, Resistenz zu vielen Antibiotika und verschiedenste Virulenzfaktoren erschweren eine erfolgreiche Behandlung. Deshalb werden neue Wege zur Diagnose und neue Ansatzpunkte für Medikamente benötigt. Das ERA-Net Pathomics Projekt zielt darauf ab diese neuen Strategien durch verschiedenste Herangehensweisen zu entwickeln. Diese Arbeit war Teil von metabolischen Studien an verschiedenen *P. aeruginosa* Infektionsmodellen.

Metabolomik, die globale Analyse des Stoffwechsels eines Organismus oder Systems, ist von allen "Omics"-Ebenen dem Phänotyp am nächsten und ermöglicht direkte Effekte einer Infektion auf den Metabolismus zu identifizieren.

Verschiedene nicht-zielgerichtete Metabolomanalysenmethoden, inklusive ICR-FT/MS und UHPLC-UHR-ToF-MS wurden entwickelt und zur Analyse von zwei verschiedenen Infektionsmodellen angewandt. Auf der einen Seite wurde *Caenorhabditis elegans*, ein praktischer Modelorganismus für die Studie von Wirt-Pathogen-Interaktionen, benutzt. Zwei verschiedene Nematodenlinien, *daf-2* und *fer-15* wurden entweder mit virulenten oder in der Virulenz abgeschwächten *P. aeruginosa* und *Salmonella enterica* infiziert. Analyse der erhaltenen Daten zeigte, dass für jeden Zustand ein spezifisches, metabolisches Muster vorhanden ist, wobei besonders der Energiestoffwechsel während der Infektion beeinträchtigt ist. Zum anderen zeigten mit *P. aeruginosa* infizierte HeLa-Zellen ein verändertes Lipidprofil. Zuletzt wurden HeLa-Zellen die den bakteriellen Virulenzfaktor ExoY exprimieren Metabolomeanalysen unterzogen, um den Effekt dieses Proteins isoliert zu studieren. Ähnlich den Ergebnissen des *C. elegans*-Modell war ein erheblicher Einfluss auf den zentralen Kohlenstoffwechsel zu sehen.

Integration aller erhaltenen Daten zeigte wie Infektion mit *P. aeruginosa* den Stoffwechsel des Wirts beeinträchtigt. Ergebnisse der verwendeten nicht-zielgerichteten Stoffwechselanalyse zeigten neue Hypothesen auf, welche mit spezialisierten Experimenten bewiesen werden müssen. Zusammenfassend lässt sich sagen, dass die Metabolomik ein hervorragend geeignetes Werkzeug für die Diagnostik und das Auffinden neuer Ansatzpunkte für Wirkstoffe darstellt.



List of publications

Ultrahigh Resolution Mass Spectrometry based Non-targeted Microbial Metabolomics

Michael Witting, Marianna Lucio, Dimitrios Tziotis, Philippe Schmitt-Kopplin

in Karsten Suhre "Genetics meets Metabolomics: From Experiment to Systems Biology", 2012, Springer

MasTRIX Reloaded: Combined Analysis and Visualization of Transcriptome and Metabolome data

Brigitte Wägele, Michael Witting, Philippe Schmitt-Kopplin, Karsten Suhre

PLoS ONE 7(7): e39860. doi:10.1371/journal.pone.0039860

A metabolomics approach to study bacterial virulence in C. elegans

Michael Witting, Marianna Lucio, Dimitrios Tziotis, Brigitte Wägele, Karsten Suhre, Romé

Voulhoux, Steve Garvis, Philippe Schmitt-Kopplin

PLoS ONE (in revision)

Poster and Oral presentations

Pathogen-host metabolomics and interactomics (Pathomics): WP3 Metabolomics (Poster)

Michael Witting, Constanze Müller, Brigitte Wägele, Agnes Fekete, Karsten Suhre, Thomas Rattei, Philippe Schmitt-Kopplin

at Metabolomics and More 2010, Freising, Germany

A mass spectrometry based metabolomic platform for the analysis of Caenorhabditis elegans (Poster)

Michael Witting, Alexander Ruf, Steve Garvis, Brigitte Wägele, Romé Voulhoux, Philippe Schmitt-Kopplin

at Anakon 2011, Zürich, Switzerland

A mass spectrometry based metabolomic platform for the analysis of Caenorhabditis elegans (Poster)

Michael Witting, Alexander Ruf, Steve Garvis, Brigitte Wägele, Romé Voulhoux, Philippe Schmitt-Kopplin

at Trend in Metabolomics 2011, Frankfurt am Main, Germany

Caenorhabditis elegans as model organism for host-pathogen interactions and probiotics research - A metabolomics approach (Poster)

Michael Witting, Alexander Ruf, Steve Garvis, Brigitte Wägele, Marianna Lucio, Saskia van



Hemert, Romé Voulhoux, Philippe Schmitt-Kopplin
at 3rd TNO Conference on Beneficial microbes 2012, Noordwijkerhout, The Netherlands

Non-targeted metabolomic analysis of Pseudomonas aeruginosa at the time points of T3 and T6 induction by ICR-FT-MS (Oral presentation)

Michael Witting, Romé Voulhoux, Steve Garvis, Agnes Fekete, Philippe Schmitt-Kopplin
at Pathomics Meeting 2009, Valencia, Spain

Metabolomics – An introduction (Oral presentation)

Michael Witting, Constanze Hermann
at Pathomics PhD Retreat 2010, Marseille, France

Metabolomic Analysis of Pseudomonas aeruginosa infection (Oral presentation)

Michael Witting, Romé Voulhoux, Steve Garvis, Brigitte Wägele, Philippe Schmitt-Kopplin
at Pathomics Meeting 2010, Paris, France

Quo Vadis Metabolomics? – Metabolomic and Systems Biology (Oral presentation)

Michael Witting
at Pathomics PhD Retreat 2011, Düsseldorf, Germany

Host-Pathogen metabolomics of Pseudomonas aeruginosa (Oral presentation)

Michael Witting, Romé Voulhoux, Steve Garvis, Brigitte Wägele, Philippe Schmitt-Kopplin
at Pathomics Meeting 2011, Vienna, Austria

Host-Pathogen interaction of Caenorhabditis elegans and Pseudomonas aeruginosa - A non-targeted metabolomics approach (Oral presentation)

Michael Witting, Steve Garvis, Alexander Ruf, Marianna Lucio, Dimitrios Tziotis, Silke Heinzmann, Norbert Hertkorn, Romé Voulhoux, Philippe Schmitt-Kopplin
at Bruker Daltonic Anwendertreffen 2012, Kassel, Germany

Caenorhabditis elegans infection model – A metabolomics view (Oral presentation)

Michael Witting, Steve Garvis, Brigitte Wägele, Karsten Suhre, Romé Voulhoux, Philippe Schmitt-Kopplin
at Pathomics Meeting 2012, Lübeck, Germany



Content

1.	Introduction.....	1
	Abstract	1
1.1	Metabolomics - Integrated studies of metabolism	2
1.1.1	Life, Biochemistry, Metabolism and Metabolomics	2
1.1.2	Metabolomics and other “omics”-science	5
1.1.3	Non-targeted and targeted metabolomics.....	8
1.1.4	Lipidomics	8
1.2	Analytical approaches – Towards ultrahigh resolution metabolomics and lipidomics	10
1.2.1	Ion cyclotron resonance – Fourier transform mass spectrometry (ICR-FT/MS).....	10
1.2.2	Hyphenated ultrahigh resolution Time of Flight Mass Spectrometry	13
1.2.3	Nuclear magnetic resonance (NMR).....	15
1.2.4	Comparison of the single methods.....	17
1.3	<i>Pseudomonas aeruginosa</i> – An opportunistic human pathogen	18
1.3.1	General overview on <i>Pseudomonas aeruginosa</i>	18
1.3.2	<i>Pseudomonas aeruginosa</i> virulence factors	20
1.3.3	Metabolomics studies on <i>Pseudomonas</i>	21
1.4	Model Systems for Host-Pathogen interactions	22
1.4.1	HeLa and other mammalian cell culture models.....	22
1.4.2	Metabolomics research in mammalian cell cultures.....	23
1.4.3	<i>Caenorhabditis elegans</i>	23
1.4.4	Metabolomics research on <i>C. elegans</i>	26
1.5	Pathomics project aims and objectives.....	28
1.6	Thesis structure	30
2.	Data analysis tools for Host-Pathogen metabolomics	33
	Abstract	33
2.1	Bio- and Chemoinformatics to assist metabolomics research	34
2.1.1	Review on chemo and bioinformatics research in metabolomics.....	34

2.2	MassTRIX: A web-based metabolite annotation tool	36
2.2.1	Annotation of mass spectrometric data.....	36
2.2.2	Analysis of transcriptomic data	37
2.2.3	Visualization of metabolomics and transcriptomic data on KEGG pathways	37
2.2.4	Comparison of different jobs.....	38
2.2.5	Comparison against other existing resources	40
2.3	Metabolic reconstructions as basis for host-pathogen metabolomics.....	41
2.3.1	Metabolic reconstruction of <i>Pseudomonas</i> species.....	42
2.3.2	Metabolic reconstruction of <i>Caenorhabditis elegans</i>	45
2.3.3	Metabolic reconstruction of HeLa cells	46
2.3.4	Improvements of reconstructions.....	47
2.3.5	Use of reconstructions in host-pathogen metabolomics.....	47
3.	Taxonomy based metabolomics of <i>Pseudomonas</i> species	51
	Abstract	51
3.1	Phylogenetics, bacterial identification and metabolomics	52
3.2	Cultivation and extraction of bacteria	54
3.2.1	Bacterial cultivation and extraction	54
3.3	Results	55
3.3.1	Statistical analysis.....	55
3.3.2	Bacterial strains compared on a metabolomic level	55
3.3.3	Different growth temperatures influence the metabolome	57
3.3.4	Core and pan metabolome of the genus <i>Pseudomonas</i>	58
3.3.5	Comparison against <i>in silico</i> metabolomes	62
4.	<i>C. elegans</i> – <i>P. aeruginosa</i> infection model	65
	Abstract	65
4.1	<i>C. elegans</i> infection model.....	66
4.2	Known and unknowns – False positive filtering of metabolite annotations	67
4.2.1	Prediction of chemical properties to build retention time models.....	70



4.2.2	Identification of fragments, adducts and isotopes	78
4.2.3	Lipid identification with AutoMS ⁿ and <i>in silico</i> fragmentation	79
4.2.4	Mass difference networking for improved metabolite annotation	84
4.2.5	Alignment of ICR-FT/MS and UHPLC-UHR-ToF-MS data	86
4.3	Biological interpretation of the <i>C. elegans</i> infection model	88
4.3.1	Statistical data analysis	88
4.3.2	Survival of <i>C. elegans</i> challenged with different pathogens	89
4.3.3	ICR-FT/MS metabolic profiling separates different sample conditions	90
4.3.4	Infection with <i>P. aeruginosa</i> leads to increased β -oxidation	95
4.3.5	Synthesis of phospholipid species during <i>Pseudomonas</i> infection	98
4.3.6	Metabolic markers of increased autophagy	100
4.3.7	Resistance of <i>daf-2</i> worms to infection	100
4.3.8	Potential infection biomarker	102
4.3.9	Metabolic signatures in long-lived <i>C. elegans</i>	105
4.3.10	Role of <i>P. aeruginosa pec7</i> during infection	106
4.3.11	Probiotic bacteria protects <i>C. elegans</i> from <i>P. aeruginosa</i> infection	107
5.	Development of extraction methods for non-targeted metabolomics of the HeLa- <i>P. aeruginosa</i> host-pathogen model	111
	Abstract	111
5.1	HeLa cells as host organism for <i>P. aeruginosa</i> infection	112
5.2	Sampling and sample preparation	112
5.2.1	Prerequisite for metabolomics samples	112
5.2.2	Quenching of metabolism	113
5.2.3	Cell lysis techniques and metabolite extraction	113
5.2.4	Solid phase extraction and sample concentration	114
5.3	Development of an extraction method for a HeLa cell infection model	114
5.3.1	Metabolite extraction from bacterial pellets	114
5.3.2	Metabolite extraction from cell culture medium containing HeLa cells	114
5.3.3	Results	115



5.3.4	Metabolite extraction from <i>P. aeruginosa</i> pellets	115
5.3.5	Metabolite extraction from HeLa cells	116
5.3.6	Final methods	117
5.4	Infection with <i>P. aeruginosa</i> affects the HeLa cell lipidome.....	118
5.5	ExoY – A potent cellular toxin	120
5.6	Effects of ExoY on the HeLa cell metabolome	121
6.	Conclusion.....	127
	Abstract	127
6.1	Metabolomics – a fast growing research area	128
6.2	Improved annotation and data analysis methods	128
6.3	Ultrahigh resolution analytical methods.....	129
6.4	Metabolomics based phylogeny of <i>Pseudomonas</i> species	130
6.5	<i>C. elegans</i> as convenient model for host-pathogen research	132
6.6	HeLa cell model	137
6.7	Comparison of the infection models.....	137
	Literature.....	141
	List of figures	156
	List of tables.....	159
	List of equations	160
	Appendix	161
	Used instruments	162
	ICR-FT/MS	162
	UHPLC-UHR-ToF-MS.....	162
	GC-FID	162
	Other instruments.....	162
	Used chemicals	163
	Solvents	163
	Buffer salts	163



Chemicals for synthesis	163
Standards	163
Columns	166
LC Columns	166
GC Columns	166
Methods	167
Cultivation and extraction of <i>C. elegans</i>	167
HeLa infection protocol	167
Lipid extraction from HeLa cells	168
Cultivation and extraction of HeLa cells expressing ExoY	168
Metabolic profiling using DI-ICR-FT/MS	168
Data processing DI-ICR-FT/MS	169
Metabolic profiling using UHPLC-UHR-ToF-MS	169
Lipid profiling using UHPLC-UHR-ToF-MS (<i>C. elegans</i>)	170
Lipid profiling using UHPLC-UHR-ToF-MS (HeLa)	170
Data processing UHPLC-UHR-ToF-MS	171
Fatty acid profiling	172
Data Analysis scripts	173
ICR-FT/MS spectra export	173
Export of MS/MS spectra generate by AutoMS ⁿ	173
Perl scripts	174
ICR-FT/MS and LC-MS alignment	174
Conversion of ASCII files to MetFrag Batchfile	175
Conversion of ASCII files to library	176
Alignment of LC-MS and LC-MS/MS features from AutoMS ⁿ	177
Curriculum vitae	178







1.1 Metabolomics - Integrated studies of metabolism

1.1.1 Life, Biochemistry, Metabolism and Metabolomics

Five attributes define life: motility, growth, testiness, reproduction and metabolism, whereas the last mentioned is the most important because it supplies energy and substances needed for all other attributes. Studying metabolism was the first objective in biochemistry and is a major part of it since decades, analyzing single enzymatic catalyzed reactions or substance classes. Through the last ten years paradigm changed to a systematic investigation of metabolism, not least analytical technologies became more powerful and “classical” biochemistry therefore was rediscovered [1]. Scientists realized that genes, transcripts or proteins cannot tell the whole story and metabolism plays a bigger role than just supplying energy and chemical building blocks. Figure 1 illustrates the central role of metabolites in different biological processes.

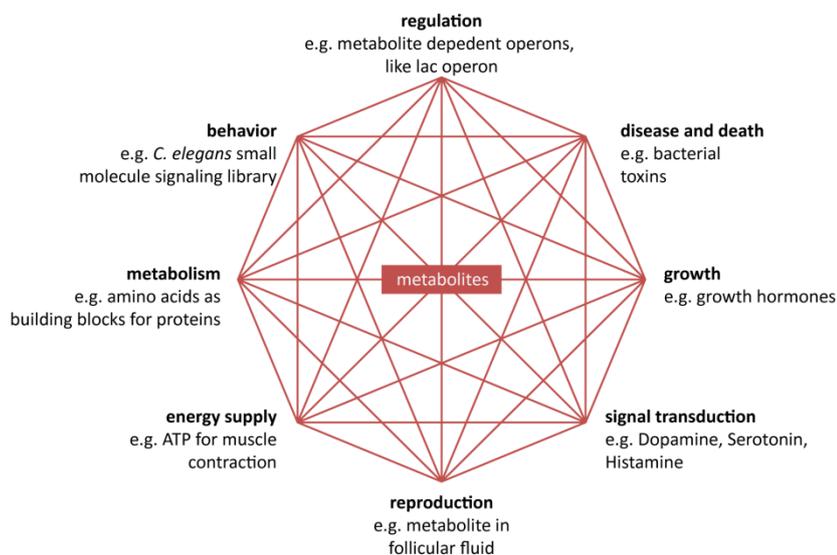


Figure 1: Metabolites play a central role in all biological processes.

Adapted with permission from PD Wolfgang Eisenreich, *Faszination Forschung* 8/11. Metabolites play a central role in life. They supply energy and chemical building blocks for all others attributes that define life. Furthermore they play essential role in different regulatory motif and in signal transduction.

The word metabolism derives from the Greek word “Μεταβολισμός”, which means change. Santorio Santorio was the first scientist doing systematic research on metabolism beginning in the year 1611. He weighed himself, the meals he ate, his urine and feces over a period of thirty years. He showed that the weight of his feces was smaller than what he ate, suggesting the theory of *insensible perspiration* for compensating this difference [2]. Louis Pasteur studied the fermentation of sugar to alcohol by yeast in the 19th century. He concluded that the conversion was performed by substances within the yeast cell he called “ferments”. Together with the publication of the synthesis of urea by Wöhler in 1824 it was shown that life is underlying the

same principles known from chemistry [3]. Another pioneer of metabolism research was Hans Krebs discovering the tricarboxylic acid cycle (TCA cycle), which is also named after him (Krebs cycle) [4]. A further breakthrough was the discovery of the regulation of the cholesterol and fatty acids metabolism, by Feodor Lynen [5]. Since this time several new metabolic pathways have been discovered in different species. One of the recent discoveries was the reverse TCA cycle found in bacteria acting as reductive CO_2 fixation [6].

Generally metabolism can be divided into two branches, anabolic, the up building, and catabolic, the degrading metabolism. Anabolic reactions deliver chemical building blocks for biological macromolecules like DNA, RNA and proteins and other molecules need for cellular functions like lipids or co-enzymes. Catabolic metabolism breaks down bigger molecules to obtain energy and precursors for anabolic metabolism and detoxifies xenobiotic substances [7]. Metabolites can be divided in several chemical classes, like sugars, small organic acids, amino acids, fatty acids, lipids, nucleotides and many more. Chemical reactions, catalyzed by enzymes, converting metabolites into each other are grouped in biochemical pathways, which again form a dense network. Examples for central pathways of metabolism in most organisms are glycolysis, the TCA cycle and the pentose phosphate pathway. All together supply energy and precursors for several amino acids and nucleotides. Furthermore the TCA cycle delivers redox equivalents in form of NADH and FADH_2 needed for the synthesis of adenosine triphosphate (ATP), the universal energy currency of cells. The pentose phosphate pathway delivers NADPH needed for biosynthesis of molecules. Excessive energy is stored in form of fats, which can be degraded if needed and fed into the TCA cycle. Several more pathways e.g. for the synthesis of purines and pyrimidine's, amino acids, lipids, hormones and many more metabolites are known. Major metabolic pathway are shared over a broad range of organisms, but each organism or organism group poses specific biosynthetic routes for secondary metabolites, which fulfill specific functions adapted to the ecological niche an organism lives in. Furthermore different species have different capabilities in degrading xenobiotics or pollutants.

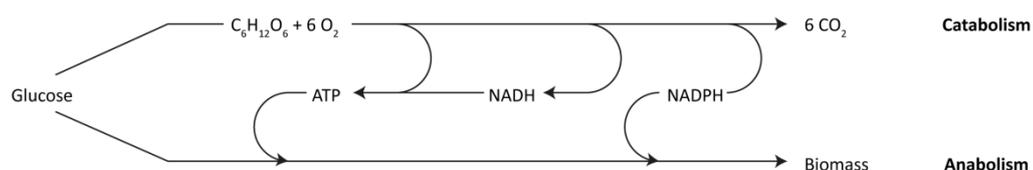


Figure 2: Catabolic and anabolic metabolism is interwoven through ATP and redox factors, like NADH or NADPH.
Adapted with permission from [8].

Anabolic and catabolic metabolisms are tightly coupled through availability of energy and redox factors, as shown in Figure 2. The cell as smallest way of life tightly regulates internal conditions through feedback loops and other regulatory motifs to stabilize health and function regardless of external conditions. This phenomenon is known as cellular homeostasis. This makes the analysis of metabolism an interesting field, as the cell is reordering metabolic fluxes as reaction to an external stimulus like infection or depletion of nutrients. Analysis of metabolism with analytical chemistry techniques is nowadays called metabolomics, growing fast, compromising to close the gap between genotype and phenotype [9, 10].

Fluxes in the metabolism are directed using regulatory enzymes. These enzymes are mostly catalyzing a rate-limiting or committed step, like early steps in a pathway or steps where energy is required. Regulatory mechanisms in metabolism can be divided into three categories: Non-covalent interactions, covalent interactions and changes in the abundance of an enzyme

Non-covalent interactions include availability of substrates and allosteric activation or inhibition. Allosteric acting molecules are either end products of a pathway (feedback inhibition), substrates of a pathway (feed-forward activation) or indicators of the cells energy status (e.g. ATP). Also protein-protein interactions are part of this category. Examples for covalent interactions are phosphorylation and dephosphorylation regulating enzyme activity. This signal is often transduced using cascades of enzymes. In the case of changing abundance gene expression of a specific enzyme is altered according to an external stimulus through a receptor or a cellular signal. Enzymes can be divided into constitutively expressed or inducible enzymes. Constitutive enzymes normally have housekeeping functions like catalyzing steps in the glycolysis pathway [7].

These regulatory events cannot be deduced from the genome or the transcriptome. Methods closer to the phenotype of a cell or organism are needed. Proteomics for examples is able to identify post-translational modified proteins, whereas metabolomics is capable of measuring metabolite levels and metabolic fluxes.

A reasonable question is why the analysis of small molecules got its warranty beside other “omics” technologies. The following points may answer this question [11].

- Many clinical assays target small molecules, from glucose in diabetes to triacylglycerol's or cholesterol
- Most of all known drugs are small molecules and a big proportion of them derive from preexisting metabolites.
- A big proportion of the identified genetic disorders involve diseases of metabolism.



- Metabolites serve as important cofactors, ligands and signaling molecules for thousands of proteins.

Although metabolomics may be still in its infancy it is rapidly catching up to the other “omics” sciences.

1.1.2 Metabolomics and other “omics”-science

Metabolomics is defined as the systematic study of metabolites in a biological system. It is the youngest approach in the “omics” family, including genomics, transcriptomics and proteomics. The metabolome is defined as the sum of all metabolites given in a biological system under particular physiological conditions. It can be divided into the exometabolome (metabolites outside the cell) and the endometabolome (intracellular metabolites). The term metabonomics was defined by Jeremy Nicholson in 1999 as "the quantitative measurement of the dynamic multiparametric metabolic response of living systems to pathophysiological stimuli or genetic modification" [12]. First measurements of metabolites in body fluids have already been carried out earlier without calling it metabolomics, e.g. the pioneering work of Linus Pauling [13]. In contrast to transcripts and proteins, metabolites are not encoded in the genome and are highly dependent on the surrounding environment, making it complicated to give definite numbers of metabolites present in an organism.

Metabolomics is applied in different research areas, from fundamental research to medical science. Metabolites are measured on routine basis from different origins like different body fluids, bacterial or cell cultures, tissues, microbiomes or even ocean water. Urine and plasma are the most employed body fluids in medical metabolomics and are used for diabetes or nutrition research, for example. Toxicological studies, biomarker and target discovery, clinical trials and studies are other topics of interest. Beside these applied areas, research in plant and bacterial physiology are occupying a big part of metabolomics science.

In the classical biological view information flows from genes via transcripts to proteins which are finally carrying out cellular functions, like metabolism. By this metabolomics is closest of all levels to an observed phenotype [9]. Redirection of metabolic fluxes takes normally seconds, while the expression of a protein from a gene ranges from several minutes to hours. Compared to the genome the metabolome is relatively small, ranging from several hundred different metabolites in simple bacteria to ten thousands in plants (Figure 3). By far not all metabolites and metabolic pathways are known yet, as functions of several genes are unknown and possible secondary functions of enzymes exist. Metabolomics gains more and more attention as a tool able to supply biomarkers for improved diagnostics and deeper

insight in pathophysiological states. Furthermore integration of all “omics”-sciences to create a model of a living system is the ultimate goal of system biology. This joining of data provides insight in transcriptional regulation, regulation on protein and metabolite level. Lastly, biology is dynamic and doesn't separate between this man-made levels.

Table 1 gives a short overview on common definition of “omics” sciences and related objects. Currently, several sub groups out of these disciplines are evolving, like lipidomics, which is discussed later or glycomics, peptidomics and others. Normally these refer to the analysis of a certain metabolite subgroup like sugars or peptides.

Table 1: Definition of "omics"-sciences and related objects

Adapted with permission from [14].

Term	Definition
<i>Genome</i>	<i>Whole set of genes of an organism.</i>
<i>Transcriptome</i>	<i>Set of all RNA molecules of a cell, organ or organism, including mRNA, tRNA, rRNA and non-coding RNA.</i>
<i>Proteome</i>	<i>Set of all proteins expressed in a cell, organ or organism under specific conditions at a given time.</i>
<i>Metabolite</i>	<i>Small molecules (low-molecular-weight (<1500)) that participate in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell</i>
<i>Metabolome</i>	<i>The total sums of metabolites of a given biological system under particular physiological conditions. The metabolome is divided into exometabolome (metabolites outside the cell) and endometabolome (intracellular metabolites)</i>
<i>Metabonome</i>	<i>The sums, products and interactions of all the individual compartments/metabolomes (including extra-genomic sources) dispersed in a complex organism; the “Global” System. (J. Nicholson)</i>
<i>Metabolite</i>	<i>Small molecules (low-molecular-weight (<1500)) that participate in general metabolic reactions and that are required for the maintenance, growth and normal function of a cell</i>
<i>Metabolomics</i>	<i>Identification and quantification of all metabolites in specified cellular, biofluid or tissue section.</i>
<i>Metabonomics</i>	<i>The quantitative measurement of the time related multi-parametric metabolic response of living system to pathophysiological stimuli or genetic modification [12]. It evaluates tissue and biological fluids for changes in endogenous metabolite levels effects of a disease or a therapeutic treatment.</i>
<i>Metabolic profiling</i>	<i>Quantitative analysis of set of metabolites or derivative products (identify or unknown) in a selected biochemical pathway or specific class of compounds. This includes target analysis, the analysis of a very limited number of metabolites, e.g. single analytes as precursors or products of specific biochemical reactions.</i>
<i>Metabolic fingerprinting</i>	<i>Unbiased, global screening approach to classify samples based on metabolite patterns or “fingerprints” that change in response to disease, environmental or genetic perturbations with the ultimate goal to identify discriminating metabolites.</i>
<i>Metabolic footprinting</i>	<i>Called also exometabolomics, it is the observation of what a cell or system excretes under controlled conditions [15]</i>

Against the classical view, nowadays interaction between all layers is thought to be tighter. Metabolites are involved in several regulatory motifs, not only in metabolism. For example polyunsaturated fatty acids play important roles in immune response of mammals and other animals [16, 17]. The goal of systems biology to understand the network of all molecules that make up life is today reachable. Next generation sequencing allows the sequencing of complete genomes in days. Transcriptomes can be determined using expression arrays, RT-qPCR and also next generation sequencing of transcripts in quantitative manner. Today, the major bottleneck in genomics is the annotation of genes, which is mostly based on homology comparison with known genes. With this approach only 40-60% of a genome can be annotated, because for a big number of genes function is unknown or no homologues exist. Metabolomics holds the opportunity to assign possible functions to orphan genes. One way to assign functions to unknown genes is the generation of knock-out mutants and the measurement of their respective metabolome. This can be repeated for all genes an organism carries and the data can be integrated in a holistic gene-metabolite correlation network, which allow the creation of hypothesis about a gene function [18].

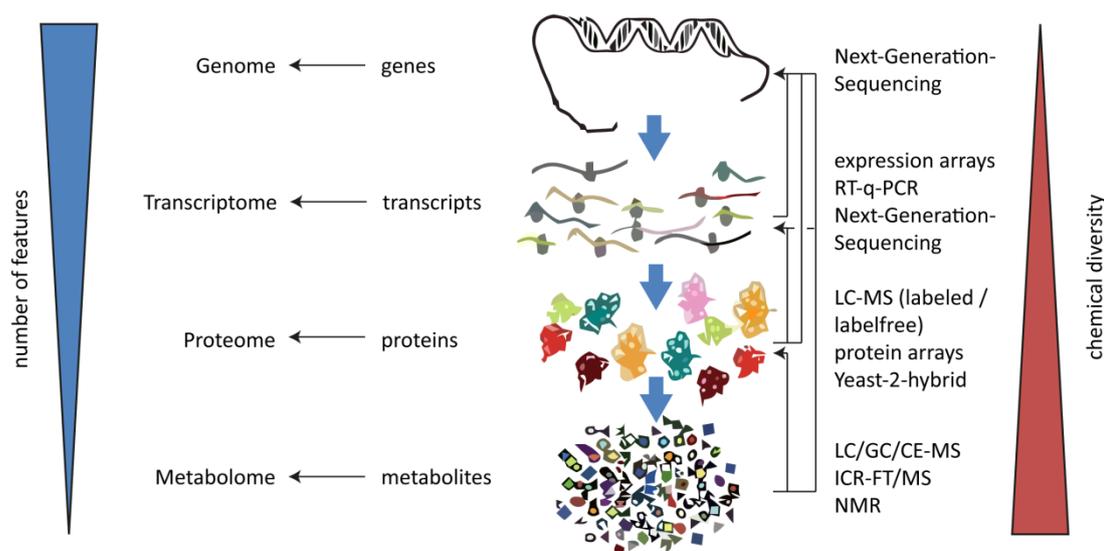


Figure 3: Genes, transcripts, proteins and metabolites are acting together in a tight network to build up life.

Adapted with permission from [19]. Information flows from genes via transcripts and proteins to metabolites, whereas several feedback loops are possible. While the number of genes, transcripts or proteins is bigger than the number of metabolites, these are more chemically diverse than the others. The right column illustrates different methods used today for measurement of the different “omics”-levels.

The same is true for the metabolites themselves. From metabolomics experiments maximum 30% of the signals can be annotated to known compounds. Some of the remaining 70% may consist of other adducts, neutral losses, homo- or heterodimers or fragments. Still, taking all this possibilities into consideration, a big proportion remains unknown. At this point

ultra-high resolution analytics come to the fore, allowing more precise separation, detection and interpretation of this unknowns.

1.1.3 Non-targeted and targeted metabolomics

The terms metabolomics and metabonomics are often used synonymous in literature. To avoid any confusion the expressions non-targeted and targeted metabolomics are widely used. Non-targeted metabolomics refers to hypothesis-free elucidation of metabolism. The aim in non-targeted metabolomics is to detect as much metabolic features from different metabolite classes as possible. The major goal is to identify metabolic alterations and pattern that correlate with a certain physiological state. Non-targeted metabolomics approaches are data and work intensive, mostly generating hypothesis during analysis, which has to be proven afterwards. The counterpart is targeted metabolomics, which is analyzing and quantifying only a subset of metabolites, e.g. compounds belonging to energy metabolism. Several analytical techniques are used in both approaches. Non-targeted metabolomics is mostly based on high or ultrahigh-resolution mass spectrometric methods, like Q-ToF-MS or ICR-FT/MS, or NMR, whereas targeted metabolomics prefers very sensitive and fast techniques, like triple quadrupole MS. The border between both is blurring.

1.1.4 Lipidomics

Lipids fulfill several cellular functions, from storage of energy, building blocks of membranes to signaling molecules. Phospholipids serve as integral parts of the membrane for separation and compartmentalization of cells; moreover these membranes serve as scaffolds for interactions along membrane-associated moieties. Also for host-pathogen interactions membranes play crucial roles, because they are the first line of defense against pathogens. Several lipids act as second messengers of signal transduction for example lysolipids, phosphoinositides and steroids. Phosphatidylcholines represent the major species of lipids in eukaryotes and they are building the lipid bilayer of membranes. An overview on different lipid classes can be found in Figure 4.

The complementary method to metabolomics measuring lipids of a system is called lipidomics. This is a challenging task, because virtually thousand to ten-thousand different lipids exist. Phospholipids for example are built out of different building blocks, like different head groups and length of fatty acids. Additionally the first side chain can be linked through an ester, enol ether or ether bond. Number and position of double bonds further boosts the chemical space of lipids. For a systematic classification of lipids the LIPIDMAPS classification system is the most used system. It divides lipids in the major classes fatty acyls (FA), glycerolipids (GL), glycerophospholipids (GP), sphingolipids (SP), sterol lipids (ST), prenol lipids



(PR), saccharolipids (SL) and polyketides (PK) with several subclasses. A comprehensive database has been accumulated by the LIPIDMAPS consortium and is accessible via the web (www.lipidmaps.org) [20]. Several lipids have similar physicochemical properties, which complicates their analysis. Nowadays two different type's lipid analyses are performed: lipid profiling or shotgun lipidomics. The later uses the raw lipid extract directly infused into MS and data dependent acquisition for precise quantification of lipid species. Normally this type of analysis is carried out on ion trap MS using MSⁿ. For selection of lipid species, neutral losses corresponding to the different head groups are used and the resulting fragment is further fragmented to obtain fatty acid chain composition. The major drawback is that this method cannot separate isomeric species. Lipid profiling uses chromatographic separation of lipid species followed by mass spectrometric detection and allows differentiation of isomeric lipid species. Several chromatographic methods exist, for example HILIC or normal phase chromatography for separation of different lipid classes or reversed phase separation for detection of single species. A common lipid profiling method uses C8 reversed phase columns and ACN/iPrOH gradient [21]. This method allows detection of phospho- and glycerolipids in one single run.

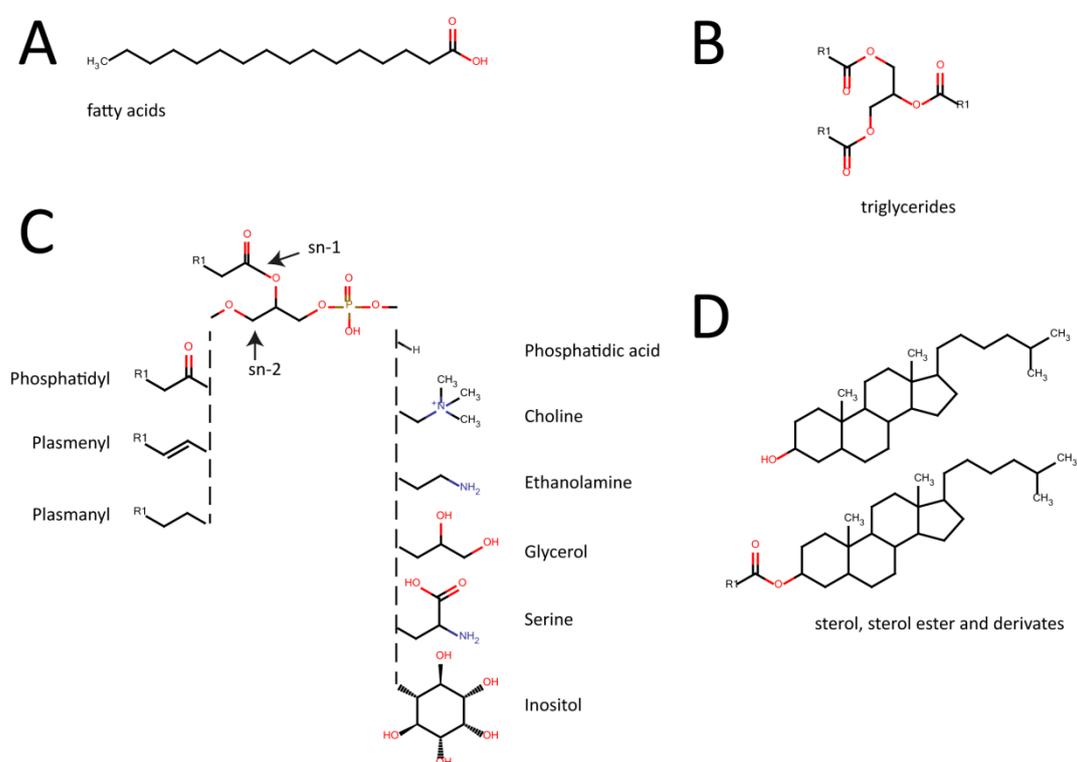


Figure 4: Examples for different lipid classes found in biological samples

(A) Fatty acids are the simplest lipids and serve as building blocks for phospholipids and glycerolipids. (B) Triacylglycerol's are highly effective energy storage molecules. (C) Different classes of phospholipid exist. They are separated by their different head groups attached to the diacylglycerol backbone. The side chain at sn1 can be bound by and ester enol-ether or ether bond. (D) Sterols, sterol ester and their derivatives present a further highly unipolar lipid group

Sterols, sterol ester and derivatives of them have special roles and are used as hormones and integral membrane constituents. These molecules are highly non-polar and hard to ionize in ESI. Special ionization methods like APCI or APPI are needed to analyze them. Steroidomics is a subcategory of lipidomics focusing on the analysis of steroid molecules. The term was created by William Griffiths [22].

1.2 Analytical approaches – Towards ultrahigh resolution metabolomics and lipidomics

To explore the metabolome or lipidome of a given organism several analytical techniques are needed. Today no method is able to cover the full metabolome. Using ultrahigh resolution analytical approaches can help to minimize the effort for metabolite identification. Current methods used in metabolomics are shortly reviewed in the following paragraphs.

1.2.1 Ion cyclotron resonance – Fourier transform mass spectrometry (ICR-FT/MS)

Ion Cyclotron Resonance Fourier Transform Mass Spectrometry (ICR-FT/MS), having a resolution up to 1,000,000 and more and a mass accuracy <0.1 ppm, is used as a profiling technique in metabolomics. It allows annotation of potential metabolites and calculation of possible elemental formulas using exact mass information. The first time ICR-FT/MS was described for non-targeted metabolomics was published by Asaph Aharoni in 2002. His group investigated the changes in ripening strawberries. The high resolution made it possible to see different metabolites between green and red strawberries even in a window of 0.1 Da [23]. Since this time a whole bunch of papers describes the use of ICR-FT/MS for metabolomics studies, which are reviewed elsewhere. Automation of data acquisition is easily accessible using autosamplers for flow injection or automated robotic devices, like the Advion TriVersa Nanomate, for sample infusion [24-26].

ICR-FT/MS is a mass analyzer based on cyclotron motion of ions in a homogeneous high magnetic field [27]. After ionization of the analytes by electrospray ionization (ESI), atmospheric pressure chemical ionization (APCI), atmospheric pressure photo ionization (APPI) or matrix-assisted laser desorption ionization (MALDI) as either positively or negatively charged ions they are focused by ion lenses transferred into the magnetic field of a superconducting magnet, where an oscillating electric field excites the ions to higher trajectories. The masses are then resolved by their different cyclotron (rotational) frequency of the ion rotation in the magnetic field. If a moving molecule with a mass m and an electric charge q ($q=n \cdot e$) is transferred into a magnetic field B which is orthogonal to the ion's velocity v , the Lorentz force F_L acts on the ion.



$$F_L = q \cdot v \cdot B$$

Equation 1: Lorentz force

The equation describes the Lorentz force that acts on point charge q that passes a magnetic field B with the velocity v . F_L = Lorentz force. q = charge. v = velocity. B = magnetic field.

In the homogenous magnetic field, the moving charge has a constant velocity and moves on a stable circular trajectory with the radius r . Their by the centrifugal force equilibrates the magnetic force.

$$\frac{m \cdot v^2}{r} = q \cdot v \cdot B \quad \text{or} \quad \frac{m}{q} = \frac{B \cdot r}{v}$$

Equation 2: Equilibrium between centrifugal and Lorentz force

Against the Lorentz force works the centrifugal force, which rule out each other. m = mass. v = velocity. r = radius of trajectory movement. q = charge. B = magnetic field.

The angular velocity ω_c , also called as cyclotron frequency is only a function of m/z -ratio and the magnetic field strength B . Because of the right-hand rule positive and negative ions have contrary courses.

$$\omega_c = \frac{q \cdot B}{m} \quad f = \frac{\omega}{2 \cdot \pi}$$

Equation 3: Angular velocity

The angular velocity ω_c is only a function of the mass-to-charge ration m/z and the magnetic field strength. ω_c = angular velocity. q = charge. B = magnetic field strength. m = mass.

An additional alternating electric field orthogonal to the magnetic field causes the cyclotron resonance of a certain m/z -ratio. This field is applied between a pair of plates, called excitation plates. If the electromagnetic wave has the same frequency as a certain ion in the cyclotron cell, the resonance (absorption of energy) as consequence increases the ion's kinetic energy hence increase the radius of its trajectory. The excited oscillating ions are detected with a second pair of plates, rotated 90° to the excitation plates. The passing ions induce an alternating current between the detection plates. This so called image current is a superimposition of several frequencies caused by several ions of different masses. Fast Fourier transformation is used to convert it to a mass spectrum.

ICR-FT/MS can reach high mass resolutions R , sometimes also referred as resolving power.

$$R = \frac{M}{\Delta M}$$

Equation 4: Definition of mass resolution

R = resolution. M = mass. ΔM = peak width at 50% of the peak height.

M is the mass of an ion and ΔM is typically the peak width at 50% of the peak height, also called full width at half maximum (FWHM). But not only resolving power is important, mass

accuracy plays a crucial role for identification of metabolites. The mass error is often reported in ppm, meaning error in part per million.

$$\Delta m/z = \frac{\text{measured mass} - \text{true mass}}{\text{true mass}} \cdot 10^6 \text{ ppm}$$

Equation 5: Definition of ppm error

The ppm error is a relative mass error used in mass spectrometry.

Working with high-resolution instruments it is needed to take also the mass of a electron ($5.485799 \cdot 10^{-4}$ u) into account. At 200 Da, 400 Da and 800 Da, the mass of an electron would yield errors of 2.74, 1.37 and 0.69 ppm, respectively.

The determined exact masses can be used to calculate possible elemental formulas. Chemical formulas can be understood as linear combination of elements with distinct monoisotopic masses, following several chemical rules [28]. Using these rules, it is possible to calculate formulas out of exact masses obtained from ICR-FT/MS. Due to the high resolving power only a few different formulas fit the exact mass, which narrows down the list of potential candidates. Additionally isotopic information can be used to confirm the predicted formulas.

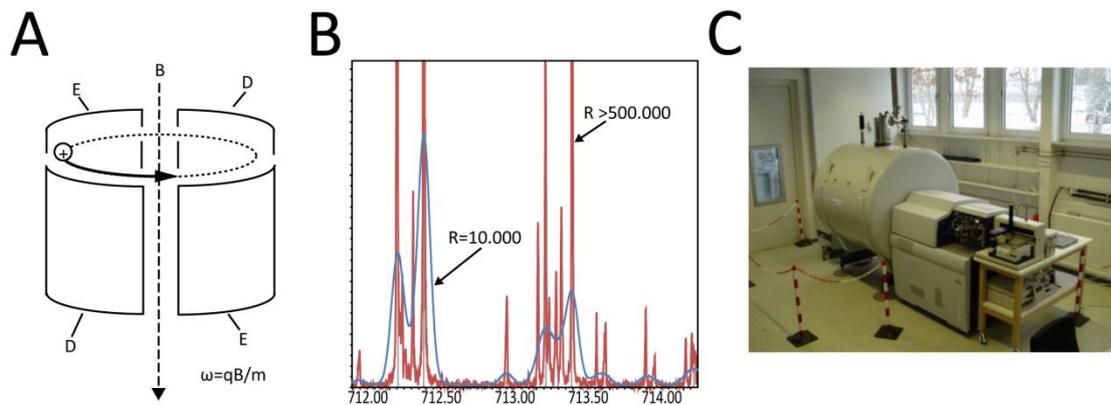


Figure 5: Principles of ICR-FT/MS

(A) Principle of ICR-FT/MS as explained in the text. E = excitation plates, D = detection plates. B = magnetic field (B) Influence of resolution on measurement accuracy. In red, a snapshot of a mass spectrum of a bacterial extract from *Pseudomonas aeruginosa* PA14 measured in positive ionization mode with a resolution of $>100,000$ is shown. In blue, the same spectrum down scaled to a resolution of 10,000. (C) Typical ICR-FT/MS instrument.

To take advantage of the high resolving power and mass accuracy of an ICR-FT/MS in metabolomics, it is mostly utilized in direct infusion (DI) mode without prior metabolite separation using chromatographic or electrophoretic methods. This technique has some drawbacks, as isomeric, like hexoses for example, and isobaric substances can't be separated. A second disadvantage is ion suppression often observed in DI-MS. Despite this, DI-ICR-FT/MS is a very robust and sensitive technique, being high-throughput capable and sample

consumption is minimal. To overcome the effects of ion suppression and to separate isomeric substances the ICR-FT/MS can be coupled at-line to a chromatographic separation. The crude metabolite extract is separated with a suitable chromatographic method and fractions are collected over the whole chromatographic run, which are in second step conducted to DI-ICR-FT/MS [29].

1.2.2 Hyphenated ultrahigh resolution Time of Flight Mass Spectrometry

Time of Flight (ToF) mass spectrometers provide high mass resolution and are fast enough for online coupling of different separation techniques. The newest generation of ToF instruments delivers resolutions up to 60,000 and mass accuracies smaller 2 ppm.

Despite gas chromatography – mass spectrometry (GC-MS), liquid chromatography – mass spectrometry (LC-MS) is widely used in metabolomics. Metabolites are separated with a suitable chromatographic method, mostly reversed phase (RP) or hydrophilic liquid interaction chromatography (HILIC), which is coupled to a MS [30-32]. Classical high-performance liquid chromatography (HPLC) is more and more replaced by ultrahigh-performance liquid chromatography (UHPLC). The most frequent used MS are triple quadrupole MS (QQQ-MS), for targeted and quadrupole-Time of Flight MS (Q-ToF-MS) for non-targeted metabolomics. UHPLC-MS allow high resolution separations of molecules using sub-2 μ m particles in the chromatographic bed. The analysis time can be significantly reduced and separation improved. It is ideally suited for complex matrices used in metabolomics. Furthermore, no or reduced ion suppression is observed in UHPLC-MS allowing more precise quantification of metabolites compared to DI-MS. A second promising approach for metabolomics is the use of two dimensional chromatographic separations, with orthogonal chemistries, increasing peak capacity dramatically [33]. In proteomics a combination of ion-exchange and RP chromatography or RP \times RP (two RP separations at different pH) is used for the separation of intact proteins or peptides from protein digests. The separation can be carried out off-line, collecting fractions from the first dimension and re-injecting them on the second or on-line using valve systems and trapping loops or columns. In both cases, analysis is time consuming and the application is rather suited for in-depth exploration of metabolomes than high-throughput analysis.

Chromatographic separation of metabolites offers several advantages over direct infusion, although they are more time consuming. Separation improves ionization yields, due to reduced ion suppression. This allows better quantification of metabolites. Furthermore, isobaric and isomeric molecules can be separated. Moreover, using fractionation strategies purification of unknown, novel metabolites is feasible.

Different modes of separation are employed, depending on the application. Reversed phase chromatography (RP) is the most used separation type. A non-polar stationary phase like C8 or C18 and mobile phase gradients from water to an organic solvent like methanol or acetonitrile are used. Typically mid- to non-polar metabolites are separated on RP columns. The use of ion pairing reagents allows also the separation of highly polar metabolites on RP columns [34]. Lipid species can be separated on a C8 column with an isopropanol gradient [21]. For chromatography of polar metabolites hydrophilic liquid interaction chromatography (HILIC) with bare silica or chemically hydrophilic modified silica columns and gradients from organic solvents to water are used. This method is not as widespread as RP, but application of metabolite separation using HILIC is growing [30].

Major problems with chromatography are overlapping peaks and retention time shifts, which can be overcome by deconvolution, addition of retention time marker and alignment of the obtained data. For this purpose several open source software like mzMine, metalign and other or commercial software like Genedata Expressionist Refiner MS or MarkerLynx exist [35, 36]. All software work in similar ways and produce data matrices with samples, chromatographic features and a response value, for example peak area or maximum intensity, which are suitable for further analysis by uni- and multivariate statistics.

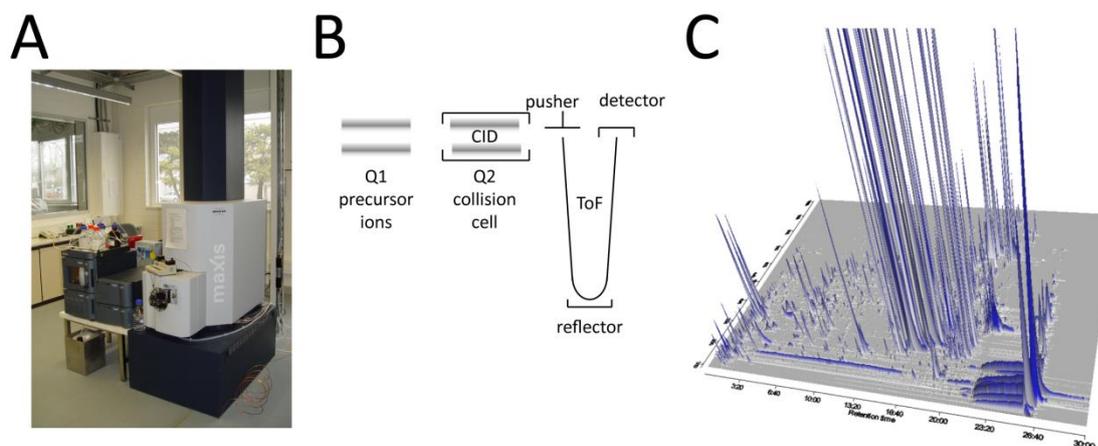


Figure 6: Principles of Q-ToF-MS

(A) Typical UHPLC-UHR-ToF-MS instrument. (B) Schematic drawing of a Q-ToF instrument. The first quadrupole serves as mass filter, the second as collision cell for MS/MS experiments and ion masses are detected in the ToF part. (C) Typical 3D chromatogram obtained from a metabolomic experiment.

Separation of ionic metabolites can be achieved by the use of capillary electrophoresis mass spectrometry (CE-MS). Buffer filled fused silica capillaries (30-100 cm long) are conducted to high voltages in the range of 10-50 kV. Metabolites are separated based on their electrophoretic mobility. Additionally the charged capillary wall induced an electro osmotic flow. Electrophoretic mobility depends on charge, size and the applied voltage; moreover

temperature, pH and ionic strength are factors influencing migration. Several pre-concentration techniques are known in CE, together with the coupling to a mass spectrometer it allows a very sensitive detection, down to nanomolar concentrations [37, 38]. Up to now it is not widely used, due to problems with performance stability. Great care must be taken of reagents, buffers and methods. If this can be guaranteed, CE-MS is a highly sensitive and powerful technique, especially for the analysis of energy and central carbon metabolism [37].

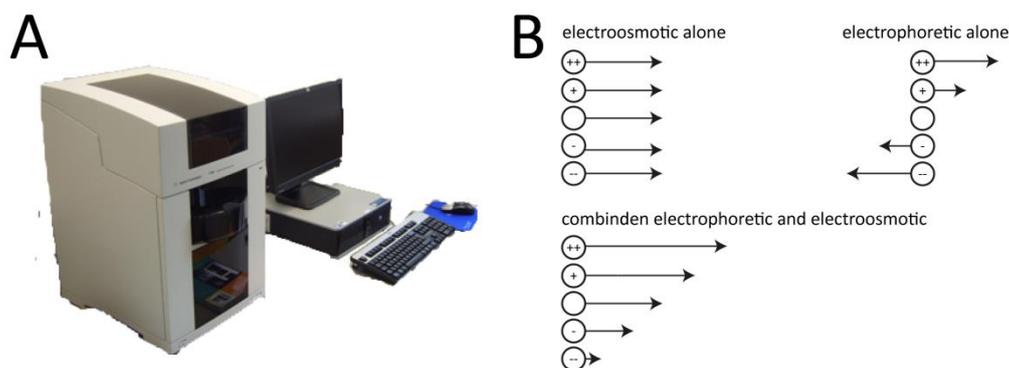


Figure 7: Principals of CE

(A) Typical CE instrument (B) Separation principle of CE. The high voltage applied to the capillary induces an electro osmotic flow, which affect all molecules, independent of their charge. The electrophoretic separation is achieved by different ion migration in the electric field. Together, electrophoretic and electro osmotic effects separate the ions.

1.2.3 Nuclear magnetic resonance (NMR)

Nuclear magnetic resonance (NMR) is used since decades as profiling method in medicine, mostly for urine and plasma. It allows the identification and quantification of the most abundant metabolites. The physical basis of this method is the measurement of the relaxation of resonating nuclei in a magnetic field. In metabolomics mostly ^1H -, ^{13}C -, ^{15}N - and sometimes also ^{31}P -NMR for phosphorylated compounds are used. A typical NMR experiment consists of two distinct steps; first magnetic nuclear spins are polarized with an external constant magnetic field. In the second step an electromagnetic pulse is applied to perturb the polarization. The frequency for this perturbation is specific for each nuclei and is called resonance frequency which directly proportional to the magnetic field strength. Magnets used in NMR are categorized according to the frequency of the proton. In a 21 T magnet a proton would have resonance frequency of 900 MHz, so that this magnet would be referred as 900 MHz magnet. The resonance frequency is influenced by the chemical environment of nuclei which shields the magnetic field to a certain extent, yielding an effective magnetic field B_{eff} . With this every nucleus in a molecule is affected by a slightly different magnetic field and has his special resonance frequency, a fact which allows determining structures of molecules. The obtained resonance frequencies are relativized to a standard compound, like tetramethylsilane

(TMS) and normalized on the operating frequency of the NMR. The normalized variable is called chemical shift δ and is measured in parts per million (ppm).

$$\delta(\text{ppm}) = \frac{\Delta\nu(\text{Hz})}{\nu_{\text{reference}}(\text{MHz})}$$

Equation 6: Definition of chemical shift

The chemical shift is relativized to standard compound, like TMS.

^{13}C - and ^{15}N -NMR are insensitive techniques as both nuclei have a low natural abundance (1.11% for ^{13}C , 0.37% for ^{15}N), but both are readily applied for structure determination of organic molecules. Sensitivity can be enhanced by labeling molecules with ^{13}C or ^{15}N . An example was published by An et al., which used *in vivo* labeling of *C. elegans* and ^{13}C heteronuclear multidimensional NMR for metabotyping of *sir-2.1* mutants. The work showed that with their developed workflow significantly enhanced sensitivity of about 2 orders of magnitude [39].

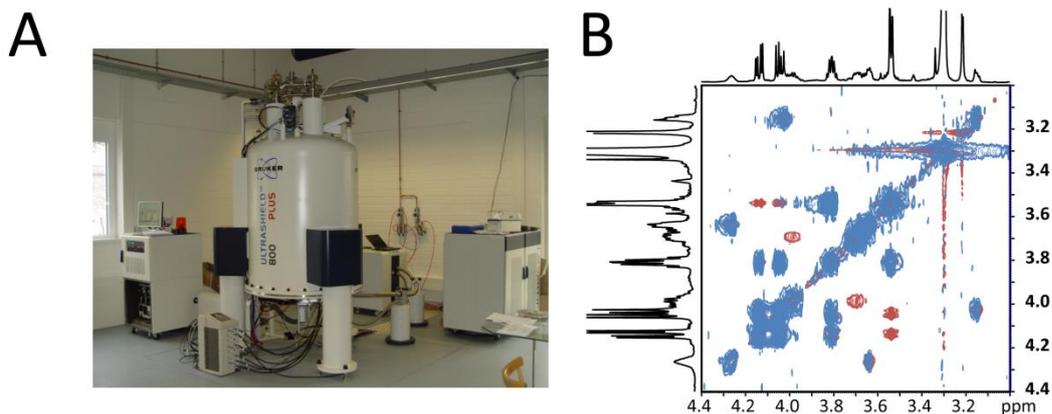


Figure 8: A) NMR B) construction

(A) Typical NMR instrument. (B) 2D-NMR experiments give more information on metabolite structures. An overlay of COSY (blue) and TOCSY (red) from a *C. elegans* extract is shown. COSY is based on spin-spin coupling and is certainly useful if in a 1D-NMR spectra are overlapping. TOCSY is also based on spin-spin coupling, but gives more information than COSY.

NMR is most likely to be the perfect profiling method; it is non-destructive, gives quantitative and qualitative information and can be easily automated. The major drawback compared to MS is its insensitivity, detecting only the most abundant metabolites. As peak shifts can overlap, a clear identification is not always possible. In metabolomics NMR people are shifting towards 2D NMR, which includes also information of spin-spin coupling. This is possible due to technical advances in NMR technology like higher magnetic fields, shortening measurement time for 2D NMR spectra.

Besides profiling, NMR is also used for structure elucidation of purified metabolites. Together with MS it allows a confirmation of possibly annotated metabolites or *de novo* description of novel metabolites.

The newest approach using NMR is a system together with LC and MS offered by Bruker as Metabolic profiler®. This system allows the fully automated collection of LC runs, MS and NMR information for complete metabolic profiling. Furthermore a recent study showed that the deuterated solvents used in NMR are not interfering with subsequent MS analysis, opening the field for combined approaches [40].

1.2.4 Comparison of the single methods

Each of the presented methods has its own advantages and disadvantages, which are summarized in Table 2.

Table 2: Advantages and Disadvantages of different metabolomics methods

The present methods have different advantages and disadvantages, making them useful for different questions. Today no ultimate solution, allowing unambiguous identification and quantification of all metabolites is available. Different methods have to be combined to enhance the coverage of the metabolome.

Method	Advantage	Disadvantage
<i>DI-ICR-FT/MS</i>	<ul style="list-style-type: none"> - <i>Fast</i> - <i>High resolution (> 500.000) and mass accuracy (<0.2ppm)</i> - <i>High sensitivity</i> 	<ul style="list-style-type: none"> - <i>No separation of isobaric and isomeric substances</i>
<i>UHPLC-UHR-ToF-MS</i>	<ul style="list-style-type: none"> - <i>High resolution (>50000) and mass accuracy (<3 ppm)</i> - <i>High sensitivity</i> - <i>Separation of isobaric and isomeric substances possible</i> 	<ul style="list-style-type: none"> - <i>Slow (gradients from 10 minutes to several hours possible)</i>
<i>GC-MS</i>	<ul style="list-style-type: none"> - <i>High sensitivity</i> - <i>Excellent chromatographic separation</i> 	<ul style="list-style-type: none"> - <i>Derivatisation needed</i>
<i>CE-MS</i>	<ul style="list-style-type: none"> - <i>Excellent separation of ionic species</i> - <i>High sensitivity</i> 	<ul style="list-style-type: none"> - <i>Unstable compared to LC-MS</i>
<i>NMR</i>	<ul style="list-style-type: none"> - <i>qualitative (structure) and quantitative (concentration) information in one technique</i> - <i>Non-destructive</i> 	<ul style="list-style-type: none"> - <i>Low sensitivity</i>

In this work DI-ICR-FT/MS and UHPLC-UHR-ToF-MS were utilized. Their high resolution and mass accuracy enable more precise metabolite annotation than the other methods. Compared

to GC and CE, UHPLC offers the advantage of a highly stable separation, even for hundreds of injections and no need for derivatisation. The combination of DI-ICR-FT/MS and UHPLC-UHR-ToF-MS offers several opportunities, which are discussed in chapter 4.

1.3 *Pseudomonas aeruginosa* – An opportunistic human pathogen

1.3.1 General overview on *Pseudomonas aeruginosa*

Pseudomonads are with over 50 different species one of the most diverse bacterial groups. The high degree of genomic variability and genetic adaptability is a great reason that they can be found in nearly every ecological niche. This proteobacterial genus was first described by the German botanist Walter Migula [41]. The genus belongs to the family of the *Pseudomonadaceae*, the order *Pseudomonadales* and the class *Gammaproteobacteria*.

Pseudomonas aeruginosa differs from all other species of the genus, because it is an opportunistic pathogen of plants and animals [42]. The rod-shaped, gram-negative, aerobic bacterium is a major burden in public health. In immunocompromised individuals it causes several diseases like pneumonia, urinary tract, surgical wound or bloodstream infections [43-45]. Furthermore *P. aeruginosa* is the most lethal gram-negative nosocomial pathogen. In 2000 the genome sequence of *P. aeruginosa* PAO1 was published. The 6.3 Mbp long genome contains over 5500 open reading frames [46]. Several other *Pseudomonas* genomes were sequenced since this time, including one of the most lethal strains *P. aeruginosa* UCBPP-PA14 [47].

Table 3: Genome size of several *Pseudomonas* species
Data taken from www.pseudomonas.com [48]

Organism	Genome Size (bp)	No. genes	Year of seq.
<i>P. aeruginosa</i> PAO1	6,264,404	5683	2000
<i>P. aeruginosa</i> UCBPP-PA14	6,537,648	5965	2006
<i>P. aeruginosa</i> PA7	6,588,339	6369	2010
<i>P. putida</i> F1	5,959,964	5403	2005
<i>P. stutzeri</i> A1501	4,567,418	4210	2008
<i>P. fluorescens</i> Pf-5	7,074,893	6233	2005
<i>P. syringae</i> pv. <i>phaseolicola</i> 1448A	5,928,787	5227	2005
<i>P. syringae</i> pv. <i>syringae</i> B728a	6,093,698	5172	2005
<i>P. syringae</i> pv. <i>tomato</i> str. DC3000	6,397,126	5685	2003

A genome wide metabolic reconstruction of the *P. aeruginosa* PAO1 metabolic network showed the broad capabilities of this pathogen. Totally 1056 genes, 1030 proteins and 883 reactions are used in this reconstruction, whereby 839 were gene-associated and 44 non gene-associated [49]. These broad metabolic capabilities together with several efflux pumps yields

in a strong resistance to antibiotics. Moreover it can even grow anaerobic on nitrate or arginine as terminal electron acceptors. *P. aeruginosa* infections are often found in patients suffering from cystic fibrosis (CF). A mutation in the CF transmembrane conductance regulator (CFTR) gene leads to a malfunction in the mucociliary clearance of inhaled microorganisms. *P. aeruginosa* is able to establish a chronic infection in the lungs of CF patients. It grows in a biofilm encapsulated by an exopolysaccharide called alginate. Beside this *P. aeruginosa* harbors several virulence factors used in infection. Table 4 gives a summary about different pathophysiological states caused by *P. aeruginosa*. One of the most studied *P. aeruginosa* strains is *P. aeruginosa* UCBPP-PA14. A genetic region, showing only partial similarity to other strains, called pathogenicity island 1 (PAPI-1), which contains a cluster of more than 100 genes could be the reason. This region is highly mobile and is maybe a reason for the evolution of variants with enhanced pathogenicity [50].

Table 4: Range of *P. aeruginosa* infections in man

Table taken from [51]

Organ	Infection	Acute/chronic	Origin	Prevention
Eye	Contact lense keratitis	Acute	Water	hygiene
Skin	Folliculitis	Acute	Water	hygiene
	wound/ulcer infection			
	Burn infections			
Ear	External otitis	Acute/chronic	Water?	hygiene
Nasal sinuses	Sinusitis	Chronic	Water?	?
Urine bladder	Urinary tract infection	Acute/chronic	?	?
Bones	Diabetic osteomyelitis in feet	Chronic	Water?	Hygiene
	Ventilator-associated pneumonia	Acute	Nosocomial humidifiers	Hygiene
Lungs/bronchi	endobronchiolitis, cystic fibrosis, bronchiectasis	Chronic	Nosocomial environment	Hygiene, early antibiotic eradication therapy of intermittent colonization
	Sepsis, neutropenic patients	Acute	?	Antibiotic prophylaxis

1.3.2 *P. aeruginosa* virulence factors

Pathogenic bacteria are using different protein secretion systems to inject effector proteins in their surrounding environment or a host cell. Six different secretion systems have been described so far [52]. *P. aeruginosa* carries five of this six and some of them also in several copies. The most prominent system is the type three secretion system or T3SS, which is shared between several pathogens, like *Chlamydia* sp., *Yersinia enterocolitica*, *Salmonella enterica* or *Shigella flexneri* [53]. The induction of T3SS can be triggered *in vitro* by Ca^{2+} chelation in the medium, upon effectors are released into the medium. Different strains of *P. aeruginosa* carry different combinations of T3SS effectors, separating cytotoxic from invasive strains. Cytotoxic strains carry *exoU*, *exoT* and sometimes *exoY*, an example is *P. aeruginosa* UCBPP-PA14, whereas invasive strains carry *exoS*, *exoT* and often *exoY*, e.g. *P. aeruginosa* PAO1 [54]. ExoY, an adenylate cyclase, converting cellular ATP to cyclic AMP (cAMP) is one of the most potent cellular toxins and will be discussed in chapter 5 in more detail [55]. ExoU is a potent phospholipase A2, ExoS and ExoT are bifunctional enzymes with GTPase activating function (N-terminus) and ADP-ribosyltransferase activity (C-terminus) [56-58]. An important point for all these effectors is that they need a eukaryotic interaction partner for activity, underlying their function as effectors in eukaryotic host cells.

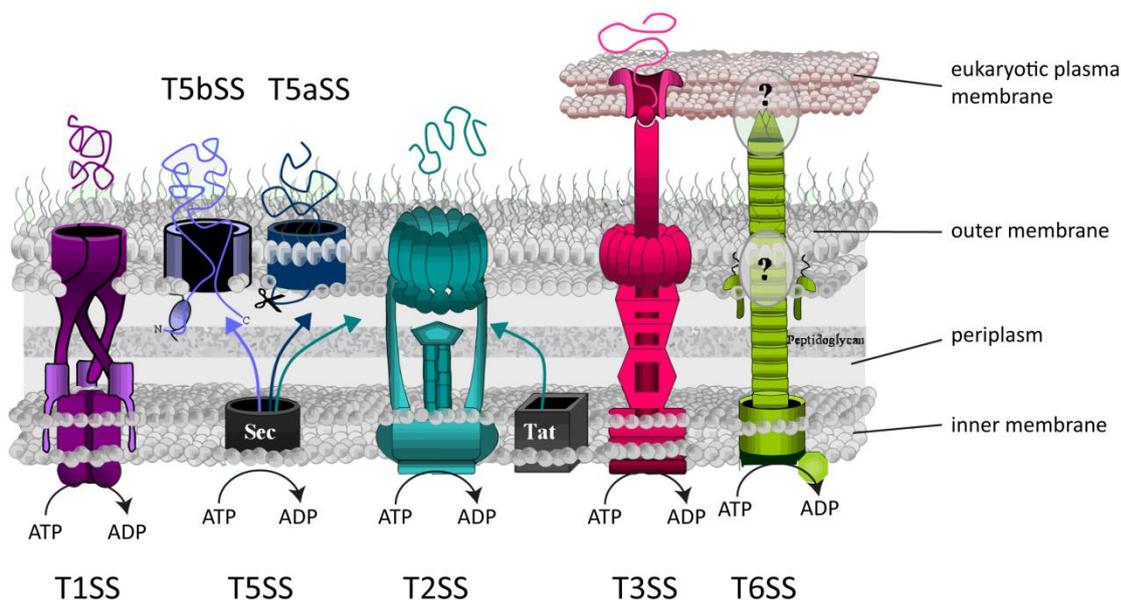


Figure 9: *P. aeruginosa* secretion systems

Figure adapted with permission from [59]. Five of the six secretion pathways found in Gram-negative bacteria are present in *P. aeruginosa*. Protein transport across the bacterial envelope can be subdivided into Sec-independent and Sec/Tat-dependent pathways. Type II (T2SS)-and type V (T5SS)-secreted exoproteins are firstly exported to the periplasm by the Sec or the Tat system before crossing the outer membrane by their dedicated secretion pathway. In contrast, type I, type III, and type VI (T1SS, T3SS, T6SS) exoproteins are directly taken over in the cytoplasm by their cognate secretion machinery.

A second major virulence factor in chronic infections, for example in CF patients, is the establishment of biofilms avoiding contact of *P. aeruginosa* with the immune system or antibiotic treatment. Biofilms are made of an exopolysaccharide called alginate, composed of mannuronic acid and guluronic acid.

Most of the virulence factors are controlled in population density dependent manner. Effectors are released and injected into the host cell after the bacterial community reached a certain density. The coordinated gene expression is regulated by a system called quorum sensing. Quorum sensing (QS) uses small, diffusible compounds as signaling molecules to coordinate certain behavior, like infection. In Gram-negative bacteria acyl homoserine lactones (AHL) serve as major signaling compounds. Biofilm formation, swarming motility and exopolysaccharide production are quorum sensing dependent in *P. aeruginosa*. A second quorum sensing system beside the AHL system exists and is based on a quinolones and regulates the T3SS.

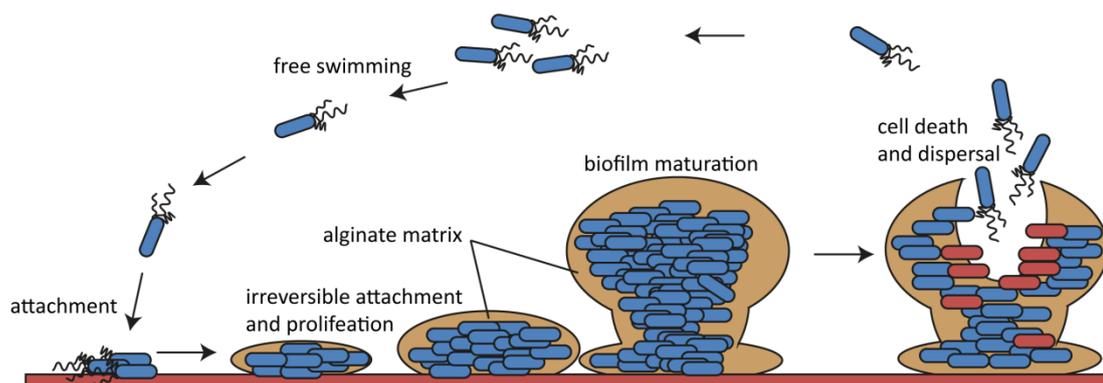


Figure 10: Infection cycle of *P. aeruginosa*

P. aeruginosa is able to attach via its pili to a surface, which is followed by an irreversible attachment, loss of the pili and proliferation. The growing colony is encapsulated by an alginate matrix and the biofilm matures. After a certain time, cell death can occur within the biofilm and cells are changing to free swimming lifestyle again to repeat the cycle.

1.3.3 Metabolomics studies on *Pseudomonas*

Several studies focusing on the metabolism of different *Pseudomonas* species have been carried out. Some are focusing on the metabolic capabilities of *P. aeruginosa* and other on differences between planktonic and biofilm growth or on production of QS molecules or other secondary metabolites.

A recent study from E. Frimmersdorf et al. used GC-MS to elucidate how *P. aeruginosa* adapts to different environments. Two different *P. aeruginosa* strains, PA01 and TBCF10839, were grown on either minimal medium with different carbon sources or a complex medium containing tryptone. During the exponential phase, metabolites directly available as carbon

sources and metabolites belonging to the central carbon metabolism were found in higher concentrations, whereas in the stationary phase metabolites connected to production of exopolysaccharides, development of biofilms and rhamnolipids were found [60]. A major feature of *P. aeruginosa* associated with CF patients is the change to a mucoid phenotype, showing overproduction of the extracellular polysaccharide alginate. The loss-of-function mutation of the anti-sigma factor MucA is commonly observed. In a NMR based metabolomics study, Behrends et al. showed that MucA modulates osmotic stress tolerance. The *mucA22* gene was mutated in *P. aeruginosa* PAO1 and supernatants and cell extracts of wildtype and mutant were subjected to NMR analysis. Higher levels of valine, methionine and the osmoprotectant glycine-betaine in supernatants are correlated with *mucA22* mutation. The *mucA22* mutants was also shown to be more susceptible to osmotic stress than the wild-type [61]. The study of biofilms in comparison with planktonic lifestyle was subject of the work of Gjersing et al. Also here NMR was employed. ^1H HRMAS NMR was used to profile biofilm and planktonic cells. PCA revealed differences between both, but significant metabolites were not further identified [62].

1.4 Model Systems for Host-Pathogen interactions

1.4.1 HeLa and other mammalian cell culture models

HeLa cells are immortal cells derived from cervical cancer cells taken from Henrietta Lacks. George Otto Gey propagated the cells and provided them to scientific community. The cells are widely used in science; over 500,000 articles refer to their use. One of the first uses was the testing of a polio vaccine by Jonas Salk in the 1950's. Beside HeLa cells several other cell types are in culture and are used for host-pathogen interactions. Human epithelia airway cells are often used for infection models of *P. aeruginosa* for example.

In a first work 1986, HeLa cells were used to assess the toxicity of *P. aeruginosa* culture supernatants. It was shown that toxicity of this supernatants is based on high molecular weight components [63]. Since this time several other studies using mammalian cell cultures have been carried out. One of these studies evaluated the cell internalization of the normally extracellular pathogen *P. aeruginosa*. Ha and Jin showed that invasion of HeLa cells by the invasive strain *P. aeruginosa* PAK is growth phase dependent and suggest that a difference in the regulatory mechanism of the T3SS [56]. Until now, no analysis of the effect of *P. aeruginosa* infection on mammalian cell metabolism using metabolomics has been carried out.



1.4.2 Metabolomics research in mammalian cell cultures

Although most applications in human and mammalian metabolomics are focusing on matrices like urine, plasma or even CSF, different mammalian cell lines have been used for metabolomics research. Most of them evaluate different quenching and extraction methods. A recent work by Lorenz et al. describes the comparison of different extraction methods for adherent cell cultures. The group used a rapid water rinse step prior to quenching with liquid nitrogen to remove remaining components of the culture medium. This rinse step together with MeOH/CHCl₃ extraction yielded metabolite extracts well suitable for further analysis with LC-MS. A closer look was taken especially to metabolites of the central carbon metabolism and labile species like ATP, GTP or NADH [64]. Beside this work, several other studies on optimized sample extraction for adherent mammalian cell cultures have been carried out. Some of them are reviewed in the previous mentioned work.

A further example for the use of mammalian cells in metabolomics was described by Sugimoto et al. In their work they used CE-ToF-MS based metabolites to identify metabolites secreted by macrophages upon stimulation with Lipopolysaccharides. Macrophages play an important role in inflammatory conditions, for example in periodontal diseases, where they lead to destruction of the infected periodontal tissue. It was shown that beside increased production of nitric oxide, TNF- α and prostaglandins, glycolysis and TCA cycle activities are increased [65].

1.4.3 *Caenorhabditis elegans*

Caenorhabditis elegans is a widely used model organism in biology. The 65 μm diameter and 1 mm long translucent nematode was introduced as model organism by Sydney Brenner in the 1960's. The second major breakthrough was the development of a technique for safe freezing and thawing of the worm by John Sulston, 1969. The genome was sequenced in 1998 and was the first of a multicellular species. Additionally 2003 the genome of the close related *C. briggsae* was sequenced. The *C. elegans* genome consists of six chromosomes, is 100,281,426 base pairs long and has over 20,000 genes. The nematode is of hermaphroditic nature, only a small number of males exist. This makes genetic modifications easy as they can be propagated through generations. Each hermaphrodite consists of exactly 959 and males of 1031 somatic cells. The normal life cycle of *C. elegans* consists of four larval stages, named L1 to L4, and an adult stage. Adult hermaphrodites can lay up to 300 eggs. Under unfavorable environmental conditions like low food availability or high population density L1 larvae can enter an alternative developmental stage called the dauer stage. This dauer stage is able to survive several months without food. If the dauer comes again to contact with food it develops

to a L4 larvae (Figure 11). *C. elegans* can be cultivated on agar plates with bacterial lawns or in liquid cultures with *Escherichia coli* OP50 or HB101 as normal food under laboratory conditions. A recent review, compared *C. elegans* against other widely used model organisms [66]. Table 5 summarizes all in the review mentioned advantages.

Sydney Brenner together with H. Robert Horvitz and John E. Sulston awarded 2002 the Nobel Prize for Medicine for their discovery of the programmed cell death using *C. elegans*. Two more Nobel Prizes were awarded using *C. elegans*, the exploration of RNA interference (2006) and the invention of the Green fluorescence protein (GFP) (2008).

Table 5: Advantages and Disadvantages of *C. elegans* as model organism

Table taken partly from [66]

(+) Advantages and (-) disadvantages of *C. elegans* as model organism

- (+) Inexpensive/easy to grow
- (+) Genome is available
- (+) Straightforward genetic tools exist
- (+) Short generation time: 2-3 week
- (+) Small; exactly 959 somatic cells
- (+) Invariant development
- (+) Transparent
- (+) Has organs/differentiated tissues
- (+) Mutants can be frozen
- (+)/(−) 50-80% of worm genes homologous to human genes

A



B

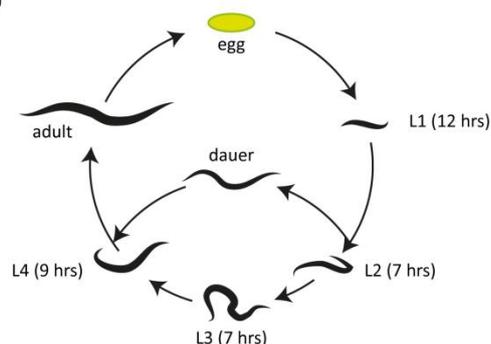


Figure 11: *C. elegans* is a widely used model organism with a short life cycle.

(A) Microscopic picture of *C. elegans*. Taken from [67]. (B) Developmental cycle of *C. elegans* is made up of four larval stages (L1-L4), an alternative dauer stage and the adult stage.

A lot of *C. elegans* genes show homology to mammal and human genes, which makes it an interesting model organism. For *C. elegans* the complete cell lineage is known. Nowadays it is used in biologic fields like physiology, ecology, genomics, neurobiology, evolutionary biology or cell biology. The nematode is also employed in pathogen-host interactions. It is known that several pathogens, including *P. aeruginosa* are able to kill *C. elegans*. For *P. aeruginosa* PA14

two different modes of killing *C. elegans* are known, depending on the cultivation medium. Bacteria grown on full medium show a “fast killing” of *C. elegans* by phenazine poisoning, leading to death in 3 to 24 hours, grown on minimal medium, they show a “slow killing” of the host with a persistent infection leading to death after a few days. Additionally two more ways of killing *C. elegans* are known from *P. aeruginosa* PAO1, lethal paralysis, leading to death within hours and “red death”, which is observed in response to PAO1 grown on phosphate depleted medium in conjunction with physiological stress on the nematodes. It is proven that certain virulence factors needed for infection in *C. elegans* have also important roles in infections in mammals. Mutation in *daf-2*, an insulin like receptor, leads to longevity and bacterial resistance. It has been shown that *P. aeruginosa* is able to suppress the initiate immune response of *C. elegans* by up-regulating *daf-2* [68].

Table 6: Essential and non-essential amino acids in *C. elegans* and humans

Taken from [66, 69] *C. elegans* has similar nutritional needs on amino acids compared to humans. E = essential. NE = non-essential.

Amino acid	<i>C. elegans</i>	Humans
<i>alanine</i>	NE	NE
<i>arginine</i>	E	E in development
<i>asparagine</i>	NE	NE
<i>aspartic acid</i>	NE	NE
<i>cysteine</i>	NE	E in development
<i>glutamic acid</i>	NE	NE
<i>glutamine</i>	NE	NE
<i>glycine</i>	NE	NE
<i>histidine</i>	E	E
<i>isoleucine</i>	E	E
<i>leucine</i>	E	E
<i>lysine</i>	E	E
<i>methionine</i>	E	E
<i>phenylalanine</i>	E	E
<i>proline</i>	NE	NE
<i>serine</i>	NE	NE
<i>threonine</i>	E	E
<i>tryptophan</i>	E	E
<i>tyrosine</i>	NE	E in development
<i>valine</i>	E	E

The genome of the worm contains genes for enzymes of fundamental metabolic pathways like glycolysis, TCA cycle, β -oxidation of fatty acids, electron transport chain and ATP synthesis. Moreover *C. elegans* is capable of ethanolic fermentation and malate dismutation, but it is not able to synthesize sterols de novo. Several amino acids are essential to *C. elegans*; Table 6 compares the essential and non-essential amino acids in the nematode and humans. At least for amino acids nutritional needs are similar between *C. elegans* and humans. Interestingly *C.*

C. elegans is able to synthesis branched chain fatty acids and uses it together with mono- and polyunsaturated fatty acids. Although *C. elegans* can be raised and maintained on a chemically defined medium, most laboratories prefer to feed the worm on bacteria. It has to be taken into account, especially when comparing results from *C. elegans* fed on different bacteria, that each of the different bacteria fed have nutritional differences. Metabolic capabilities and a draft metabolic reconstruction of *C. elegans* are discussed in chapter 2.

1.4.4 Metabolomics research on *C. elegans*

Several metabolomics studies using *C. elegans* have been carried out, mostly using NMR. As *C. elegans* has a hard cuticle disruption is a key issue in metabolite extraction. Standard lysis methods like the alkaline lysis method used for DNA extraction interfere with metabolite analysis. Basic experiments for optimization of metabolite extraction were carried out by Geier et al. Two different extraction solvent systems and six different disruption methods were tested with GC-MS, UHPLC-MS or NMR. They group concluded that 80% MeOH with bead beating is a good trade-off between extraction efficiency and stability [70]. One widely used method in *C. elegans* metabolomics is MAS-NMR using whole animal cultures. This method is simple and robust in sample preparation. A major work showed the capability of ^1H -MAS-NMR to reveal latent phenotypes in *C. elegans* mutants showing no change in morphology. The study used *sod-1* mutants to map the consequences of oxidative stress on the metabolome [71]. Major work was investigated in elucidating the metabolome of long-lived worms. A first work compared dauer larvae, worms with mutations in the insulin/insulin-like signaling pathway and translation-defective mutants. In this work it was shown that these three types of worms share a metabolic profile. Moreover dauer larvae showed higher levels of phosphoserine, hydroxyproline and choline, which can be an indication that autophagy plays a role in survival of dauers [72]. The group around Martin et al. employed different mutants in a ^1H -NMR study. Different combinations of *daf-2*, *daf-16* and *pept-1* mutants and their culture supernatant were profiled to evaluate their alterations of insulin/insulin-like signaling pathway and the amino acid absorption. Alterations in BCAA metabolism were the dominant features found in long-lived *daf-2* or *daf-2;pept-1* mutants, being consistent with the results from Fuchs et al. Moreover they show peculiar choline and acetate metabolism, which speaks for a dauer like metabolism with, increased autophagy. Lastly decreased one carbon pool metabolism is descriptive for *pept-1* mutants [73]. Butler et al. used HPLC and PCA to discriminate the differences between long-lived and short lived mitochondrial electron transport chain mutants. Metabolites were separated by HPLC with PDA detection, followed by fraction collection. Identity of metabolites was confirmed using LC-ESI-MS/MS. This approach showed

that pyruvate is enriched in long-lived mutants and lactate, α -ketoglutarate and Gly-Pro in short-lived mutants [74].

Other work used 2D-NMR spectroscopy to identify small signaling molecules. The group of Frank Schröder identified 3 known and 4 previously unknown ascarosides. These molecules, glycosides of different hydroxy fatty acids, are important signaling molecules in *C. elegans* biology. They serve as attractants and are responsible hormones for the entry into the dauer phase. Wild-type and *daf-22* worms were compared on the secreted metabolites. One of the novel identified ascarosides contained previously not described para-aminobenzoic moiety [75]. In a further work biogenesis of the ascarosides was evaluated. For this purpose several mutants were grown and subjected to LC-MS/MS analysis. The ascarosides undergo peroxisomal β -oxidation to yield different carbon side chain length. Afterwards the molecules are selectively derivatized with different moieties of varied biogenetic origin, for example with indol-3-carboxylic acid as degradation product of tryptophan. Summing up ascarosides are built from a modular library with building blocks from sugar metabolism, peroxisomal β -oxidation and amino acid metabolism [76].

Beside the here mentioned articles, several other metabolomic work was carried out on *C. elegans*, but until now, no work has been described focusing on the influence of pathogens on the *C. elegans* metabolism.



1.5 Pathomics project aims and objectives

The goal of the Pathomics project was to develop novel biomarkers and drug target candidates of two different pathogenic bacteria. On the one hand, the obligate intracellular pathogens *Chlamydiae* and on the other hand *P. aeruginosa*, an opportunistic extracellular human pathogen, were investigated. Both pathogens share a common principle for host-pathogen interaction: secretion of effector proteins in their host cells. To achieve this aims the Pathomics project was separated into three work packages (WP):

WP1: Transcriptional context

WP2: Host-pathogen interactome

WP3: Host-pathogen metabolome

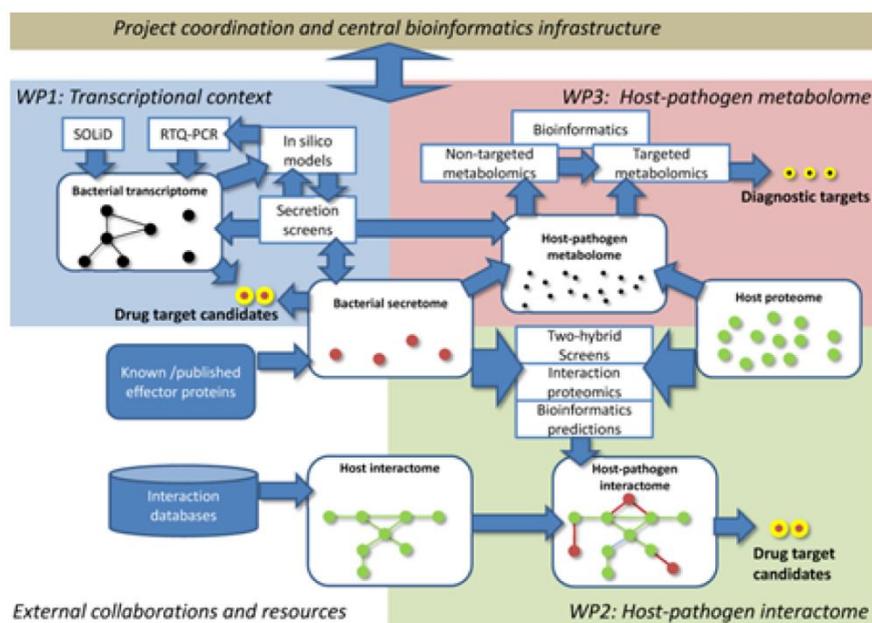


Figure 12: Project structure of Pathomics Project

The three different work packages will be integrated in a complete model of the host-pathogen interactions of *P. aeruginosa* and *Chlamydiae*. Metabolomics plays a central role, because in the “omics” hierarchy it is the closest to the observed phenotype.

WP1 focused on studying the bacterial transcriptome during infection and used Next Generation sequencing of transcripts. In the second work package (WP2) different methods were used to identify interaction partner of bacterial effector proteins, based on yeast two hybrid screens, protein pull downs and bioinformatics prediction. WP3 uses metabolomics to identify metabolic alterations occurring in the host-pathogen system during infection.

This work was settled in WP3. To reach the goal different host-pathogen systems were used. On the one side a cell culture model using HeLa cells as host and on the other *C. elegans*, a widely used model organism for host-pathogen interactions was used to study the influence

of *P. aeruginosa* on the host metabolism. Additionally to the analysis of the interactions this work also focused on the metabolism of different *Pseudomonads* to reveal their metabolic capabilities.

For this, several sample preparation, mass spectrometric and chromatographic methods were developed and used on routine basis for elucidation of metabolic alterations during infection. Furthermore novel and improved data analysis techniques were conducted for improved metabolite annotation, especially in the context of host-pathogen interactions.



1.6 Thesis structure

To achieve the aims different analytical methods were implemented and used on a routine basis. The existing MassTRIX web server for metabolite annotation was further optimized for more comfortable use and a joined analysis of metabolome and transcriptome data. Simple metabolic reconstructions helped to understand the capabilities of the studied organisms. This work is presented in chapter 2. Chapter 3 discusses metabolomes of different closely related *Pseudomonas* species. This data is compared to the metabolic reconstructions obtained in chapter 2. Beside the bioinformatic tools from chapter 2, chemoinformatic methods were implemented for further false positive filtering of metabolite annotations. These tools are presented in chapter 4, together with the used *C. elegans* infection model on which they were developed. This chapter also includes the biological interpretation of the obtained data. For a HeLa cell based infection model methods for sample preparation and extraction had to be developed. Results from this development are depicted in chapter 5 together with results from this infection model. Furthermore the bacterial effector ExoY and its effects on the host metabolism were studied in HeLa cells. The final conclusion in chapter 6 compares the two different infection models and lists their similarities and differences. Additionally further ways of optimization and topics for future investigations are demonstrated.

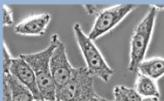
<p>Chapter 1:</p> <ul style="list-style-type: none"> - Introduction to Metabolomics - Introduction to <i>P. aeruginosa</i>, <i>C. elegans</i> and HeLa cells - Project aims 	
<p>Chapter 2:</p> <ul style="list-style-type: none"> - Review on chemo- and bioinformatics research in metabolomics - Optimization of MassTRIX - Metabolic reconstructions for host-pathogen metabolomics 	
<p>Chapter 3:</p> <ul style="list-style-type: none"> - Comparison of metabolomes from different <i>Pseudomonas</i> species - Comparison of real life metabolomes against <i>in silico</i> predictions 	
<p>Chapter 4:</p> <ul style="list-style-type: none"> - Tools for filtering of false positive annotations - Quantitative structure retention relationship (QSRR) and mass difference networking - Lipid identification based on data-dependent tandem mass spectrometry and <i>in silico</i> fragmentation - biological interpretation of <i>C. elegans</i> infection model 	
<p>Chapter 5:</p> <ul style="list-style-type: none"> - Development of sample preparation and extraction method for HeLa cell infection model - biological interpretation of HeLa cell infection model - effects of ExoY on the host metabolome 	
<p>Chapter 6:</p> <ul style="list-style-type: none"> - Comparison of <i>C. elegans</i> and HeLa infection model - Discussion of applicability of metabolomics in phylogenetics and taxonomy of bacteria - Outlook 	

Figure 13: Overview on thesis structure





2.1 Bio- and Chemoinformatics to assist metabolomics research

Metabolomics like other “omics” data sets tend to be rich in both information and size. On instrumental side a lot of effort has been carried out to automate sample preparation and data acquisition. On the side of data analysis several software solutions were developed for workflow based preprocessing of MS and NMR data. For biological interpretation of measured data only a small amount of software is available, mostly being free-ware and webserver programmed by research groups due to their own needs.

More groups facing the problem of identification of novel metabolites in large datasets. In this case bio- and chemoinformatics can help. On the one side bioinformatics helps to explore and understand the metabolic capabilities of a studied organism. Several biological databases contain information on enzymes, metabolites and their relation, e.g. KEGG, MetaCyc or BRENDA [77-79]. These can be used to predict *in silico* metabolomes, which can be compared against real measurement data. On the other side chemoinformatics can help in data interpretation of measured data. One of the easiest cases is the calculation of possible formulas for a given mass. Other examples are prediction of physical and chemical properties of novel structures, like pK_a , logP or isoelectric point.

The two most wide-spread programming languages in bio- and chemoinformatics are Perl and Java. Perl is a platform-independent and interpreted language and is the most used language in bioinformatics. Perl was developed in the 1980's by Larry Wall and has influences from Awk, BASIC-PLUS, C, C++, Python and many more. It is flexible enough to be used also for bigger projects. The basic Perl distribution can be expanded with several modules, stored in central resource called CPAN (Comprehensive Perl Archive Network). The second language, Java, is an object oriented programming language and was developed by Sun Microsystems, which belongs to Oracle since 2010. It is robust, safe, portable, easy to interpret and highly dynamic. Several API's (application programming interface) for chemoinformatics exist, whereas ChemAxon delivers one of the biggest, which is also used by several pharmaceutical companies, e.g. GlaxoSmithKline. ChemAxon includes programs and an API for prediction of several chemical and physical properties, as well as clustering of chemical structures [80-82]. Additionally it includes possibilities for warehousing of chemical structures in databases like MySQL or Oracle.

2.1.1 Review on chemo and bioinformatics research in metabolomics

A first paper dealing with chemoinformatics approaches in metabolomics was published 2003 by Nobeli et al. In their work they predicted several physicochemical properties for the *Escherichia coli* metabolome. 745 two-dimensional structures were taken from BioCyc and

KEGG and characterized on their molecular properties. Most metabolites have a molecular weight smaller than 500 Da and 80% a logP smaller than 0. Additionally a library of 57 fragment structures has been used to further characterize the metabolome [83]. Such approaches can be used in the process of unknown identification. Knowledge about chemical building blocks occurring in organisms reduces possibilities for structures of unknown molecules. Indeed chemical fingerprinting was used in conjunction with tandem mass spectrometry to develop a tool for querying large chemical databases for identification of unknown molecules [84].

A common approach to identify unknown molecules is the calculation of possible elemental formulas. Kind and Fiehn developed a set of filters known as the seven golden rules to minimize false positive formulas. These rules include restricting numbers of elements (rule 1), following the LEWIS and SENIOR chemical rules for valency checks (rule 2), isotopic pattern (rule 3), hydrogen/carbon ratios (rule 4), elemental ratios of nitrogen, phosphorous, sulfur and oxygen against carbon (rule 5), element ratio probabilities (rule 6) and presence of trimethylsilylated compounds for GC-MS (rule 7). These rules were derived from 68237 existing chemical formulas and checked for consistency using different chemical databases [28]. To overcome the major bottleneck, the identification of unknown compounds, Wolf et al. presented an *in silico* fragmentation approach. The MetFrag suite searches for candidates in different compound databases and ranks them by the agreement between measured and *in silico* fragments. The calculation of *in silico* fragments is based on bond disconnection and produces all possible topological fragments of a candidate. Run times for a single compound are typically less than one second. This fast implementation enables to screen large databases for possible candidates [85].

These are examples for the effort of the metabolomics research community to make data analysis easier. However, the proposed methods are used for annotation and identification of metabolites. Calculation of molecular formulas is useful, but for most of the calculated formulas no or hits with low biological significance will be found in the public chemical databases like PubChem or ChemSpider [8, 86]. A usual first step in data analysis is the annotation of known metabolites from different biological databases, like KEGG, LipidMaps or HMDB [20, 77, 87, 88]. Although useful, two of the three mentioned databases have no interface for a mass spectrometry based search of metabolites. HMDB allows the user to enter masses, tandem mass spectra, GC-MS and 1D and 2D NMR data for metabolite search. KEGG only offers a formula based search and LipidMaps a prediction tools for mass spectrometry.

To overcome this issue, the MassTRIX webserver has been developed.



2.2 MassTRIX: A web-based metabolite annotation tool

MassTRIX is a webserver for direct annotation of ICR-FT/MS data to metabolic pathways originally developed by Karsten Suhre and Philippe Schmitt-Kopplin [89]. The newest version was updated to allow the combined analysis of transcriptome and metabolome data.

2.2.1 Annotation of mass spectrometric data

The core functionality uses a given mass list and compares it against theoretical masses of adducts of metabolites from a chosen database within a certain error range. Table 7 shows the mass spectrometric adducts that are covered by MassTRIX.

Table 7: MS adducts covered by MassTRIX

All possible adduct masses are calculated based on exact atomic masses. For atoms with significantly abundant isotopes all isotopes were included ($e^- = 5.48579 \times 10^{-4}$ u)

Scan mode	Adduct	calculation
Negative	$[M-H]^-$	$M - 1.007825037 \cdot e^-$
	$[M+Br]^-$	$M + 78.9183361 \cdot e^-$ (^{79}Br , 50.69%)
		$M + 80.91629 \cdot e^-$ (^{80}Br , 49.31%)
	$[M+Cl]^-$	$M + 34.96885273 \cdot e^-$ (^{35}Cl , 75.77%)
$M + 36.96590262 \cdot e^-$ (^{37}Cl , 24.23%)		
Neutral	$[M]$	M
Positive	$[M+H]^+$	$M + 1.007825037 \cdot e^-$
	$[M+Na]^+$	$M + 22.9897697 \cdot e^-$
	$[M+K]^+$	$M + 38.9637079 \cdot e^-$ (^{39}K , 93.26%)
		$M + 40.9618254 \cdot e^-$ (^{41}K , 6.73%)

Metabolites from different databases are used by MassTRIX for the annotation process. The monoisotopic masses were recalculated based on exact atomic masses using the molecular formulas stored in the respective database [90]. At the moment databases supported by MassTRIX are: KEGG, HMDB, LIPIDMAPS and MetaCyc [20, 77, 78, 88] in different combinations. Also metabolomic databases are growing fast, not all known metabolites are inventoried. Several signaling molecules from *C. elegans* for example or plant secondary cannot be found in KEGG for example [77]. To overcome this circumstance in the analysis of metabolomics data, the new version of MassTRIX includes the possibility to upload a list of own molecules as pre-calculated adducts, which will be included in the annotation process. If KEGG ID's are supplied with this list, pathway mapping of these compounds is possible. Moreover with this function adducts not covered by MassTRIX, e.g. $[M+H-H_2O]^+$ or $[M+2H]^{2+}$, can be included. Uploaded masses are matched against the theoretical adduct mass of metabolites from the chosen database within a certain error range, usually expressed in ppm. A maximum error up to 3 ppm is possible, for instruments with lower resolution an absolute

error range has been added in the new version. Several elements of this adducts have isotopes with significant natural abundances. To avoid false positive annotations, adducts are filtered according to isotopes. Bromine for example has two different isotopes (^{79}Br and ^{81}Br) with a natural abundance of about 50%. Peaks identified as $[\text{M}+\text{Br}]^-$ adduct are only kept if both isotopes were found. Isotopic filtering is also applied to ^{13}C , ^{15}N , ^{18}O and ^{34}S species in molecules, meaning an isotope peak is considered as true if the corresponding monoisotopic peak is also found. Figure 14 shows the main workflow of MasSTRIX. As alternative a list of KEGG compound ID's can be submitted, bypassing the whole annotation procedure.

2.2.2 Analysis of transcriptomic data

Transcriptomic data can be submitted to MasSTRIX in two different formats, either a self-annotated file or *.cel* files for Affymetrix gene chips. The first one contains KEGG ID's, KEGG KO numbers, EC number or gene identifiers and a foldchange or UP and DOWN as keywords. The submitted values are used for coloring of the respective enzyme on pathway metabolic pathway maps together with annotated compounds. This format allows the use of non Affymetrix gene expression chips or other techniques like serial analysis of gene expression (SAGE) or next generation sequencing of transcripts (NGS). In the second variant two *.cel*-files as output of Affymetrix gene expression chips are submitted. One serves as reference file and the other is specific for the sample state. The data is analyzed with the GCRMA package in R. Gene chip robust multi array averaging (GCRMA) is an improved version of the robust multi array (RMA) method of normalization and summarization. GCRMA uses sequence specific probe affinities of gene chip probes for more accurate gene expression values. Results from this analysis are fully downloadable for further investigations.

2.2.3 Visualization of metabolomics and transcriptomic data on KEGG pathways

KEGG provides a programmatic access to their database via an API. The KEGG API is a SOAP based webservice, which can be used to submit a list of KEGG compound ID's, KEGG KO numbers, EC numbers or gene identifier together with a list of colors to obtain colored pathway maps of a chosen organism. MasSTRIX uses the KEGG API to visualize results from the compound annotation and transcriptomic data analysis on metabolic pathway maps. The displayed maps are fully clickable, different result pages are cross-linked and ID's are linked out to the respective pathway. Because the step of pathway mapping is the slowest in the whole process, by default only the "Glycolysis and Gluconeogenesis" pathway (map00010) is used by MasSTRIX to save calculation time. If other additional pathway should be colored this has to be pointed out in the input window by giving the corresponding pathway numbers. It should be noted, even if only the default is used, that all metabolite-pathway relationships are

included in the download files. Current limitations are the missing multicoloring and multiplexing in the KEGG API. Due to this fact, more complex data, e.g. combination of transcriptomics, proteomics and metabolomics data cannot be performed without own programming work. Several visualization tools rely on custom rewritten SVG and XML files, Paintomics for example [91]. This may allow a more complex data representation, but the creation of these files is a tedious work, especially if newer version of a map is released, everything has to be changed. It would be a promising feature if KEGG would provide an interface for more complex data, like pathway tools from BioCyc. This tool allows to map e.g. bar plots to the MetaCyc pathways.

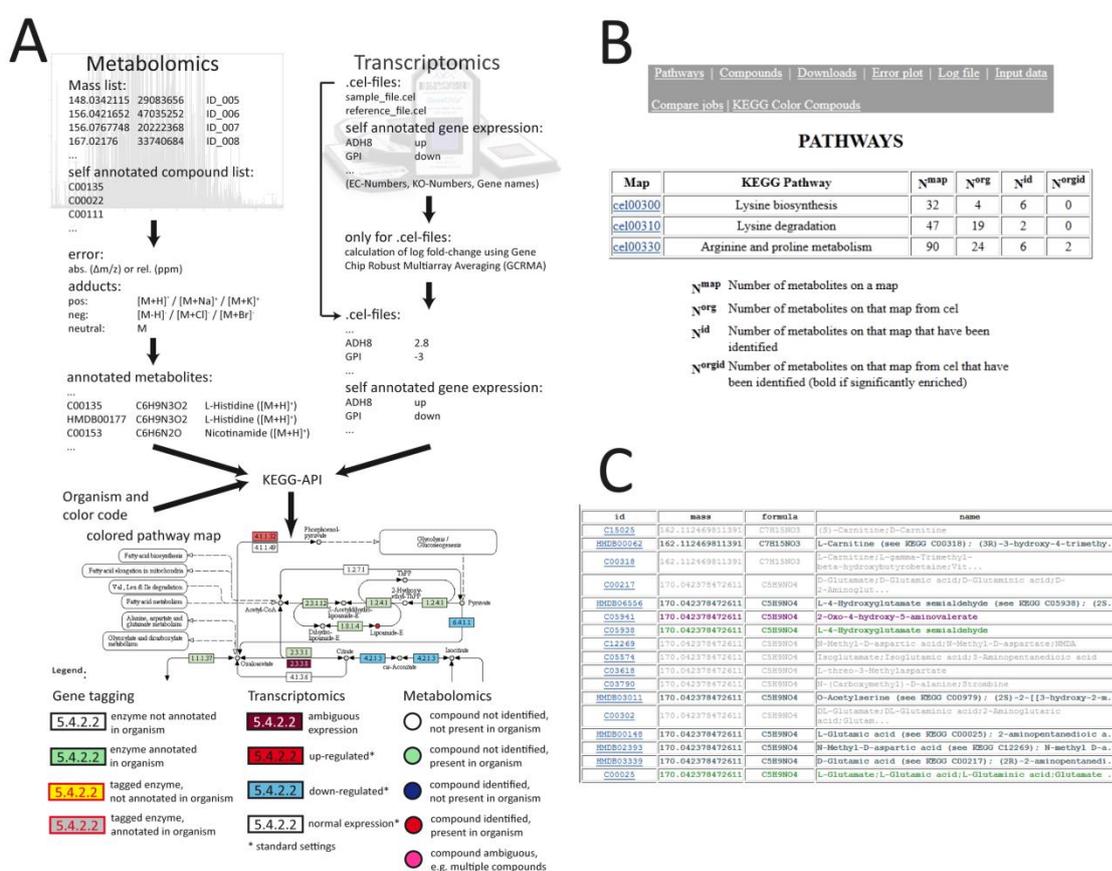


Figure 14: Principal workflow and outputs of MassTRIX.

(A) Workflow for metabolomics and transcriptomic data. Results from both data types are mapped together on metabolic pathways obtained from KEGG. (B) Pathway section of a MassTRIX job, showing number of annotated metabolites per pathway map. (C) Compound section showing annotated compounds. Green entries were found on minimum one of the selected pathways.

2.2.4 Comparison of different jobs

Different MassTRIX jobs can be compared on either pathway or metabolite level. This allows the analysis of time series experiments or different experimental conditions. The pathway comparison tool uses the 5 digit numbers of pathway maps in KEGG, without organism information. This enables interorganismic comparison and it is especially useful for

host-pathogen metabolomics. The comparison on pathway level gives a quick overview on metabolic pathways that are possibly up- or down regulated in pathophysiological state. This approach has been used in several publications to obtain a first overview on altered pathway in certain sample state [92, 93], e.g. diseased or treated.

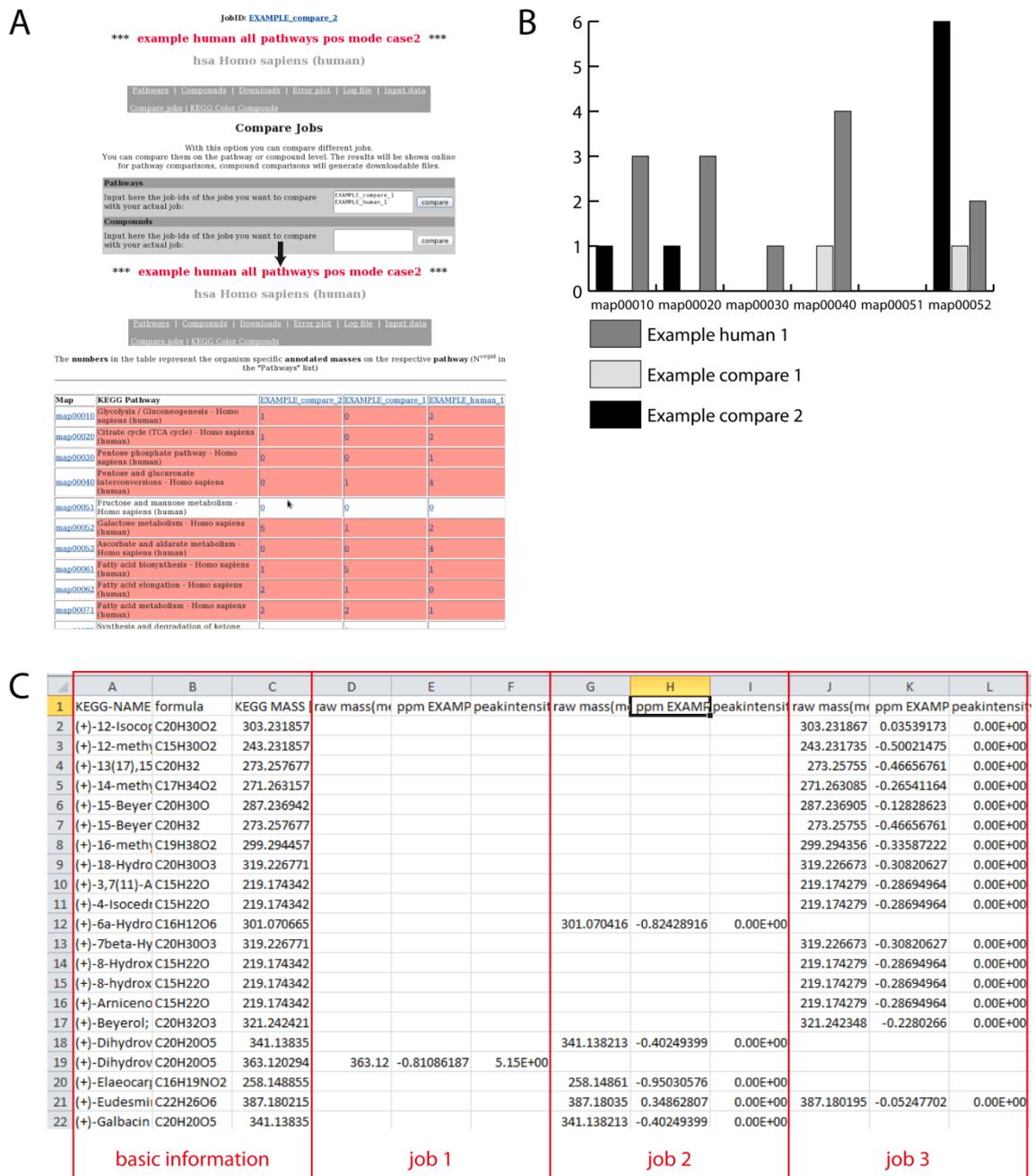


Figure 15 Possible comparisons between jobs in MassTRIX

(A) Screen shot from the "Compare jobs" functionality and the obtained result page comparing jobs on pathway level. (B) Result from comparing 3 different jobs on pathway level represented in a barplot. X-Axis represents the different pathway maps and Y-Axis the number of annotated compounds on this pathway. (C) Result from comparing 3 different jobs on compound level. The file can directly be downloaded and opened in MS Excel.

2.2.5 Comparison against other existing resources

Beside MasSTRIX several other solutions for annotation of mass spectrometric data exist. Two examples are the Pathos webserver (<http://motif.gla.ac.uk/Pathos/pathos.html>) [94] and Paintomics (www.paintomics.org) [91]. Pathos principally is based on the same functionality as MasSTRIX and is written in Java and uses an underlying MySQL databases. It annotates possible metabolites within an error range to experimental masses. To compare both, a list of 25644 masses from a *C. elegans* metabolome extract measured on a Bruker solariX ICR-FT/MS were subjected to the two webserver. Search was carried out in positive mode with correction for $[M+H]^+$ and $[M+Na]^+$ adducts and 3 ppm maximum error. Pathos is capable to correct for all adducts listed in [24]. Pathos annotated 1229 metabolites from the input masses in two minutes. Colored pathway maps are created on demand, after the annotation process. However this pathway maps are not cross-linked with other result pages like in MasSTRIX. Additionally submitted jobs are not stored on server and have to be recalculated every time from the beginning. A basic comparison between different samples states is possible, mapping masses from different samples with different colors on the pathways. With Pathos no joined analysis and visualization of metabolomics and transcriptomics data is possible. MasSTRIX needed for the same calculation 15 minutes and yielded 7412 annotated masses. Because Pathos is only using masses occurring on metabolic pathways it is limited to a certain subset of KEGG. MasSTRIX uses a flat file database, which slows down performance compared to Pathos. Transcriptomic data was not included in the MasSTRIX analysis to have similar demands on calculation and only the default settings for pathway mapping are used. To obtain additional colored pathways, two to three minutes more per pathway are needed, due to connection via the KEGG API to the KEGG database. Additional transcriptome data will just need several minutes more for calculation. The last webserver, Paintomics, only allows the joint visualization of pre-analyzed and identified metabolites and genes, making it different from the two previous webserver. Representations based on the KEGG pathways are completely rewritten with XML and SVG technology to allow a multiplexed data representation. This is the big advantage of Paintomics.

Summing up, MasSTRIX is a valuable resource for a first glance analysis of mass spectrometric data. In terms of speed it is out performed by Pathos, but the data obtained is more comprehensive. Changing the underlying database of MasSTRIX from flat file to a SQL based database would greatly improve speed.



2.3 Metabolic reconstructions as basis for host-pathogen metabolomics

Metabolic reconstructions are using genomic data to retrieve information about enzymes and regulatory motifs occurring in an organism. They allow the determination of metabolic reaction networks to obtain organism specific metabolite information, which can be certainly useful in host-pathogen metabolomics. Often it cannot be determined if a metabolites origin is the host or the pathogen. Metabolic reconstructions can, for known metabolites, show if an organism is able to synthesize such a metabolite. In case of novel unknown metabolites it can deliver cues about a possible synthesis pathways and which enzymes may be involved. For several organisms metabolic reconstructions already exist and are stored in databases like MetaCyc [78] or Model SEED [95].

Basis for a metabolic reconstruction is a fully annotated genome. Mostly, data stored in the KEGG database is used. Information about genes is converted to proteins and protein complexes and their enzymatic function. Here BRENDA (Braunschweig Enzyme database), beside other databases, is one of the most utilized [79]. The only major drawback is that it is not as comprehensive as other databases like KEGG. Possible substrates, products and cofactors are added to the network. Until this point the reconstruction can be completely automated. This reconstruction can contain several false positives, due to assignment of metabolites to enzymes that are not present in the studied organism, but present in other organisms stored in the used database. To overcome this, an expert has to evaluate the automatically created draft reconstruction. In most of the cases this is the time limiting step, because primary literature is used to refine the model. One example for *Pseudomonas aeruginosa* PAO1 reconstruction has been already mentioned in the introduction [49].

Reconstructions in this work should rather help to understand the metabolic capabilities and possible metabolic interactions between the studied organisms than to represent a 100% correct model for *in silico* predictions. Pathogenic bacteria infect host for their own survival. A host represents an ecological, nutrient rich niche, in which a pathogen can grow. Sometimes pathogens are completely dependent on the host metabolism. An extreme example is the missing biosynthetic pathway for tryptophan in *Chlamydia pneumonia* [96]. To unravel metabolic interactions, metabolic reconstructions can help to understand the interplay between host and pathogen. *P. aeruginosa* is a special case, because it has broad metabolic capabilities and survives in broad range of ecological niches. In this case nutrient supply may be not the first reason to infect higher organisms, but the host may synthesize different metabolites easier than the bacteria itself. Additionally, energy resources of the host and not

the pathogen own are used for synthesis. Virulence factors secreted by *P. aeruginosa* need to be highly potent in re-ordering metabolic fluxes or to perturb pathways. Here highly effective effectors are needed to disturb the cell homeostasis, because the pathogen can only synthesize a limited amount of effector and these have to lead a maximum effect.

2.3.1 Metabolic reconstruction of *Pseudomonas* species

The genus *Pseudomonas* is a very diverse branch of the bacterial clade. All *Pseudomonas sp.* are extracellular bacteria, including human, animal and plant pathogens. This leads to the suggestion that the metabolic capabilities in this genus are very broad.

An approach to compare metabolic capabilities was published by Kastenmüller et al., which originally was used to compare the metabolic capabilities of the 214 sequenced genomes and to compare different environmental conditions and diseases. Analysis showed clusters of metabolically similar microbes deriving from similar ecological niches [97]. The same analysis was carried out on the *Pseudomonas* genomes stored in the KEGG database. This analysis showed a cluster of the three strains of the human pathogen *P. aeruginosa*, which make this pathogen different from the other species. To uncover metabolic traits special to the pathogen metabolic reconstructions of different *Pseudomonas* species were compared.

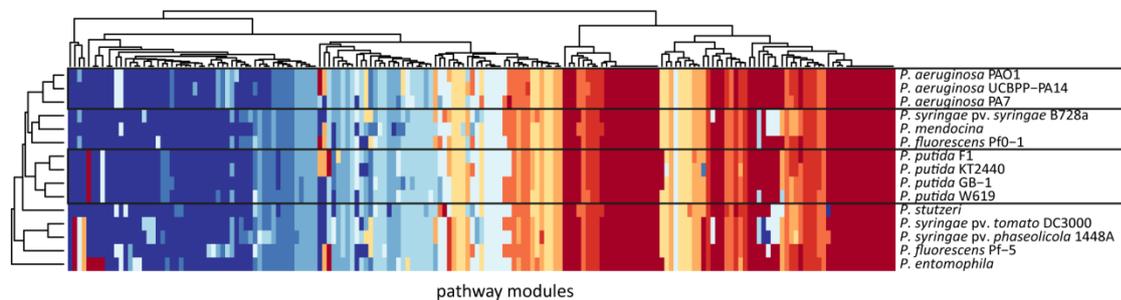


Figure 16: Metabolic capabilities of different *Pseudomonas* species

Different *Pseudomonas* species were compared on their metabolic capabilities as described in [97]. Blue color represents a low score, red color a high score on the respective pathway module. *P. aeruginosa* and *P. putida* form clear species specific cluster, whereas all other species are spread over the dendrogram.

To generate a first overview data set, KEGG database was used. Draft reconstructions were created for all *Pseudomonas* species listed in Table 8. These different strains and species are also present in BioCyc and Pathosystems Resource Integration Center (PATRIC) [98]. Genes for each species were collected from the genome part of KEGG using the KEGG API, together with associated EC numbers. From these compounds linked to the corresponding enzymes were obtained and all information were stored in one flat file per organism. Metabolite information of all available *Pseudomonas* species was integrated in one table containing all metabolites and species. Presence or absence of a metabolite in one species is indicated with 1 and 0. This

table was used to compute an overview on the *in silico* pan and core metabolome. A core metabolome represents metabolites that are present in all studied species, whereas the pan metabolome are all metabolites of all species. The terms were borrowed from genomics, where they are referred to core and pan genomes.

Table 8: Number of genes of different *Pseudomonas* species

Data was directly taken from PATRIC, KEGG and BioCyc. Between PATRIC and KEGG numbers only differ slightly, whereas BioCyc shows bigger differences.

Organism	Number of genes		
	PATRIC	KEGG	BioCyc
<i>Pseudomonas aeruginosa</i> PAO1	5836	5571	5646
<i>Pseudomonas aeruginosa</i> UCBPP-PA14	6022	5892	5964
<i>Pseudomonas aeruginosa</i> PA7	6064	6286	6361
<i>Pseudomonas aeruginosa</i> LESB58	6137	5925	6005
<i>Pseudomonas fluorescens</i> Pf-5	6363	6108	6274
<i>Pseudomonas fluorescens</i> Pf0-1	5763	5722	5827
<i>Pseudomonas fluorescens</i> SBW25	6256	6395	6475
<i>Pseudomonas putida</i> GB-1	5438	5408	5529
<i>Pseudomonas putida</i> F1	5377	5250	5345
<i>Pseudomonas putida</i> W619	5246	5182	5279
<i>Pseudomonas putida</i> KT2440	5777	5350	5516
<i>Pseudomonas syringae</i> pv. <i>syringae</i> B728a	5293	5089	5169
<i>Pseudomonas stutzeri</i> A1501	4203	4128	4209

The BioCyc database allows the direct download of compounds predicted to be metabolized in a given species. This information was directly used and combined in a single table. Reconstructions stored in the BioCyc database collection are generated by the PathoLogic program, meaning that has not undergone any manual curation [99, 100]. A major advantage of BioCyc over KEGG is the storage of secondary metabolites and the linkage to a respective organism. KEGG also stores several secondary metabolites, but missing linkage to producing organisms makes this information for this purpose useless.

A reconstruction from PATRIC was created by downloading genes and their corresponding EC numbers. For annotation of genomes PATRIC uses the SEED with its annotation server RAST [101]. Corresponding compounds to the EC numbers were downloaded from KEGG via the KEGG API.

In general, metabolite lists from KEGG contain more compounds than the corresponding lists obtained from BioCyc and reconstruction from PATRIC contains always the most metabolites. This might be due to the fact that KEGG is a more general database and many more compounds are linked to a single enzyme, whereas BioCyc is more focused on bacterial metabolism, with a few exceptions, like HumanCyc or AraCyc. Beside its general usability and size, the KEGG database contains several errors, possibly due to the automated generations of the single organism databases. An example is an incomplete TCA cycle in *P. aeruginosa* UCBPP-PA14 in KEGG. The last reaction converting malate to oxaloacetate and closing the cycle is missing in the KEGG database. In contrast, BioCyc contains this reaction and refers to a gene (PA14_19190) that has the same annotation in PATRIC. However searching for this gene in KEGG, the system returns the right annotation and EC number, but it is not linked to the respective pathway map. Figure 17 shows a comparison of the core and pan metabolome of the data obtained from BioCyc.

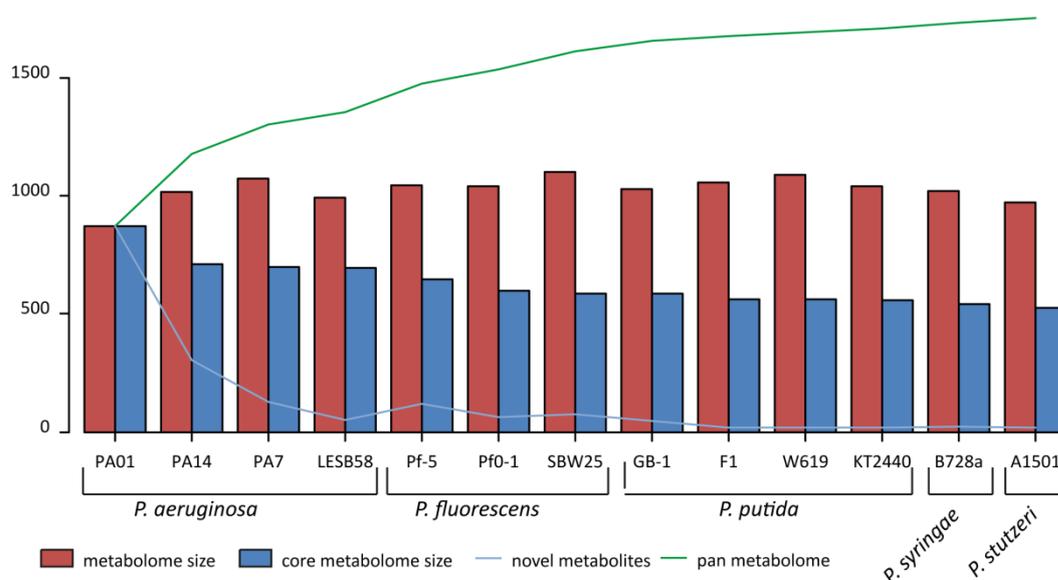


Figure 17: Comparison of the in silico metabolomes downloaded from BioCyc.

Data was compared on basis of metabolome size and occurrence of metabolites in the different species. Core metabolites are defined as metabolites that are found across all species. The pan metabolome are all compounds in all species, which are totally 1755. Novel metabolites are defined as metabolites that were not found in previous species.

The core metabolome represents central carbon metabolites, like glucose, intermediates of glycolysis and TCA cycle, amino acids, nucleotides and nucleosides etc. The pan metabolome includes mainly secondary metabolites like pyochelin or pyocyanin. The core metabolome counts 527 metabolites for BioCyc, 932 for KEGG and 1966 for PATRIC respectively. It is interesting that with PATRIC the number is surprisingly high. *P. aeruginosa* UCBPP-PA14 was used to compare results from KEGG and PATRIC. Most of the metabolites using the gene annotation stored in KEGG for reconstruction overlap with PATRIC. Only 99 were unique for

KEGG and 1100 for PATRIC. These 1100 were mapped onto pathways and results corresponded mostly to secondary metabolites biosynthesis, ubiquinone and other terpenoid-quinone biosynthesis.

Although the metabolite lists from PATRIC are similar, they are most comprehensive overview on metabolites that can be generated automatically. However this created lists may still contain metabolites that are not present in the respective species. Table 9 compares the metabolome sizes of the different *Pseudomonas* species. PATRIC is specialized to pathogens and therefore contains the better annotation than KEGG or BioCyc. The other possibility is that in KEGG or BioCyc assignments of the gene to its respective EC number are missing.

Table 9: Metabolome sizes of different *Pseudomonas* species

In silico metabolomes were produced from three different sources, whereby data in BioCyc was prestored and metabolomes for KEGG and PATRIC have to be predicted with own tools.

Organism	Metabolome size		
	PATRIC	KEGG	BioCyc
<i>Pseudomonas aeruginosa</i> PAO1	2352	1316	873
<i>Pseudomonas aeruginosa</i> UCBPP-PA14	2353	1342	1018
<i>Pseudomonas aeruginosa</i> PA7	2356	1381	1075
<i>Pseudomonas aeruginosa</i> LESB58	2355	1331	994
<i>Pseudomonas fluorescens</i> Pf-5	2473	1358	1044
<i>Pseudomonas fluorescens</i> Pf0-1	2333	1330	1042
<i>Pseudomonas fluorescens</i> SBW25	2372	1375	1102
<i>Pseudomonas putida</i> GB-1	2301	1304	1028
<i>Pseudomonas putida</i> F1	2344	1307	1060
<i>Pseudomonas putida</i> W619	2294	1297	1089
<i>Pseudomonas putida</i> KT2440	2280	1261	1041
<i>Pseudomonas syringae</i> pv. <i>syringae</i> B728a	2186	1284	1021
<i>Pseudomonas stutzeri</i> A1501	2141	1221	975

It should be mentioned that not only presence or absence of a metabolite is important, but also its amount. Additionally several metabolic pathways will be only active under certain conditions and possibly play only a minor role. The obtained data is used in chapter 3 to compare real life metabolomics data with the predictions for *P. aeruginosa* UCBPP-PA14.

2.3.2 Metabolic reconstruction of *Caenorhabditis elegans*

A reconstruction from genes stored in the KEGG database was generated as described above. Using the genome from *C. elegans* stored in KEGG 1363 compounds were found that

could be metabolized. Some reactions described in *C. elegans* are missing in KEGG, for example the glyoxylate cycle. The nematode carries one bifunctional enzyme having both activities (isocitrate lyase and malate synthase) needed for this reaction. In embryos the activity of this enzyme is high, because it is needed to break down storage lipids for energy production. A second example is the missing phosphoethanolamine methylation pathway in the glycerophospholipid metabolism pathway. *C. elegans* has two functional phosphoethanolamine methyltransferases. The first one PMT-1, catalyzes the methylation of phosphoethanolamine to N-methylphosphoethanolamine and the second, PMT-2, catalyzes the subsequent methylation to phosphocholine [102]. These are just two examples, where automated reconstructions need an expert judging on the obtained data. Furthermore the complete family of ascaroside signaling molecules are absent in the KEGG database. 154 metabolites were added from SMID database, to close this gap. Still not all biosynthetic routes for building blocks of the ascarosides are known, so still major gaps are present in the metabolite list for *C. elegans*.

2.3.3 Metabolic reconstruction of HeLa cells

For human metabolites the major resource is the human metabolome database (www.hmdb.ca). It includes chemical, clinical and molecular biology data for over 40,000 metabolites, including lipids [88, 103]. Several thousand entries have been measured and were detected with one or more methods at the Metabolomics Innovation Centre in Canada. Although, this database is comprehensive and useful tool, it includes not only endogenous, but also exogenous small molecules like drugs or food additives, which are commonly detected in human samples. Because in HeLa cells used in this work will not contain most of these metabolites, an automated draft reconstruction using KEGG was carried out. The reconstruction yielded 1902 compounds on all major pathways. Some metabolites present in this dataset are non-endogenous metabolites like chlorinated compounds or entries containing an R-group in the formula. These unphysiological compounds were removed from the KEGG list to obtain an overview on the HeLa cell metabolome. Totally 1383 compounds remain after filtering.

The Small Molecule Pathway Database (SMPDB) is a sub database of HMDB. Metabolites present in this database were included into the dataset [104]. SMPDB states that more than 2/3 of the stored pathways are not found in any other pathway databases. Therefore, metabolites found in this database were downloaded and unique metabolites should be included into the metabolic reconstruction. Totally 798 unique metabolites were found in the SMPDB. From this 682 overlap with the metabolites obtained from KEGG. The final list



contains 1499 unique metabolites. In HumanCyc, a second major resource of metabolic networks of humans, totally 1321 metabolites are stored for the human metabolism. Out of this 1321, 819 had a corresponding KEGG identifier and 580 are overlapping with metabolites from the KEGG and SMPDB reconstruction. Interestingly, the overlap between this two databases is surprisingly low. Recently, Stobbe et al. evaluated the overlap of metabolic network databases on human metabolism: Edinburgh Human Metabolic Network (EHMN), KEGG, Reactome, HumanCyc and BiGG. They found that the consensus for metabolites between the five databases is only 9% [105].

2.3.4 Improvements of reconstructions

Although metabolite databases are growing fast and reconstructions of sequenced organisms are improved on a regular basis, they only can cover a small part of real metabolomes. Using the MassTRIX server for annotation of ICR-FT/MS data only up to maximal 30% of the obtained signals can be annotated to metabolites. This shows that still major knowledge about metabolism is missing. In case of *Pseudomonas* species, different secondary metabolites, described in literature are missing. This data has to be extracted from primary literature and mapped to the respective organism. Additionally data from metabolomics experiments has to be evaluated to identify possible unknown secondary metabolites. An example is Myxobase, a database of secondary metabolites from myxobacteria at the Helmholtz Zentrum für Infektionsforschung in Braunschweig. This database contains information on more than 1000 samples. Unfortunately, this database is not public accessible.

In all cases harmonization of database content is needed. Unique and up-to-date identifiers for metabolites are needed. The IUPAC representation for chemical structures, the InChI key is machine readable and unique for each molecule, thus providing opportunity to be used in metabolomics.

2.3.5 Use of reconstructions in host-pathogen metabolomics

In silico predicted metabolomes can be applied to metabolomics datasets to determine the possible origin of a metabolite. Generally, metabolites from central metabolic pathways will be present in both, the host and the pathogen. For example precursors of branched chain amino acids will be present in bacterial samples, but not in *C. elegans* or human samples. If such metabolites are present they possibly derive from the ingested bacterial food or the infecting bacteria. Figure 18 shows a combined pathway map of *P. aeruginosa* UCBPP-PA14 (red) and *C. elegans* (green). Reactions present in both are marked in blue. To mark examples of errors present in KEGG, data obtained from this database was used for the comparison in Figure 18.



Although still gaps are present in the KEGG database for further work it was used due to possibility to generated colored maps of the obtained results. Where no results were found in KEGG, other databases were used.

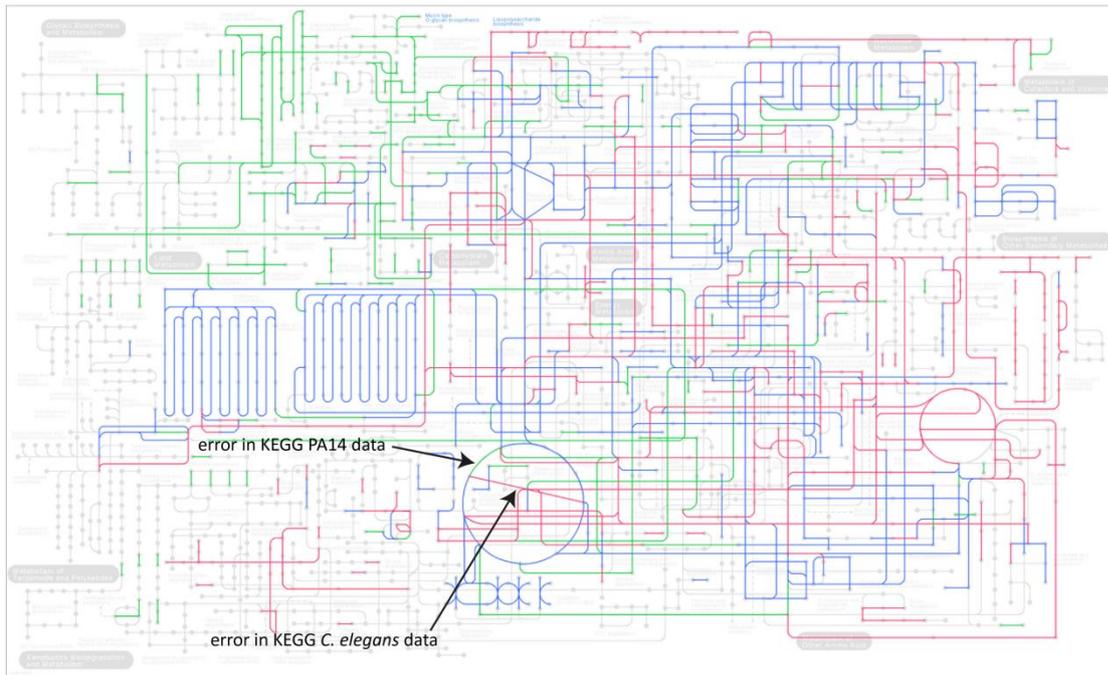


Figure 18: Combined metabolic pathway map obtained directly from KEGG for *P. aeruginosa* PA14 and *C. elegans* Arrows indicated two errors mentioned in the text. First the missing last reaction converting malate to oxaloacetate and closing the cycle in *P. aeruginosa* PA14 (red) and the missing Glyoxylate shunt in *C. elegans* (green)

The obtained *in silico* metabolomes can be compared against real life data to determine origin of a metabolite. Indeed, such comparisons are used quite often in this work in the following chapters. Additionally, chapter 3 compares different *Pseudomonads* grown in different medium based on their metabolomes to determine core and pan metabolomes of this genus.





3.1 Phylogenetics, bacterial identification and metabolomics

Phylogenetics is used to group bacteria and other organisms and their evolutionary relation. Phylogenetic trees are based on genetic data, like 16sRNA, in contrast to taxonomy which groups organism by their similar phenotype. For identification of bacteria in routine analysis, the so called multicolored row is used. This row uses different color reaction and visible observations to identify several metabolic pathways. The Kligler agar for example is used for the presence of lactases, fermentation and formation of H₂S. Other tests include the formation of indole from tryptophan, motility, presences of ornithine decarboxylase or the possibility to use Citrate as energy source. The obtained information is used to identify or at least group a bacterial isolate into a genus. A system offered by the BiOLOG company uses 71 carbon sources and 23 chemical sensitivity assays in parallel in a 96 well plate to identify bacteria. If a bacterium can grow under a certain condition it produces a color from a color reaction and this used as read-out for identification.

A newer method for identification of bacteria is the MALDI Biotyper[®], developed and distributed by Bruker. The instrument is based on a MALDI-ToF-MS to produce unique molecular fingerprints of an organism. Similarity between fingerprints measured and stored in a database allows the identification of different bacteria. This method is faster than the traditional methods and already widely used.

Metabolomics can be understood and used as future development of the multicolored row, not only testing single reaction, but testing for the whole metabolism. Comparison of different closely related yeast species based on their lipidome was recently carried out. The work showed that genetically closer species have similar patterns in their glycerophospholipids [106]. Although, very interesting this study did not target the central metabolism or major precursors or building blocks. In order to compare different species also abundance and usage of central metabolites play an important role. Furthermore, secondary metabolites are important in separating different species. One example is the in the introduction mentioned study from E. Frimmersdorf et al., which used GC-MS to elucidate how *P. aeruginosa* adapts to different environments. However only two strains were compared [60].

The comparison of the metabolome from the halophilic bacterium *Salinibacter ruber* using ICR-FT/MS was investigated by Rossello-Mora et al. to determine a geographical discrimination between different isolates. *S. ruber* can found in different parts of the world. Totally 28 isolates, 10 Mediterranean, 13 Atlantic and 5 Peruvian, were cultivated under same conditions and both supernatants and cell pellets were analyzed on a Bruker APEX Qe ICR-FT/MS with 12 T superconducting magnet and Apollo II ESI source. Multivariate statistical analysis revealed a

good separation of three different isolation regions. Furthermore, PLS-DA was able to separate the Mediterranean stains into their different origin locations. Several sulfonolipid species were shown to be differentially present, according to the region of isolation [107]. A closer look at two Mediterranean strains, M8 and M31 showed that 10% of the genes encoded in M8 are absent in M31. Moreover metabolomic analysis, phage susceptibility and competition experiments revealed that these differences are not neutral [108]. The most recent work focused on the response of these two strains to environmental changes. ICR-FT/MS, together with multivariate statistics, separated different growth states and assigned significantly different metabolites. For the stationary phase, for example, metabolites belonging to the aminosugar, glycerolipid and glycerophospholipid metabolism showed decrease or increase [109].

Table 10: *Pseudomonas* strains used in this work

Five different *Pseudomonas* species were selected. The table indicates the different used strains and their origin of isolation. Most of the species are clinical isolates. Type strains are indicated in bold.

<i>species</i>	<i>strain</i>	<i>origin</i>
<i>P. aeruginosa</i>	CCUG 551	type strain
	PA14	clinical, wound isolate
	CHA	clinical, CF isolate
	PA103	clinical
<i>P. fluorescens</i>	CCUG 1253	type strain
	CCUG 58749	clinical
	CCUG 61063	clinical
<i>P. putida</i>	CCUG 12690	type strain
	CCUG 2480 = ATCC 23973	soil
	CCUG 33664	plant
<i>P. syringae</i>	CCUG 14279	type strain
	CCUG 1826	plant
	CCUG 14280	plant
<i>P. stutzeri</i>	CCUG 11256 gv 1	clinical

Different strains and species have different metabolic capabilities as shown in chapter 2. These different capabilities are evaluated in two different ways. First, common metabolites, like amino acids, sugars or metabolites of the central carbon metabolism are compared between the different species to reveal how good each species can adapt to the given growth conditions. Second, chromatograms, especially results from the reversed phase separation are searched for precursor of secondary metabolites and the secondary metabolites themselves to find species specific metabolic features. The review article by Gross and Loper gives nice hints on the possibilities of secondary metabolite production in *P. aeruginosa*, *P. entomophila*, *P. fluorescens* and *P. syringae* [110]. Secondary metabolites include non-ribosomal peptides and

amino acid derived molecules, polyketides, lipopeptides and others. Only a small portion of these metabolites are stored in major metabolomics databases like KEGG or MetaCyc. To overcome this issue the Dictionary of Natural Products (DNP) was utilized. Free text search found 953 entries associated with the query “*Pseudomonas*”. Out of these 687 were metabolites with a unique sum formula were selected. By far not all secondary metabolites are stored in databases; most of them have to be curated from primary literature.

A major question is how a bacterium adapts to its surrounding environment. This is especially important for pathogens, like *P. aeruginosa*. To evaluate metabolic adaptation of different *Pseudomonas* species (Table 10) were compared them based on their metabolomes in different growth conditions. Bacteria were grown on either LB medium, representing a rich medium, or M9 buffer with glucose as sole carbon source. The selected strains range from clinical to environmental isolates, also isolated from extreme environments, like hydrothermal vents. Totally 13 different species and strains were compared. The obtained data was used to calculate phylogenetic trees based on metabolic phenotypes, afterwards referred as “phylometabolomic trees”.

3.2 Cultivation and extraction of bacteria

3.2.1 Bacterial cultivation and extraction

M9 and M9 with glucose were prepared from the stock solution depicted in Table 11. When all the solutions are cold, the stocks solutions are mixed in sterile water as followed. For 1000 ml M9 870 ml sterile water, 100 ml solution A, 10 ml solution B, 10 ml solution C, 10 ml solution D and for 1000 ml M9 with glucose mix 860 ml sterile water, 100 ml solution A, 10 ml solution B, 10 ml solution C, 710 ml solution D and 10 ml solution E were mixed.

Table 11: M9 stock solutions
All solutions were prepared freshly.

Solution	Preparation
<i>Solution A</i>	60 g Na_2HPO_4 , 30 g KH_2PO_4 , 5g NaCl , 10g NH_4Cl in 1l water; sterilize at 121°C/ 15min
<i>Solution B</i>	2.46 g $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$ (0.1M) in 100ml water; sterilize separately at 121°C/ 15min
<i>Solution C</i>	0.11 g CaCl_2 (0.01M) g in 100 ml water; sterilize separately at 121°C/ 15min
<i>Solution D</i>	0.47 g $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$ (0.017 M) in 100ml water; sterilize separately by filtration (Millipore 0.22 μm filter)
<i>Solution E</i>	20% of carbon source (glucose): 20g/100 ml water, sterilize separately at 121°C

For LB medium 5 g yeast extract were dissolved together with 10 g tryptone and 10 g NaCl in 900 ml ddH₂O and pH was adjusted to 7. ddH₂O was added to make up 1 L and the solution was autoclaved for 20 minutes at 121°C.

Different *Pseudomonas* strains were obtained from CNRS Marseille, Universidad Illes de Balears (UIB) and the Culture Collection University of Göteborg (CCUG). Lysogeny broth (LB) medium and M9 buffer with glucose were prepared as described in the literature. All *Pseudomonas aeruginosa* strains were grown in LB medium at 30 and 37°C and additionally in M9 with glucose at 30°C. All other strains were grown in LB medium and M9 with glucose at 30°C. Samples were collected at OD=1. From each sample biological duplicates were created. Bacterial pellets were collected by centrifugation, washed with M9 buffer and 2 ml of the supernatant of the original medium and the pellet were snap frozen in liquid nitrogen. All samples were stored at -80°C prior to extraction. Metabolites were extracted from the bacterial pellets with 50% MeOH. Concentration was adjusted to yield 1e9 cells / ml. After vigorous vortexing to suspend the pellet, the sample was sonicated for 20 minutes in an ice-cold sonic bath. After centrifugation at 4°C and 14,000 rpm the supernatant was transferred to an autosampler vial.

Data was obtained from reversed phase separation on BEH C8 column and positive mode ionization and HILIC separation on a BEH Amide column with positive and negative ionization. All measurements and data-preprocessing steps were carried out as described in the appendix.

3.3 Results

3.3.1 Statistical analysis

Statistical analysis was performed with Genedata Expressionist for Mass Spectrometry 7.6. Several PLS models were built to reveal species and condition specific metabolites. A metabolite was concerned as significant, if the VIP score of the corresponding model was bigger than 1 and the p-value < 0.05. All species grown in M9 buffer were compared against all grown in LB medium at 30°C. To reveal species specific patterns, always one species was compared against all others in the same growth condition. Lastly the effect of growth temperature was evaluated by comparing *P. aeruginosa* grown in LB medium at 30°C and 37°C.

3.3.2 Bacterial strains compared on a metabolomic level

HCA and PCA show that the obtained data is first clustered by the growth medium and temperature and afterwards by species. These results serve as basis for different PLS models. In a first attempt samples were compared on growth medium independent of single species.

Several amino acids were found in higher amounts in bacteria grown in LB medium. The four most significant were aspartate, leucine, threonine and valine, which are also in the top twenty of most significantly changed metabolites in the BEH Amide dataset with negative ionization. The amino acids are intracellular, because remaining LB medium was washed away with M9 buffer. Levels of threonine, leucine and valine are linked through the biosynthesis of valine, leucine and isoleucine pathway. However, amino acids are not the limiting factor of growth in LB medium, because they are readily available. It is especially useful to take up these amino acids directly from the medium instead of synthesizing them *de novo*. In contrast, bacteria grown in M9 medium have to synthesize these amino acids for protein synthesis. A closer look was taken to the biosynthetic precursors of leucine and valine. Unfortunately, most of the biosynthetic precursors can be also found on the respective degradation pathways. Most of these metabolites showed significant higher levels in bacteria grown in LB medium. An exception from this is (S)-2-acetolactate, which is only found on the biosynthetic route to valine and leucine. It is significantly higher in M9 grown bacteria (p-value = $7.87e-10$). The enzyme acetolactate synthase is highly feedback regulated by branched chain amino acids. This suggests that in M9 grown bacteria the biosynthetic route is active, whereas in LB medium grown bacteria utilize excessive branched chain amino acids. Furthermore in LB grown *Pseudomonads* increased levels of metabolites belonging to the arginine metabolism were found. N-acetyl-ornithine, ornithine, aspartate, argininosuccinate and arginine were significantly increased, suggesting that excessive nitrogen is bound to arginine. Lastly arginine and ornithine are needed for synthesis of the siderophore pyoverdine.

In contrast to this, pyochelin, another iron siderophore was found in higher levels in bacteria grown in M9 buffer with glucose. From the structural point of view, pyochelin is the simpler siderophore. Bacteria grown in M9 buffer possibly prefer pyochelin over pyoverdine for iron acquisition, to save valuable amino acids. Interestingly, *P. fluorescens* CHA0 was shown to produce enantio-pyochelin, the optical antipode of pyochelin [111]. Because no chiral separation was employed it could not be determined if the detected peak corresponds to pyochelin or its optical isomer.

Further examples of metabolites found higher in M9 grown bacteria are D-glucose-6 sulfate (p-value = 0.004) or nicotinurate (p-value = $4.80e-07$). Several more markers for either LB medium or M9 buffer with glucose grown bacteria were found, but a big proportion of them had no putative annotation.



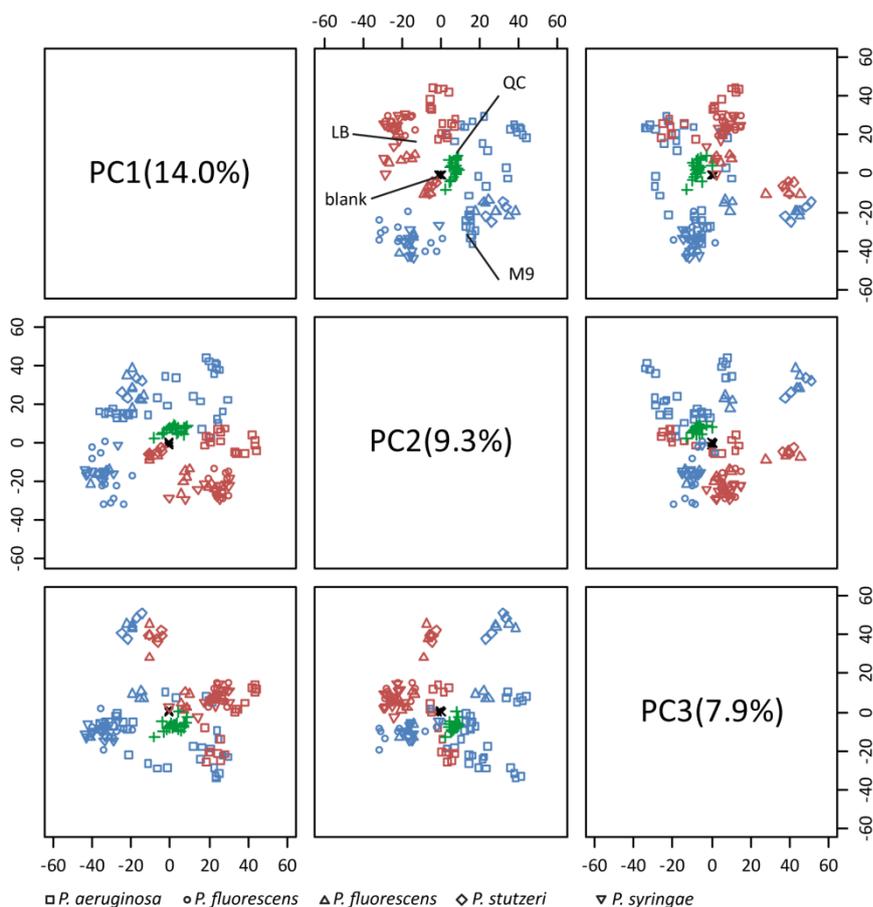


Figure 19: Scores plots of Principal component analysis of BEH C8 separation

After filtering, data was processed with principal component analysis (PCA). The first two principal components were able to roughly separate the bacterial pellets by the used growth medium.

Using Metlin several unknown masses found higher in *Pseudomonads* derived from LB medium, could be annotated as peptides, possibly ingested from the rich culture medium. Because in the used methods, metabolites from the central carbon metabolism like citrate are only rarely detected, no information on the different metabolic fluxes between bacteria grown in rich LB medium or minimal M9 buffer can be given. Results like the degradation of branched chain amino acids above show that LB medium grown bacteria directly utilize uptaken nutrients, whereas bacteria grown in M9 have to synthesize all metabolites *de novo*. This suggests that growing bacteria in M9 generally gives a good overview on metabolomic lineages of different species concerning the core metabolism. In contrast, LB medium grown ones give information about species relationship on secondary metabolism.

3.3.3 Different growth temperatures influence the metabolome

P. aeruginosa strains grown in LB medium at two different temperatures were used to evaluate the influence of growth temperature on the cellular metabolome. PCA analysis showed a good separation of the two temperatures along the first principal component. A PLS

model was used to identify masses contributing to the group separation. Several masses having a VIP score > 1 and p -values < 0.05 were identified.

Bacteria grown at 30°C show higher levels of tryptophan metabolites, like anthranilate (p -value = 0.001), indole-3-acetate (p -value = 0.004), indolpyruvate (p -value = $7.2e-06$) and kynurenine (p -value = 0.03). No accumulation of tryptophan in bacteria grown at 37°C was observed. Still, results suggest that tryptophan metabolism is only active at lower temperatures.

Interestingly, intracellular pyochelin showed also a temperature-dependency, with higher levels at 30°C. Again several unknowns were found to be different along the two used temperatures with low p -values, but no clear evidence for redirection of metabolic fluxes except for tryptophan were found. Possibly the temperature difference is too small. Going to deeper or higher temperature will induce a more different metabolism. Also growth in M9 buffer can give better insights on the changed metabolism upon different temperatures.

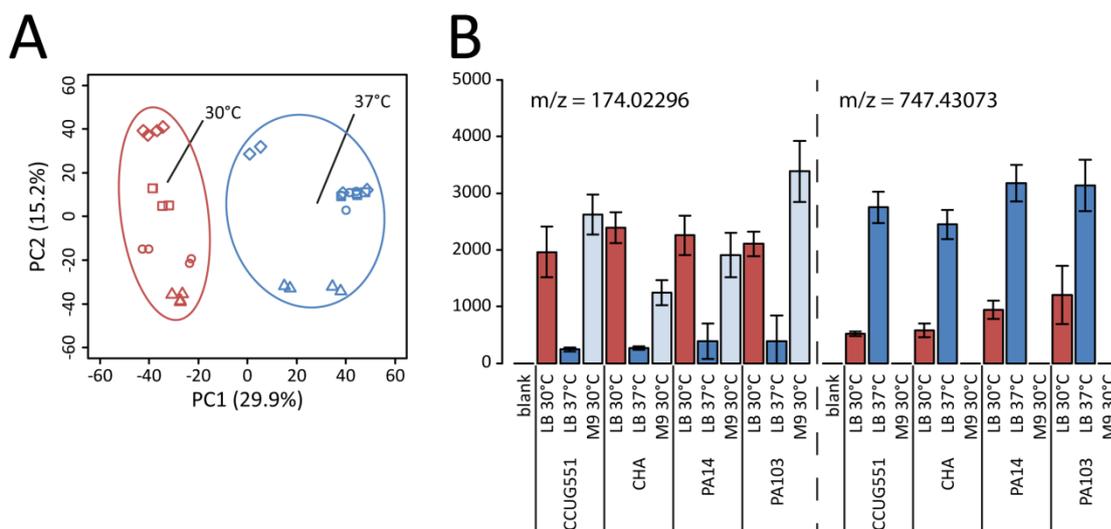


Figure 20: Different growth temperatures lead to differences in the intracellular metabolome

(A) Scores plot of principal component analysis of BEH C8 data. The first principal component was able to separate bacteria grown at different temperatures, whereas the second separated different strains of *P. aeruginosa* from each other. (B) Examples of two masses significant for bacteria grown at 30°C or 37°C.

Lastly, a lipidomics approach can tell if lipid composition changes with temperature. This could be of particular interest to understand how bacteria are adapting to changing environments and how thermophile bacteria can survive elevated temperatures.

3.3.4 Core and pan metabolome of the genus *Pseudomonas*

In order to reveal species specific metabolite patterns, data was split into single datasets according to growth medium. PCA was used to reveal group separations based on the

metabolomes. Figure 21 shows the first three principal components obtained from the BEH C8 separation with positive mode ionization from the LB medium dataset. The first two principal components nicely separate *P. aeruginosa* from all other species. The additional third component showed a cluster of *P. aeruginosa* samples that are different from the remaining. This cluster corresponds to the strain PA14. Furthermore a mixed cluster of *P. putida* CCUG12690 and *P. stutzeri* CCUG11256gv1 was revealed.

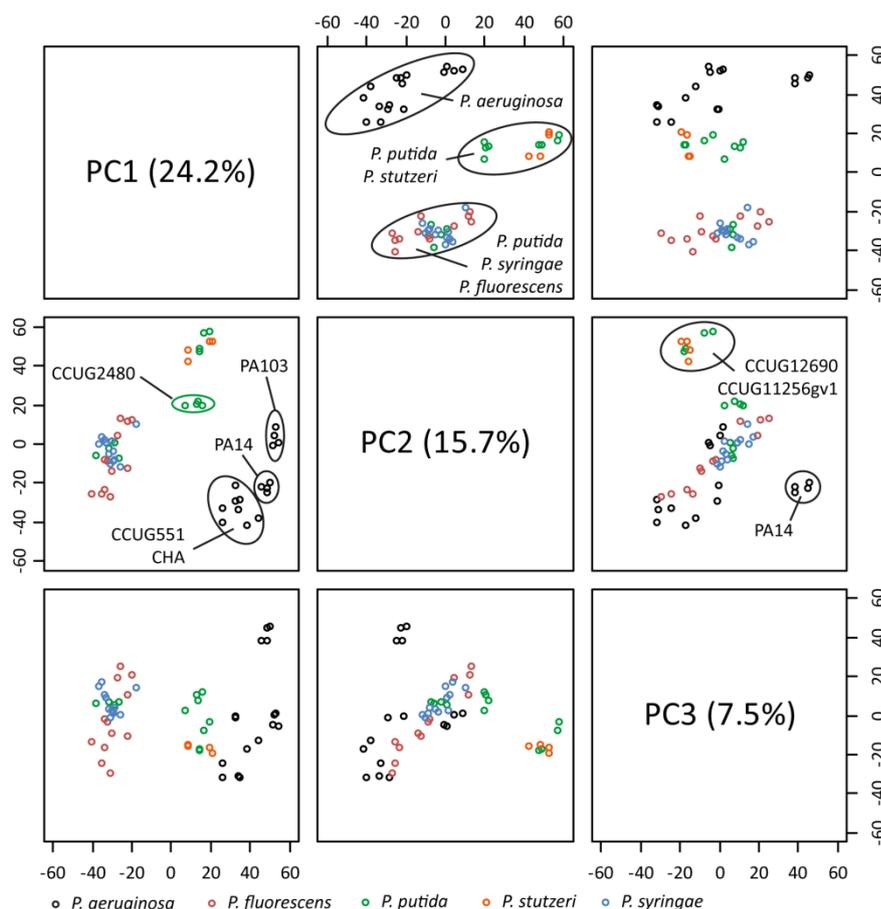


Figure 21: PCA scores plots from BEH C8 separation of bacterial pellets grown in LB medium

In the first two principal components, three different clusters are visible. The first one corresponds only to *P. aeruginosa* strains, whereas the second is a mixture of *P. putida* and *P. stutzeri* and the third a mixture of *P. putida*, *P. syringae* and *P. fluorescens*. The second and third principal component show that *P. putida* CCUG12690, *P. stutzeri* CCUG11256gv1 and *P. aeruginosa* PA14 are different from all other strains.

Overall separation was good enough to build PLS models. Within a single dataset always one species was compared against the remaining. Although several masses could be annotated to a putative metabolite, most of the specific metabolites have no possible annotation using MassTRIX. Central carbon metabolites were shown to be different between bacteria grown in M9 buffer with glucose or LB medium. However, some amino acids and other metabolites showed higher levels in some species, suggesting a better adaptation to the surrounding environment. For example *P. fluorescens* showed in M9 buffer higher levels of aspartate (p-value = 0.04) and glutamate (p-value = 0.0004). Both amino acids are directly deriving from

acids present in the TCA cycle. Levels of glutamine remained unchanged. Aspartate is an important precursor for the synthesis of lysine, methionine, threonine and isoleucine. O-succinylhomoserine (p-value = 0.0005), a further precursor of methionine was found higher in *P. fluorescens*, together with threonine (p-value = 0.002) and leucine (p-value = 0.02). Analysis of metabolic capabilities carried out according to Kastenmüller et al. showed higher capabilities in methionine synthesis for *P. fluorescens* Pf-5. This suggests that *P. fluorescens* is well adapted for growth in minimal medium.

In contrast to this, in LB medium grown bacteria most of the metabolites significant for a species are belonging to unknowns of the higher mass range with masses > 500 Da. These masses are possibly belonging to secondary metabolites or alginate precursors for biofilm formation. A metabolite separating *P. putida* from all other species is intracellular pyocheline. Figure 22 shows the different levels of intracellular pyocheline and 2,4-diacetylphloroglucinol (DAPG). DAPG is a known secondary metabolite from *P. fluorescens* with antibiotic properties and major determinant in biological control of plant diseases [112].

A metabolome based tree was constructed based on the obtained data, for each method. Intensities of masses occurring in more than 2 samples per condition were averaged and data was split into a LB medium and M9 medium dataset. Based on the averaged intensities distance between the single species was calculated and a tree was plotted based on the calculated distances to show functional relationships between the different species. Most metabolites from the intermediary metabolism, like amino acid precursors were detected in BEH Amide separation with negative ionization. Together with bacteria grown in M9 buffer this method seems to be suitable to derive a tree based on the central metabolic pathways. The tree derived from this method showed a good separation of *P. aeruginosa* from all other species. Also in the prediction of metabolic capabilities *P. aeruginosa* showed a subcluster different from all other. However the subcluster of *P. putida* species could not be found in the tree. Interestingly on a metabolomic level the type strains of *P. stutzeri* and *P. putida* (*P. stutzeri* CCUG11256gv1 and *P. putida* CCUG12690), cluster together, away from all other strains.

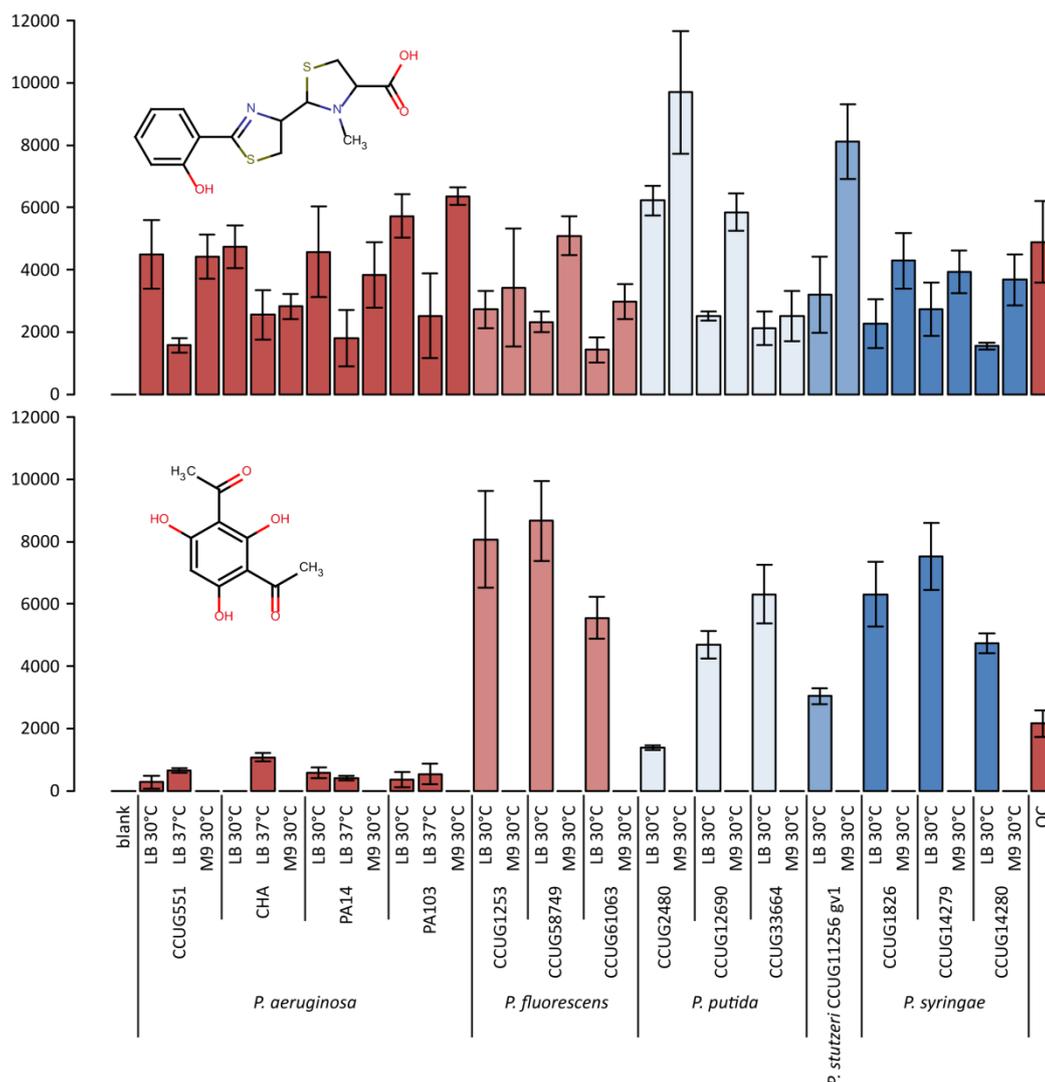


Figure 22: Levels of intracellular pyochelin and 2,4-diacetylphloroglucinol

Pyochelin is found in higher amounts in bacteria grown in M9 medium, whereas *P. aeruginosa* seems to be an exception, with similar or lower levels compared to LB medium. 2,4-diacetylphloroglucinol is a secondary metabolite found in *P. fluorescens*. It was detected in LB medium grown bacteria, with highest levels in *P. fluorescens*, followed by *P. syringae* and *P. putida*.

Compared to the 16sRNA lineage, the metabolome derived lineage is substantially different. Ribosomal RNA was probably one of the first parts of early life on earth and is present in cells of all organisms. This makes it a good marker for evolution of species. A metabolome based tree present more of a functional rather than an evolutionary relationship. Although the metabolome coverage of the used method is incomplete, they can provide first idea; how a metabolomics based taxonomy approach can look like. Methods that cover more parts of the central carbon metabolomics can give a better insight on adaptation of different species to different growth conditions. Additionally it was shown that not only ICR-FT/MS metabolic profiling as used by Rosella-Mora et al., but also UHPLC-UHR-ToF-MS is able to separate different species and strains from each other.

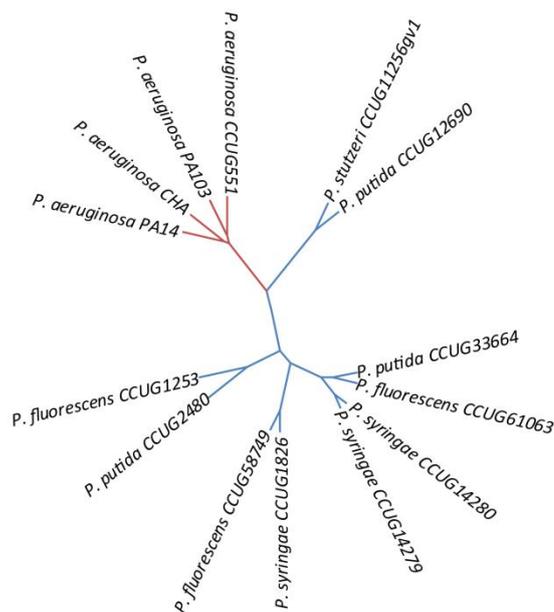


Figure 23: Distances of different *Pseudomonas* species based on metabolome

Data from BEH Amide separation with negative ionization was used to compute distances between the different species. All measured species from *P. aeruginosa* clustered together, whereas no clear clusters were observed for all other species

A further approach that could be possibly tested in future is to use lipidomics. A widely used method in microbiology for typing bacterial cultures is the use of fatty acid profiles obtained by GC analysis. Lipidomics approaches could go a step beyond and also reveals how these fatty acids are used in different lipid species. The above mentioned article from Hein et al. using different yeast strains proofed this concept.

3.3.5 Comparison against *in silico* metabolomes

Annotated metabolites from the different methods were used for comparison against the *in silico* metabolomes obtained from PATRIC, described in Chapter 2. A list of unique KEGG identifiers of annotated metabolites from all methods was created, which included 5596 unique metabolites. The theoretical list from PATRIC contained 2353 metabolites. Overlap between both was 818. From these 818, 383 could be mapped on the main metabolic pathway map of *P. aeruginosa* PA14 and included amino acids, some central carbon metabolites, several breakdown products and precursor metabolites and secondary metabolites. However, most detected metabolites could not be mapped to a pathway and represent possibly secondary metabolites. On the side of the prediction, 1535 metabolites could not be found in any method, including metabolites from the TCA cycle, Pentose phosphate pathway and energy metabolites like ATP. Furthermore the prediction included metabolites not amenable to mass spectrometry, like acetate, sulfate or other inorganic ions. From the remaining detected metabolites a big proportion could be mapped to the KEGG secondary metabolite

pathway, but in branches not present in *P. aeruginosa*. These metabolites represent an interesting target for identification and structural characterization, because they are possibly novel secondary metabolites.

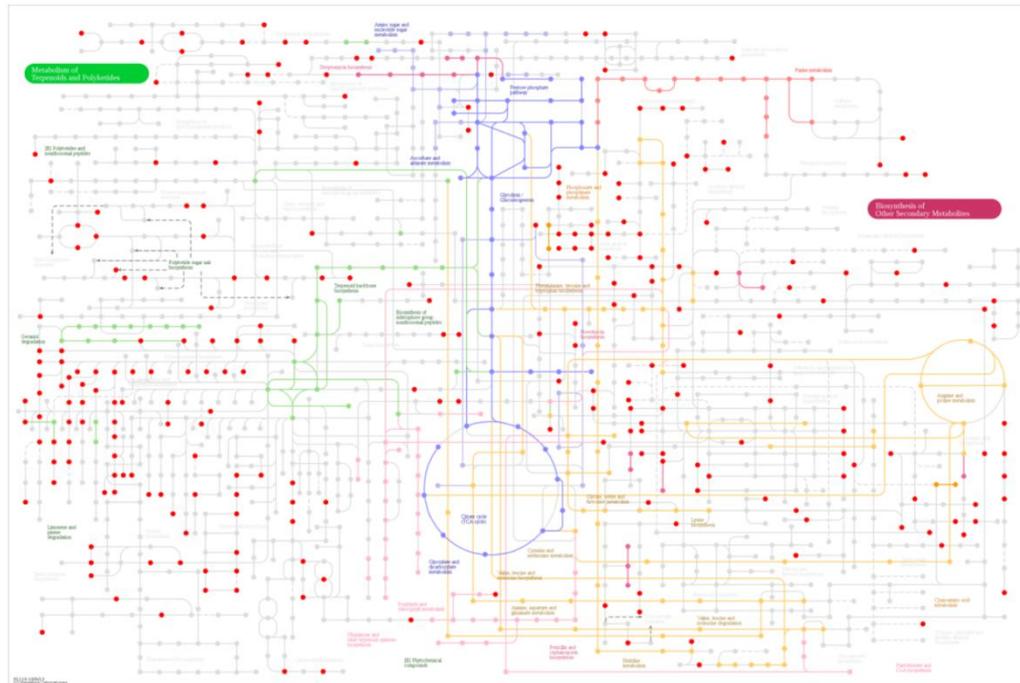


Figure 24: Metabolites found in measurements, but not in prediction

Metabolites that were not found in the prediction could be mapped mainly onto secondary metabolite pathways.



4.1 *C. elegans* infection model

There remains a need for robust and simple animal models of infection to allow for better understanding of host-microbe interactions, bacterial pathogenesis as well as development of antimicrobial therapies. An attractive model is the invertebrate nematode *C. elegans*, now well recognized as relevant for the study of bacterial pathogenesis as well as a model for other conditions. The worm *C. elegans* is susceptible to a number of bacterial pathogens, which are able to kill or induce a range of symptoms of disease. Human pathogens already known to affect *C. elegans* include Gram-negative bacteria such as, *Pseudomonas* and *Salmonella* as well as a number of Gram-positive bacteria such as *Enterococcus* and *Staphylococcus*. Many of these are able to colonize the worm intestine, and the pathogenic effect can be measured by marking the decrease in lifespan of the nematode. Several studies have already been undertaken in which mutant libraries of a bacterial pathogen have been tested in the worm to identify candidate virulence genes [113]. *C. elegans* also offers researchers a simpler model for analysis of the host side of pathogen-host interactions. Genetic analysis of the host is relatively straight forward, for example by screening for worm mutants either more resistant or hypersensitive to a given pathogen.

To prove that non-targeted metabolomics is a useful tool in host-pathogen research a set of experiments using the *C. elegans* strains *fer-15* and *daf-2* cultivated either on bacterial lawns of *E. coli* OP50, *P. aeruginosa* PA14wt, *P. aeruginosa* PA14Δ*gacA*, *Salmonella enterica subsp. enterica* or without food for 24 hours to exploit the effect of *P. aeruginosa* infection on the host metabolome of *C. elegans* was performed. *P. aeruginosa* PA14Δ*gacA* is highly attenuated in killing *C. elegans* under slow killing conditions (Figure 36). *gacA* encodes for the global activator protein A (GacA), the response regulator of the GacA/GacS two-component regulatory system. It has been demonstrated that GacA positively regulates the production of several virulence factor in *P. aeruginosa* and is needed for full virulence in mammals and *C. elegans* [114]. Furthermore, infection with *S. enterica subsp. enterica* leads to a different gene expression pattern in the host compared with *P. aeruginosa*, leading to the idea, that also the metabolome will be affected differentially. Mutants of *C. elegans* lacking *daf-2*, the only homolog of a mammalian insulin/IGF-1-family receptor in the genome, were indicated to be resistant to bacterial pathogens [115]. *P. aeruginosa* is able to suppress host immunity by activating the *daf-2* signaling pathway, which also regulates metabolism [68, 116]. Starvation was chosen as second kind of stress, other than infection, to put the metabolome in a more general picture. An additional sample set containing *daf-2* and *fer-15* worms fed with *P. aeruginosa* PA14wt, *P. aeruginosa* PA14Δ*gacA* and *P. aeruginosa* PA14Δ*pec7* was included for lipidome analysis.

Metabolic profiling on these samples was carried out using DI-ICR-FT/MS and UHPLC-UHR-ToF-MS. The used methods included reversed phase and HILIC separation for metabolite extracts and a reversed phase separation of lipids and are described in the appendix. This comprehensive data collection was used for development and application of data analysis tools for filtering of false positive annotations, based on chemo- and bioinformatics approaches, which are presented in the first part of this chapter.

Retention time modeling for false positive filtering of metabolite annotation and lipid separation was subjected to data dependent acquisition of MS/MS spectra together with *in silico* fragmentation for lipid identification. Mass difference networking was used to obtain novel putative formulas for unknown masses in the ICR-FT/MS datasets. Finally UHPLC-UHR-ToF-MS and DI-ICR-FT/MS datasets were aligned, to combine exact mass and retention time information.

All results were combined in a comprehensive metabolite map, which was used for biological interpretation of the infection model presented in the second part of this chapter.

4.2 Known and unknowns – False positive filtering of metabolite annotations

A major problem in metabolomics using mass spectrometry is the annotation and identification of masses. With currently used databases and approaches only up to 30% of all masses can be annotated with a putative metabolite. The remaining 70% may correspond to adducts not covered by the used data analysis tool, homo- or heterodimers, fragments or novel metabolites. Especially in direct infusion mass spectrometry identification of fragments is a difficult task, due to missing chromatographic separation. In this situation, masses cannot be assigned as fragment to one single mass. Chromatographic separation greatly enhances metabolite annotation and identification. First reduced ion suppression enhances sensitivity for low abundant molecules. Second, chromatographic retention gives information about different physicochemical properties of the detected masses and allows filtering of possible false positive annotations. In reversed phase separations, molecules are eluting according to the hydrophobicity, starting with the most hydrophilic substances in the mixture. If, for example, carnitine, a small tertiary amine is detected at higher retention times, this may be due to wrong annotation of the mass or it corresponds to a fragment of a more hydrophobic molecule, e.g. a long-chained Acyl-carnitine. To improve identification, one possible approach is to measure retention time of single metabolites as pure chemical standard to obtain information about correct retention time and possible fragment and adducts. This is virtually impossible due to several reasons. Not all metabolites that might be present in a sample can

be obtained as pure chemical standard. Moreover buying all this standards consumes a lot of money and would be not feasible for most metabolomics laboratories. To overcome limitations several approaches are used. First MS/MS of a mass can help to identify a metabolite or at least to reduce the number of possible candidates. For final identification the fragmentation pattern of the metabolite detected in the sample should be compared to a chemical standard. If this is not possible, *in silico* calculation of theoretical fragmentation can help. Several fragmentation rules of organic compounds are known and can be used for such *in silico* approaches. This is implemented in commercial products like ACD/MS Fragmenter or Mass Frontier [117, 118]. However calculations can take up to several hours for complex molecules. A promising tool to enable fast calculations is the bond disconnection algorithm implemented in MetFrag [85]. This algorithm not always produces mechanistically correct fragments, but most fragments produced can be found in tandem mass spectra. MetFrag is implemented as webserver (<http://msbi.ipb-halle.de/MetFrag/>) or can be downloaded as command-line tool. Furthermore the complete source code, written in Java, can be downloaded and used under GNU license.

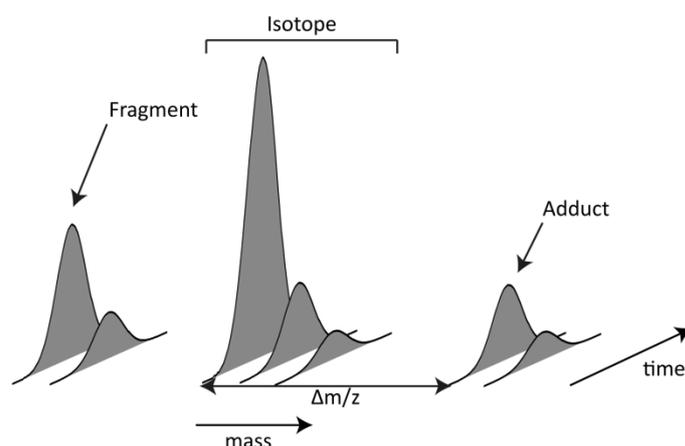


Figure 25 Complex mass spectrometric patterns can arise from a single metabolite species, including fragments, isotopes and adducts, which complicates data analysis.

Adapted with permission from [119]. A single metabolite not always produces a single signal. First different isotopic species will be also detected. Second, the ionization process can not only produce a single pseudo molecular ion, but also other adducts. One of the most detected adduct is the $[M+Na]^+$ adduct in positive ionization mode. Finally, under certain settings in-source fragmentation can occur.

Another method to filter false positive assignment is the calculation of a quantitative structure retention relationship (QSRR). Physicochemical properties of molecules are used to predict retention time. This method exists for a long time for GC-MS, based on different retention indices. Kumari et al. used *in silico* derivatisation and retention time prediction to identify unknown substances measured on exact mass CI-GC-MS [120]. Recently QSRR was adapted by Creek et al. for removal of false positive assignment of metabolite annotation. The model is based on the measurement of 120 authentic standard substances on a ZIC-HILIC

column with positive/negative ionization switching Orbitrap-MS. The retention time model was able to filter out 40% of false positive annotations [121].

The Metabolomics Standards Initiative published several levels of confidence of identity, which serve as guideline of metabolite identification. The different levels are shown in Table 12. Bowen and Northen used in their work the term metabolite atlas to describe a matrix of several experiments using the same method. This metabolite atlas can be used to reveal common artifacts like adducts or fragments across all biological samples [119]. Using global correlation analysis such features can be identified and separated from biologic entities. Furthermore mass difference can be used to compute possible neutral losses or building blocks [122].

Table 12: The four levels of metabolite identification confidence defined by the Metabolomics Standards Initiative (Sumner et al. 2007)

Taken from [123]

Level	Confidence of identity	Level of evidence
1	<i>Confidently identified compounds</i>	<i>Comparison of two or more orthogonal properties with an authentic chemical standard analyzed under identical analytical conditions</i>
2	<i>Putatively annotated compounds</i>	<i>Based upon physicochemical properties and/or spectral similarity with public/commercial spectral libraries, without reference to authentic chemical standards</i>
3	<i>Putatively annotated compound classes</i>	<i>Based upon characteristic physicochemical properties of a chemical class of compounds, or by spectral similarity to known compounds of a chemical class</i>
4	<i>Unknown compounds</i>	<i>Although unidentified and unclassified, these metabolites can still be differentiated and quantified based upon spectral data</i>

For more accurate annotation and identification of metabolites and lipids of the in this work used *C. elegans* infection model different approaches were employed. Quantitative structure retention relationship models (QSRR) were built for the chromatographic separations used and lipid identification was achieved by using a data-dependent acquisition approach based on the AutoMSⁿ algorithm of Bruker.

4.2.1 Prediction of chemical properties to build retention time models

Identification of metabolites is a key issue in metabolomics. To unambiguously identify a metabolite, reference standard compounds are needed for comparison. This is virtually impossible, because not all metabolites are available as pure standard. Even if standards for all metabolites would be available, for most laboratories it is not feasible to run standards of all metabolites with each chromatographic sample batch due to tremendous costs. One way to improve the identification is MS/MS during chromatography and subsequent comparison against tandem MS databases. This works well for pure peaks of sufficient length to obtain good MS/MS spectra. If chromatographic peaks are overlapping problems can arise, that the mass underlying the first peak is still fragmented and the second peak will be overseen. Summing up this approach is more useful for targeted metabolomics, where metabolites that should be profiled are already known and reference standards exist.

A second approach offers bigger flexibility in annotation of metabolites. Beside the mass and/or fragmentation patterns information on physicochemical parameters of a metabolite are used to predict retention behavior. The work of Creek et al. brought this approach from pharmaceutical application to metabolomics. In their publication they modeled the retention behavior on a ZIC-HILIC column with positive/negative polarity switching Orbitrap-MS. This work was used as a basis for the calculation of a QSRR model for the metabolite cartography of *C. elegans*. The used analytical platform consisted of UHPLC-UHR-ToF-MS measurements on a BEH C8 and a BEH Amide column to increase the metabolome coverage with the usage of these orthogonal methods. Both methods offer different selectivity and allow a comprehensive snapshot of the *C. elegans* metabolome, whereas BEH Amide separation is well suited for central metabolites like amino acids and other and BEH C8 separates more lipophilic substances. The reversed phase method was able to separate over 20 different acylcarnitines, whereas chromatography on the BEH Amide column could resolve L-Isoleucine and L-Leucine. Several compounds were tested on these methods. Problems arise specifically from polyacids like Citrate or phosphorylated compounds like ATP, which are interacting strongly with the stationary phase and give distorted peaks or no signal. Overall, the metabolome coverage was good, with annotations on all major pathways. To filter false positive annotations a QSRR model was calculated.

4.2.1.1 Characteristics of the *C. elegans in silico* metabolome

To get an overview on the physicochemical properties of the *C. elegans* metabolome, several parameters were predicted for the *in silico* metabolome described in chapter 2. This list included 1363 metabolites from KEGG and 154 secondary metabolites from SMID database

For prediction of several parameters SMILES strings were imported to MS Excel from a group internal database. Using JChem for Excel these SMILES were converted to structures and standardized using the Standardizer Plugin. From this standardized structures different physicochemical parameters were predicted. All calculations were carried out in MS Excel with JChem for Excel.

Following parameter comparably to Peironcelly et al. were calculated: count of C, H, O, N, S and P atoms, heavy atom count, molecular weight, logP, logD at pH 7.2 (approximate pH of cytoplasm), number of H acceptors and donors, number of rings, number of rings with heteroatoms, rotatable bonds and number of aromatic rings [124]. logP and logD values give information about the hydrophobicity of a molecule. The logP value accounts the octanol/water partition coefficient for a unionized molecule, whereas the logD give the partition coefficient for ionization at a certain pH. Equations for the definition of both values are given in the equations 7 and 8.

$$\log P_{\frac{oct}{wat}} = \log \left(\frac{[solute]_{octanol}}{[solute]_{water}^{un-ionized}} \right)$$

Equation 7: Definition of logP

The octanol/water partition coefficient is a measure of the hydrophobicity of a molecule

$$\log D_{\frac{oct}{water}} = \log \left(\frac{[solute]_{octanol}^{ionized} + [solute]_{octanol}^{neutral}}{[solute]_{water}^{ionized} + [solute]_{water}^{neutral}} \right)$$

Equation 8: Definition of logD

The logD value takes also the micro species of a molecule at a certain pH into account.

The logP for metabolites from the *C. elegans* metabolic reconstruction are ranging between -15 and 15. Figure 26 shows a density plot of the logP and logD values. ATP for examples, a very polar metabolite and universal energy currency of cells shows a logP value of -6.2. The opposite range can be represented by Cholesterol with a value of 7.11. A second important value for the description of a metabolome *in silico*, especially when MS should be used for analysis, is the molecular weight. Most of the *C. elegans* metabolites have a mass below 500 Da, as shown in Figure 27A. Different lipid species are not included in the analysis and present a special case, discussed in a later part of this chapter. With different carbon chain length and combination of different fatty acid moieties in PC's, PE's, TG and other lipid classes several thousand to ten thousands of lipids would be possible, especially as *C. elegans* is able to synthesis branched chain fatty acids *de novo*.

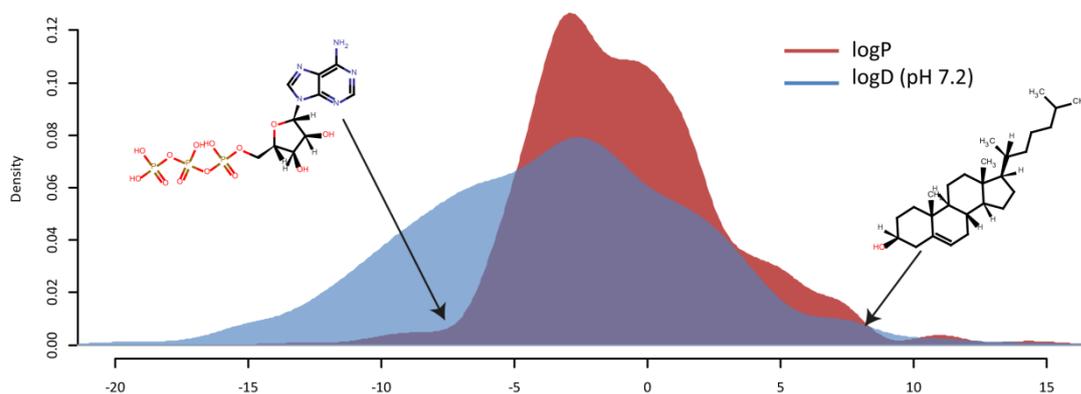


Figure 26: logP and logD values of the *C. elegans in silico* metabolome

logP values are ranging from -15 to 15, whereas logD values show a shift towards lower values., which is reasonable, because most metabolites have an ionizable group.

The number of carbon atom for metabolites ranges between 2 and over 50, where most metabolites have 30 or lower. Interestingly the metabolites show low counts of nitrogen and sulfur, but a proportion of them showed higher counts on phosphor and oxygen. Peroceilyn et al. compared the HMDB against ZINC, a database for non-metabolites, finding that the most metabolites have lower counts in nitrogen and sulfur, but high proportion of phosphor and oxygen. Second, the number of rings and rotatable bonds was lower for metabolites compared to non-metabolites, showing that metabolites tend to have simpler structures. A maximum ring count of eleven was found in the *C. elegans* metabolite dataset. The most complex aromatic ring systems were found in FAD and heme. Also a lower number of rotatable bonds were found. Metabolites tend to have a more rigid structure compared to pharmaceuticals. Exceptions from this are fatty acids and aliphatic molecules like lipids.

Figure 26 shows the shift of hydrophobicity of *C. elegans* metabolites towards hydrophilic part of the diagram, when using the logP instead of the logD value. From this it can be concluded that the logD value is much more valuable for metabolite prediction and building of a structure retention time model, because most metabolites have ionizable groups. The different parameters have also a high relevance for mass spectrometry, because they influence the ionization in an ion source. Molecules like Cholesterol with no ionizable group will be hard to ionize under ESI conditions and can possibly only be observed as $[M+Na]^+$ adduct, whereas it should be readily ionized using APCI or APPI. A high content of nitrogen will make a molecule amenable to positive ionization mode. Contrary, metabolites containing a phosphate group have a high prevalence to be detected under negative ionization conditions. Other parameter like the number of rotatable bonds will influence the chromatographic behavior.

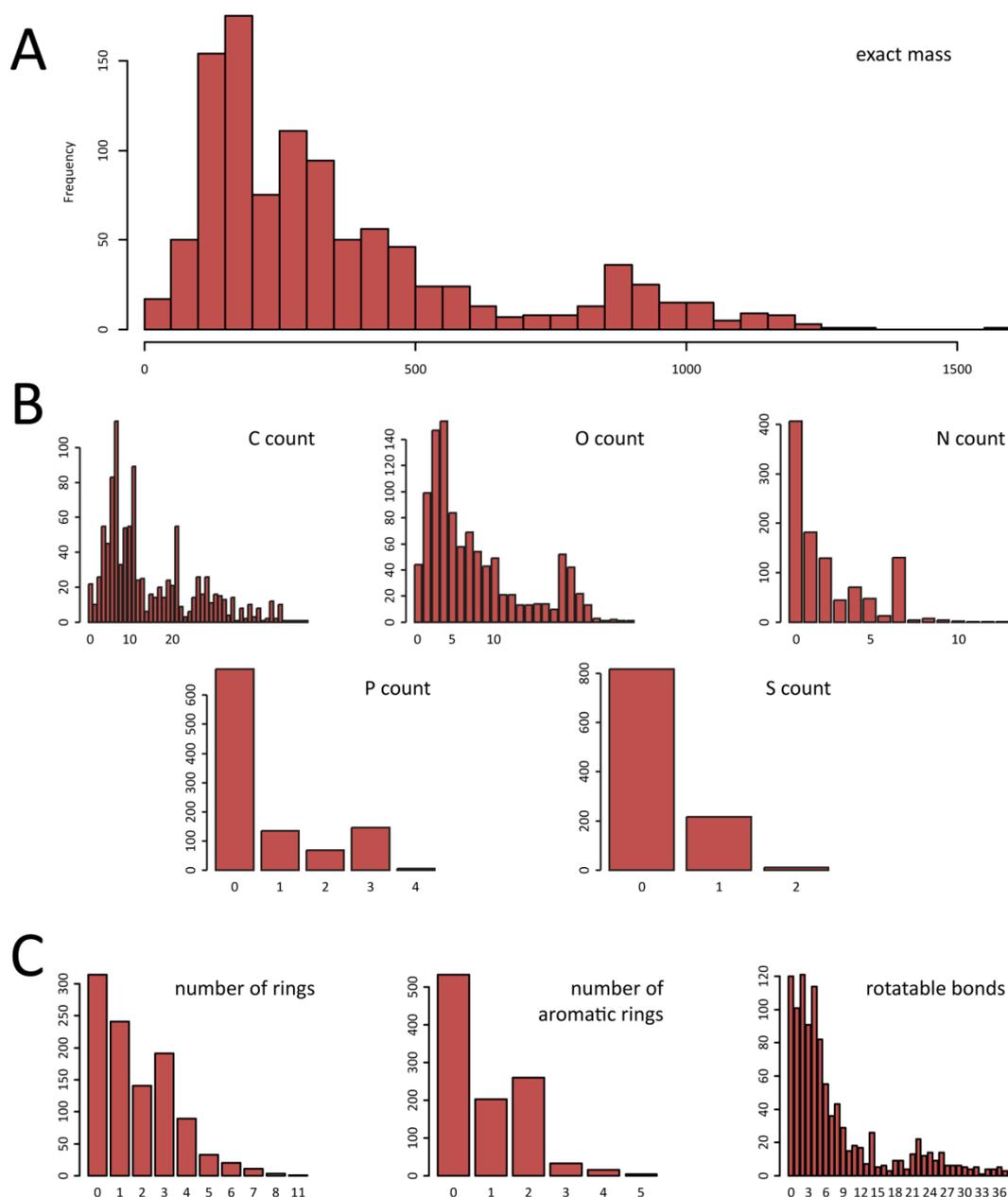


Figure 27: Physicochemical properties of the *C. elegans in silico* metabolome

(A) Masses of metabolites are ranging until 1500. Most metabolites have a mass below 500 Da. (B) Atomic counts show that the majority of the metabolites have lower than 15 carbon atoms and between 0 and 10 oxygen atoms. The second heteroatom found is nitrogen followed by phosphor and sulfur. (C) Also rings and aromatic rings play together with the number of rotatable bond an important role. Especially nucleosides and related molecules have several rings. Normally metabolites are very rigid with a low number of free rotatable bonds. Exceptions from this are fatty acids, which usually have higher amounts of such bonds.

To reveal building blocks of the metabolome, the maximum common substructure (MCS) of the *C. elegans* metabolome was searched using the Library MCS tool from ChemAxon. Figure 28A shows the most common substructures in the used dataset. It included phosphate groups, the chiral center of amino acids, dicarboxylic acids and β -keto acids for example. To further evaluate building blocks the FRAG57 library from Nobeli et al. was used to search for other building blocks, they used for the structural anatomy of the *E. coli* metabolome [83]. The most

becomes more dissimilar. This information can be used for identification of novel biosynthetic routes. If a novel metabolite is identified with its chemical structure, calculation of similarity to known metabolites can possibly reveal biosynthetic precursor and building blocks. In *C. elegans* Ascaroside signaling molecules are made up of building from different metabolic pathways. For example the indole-3-carbonyl unit in indole containing ascarosides may derive from tryptophan, while the p-hydroxybenzoic acid residue derives from tyrosine metabolism. However this approach can be only used for molecules with an appropriate size. In bigger molecules for example heme or related species it can be hard to detect precursors.

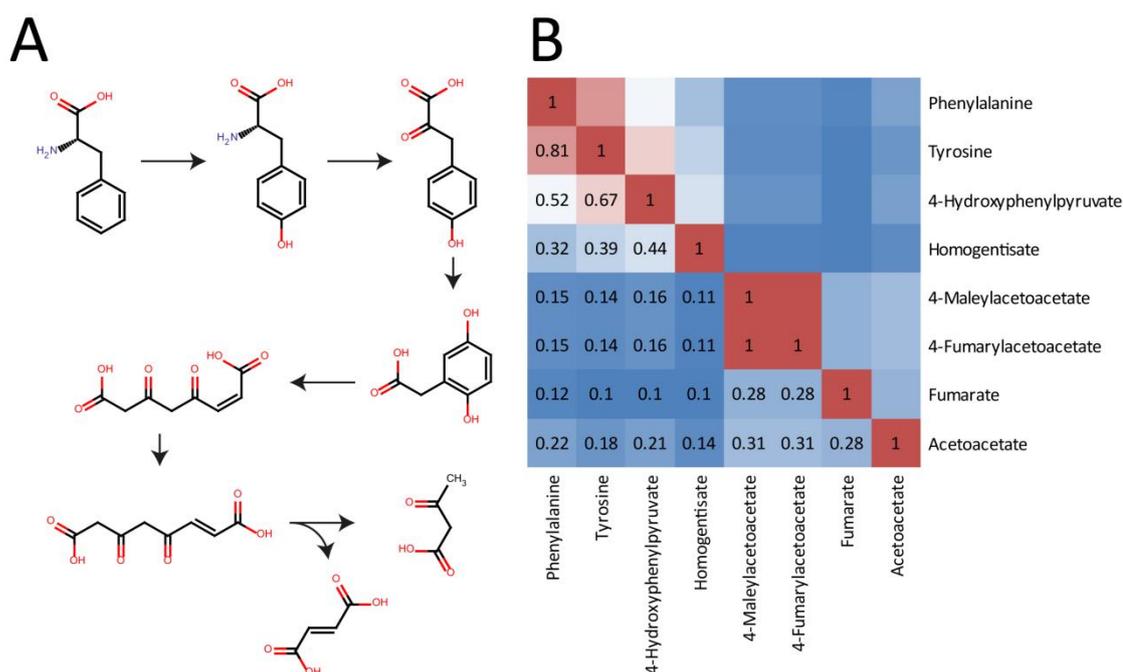


Figure 29: Metabolites on the same metabolic pathway show chemical similarity

(A) Common degradation pathway of phenylalanine via tyrosine in *C. elegans*. The pathway can be split into two parts. The first contains molecules with intact aromatic ring, whereas the second contains ring-opened molecules. (B) Heatmap of chemical similarity based on the Tanimoto coefficient (similarity = 1-Tanimoto). The two parts of the pathway can be seen in the heatmap.

All these parameters influence the retention of molecule in a chromatographic system. For a meaningful retention time model the most significant have to be discovered.

4.2.1.2 Quantitative structure retention relationship for *C. elegans* metabolites

Creek et al. reported a QSRR model based on measurement of 120 authentic standards on a ZIC-HILIC column. The average prediction error of this model was around 10-40%.

57 Standards were measured under RP and HILIC conditions to determine their retention to build a retention time model. To avoid possible overlap of isobaric substances or molecules that can generate similar MS patterns, all metabolites were distributed into six standard mixtures. The same model used by Creek et al. was used for prediction of retention times on

the BEH Amide column. For reversed phase separation a simple regression against the logD value was carried out. The buffers used with BEH Amide had a pH of 6.0 and 4.2, for buffer A and B respectively. For calculation of the physicochemical properties the mean of 5.1 was chosen.

Table 13: Used standards with retention times under RP and HILIC conditions

Standard compounds were injected in six different standard mixes to avoid isobaric and isomeric substance, which could lead to misinterpretation. Retention times were determined by creating EIC traces for each metabolite and taking the time at the peak apex.

Metabolite	Std. mix	RT BEH C8	RT BEH Amide
<i>alanine</i>	A	1.5(+)/n.d.(-)	10.71(+)/n.d.(-)
<i>arginine</i>		1.4(+)/1.3(-)	12.93(+)/12.97(-)
<i>asparagine</i>		1.5(+)/1.4(-)	11.62(+)/11.64(-)
<i>glutamate</i>		1.5(+)/1.4(-)	11.60(+)/11.54(-)
<i>glycine</i>		n.d.	n.d.
<i>histidine</i>		1.4(+)/1.3(-)	12.98(+)/12.97(-)
<i>leucine</i>		2.6(+)/2.5(-)	8.78(+)/8.73(-)
<i>serine</i>		1.5(+)/1.4(-)	11.50(+)/11.51(-)
<i>tryptophan</i>		5.1(+)/4.9(-)	8.73(+)/8.69(-)
<i>tyrosine</i>		1.9(+)/1.9(-)	9.98(+)/9.84(-)
<i>aspartate</i>	B	1.5(+)/1.5(-)	12.30(+)/12.26(-)
<i>isoleucine</i>		2.4(+)/2.3(-)	9.00(+)/8.95(-)
<i>lysine</i>		1.5(+)/1.3(-)	13.12(+)/13.14(-)
<i>methionine</i>		1.7(+)/1.7(-)	9.17(+)/9.16(-)
<i>phenylalanine</i>		3.4(+)/3.3(-)	8.67(+)/8.86(-)
<i>proline</i>		1.5(+)/1.5(-)	12.28(+)/n.d.(-)
<i>valine</i>		n.d.(+)/1.6(-)	9.95(+)/9.84(-)
<i>(S)-malate</i>	C	n.d.(+)/1.7(-)	10.65(-)
<i>2,3-pyrazinedicarboxylic acid</i>		n.d.(+)/n.d.(-)	n.d.
<i>benzoylformic acid</i>		n.d.(+)/4.9(-)	n.d.
<i>citrate</i>		1.9(+)/1.9(-)	n.d.
<i>cumarin 3-carbonic acid</i>		7.6(+)/7.5(-)	2.75(-)
<i>malic acid</i>		n.d.(+)/1.7(-)	10.43
<i>fumarate</i>		n.d.(+)/2.2(-)	n.d.
<i>glutaconic acid</i>		3.2(+)/3.3(-)	9.28(-)
<i>glutaric acid</i>		4.4(+)/2.9(-)	8.66(-)
<i>maleic acid</i>		n.d.(+)/2.2(-)	n.d.
<i>mesaconic acid</i>		3.2(+)/3.3(-)	9.28(-)
<i>meso-tartaric acid</i>		n.d.(+)/n.d.(-)	n.d.
<i>oxalic acid</i>		n.d.(+)/1.8(-)	n.d.
<i>pyruvate</i>		n.d.(+)/1.7(-)	n.d.
<i>succinate</i>		n.d.(+)/2.2(-)	n.d.
<i>tartronic acid</i>		n.d.	n.d.
<i>D-glucose</i>		D	12.2(+)/n.d.(-)
<i>D-glucose 6-phosphate</i>	1.5(+)/1.5(-)		13.53(-)
<i>D-ribose</i>	n.d.(+)/1.5(-)		n.d.(+)/5.70(-)
<i>rhamnose</i>	13.7(+)/1.5(-)		8.86(+)/6.12(-)
<i>arachidic acid(C20:0)</i>	E	15.6(+)/15.5(-)	1.50(-)

Metabolite	Std. mix	RT BEH C8	RT BEH Amide
<i>capric acid (C10:0)</i>		13.1(+)/13.0(-)	1.54(-)
<i>caprylic acid (C8:0)</i>		11.9(+)/11.8(-)	1.55(-)
<i>cholesterol</i>		n.d.	1.54(-)
<i>heneicosanoic acid (C21:0)</i>		15.7(-)	1.50(-)
<i>lauric acid (C12:0)</i>	E	13.9(+)/13.8(-)	1.52(-)
<i>myristic acid (C14:0)</i>		14.5(+)/14.4(-)	1.52(-)
<i>palmitic acid (C16:0)</i>		15.0(+)/14.9(-)	1.52(-)
<i>perlagonic acid (C9:0)</i>		12.5(-)	1.54(-)
<i>stearic acid (C18:0)</i>		15.3(+)/15.2(-)	1.52(-)
<i>undecanoic acid (C11:0)</i>		13.4(-)	1.54(-)
<i>biotin</i>			6.7(+)/6.7(-)
<i>cytosine</i>		1.2(+)	n.d.
<i>D-fructose</i>		12.2.(+)/1.5(-)	9.28(-)
<i>folate</i>	F	5.6(+)/5.6(-)	9.33/10.34(+)/ 9.26/10.26(-)
<i>carnitine</i>		1.4(+)	9.35/10.17(+)
<i>nicotinamide</i>		1.8(+)	n.d.
<i>orthophosphate</i>		n.d.	n.d.
<i>phenol</i>		n.d.	n.d.
<i>spermidine</i>		1.2(+)/1.5(-)	n.d.

Major problems during chromatography arise from poly acid, like citrate, which shows a broad peak over almost two thirds of the chromatogram. This retention behavior can be explained by having a closer look of the micro species present at a certain pH. At a pH of 5.1 used for the calculation of parameters for the BEH Amide separation, citrate has six different micro species, three of them in significant amounts. Out of the measured 57, 41 could be used for calculation of a QSRR model. For modeling the parameter and equation used by Creek et al. was used. They determined the most significant properties using multi linear regression. Out of 11 physicochemical parameters, the following 6 were the most significant. The logD values describes the polarity of a molecule, Neg and Pos indicate the charge of a molecule at the pH used for modeling, Rot specifies the number of rotatable bond and Phos the number of phosphate groups. The last parameter HBD/MS is an indicator for the molecule to cater hydrogen bond [121].

$$\log(RF) = k_1(\log D) + k_2(Neg) + k_3(Pos) + k_4(Rot) + k_5(Phos) + k_6\left(\frac{HBD}{MS}\right) + c$$

Equation 9: Model equation of QSRR model used by Creek et al.

The equation is taken from [121]. RF = retention factor. Neg/Pos = negative or positive charge at the given pH. Rot = number of rotatable bonds. Phos = number of phosphate groups. HBD = hydrogen bond donors. MS = mass.

Parameter coefficients k_1 to k_6 and the constant c were optimized using the Solver plugin in Excel and square sum minimization. With the optimized model equation retention times for all metabolites found in the *C. elegans* metabolic reconstruction were predicted.

Table 14: Calculated parameter for model equation 9 for the BEH Amide separation

Similar parameters as described by Creek et al. [121] were used and optimal solution was found using the Excel Solver function.

<i>parameter</i>	<i>BEH Amide</i>
k_1	-0.162619663
k_2	0.238574045
k_3	-0.004186964
k_4	-0.077000521
k_5	-0.387499906
k_6	3.412526988
c	0.148447755

Mean error between predicted and real retention of the used standards was lower than 5%. However for filtering of annotations an error of maximum 25% was assumed, due to the relatively low number of standards used.

For the RP separation regression of the retention time against the logD value at a pH of 3.5 was carried out. R^2 was 0.88 and the mean error between real and predicted retention was around 35%. Higher errors can be explained by missing metabolites in the middle range of the chromatogram. For further optimization more standards eluting in the region between 5 and 15 minutes have to be included.

All in all, this method was used as additional filter to avoid false positive annotations. It was used for filtering of potential biomarker candidates, were indicated in the further work. Nevertheless, for final identification retention time matching against an available reference standard was used if possible.

4.2.2 Identification of fragments, adducts and isotopes

To further identify possible adducts, neutral losses or fragments correlation analysis was carried out. Data was exported to R and Pearson correlation coefficients were calculated between all detected features. Signal arising from the same metabolite should have a positive correlation coefficient. Features with a correlation coefficient >0.9 and the same retention time were further considered. Mass differences between the related masses were calculated to compare them against theoretical mass differences of neutral losses and adducts. The most frequent observed mass difference were 1.003 Da and 21.982 (± 0.005 Da), which are related to ^{13}C isotopes and sodium adducts in positive mode. Other common neutral losses, like loss of

water, CO₂, phosphoric acid or acetyl groups were found. Figure 30 shows an example. Histidine was detected at two different retention times in the BEH Amide separation with negative ionization. The peaks were compared against an authentic standard of histidine, which showed that only the second peak corresponds to histidine. Correlation analysis showed that the first peak highly correlates with several other peaks, annotated as N-acetyl-histidine, N-acetylhistamine and histamine and their isotopes. Mass differences between the peaks are identified as neutral losses of C₂H₂O (loss of acetyl group) or loss of CO₂ (decarboxylation). All fragments showed intensities lower than the first isotopic peak of N-acetyl-histidine.

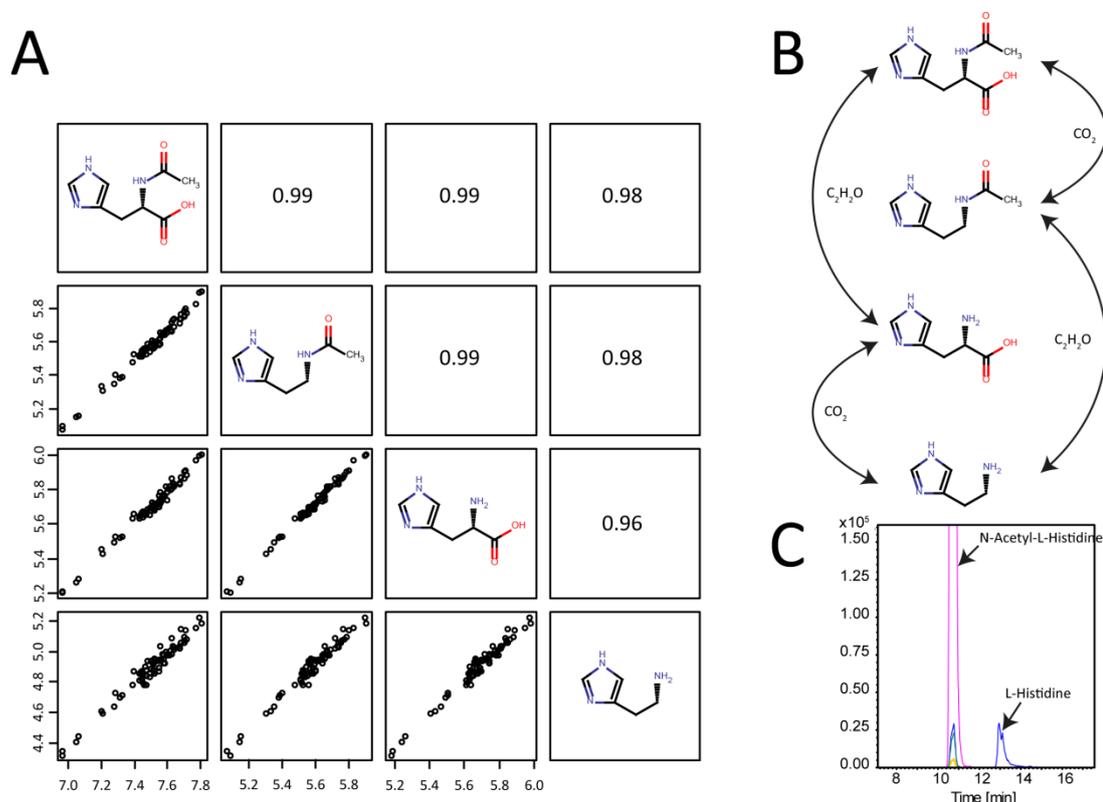


Figure 30: Example of cluster found using correlation analysis and RT filtering

(A) Paired scatterplots of different peaks show high correlation. (B) In source fragmentation with different neutral losses yields different ion species originating from the same metabolite. (C) Chromatograms of N-acetylhistidine and associated fragment peaks and an authentic histidine standard.

4.2.3 Lipid identification with AutoMSⁿ and *in silico* fragmentation

A major problem in lipidomics is the identification of different lipid species, isomers and isobars. One possible way to reduce the candidate structures for identification is to analyze the complete dataset and select a number of markers after statistical analysis, which are subjected to further analysis. A second approach is the identification of as much as possible lipid species in time with the first analysis. This is possible by using a so called data dependent approaches. Here detected peaks meeting certain criteria are directly selected and fragmented between normal MS scans.

A special attention has to be drawn to the two different lipid class's phosphatidylcholine (PC) and phosphatidylethanolamine (PE) species, because between them several isobaric molecules exist, as shown in Figure 31. In *C. elegans* also monomethyl (PE-NMe) and dimethyl (PE-NMe₂) phosphatidylethanolamines can occur as intermediates of PC synthesis. A second important class of lipids are triacylglycerols (TG), which serve as energy storage. In the LipidMaps database totally 6878 different TG species are listed, which have in total 300 different formulas. Diversity is achieved by different fatty acid composition and positional isomers. The maximum number of isomers and isobars for one formula was determined to be C₅₉H₁₀₄O₆, with totally 73 different species.

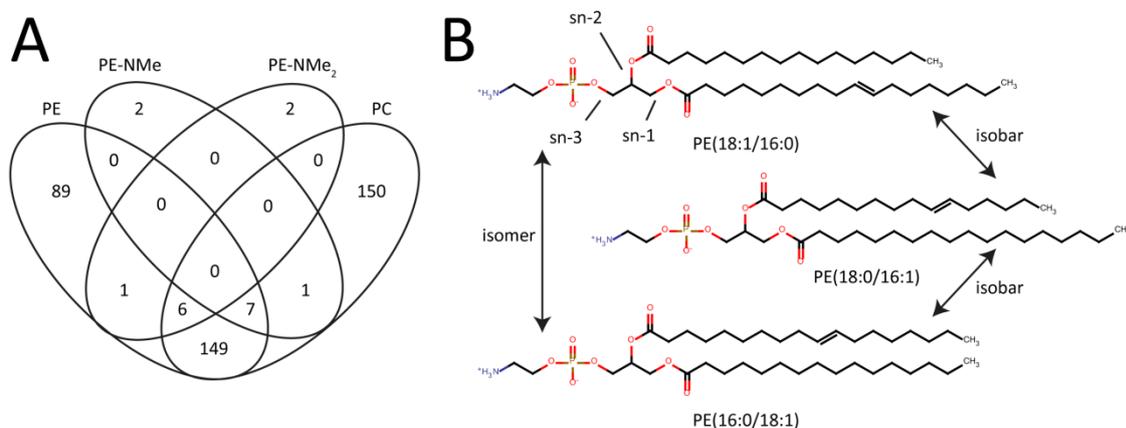


Figure 31: Isomeric and isobaric lipid species complicate analysis of lipidome data

(A) Venn diagram showing overlap of unique formulas from PE's, PE-NMe's, PE-NMe₂'s and PC's. PE-NMe's and PE-NMe₂'s are intermediates of one PC synthesis pathway in *C. elegans* (B) Difference between isomers and isobars explained across three different PE species. All molecules have the same exact mass.

If a chromatographic separation is used prior to mass spectrometry, retention time can be used as second filter to reduce false positive assignments. On a reversed phase column lipids elute based on their hydrophobicity, caused by different fatty acid chain length. Number and position of double bonds further influence retention behavior. Generally the effect of one double bond is a little bit weaker than difference based one carbon atom in the side chain.

For data dependent acquisition of MS/MS the AutoMSⁿ algorithm was used. This algorithm, build in Bruker microTOF control, enables automated generation of MS/MS spectra during a chromatographic separation or direct infusion. The algorithm first records a survey MS scan. From this spectrum the n highest mass are selected as precursor for MS/MS experiments. In the next n scans these masses are fragmented starting with the most intense. After fragmentation already used masses are excluded for a selected time range, usually the typical width of a chromatographic peak. After this the cycle starts again. AutoMSⁿ was used to identify major lipid species in conjunction with MetFrag. Lipid species are identified based on

characteristic fragments and neutral losses. In positive mode fragmentation of phosphocholines yields a characteristic mass of 184.1 Da, representing the phosphocholine headgroup. Another example is the neutral loss of a fatty acid side chain plus ammonium, which is detected when fragmenting $[M+NH_4]^+$ adducts of triacylglycerols [125].

Samples of the *C. elegans* infection model with *P. aeruginosa* PA14wt, *P. aeruginosa* PA14 Δ gacA and *P. aeruginosa* PA14 Δ pec7 was used to test the applicability of the AutoMSⁿ algorithm build in the Bruker microTOF control software. Totally 78 injection in positive mode were performed.

Automated fragmentation was carried out on all samples in this *C. elegans* infection model. Totally 30276 MS/MS spectra were collected. After removing duplicate spectra from duplicate chromatographic features 24563 remained. If such an approach would be only carried out on QC samples only 2362 MS/MS spectra would be collected. Due to intensity differences of the same species in different samples, in several cases it meets the criteria for fragmentation and in others not. This yields in much higher number of MS/MS spectra, based on the natural sample inhomogeneity and differences between sample states.

Lipid identification was achieved by using the MetFrag batch client on a grid computing system at the IBP Halle. LipidMaps served as input for molecule structures. A search with 0.005 Da maximum error was carried out and MassTRIX was used for preliminary annotation of lipids to the precursor masses and neutral masses needed for MetFrag were obtained from the corresponding annotation. From 19214 subjected MS/MS spectra, 6185 could be matched against a theoretical fragmentation pattern. For each subjected and matched MS/MS spectrum, MetFrag returns an .sdf file. The files were merged into a single .sdf file, which was imported into Excel using JChem for Excel. MetFrag scores ranged from 0 to 0.94. Totally 4388 MS/MS spectra could be used for identification with a MetFrag score ≥ 0.75 . Results from randomly selected spectra were manually inspected and most obtained hits were correctly assigned. The identified lipids included 439 PC's, 223 PE's and 109 PS's possible species. The most common fragments detected were head groups and their respective loss, e.g. 184.1 and M-183 for PC species.

For most phospholipids several MS/MS spectra at different retention time were found, which correspond to different isomers eluting at different time points. In total 2 to 6 different isomers were detected. An example can be seen in Figure 32A, where 5 different peaks were detected in the retention time region corresponding to phospholipids. The mass is corresponding to PC(40:6).

Triacylglycerols were mainly detected as $[M+NH_4]^+$ adducts, which are not covered in MassTRIX. Theoretical adduct masses of all TG's covered by LipidMaps were calculated and used for the annotation process. Totally 1006 different peaks were annotated in *C. elegans* samples as TG's, with carbon counts from 37 to 66 and up to 15 double bonds. The longest detected homologous series contained 46 carbon atoms in the side chain. Most isomers were detected with the TG(46:1). Interestingly 27 peaks were annotated as TG(46:1) but only 24 are listed in LipidMaps. TG's containing for example alicyclic fatty acids are not listed in LipidMaps, but *C. elegans* is known to use them from the ingested bacteria. TG's containing 46 carbons in the side chain were used to exploit retention behavior. From the obtained MS/MS spectra 128 had a putative annotation as TG and were subjected to MetFrag for *in silico* fragmentation. Most of the subjected mass spectra could be used for identification and confirmation of the annotation.

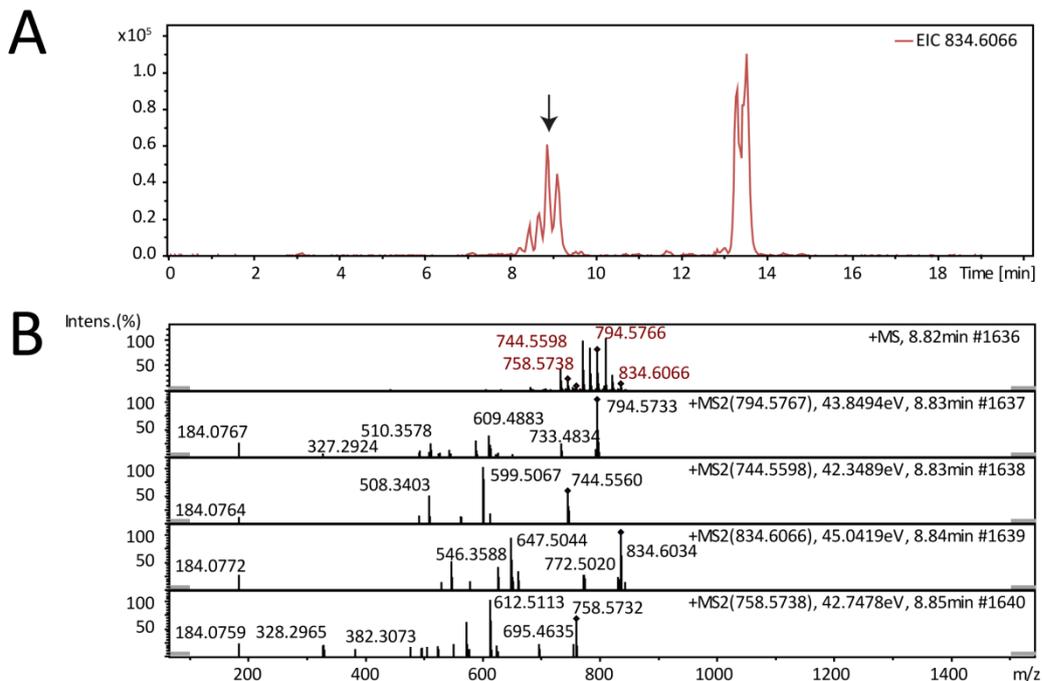


Figure 32: The AutoMSⁿ algorithm is a data dependent algorithm for the automated acquisition of MS/MS spectra during a chromatographic run

(A) Extracted ion chromatogram of mass 834.6066, which is possible corresponding to PC(40:6). For this mass 4 different chromatographic peaks in the region where phospholipids are eluting are found. The arrow indicates the time point where the MS/MS spectra shown in panel B were obtained. (B) MS and MS/MS obtained by the AutoMSⁿ algorithm. The first panel shows the survey MS scan with the masses which are fragmented in the 4 upcoming scans in red. MS/MS scan are depicted in the lower four panels, whereas the third corresponds to PC(40:6), showing the typical 184.1 fragment special for PC.

Visual inspection of a retention time versus mass plot of all peaks annotated with 46 carbons in the side chain showed a linear trend with increasing double bonds. One additional double bond shifts retention of about 0.4 minutes. The effect of an additional carbon atom in the side chain was similar. Figure 33 shows exemplary the retention behavior of selected TG

species. This methodology allows finding corner points from which a possible fatty acid composition can be deduced. In several cases, at the cross points of the double and carbon line, two instead of one point exist, which are shifted left and right from the expected position. In these cases possibly *cis*- and *trans* or positional isomers are separated. Due to the short gradient time of 20 minutes, this separation is only visible in certain rare cases and mostly these isomers elute at the same time. Bird et al showed separation of *cis-trans* Phospholipid isomers with an established lipid separation method and high resolution mass spectrometry [126].

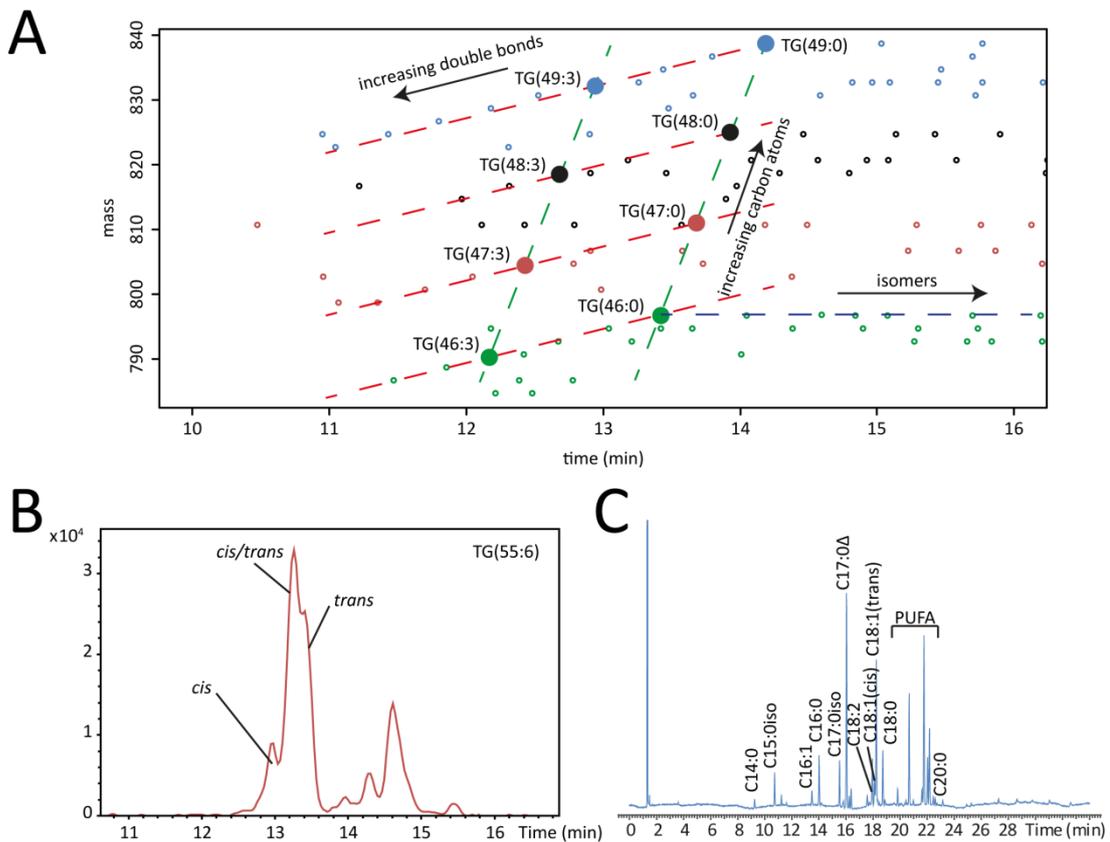


Figure 33: Retention time shifts and mass differences allow identification of members of homologous series.

(A) Mass difference of 2 Da indicates an additional double bond, which leads to a retention time shift of 0.4 minutes. Increasing carbon numbers have a mass difference of 14 Da for each additional carbon atom, which also leads to a retention time shift of 0.4 minutes. Members of homologous series can be found on straight lines in the 2D chromatogram. (B) Sometimes for different peaks of the same molecule are found in a narrow retention time range. The figure shows an extracted ion chromatogram of TG(55:6). The small peak in the front and the shoulder at the end are possibly due to partial separation of *cis-trans* lipid isomers. (C) GC-FID analysis of total fatty acids showed that C17:0Δ is the major fatty acid species in *C. elegans*.

To obtain a starting point for identification normally lipid species not occurring in the studied organism are used as internal standards, e.g. PC(17:0/17:0) or TG(17:0/17:0/17:0) in human cells. Because *C. elegans* is able to synthesize these molecules on its own or they are taken up from bacterial food, these standards cannot be used. Total fatty acid profiling using GC-FID showed C17:0 as the predominant fatty acid in *C. elegans* followed by C18:1n7 and

C20:4n6. Suitable alternatives for these standards could be heavy isotope labeled analogues, e.g. molecules where hydrogen atoms of the carbon backbone are exchanged with deuterium.

Although a high amount of initial MS/MS spectra were obtained, only a small fraction of them could be finally used for identification. From initially 19214 subjected spectra only 4388 returned a usable result, which is about 23%. Due to the relative high isolation window for fragmentation, possibly substance with a similar mass and retention time are fragmented at the same time, which would lead to a mixed MS/MS spectrum and possible confusion in data interpretation. Also a longer chromatography method would lead to better results and separation of interfering molecules.

Summing up, even though more optimization work is needed the AutoMSⁿ algorithm together with MetFrag is a useful tool for on-the-fly identification of lipid species. It allowed the identification of several hundred different lipids from different classes. For the non-targeted approach lipidomics approach used in this work, this was certainly useful to confirm at least partly the annotations from MasSTRIX. For final identification of a lipid species and the correct isomer, pure chemical standards are needed. Currently only a limited number of such standards are available.

4.2.4 Mass difference networking for improved metabolite annotation

Using MasSTRIX for annotation of ICR-FT/MS masses about 30% of all peaks could be annotated as either $[M+H]^+$ or $[M+Na]^+$ adduct in positive or $[M-H]^-$ in negative ionization mode. The mean annotation error was 0.03 and 0.13 ppm for negative and positive ionization mode, respectively. Metabolites from all major pathways known to be present in *C. elegans* were annotated.

To obtain chemical formulas, which can be used for searching chemical databases like PubChem or Chempider masses were subjected to mass difference networking described by Tziotis et al. using NetCalc, an in-house written MatLab script for calculating mass difference networks [127]. Mass differences according to Table 15 were used for the calculation of formulas. Main network graphs were visualized using yED (version 3.8, http://www.yworks.com/en/products_yed_about.html) and network characteristics were calculated using Cytoscape (version 2.8.2, <http://www.cytoscape.org>). These network characteristics include node degree, number of neighbors and path length. Overall data shows the calculated networks are not random, which has already been shown for natural organic matter by Tziotis et al.

Totally 1184 and 778 additional formulas for positive and negative ionization mode respectively could be annotated to the MassTRIX formulas in the main network. Several sub networks disconnected from the main graph exist, suggesting that the input mass differences may be incomplete or insufficient to cover all metabolic transformations or novel not known mass transformations. Only basic transformations like oxidation can be represented in mass differences, special cases like condensation or cleavage reaction are not amenable to this approach. Frequencies of individual transformations together with shapes of the main networks from positive and negative ionization mode can be found in Figure 34. Highest frequencies were found within homologous series, oxidation either by gaining an oxygen or loss of H₂, hydration/dehydration and decarboxylation. Many metabolic pathways like β -oxidation rely on these reactions.

Table 15: Mass differences used for network calculation

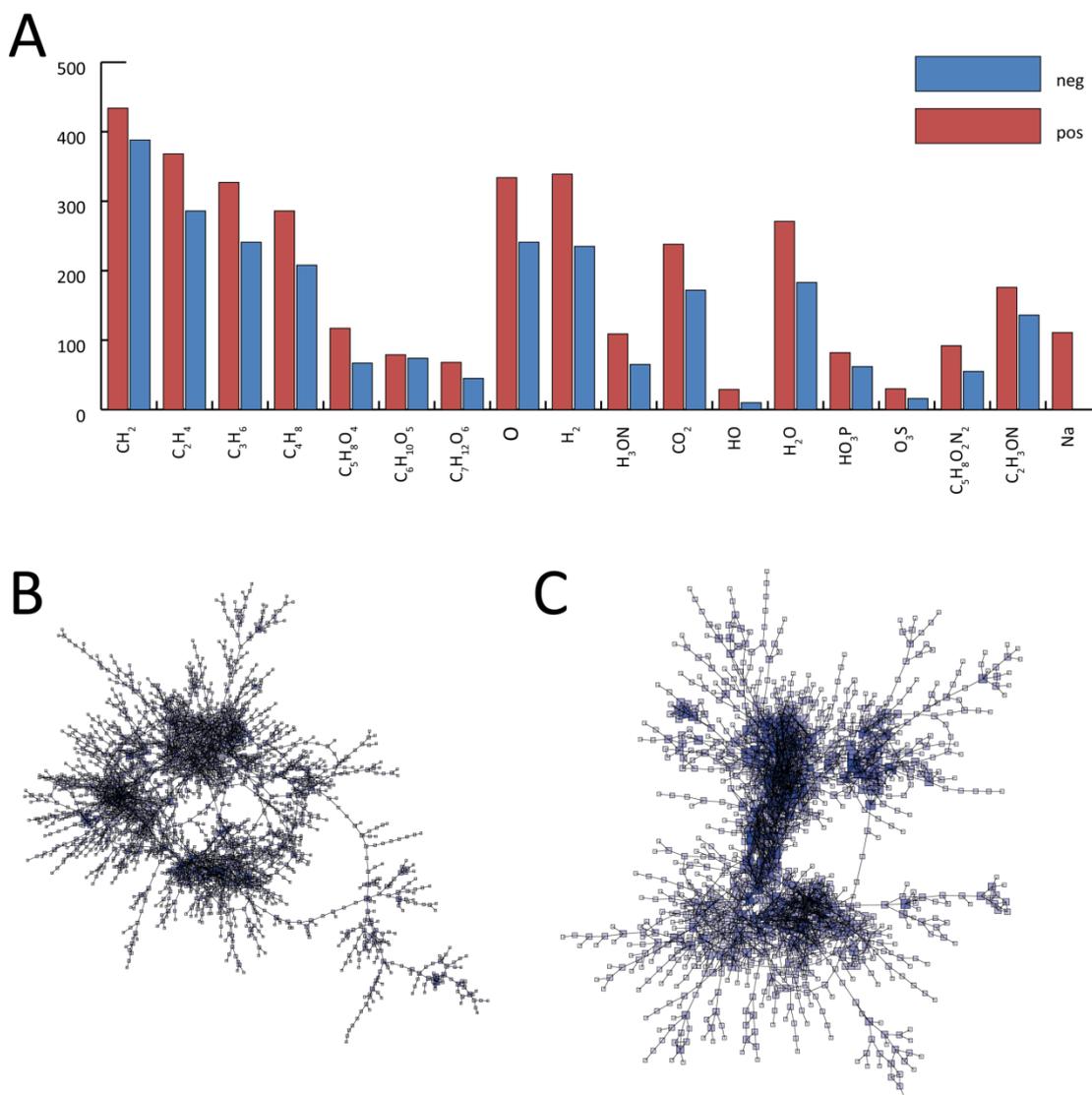
The mass differences correspond to homologous series, which can be found between different fatty acid in lipids, or common metabolic transformations. In positive mode the mass difference between the [M+H]⁺ and [M+Na]⁺ adduct were added to identify possible sodium adducts.

<i>difference formula</i>	<i>Biological meaning</i>	<i>mass difference</i>	<i>ionization mode</i>
CH ₂	<i>Homologous series</i>	14.0156501	+/-
C ₂ H ₄		28.0313001	+/-
C ₃ H ₆		42.0469502	+/-
C ₄ H ₈		56.0626003	+/-
C ₅ H ₈ O ₄	<i>O-linked sugars</i>	132.042259	+/-
C ₆ H ₁₀ O ₅		162.052823	+/-
C ₇ H ₁₂ O ₆		192.063388	+/-
O	<i>Oxidation (O)</i>	15.9949146	+/-
H ₂	<i>Reduction/oxidation (H2)</i>	2.01565006	+/-
H ₂ O	<i>Loss / gain of water</i>	18.0105647	+/-
H ₃ ON	<i>transamination</i>	1.03163448	+/-
Na	<i>ESI sodium adduct</i>	21.9819442	+
HO	<i>N-hydroxylation</i>	17.0027397	+/-
HO ₃ P	<i>Phosphate ester/ anhydride</i>	79.9663305	+/-
O ₃ S	<i>Sulfate ester / anhydride</i>	79.9568149	+/-
C ₅ H ₈ O ₂ N ₂	<i>glutamine</i>	128.058578	+/-
C ₂ H ₃ ON	<i>glycine</i>	57.0214637	+/-
CO ₂	<i>decarboxylation</i>	43.9898292	+/-

Table 16: Overview on masses in main network and the annotation of new formulas

The NetCalc approach was able to predict several new formulas that were not found before in the used databases. These formulas can be used to query bigger databases like Pubchem or Chempider or serve as basis for *de novo* structure elucidation.

Features	Pos	Neg
No of total features	13080	11738
No of features in main network	2497	1301
No of MassTRIX annotation in main network	1313	523
Novel formulas (NetCalc)	1184	778

**Figure 34: Overview on NetCalc networks**

(A) Counts of different mass differences observed in the main networks from negative and positive ionization mode.

(B) Main network from positive ionization mode (C) Main network from negative ionization mode

4.2.5 Alignment of ICR-FT/MS and UHPLC-UHR-ToF-MS data

ICR-FT/MS and UHPLC-UHR-ToF-MS data were aligned with a custom written Perl script. The advantage of this alignment is to have also exact masses for unknown entities in the

chromatograms, without utilizing an online coupling of ICR-FT/MS. To find optimum alignment conditions, different errors from 1 to 15 ppm were tested and numbers of hits were counted. Due to the lower resolution of UHR-ToF-MS several ICR-FT/MS masses can correspond to a single chromatographic feature. The mass with the lowest error was chosen as the true mass and the number of different possible masses was stored. Figure 35 gives an overview on the alignment of BEH Amide separation with negative ionization as example.

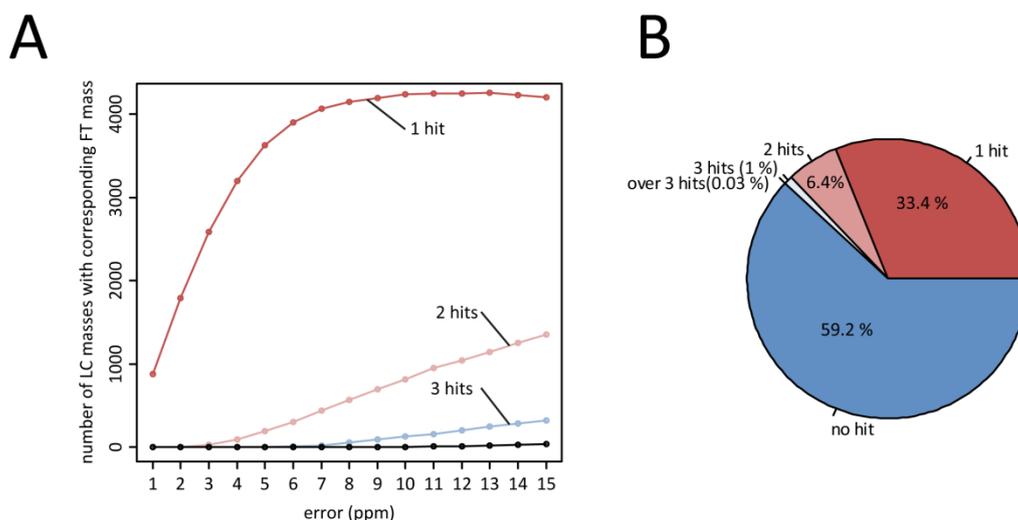


Figure 35: Results of ICR-FT/MS and UHPLC-UHR-ToF-MS alignment

(A) Number of LC masses that have a corresponding FT mass as function of the ppm error. Different lines show counts of masses that have either 1 hit, 2 hits, 3 hits or more (black line). (B) Percentage of masses with no, 1, 2, 3 or more hits at 10 ppm.

Interestingly, 59% of the observed LC peaks had no corresponding FT mass at 10 ppm error. Because mass ranges were different, masses out of the ICR-FT/MS mass range were removed, but still 57% had no FT mass. Over 50% of these masses have a mass small than 500 Da. For all UHPLC-UHR-ToF-MS post-processing steps a mass error of 0.005 Da was used, including annotation using MasSTRIX. For masses lower than 500 Da this would yield an error > 10 ppm. Because metabolic profiling was carried out with a large mass range, calibration of UHPLC-UHR-ToF-MS data was only optimal for the middle of the mass range. Most biological relevant molecules have a mass below 500 Da. For this purpose alignment of ICR-FT/MS and UHPLC-UHR-ToF-MS was repeated with an absolute error 0.005 Da. Using an absolute error of 0.005 Da for alignment enhanced performance for masses lower than 500 Da. In the higher mass range this error didn't perform as good as the relative ppm scale. These results are valid for all other used methods. For further work a 10 ppm error was chosen, being aware that below 500 Da possibly more hits exist with an absolute error below 0.005 Da.

Splitting the mass ranges of both methods into two methods, one for low masses until 500 Da and for masses >500 Da can be a suitable solution to overcome the calibration problem. However, this would double the effort for measurement and data analysis time.

4.3 Biological interpretation of the *C. elegans* infection model

The metabolite maps obtained from the first part of this chapter were subjected to statistical evaluation to reveal metabolites significant for each sample condition. This evaluation is based on non-supervised and supervised multivariate statistics. Results from the different methods are grouped together according to their biological interpretation.

4.3.1 Statistical data analysis

4.3.1.1 DI-ICR-FT/MS

HCA and PCA were performed in R or SIMCA-P 9.0 and used to reveal group separation and data quality. To assess the class separation different orthogonal partial least square models (OPLS/O2PLS-DA) have been built up in SIMCA-P 11.5. The contribution of the masses in the separation of the different groups has been evaluated through the examination of the different S-PLOT. Always one group of a given genotype was compared against the remaining four. Masses with the highest magnitude in covariation and correlation were chosen as potential candidates for the class separation. P-values were calculated using Welch's t-test and results were combined in MS Excel for further analysis.

4.3.1.2 UHPLC-UHR-ToF-MS

Data analysis of UHPLC-UHR-ToF-MS data was carried out in Genedata Analyst 7.5. In the case of BEH Amide separation with positive ionization for some chromatograms no data was obtained. Visual inspection of sample vials showed no remaining sample. Possibly no sample was injected. To overcome the problem of missing values, these were replaced with the arithmetic mean of the remaining values of a group. To test if this interferes with further analysis, PCA using the non-filled data without the chromatograms and the filled data was carried out. Data filling did not lead to a distortion of the data, because group separation was similar between the two datasets. For further analysis the mean filled data set was used.

Several PLS models, comparing always one condition of a genotype were compared against the remaining four. Importance for the group separation in the models was evaluated by a VIP score > 1. All results were combined in MS Excel for further analysis. Results were tested for significant differences between the sample groups using Welch's t-test. Changes with a p-value < 0.05 were considered as significant.



4.3.2 Survival of *C. elegans* challenged with different pathogens

The nematode strains used in this work were provided by the Caenorhabditis Genetics Center, which is funded by the NIH National Center for Research Resources (NCRR). The Strains *fer-15* (b26) and *daf-2* (e1370) of *C. elegans* were used in this work. Worms were cultured as described previously [128]. Eggs were isolated by hypochlorite treatment of gravid adults and hatched on NGM at 25°C to obtain synchronized sterile worms for use in the slow killing and feeding studies. The *P. aeruginosa* strain PA14, *Salmonella enterica* serovar Typhimurium strain 12023 and *Escherichia coli* OP50 have been described. The *P. aeruginosa* PA14 Δ *gacA* mutant was generated using the vector pKNG101 as described [129]. Media for routine bacterial culture and maintenance was Luria–Bertani broth. Bacteria were cultured on nematode growth medium (NGM) as described below for nematode feeding and killing assays. The slow killing assay was performed as described previously [130]. Each independent assay consisted of three replicates. *E. coli* OP50 was used as a negative control. L4 stage *C. elegans* were picked onto plates containing overnight growth of each bacterial strain, and on a daily basis worms were evaluated for viability. Worms were considered dead when they no longer responded to physical stimuli. The statistical analyses indicated were carried out using the Graph Pad Prism 4 software.

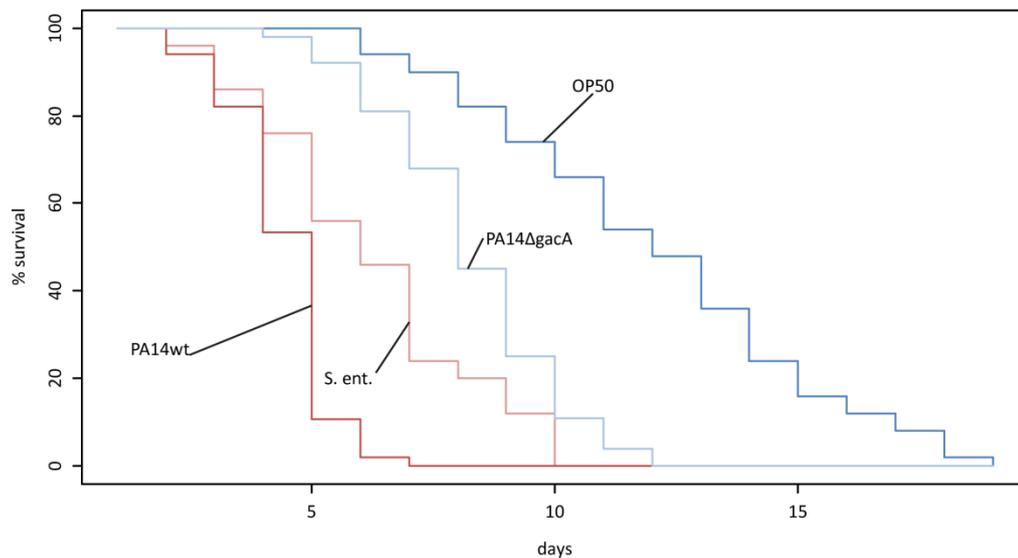


Figure 36: Typical killing curves obtained from *C. elegans* fed on different pathogenic bacteria.

L4 *C. elegans* grown on *E. coli* OP50 were switched to a pathogenic diet and were evaluated on a daily basis worms for viability. Worms were considered dead when they no longer responded to physical stimuli. Pathogenic bacteria significantly reduce lifespan of the nematode.

Under standard laboratory conditions *C. elegans* is fed on different strains of *E. coli*, like OP50, a uracil auxotroph, which was chosen by Sydney Brenner, because it grows as thin layers on a solid surface. However if diet is switched to pathogenic bacteria survival is greatly

reduced. When a population of *fer-15 C. elegans* is raised on *E. coli* OP50, the median survival of the group is 10 days, for *P. aeruginosa* PA14wt the median survival is reduced 3.5 days, for the *P. aeruginosa* PA14Δ*gacA* mutant the survival to 7 days and for *Salmonella enterica* 12023 the median survival to 5.5 days. These results reflect previously obtained results and demonstrate that each of the pathogens used is able to kill *C. elegans*. Typical killing curves are depicted in Figure 36. For *daf-2* worms survival was similar to the literature.

4.3.3 ICR-FT/MS metabolic profiling separates different sample conditions

The obtained ICR-FT/MS datasets were first evaluated by unsupervised statistical methods like PCA and HCA. In PCA no clear separation of the ten different groups was obtained, so data was split into two data sets, according to the genetic background of *C. elegans*.

In the *fer-15* data set the first principal component (PC1) separated *E. coli* OP50 (OP50) and *P. aeruginosa* PA14wt (PA14wt) fed and starved worms from *S. enterica* (*S. ent*) and *P. aeruginosa* PA14Δ*gacA* (PA14Δ*gacA*) fed worms. The second principal component (PC2) separates starvation, PA14wt and PA14Δ*gacA* fed worms from nematodes fed with OP50 and *S. ent*. Because infection of *C. elegans* with PA14wt is dependent on *daf-2*, this clustering suggests a separation according to *daf-2* dependency of the metabolic phenotypes. Also, PA14wt and PA14Δ*gacA* are grouped together with starvation separated from the remaining two groups from, making them similar to the starvation response. Looking into the scores plot of the *daf-2* dataset, group separation was different. PC1 separates starved, PA14wt and PA14Δ*gacA* fed worms from *S. ent* and OP50 fed worms. Furthermore, *S. ent* and OP50 are separated along PC2. A similar pattern was found also in positive mode analysis with ICR-FT/MS and BEH C8 separation with positive ionization.

A major feature of *P. aeruginosa* infection in *C. elegans* is the up-regulation of *daf-2*. This event is abolished in *daf-2* mutants and PA14wt and PA14Δ*gacA* responses are becoming more similar, so that in the *daf-2* mutant, the first PC separates PA14 fed and starved worms from OP50 and *S. ent* fed *C. elegans*. However, the responses between PA14wt and PA14Δ*gacA* can still not be the same, because PA14wt is still able to secrete protein effectors. This makes a comparison of the responses of the two different genotypes interesting, because it is possible to differentiate between the metabolic responses according to *daf-2* up-regulation and the response due to bacterial effectors.

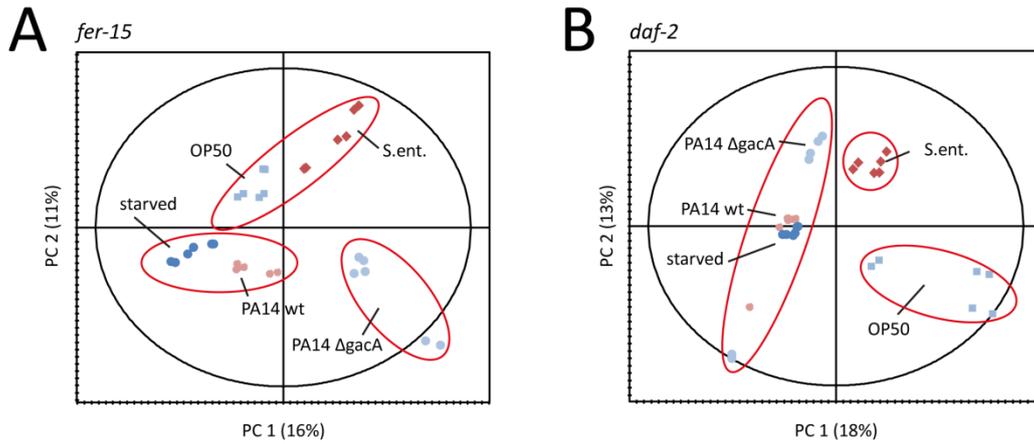


Figure 37: Principal component analysis of negative ICR-FT/MS dataset

(A) In the *fer-15* genetic background the first principal component separates *E. coli* OP50 and *P. aeruginosa* PA14 wt fed and starved worms from PA14Δ*gacA* and *S. ent.* fed ones. The second principal component separates *E. coli* OP50 and *S. ent.* fed *C. elegans* from the remaining groups. (B) In the *daf-2* mutants, a similar separation across the first principal component was observed as with the second one in the *fer-15* worms, suggesting this separation is becoming more important.

Masses were uploaded group specific to MasSTRIX to obtain counts of metabolites per pathway to get a first idea about altered metabolic pathways. Masses occurring four or more times in one group were uploaded per group and different conditions were compared on pathway level using the compare function in MasSTRIX. Interestingly, *fer-15* worms fed with OP50 have the highest counts in most amino acid metabolism pathways. Amino acid metabolism may be down regulated in all kind of stress, like infection or starvation. However, also *daf-2* mutants fed with *E. coli* OP50 have lower counts on these pathways.

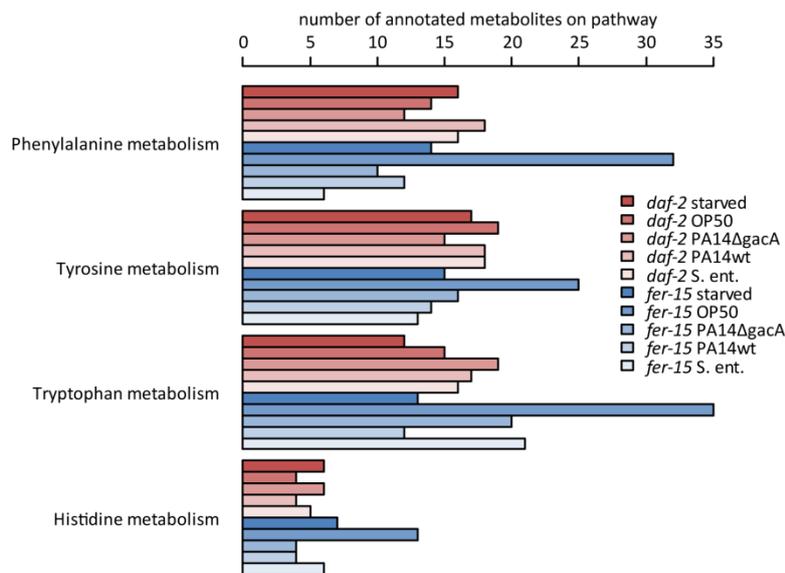


Figure 38: Pathway comparison from positive ionization ICR-FT/MS data

fer-15 *C. elegans* fed with OP50 show the highest count of annotated metabolites for most amino acid metabolic pathways, whereas all other have lower counts, including *daf-2* worms fed with OP50.

To reveal metabolites significant for the separate groups several orthogonal partial least square models (OPLS/O2PLS-DA) were built. Always one group of a genotype was compared against the remaining four. As an illustration of this, in OP50 fed worms, degradation products of phenylalanine like phenyl pyruvate, phenyl lactate, and 2-hydroxy-3-phenylpropanoate were found to be significantly higher than in worms in the other feeding schemes. This is in agreement with the counts of annotated metabolites on the phenylalanine metabolism pathway. The pathway map in Figure 39 shows a cut-out of phenylalanine metabolism with markers for OP50 fed *fer-15* nematodes mapped in green color and non-specific metabolites in gray color. Some of these metabolites might be of bacterial origin, from ingested *E. coli* OP50, as *C. elegans* lacks enzymes for their production (green frame in figure 3B). Further, the data shows the linked metabolic pathway of Tyrosine metabolism is also affected. Additionally tryptophan, histidine, glutamine and glutathione were found with significantly higher intensities with OP50 diet, suggesting a large scale down regulation in amino acid metabolism in *C. elegans* during bacterial infection and starvation. In case of *daf-2* worms it is known that they have a down-regulated amino acid metabolism and elevated amino acid levels. However, down-regulation of amino acid metabolism seems to be common to all infections, suggesting that it is possible a response to pathogen associated patterns (PAMP's). Because no elevated levels of amino acids are found, these are possibly used for protein biosynthesis or they are excreted.

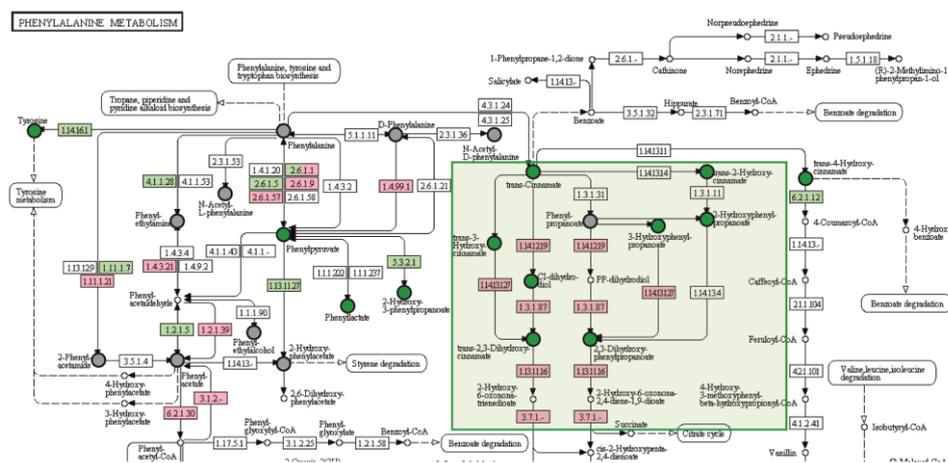


Figure 39: Several metabolites of phenylalanine were found significant for *fer-15* worms fed with *E. coli* OP50
Green dots correspond to metabolites found to significant for *fer-15* *C. elegans* fed with *E. coli* OP50, whereas grey ones where detected, but not specific for any group. Several of these hits are laying a part of phenylalanine metabolism pathway map, not present in *C. elegans*, represented by the green shading. Enzymes marked in red are present in *E. coli*, suggesting that these metabolites originate from ingested bacteria.

The glycolysis and the TCA cycle are the central metabolic pathways and many metabolites derive from molecules present in these pathways. Negative ICR-FT/MS data was used to

evaluate the effect of infection on these central pathways. Figure 40 compares the level of the 9 detected metabolites. Out of these 9 metabolites, 8 had a p-value < 0.01, with higher levels in PA14 Δ gacA and one had a p-value < 0.05. All of the detected metabolites show lower intensities in PA14wt worms, which could be an indicator for low consumption of these metabolites. If this would be the case, oxaloacetate and the precursor malate would possibly accumulate in the cells. Oxaloacetate, was not detected, possibly because to decarboxylation in the ESI spray. Malate the direct precursor has also lower intensities in infected *C. elegans*. However, oxaloacetate is a strong inhibitor of complex II, converting succinate to fumarate. Feasibly it is converted to aspartate to overcome inhibition, but no changes in aspartate were detected. A more quantitative method like UHPLC-MS is needed to better reveal changes in amino acids, like aspartate. Additionally, search for the molecular formula of aspartate showed several isobaric species and the intensity profile in ICR-FT/MS may show a superimposition of more than one molecule. Lastly, *C. elegans* is capable of performing ethanolic fermentation, which could be the preferred pathway for generation of energy during infection. *fer-15 C. elegans* fed with *S. ent* show high level comparable to OP50 or PA14 Δ gacA., showing that down regulation of the TCA cycle is a PA14wt specific event. All this effects are erased in *daf-2* worms, where most of the levels are similar, independent of food. Fuchs et al. showed that in *daf-2* mutants the glyoxylate shunt, gluconeogenesis and starch metabolism are up regulated and glycolysis and TCA cycle are down regulated [72].

In contrast to *P. aeruginosa* infection, in *S. enterica* infected worms significant masses were annotated on starch and sucrose metabolism, amino sugar and nucleotide sugar metabolism and galactose metabolism. Glycogen (as listed in KEGG, C00182), glucose, sucrose, and glucose 6-phosphate show higher intensities in the spectra of *S. enterica* fed worms. Higher levels of these metabolites suggests that *C. elegans* is mobilizing energy to fight the infection. A recent study by Antunes et al. used a mouse model to evaluate effects of *Salmonella* infection on the host metabolome [92]. In this study the authors reported a predominant effect on the host hormone metabolism; they also found lower levels of these energy utilization metabolites. The group used a significantly longer infection time before sampling, which can explain lower levels. The authors also hypothesize about a general energy mobilization response to infection, which could well be preserved during evolution.

To examine if the metabolome data of *P. aeruginosa* infected worms obtained by ICR-FT/MS is consistent with previously reported data, it was overlaid with gene expression data obtained from Troemel et al. [131].

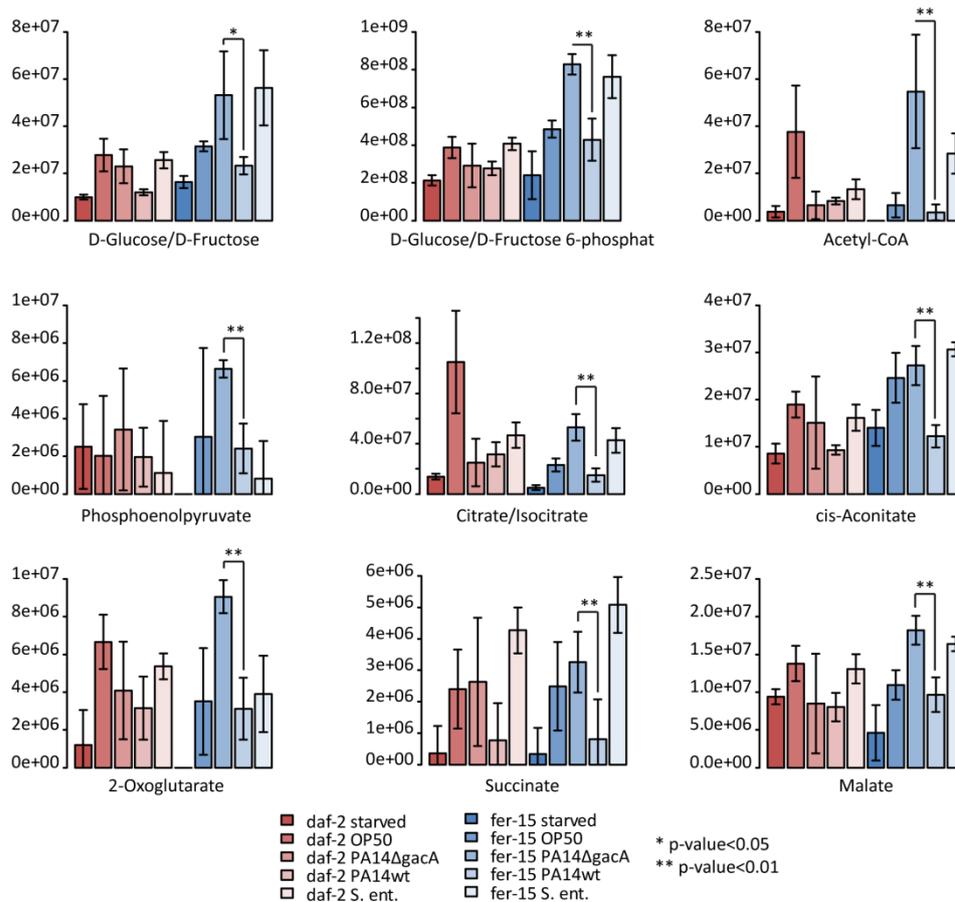


Figure 40: Metabolites from Glycolysis and TCA cycle detected in negative ICR-FT/MS analysis

Several metabolites could be detected, but most were below the mass range of the ICR-FT/MS. From the detected all show significantly lower levels in *fer-15* worms infected with PA14wt, whereas *S. ent* infected worms have similar levels to OP50 and PA14ΔgacA. In *daf-2* worms this response is erased.

An example consistent with previous results is the down regulation of *hsd-1* (Y6B3B.11) in *P. aeruginosa* infected worms. *hsd-1* encodes for a hydroxysteroid dehydrogenase, which converts cholesterol to 4-cholesten-3-one, a precursor of Δ4-dafachronic acid, a major signaling molecule regulating *C. elegans* development and lifespan. For PA14wt, PA14ΔgacA and *S. ent* infected worms intensities for 4-cholesten-3-one are lower and for starved worms slightly higher compared to OP50 fed worms (PA14wt p-value = 0.04, PA14ΔgacA p-value = 0.03, *S. ent.* p-value = 0.02). *daf-2* mutants generally have lower levels of 4-cholesten-3-one. Furthermore other sterols and sterol derivatives were found in significantly higher levels in starved worms, confirming the role that steroid hormone signaling plays in *C. elegans* during the stress of starvation.

These examples show how non-targeted metabolic profiling using DI-ICR-FT/MS is able to separate different metabolic phenotypes. Nevertheless, ion suppression is a major problem in direct infusion experiments and methods for better quantification are needed. Furthermore, isobaric and isomeric substances cannot be separated. Both facts are possibly leading to false

positive results. For confirmation or rejection of this results non-targeted metabolomics using UHPLC-UHR-ToF-MS was additionally employed.

4.3.4 Infection with *P. aeruginosa* leads to increased β -oxidation

Statistical analysis of the UHPLC-UHR-ToF-MS revealed higher levels of l-carnitine in *fer-15* worms infected with *P. aeruginosa* PA14 wt. This led to the suggestion that infection may interfere with fatty acid metabolism. In reversed phase separation with positive mode ionization, several acylcarnitines were detected. Levels of C2 to C17 acylcarnitines are significantly lower ($p < 0.05$) in PA14 wt infected worms compared to PA14 Δ *gacA* fed ones, independent of the worms genotype, shown in Figure 41. The only exceptions are C12 and C16 acylcarnitine in *daf-2* mutants. Acylcarnitines are the transport form of fatty acids across the membrane of mitochondria, where β -oxidation takes place. To generate energy from lipid storage, fatty acids are first released from triglycerides by a lipase. After activation with coenzyme A, fatty acids are transferred to carnitine to yield acylcarnitines. These are transported across the mitochondrial membrane. Inside the mitochondria they are transferred back to Acyl-CoA's and undergo β -oxidation. The free carnitine is transported back to cytosol, by antiport with the acylcarnitines. Higher levels of carnitine may be explained by increased β -oxidation. Indeed gene expression analysis carried out by another group, showed that several genes corresponding to β -oxidation are up-regulated in worms fed with PA14 wt compared to PA14 Δ *gacA* [131]. Up-regulated genes include fatty acid CoA synthetases, generating acyl-CoA's from free fatty acids and 3-hydroxyacyl-CoA dehydrogenases, which are enzymes of the β -oxidation. Furthermore, as mentioned before, PA14 wt is able to up regulate *daf-2*. Since levels of acylcarnitines are comparable between *daf-2* and *fer-15* worms, this seems to be a *daf-2* independent event.

To further follow this idea, the lipidome data was searched for indicators for increased β -oxidation. Levels of Triglycerides were compared between PA14wt and PA14 Δ *gacA* fed worms. Totally 100 peaks, that correspond to Triglycerides were found in significantly lower levels ($p < 0.05$) in infected worms. Taken together, these results imply that infection with PA14wt results in increased β -oxidation for energy generation.

If β -oxidation dominates, ketone bodies are built out of excessive acetyl-CoA. Two acetyl-CoA's are condensed to build acetoacetyl-CoA. In a further reaction acetoacetate is reduced to D-3-hydroxybutyrate or decarboxylates spontaneous to acetone. Acetoacetate was not detected, but D-3-hydroxybutyrate and its carnitine conjugate. Both are lower during infection, suggesting that β -oxidation is taking place beside normal energy metabolism. Triacylglycerides are mobilized in a cAMP dependent manner during starvation or conditions

with low amount of other nutrients available. Possibly, this is a reaction of *C. elegans* to increased damage of the intestine and accumulation of bacteria in the gut. However, comparison of gene expression showed only a partial overlap with starvation induced expression alteration. It is more likely that this is an effect of the bacterial effector ExoY, a potent adenylate cyclase, increasing the intracellular cAMP level, which increased activity of cAMP dependent lipases and increased carnitine synthesis.

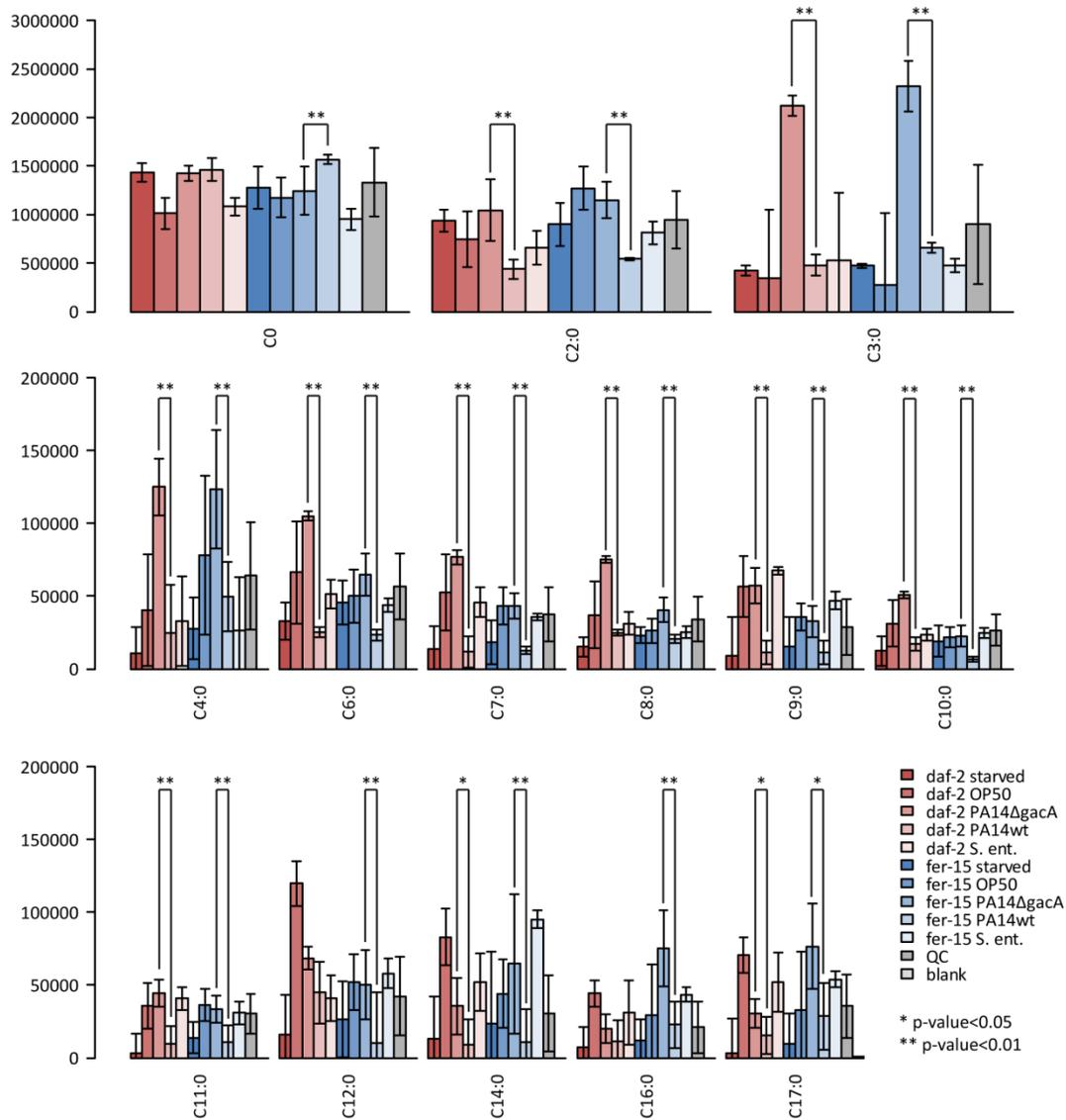


Figure 41: Acylcarnitines are affected by infection independent of *C. elegans* genotype

Carnitine was found in higher levels in *P. aeruginosa* PA14 wt infected *fer-15* worms, whereas acylcarnitines are lower compared to PA14 ΔgacA fed worms. Also starvation shows lower levels of acylcarnitines.

Additionally to acylcarnitines with saturated fatty acids, species with unsaturated chains were detected. Most of them were not significantly different, except two C18:3 species. These two species, corresponding to α - and γ -linolenyl-carnitine, could be base line separated and show different behavior. While the γ -isoform is higher in infection, the α -isoform is lower.

Nandakumar et al previously showed that γ -linolenic acid together with stearidonic acid is needed in *C. elegans* for immunity against *P. aeruginosa*. Stearidonyl carnitine was also detected, being lower in infected *C. elegans*.

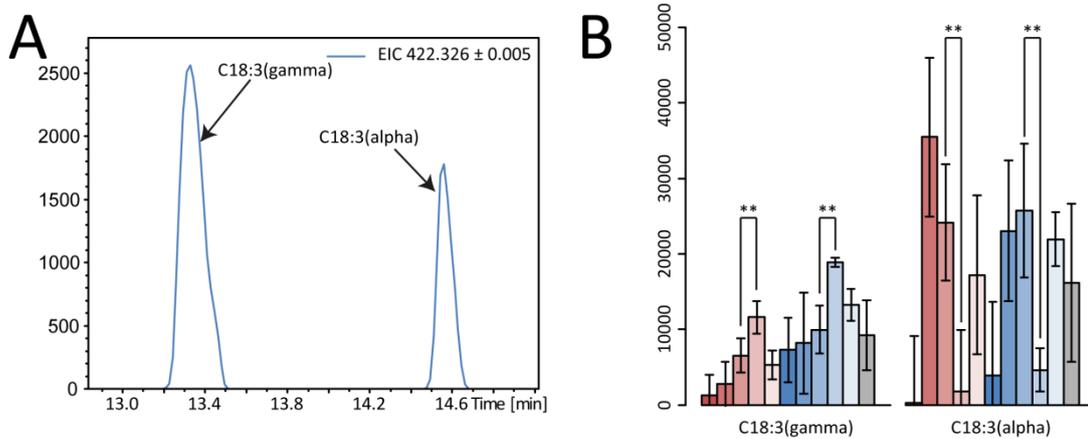


Figure 42: Two different isomers of C18:3 Carnitine were detected

(A) Extracted ion chromatogram corresponding to C18:3-carnitine, showing baseline separation of the two isomers. The γ -isomer elutes first due to its more bended fatty acid side chain, compared to the α -isomer. (B) Peak areas of the two different isomers. The γ -isomer has higher levels in infected worms compared to the α -isomer, which is showing the opposite pattern. (** p-value < 0.01)

Finally, the biosynthesis and degradation of carnitine was target of further investigations. All precursors from trimethyl lysine, hydroxytrimethyl lysine, 4-trimethylammoniumbutanal and 4-trimethylammoniumbutanoate were detected. From the degradation products, 3-dehydrocarnitine could be found. Some metabolites show different levels, but exact patterns could not be explained. Markly, the direct precursor of carnitine, 4-trimethylammoniumbutanoate is higher in PA14 wt infected worms. It seems that during infection *C. elegans* synthesizes more carnitine to further enhance the transport of fatty acids. In *E. coli* it was shown that cAMP mediates the induction of carnitine biosynthesis [132, 133]. Possibly this is also the case in *C. elegans*. 3-Dehydrocarnitine is a degradation product of carnitine that can be found in prokaryotes. It is further metabolized to glycinebetaine and finally to glycine. It could be detected in *C. elegans* suggesting that the nematode has a functional carnitine dehydrogenase, which has been until now not described in *C. elegans*.

Using the BLAST tool on the Wormbase homepage, several candidates for a possible carnitine dehydrogenase could be identified. One of the best homologies was found to the carnitine 3-dehydrogenase from *Burkholderia sp.* KJ006. Two isoforms a and b of the protein Y71F9B.9 were found with an E-value of $5e-32$. Both amino acid sequences were search for domains using the Pfam database. The two domains 3HCDH_N and 3HCDH, also present in the carnitine 3-dehydrogenase from *Burkholderia sp.* KJ006 were found in both isoforms. Looking to the corresponding gene in KEGG, it is annotated as hypothetical protein and possible l-

gulonate 3-dehydrogenase. Anyway, all evidences indicated a possible 3-dehydrogenase activity, which may also convert carnitine to 3-dehydrocarnitine. More experiments, for example creation of a deletion mutant and/or purification of the protein are needed to confirm this activity, but was out of the scope of this work.

In contrast to this, *S. enterica* infected worms have higher levels of sugar metabolites and breakdown products of glycogen and similar levels of acylcarnitines to PA14Δ*gacA* or OP50 fed worms. The above mentioned work of Antunes et al. suggested energy mobilization from sugar and glycogen storages during infection with *S. enterica* [92]. The obtained data confirms this idea, because lipid storages were not affected. Statistical analysis showed furthermore higher levels of D-glucose and D-ribose, glycerone phosphate and dihydroxyacetone phosphate in *fer-15* worms infected with *S. ent.* UDP-glucose as major precursor of glycogen, showed no changes, suggesting that higher levels of sucrose and glycogen are derived from degradation of storage sugars. These are broken down and fed into the TCA cycle via the glycolysis pathway as shown in 4.3.3.

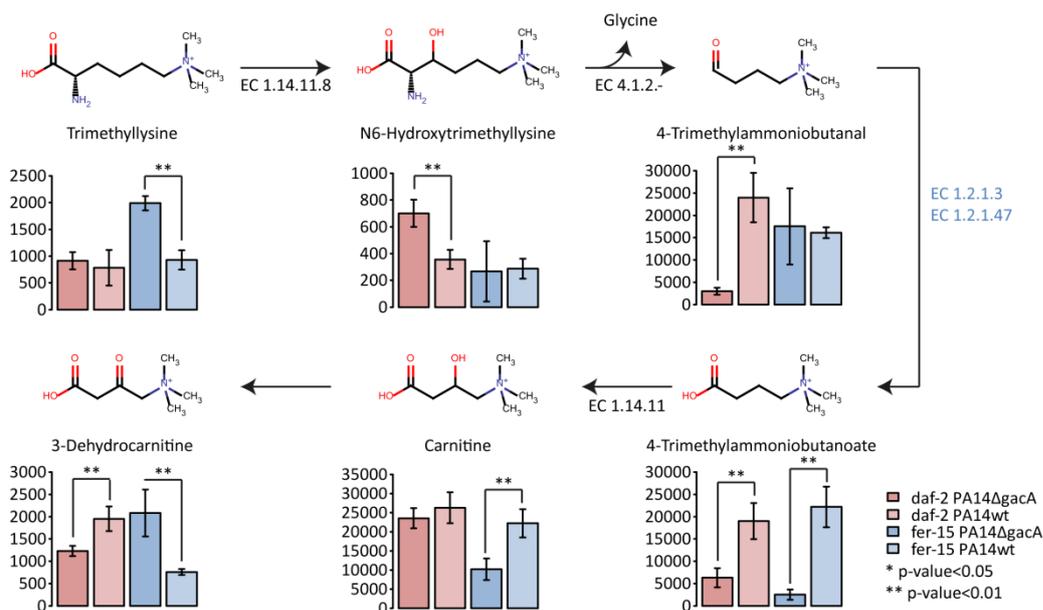


Figure 43: Carnitine synthesis and degradation pathways in *C. elegans*

L-carnitine is synthesized from trimethyllysine, obtained from degraded proteins. No specific pattern for infection was observed until the direct precursor of 4-trimethylammoniobutanoate, which is significantly higher PA14wt infected worms. The degradation product 3-dehydrocarnitine is lower in *fer-15* worms infected with PA14wt, in contrast to *daf-2* worms, which possible indicates a *daf-2* dependent degradation of l-carnitine.

4.3.5 Synthesis of phospholipid species during *Pseudomonas* infection

Beside lower levels of triacylglycerols, lipidomics analysis showed that PA14wt infected worms have lower amount of phosphocholine lipid species, whereas phosphoethanolamine and phosphoserine lipids remained unchanged. In both, *daf-2* and *fer-15* worms, the PC/PE

ratio changed from PC's a predominant species in PA14Δ*gacA* fed to PE's in PA14wt fed *C. elegans*. PC species are possibly cleaved by ExoU, a phospholipase A2 enzyme secreted by *P. aeruginosa* or synthesis of PC species is altered during infection.

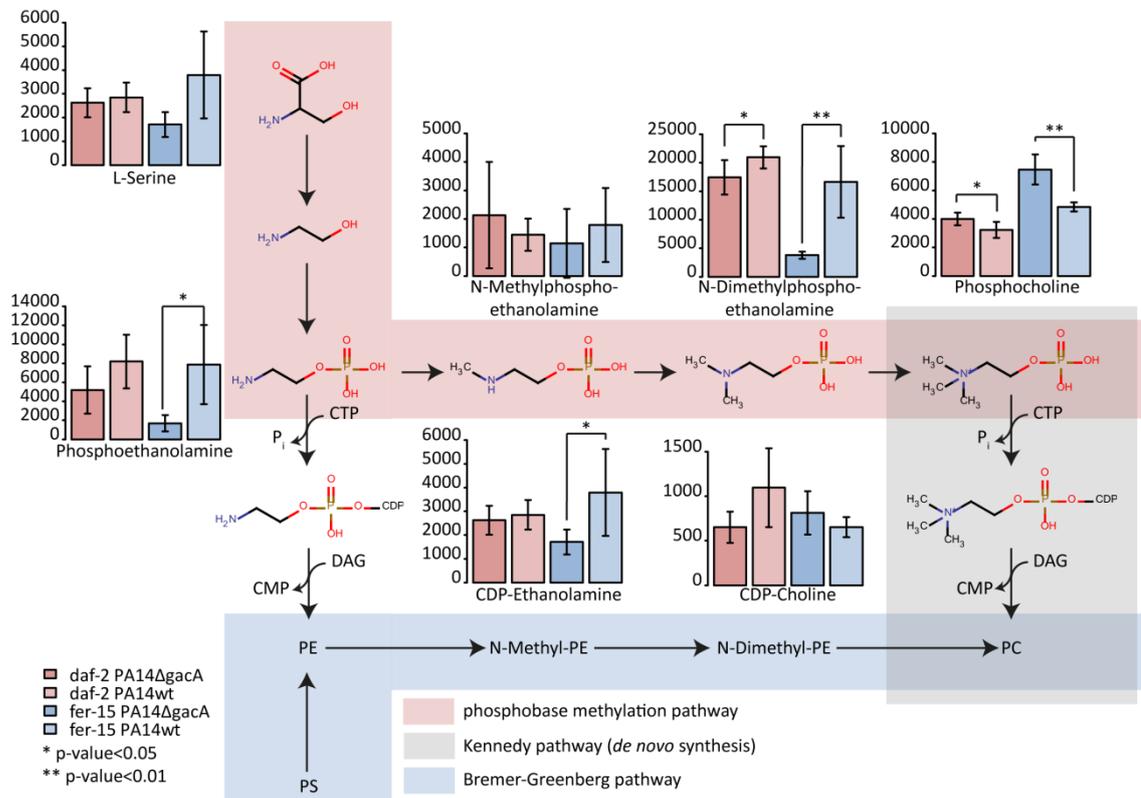


Figure 44: Synthesis of different phospholipid species in *C. elegans*

In *C. elegans* three different pathways for the synthesis of PC's exist. The Kennedy pathway uses Choline from the nutrition and couples it to CDP-DAG. The Bremer-Greenberg pathway subsequently methylates PE to PC. The phosphobase methylation pathway methylates Phosphoethanolamine to Phosphocholine, which is afterwards coupled to CDP-DAG. The obtained data suggest a cut-off of the phosphobase methylation pathway during *P. aeruginosa* infection. Choline levels were unchanged. PE species are synthesized normally. This leads to a shifted PC/PE ratio during infection.

C. elegans has three different pathways for the synthesis of PC's. The easiest is the Kennedy pathway which uses choline from dietary intake and couples it to DAG via activation with CDP. The other two pathways rely on methylation of ethanolamine, either before or after its reaction with DAG. The reactions are catalyzed by two enzymes, PMT-1 and PMT-2. Several metabolites found in the synthesis of PC species were found in BEH Amide separation with positive and negative ionization. The obtained results suggest a cut-off of the phosphobase methylation pathway during the last methylation step to yield phosphocholine. In contrast synthesis of PE species is enhanced in infected worms. Choline is lower in *daf-2* and *fer-15* worms infected with PA14wt, possibly due to lower uptake during infection. Trends for all other metabolites are similar between *daf-2* and *fer-15* worms, whereby in the *daf-2* mutants, all effects seem to be damped. To proof that this shift is valid, lipids have to be quantified to calculate molar ratios of single lipid classes. However, levels of lipid precursor strengthen this

theory. This lipid remodeling possibly aims to change properties of the *C. elegans* cell membranes for easier injection of bacterial effectors or infiltration of deeper tissues.

4.3.6 Metabolic markers of increased autophagy

Autophagy, the process known as self-eating, is an essential part of cellular remodeling and energy generation during starvation. Jia et al. discovered that autophagy is required for protection against the intracellular pathogen *Salmonella typhimurium* [134]. Additionally Irazoqui et al. found abnormal autophagosomal activity after 48 h of infection with *P. aeruginosa* PA14 [135]. Common markers for autophagy are modified amino acids like L-hydroxyproline or o-phosphoserine. The posttranslational modifications occur in the protein and levels of corresponding free amino acids can indicate increased autophagy. One of the first hits found was glycylprolylhydroxyproline, which was discovered to have higher levels in *S. ent.* fed *fer-15* worms (p-value = 8.4E-05), whereas levels in PA14 wt and PA14 Δ *gacA* were similar. A common motif found in the amino acid sequence is Gly-Pro-X and Gly-X-Hyp (X = any other amino acid except glycine, proline or hydroxyproline, Hyp = hydroxyproline). Glycine is the most common amino acid in collagen. Several unknown masses significantly higher in *S. ent.* fed worms were annotated as possible glycine containing tri-peptides using Metlin. A second indicator for increased autophagy can be choline, derived from PC's. Again *fer-15* worms fed with *S. ent.* have significantly higher levels (p-value = 0.00051). Lastly spermidine had increased levels in *S. ent.* fed *fer-15* worms (p-value = 0.0003). Spermidine is known to induce autophagy in *C. elegans* and is possibly a mediating molecule to signal increased autophagy. In *daf-2* worms most of these effects are lost. *daf-2* worms generally have higher autophagic activity [136], where possibly all this metabolites are further degraded, which is not the case in *fer-15* animals. Hansen et al. described that autophagy and DAF-16/FOXO are needed for lifespan extension. Because autophagy still takes place in a *daf-2/daf-16* double mutants, *daf-16* seems not be directly involved in autophagy. However for lifespan extension DAF-16/FOXO is needed and is possibly involved in further breakdown and recycling of degradation products [137].

No increase of metabolic autophagy marker where found in PA14wt fed worms, which is conform with the results from Irazoqui et al., which observed autophagosomes at 48 h post infection [135].

4.3.7 Resistance of *daf-2* worms to infection

One of the major questions is the resistance of *daf-2* worms to *P. aeruginosa*. Metabolites that show a dependency on *daf-2* are good candidates to mediate the resistance. To identify such metabolites peaks with opposite behavior in *fer-15* and *daf-2* nematodes fed on PA14 wt were searched in all datasets. Because *daf-2* is up regulated in PA14wt infected *C. elegans*,

metabolites that are *daf-2* dependent are either repressed or induced in infection. This effect is erased in *daf-2* deletion mutants. *fer-15* worms fed with PA14 Δ *gacA* serve as control. Furthermore metabolites found in high concentrations in *daf-2* worms infected with PA14wt and lower in *daf-2* mutants fed with PA14 Δ *gacA* and *fer-15* worms fed with either PA14 wt or PA14 Δ *gacA* possibly mediates pathogen resistance. Totally in BEH Amide separation with positive and negative ionization 3235 and 1536 and in BEH C8 separation 8682 and 581 showed this behavior respectively.

One of the metabolites identified by this analysis is adenosine. *P. aeruginosa* fed worms have low levels of adenosine except *daf-2* worms fed on PA14wt. The used QSRR model predicted a retention time of 5.4 ± 1.5 minutes for adenosine, the real retention time in the samples was 6.8 minutes. An extracted ion chromatogram corresponding to adenosine is shown in Figure 45. Retention time was confirmed with an authentic standard. Indeed it was shown that adenosine disturbs biofilm formation of *P. aeruginosa* PA14. Furthermore it reduces pyocyanin, extracellular polysaccharide, quinolone signaling and siderophore production. The gene expression response of *P. aeruginosa* to adenosine shows repression of 79 genes involved in iron acquisition. Although it was thought that adenosine might bind to the Ferric uptake regulator (Fur), this is not the case, as shown by EMSA [138]. It is possible that adenosine is secreted into the lumen of *C. elegans*. Markly only PA14wt fed worms show higher levels, which suggests that deletion of *daf-2* together with another response to infection is needed. Consistent with this, *daf-2* mutants fed on *S. enterica* have also higher adenosine levels than *fer-15* worms.

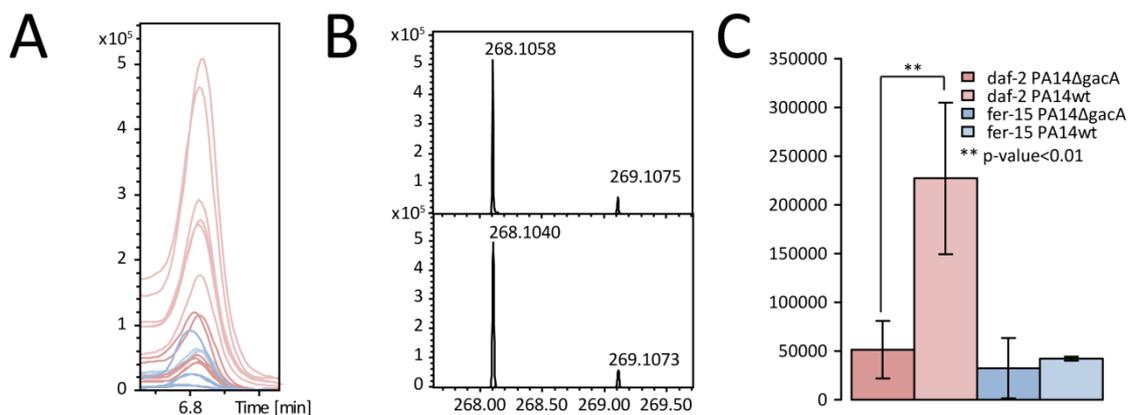


Figure 45: Adenosine possibly mediates resistance in *daf-2* worms

(A) Extracted ion chromatogram of adenosine. The peak is distorted at the beginning; this shape is also seen by injecting an authentic standard. The distortion is maybe due to different protonation, at a pH of 5.1 adenosine shows 3 different micro species. (B) Mass spectrum of adenosine at peak apex (upper panel) compared to theoretical isotopic pattern (lower panel). (C) Peak areas of adenosine in worms fed with *P. aeruginosa* show higher levels in *daf-2* mutants infected with PA14wt.

The question that arises from this data is if adenosine is synthesized *de novo* in a higher rate or it originates from AMP for example. Looking again into gene expression data from Ausubel et al. the first reaction forming 5-phosphoribosylamine from l-glutamine and 5-phospho-alpha-D-ribose 1-diphosphate is repressed. Gene expressions of all other enzymatic genes seem not to be altered. So it is likely that adenosine is produced from AMP. Unfortunately, the precursors of adenosine were not detected or were close the noise. Beside its known disruption of *P. aeruginosa* biofilms, adenosine is more generally known as danger signal, signaling tissue and cell disruption.

4.3.8 Potential infection biomarker

ICR-FT/MS and BEH C8 analysis with negative ionization revealed a mass (470.23136), which is specifically increased in *C. elegans* fed with pathogens. No corresponding mass was detected in positive ionization mode, which is a hint for a structure with highly negative groups. No annotation for this mass was found by MassTRIX or Metlin. Using the NetCalc approach the putative formula of $C_{23}H_{38}NO_7P$ (neutral molecule) could be annotated to this mass. Searches in Pubchem and ChemSpider revealed N-arachidonoyl-phospho-l-serine as possible chemical structure. This molecule shares a common substructure with N-acylethanolamines, which have been shown to mediate the effect of diet on lifespan of *C. elegans* [139].

N-arachidonoyl-phospho-l-serine is not commercially available and was synthesized from arachidonic acid and phospho-l-serine for confirmation. First, arachidonic acid was activated with N-hydroxysuccinimide (NHS) and dicyclohexylcarbodiimide (DCC). 500 mg arachidonic acid were added to a solution of 239 mg NHS in 10 ml dry ethyl acetate. After five minutes of stirring 438 mg DCC in 2 ml dry ethyl acetate were added and the mixture was stirred overnight under a N_2 atmosphere. During the reaction the solution became turbid due to precipitation of dicyclohexylurea. After filtration and evaporation the residue was dissolved in $CHCl_3$ and chromatographed on a custom made glass column filled with 5g Silica gel 60. Fractions were checked for purity using TLC. Pure fractions were mixed and $CHCl_3$ was evaporated under a gentle nitrogen steam to yield a colorless oil. The oil was dissolved in 1.5 ml THF and added to 911 mg phospho-l-serine in 8.5 ml 0.1 M Na_2CO_3 / 0.1 M $NaHCO_3$. The mixture was allowed to react to overnight under N_2 atmosphere. The reaction mixtures was dried to a volume of about 5 ml and acidified to pH = 1 with 1 N HCl. The product was extracted two times with 10 ml dichloromethane. Solvent was evaporated with a gentle stream of nitrogen and the residue was dissolved in MeOH and purified on a Waters BEH C8 column with a MeOH gradient. Fractions corresponding to N-arachidonoyl-phospho-l-serine were collected and solvent was evaporated in a Speed Vac vacuum concentrator.

window was set to a smaller window, ion transmission was not sufficient to obtain a clear tandem mass spectrum. However, even if fragments of PE(O-18:0/0) are major masses detected, also peaks corresponding to N-arachidonoyl-phospho-l-serine could be found (m/z 328 and 372). For the standard a retention time of 14.5 minutes was determined, whereas the corresponding peak in the sample showed a retention time of 15.0 minutes. Because the metabolite extract represents a highly complex matrix and retention time can differ from pure chemical substances, the standard was spiked into the sample. The spiked sample showed two peaks with a similar fragmentation pattern at 14.5 and 15.0 minutes, suggesting that the peak observed in the samples is not N-arachidonoyl-phospho-l-serine. Because retention time differs only slightly and MS/MS pattern are similar, a similar structure with an isomeric C20:4 fatty acid side chain could be the true structure.

Further work has to be carried out, to reveal the true structure of this unknown. However, retention time and MS/MS pattern show that the synthesized N-arachidonoyl-phospho-l-serine is close to the real structure. Possible other side chains are 5(E)-arachidonic acid or ω -3-arachidonic acid. Finally, in the synthesized standard, a small peak at the retention time of the unknown substance is visible, which possibly derives from another C20:4 fatty acid impurity in the used arachidonic acid. Another possibility for structure determination is purification from raw *C. elegans* extracts using preparative chromatography. For this approach not enough material to obtain an amount sufficient for NMR analysis was available.

After final structural elucidation of the true structure, possible biosynthetic routes have to be elucidated. Fatty acid amides can be obtained by the condensation of an amine with a fatty acid, but mostly they are produced by condensation of fatty acids with the terminal nitrogen of a PE or PS and subsequent cleavage of the ester bond to the phosphate group. Possibly also this molecule is synthesized by the second pathway. The major question is, if the phosphate group is derived from the phospholipid or is added after cleavage.



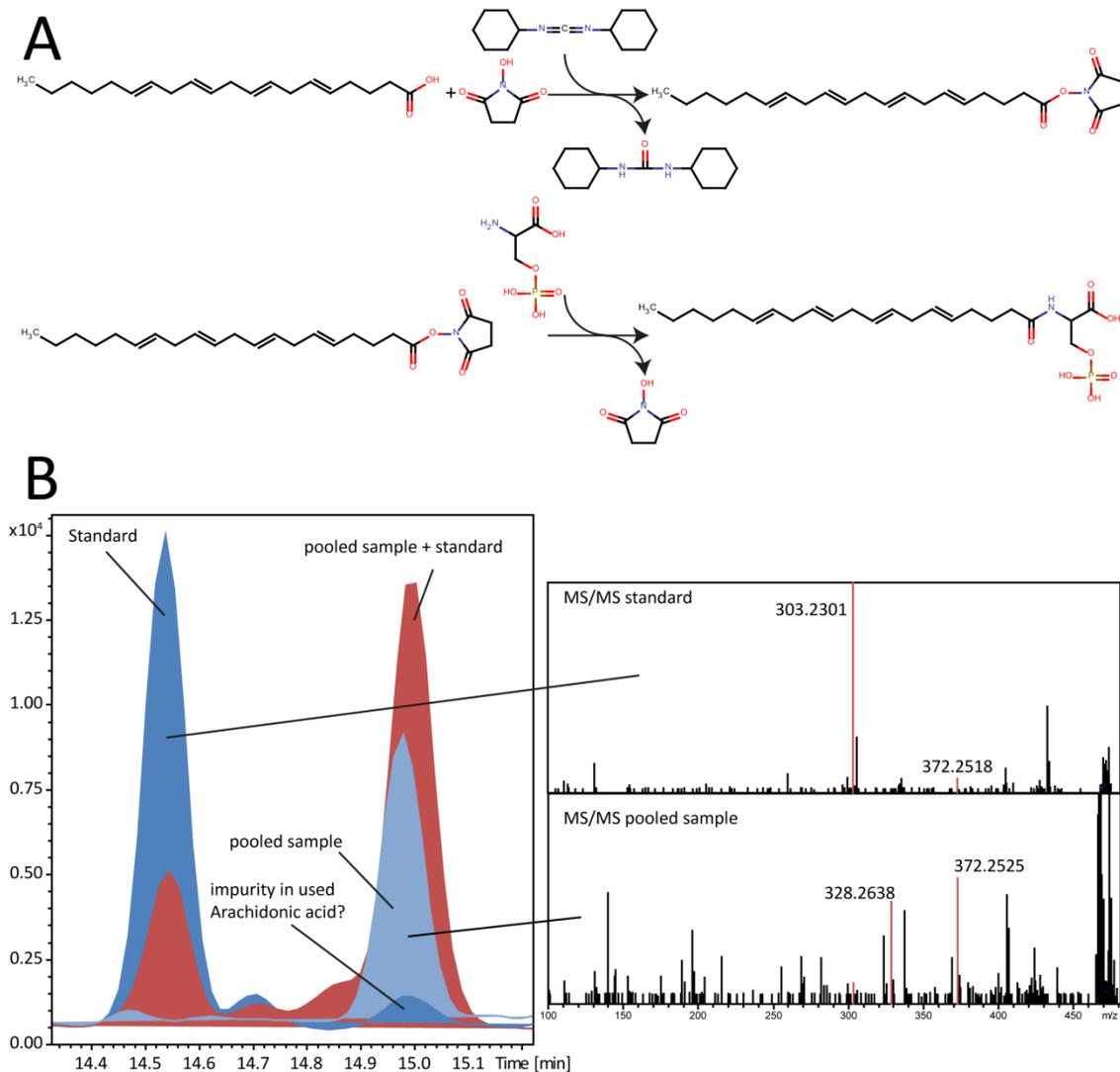


Figure 47: Synthesis strategy and results from N-arachidonoyl-phospho-l-serine synthesis

(A) Reaction scheme of the synthesis strategy. In the first step arachidonic acid was activated as NHS-Ester with the help of DCC. Phospho-l-serine was coupled to arachidonic acid NHS-Ester in a basic buffer to deprotonate all functional groups and to form N-arachidonoyl-phospho-l-serine. (B) Chromatogram and tandem mass spectra reveal that the unknown peak is not N-arachidonoyl-phospho-l-serine, although MS/MS pattern are similar. In the pooled sample, the ionization is suppressed by the co-eluting PE(O-18:0/0). Most of the fragments correspond to this molecule. Direct infusion experiments of the synthesized standard showed that the main fragments of N-arachidonoyl-phospho-l-serine are 303, 328 and 372. All of them could be detected in the pooled sample.

4.3.9 Metabolic signatures in long-lived *C. elegans*

An experiment which was carried out several times in literature using NMR was the comparison of long-lived *daf-2* mutants versus wildtype *C. elegans*. The experimental setup of this work allowed comparing *daf-2* and *fer-15* worms fed on *E. coli* OP50 to find metabolic signatures of long life using mass spectrometry, beside the original topic of finding metabolic alterations in *P. aeruginosa* infection. One of the most significant changes is found in branched chain amino acids leucine, isoleucine and valine, which is already known from previous work [72, 73]. In contrast to these publications, which were carried out with NMR detecting only the most abundant metabolites, the used mass spectrometric approach allowed a deeper

understanding. Interestingly, *daf-2* worms have higher levels of biosynthetic precursors of these branched amino acids and almost no degradation products. Since *C. elegans*, like other animals is not able to synthesize these amino acids [140], the precursors have to originate from *E. coli* OP50 and were ingested.

The endogenous polyamine spermidine has been described to promote longevity by the induction of autophagy. In the work of Eisenberg et al. *C. elegans* had an enhanced survival rate of about 15% with a p-value < 0.0001, when fed additionally with 0.2 mM spermidine in the medium. However, this effect was abolished in *bec-1* RNAi animals, deficient in autophagy induction [141]. Spermidine is synthesized from ornithine via putrescine. Ornithine and N-acetylputrescine, together with two other polyamines agmatine and cadaverine, were found in lower levels in *daf-2* worms. Putrescine was unchanged and spermidine was found in higher levels. Possibly, higher levels of spermidine may contribute to the longevity of *daf-2* mutants.

Generally, *daf-2* worms are described in literature to have a lower metabolism and increased fat storage. From the lipidomic data set, it could be seen that *daf-2* worms have higher levels of triacylglycerols.

4.3.10 Role of *P. aeruginosa* *pec7* during infection

The gene *pec7* was discovered by Rafat Zrieq at the Heinrich Heine University in Düsseldorf in the Pathomics project as possible new effector gene by screening the whole *P. aeruginosa* PA14 genome for abnormal growth of yeast [personal communication]. Totally 54 genes were identified using this approach, out of this 8 were selected for further experiments (*pec1* to *pec8*). The encoded protein Pec7 is probably a SUFU-like protein, responsible for expression of diverse genes in the nucleus. Das et al. discovered a SUFU-like protein in *Neisseria gonorrhoeae* [142]. Until now, no clue about the function of Pec7 in *P. aeruginosa* exists. However, using a *C. elegans* slow killing assay showed a significant reduced pathogenicity of a PA14Δ*pec-7* mutant, but the effect was not as strong as with PA14Δ*gacA*. Furthermore, co-localization of Pec7 with lipid droplets was shown by Rafat Zrieq [personal communication]. Due to this fact, a lipidomics analysis was carried out on *daf-2* and *fer-15* *C. elegans* fed with PA14wt, PA14Δ*gacA* or PA14Δ*pec7*. In PCA analysis the PA14Δ*pec7* fed worms clustered close to PA14wt fed ones independent of the worm genotype, as shown in Figure 48. First analysis showed that the observed shift in the PE/PC ratio between PA14Δ*gacA* and PA14wt, is still valid between PA14Δ*gacA* and PA14Δ*pec7*.



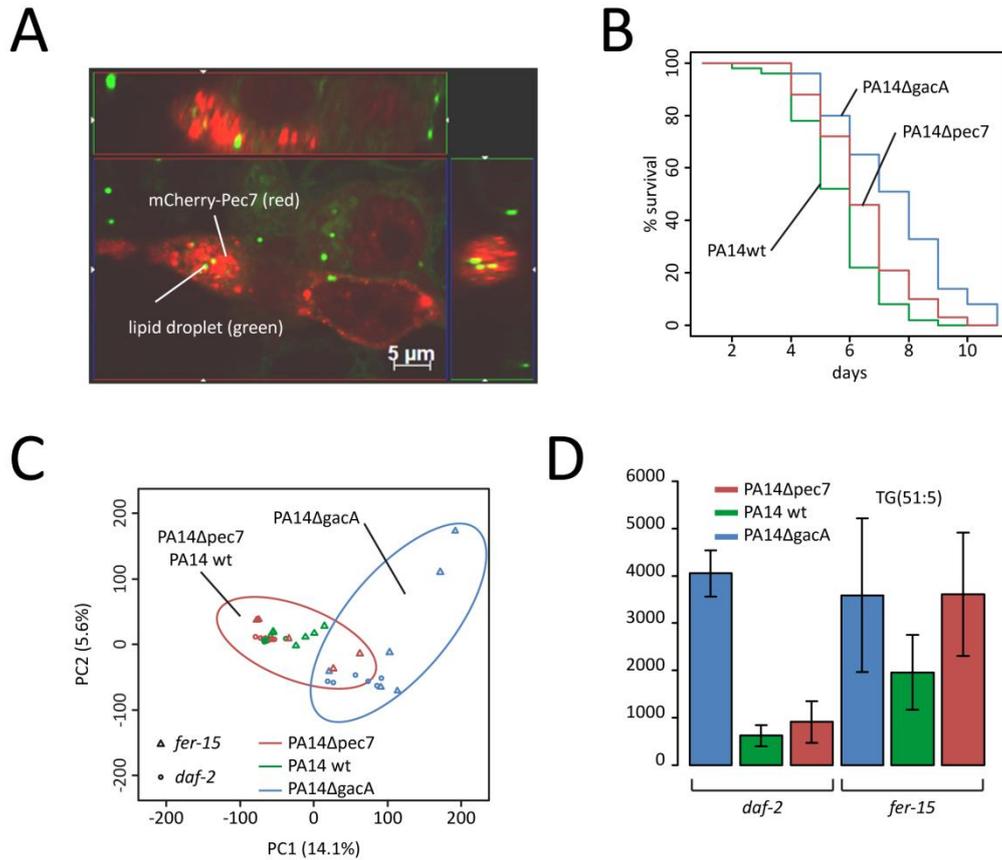


Figure 48: Pec7 is a possible novel bacterial effector or protein needed for virulence

(A) Co-localization of Pec7 and lipid droplets shown by confocal microscopy. Pec7 was expressed as mCherry-tagged protein and lipid droplets were visualized with BODIPY (picture courtesy of Dr. Rafat Zrieq, HHU Düsseldorf) (B) Killing curve of slow killing assay comparing PA14wt, PA14 Δ gacA and PA14 Δ pec7. PA14 Δ pec7 fed worm's significant enhanced survival compared to PA14wt, but the effect is not as pronounced as for PA14 Δ gacA (C) PCA from positive ionization lipidome data shows that PA14 Δ pec7 behaves like the wildtype (D) Example of one lipid that was found different between PA14wt and PA14 Δ pec7. However no clear differences or direction was found for to discriminate between the two.

To reveal differences between PA14wt and PA14 Δ pec7 levels of lipid were compared based on their log-foldchange and p-value. 1134 features showed a significant difference between the two sample states. One example is shown in Figure 48D. However, the different masses gave no clear trend, because annotations were from across all lipid species and levels were at lower range of the intensity scale. It is likely that these results are observed by chance. Still the question is open, why Pec7 does co-localize with lipid droplets. The *C. elegans* model might be the wrong approach to test possible alterations in lipid droplet composition, because lipid droplets disappear fast during infection. This is also the case for the PA14 Δ pec7 mutant, allowing no comparison of TG composition. A better approach would be to express *pec7* in HeLa or yeast cells to determine their lipid composition.

4.3.11 Probiotic bacteria protects *C. elegans* from *P. aeruginosa* infection

Probiotic bacteria are known to offer positive effects in different diseases. *C. elegans* recently entered the field of probiotics research. *B. subtilis* for example is known to enhance

the lifespan of the nematode. An example was published by Wang et al., where *C. elegans* was used to preselect *Lactobacillus* isolates to control *Salmonella typhimurium* infection in piglets [143]. This work led to the suggestion that probiotic bacteria can offer positive effects in *P. aeruginosa* infection. 32 probiotic strains were provided from Winclove Industries in the Netherlands, to test them in *C. elegans* slow killing assays.

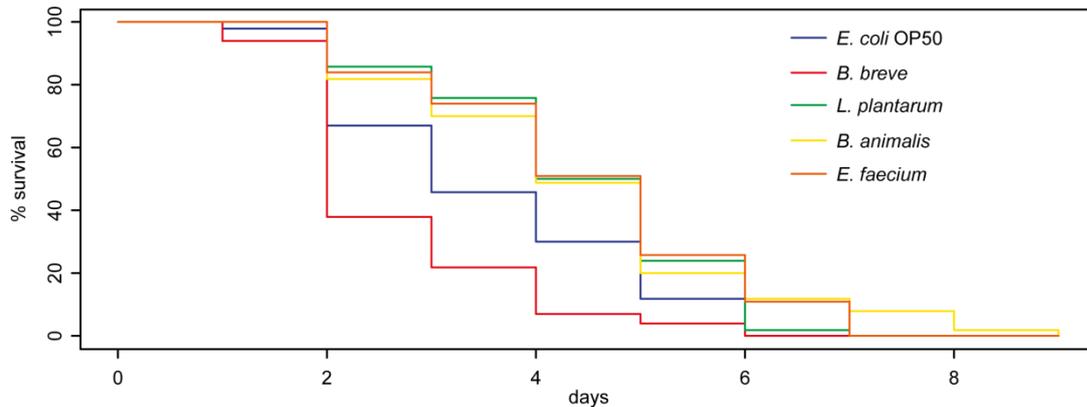


Figure 49: Killing curves of *C. elegans* pre-fed with probiotics

C. elegans were grown on *E. coli* OP50 and afterwards fed for 24 hours on probiotic strains. After this time, diet was switched to *P. aeruginosa* PA14 and survival was monitored. For all four strains, survival was significantly different from *E. coli* OP50 with p-values < 0.01.

C. elegans were grown on *E. coli* OP50 until adulthood and switched to the different probiotic strains for 24 hours. After this pre-feeding, *C. elegans* food was changed to the pathogenic *P. aeruginosa* PA14 and survival was evaluated as described in chapter 4.3.2. Out of the tested strains one, *Bacillus breve*, showed a negative effect and three, *Lactobacillus plantarum*, *Bacillus animalis* and *Enterococcus faecium* significantly enhanced *C. elegans* survival on *P. aeruginosa*.





5.1 HeLa cells as host organism for *P. aeruginosa* infection

To examine metabolic effects of infection on HeLa cells ICR-FT/MS was conducted for non-targeted profiling. For this purpose a metabolite extraction procedure was developed for both HeLa cells and *P. aeruginosa*. In parallel Steve Garvis developed a dedicated sample preparation to separate both organisms after infection and reducing the bacterial load in HeLa cell samples. HeLa cells were infected with *Pseudomonas aeruginosa* PA14 for a half hour. Because *P. aeruginosa* PA14 belongs to the cytotoxic strains of *P. aeruginosa* and cell lysis is observed fast after infection. After the infection a mixture of bacteria, lysed and unlysed HeLa cells are found in the medium. The final method for clean-up is a good compromise between fast sample preparation needed in metabolomics and the reduction of bacteria in the sample.

In the second part of this chapter HeLa cells expressing ExoY, a T3SS effector of *P. aeruginosa* were subjected to metabolic profiling to reveal metabolic alterations caused by this enzyme.

5.2 Sampling and sample preparation

Following passage discuss the needs for a sample preparation in metabolomics. Care was taken to apply all this needs to the developed method.

5.2.1 Prerequisite for metabolomics samples

Like for all other analytical chemistry applications, also in metabolomics sample preparation and clean-up is the essential first step. Most metabolomics applications are starting from cell pellets or a piece of tissue. Protocols for sample preparation are dependent on the cell-structure of the biological sample. Plant cell walls need a different treatment than a mammalian cell culture sample. One fact is similar for all protocols: After sample preparation the sample should be highly enriched and pure in the targeted analytes. For metabolomics, especially non-targeted metabolomics, this is a tough task, as different classes of molecules will be analyzed simultaneously and each shows different physicochemical parameters and concentrations, which are ranging over several orders of magnitude from pico- to millimolar. It should be mentioned that a truly non-targeted extraction doesn't exist, as an extraction system is always more discriminant for some metabolite class as for another. Lastly, an optimal sample preparation method should include only as minimal steps that are needed to achieve this goal, to keep the samples as close as possible to its native state.

Major task in metabolomics sample preparation are quenching of metabolism, lysis of cells, extraction of metabolites and removal of interfering substances, e.g. proteins and salts.



5.2.2 Quenching of metabolism

Metabolome analysis can be compared to a photo, which is a static picture of a dynamic environment. Metabolism is highly variable and time and condition dependent. Metabolites from the primary metabolism are produced and consumed with high turnover rates, e.g. 1.5 mM/s for ATP [144, 145], whereas secondary metabolites accumulate in the cell or are excreted. To obtain a snapshot of a system physiology, metabolism has to be stopped quickly, meaning inactivation of metabolic enzymes. This can be carried out by using organic solvent (hot or cold), extreme pH conditions ($\text{pH} < 3$ or $\text{pH} > 11$) or snap-freezing in liquid nitrogen. A problem is the separation of intra- and extracellular matrix. Methods used in microbial metabolomics, like spraying culture into cold methanol leads to mixture of both matrices. This is especially complicated if rich culture media with unknown exact formulation are used. These can contaminate the sample too much, which makes a reliable data interpretation impossible. Separation by centrifugation or filtration may alter the cellular metabolome. A trade-off has to be found between both possibilities.

5.2.3 Cell lysis techniques and metabolite extraction

After the metabolism has been stopped cells have to be lysed to get access to the intracellular metabolites. Different protocols exist for this. Roughly separated mechanical and non-mechanical methods exist, including ultrasonic, grinding, enzymatic or chemical lysis or osmotic shock. During the whole process attention to temperature in the sample should be given. Most mechanical methods heat the samples up, so lysis has to be carried on dry ice or ice. In case of chemical lysis some metabolites classes will be degraded or converted to other forms [145]. Together with the lysis often metabolites are directly extracted into an appropriate solvent. For metabolites from primary metabolism and polar to mid polar secondary metabolites solvent mixtures of water and a miscible organic solvent are used with concentrations $\geq 50\%$ organic. Commonly methanol, ethanol, acetonitrile or isopropanol are used. Using organic solvents precipitates proteins, a major interference in metabolomics studies. If more non-polar substances like lipids should be extracted isopropanol, chloroform or methyl-tert-butyl ether (MTBE) can be used. A method to obtain a total lipid extract from solid material was described by Folch using a chloroform/methanol (2/1) mixture [146], while the Bligh and Dyer method is used for lipid extraction from aqueous samples [147]. Both methods are using chloroform for extraction, meaning the lipid rich layer after centrifugation will be the lower phase, making it hard to automate this procedure. An alternative was described by Matyash using MTBE. In this method the organic solvent forms the upper layer is useful for automation on robotic system. Extraction yields are comparable to the other mentioned methods [148].

5.2.4 Solid phase extraction and sample concentration

Bacteria and mammalian cells are normally cultivated in nutrient rich and highly salty media. If the exometabolome of an organism should be studied this can lead to problems, especially in DI-MS. Precipitating salts or proteins can cause clogging of the ESI sprayer yielding an unstable spray current and non-reproducible measurements. Therefore interfering salts and other substances have to be removed from the sample. Such a clean-up can be performed with solid phase extraction (SPE). The principal is similar to chromatography and is based on the distribution of analytes between a solid and mobile phase. Analytes of interest are trapped on a suitable solid phase and interfering substances are washed away. Afterwards compounds are eluted with a suitable organic solvent. Several materials for SPE exist, including reversed-phase materials, ion exchanger and mixed mode. As base material mostly silica gel is used, but polymer based material are spreading more and more. SPE is not only useful for removal of interfering salts, but also for targeted analysis and clean-up of the targeted compounds and their concentration by using a smaller elution volume compared to original applied sample volume. Also other methods for concentration of a sample can be used if necessary, e.g. lyophilisation, gentle streams of nitrogen or vacuum centrifuges to remove solvents. Attention has to be drawn to this step, because some analytes may get lost during this procedure. With this method also starting conditions of a sample can be optimized for a specific analytical method, e.g. the change to deuterated solvents for NMR. Also sensitivity is increased if the final volume is smaller than the previous sample volume.

5.3 Development of an extraction method for a HeLa cell infection model

5.3.1 Metabolite extraction from bacterial pellets

For metabolite extraction several different extraction solvents were tested. All samples were prepared from aliquots of the same *P. aeruginosa* culture. Residual medium was removed by washing with 1 ml water in three steps. After addition of 400 μ l extraction solvent, samples were sonicated in an ice-cold sonic bath for 15 minutes. After centrifugation with 14,000 rpm at 4°C for 20 minutes, supernatant was transferred to a fresh eppendorf cup. To make different extraction solvents comparable, solvent was evaporated and the residue was reconstituted in 400 μ l 70% MeOH and stored at -80°C until analysis.

5.3.2 Metabolite extraction from cell culture medium containing HeLa cells

Different SPE materials were tested for metabolite extraction from medium. 1 ml cell culture medium was subjected to pre-washed 1 ml SPE column. Salts were washed away with

1 ml water and metabolites were eluted with 500 μ l solvent. Solvent was evaporated and the residue was reconstituted in 250 μ l 70% MeOH.

5.3.3 Results

A major problem working with host-pathogen system is the mixture of two different organisms. For transcriptomics analysis this is no problem as specificity is achieved through differences in sequences. In metabolomics such a differentiation is not possible, as long as no isotopic labeling of one species is used. This is a difficult task, because different isotopomers would increase the number potential features and data annotation and identification of metabolites, especially of unknown masses would become increasingly complex. Chapter 2 showed that several parts of the metabolism are shared between *P. aeruginosa* and humans, but also that each of the organism has special metabolic pathways. Problems arise from the shared metabolic pathways and unknown entities, yet measurement of independent cultures of both organisms separately under same conditions can help in data analysis to exploit the possible origin of an unknown metabolite.

A second problem arises from the used host-pathogen system. The used *P. aeruginosa* strain PA14 belongs to the cytolytic strains and HeLa cells are lysed quickly after contact with the bacteria. After the infection time, the cell culture contains a mixture of unlysed and lysed HeLa cells, bacteria and the culture medium. This mixture can be separated into two fractions by low speed centrifugation, the medium containing the HeLa cells and cell debris and a bacterial pellet. For each of this fraction a metabolite extraction procedure was developed.

5.3.4 Metabolite extraction from *P. aeruginosa* pellets

Following solvents and mixtures were tested for extraction of metabolites from *P. aeruginosa*: 100% water, 100% MeOH, 100% ACN, 50% MeOH and 50% ACN. To avoid differences based on different ionization, all samples were evaporated to dryness and reconstituted in 70% MeOH. The different extraction methods were compared using DI-ICR-FT/MS as described in the appendix. Table 17 gives an overview on the number of detected masses in the different extraction solvents and their overlap with other extraction solvents.

100% ACN yielded in both, positive and negative ionization the most features. Interestingly, 100% water has the highest count in unique features. Inspection of van-Krevelen diagrams showed that 100% ACN extracts mostly mid- to non-polar metabolites. To extract more polar metabolites, water has to be added to the extraction solvent. 50% MeOH showed the lowest overlap with 100% ACN, suggesting that a two-step extraction procedure with first 50% MeOH followed by 100% ACN is a suitable solution. Indeed 50% MeOH is able to extract metabolites

from the central carbon metabolism and other important biological pathways. This result is valid for both ionization modes.

Table 17: Detected masses in the different tested extraction solvents

The different extraction solvents were compared based on the total number of peaks in positive and negative ionization mode and on their overlap with other extraction solvents. Numbers in bold represent the highest numbers in total or unique peaks and underlined numbers the overlap between 100% ACN and 50% MeOH.

<i>ion mode</i>	<i>Solvent</i>	<i>Total peaks</i>	<i>Unique peaks</i>	<i>50% ACN</i>	<i>50% MeOH</i>	<i>100% ACN</i>	<i>100% MeOH</i>	<i>100% water</i>
<i>pos</i>	50% ACN	3590	299	-				
	50% MeOH	3447	340	1456	-			
	100% ACN	3644	393	1351	1307	-		
	100% MeOH	3027	366	1061	950	1204	-	
	100% water	2903	448	1035	919	955	800	-
<i>neg</i>	50% ACN	2820	265	-				
	50% MeOH	2672	435	1523	-			
	100% ACN	4378	531	1645	1476	-		
	100% MeOH	3653	394	1157	1177	2930	-	
	100% water	2578	664	1262	995	1161	940	-

5.3.5 Metabolite extraction from HeLa cells

After infection, HeLa cells are recovered as lysed and unlysed cells in the cell culture medium and the sample has to be homogenized. The sample was homogenized by lysing remaining intact HeLa cells using either sonic finger or sonic bath. The sonic finger was found to heat the samples too much; even when they were chilled on ice during cell lysis. Therefore it was excluded from further investigation and HeLa cells were lysed in an ice-cold sonic bath. Four different SPE phases, C2, C8, C18 and HLB, were tested for desalting of medium samples containing lysed HeLa cells. Again DI-ICR-FT/MS was used for comparison of the different methods. Table 18 compares the characteristics of these materials. C18, C8 and C2 are silica based reversed phase materials used in routine analysis and HLB is hydrophilic-hydrophobic balanced co-polymer. All phases were washed and pretreated in the same way. For elution either MeOH or ACN was used. A similar comparison like for the optimization of pellet extraction method was used. Figure 50 shows a comparison of the masses detected in the different methods for positive ionization mode. Similar results were obtained from negative mode.

Although the C18 phases yielded the highest amount of masses with MeOH elution, the HLB material with ACN elution was chosen for further work, due to two reasons. First, looking into van-Krevelen diagrams of the detected masses, the HLB material showed a higher dispersion of the polarity range and covered more metabolite classes than the C18 phase, which retained mainly lipophilic compounds. Second the HLB copolymer is allowed to dry out

between subsequent washing and elution steps, which mean that previous applied solvents can be removed more efficiently from the sorbent.

Table 18: Characteristics of used SPE materials

EC = end capping

Brand	Functional group	Base-material	Carbon load	EC	Surface area (m²/g)	Particle size (μm) and shape	Mean pore size (Å)
Varian Bond Elut C18	C18	Silica	17.4	y	500	40, irregular	60
Varian Bond Elut C8	C8	Silica	12.1	y	500	40, irregular	60
Varian Bond Elut C2	C2	Silica	5.6	y	500	40, irregular	60
Waters Oasis HLB	HLB (hydrophilic -lipophilic balance)	Co-polymer	-	-	-	30	80

5.3.6 Final methods

The described methods were optimized on single samples of each organism. In a real infection model, these organisms are mixed to mimic and infection. A method for separation of bacteria and HeLa cells was developed by Steve Garvis at the CNRS in Marseille and is depicted in the appendix. The method was finally applied to these samples and consisted of a two-step extraction with 50% MeOH, followed by 100% ACN for bacterial pellets and SPE using the HLB material with ACN elution for the HeLa cells. Because later analysis showed that lipids are mainly affected by infection, a lipid extraction was used as described in the appendix.

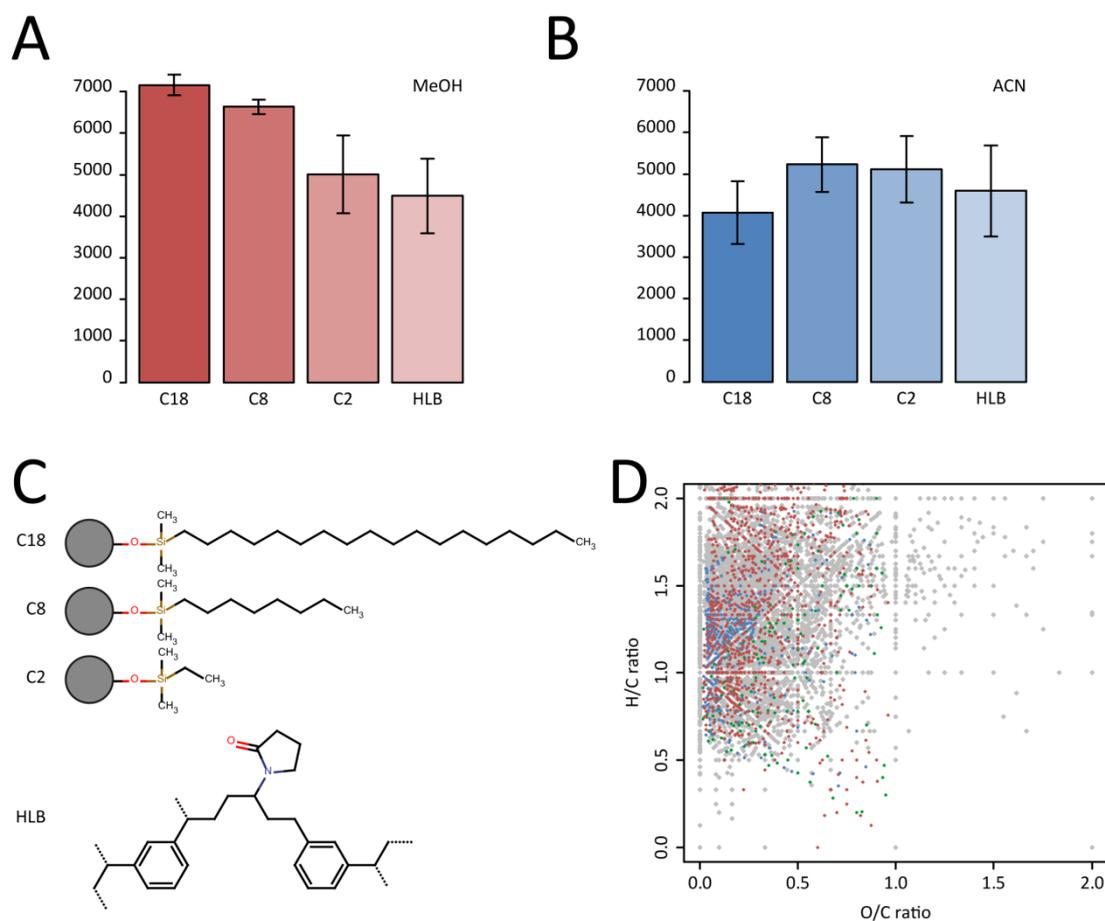


Figure 50: Results of method optimization for HeLa cell preparation

(A) Barplot of number of masses obtained in positive ICR-FT/MS analysis from elution with MeOH. (B) Barplot of number of masses obtained in positive ICR-FT/MS analysis from elution with ACN. (C) Structures of the used SPE material. C18, C8 and C2 are silica based material, whereas HLB is a copolymer. (D) Van-Krevelen diagram of HeLa extract eluted from HLB material with ACN. Green points represent CHO, blue CHOS and red CHON molecules. Gray points represent data from the KEGG database.

5.4 Infection with *P. aeruginosa* affects the HeLa cell lipidome

To mimic infection of human cells, HeLa cells and *P. aeruginosa* has to been in contact. For practical reasons this means co-cultivation. Problems arising from this co-cultivation are contamination of HeLa cell preparations with bacteria. A dedicated sample preparation reduces the bacterial load in the samples for metabolome analysis, but a 100% clean sample cannot be produced. To evaluate if bacterial lipids are interfering with analysis of the HeLa cell lipidome, extracts of both species separately were subjected to lipidome analysis and the overlap of peaks was determined. Major membrane lipids in *P. aeruginosa* are PE, whereas it can also contain several percent of PC. Little or no contamination was observed for phospholipid or triacylglycerol (TG) regions in the chromatogram. This is in good agreement with literature, as TG have only rarely described in bacteria, mainly in actinomycetes [149]. Figure 51 shows the overlap of the different lipidomes.

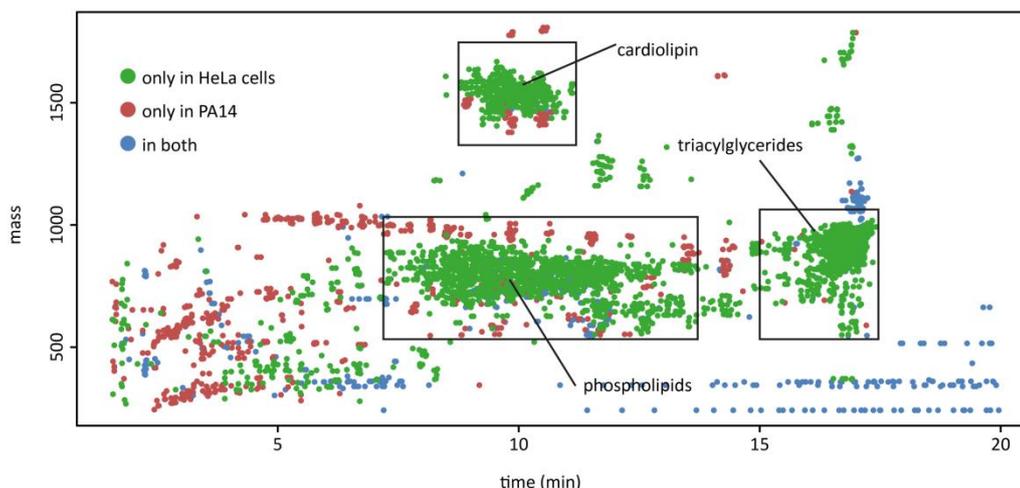


Figure 51: Overlap of HeLa and *P. aeruginosa* lipidome

Features overlapping between HeLa and *P. aeruginosa*. Features in blue are found in both organisms, whereas green are on found in HeLa cells and red only in PA14.

Analysis of the obtained ICR-FT/MS spectra showed that in infected HeLa cells more LysoPC and LysoPE species are found. This is possibly caused by ExoU activity. ExoU is a potent phospholipase A2. ExoU may alter lipid composition to lyse host cells. *exoU* is found together with *exoT* and *exoY* in cytotoxic strains.

To identify lipids altered during infection HeLa cells infected with *P. aeruginosa* PA14 Δ *gacA* were compared against *P. aeruginosa* PA14wt infected ones. The foldchange was calculated and significant features with a p-value < 0.05 were further evaluated. Several triacylglycerols are lower in PA14 wt. This is agreement with the results obtained from *C. elegans*, which also showed lower TG levels during infection.

In contrast to this, diacylglycerols and ceramides were found in higher levels. Interestingly both molecules are linked to apoptosis. Ceramides are enriched in distinct parts of the membrane called lipid rafts. Lipid rafts play important roles in infection of mammalian cells with *P. aeruginosa*, *N. gonorrhoeae* or *S. aureus*. Riethmüller et al. showed that infection with *P. aeruginosa* leads to activation of acid sphingomyelinase. The ceramide enriched lipid rafts are possibly needed for bacterial internalization [150]. Furthermore acid sphingomyelinase is linked to redox signaling in *P. aeruginosa* induced apoptosis of macrophages [151]. Figure 52A shows a barplot of Cer(d18:1/25:0) as example.

Lastly, PA14 wt infected cells had different levels of phospholipids with saturated fatty acid side chains, whereas lipids significant for PA14 Δ *gacA* included phospholipids with highly unsaturated side chains. Unfortunately, lysophospholipids were not readily detected in this sample set. It seems that ExoU has a preference for lipids with highly unsaturated side chains.

In fact, ExoU increases intracellular levels of unsaturated fatty acids from which prostaglandins are produced. The group of Plotkowski et al. recently reported the ability of ExoU to induce overproduction of prostaglandin E2 in airway epithelial cell [152]

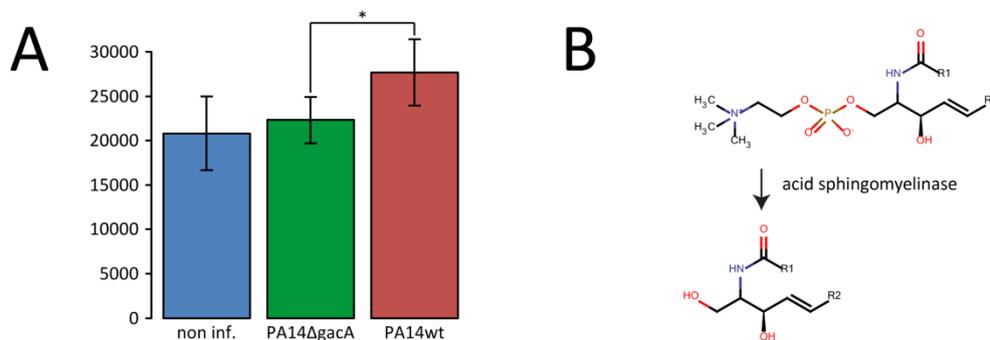


Figure 52: Several Ceramides were found in higher amount in *P. aeruginosa* infected HeLa cells.

(A) Cer(d18:1/25:0) was found in higher levels along with other Ceramide species in PA14 wt infected HeLa cells (B) Ceramides are produced from sphingolipids by the enzyme acid sphingomyelinase, which is up regulated in *P. aeruginosa* infection.

From the used infection model only small numbers of HeLa cells, in the range of 9e4 to 2e5 cells, could be recovered from the culture plates. Probably most cells are lysed and lost during the washing step to remove bacteria. After extraction in the appropriate extraction medium, either with SPE or lipid extraction only small amounts of extract of with about 50 to 100 μl could be obtained. Due to high sensitivity of ICR-FT/MS the amount is sufficient for direct infusion experiments. For UHPLC-UHR-ToF-MS analysis the obtained amount was not optimal. One solution to overcome this problem would be to avoid the step of washing away bacteria and prepare an extract from the whole system. This yields an extract containing metabolites and lipids from both organisms. Therefore metabolites present in both, have to be excluded, e.g. information about PE species, major membrane constituents of *P. aeruginosa*, could not be used. The other possibility is to upscale the whole infection model and use bigger cultivation plates. This certainly increases also costs, due to the used special cultivation plates. Lastly, the analytical method could be improved. The method used for lipid profiling of *C. elegans* offers higher sensitivity, but sample consumption is also higher. Due to lack of time and material this method could not be tested with HeLa lipid extracts.

5.5 ExoY – A potent cellular toxin

Isolated studies on single virulence factors are useful to reveal their role in the multi-parametric process of infection. These effectors have to be highly potent effectors, because bacteria cannot produce vast amounts of them to achieve their aims. ExoY is one of the five T3SS effectors of *Pseudomonas aeruginosa*. The protein is 378 amino acids long and contains two ATP binding sites which are similar to the domains of the extracellular adenylyl cyclases

from *Bordetella pertussis* (CyaA) and *Bacillus anthracis* (edema factor). Upon injection into mammalian cells ExoY raises the intracellular cAMP level, leading to differential expression of multiple genes including many that are known to be regulated by cAMP. The eukaryotic interaction partner of ExoY is still unknown. Recent work presented on the 5th International Conference on cGMP: Generators, Effectors and Therapeutic Implications, 2011, suggested that ExoY has a preference for GTP. Using a sensitive HPLC-MS/MS method it was shown the GTP and UTP are more preferred than ATP in both cell culture with intact cells or by the purified enzyme [153]. cAMP, produced by ExoY is a well-known second messenger involved in regulation of glycogen, sugar and lipid metabolism. cAMP dependent lipases convert TG's to free fatty acids and glycerol. Furthermore in *E. coli* it has been shown that cAMP regulated the biosynthesis of carnitine, suggesting cAMP has broader functions than just being a hunger signal [132]. Pulldown experiments at the CeMM in Vienna gave several hits for putative ExoY interaction partners. Table 9 summarizes the Pulldown results using the TAP-TAG strategy [154]. Parallel to these experiments a set of samples was prepared to elucidate the effect of ExoY on the host cell metabolome. Cells with a vector carrying full length *exoY* or an empty vector were subjected to metabolic profiling using ICR-FT/MS.

Table 19: Possible interaction partner for ExoY

Possible interaction partner for ExoY were determined using pulldown experiments at CeMM in Vienna.

Experimental interactor	interactor Description	Pathomics experiment	Confidence	Pathways
TBB3_HUMAN	Tubulin beta-3 chain	Pulldown first	Low	3
EFTU_HUMAN	Elongation factor Tu, mitochondrial	Pulldown first	low	1
EF1A3_HUMAN	Putative elongation factor 1-alpha-like 3	Pulldown first	low	1
ACACB_HUMAN	Acetyl-CoA carboxylase 2	Pulldown first	low	5
UBB_HUMAN	Polyubiquitin-B	Pulldown first	low	1
2ABA_HUMAN	Serine/threonine-protein phosphatase 2A 55 kDa regulatory subunit B alpha isoform	Pulldown first	high	5

5.6 Effects of ExoY on the HeLa cell metabolome

Data from negative mode experiments confirmed previous knowledge, that ExoY increases the intracellular cAMP concentration and decreases ATP, and also ADP and AMP levels. Despite this also higher levels in adenosine were found. Similar profiles were found for GTP and CTP and their metabolites. This shows that the results are consistent with published

results. Figure 53 shows the levels of different nucleotides and nucleosides, detected. Together with higher levels of adenosine, a general danger signal [155], higher levels of inosine were found. Inosine is directly built from adenosine, suggesting that cells are possibly trying to decrease adenosine signaling.

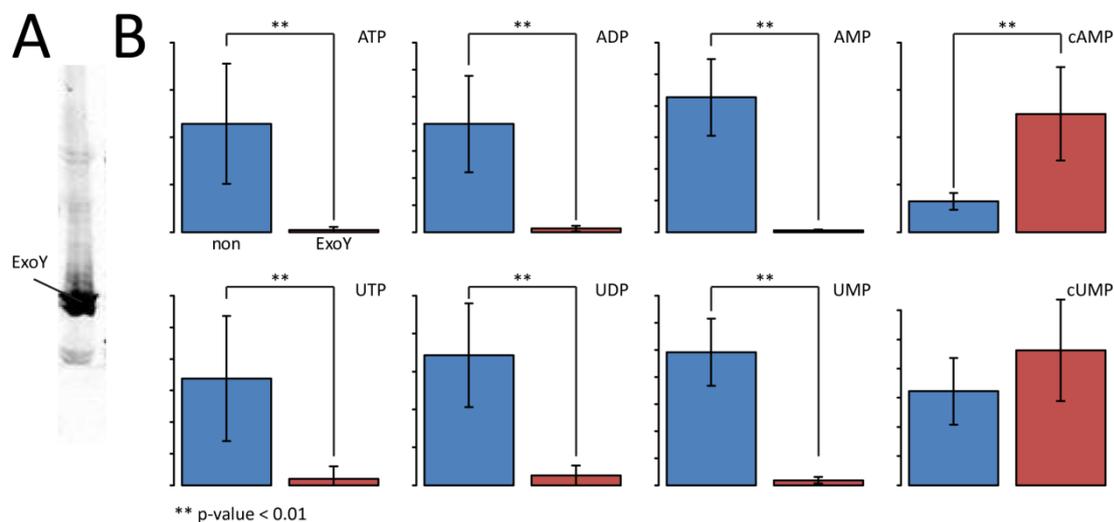


Figure 53: Expression of ExoY leads to decreased nucleoside levels and higher levels of cyclic NMP's
 (A) Western Blot of a protein extract from transfected shows expression of ExoY. (B) Levels of nucleosides where lower in *exoY* transfected cells, whereas cyclic nucleotide monophosphates increased.

To reveal further metabolic alterations occurring in cells expressing ExoY log-ratio analysis was performed. Citrate/isocitrate and cis-aconite were found in significantly lower amounts. Moreover lactate and malate were found as unique masses in ExoY transfected cells, whereas no NADH was detected in these samples. This findings lead to the suggestion that ExoY is interfering with the TCA cycle. Normally an increased level of cAMP acts positively on the conversion of pyruvate to acetyl-CoA, which enters the TCA cycle by condensation with oxaloacetate yielding citrate. Interestingly one of the partners of ExoY in the pulldown experiments is acetyl-CoA carboxylase 2, converting acetyl-CoA to malonyl-CoA. Unfortunately, malonyl-CoA could not be detected. The lower amount of NADH in ExoY transfected cells leads to the suggestion that the TCA cycle is not passed through. Instead pyruvate is metabolized to lactate.

From glycolysis, D-glucose, fructose/glucose 6-Phosphate and fructose 1,6-bis-Phosphate were detected in significantly lower levels in ExoY transfected cells. Other metabolites of the glycolysis pathway could not be detected, due to the high mass cut-off of the ICR-FT/MS. Two possibilities exist. First levels of the three mentioned metabolites are lower because they are consumed the following reactions. This means, they are fed through the glycolysis to yield pyruvate. The second alternative would be that glycogen is produced from glucose for storage

of excessive energy. A marker of increased glycogen synthesis is UDP-glucose. However, also UDP-glucose levels are low in cells expressing ExoY. These results suggest that normal glycolysis takes place and pyruvate is further metabolized to lactate.

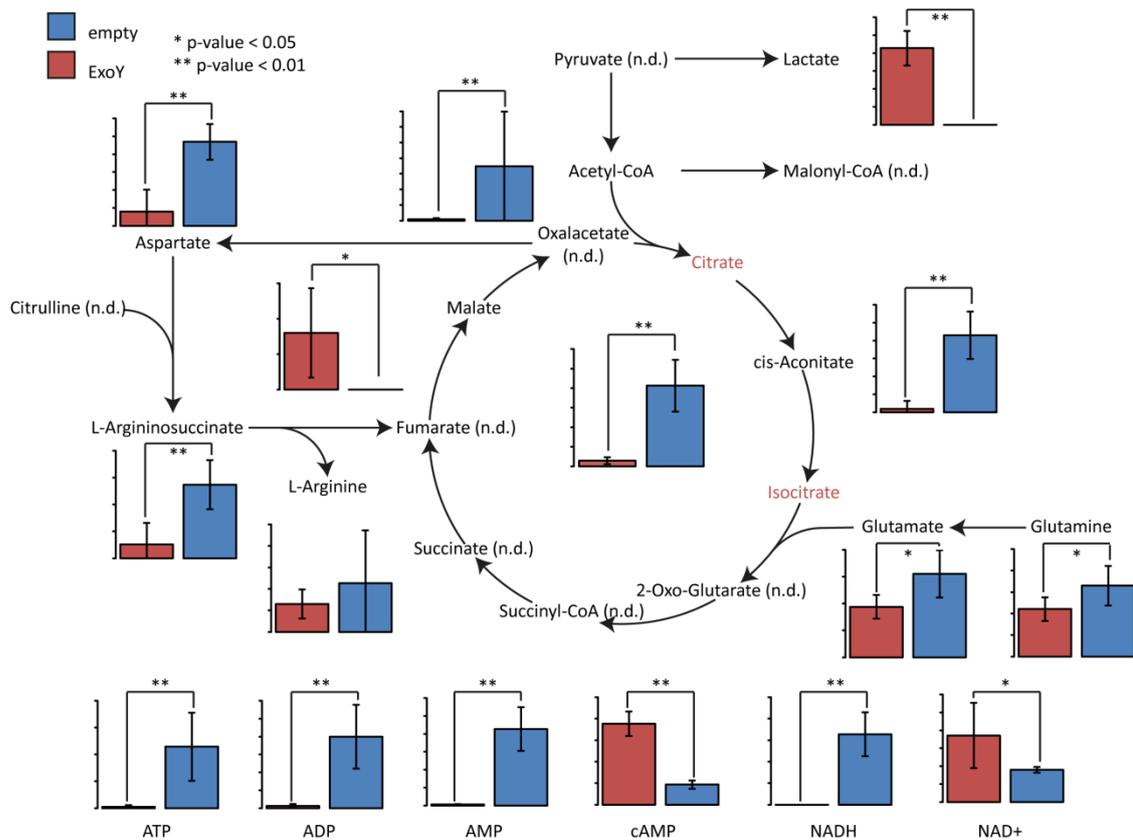


Figure 54: Cells expressing ExoY have an interrupted TCA cycle

Differences in detected metabolites suggest an interruption of the TCA cycle with lower levels of citrate and cis-aconitate and higher levels of malate in ExoY expressing cells. Certainly, acetyl-CoA should be also increased, which is not the case. Possibly it is cleaved to CoA and acetate.

Additionally to TCA cycle metabolites, some acylcarnitines could be detected in positive ionization mode. Similar to infected *C. elegans* samples, cells transfected with ExoY have lower levels in acylcarnitines. Possibly fat storage of the cells is broken down by β -oxidation, with goal to generate energy via the TCA cycle. However, blockage of the TCA cycle, would lead to an increased level of acetyl-CoA which was not detected. In contrary, levels of acetyl-CoA are lower in ExoY transfected cells. Acetylcarnitine is known as buffer for acetate to free CoA for other metabolic processes. Both, CoA and acetylcarnitine have lower levels cell expressing ExoY. Possibly, acetyl-CoA cleaved into acetate and CoA, which is used in other metabolic reactions. Hoerr et al. described elevated acetate levels in serum of mice infected with *P. aeruginosa* PA01 [156]. Because no lipidome data was obtained from these cells, it could not be seen if TG's are lower. Certainly, HeLa cells infected with PA14wt show decreased levels of TG's.

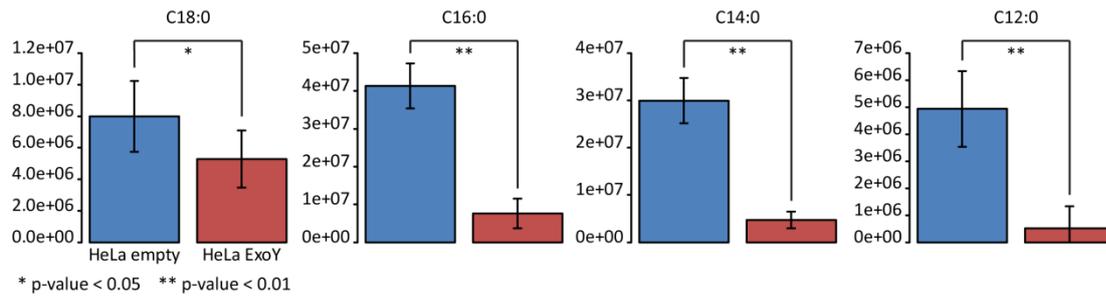


Figure 55: Acylcarnitines detected in HeLa cells

The detected acylcarnitines have lower levels in ExoY transfected HeLa cells, which is a similar trend as found in *C. elegans*

For confirmation of these results further experiments are needed, e.g. targeted methods for quantification of glycolysis and TCA cycle metabolites. Additionally, a lipidomic approach to reveal if ExoY is responsible for TG reduction would be useful. Lastly, a non-catalytic ExoY mutant has to be included as additional control.





6.1 Metabolomics – a fast growing research area

The systematic study of metabolism, today called metabolomics, currently revives. It is a fast growing discipline of the “omics”-technologies. Standard protocols are currently available for several standard biology organisms, like mice, *C. elegans*, *A. thaliana* or other model organisms. Despite this, new protocols are published on a monthly basis. Also, more and more high-throughput platforms are established all over the world, allowing metabolic profiling of even large population cohorts [157].

Metabolomics holds great opportunities as functional genomics tools, to describe novel gene functions, enzymes and metabolic pathways. The described methods in this rely on the ultrahigh-resolution of current mass spectrometric technology and allowed comprehensive detection of both known and unknown metabolites. Beside the used non-targeted profiling approach, the targeted investigation of different metabolites and metabolic flux analysis hold great potentials to further elaborate the used *P. aeruginosa* infection models.

6.2 Improved annotation and data analysis methods

The extensively used MasSTRIX metabolite annotation server was further improved in this work. It was developed to meet one of the key aspects of metabolomics, conversion of mass spectrometric data into biological meaning. The server uses different metabolomics databases, like KEGG, HMDB and LIPIDMAPS to putatively annotate metabolites. A key improvement was implementation and possibility to upload own mass lists of metabolites not covered by the used databases. This feature enabled to include metabolites like the ascaroside signaling molecules of *C. elegans* or secondary metabolites from different *Pseudomonas* species in the annotation process that cannot be found in any of the above mentioned databases. Furthermore comparing different sample states with the compare pathways functionality turned out to be a useful tool to retrieve a first overview on altered metabolic pathways. Comparison against other existing tools with the same functionality showed that MasSTRIX is a useful resource for metabolomics researches, which is also used extensively by other users. The only major drawback is the slowness compared to other tools, which will be changed in further releases by the use of a MySQL database instead of flat files.

In order to understand metabolic capabilities of the studied organisms' simple metabolic reconstructions were carried out. The lists were assembled by using data retrieval tools from different databases. These list rather present metabolites that are expected to be found in an organism, than a complete reaction network for *in silico* modeling.



Comparison of different species from the genus *Pseudomonas*, showed the great metabolic flexibility of these bacteria. The extracellular pathogen *Pseudomonas aeruginosa* possess an extensive arsenal of genes, encoding for several xenobiotic degradation pathways and efflux pumps, making it hard to treat with antibiotics. Because *P. aeruginosa* can synthesize all needed building blocks, like amino acids itself the major drive for infecting higher organisms doesn't seem to be the search for nutrients. On the first sight, this seems to be clear, but hosts normally present a nutrient rich environment. A study on the gene expression of metabolic genes of *P. aeruginosa* isolates from CF patients lungs over several months showed high metabolic activity in the begin of infection, which declines [158]. This suggests that *P. aeruginosa* is rather taking up nutrients from the surrounding environment than synthesizing them itself.

C. elegans in contrast is dependent on external sources of branched chain amino acids and cholesterol. Also different metabolic processes are not present in *C. elegans* and if metabolites related to these are found in measurements, they possible derive from ingested bacteria. One example is depicted in chapter 4, where metabolites of phenylalanine were found in *fer-15 C. elegans* fed with *E. coli* OP50 that cannot be produced by the worm.

6.3 Ultrahigh resolution analytical methods

Two different methods of metabolic profiling were used, DI-ICR-FT/MS and UHPLC-UHR-ToF-MS. DI-ICR-FT/MS, with superior resolution and mass accuracy, was shown to be a good screening tool for metabolic difference between certain sample conditions. It provided good group separation and revealed metabolic phenotypes with significantly different metabolites, as possible marker for each group. Nevertheless, some drawbacks exist. First, in DI-MS ion suppression is a commonly observed phenomenon, in which ionization of lower abundant molecules is suppressed by higher abundant ones or molecules that ionize more readily. Second, due to missing chromatography isomeric and isobaric substances cannot be resolved. Despite this drawbacks, DI-ICR-FT/MS is a useful tool for first metabolic profiling, because measurement time and sample consumption are much smaller than for other techniques.

For calculation of molecular formulas a mass difference network based approach, described by Tziotis et al., was used. This approach revealed several novel formulas which were previously not found in the major metabolite database like KEGG or HMDB. With these, bigger and more comprehensive databases like PubChem or Chemspider can searched. Calculation is fast and only dependent on the input list of allowed chemical transformations.



The second method, UHPLC-UHR-ToF-MS added additional information of retention time and separation of isobaric and isomeric substances. Totally, 3 different methods, 2 for metabolite profiling and 1 for lipid profiling were used. Usage of positive and negative ionization mode further enhanced the metabolome and substance coverage. A QSRR model was developed for enhanced metabolite annotation and identification. Although the total model was only based on 41 standard substances, prediction errors of about 25% were achieved. Further inclusion of more substances in a future model, will possibly reduce this error. Creek et al. already stated in their work on retention time prediction for metabolomics, that the error is substance class specific, so prediction will always work better for one or the other substance class. Furthermore, prediction of retention time is always coupled to the goodness of the prediction of physicochemical properties of the molecules. Some chemoinformatics tools like the ChemAxon toolkit offer the possibility to teach their prediction algorithms with an own library of molecules. The closer such a library is to the real metabolites present in the sample, the better the prediction will be. Chemoinformatics currently enters the metabolomics field in different ways, as reviewed shortly in chapter 2. It holds great opportunities to assist in identification of known and unknown metabolites.

The used separation methods allowed the separation and detection of metabolites from several metabolic pathways. Reversed phase separation using a BEH C8 column could separate different acylcarnitines, including isomeric species like α - and γ -linolenyl-carnitine. In contrast to this, chromatography on a BEH Amide column could separate leucine and isoleucine. The only major drawbacks of these methods are missing metabolites from TCA cycle and nucleotides, e.g. ATP.

Alignment between ICR-FT/MS and UHPLC-UHR-ToF-MS was used to obtain exact masses together with exact mass measurements. This approach added to possibility to obtain exact masses for formula calculation of unknown molecules together with retention time information, which can add hints about possible molecule structures.

6.4 Metabolomics based phylogeny of *Pseudomonas* species

The genus *Pseudomonas* is a very broad bacterial clade with species posing different metabolic capabilities, as shown in chapter 2. Current standard methods in bacterial taxonomy include phenotypic analysis, as far as possible, testing for different growth conditions and presence of reactions. Phylogeny is based on sequence similarity of 16S RNA. Recently the MALDI Biotyper[®] of Bruker enabled a more rapid identification of microbes based on molecular profiles obtained from MALDI-ToF-MS. Metabolomics based on ICR-FT/MS profiling



was shown by Rosella-Mora et al. to be a useful tool to reveal differences between different geographic isolates of the halophile *Salinibacter ruber*.

In this work different *Pseudomonas* species were used in order to first proof that also UHPLC-UHR-ToF-MS is also applicable and second that metabolomics can be a useful tool in microbiology.

For this purpose 14 different strains were grown in either LB medium or M9 buffer with Glucose as sole carbon source. Statistical analysis revealed a separation according to growth medium and afterwards in the single media a separation according to the single species. HILIC separation using a BEH Amide column together with negative ionization and growth in M9 buffer was shown to be a good method to cluster the different species based on their central metabolites, as far as they were detected. Although several metabolites were missing in this method; it provides a good starting point. Chromatographic separation that also allow the detection of Citrate and related metabolites from the TCA cycle, metabolites from Glycolysis, Pentose phosphate pathway and many others can give a better picture and a more clear separation of the different species. Reversed phase separation of metabolite extracts derived from bacteria grown in LB medium gave hints about synthesis of secondary metabolites.

A major problem is the measurement of LB medium samples. Here the medium is already a complex mixture of peptides and metabolites, making a non-targeted investigation of secreted metabolites nearly impossible. A better solution would be a complex, but chemically defined medium. Furthermore samples were only taken at the exponential phase of growth. A second interesting point would be the stationary phase because here most *Pseudomonads* start to develop biofilms at this time point. Additional isotopic labeling of certain amino acids will allow revealing their metabolic faith and how the different strains are using them.

In most of the methods *P. aeruginosa* clustered away from all other species. This is possible of clinical relevance and can be possibly used for diagnosis in future. Lastly a better understanding of *P. aeruginosa* metabolism can be used for development of novel therapeutics.

A standardized method including even more than one or two growth conditions, like in the BiOLOG microbial identification system can be a useful tool for microbiologists. In the BiOLOG system several carbon sources are tested in microtiter plates. A similar system could be set up for metabolomics, where bacteria are grown in microtiter plates in chemically defined medium with different carbon sources, e.g. M9 medium with Glucose or Succinate. Quantification of central metabolites or even flux analysis using ^{13}C labeled carbon sources or ^{15}N labeled amino

acids can give information of their direct use. The obtained data can be stored in public available database which can be used by biologist all over the world. Such use of public data repositories of all kind of data was already emphasized by Ramon Rosello-Mora in his review on taxonomy of bacteria and archaea [159].

6.5 *C. elegans* as convenient model for host-pathogen research

50-80% percent of the *C. elegans* genes show homology to human genes. Short reproductive cycle, a translucent body, possibility to generate knock-out mutants and the low cost made it to an ideal model organism to study several diseases. *C. elegans* is widely used in host-pathogen research as alternative host organism. Comparing the preparation steps for *P. aeruginosa* infection models used in this work, *C. elegans* allowed the easier maintenance. HeLa cells were rapidly lysed already after several minutes upon contact with the pathogen, whereas first *C. elegans* died after several hours to one day. Second, the removal of interfering bacteria was easier with *C. elegans* than with HeLa cells, which needed careful washing steps and the used of special cultivation plates. In contrast, *C. elegans* just needed simple shaking in M9 buffer to remove residual bacteria from the intestinal tract. Lastly, the nematode is a whole organism model, with differentiated organs, which possibly allows better to study the influence of infection on the host metabolic system.

The presented work showed the usability of *C. elegans* and non-targeted metabolomics in host-pathogen research. Metabolic profiling of *C. elegans* was carried out using DI-ICR-FT/MS and UHPLC-UHR-ToF-MS.

Usage of *C. elegans* as host model organism allowed identifying key events in metabolic reactions upon infection. Data analysis showed that down-regulation of amino acid metabolism is a major contributor to the metabolic phenotype of infected or stressed worms. Also *daf-2* worms fed on *E. coli* OP50 show lower activities on these metabolic pathways. This reaction is shared between *P. aeruginosa* PA14 and *S. enterica* infection. Moreover it is found in worms fed with *P. aeruginosa* PA14 Δ *gacA*, suggesting that this is a common response, possible mediated by pathogen associated molecular patterns (PAMPs). On top of this common reaction, different metabolic phenotypes evolve. The inclusion of the PA14 Δ *gacA* mutant was certainly useful, because nutritional differences between *E. coli* and *P. aeruginosa* would have led to false positive discoveries. In *P. aeruginosa* infection, comparing the fully virulent wildtype against the highly attenuated Δ *gacA* mutant, lipid metabolism seems to play a central role. Low levels of acylcarnitines and triacylglycerols argue for increased β -oxidation in *C. elegans* infected with PA14 wt. β -oxidations takes place in parallel to glycolysis, because under conditions of pure β -oxidation ketone bodies are formed from acetyl-CoA, which are

missing during infection. In contrast to this, *S. enterica* infected worms have normal acylcarnitine levels comparable to *E. coli* OP50 or PA14ΔgacA, but increased levels of sugar metabolites. Contrary to *P. aeruginosa* infection, energy seems to be utilized by metabolizing stored sugars.

The open question is what molecular mechanisms are responsible for this metabolic phenotype? Because this trend is similar in *fer-15* and *daf-2* worms, regulation independent of *daf-2* or a bacterial effector are plausible causes. ExoY is a good candidate as effector causing these metabolic alterations. cAMP produced by ExoY acts as second messenger, regulating metabolism of glycogen, sugar and lipid metabolism. Indeed, the major storages of lipids in *C. elegans* are found in intestinal cells, which are in direct contact with the pathogen. Secretion of bacterial effectors into this cells lead to increased energy utilization from lipids. In adipocytes, cAMP is known to stimulate lipolysis through enhanced lipase activity.

HeLa cells expressing ExoY show partially similar patterns to PA14wt infected *C. elegans*. In these cells, also a cut-off of the TCA cycle was observed, with low levels of citrate/isocitrate and cis-aconitate. Interestingly, lactate was found only in cells expressing ExoY, which leads to the conclusion that pyruvate obtained from glycolysis is further metabolized by lactic fermentation. Additionally, decreased levels of several acylcarnitines were found. Because no lipidomic data was collected it is not possible to tell if TG's are depleted like in *C. elegans*. HeLa cells infected with PA14wt showed low levels of TG's, consistent with the nematode model.

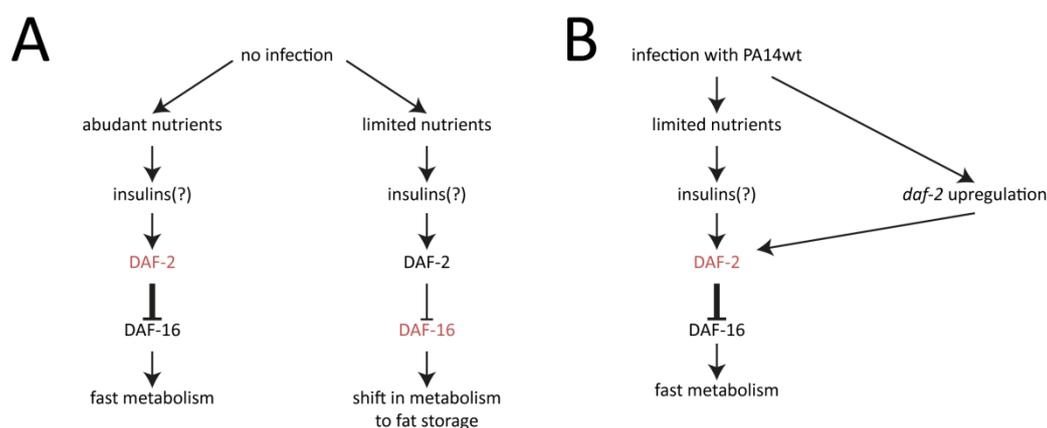


Figure 56: Genetic model of metabolism regulation during infection

(A) DAF-2 is one of the major regulators in *C. elegans* metabolism under non-infection conditions. If nutrients are highly abundant DAF-2 is highly active and inhibits DAF-16, leading to a normal metabolism. However if nutrients are limited DAF-2 activity decreases and DAF-16 is active, leading to a shifted metabolism towards fat storage and slower metabolism. (B) During infection also nutrients are limited and normally DAF-2 would be not active, but PA14wt is able to up-regulate expression of *daf-2* and possibly tries to mimic high nutrient conditions.

However, ExoY together with the two other type 3 effectors ExoU and ExoT, are dispensable for *C. elegans* killing, as shown before [160, 161]. From this point of view, the

observed metabolic phenotype, possibly caused by ExoY seems to be not the only driving force for *C. elegans* killing. A second major event is the up regulation of *daf-2* by *P. aeruginosa* to suppress host immunity. Daf-2 inhibits Daf-16 and therefore down regulates Daf-16 target genes. Because Daf-16 also controls the transcription of several metabolic genes, *daf-2* up-regulation also affects metabolism. This up-regulation requires the two-component regulator GacA.

Because during infection no or less bacteria are digested by *C. elegans* nutrients are limited, which would normally lead to decreased *daf-2* activity and a slower metabolism. One of the goals of *daf-2* up-regulation is possibly to maintain a normal metabolism in the host to make use of this nutrient rich niche.

In *daf-2* worms the TCA cycle is inactive and the glyoxylate shunt, gluconeogenesis and starch metabolism are dominating, as shown by Fuchs et al. [72]. Reduced levels of triacylglycerols and acylcarnitines were also found in PA14wt infected *daf-2* worms, suggesting that increased lipolysis and β -oxidation is a *daf-2* independent event. The adaptation of metabolic fluxes in *daf-2* worms, possibly gives them an advantage over *fer-15* worms to cope with increased acetyl-CoA levels. In *fer-15* worms the fate of acetyl-CoA in PA14 wt infected worms is still an open question. Cells can use carnitine to restore CoA levels. However, acetylcarnitine is lower in infected *daf-2* and *fer-15* worms. Eventually acetyl-CoA is cleaved into CoA and acetate which is excreted into the lumen, where it possibly serves as carbon source for *P. aeruginosa*.

Following model of metabolic alterations in *C. elegans* during *P. aeruginosa* infection, shown in Figure 57, is proposed, which has to be proven by further targeted metabolome and transcriptome analysis. After ingestion and attachment *P. aeruginosa* injects type three effectors into the intestinal cells of *C. elegans*. Adenylate cyclase activity of ExoY increases the intracellular cAMP level, leading to increased lipolytic activity. Genes of β -oxidation are up-regulated and increased β -oxidation leads to decreased levels of TG's and Acylcarnitines.

This leads first to an increased level of acetyl-CoA. Additional up-regulation of *daf-2* reduces gluconeogenesis. Because TCA cycle is blocked due *P. aeruginosa* effectors, acetyl-CoA cannot be further metabolized and energy metabolism is disturbed. Excessive acetyl-CoA is cleaved to CoA and acetate, which is excreted. Because in *daf-2* worms down regulation of Gluconeogenesis is not possible, acetyl-CoA can be converted via the glyoxylate shunt to malate and oxaloacetate and is possibly used for gluconeogenesis.

Quantification of metabolites present on the glycolysis and gluconeogenesis pathway, TCA cycle and glyoxylate shunt can proof this hypothesis. The key enzyme for understanding this metabolic alteration seems to be citrate synthase.

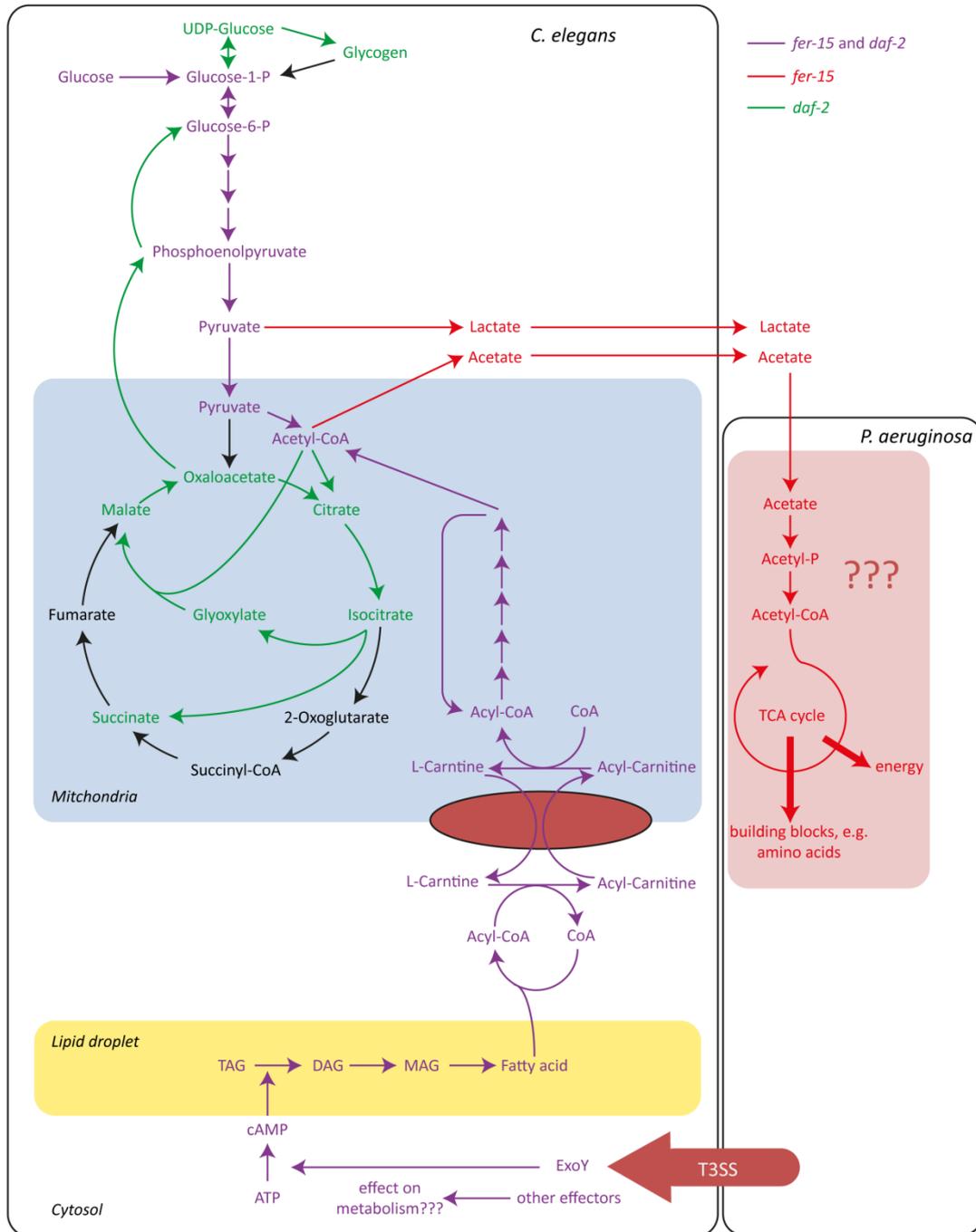


Figure 57: Proposed model of central carbon metabolism in *C. elegans* during *P. aeruginosa* infection

P. aeruginosa attaches to the intestinal cells of *C. elegans* and injects different T3SS effectors, like ExoY into these cells. Increasing cAMP levels lead to enhanced lipolysis and β -oxidation (response shared between *daf-2* and *fer-15* worms, lila color). Due to a cut-off of the TCA cycle, the obtained acetyl-CoA cannot be further metabolized in the TCA cycle and up regulation of *daf-2* leads to decreased gluconeogenesis. In *fer-15* mutants (red color), acetyl-CoA is cleaved to CoA and acetate, which is excreted and can be taken up by *P. aeruginosa*. In contrast, in *daf-2* worms, higher gluconeogenesis converts acetyl-CoA to sugars.

Furthermore it must be proven that *C. elegans* in fact secretes excessive acetate, which can be used by *P. aeruginosa*. Looking to the other side of the host-pathogen equation will be certainly useful. The protocol to remove remaining bacteria in the worm gut can be used to recover these bacteria and study their metabolome. If increased acetate utilization by *P. aeruginosa* is found, this will finally proof this model. E.g. growing *C. elegans* for certain generations on bacteria fed with uniformly ^{13}C labeled glucose before switching to *P. aeruginosa* will allow accumulating metabolites enriched in ^{13}C . If $^{13}\text{C}_2$ -acetate is afterwards found in *P. aeruginosa* recovered from the worm gut, this would be the ultimate proof.

A comparison of infection with different *P. aeruginosa* strains or mutants will be certainly useful to reveal if the metabolic host response is similar to all strains or if different metabolic phenotypes evolve. Not all strains are carrying ExoY. The hypothesis that ExoY is the effector affecting at least the central carbon metabolism holds true, strains where ExoY is absent should show a different phenotype.

A second major question is the resistance of *daf-2* worms to bacterial pathogens. Metabolome analysis revealed higher levels of adenosine in *daf-2 C. elegans* infected with *P. aeruginosa* or *S. enterica* compared to *fer-15* worms. Adenosine was shown to reduced pathogenicity of *P. aeruginosa* [138]. If adenosine is a mediator of this resistance, an adenosine deficient and *daf-2* double mutant should have reduced resistance.

Anyway, it was shown that metabolomics is a useful tool for host-pathogen research. It can be used in future research, to reveal contribution of different virulence genes. *C. elegans* can be fed on mutants of certain virulence genes and following cluster analysis can reveal similar metabolic phenotypes and if they are comparable to PA14 wt or the highly attenuated PA14 ΔgacA mutant or novel phenotypes can be found. *C. elegans* fed with PA14 Δpec7 is an example used in this work. The obtained lipidomics data showed that PA14 Δpec7 is behaving like the parental wildtype.

Furthermore metabolomics can assist in annotation of orphan genes. Data in this work suggest a possible carnitine 3-dehydrogenase activity in *C. elegans*, which is not described until now. Mutation of the corresponding gene and cultivation of the nematode in axenic culture medium containing isotope labeled carnitine can proof this hypothesis.

The Million Mutation Project of *C. elegans* would be a good opportunity for in-depth metabolome analysis of *C. elegans*. Currently the project reports 822,138 SNP's in 20,117 genes and 12,534 knockouts in 6774 genes [162]. Screening of this library in a metabolomics approach would be certainly useful to generate deeper understanding of metabolism in

eukaryotes. Such high-throughput studies have already been used with prokaryotes like *E. coli* [18].

6.6 HeLa cell model

Most work in the HeLa cell model was focusing on a dedicated sample preparation for this infection model. Problems arise from the mixture of two different organisms, because metabolites cannot be mapped back to the original species, like it is possible for DNA-, RNA or protein sequences. The developed sample preparation methods represent a good compromise between fastness needed in metabolomics to obtain an instantaneous snapshot of a cells metabolism and the reduction of the bacterial load in the cell culture preparation.

This method was applied to a HeLa infection model and DI-ICR-FT/MS and lipid profiling using UHPLC-UHR-ToF-MS showed alteration in the host cell metabolome. A major feature was reduced levels of TG's, similar to infected *C. elegans*. Second, ceramide levels were increased, which is known from previous studies, showing that the developed workflow and methods are valid. A major problem from the used methodology is the low recovery of cells. For higher sensitivity in the used analytical methods, higher cell numbers are needed. Scaling up the whole model, could be one solution. Another possibility would be to go to nano flow LC-MS, which is well suited for low volume applications, needing high sensitivity.

HeLa cells were also used to express the bacterial effector ExoY to enlighten its effects on the host metabolome. Analysis was carried with DI-ICR-FT/MS and revealed changes in the TCA cycle. In *exoY* expressing cells this pathway seems to be cut off and pyruvate and acetyl-CoA are metabolized in alternate routes. Also lower levels of acylcarnitines were found, which can be an indicator of increased β -oxidation.

6.7 Comparison of the infection models

Results from the *C. elegans* and HeLa infection models show a striking overlap, especially at the TCA cycle and low levels of TG's. This suggests that this response is possibly conserved during evolution. The used HeLa infection model has to be further optimized to obtain cleaner preparations in higher amounts. Using the more sensitive lipidome methods used for lipid profiling in *C. elegans*, can further enhance results. Still the major problem with this model, is the fast lysis of HeLa cells. Moving to a different of *P. aeruginosa*, e.g. PAO1, can solve this issue. However, the question is still open, if the metabolic phenotype is the same as for PA14. Table 20 compares some effects seen in the different infection models.

Because most effects are found in both model systems, it seems that the affection of energy and lipid metabolism is one of the major events in *P. aeruginosa* infection. Several

results published by other groups support this idea. In a mouse model of *P. aeruginosa* used by Hoerr et al, reduced levels of citrate, 2-oxoglutarate, glucose, fumarate, pyruvate and succinate were found, whereas acetylcarnitine, glycerol, lactate and acetate were elevated [156]. Interestingly, Behrends et al. compared 179 strains of *P. aeruginosa* isolated from cystic fibrosis patients based on their exometabolome. They found a highly significant negative correlation between time of infection and acetate levels in the cell supernatants. Acetate levels of wildtype *P. aeruginosa* PAO1, which was used as control, exceeded most clinical isolates. The results either suggest a decreased excretion of acetate or an increased uptake from the medium [163]. In another work, the group around Nina Jagmann described parasitic growth of *P. aeruginosa* PAO1 with *Aeromonas hydrophila* on chitin as carbon and nitrogen source. *P. aeruginosa* PAO1 is not able to use chitin. The growth consisted of two phases. First both organisms grew in a commensal way. In the second phase a massive acetate release of about 10 mM by *A. hydrophila* was observed, which was caused by pyocyanin from *P. aeruginosa*. Pyocyanin inhibited aconitase and the TCA cycle, but not pyruvate dehydrogenase [164].

Table 20: Comparison of different infection models

The effects were observed between PA14ΔgacA and PA14wt. n.d. = not detected

Effect	<i>C. elegans</i> (fer-15)	<i>C. elegans</i> (daf-2)	<i>HeLa</i> infection	<i>HeLa</i> ExoY
Altered TCA cycle?	Yes		n.d.	Yes
Low acyl-carnitines ?	Yes	Yes	n.d.	Yes
Low TG's?	Yes	Yes	Yes	n.d.
Amino acid metabolism affected?	Yes	Yes	n.d.	no

Apart from this, results have to be validated using targeted methods. The presented methods represent good starting points for further work in the field of host-pathogen metabolomics. They can be used in a broader screening approach using different mutants and compare the observed metabolic phenotypes. This makes it hopefully possible to identify molecular actors that are responsible for this phenotypes and supply future targets for drug development or diagnostics.





Literature

Literature



1. Blow, N., *Metabolomics: Biochemistry's new look*. Nature, 2008. **455**(7213): p. 697-700.
2. Santorio, S., *De Statica Medicina*. 1675.
3. Wöhler, F., *Ueber künstliche Bildung des Harnstoffs*. Annalen der Physik, 1828. **88**(2): p. 253-256.
4. Krebs, H.A. and W.A. Johnson, *Metabolism of ketonic acids in animal tissues*. Biochemical Journal, 1937. **31**(4): p. 645-660.
5. Lynen, F., *Lipide Metabolism*. Annual Review of Biochemistry, 1955. **24**(1): p. 653-688.
6. Buchanan, B. and D. Arnon, *A reverse KREBS cycle in photosynthesis: consensus at last*. Photosynthesis Research, 1990. **24**(1): p. 47-53-53.
7. Berg, J.M., T.J. L., and S. Lubert, *Biochemistry*. 2002, W H Freeman: New York.
8. 2012; Available from:
<http://www.imsb.ethz.ch/researchgroup/sauer/research/centralmetabolism>.
9. Fiehn, O., *Metabolomics – the link between genotypes and phenotypes*. Plant Molecular Biology, 2002. **48**(1): p. 155-171.
10. Weckwerth, W., *Metabolomics in Systems Biology*. Annual Review of Plant Biology, 2003. **54**(1): p. 669-689.
11. Wishart, D., *Systems Biology Resources Arising from the Human Metabolome Project, Genetics Meets Metabolomics*, K. Suhre, Editor. 2012, Springer New York. p. 157-175.
12. Nicholson, J.K., J.C. Lindon, and E. Holmes, *'Metabonomics': understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data*. Xenobiotica, 1999. **29**(11): p. 1181-1189.
13. Pauling, L., et al., *Quantitative Analysis of Urine Vapor and Breath by Gas-Liquid Partition Chromatography*. Proceedings of the National Academy of Sciences, 1971. **68**(10): p. 2374-2376.
14. Marianna, L., *Datamining metabolomics: the convergence point of non-target approach and statistical investigation*. 2009.

15. Kell, D.B., et al., *Metabolic footprinting and systems biology: the medium is the message*. Nat Rev Micro, 2005. **3**(7): p. 557-565.
16. Calder, P.C., *Dietary Fatty Acids and the Immune System*. Nutrition Reviews, 1998. **56**(1): p. S70-S83.
17. Nandakumar, M. and M.-W. Tan, *Gamma-Linolenic and Stearidonic Acids Are Required for Basal Immunity in Caenorhabditis elegans through Their Effects on p38 MAP Kinase Activity*. PLoS Genet, 2008. **4**(11): p. e1000273.
18. Fuhrer, T., et al., *High-Throughput, Accurate Mass Metabolome Profiling of Cellular Extracts by Flow Injection–Time-of-Flight Mass Spectrometry*. Analytical Chemistry, 2011. **83**(18): p. 7074-7080.
19. Goodacre, R., *Metabolomics – the way forward*. Metabolomics, 2005. **1**(1): p. 1-2.
20. Sud, M., et al., *LMSD: LIPID MAPS structure database*. Nucleic Acids Research, 2007. **35**(suppl 1): p. D527-D532.
21. Hu, C., et al., *RPLC-Ion-Trap-FTMS Method for Lipid Profiling of Plasma: Method Validation and Application to p53 Mutant Mouse Model*. Journal of Proteome Research, 2008. **7**(11): p. 4982-4991.
22. Ogundare, M., et al., *Cerebrospinal Fluid Steroidomics: Are Bioactive Bile Acids Present in Brain?* Journal of Biological Chemistry, 2010. **285**(7): p. 4666-4679.
23. Aharoni, A., et al., *Nontargeted Metabolome Analysis by Use of Fourier Transform Ion Cyclotron Mass Spectrometry*. OMICS: A Journal of Integrative Biology, 2002. **6**(3): p. 217-234.
24. Huang, N., et al., *Automation of a Fourier transform ion cyclotron resonance mass spectrometer for acquisition, analysis, and e-mailing of high-resolution exact-mass electrospray ionization mass spectral data*. Journal of the American Society for Mass Spectrometry, 1999. **10**(11): p. 1166-1173.
25. Southam, A.D., et al., *Dynamic Range and Mass Accuracy of Wide-Scan Direct Infusion Nanoelectrospray Fourier Transform Ion Cyclotron Resonance Mass Spectrometry-Based Metabolomics Increased by the Spectral Stitching Method*. Analytical Chemistry, 2007. **79**(12): p. 4595-4602.

26. Kim, S., et al., *Automated electrospray ionization FT-ICR mass spectrometry for petroleum analysis*. Journal of the American Society for Mass Spectrometry, 2009. **20**(2): p. 263-268.
27. Marshall, A.G., *Milestones in Fourier transform ion cyclotron resonance mass spectrometry technique development*. Int J Mass Spectrom, 2000. **200**: p. 331-356.
28. Kind, T. and O. Fiehn, *Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry*. BMC Bioinformatics, 2007. **8**(1): p. 105.
29. Li, X., et al., *At-line coupling of UHPLC to chip-electrospray-FTICR-MS*. Analytical and Bioanalytical Chemistry, 2007. **389**(5): p. 1439-1446.
30. Cubbon, S., et al., *Metabolomic applications of HILIC-LC-MS*. Mass Spectrometry Reviews. **29**(5): p. 671-684.
31. Rainville, P.D., et al., *Novel Application of Reversed-Phase UHPLC-*oa*TOF-MS for Lipid Analysis in Complex Biological Mixtures: A New Tool for Lipidomics*. Journal of Proteome Research, 2006. **6**(2): p. 552-558.
32. Wilson, I.D., et al., *High Resolution Ultra Performance Liquid Chromatography Coupled to *oa*-TOF Mass Spectrometry as a Tool for Differential Metabolic Pathway Profiling in Functional Genomic Studies*. Journal of Proteome Research, 2005. **4**(2): p. 591-598.
33. Stoll, D.R., et al., *Fast, comprehensive two-dimensional liquid chromatography*. Journal of Chromatography A, 2007. **1168**(1-2): p. 3-43.
34. Buescher, J.M., et al., *Ultrahigh Performance Liquid Chromatography-Tandem Mass Spectrometry Method for Fast and Robust Quantification of Anionic and Aromatic Metabolites*. Analytical Chemistry, 2010. **82**(11): p. 4403-4412.
35. Pluskal, T., et al., *MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data*. BMC Bioinformatics. **11**(1): p. 395.
36. Lommen, A. and H. Kools, *MetAlign 3.0: performance enhancement by efficient use of advances in computer hardware*. Metabolomics, 2012. **8**(4): p. 719-726.
37. Soga, T., et al., *Quantitative Metabolome Analysis Using Capillary Electrophoresis Mass Spectrometry*. Journal of Proteome Research, 2003. **2**(5): p. 488-494.



38. Monton, M.R.N. and T. Soga, *Metabolome analysis by capillary electrophoresis-mass spectrometry*. Journal of Chromatography A, 2007. **1168**(1-2): p. 237-246.
39. An, Y.J., et al., *Metabotyping of the C. elegans sir-2.1 Mutant Using in Vivo Labeling and 13C-Heteronuclear Multidimensional NMR Metabolomics*. ACS Chemical Biology, 2012.
40. Beltran, A., et al., *Assessment of Compatibility between Extraction Methods for NMR- and LC/MS-Based Metabolomics*. Analytical Chemistry, 2012. **84**(14): p. 5838-5844.
41. W., M., *Über ein neues System der Bakterien*. Arb Bakteriol Inst Karlsruhe, 1894. **1**: p. 235-328.
42. N., P.J., *Family I. Pseudomonadaceae*, in *Bergey's Manual of Systemative Bacteriology*, Bergey, Editor. 1984. p. 141-172.
43. Willenbrock, H., et al., *An environmental signature for 323 microbial genomes based on codon adaptation indices*. Genome Biology, 2006. **7**(12): p. R114.
44. Willenbrock, H. and D. Ussery, *Prediction of highly expressed genes in microbes based on chromatin accessibility*. BMC Molecular Biology, 2007. **8**(1): p. 11.
45. Lyczak, J.B., C.L. Cannon, and G.B. Pier, *Establishment of Pseudomonas aeruginosa infection: lessons from a versatile opportunist*. Microbes and Infection, 2000. **2**(9): p. 1051-1060.
46. Stover, C.K., et al., *Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen*. Nature, 2000. **406**(6799): p. 959-964.
47. Lee, D., et al., *Genomic analysis reveals that Pseudomonas aeruginosa virulence is combinatorial*. Genome Biology, 2006. **7**(10): p. R90.
48. Winsor, G.L., et al., *Pseudomonas Genome Database: improved comparative analysis and population genomics capability for Pseudomonas genomes*. Nucleic Acids Research, 2011. **39**(suppl 1): p. D596-D600.
49. Oberhardt, M.A., et al., *Genome-Scale Metabolic Network Analysis of the Opportunistic Pathogen Pseudomonas aeruginosa PAO1*. The Journal of Bacteriology, 2008. **190**(8): p. 2790-2803.

50. He, J., et al., *The broad host range pathogen Pseudomonas aeruginosa strain PA14 carries two pathogenicity islands harboring plant and animal virulence genes*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(8): p. 2530-2535.
51. Rehm, B.H.A., *Pseudomonas*. 1 ed. 2008: Wiley. 402.
52. Economou, A., et al., *Secretion by numbers: protein traffic in prokaryotes*. Molecular Microbiology, 2006. **62**(2): p. 308-319.
53. Cornelis, G.R., *The type III secretion injectisome*. Nat Rev Micro, 2006. **4**(11): p. 811-825.
54. Wolfgang, M.C., et al., *Conservation of genome content and virulence determinants among clinical and environmental isolates of Pseudomonas aeruginosa*. Proceedings of the National Academy of Sciences, 2003. **100**(14): p. 8484-8489.
55. Yahr, T.L., et al., *ExoY, an adenylate cyclase secreted by the Pseudomonas aeruginosa type III system*. Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(23): p. 13899-13904.
56. Ha, U. and S. Jin, *Growth Phase-Dependent Invasion of Pseudomonas aeruginosa and Its Survival within HeLa Cells*. Infection and Immunity, 2001. **69**(7): p. 4398-4406.
57. Shaver, C.M. and A.R. Hauser, *Relative Contributions of Pseudomonas aeruginosa ExoU, ExoS, and ExoT to Virulence in the Lung*. Infection and Immunity, 2004. **72**(12): p. 6969-6977.
58. Kulasekara, B.R., et al., *Acquisition and Evolution of the exoU Locus in Pseudomonas aeruginosa*. Journal of Bacteriology, 2006. **188**(11): p. 4037-4050.
59. Bleves, S., et al., *Protein secretion systems in Pseudomonas aeruginosa: A wealth of pathogenic weapons*. International Journal of Medical Microbiology, 2010. **300**(8): p. 534-543.
60. Frimmersdorf, E., et al., *How Pseudomonas aeruginosa adapts to various environments: a metabolomic approach*. Environmental Microbiology, 2010. **12**(6): p. 1734-1747.

61. Behrends, et al., *Metabolic profiling of Pseudomonas aeruginosa demonstrates that the anti-sigma factor MucA modulates osmotic stress tolerance*. Vol. 6. Cambridge, ROYAUME-UNI: Royal Society of Chemistry. 8.
62. Gjersing, E.L., et al., *NMR Metabolomics of Planktonic and Biofilm Modes of Growth in Pseudomonas aeruginosa*. *Analytical Chemistry*, 2007. **79**(21): p. 8037-8045.
63. Muller, H., et al., *Variation and adaptation of Pseudomonas aeruginosa toxicity to HeLa cells and fibroblasts*. *Journal of Clinical Microbiology*, 1986. **24**(3): p. 317-323.
64. Lorenz, M.A., C.F. Burant, and R.T. Kennedy, *Reducing Time and Increasing Sensitivity in Sample Preparation for Adherent Mammalian Cell Metabolomics*. *Analytical Chemistry*, 2011. **83**(9): p. 3406-3414.
65. Sugimoto, M., et al., *Non-targeted metabolite profiling in activated macrophage secretion*. *Metabolomics*, 2012. **8**(4): p. 624-633.
66. Hulme, S.E. and G.M. Whitesides, *Chemistry and the Worm: Caenorhabditis elegans as a Platform for Integrating Chemical and Biological Research*. *Angewandte Chemie International Edition*, 2011. **50**(21): p. 4774-4807.
67. 2012. Available from:
<http://blog.neuinfo.org/wp-content/uploads/2011/06/celegans.jpg>.
68. Evans, E.A., T. Kawli, and M.-W. Tan, *Pseudomonas aeruginosa Suppresses Host Immunity by Activating the DAF-2 Insulin-Like Signaling Pathway in Caenorhabditis elegans*. *PLoS Pathog*, 2008. **4**(10): p. e1000175.
69. Szewczyk, N., E. Kozak, and C. Conley, *Chemically defined medium and Caenorhabditis elegans*. *BMC Biotechnology*, 2003. **3**(1): p. 19.
70. Geier, F.M., et al., *Cross-Platform Comparison of Caenorhabditis elegans Tissue Extraction Strategies for Comprehensive Metabolome Coverage*. *Analytical Chemistry*, 2011. **83**(10): p. 3730-3736.
71. Blaise, B.J., et al., *Metabotyping of Caenorhabditis elegans reveals latent phenotypes*. *Proceedings of the National Academy of Sciences*, 2007. **104**(50): p. 19808-19812.
72. Fuchs, S., et al., *A metabolic signature of long life in Caenorhabditis elegans*. *BMC Biology*, 2010. **8**(1): p. 14.

73. Martin, F.-P.J., et al., *Metabotyping of Caenorhabditis elegans and their Culture Media Revealed Unique Metabolic Phenotypes Associated to Amino Acid Deficiency and Insulin-Like Signaling*. Journal of Proteome Research, 2011. **10**(3): p. 990-1003.
74. Butler, J.A., et al., *Long-lived mitochondrial (Mit) mutants of Caenorhabditis elegans utilize a novel metabolism*. The FASEB Journal. **24**(12): p. 4977-4988.
75. Pungaliya, C., et al., *A shortcut to identifying small molecule signals that regulate behavior and development in Caenorhabditis elegans*. Proceedings of the National Academy of Sciences, 2009. **106**(19): p. 7708-7713.
76. von Reuss, S.H., et al., *Comparative Metabolomics Reveals Biogenesis of Ascarosides, a Modular Library of Small-Molecule Signals in C. elegans*. Journal of the American Chemical Society, 2012. **134**(3): p. 1817-1824.
77. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes*. Nucleic Acids Res, 2000. **28**(1): p. 27-30.
78. Caspi, R., et al., *The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases*. Nucleic Acids Research, 2008. **36**(suppl 1): p. D623-D631.
79. Schomburg, I., A. Chang, and D. Schomburg, *BRENDA, enzyme data and metabolic information*. Nucleic Acids Res, 2002. **30**(1): p. 47-49.
80. Csizmadia, F., *JChem: Java applets and modules supporting chemical database handling from web browsers*. Journal of Chemical Information and Computer Sciences, 2000. **40**(2): p. 323-324.
81. Steinbeck, C., et al., *The Chemistry Development Kit (CDK): an open-source Java library for chemo- and bioinformatics*. J Chem Inf Comput Sci, 2003. **43**(2): p. 493-500.
82. Steinbeck, C., et al., *Recent developments of the Chemistry Development Kit (CDK) - An open-source Java library for chemo- and bioinformatics*. Current Pharmaceutical Design, 2006. **12**(17): p. 2111-2120.
83. Nobeli, I., et al., *A Structure-based Anatomy of the E. coli Metabolome*. Journal of Molecular Biology, 2003. **334**(4): p. 697-719.
84. Heinonen, M., et al., *Metabolite identification and molecular fingerprint prediction through machine learning*. Bioinformatics, 2012. **28**(18): p. 2333-2341.

85. Wolf, S., et al., *In silico fragmentation for computer assisted identification of metabolite mass spectra*. BMC Bioinformatics, 2010. **11**(1): p. 148.
86. Wang, Y., et al., *PubChem: a public information system for analyzing bioactivities of small molecules*. Nucleic Acids Res, 2009(37 Web Server): p. W623-633.
87. Kanehisa, M., et al., *From genomics to chemical genomics: new developments in KEGG*. Nucleic Acids Research, 2006. **34**(suppl 1): p. D354-D357.
88. Wishart, D.S., et al., *HMDB: a knowledgebase for the human metabolome*. Nucleic Acids Res, 2009(37 Database): p. D603-610.
89. Suhre, K. and P. Schmitt-Kopplin, *MassTRIX: mass translator into pathways*. Nucleic Acids Research, 2008. **36**(suppl 2): p. W481-W484.
90. Wapstra, A.H., G. Audi, and C. Thibault, *The Ame2003 atomic mass evaluation: (I). Evaluation of input data, adjustment procedures*. Nuclear Physics A, 2003. **729**(1): p. 129-336.
91. García-Alcalde, F., et al., *Paintomics: a web based tool for the joint visualization of transcriptomics and metabolomics data*. Bioinformatics, 2011. **27**(1): p. 137-139.
92. Antunes, L.C.M., et al., *Impact of Salmonella Infection on Host Hormone Metabolism Revealed by Metabolomics*. Infection and Immunity, 2011. **79**(4): p. 1759-1769.
93. Jansson, J., et al., *Metabolomics Reveals Metabolic Biomarkers of Crohn's Disease*. PLoS ONE, 2009. **4**(7): p. e6386.
94. Leader, D.P., et al., *Pathos: A web facility that uses metabolic maps to display experimental changes in metabolites identified by mass spectrometry*. Rapid Communications in Mass Spectrometry, 2011. **25**(22): p. 3422-3426.
95. Henry, C.S., et al., *High-throughput generation, optimization and analysis of genome-scale metabolic models*. Nat Biotech, 2010. **28**(9): p. 977-982.
96. Bellmann-Weiler, R., et al., *Divergent modulation of Chlamydia pneumoniae infection cycle in human monocytic and endothelial cells by iron, tryptophan availability and interferon gamma*. Immunobiology, 2010. **215**(9-10): p. 842-848.

97. Kastenmüller, G., J. Gasteiger, and H.-W. Mewes, *An environmental perspective on large-scale genome clustering based on metabolic capabilities*. *Bioinformatics*, 2008. **24**(16): p. i56-i62.
98. Gillespie, J.J., et al., *PATRIC: the Comprehensive Bacterial Bioinformatics Resource with a Focus on Human Pathogenic Species*. *Infection and Immunity*, 2011. **79**(11): p. 4286-4298.
99. Dale, J., L. Popescu, and P. Karp, *Machine learning methods for metabolic pathway prediction*. *BMC Bioinformatics*, 2010. **11**(1): p. 15.
100. Karp, P.D., et al., *Pathway Tools version 13.0: integrated software for pathway/genome informatics and systems biology*. *Briefings in Bioinformatics*, 2010. **11**(1): p. 40-79.
101. Aziz, R., et al., *The RAST Server: Rapid Annotations using Subsystems Technology*. *BMC Genomics*, 2008. **9**(1): p. 75.
102. Brendza, K.M., et al., *Phosphoethanolamine N-methyltransferase (PMT-1) catalyses the first reaction of a new pathway for phosphocholine biosynthesis in *Caenorhabditis elegans**. *Biochem J*, 2007. **404**(3): p. 439-448.
103. Wishart, D.S., et al., *HMDB 3.0—The Human Metabolome Database in 2013*. *Nucleic Acids Research*, 2012.
104. Frolkis, A., et al., *SMPDB: The Small Molecule Pathway Database*. *Nucleic Acids Research*, 2010. **38**(suppl 1): p. D480-D487.
105. Stobbe, M., et al., *Critical assessment of human metabolic pathway databases: a stepping stone for future integration*. *BMC Systems Biology*, 2011. **5**(1): p. 165.
106. Hein, E.-M. and H. Hayen, *Comparative Lipidomic Profiling of *S. cerevisiae* and Four Other Hemiascomycetous Yeasts*. *Metabolites*, 2012. **2**(1): p. 254-267.
107. Rossello-Mora, R., et al., *Metabolic evidence for biogeographic isolation of the extremophilic bacterium *Salinibacter ruber**. *ISME J*, 2008. **2**(3): p. 242-253.
108. Pena, A., et al., *Fine-scale evolution: genomic, phenotypic and ecological differentiation in two coexisting *Salinibacter ruber* strains*. *ISME J*. **4**(7): p. 882-895.



109. Brito-Echeverría, J., et al., *Response to adverse conditions in two strains of the extremely halophilic species <i>Salinibacter ruber</i>*. *Extremophiles*. **15**(3): p. 379-389-389.
110. Gross, H. and J.E. Loper, *Genomics of secondary metabolite production by Pseudomonas spp.* *Natural Product Reports*, 2009. **26**(11): p. 1408-1446.
111. Youard, Z.A., et al., *Pseudomonas fluorescens CHA0 Produces Enantio-pyochelin, the Optical Antipode of the Pseudomonas aeruginosa Siderophore Pyochelin*. *Journal of Biological Chemistry*, 2007. **282**(49): p. 35546-35553.
112. Bonsall, R.F., D.M. Weller, and L.S. Thomashow, *Quantification of 2,4-Diacetylphloroglucinol Produced by Fluorescent Pseudomonas spp. In Vitro and in the Rhizosphere of Wheat*. *Applied and Environmental Microbiology*, 1997. **63**(3): p. 951-5.
113. Garvis, S., et al., *Caenorhabditis elegans Semi-Automated Liquid Screen Reveals a Specialized Role for the Chemotaxis Gene cheB2 in Pseudomonas aeruginosa Virulence*. *PLoS Pathog*, 2009. **5**(8): p. e1000540.
114. Tan, M.-W., et al., *Pseudomonas aeruginosa killing of Caenorhabditis elegans used to identify P. aeruginosa virulence factors*. *Proceedings of the National Academy of Sciences*, 1999. **96**(5): p. 2408-2413.
115. Garsin, D.A., et al., *Long-Lived C. elegans daf-2 Mutants Are Resistant to Bacterial Pathogens*. *Science*, 2003. **300**(5627): p. 1921-1921.
116. Houthoofd, K., et al., *DAF-2 pathway mutations and food restriction in aging Caenorhabditis elegans differentially affect metabolism*. *Neurobiology of Aging*, 2005. **26**(5): p. 689-696.
117. *ACD/MS Fragmenter*. Mass spectral fragmentation analysis software.
118. *MassFrontier*. Mass spectral fragmentation analysis software.
119. Bowen, B.P. and T.R. Northen, *Dealing with the Unknown: Metabolomics and Metabolite Atlases*. *Journal of the American Society for Mass Spectrometry*. **21**(9): p. 1471-1476.
120. Kumari, S., et al., *Applying In-Silico Retention Index and Mass Spectra Matching for Identification of Unknown Metabolites in Accurate Mass GC-TOF Mass Spectrometry*. *Analytical Chemistry*, 2011. **83**(15): p. 5895-5902.

121. Creek, D.J., et al., *Toward Global Metabolomics Analysis with Hydrophilic Interaction Liquid Chromatography–Mass Spectrometry: Improved Metabolite Identification by Retention Time Prediction*. *Analytical Chemistry*, 2011. **83**(22): p. 8703-8710.
122. Brown, M., et al., *Mass spectrometry tools and metabolite-specific databases for molecular identification in metabolomics*. *Analyst*, 2009. **134**(7): p. 1322-1332.
123. Dunn, W., et al., *Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics*. *Metabolomics*: p. 1-23.
124. Peironcely, J.E., et al., *Understanding and Classifying Metabolite Space and Metabolite-Likeness*. *PLoS One*, 2011. **6**(12): p. e28966.
125. Murphy, R.C., et al., *Detection of the abundance of diacylglycerol and triacylglycerol molecular species in cells using neutral loss mass spectrometry*. *Analytical Biochemistry*, 2007. **366**(1): p. 59-70.
126. Bird, S.S., et al., *Separation of Cis–Trans Phospholipid Isomers Using Reversed Phase LC with High Resolution MS Detection*. *Analytical Chemistry*, 2012. **84**(13): p. 5509-5517.
127. Tziotis, D., Norbert Hertkorn, and P. Schmitt-Kopplin, *Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity*. *European Journal of Mass Spectrometry*, 2011. **17**(4): p. 6.
128. Lewis, J.A. and J.T. Fleming, *Chapter 1 Basic Culture Methods*, in *Methods in Cell Biology*, F.E. Henry and C.S. Diane, Editors. 1995, Academic Press. p. 3-29.
129. Kaniga, K., I. Delor, and G.R. Cornelis, *A wide-host-range suicide vector for improving reverse genetics in Gram-negative bacteria: inactivation of the blaA gene of Yersinia enterocolitica*. *Gene*, 1991. **109**(1): p. 137-141.
130. Kurz, C.L. and J.J. Ewbank, *Caenorhabditis elegans: an emerging genetic model for the study of innate immunity*. *Nat Rev Genet*, 2003. **4**(5): p. 380-390.
131. Troemel, E.R., et al., *p38 MAPK Regulates Expression of Immune Response Genes and Contributes to Longevity in C. elegans*. *PLoS Genet*, 2006. **2**(11): p. e183.
132. Hormiga, J., et al., *Quantitative analysis of the dynamic signaling pathway involved in the cAMP mediated induction of l-carnitine biosynthesis in E. coli cultures*. *Molecular BioSystems*, 2010. **6**(4): p. 699-710.

133. Sevilla, A., et al., *Impairing and Monitoring Glucose Catabolite Repression in L-Carnitine Biosynthesis*. Biotechnology Progress, 2007. **23**(6): p. 1286-1296.
134. Jia, K., et al., *Autophagy genes protect against Salmonella typhimurium infection and mediate insulin signaling-regulated pathogen resistance*. Proceedings of the National Academy of Sciences, 2009. **106**(34): p. 14564-14569.
135. Irazoqui, J.E., et al., *Distinct Pathogenesis and Host Responses during Infection of C. elegans by P. aeruginosa and S. aureus*. PLoS Pathog, 2010. **6**(7): p. e1000982.
136. Hansen, M., et al., *A Role for Autophagy in the Extension of Lifespan by Dietary Restriction in C. elegans*. PLoS Genet, 2008. **4**(2): p. e24.
137. Hansen, M., et al., *A Role for Autophagy in the Extension of Lifespan by Dietary Restriction in C. elegans*. PLoS Genet, 2008. **4**(2): p. e24.
138. Sheng, L., et al., *Interkingdom adenosine signal reduces Pseudomonas aeruginosa pathogenicity*. Microbial Biotechnology, 2012. **5**(4): p. 560-572.
139. Lucanic, M., et al., *N-acylethanolamine signalling mediates the effect of diet on lifespan in Caenorhabditis elegans*. Nature, 2011. **473**(7346): p. 226-229.
140. Payne, S.H. and W.F. Loomis, *Retention and Loss of Amino Acid Biosynthetic Pathways Based on Analysis of Whole-Genome Sequences*. Eukaryotic Cell, 2006. **5**(2): p. 272-276.
141. Eisenberg, T., et al., *Induction of autophagy by spermidine promotes longevity*. Nat Cell Biol, 2009. **11**(11): p. 1305-1314.
142. Das, D., et al., *The crystal structure of a bacterial Sufu-like protein defines a novel group of bacterial proteins that are similar to the N-terminal domain of human Sufu*. Protein Science, 2010. **19**(11): p. 2131-2140.
143. Wang, C., et al., *Use of Caenorhabditis elegans for Preselecting Lactobacillus Isolates To Control Salmonella Typhimurium*. Journal of Food Protection 174;, 2011. **74**(1): p. 86-93.
144. Rizzi, M., et al., *In vivo analysis of metabolic dynamics in Saccharomyces cerevisiae: II. Mathematical model*. Biotechnology and Bioengineering, 1997. **55**(4): p. 592-608.
145. Silas G. Villas-Bôas, U.R., Michael A. E. Hansen, Jorn Smedsgaard, Jens Nielsen, *Metabolome Analysis - An Introduction*. 1 ed. 2007: John Wiley & Sons Inc.

146. Folch, J., M. Lees, and G.H.S. Stanley, *A Simple Method For The Isolation And Purification Of Total Lipids From Animal Tissues*. Journal of Biological Chemistry, 1957. **226**(1): p. 497-509.
147. Bligh, E.G. and W.J. Dyer, *A rapid method of total lipid extraction and purification*. Canadian Journal of Biochemistry and Physiology, 1959. **37**(8): p. 911-917.
148. Matyash, V., et al., *Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics*. Journal of Lipid Research, 2008. **49**(5): p. 1137-1146.
149. Alvarez, H.A. and A.S. Steinbüchel, *Triacylglycerols in prokaryotic microorganisms*. Applied Microbiology and Biotechnology, 2002. **60**(4): p. 367-376.
150. Riethmüller, J., et al., *Ceramide in Pseudomonas aeruginosa infections*. European Journal of Lipid Science and Technology, 2007. **109**(10): p. 998-1002.
151. Zhang, Y., et al., *Acid Sphingomyelinase Amplifies Redox Signaling in Pseudomonas aeruginosa-Induced Macrophage Apoptosis*. The Journal of Immunology, 2008. **181**(6): p. 4247-4254.
152. Plotkowski, M.-C., et al., *Lipid body mobilization in the ExoU-induced release of inflammatory mediators by airway epithelial cells*. Microbial Pathogenesis, 2008. **45**(1): p. 30-37.
153. Voigt, U., et al., *The effector protein ExoY secreted by Pseudomonas aeruginosa is a nucleotidyl cyclase with preference for GTP*. BMC Pharmacology, 2011. **11**(Suppl 1): p. P74.
154. Burckstummer, T., et al., *An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells*. Nat Meth, 2006. **3**(12): p. 1013-1019.
155. la Sala, A., et al., *Alerting and tuning the immune response by extracellular nucleotides*. Journal of Leukocyte Biology, 2003. **73**(3): p. 339-343.
156. Hoerr, V., et al., *Gram-negative and Gram-Positive Bacterial Infections Give Rise to a Different Metabolic Response in a Mouse Model*. Journal of Proteome Research, 2012. **11**(6): p. 3231-3245.
157. Gieger, C., et al., *Genetics Meets Metabolomics: A Genome-Wide Association Study of Metabolite Profiles in Human Serum*. PLoS Genet, 2008. **4**(11): p. e1000282.

158. Oberhardt, M.A., et al., *Metabolic Network Analysis of Pseudomonas aeruginosa during Chronic Cystic Fibrosis Lung Infection*. The Journal of Bacteriology: p. JB.00900-10.
159. Rosselló-Móra, R., *Towards a taxonomy of Bacteria and Archaea based on interactive and cumulative data repositories*. Environmental Microbiology, 2012. **14**(2): p. 318-334.
160. Miyata, S., et al., *Use of the Galleria mellonella Caterpillar as a Model Host To Study the Role of the Type III Secretion System in Pseudomonas aeruginosa Pathogenesis*. Infection and Immunity, 2003. **71**(5): p. 2404-2413.
161. Wareham, D.W., A. Papakonstantinou, and M.A. Curtis, *The Pseudomonas aeruginosa PA14 type III secretion system is expressed but not essential to virulence in the Caenorhabditis elegans-P. aeruginosa pathogenicity model*. FEMS Microbiology Letters, 2005. **242**(2): p. 209-216.
162. Flibotte, S., et al., *Whole-Genome Profiling of Mutagenesis in Caenorhabditis elegans*. Genetics, 2010. **185**(2): p. 431-441.
163. Behrends, V., et al., *Metabolic adaptations of Pseudomonas aeruginosa during cystic fibrosis chronic lung infections*. Environmental Microbiology, 2012: p. no-no.
164. Jagmann, N., H.-P. Brachvogel, and B. Philipp, *Parasitic growth of Pseudomonas aeruginosa in co-culture with the chitinolytic bacterium Aeromonas hydrophila*. Environmental Microbiology, 2010. **12**(6): p. 1787-1802.
165. Watts, J.L. and J. Browse, *Genetic dissection of polyunsaturated fatty acid synthesis in Caenorhabditis elegans*. Proceedings of the National Academy of Sciences, 2002. **99**(9): p. 5854-5859.

List of figures

Figure 1: Metabolites play a central role in all biological processes.....	2
Figure 2: Catabolic and anabolic metabolism is interwoven through ATP and redox factors, like NADH or NADPH.....	3
Figure 3: Genes, transcripts, proteins and metabolites are acting together in a tight network to build up life.....	7
Figure 4: Examples for different lipid classes found in biological samples	9
Figure 5: Principles of ICR-FT/MS	12
Figure 6: Principals of Q-ToF-MS	14
Figure 7: Principals of CE	15
Figure 8: A) NMR B) construction.....	16
Figure 9: <i>P. aeruginosa</i> secretion systems	20
Figure 10: Infection cycle of <i>P. aeruginosa</i>	21
Figure 11: <i>C. elegans</i> is a widely used model organism with a short life cycle.....	24
Figure 12: Project structure of Pathomics Project	28
Figure 13: Overview on thesis structure	30
Figure 14: Principal workflow and outputs of MassTRIX.	38
Figure 15 Possible comparisons between jobs in MassTRIX.....	39
Figure 16: Metabolic capabilities of different <i>Pseudomonas</i> species.....	42
Figure 17: Comparison of the in silico metabolomes downloaded from BioCyc.	44
Figure 18: Combined metabolic pathway map obtained directly from KEGG for <i>P. aeruginosa</i> PA14 and <i>C. elegans</i>	48
Figure 19: Scores plots of Principal component analysis of BEH C8 separation	57
Figure 20: Different growth temperatures lead to differences in the intracellular metabolome	58
Figure 21: PCA scores plots from BEH C8 separation of bacterial pellets grown in LB medium.	59
Figure 22: Levels of intracellular pyochelin and 2,4-diacetylphloroglucinol.....	61
Figure 23: Distances of different <i>Pseudomonas</i> species based on metabolome.....	62
Figure 24: Metabolites found in measurements, but not in prediction	63
Figure 25 Complex mass spectrometric patterns can arise from a single metabolite species, including fragments, isotopes and adducts, which complicates data analysis.	68
Figure 26: logP and logD values of the <i>C. elegans in silico</i> metabolome	72
Figure 27: Physicochemical properties of the <i>C. elegans in silico</i> metabolome	73
Figure 28: Building blocks of the <i>C. elegans in silico</i> metabolome	74
Figure 29: Metabolites on the same metabolic pathway show chemical similarity	75



Figure 30: Example of cluster found using correlation analysis and RT filtering	79
Figure 31: Isomeric and isobaric lipid species complicate analysis of lipidome data	80
Figure 32: The AutoMS ⁿ algorithm is a data dependent algorithm for the automated acquisition of MS/MS spectra during a chromatographic run	82
Figure 33: Retention time shifts and mass differences allow identification of members of homologous series.	83
Figure 34: Overview on NetCalc networks.....	86
Figure 35: Results of ICR-FT/MS and UHPLC-UHR-ToF-MS alignment.....	87
Figure 36: Typical killing curves obtained from <i>C. elegans</i> fed on different pathogenic bacteria.	89
Figure 37: Principal component analysis of negative ICR-FT/MS dataset	91
Figure 38: Pathway comparison from positive ionization ICR-FT/MS data	91
Figure 39: Several metabolites of L-Phenylalanine were found significant for <i>fer-15</i> worms fed with <i>E. coli</i> OP50	92
Figure 40: Metabolites from Glycolysis and TCA cycle detected in negative ICR-FT/MS analysis	94
Figure 41: Acylcarnitines are affected by infection independent of <i>C. elegans</i> genotype	96
Figure 42: Two different isomers of C18:3 Carnitine were detected	97
Figure 43: Carnitine synthesis and degradation pathways in <i>C. elegans</i>	98
Figure 44: Synthesis of different phospholipid species in <i>C. elegans</i>	99
Figure 45: Adenosine possibly mediates resistance in <i>daf-2</i> worms	101
Figure 46: Intensity pattern of unknown mass 470.23132 in DI-ICR-FT/MS experiments	103
Figure 47: Synthesis strategy and results from N-arachidonoyl-phospho-l-serine synthesis... ..	105
Figure 48: Pec7 is a possible novel bacterial effector or protein needed for virulence	107
Figure 49: Killing curves of <i>C. elegans</i> pre-fed with probiotics.....	108
Figure 50: Results of method optimization for HeLa cell preparation.....	118
Figure 51: Overlap of HeLa and <i>P. aeruginosa</i> lipidome.....	119
Figure 52: Several Ceramides were found in higher amount in <i>P. aeruginosa</i> infected HeLa cells.	120
Figure 53: Expression of ExoY leads to decreased nucleoside levels and higher levels of cyclicNMP's	122
Figure 54: Cells expressing ExoY have an interrupted TCA cycle	123
Figure 55: Acylcarnitines detected in HeLa cells.....	124
Figure 56: Genetic model of metabolism regulation during infection	133

Figure 57: Proposed model of central carbon metabolism in <i>C. elegans</i> during <i>P. aeruginosa</i> infection	135
Figure 58: Workflows used in Genedata Refiner MS	171



List of tables

Table 1: Definition of "omics"-sciences and related objects	6
Table 2: Advantages and Disadvantages of different metabolomics methods	17
Table 3: Genome size of several <i>Pseudomonas</i> species	18
Table 4: Range of <i>P. aeruginosa</i> infections in man	19
Table 5: Advantages and Disadvantages of <i>C. elegans</i> as model organism.....	24
Table 6: Essential and non-essential amino acids in <i>C. elegans</i> and humans.....	25
Table 7: MS adducts covered by MassTRIX.....	36
Table 8: Number of genes of different <i>Pseudomonas</i> species	43
Table 9: Metabolome sizes of different <i>Pseudomonas</i> species.....	45
Table 10: <i>Pseudomonas</i> strains used in this work	53
Table 11: M9 stock solutions	54
Table 12: The four levels of metabolite identification confidence defined by the Metabolomics Standards Initiative (Sumner et al. 2007).....	69
Table 13: Used standards with retention times under RP and HILIC conditions	76
Table 14: Calculated parameter for model equation 9 for the BEH Amide separation.....	78
Table 15: Mass differences used for network calculation	85
Table 16: Overview on masses in main network and the annotation of new formulas	86
Table 17: Detected masses in the different tested extraction solvents	116
Table 18: Characteristics of used SPE materials	117
Table 19: Possible interaction partner for ExoY.....	121
Table 20: Comparison of different infection models	138
Table 21: Lipid gradient program used for lipid profiling of <i>C. elegans</i> samples.....	170



List of equations

Equation 1: Lorentz force	11
Equation 2: Equilibrium between centrifugal and Lorentz force	11
Equation 3: Angular velocity	11
Equation 4: Definition of mass resolution.....	11
Equation 5: Definition of ppm error.....	12
Equation 6: Definition of chemical shift	16
Equation 7: Definition of logP	71
Equation 8: Definition of logD	71
Equation 9: Model equation of QSRR model used by Creek et al.....	77



Appendix

Appendix



Used instruments

ICR-FT/MS

Name	Manufacturer	Description
<i>solariX ICR-FT/MS</i>	<i>Bruker</i>	<i>Apollo II ESI source, Magnex Scientific 12 T superconducting magnet</i>

UHPLC-UHR-ToF-MS

Name	Manufacturer	Description
<i>Acquity UHPLC Binary Solvent Manager</i>	<i>Waters</i>	<i>Part No. 186015001, Serial No. L09UPB 919M</i>
<i>Acquity UHPLC Sample Manager</i>	<i>Waters</i>	<i>Part No. 186015006, Serial No. L09UPA 507M</i>
<i>Acquity UHPLC Column Manager</i>	<i>Waters</i>	<i>Part No. 186015009, Serial No. L09UPM 501G</i>
<i>Acquity UHPLC PDA</i>	<i>Waters</i>	<i>Part No. 186015026, Serial No. A10UPD 473A</i>
<i>maXis 3G UHR-ToF-MS</i>	<i>Bruker</i>	<i>Apollo II ESI source, Serial No. 10101</i>

GC-FID

Name	Manufacturer	Description
<i>GC 6890 with Autosampler</i>	<i>HP</i>	<i>Part No. G1530A, Serial No. DE00022295</i>

Other instruments

Name	Manufacturer	Description
<i>SpeedVac</i>	<i>Thermo Scientific</i>	<i>Savant SPD121P SpeedVac Concentrator</i>
<i>refrigerated Centrifuge</i>	<i>Eppendorf</i>	<i>5804R</i>
<i>sonic bath</i>	<i>Bandelin Sonorex</i>	<i>RK100H</i>
<i>sonic bath</i>	<i>Bandelin Sonorex</i>	<i>RK510</i>
<i>laboratory shake</i>	<i>Edmund Bühler</i>	<i>KS10 Laboratory shaker</i>
<i>vortex</i>	<i>Scientific Industry</i>	<i>Vortex Genie 2</i>
<i>balance</i>	<i>Kern</i>	<i>ABT120-4M</i>

Used chemicals

Solvents

Name	Manufacturer	Description
<i>methanol</i>	<i>Fluka</i>	<i>LC-MS Chromasolv, 34966-1L</i>
<i>acetonitrile</i>	<i>Fluka</i>	<i>LC-MS Chromasolv, 34967-1L</i>
<i>2-propanol</i>	<i>Fluka</i>	<i>LC-MS Chromasolv, 34965-1L</i>
<i>MTBE</i>	<i>Fluka</i>	<i>Chromasolv Plus for HPLC, 99.9%, 650560-1L</i>
<i>water</i>	---	<i>Water was obtained from a Merck Millipore Integral purification system on demand 18.2 MΩ, TOC < 5 ppb</i>
<i>dichloromethane</i>	<i>Sigma-Aldrich</i>	<i>ether Chromasolv Plus for HPLC, 99.9%, 650463-1L</i>
<i>hexane</i>	<i>Fluka</i>	<i>Chromasolv Plus for HPLC, 99.9%, 650552-1L</i>
<i>ethyl acetate</i>	<i>Sigma-Aldrich</i>	<i>Anhydrous, 99.8%, 270989</i>
<i>THF</i>	<i>Sigma-Aldrich</i>	<i>Chromasolv Plus for HPLC, ≥ 99.9%, 34865</i>
<i>chloroform</i>	<i>Sigma-Aldrich</i>	<i>Chromasolv Plus for HPLC, ≥99.9%, 650498</i>

Buffer salts

Name	Manufacturer	Description
<i>ammonium acetate</i>	<i>Biosolve</i>	<i>ULC/MS grade, 01244156</i>
<i>formic acid</i>	<i>Fluka</i>	<i>for mass spectrometry, 94318-50ML-F</i>
<i>sodium carbonate</i>	<i>Sigma-Aldrich</i>	<i>Anhydrous, 451614</i>
<i>sodium bicarbonate</i>	<i>Sigma-Aldrich</i>	<i>≥ 99.5%, S8875</i>

Chemicals for synthesis

Name	Manufacturer	Description
<i>phospho-L-serine</i>	<i>Sigma-Aldrich</i>	<i>≥ 98%, P0878</i>
<i>arachidonic acid</i>	<i>Sigma-Aldrich</i>	<i>≥ 95%, 10931</i>
<i>DCC</i>	<i>Sigma-Aldrich</i>	<i>99%, D80002</i>
<i>NHS</i>	<i>Sigma-Aldrich</i>	<i>98%, 130672</i>
<i>Silica gel</i>	<i>Sigma-Aldrich</i>	<i>High purity grade, ≥ 99%, 236772</i>
<i>HCl</i>	<i>Sigma-Aldrich</i>	<i>1.0 N, H9892</i>

Standards

Following standards were used for retention time determination for usage in a QSRR model. Stock solutions were prepared in a suitable solvent with a 1000 ppm concentration and stored at -20°C.

Name	Manufacturer	Description
alanine	Merck	for biochemistry, 1010070025
arginine	Sigma-Aldrich	L-Arginine, ≥98% (TLC), A5006-100G
asparagine	Sigma-Aldrich	≥98%, A0884
glutamate	Merck	3510
glycine	Merck	≥ 99%, 1042010100
histidine	Sigma-Aldrich	≥98%, H8000
leucine	Sigma-Aldrich	≥98%, L8000
serine	Sigma-Aldrich	≥ 99%, S4500
tryptophan	Sigma-Aldrich	≥ 98%, T025
tyrosine	Sigma-Aldrich	≥ 98%, T3754
aspartate	Sigma-Aldrich	≥ 98%, A9256
isoleucine	Sigma-Aldrich	≥ 98%, I2752
lysine	Sigma-Aldrich	≥ 98%, L5626
methionine	Sigma-Aldrich	≥ 99%, 64320
phenylalanine	Sigma-Aldrich	≥ 98%, P2126
proline	Sigma-Aldrich	≥ 99%, P0380
valine	Sigma-Aldrich	≥ 99%, V0375
(S)-malate	Merck	for synthesis, 814737
2,3-pyrazinedicarboxylic acid	Merck	for synthesis, 800214
benzoylformic acid	Fluka	≥ 98%, 78610
citrate	Sigma-Aldrich	≥ 99%, C0759
cumarin-3-carbonic acid	Merck	for synthesis, 818273
DL-malic acid	Merck	for synthesis, 814737
fumarate	Sigma-Aldrich	≥ 99%, 240745
glutaconic acid	Sigma-Aldrich	≥ 97.0%, 49360
glutaric acid	Alfa Aesar	≥ 99%, A14595
maleic acid	Sigma-Aldrich	≥ 99%, M0375
mesaconic acid	Sigma-Aldrich	≥ 99%, 131040
meso-tartaric acid	Sigma-Aldrich	≥ 99%, 95350
oxalic acid	Merck	extra pure, 100492
pyruvate	Merck	for biochemistry, 106619
succinate	Sigma-Aldrich	≥ 99%, 239682
tartronic acid	Sigma-Aldrich	≥ 97%, 86320
D-glucose	Sigma-Aldrich	≥ 99.5%, G7021
D-glucose 6-phosphate	Sigma-Aldrich	≥ 99%, G7879
D-ribose	Sigma-Aldrich	≥ 99%, R7500
rhamnose	Sigma-Aldrich	≥ 99%, R3875
arachidic acid C20:0	Sigma-Aldrich	≥ 99%, A3631
capric acid C10:0	Sigma-Aldrich	≥ 98%, C1875
caprylic acid C8:0	Sigma-Aldrich	≥ 99%, C2875
cholesterol	Sigma-Aldrich	≥ 99%, C8667
heneicosanoic acid C21:0	Sigma-Aldrich	≥ 98%, 51530
lauric acid C12:0	Sigma-Aldrich	≥ 98%, W261416
myristic acid C14:0	Fluka	≥ 98%, 70082
palmitic acid C16:0	Sigma-Aldrich	≥ 99%, P0500
perlagonic acid C9:0	Sigma-Aldrich	≥ 96%, W278408
stearic acid C18:0	Sigma-Aldrich	≥ 95%, W303518
undecanoic acid C11:0	Sigma-Aldrich	≥ 97%, W324507
biotin	Sigma-Aldrich	≥ 99%, B4501
cytosine	Sigma-Aldrich	≥ 99%, C3506
D-fructose	Fluka	Ph Eur, 47748

Name	Manufacturer	Description
<i>folate</i>	<i>Sigma-Aldrich</i>	<i>BioReagent, F8758</i>
<i>carnitine</i>	<i>Sigma-Aldrich</i>	<i>≥ 98%, C0158</i>
<i>nicotinamide</i>	<i>Sigma-Aldrich</i>	<i>≥ 98%, N3376</i>
<i>orthophosphate</i>	<i>Sigma-Aldrich</i>	<i>1.0 M, P0180</i>
<i>phenol</i>	<i>Merck</i>	<i>Ph Eur, 1002060250</i>
<i>spermidine</i>	<i>Sigma-Aldrich</i>	<i>≥ 99%, S2626</i>
<i>adenosine</i>	<i>Sigma-Aldrich</i>	<i>≥ 99%, A9251</i>
<i>ESI Tune Mix</i>	<i>Agilent</i>	<i>G1969-85000</i>

Columns

LC Columns

Name	Manufacturer	Description
<i>BEH C8</i>	<i>Waters</i>	<i>130 Å, 1.7 µm, 2.1 mm x 150 mm, 186003377</i>
<i>BEH Amide</i>	<i>Waters</i>	<i>130 Å, 1.7 µm, 2.1 mm x 150 mm, 186004802</i>

GC Columns

Name	Manufacturer	Description
<i>DB-5</i>	<i>J&W Scientific</i>	<i>123-5032, DB-5, 0.32 mm ID x 30 m, 0.25 µm film thickness</i>



Methods

Cultivation and extraction of *C. elegans*

Overnight LB broth cultures of the test bacterial strains were spread onto a 9 cm diameter NGM plates and incubated at 37°C for 24 hours. Briefly, three day old L4 stage or adult hermaphrodite worms raised on *E. coli* OP50 were washed extensively in M9 buffer to remove residual bacteria and 1500 worms were transferred onto feed plates seeded with one of the bacterial cultures, or onto non-seeded plates to obtain starved worms. Plates were incubated at 25°C for 24 hours and worms were then washed extensively with multiple M9 buffer changes to remove residual bacteria. After the final wash, the buffer was replaced with 1 mL of cold 50:50 (v/v) MeOH:water and flash frozen in liquid nitrogen.

Samples were thawed on ice prior to analysis and all subsequent steps of the extraction were carried out on ice. To disrupt the worms, the mixture was sonicated for 15 min in a sonic bath with ice-cold water and vortexed every 3 to 5 minutes for metabolite extraction. After centrifugation at 14,000 rpm for 10 min at 4°C the supernatant was transferred to an on ice pre-chilled eppendorf cup and snap-frozen in liquid nitrogen and stored at -80°C prior to analysis. The remaining debris pellet was transferred to a 4 ml Teflon-lined glass vial with 100 µl MeOH and 1 ml MTBE was added to extract lipids. After shaking at 450 rpm for one hour at room temperature 500 µl water were added and the sample were shaken again for a half hour. Phases were separated by low speed centrifugation. The upper organic phase was transferred to a fresh 4 ml Teflon-lined glass vial and solvent was evaporated in a vacuum centrifuge. The residue was dissolved in 10 parts water / 12 parts MeOH and 4 parts MTBE to yield a concentration of 5000 worms/ml.

HeLa infection protocol

HeLa cells were seeded with 3e5 cells in 5 ml DMEM with FCS into each well UpCell™ 6 well plates and allowed to grow overnight. Bacterial cultures were prepared also overnight in LB medium. Phenol red free DMEM without FCS and sterile PBS were warmed to 37°C. Additionally, ice-cold PBS was prepared. Bacteria were prepared in DMEM without phenol red with a MOI of 1:20. DMEM was removed from cells and 3 ml DMEM without phenol red and FCS were added gently per well. 1 ml of bacterial suspension was added to each and plates were centrifuged at 500 x g for 5 minutes. Infection was allowed to proceed for 30 minutes at 37°C. Non adherent or internalized bacteria were removed by washing cells gently with 37°C warm PBS. After washing 3 ml ice cold PBS was added to each well and the plates were placed onto cold blocks to lift cells from the surface. The suspension was transferred to two 15 ml conical tubes and remaining cells in the wells were transferred with 1 ml of additional ice cold

PBS. Samples were shortly vortexed and centrifuged at 1000 x g for 2 minutes. Samples were either directly frozen in liquid nitrogen or freeze dried.

Lipid extraction from HeLa cells

Lipids from lyophilized samples were extracted with methyl tert-butyl ether (MTBE). The samples were suspended in 200µl MeOH, transferred to a 4 ml glass vial and 2ml MTBE were added. Samples were extracted for 30 min at RT with 430 rpm in a rotary shaker. 1ml water was added for phase separation. After centrifugation the upper organic phase was transferred to fresh 4ml glass vial and dried under vacuum. The residue was dissolved in 10 parts water / 12 parts MeOH and 4 parts MTBE to yield a concentration of 12.5e6 cells/ml. To compare if remaining *P. aeruginosa* in the infected HeLa cells are disturbing subsequent analysis, separate cultures of *P. aeruginosa* PA14 and HeLa cells were prepared and extracted as mentioned above.

Cultivation and extraction of HeLa cells expressing ExoY

HeLa cells transfected with a vector containing full length *exoY* or with an empty vector were cultured in DMEM for 24 hours. Expression of ExoY was controlled with Western blotting. Cells were quenched with cold MeOH/water, scrapped of the plates and directly frozen in liquid nitrogen and stored at -80°C until extraction and analysis. For extraction cells were thawed on ice and lysed in an ice cold sonic bath for 15 minutes. Cellular debris was removed by centrifugation at 14,000 rpm and 4°C for 15 minutes. The supernatant was directly transferred to a fresh eppendorf cup.

Metabolic profiling using DI-ICR-FT/MS

Metabolite extracts were diluted 1:50 with 70% MeOH before analysis. The analysis was performed on a Bruker solariX equipped with a 12T magnet and an Apollo II ion source. The samples were introduced with a syringe pump at a flow rate of 120µl/h. Settings for the ion source were as followed: drying gas temperature = 200°C, drying gas flowrate = 2.4 L/min, nebulizer gas flow = 1.1 L/h, capillary voltage = 4500 V, spray shield = -500V. The mass spectrometer was externally calibrated on cluster of arginine (10 ppm in MeOH). Spectra were obtained in positive and negative mode with an m/z range from 100 to 1000 at 2 mega words and 300 scans were accumulated for one spectrum. All tubing's, the syringe and the sprayer needle were washed with first 50% MeOH, followed by 100% water with 0.1% formic acid, 50% MeOH, 100% ACN and 100% isopropanol and finally 50% MeOH again between two different samples to avoid cross contamination of the spectra.

Data processing DI-ICR-FT/MS

Spectra were internally recalibrated and exported as mass list file (.asc) at a signal-to-noise ratio of 3 using Bruker Data Analysis 4.0 (Bremen, Germany). A matrix, containing an aligned mass list with the corresponding intensities for each sample was generated using an in-house written software tool with a window width of 1.0 ppm for peak alignment. All further calculations and filtering were done in Microsoft Excel 2010. Masses with an occurrence less than 5% in all samples were excluded from further analysis. All masses were uploaded to MassTRIX with a search range of 3.0 ppm for possible metabolite annotation.

Metabolic profiling using UHPLC-UHR-ToF-MS

For UHPLC-UHR-ToF-MS samples were concentrated four times by drying aliquots of the original extract in a SpeedVac vacuum concentrator and re-dissolving in 50% MeOH. Pooled samples served as quality control and were injected ten times before each sample batch to condition the column and after each tenth sample injection. Two modes of separation were employed for metabolite detection, Reversed phase chromatography for separation of mid-to nonpolar metabolites and HILIC for polar metabolites. For analysis a Waters Acquity UHPLC was coupled to Bruker maXis UHR-ToF-MS. 10 µl of sample were injected either onto a BEH C8 or BEH Amide column with 150 mm length and 2.1 mm inner diameter. Column temperature was maintained at 40°C.

During reversed phase chromatography metabolites were eluted using a water-MeOH gradient with 10% MeOH and 0.1% formic acid in buffer A and 100% MeOH and 0.1% formic acid in buffer B. The gradient started at holding 100% A for 2 min and increase to 100% B at 14 min, holding 100% B for 2 min and returning to initial conditions in 0.5 min and re-equilibrating for 3.5 min.

For chromatography of metabolites on the BEH Amide column buffer A consisted of 95% ACN, 5 % water, 10 mM NH₄Ac and 0.1% formic acid. Buffer B was 50% ACN, 50% water, 10 mM NH₄Ac and 0.1% formic acid. The gradient started at 100% A holding for 2 min and an linear increase to 100 %B at 15 min, holding 100% B for 4.5 min and returning to initial conditions in 0.5 min and re-equilibrating for 3.5 min.

Settings for the ion source were as followed: drying gas temperature = 200°C, drying gas flowrate =8.0 L/min, nebulizer gas pressure = 2.0 bar, capillary voltage = 4500 V in positive and 3500 V in negative mode, end plate offset = -500V, acquisition speed = 2 Hz with rolling average of 0.5 Hz. The mass range was set from 50 to 1200 m/z. From 0.05 to 0.3 minutes of chromatography a calibrant mixture consisting of 1:4 diluted ESI tune mix was injected using a

six port valve at the MS each run. Injection order was randomized and every tenth injection was a QC. Prior the sample batch one solvent blank and ten times QC was injection to equilibrate the column.

Lipid profiling using UHPLC-UHR-ToF-MS (*C. elegans*)

Lipids were separated with modified version of the method described by Hu et al. [21] on a Waters Acquity BEH C8 column with 1 mm inner diameter and 150 mm length. For elution the gradient program shown in Table 21 at a flow rate of 70 μ l/min was used with buffer A consisting of 60% ACN and 40% water and buffer B consisting of 90% iPrOH and 10% ACN, both with 10 mM NH₄Ac and 0.1 % formic acid. Detection was carried out in positive mode using a Bruker maXis UHR-ToF-MS with an Apollo II ESI source. For data-dependent acquisition the Apollo II ESI source and the AutoMSⁿ functionality build in microTOF control were used. ESI setting were similar to the metabolic profiling method above. Mass range was set from 200 to 1500 m/z. For AutoMSⁿ scan rate was set to 5 Hz and 4 precursors were selected with an active exclusion time of 0.25 min after fragmentation.

Table 21: Lipid gradient program used for lipid profiling of *C. elegans* samples.

The gradient was adapted from [21]. Times described in the publication were scaled down to a total runtime of 20 minutes.

<i>Time</i>	<i>%B</i>
0	32
1	32
2.7	45
3.5	52
5.5	58
7.5	66
9.5	70
12	75
14	97
16.7	97
20	32

Lipid profiling using UHPLC-UHR-ToF-MS (HeLa)

Lipids were separated on a BEH C8 column (2.1 x 150 mm) with an iPrOH gradient. Buffer A was 60% ACN and 40% water with 10mM NH₄Ac and 0.1% formic acid and buffer B was 90% iPrOH and 10% ACN with 10mM NH₄Ac and 0.1% formic acid. The gradient started at 0% B, which was held for 2 minutes, followed by a linear increase to 100% B at 16 minutes. 100% B was held for 18 minutes and the gradient returned to 0% B at 18.5 minutes. Column was re-equilibrated for 1.5 minutes. Sample was injected using full loop injection of 10 μ l. Flowrate was set to 0.2 ml/min and column temperature was 40°C. The UHPLC was coupled to a Bruker

maXis equipped with an Apollo II ESI source and data was obtained from positive ionization mode at a mass range of 150-2000 Da.

Data processing UHPLC-UHR-ToF-MS

Data was preprocessed using Genedata Refiner MS 7.5. Several workflows were created to split the system load and increase performance. The first workflow loads the Bruker raw data, applies a chromatogram grid and subtracts chemical noise. Afterwards the file is exported into *.sbf* format. This significantly reduces file size. Each file of a data set was processed separately within a loop activity. The second workflow aligns the different chromatograms to overcome retention time shifts between the samples. Again the aligned data files are exported to *.sbf* format. In the last workflow peaks are detected using the “Summed Peak Detection” algorithm, filtered for an occurrence in minimum 10% of the samples and subjected to isotope clustering. The obtained data matrix was exported to *.gda* and *.xlsx* format for further work and statistical analysis (Figure 58). For the lipid data set different workflows were used. The used AutoMSⁿ algorithm collects data in centroid mode. To make this data usable with Genedata Refiner MS files were imported and profile data was generated using the “centroid to Profile” function between “Load from File” and “Chromatogram Grid”. All other workflows were similar. The generated MS/MS spectra were exported using a VB-Script within Bruker Data Analysis 4.0. The generated files were sorted and aligned with the UHPLC-UHR-ToF-MS data using an in-house written Perl script.

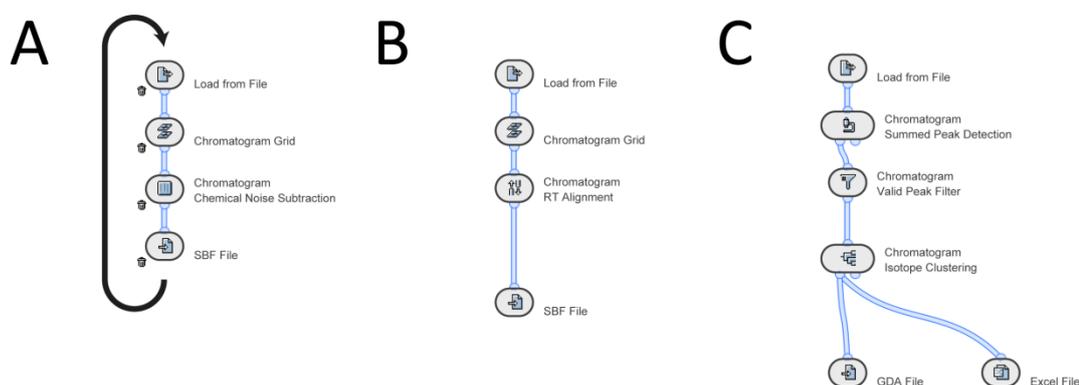


Figure 58: Workflows used in Genedata Refiner MS

(A) The first workflow subtracts chemical background from the chromatograms, which significantly reduces file size. The single steps are processed within a loop for each single data file to reduce memory usage. (B) In the second workflow all chromatograms are aligned to overcome retention time shifts. A tree-guided pairwise alignment algorithm is used. (C) The aligned chromatograms are used for peak detection and isotope clustering. The “Summed Peak Detection” algorithm sums up all chromatograms and uses this summed chromatogram to find chromatographic peaks. These peaks are searched in a second step in the single chromatograms, within certain borders in retention time and mass. Finally the obtained matrix is subjected to identification of isotope clusters and exported to *.gda* and *.xlsx* files.

Fatty acid profiling

For fatty acid profiling a method was adapted from Watts et al. [165]. Lipids from 2000 worms were transesterified using 1 ml 2.5% H₂SO₄ in MeOH. The samples were boiled at 80°C for 1 h in 4 ml glass vials. After cooling to ambient temperature 1.5 ml H₂O and 0.2 ml hexane were added and samples were shaken for 10 minutes at room temperature. The hexane layer was transferred to a GC vial with micro insert and 5 µl were injected using a PTV injector with a split of 50:1 and temperature of 250°C onto a DB-5 column (30 m x 0.32 mm ID, 0.25 µm film thickness). After 4 minute long isothermal step temperature was increased linearly with 4°C/min to 250°C, which was hold for 4 minutes. Helium was used as carrier gas at a velocity of 30ml/min. The FID temperature was set to 300°C and Hydrogen flow was 30 ml/min and synthetic air was 400 ml/min. Helium was used as make-up gas with 30 ml/min.

Raw data was imported into ACD/Labs Spectrus Processor and peaks were detected after baseline correction. Identity of fatty acids was confirmed by matching retention time with an authentic standard.



Data Analysis scripts

ICR-FT/MS spectra export

```
Analysis.Spectra.MassListClear
val=Analysis.Spectra.MassListFind(50, 1000)
Analysis.Spectra(1).ExportMassList "H:\Export\" + analysis.name, daASCII
form.close
```

Export of MS/MS spectra generate by AutoMSⁿ

```
'This script exports a ASCII file for each compound spectrum
'iterate thru all compounds

Filename=left(Analysis.Name, len(Analysis.Name)-2)

Analysis.Compounds.Clear
Analysis.FindAutoMSn

Analysis.Export "F:/AutoMSMS/Chrom/" & Filename & ".mzxml", daMzXML, daLine
Analysis.Compounds.Export "F:/AutoMSMS/Compound/" & Filename & ".csv", daCSV

Dim compound

For i=1 to Analysis.Compounds.Count

    Compoundname=Analysis.Compounds(i).Name
    lower = 0
    upper = 0
    For j = 1 to len(Compoundname)

        if Mid(Compoundname, j, 1)="(" then lower = j+1

    Next

    compound = Mid(Compoundname, lower, 8)
    compound = Replace(compound, ".", ",")

    'Analysis.Compounds(i)(1).Export "F:/AutoMSMS/Result/" & Filename & "_MS1_Cmpd" & CStr(i) & "_" & compound ,
    daASCII
    Analysis.Compounds(i)(2).Export "F:/AutoMSMS/Result/" & Filename & "_MS2_Cmpd" & CStr(i) & "_" & compound ,
    daASCII

Next

Analysis.close

form.close
```

Perl scripts

ICR-FT/MS and LC-MS alignment

```

#!/usr/bin/perl

#read ft masses
print "reading FT masses...";
open(FTIN, "F:/neg/FT_neg.txt");
while(<FTIN>) {
    chomp($_);
    @split = split("\t", $_);
    $ftMass{$split[0]}="";
}
close(FTIN);
print "done\n";

#read lc masses
print "reading LC masses and RT...";
open(LCIN, "F:/neg/LC_amide_neg.txt");
while(<LCIN>) {
    chomp($_);
    @split = split("\t", $_);
    $lcMass{$split[0]}=$split[1];
    $lcMass2{$split[0]}=$split[2];
}
close(LCIN);
print "done\n";

#write results to file
open(OUT, ">amide_neg_stat.txt");
print "comparing datasets...\n";
print "LC mass\tLC RT\tID";
print OUT "LC mass\tLC RT\tID";

for my $ppm (1..15) {
    print "\tmass_$ppm\trealppm_$ppm\tnumber of hits_$ppm";
    print OUT "\tmass_$ppm\trealppm_$ppm\tnumber of hits_$ppm";
}

print "\n";
print OUT "\n";

foreach my $lc (sort keys %lcMass) {

    print "$lc\t$lcMass{$lc}\t$lcMass2{$lc}";
    print OUT "$lc\t$lcMass{$lc}\t$lcMass2{$lc}";

    for my $ppm (1..15) {

        $roundLc=sprintf("%.f", $lc);
        $result{$lc}{$ppm}="--"."#".$ppm;
        $number{$lc}{$ppm}=0;

        foreach my $ft (sort keys %ftMass) {
            $roundFt=sprintf("%.f", $ft);

            @split=split("#", $result{$lc}{$ppm});
            $ppmOld=$split[1];

            if ($roundLc eq $roundFt && $ft>0) {
                $ppmReal = ($lc-$ft)/$ft*1000000;
                if (abs($ppmReal) < $ppm) {
                    print "found\n";
                    $number{$lc}{$ppm}++;
                    if (abs($ppmReal) < abs($ppmOld)) {
                        $result{$lc}{$ppm}=$ft."#".$ppmReal;
                    }
                }
            }
        }

        @split=split("#", $result{$lc}{$ppm});
        $massPrint=$split[0];
        $ppmPrint=$split[1];
        print "\t$massPrint\t$ppmPrint\t$number{$lc}{$ppm}";
        print OUT "\t$massPrint\t$ppmPrint\t$number{$lc}{$ppm}";
    }

    print "\n";
    print OUT "\n";
}
close(OUT);

```

Conversion of ASCII files to MetFrag Batchfile

```
#!/usr/bin/perl -w

open(IN, "H:/lipid.annotated");

while(<IN>) {

    chomp($_);
    @split = split("\t", $_);
    $anno{$split[0]}=$split[2];
    print "$_\n"

}

close(IN);

# create a list of all *.ascii files in
# the current directory

opendir(DIR, "H:/AutoMSMS/Result");
@files = grep(/\.ascii$/,readdir(DIR));
closedir(DIR);

# print all the filenames in our array
foreach $file (@files) {

    @filesplit=split("_", $file);

$filename="$filesplit[0]_filesplit[1]_filesplit[2]_filesplit[3]_filesplit[4]_filesplit[5]_filesplit[6]_filespli
t[7]_filesplit[9]";

    open(IN, "H:/AutoMSMS/Result/$file");

    while(<IN>) {

        chomp($_);
        @line = split(",", $_);

        if ($anno{$line[4]} ne "") {
            open(OUT, ">H:/AutoMSMS/batchfiles2/$filename.mb");
            print OUT "# Parent Mass:$anno{$line[4]}\n";
            print OUT "# Search PPM:50\n";
            print OUT "# Mode:1\n";
            print OUT "# Charge:+\n";

            $length=@line;

            for ($i=8;$i<$length;$i++) {
                @split=split(" ", $line[$i]);
                print OUT "$split[0] $split[1]\n";

            }

            close(OUT);

        }

    }

    close(IN);

}
```


Alignment of LC-MS and LC-MS/MS features from AutoMSⁿ

```

#!/usr/bin/perl

$ppm=10;
$rtdev=10;

#read ft masses
print "reading LC/MS features...";
open(MS, "F:/MS.txt");
while(<MS>) {
    chomp($_);
    @split = split("\t", $_); #input format: mass\rt\ID
    $MS1{$split[0]}=$split[1]; #MS1 = RT
    $MS2{$split[0]}=$split[2]; #MS2 = ID
}
close(MS);
print "done\n";

#read lc masses
print "reading LC/MSMS features...";
open(MSMS, "F:/MSn.txt");
while(<MSMS>) {
    chomp($_);
    @split = split("\t", $_); #input format: mass\rt\ID
    $MSMS1{$split[0]}=$split[1]; #MSMS1 = RT
    $MSMS2{$split[0]}=$split[2]; #MSMS2 = ID
}
close(MSMS);
print "done\n";

#write results to file
open(OUT, ">F:/AutoMSMSmatch.txt");
print "comparing datasets...\n";
print OUT "MS ID\tMSMS ID\tMS mass\tMSMS mass\tMS rt\tMSMS rt\n";

#iterate through masses and compare
$i=0;
foreach my $msFeature (sort keys %MS1) {
    $roundmsFeature=sprintf("%.f", $msFeature);

    foreach my $msmsFeature (sort keys %MSMS1) {
        $roundmsmsFeature=sprintf("%.f", $msmsFeature);

        @split=split("#", $result{$lc});
        $ppmOld=$split[1];

        if ($roundmsFeature eq $roundmsmsFeature) {

            $ppmReal = ($msmsFeature-$msFeature)/$msFeature*1000000;

            if (abs($ppmReal) < $ppm) {

                $rtmsFeature = $MS1{$msFeature};
                $rtmsmsFeature = $MSMS1{$msmsFeature};

                $rtdevReal = ($rtmsFeature-$rtmsmsFeature)/$rtmsFeature*100;

                if (abs($rtdevReal) < $rtdev) {
                    print
"$MS2{$msFeature}\t$MSMS2{$msmsFeature}\t$msFeature\t$msmsFeature\t$MS1{$msFeature}\t$MSMS1{$msmsFeature}\n";
                    print
"$MS2{$msFeature}\t$MSMS2{$msmsFeature}\t$msFeature\t$msmsFeature\t$MS1{$msFeature}\t$MSMS1{$msmsFeature}\n";
                }
            }
        }
    }
}

close(OUT);

```

Curriculum vitae

Born on the 6th of May 1986; German

Present address:

Helmholtz Zentrum München

Research Unit Analytical BioGeoChemistry

Ingolstädter Landstraße 1

85764 Neuherberg

Tel: +49(0)89-3187-4755

Fax: +49(0)89-3187-3358

E-Mail: michael.witting@helmholtz-muenchen.de

Lab website: www.helmholtz-muenchen.de/bgc



Current position:

Postdoctoral fellow at Research Unit Analytical BioGeoChemistry, Helmholtz Zentrum München

Education and professional experience:

2005 – 2009: Studies of Applied Chemistry, with functional direction to Biochemistry, Georg-Simon-Ohm Hochschule Nürnberg

2008: Research term, Institute of Biochemistry, University of Innsbruck, Project: “Functional analysis of the chicken BASP1 promotor”

2009: Diploma Thesis, Georg-Simon-Ohm Hochschule Nürnberg, Laboratory of Bioanalytical Chemistry, Project: “Bioanalytischer Nachweis von Histamin in biologischen Matrices” (engl. „Bioanalytical analysis of histamine in biological matrices”)

2009 – 2012: PhD thesis at Research Unit Analytical BioGeochemistry, Helmholtz Zentrum München

2009 – 2012: Research Assistant, Department of Genome oriented Bioinformatics, TU München

2013 – Now: Postdoctoral fellow at Research Unit Analytical BioGeoChemistry, Helmholtz Zentrum München



Scientific communications:**Peer reviewed:**

MassTRIX Reloaded: Combined Analysis and Visualization of Transcriptome and Metabolome data

Brigitte Wägele, Michael Witting, Philippe Schmitt-Kopplin, Karsten Suhre

PLoS ONE 7(7): e39860. doi:10.1371/journal.pone.0039860

A metabolomics approach to study bacterial virulence in C. elegans

Michael Witting, Marianna Lucio, Dimitrios Tziotis, Brigitte Wägele, Karsten Suhre, Romé

Voulhoux, Steve Garvis and Philippe Schmitt-Kopplin

PLoS ONE (in revision)

Phenotype of htgA (mbiA), a recently evolved orphan gene of Escherichia coli and Shigella, completely overlapping in antisense to yaaW

Lea Fellner, Niklas Bechtel, Michael Witting, Svenja Simon, Philippe Schmitt-Kopplin, Daniel Keim, Siegfried Scherer and Klaus Neuhaus

FEMS Microbiology Letters, in press

Ultrahigh resolution mass spectrometry-based metabolic characterization reveals cerebellum as a disturbed region in two animal models

Shuhai Lin, Basem Kanawati, Liangfeng Liu, Michael Witting, Min Li, Jiandong Huang, Philippe Schmitt-Kopplin and Zongwei Cai

Talanta 118(0): 45-53

Book chapter:

Ultrahigh Resolution Mass Spectrometry based Non-targeted Microbial Metabolomics

Michael Witting, Marianna Lucio, Dimitrios Tziotis and Philippe Schmitt-Kopplin

in Karsten Suhre “Genetics meets Metabolomics: From Experiment to Systems Biology”, 2012, Springer

Transcriptome and Metabolome Data Integration – Technical Perquisites for Successful Data Fusion and Visualization

Michael Witting and Philippe Schmitt-Kopplin

Carolina Simo, Alejandro Cifuentes, Virginia Garcia-Canas “Fundamentals of Advanced Omics Technologies: From Genes to Metabolites”, in press