

Mass Spectrometry based Bioanalytics on Model Organisms

Dissertation zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von

Christian Treitz

Kiel, 2017

Erster Gutachter: Prof. Dr. Andreas Tholey

Zweiter Gutachter: Prof. Dr. Matthias Leippe

Tag der mündlichen Prüfung: 22.02.2018

Zum Druck genehmigt: 22.02.2018

Danksagung

An dieser Stelle möchte ich mich bei allen bedanken, die mit ihrer Unterstützung zu der Fertigstellung dieser Dissertation beigetragen haben.

Mein herzlicher Dank gilt meinem Doktorvater, Professor Andreas Tholey, der mit seiner kompetenten Betreuung und unermüdlichen Motivation maßgeblich zur erfolgreichen Fertigstellung dieser Arbeit beigetragen hat. Das entgegengebrachte Vertrauen, die interessanten Projekte und den konstruktiven Austausch in zahlreichen wissenschaftlichen und persönlichen Gesprächen weiß ich sehr zu schätzen.

Bei den jetzigen und auch ehemaligen Kollegen der Arbeitsgruppe möchte ich mich nicht nur für das positive Arbeitsklima sondern auch für die bereitwillige Anteilnahme an Freizeitaktivitäten und dem damit gelieferten Ausgleich bedanken.

Für die Durchsicht, Korrektur und Diskussion meiner Arbeit bedanke ich mich bei Dr. Liam Cassidy, Dr. Diego Yepes, Dr. Tomas Koudelka, Jan Leipert und Patrick Kaleja. Insbesondere Liam hat einen Großteil der Arbeit korrekturgelesen und meine Vergehen an der englischen Sprache tapfer ertragen; *thank you*. Beim Thema Korrekturlesen darf ich auch Joanna Tucher nicht vergessen, die in der kurzen Zeit, zurück zu Besuch im hohen Norden, so lieb war die Zusammenfassung zu lesen.

Des Weiteren möchte ich mich bei den zahlreichen Kooperationspartnern bedanken, ohne deren Hilfe die behandelten Forschungsprojekte nicht zustande gekommen wären.

Dabei gilt mein Dank Professor Matthias Leippe und den Mitarbeitern des Zoologischen Instituts der Christian-Albrechts-Universität für das Konzept zur Erforschung erkrankter Fadenwürmer. Ohne die zahlreichen Ideen und richtungsweisenden Ratschläge zur Umsetzung des Projekts wäre es zweifellos nicht zu einem erfolgreichen Abschluss gekommen. Insbesondere bedanke ich mich auch bei der ehemaligen Mitarbeiterin Aylin Höckendorf für ihre Unterstützung bei der tagtäglichen (und nächtlichen), scharenweisen Anzucht und Ernte von Würmern.

Den Freunden und Kollegen aus Toulouse, Brice Enjalbert, Fabien Létisse und Pierre Millard, die unter der Schirmherrschaft von Professor Jean Charles Portais forschen, möchte ich für die Gastfreundschaft im sonnigen Süden Frankreichs einen Tost aussprechen. Obwohl der kurze Forschungsaufenthalt der aeroben Kultivierung von Mikroben gewidmet war, erinnere ich mich mit besonderer Wehmut an die exzellenten Gärungsprodukte des Landes.

Des Weiteren möchte ich mich bei Dr. Cleo Pietschke und Dr. Sebastian Fraune für meine Einbeziehung in ein weiteres sehr interessantes Forschungsprojekt, die erfolgreiche Kooperation und ihr persönliches Engagement bedanken.

Tief verbunden und dankbar bin ich auch meinen Freunden und meiner Familie für ihre Unterstützung und ihr Verständnis dafür, dass sie in den letzten Jahren viel auf meine Anwesenheit verzichten mussten.

Herzlichen Dank, Christian Treitz

Lebenslauf

Persönliche Daten

Name: Christian Treitz
Adresse: Olshausenstraße 7
24118 Kiel
Mail: c.treitz@gmx.net
Geburtsdatum: 25.05.1981
Geburtsort: Landau
Staatsangehörigkeit: Deutsch

Schulbildung

09 / 1991 – 06 / 2000: Realgymnasium Völklingen
06 / 2000: Allgemeine Hochschulreife

Hochschulstudium

11 / 2001 – 11 / 2010: Diplomstudium Biologie
Schwerpunkt Human- und Molekularbiologie
Universität des Saarlandes

11 / 2003: Vordiplom

02 / 2004 – 07 / 2006: Laborassistent am Korea Institute of Science and
Technology Europe

07 / 2009 – 10 / 2009: Diplomprüfungen
Nebenfach: Genetik, Molekularbiologie
Hauptfach: Biochemie

10 / 2009: Hilfwissenschaftler am Institut für Experimentelle
Medizin (IEM), Arbeitsgruppe Systematische
Proteomforschung, Christian-Albrechts-Universität
zu Kiel (CAU)

12 / 2009 – 10 / 2010: Diplomarbeit zum Thema „Quantitative Analyse von
Proteinen der Hefe Zellwand“ am IEM, AG
Systematische Proteomforschung (CAU)

seit 01 / 2011: Beginn des Promotionsstudiums am IEM, AG
Systematische Proteomforschung (CAU)

Publikationen

Teile dieser Dissertation finden sich in folgenden Veröffentlichungen wieder:

Treitz C, Cassidy L, Höckendorf A, Leippe M, Tholey A (2015).
Quantitative proteome analysis of *Caenorhabditis elegans* upon exposure to nematicidal *Bacillus thuringiensis*.
J Proteomics, 113: 336-350.

Treitz C, Enjalbert B, Portais JC, Letisse F, Tholey A (2016).
Differential quantitative proteome analysis of *Escherichia coli* grown on acetate vs. glucose.
Proteomics, 16: 2742-2746.

Pietschke C, Treitz C, Forêt S, Schultze A, Künzel S, Tholey A, Bosch T, Fraune S (2017).
Host modification of a bacterial *quorum* sensing signal induces a phenotypic switch in bacterial symbionts.
Proc Natl Acad Sci, 114: E8488-E8497, doi: 10.1073/pnas.1706879114.

Tholey A, Treitz C, Kussmann M, Bendixen E, Schrimpf SP, Hengartner MO. (2013).
Model organism proteomics - from holobionts to human nutrition.
Proteomics, 13: 2537 - 2541.

Mitwirkung an Veröffentlichungen zusätzlich zu den Teilen dieser Dissertation:

Wilde SC, Treitz C, Keppler JK, Koudelka T, Kalpani K, Tholey A, Rawel HM, Schwarz K (2016).
 β -Lactoglobulin as nanotransporter - part II: characterization of the covalent protein modification by allicin and diallyl disulfide.
Food Chem, 197: 1022-1029.

Leipert J, Treitz C, Leippe M, Tholey A (2017).
Identification and quantification of N-acyl homoserine lactones involved in bacterial communication by small-scale synthesis of internal standards and matrix-assisted laser desorption/ionization mass spectrometry.
J Am Soc Mass Spectrom, in press, doi: 10.1007/s13361-017-1777-x.

Patent:

Wilde SC, Keppler JK., Schwarz K. Palani K, Treitz C, Koudelka T, Tholey A (2017).
Food additive comprising bio-active components of garlic
[google.com/patents/WO2017050314A1](https://www.google.com/patents/WO2017050314A1)

Table of contents

Table of contents	1
List of abbreviations	5
Zusammenfassung	7
Summary	9

Chapter 1. General introduction

1.1 Bioanalytical research on model organisms	11
1.2 Proteomics sample preparation	13
1.3 Protein and peptide separation	15
1.4 High-performance liquid chromatography.....	18
1.5 Mass spectrometry.....	22
1.6 Peptide identification.....	24
1.7 Quantitative proteomics.....	26

Chapter 2. Quantitative proteome analysis of *Caenorhabditis elegans* upon exposure to nematocidal *Bacillus thuringiensis*

2.1 Abstract.....	30
2.2 Introduction.....	32
2.3 Material and methods	
2.3.1 Materials.....	35
2.3.2 <i>C. elegans</i> cultivation and sample preparation.....	35
2.3.2.1.1 Synchronization by hypochlorite treatment.....	35
2.3.2.1.2 Saccharose floating	36
2.3.2.1.3 Worms sifting using nylon net filters.....	36
2.3.2.1.4 Validation of sample preparation procedures by microscopic survey..	36
2.3.2.1.5 Cultivation of <i>C. elegans</i> and incubation with <i>Bacillus thuringiensis</i> ...	37
2.3.3 Protein extraction, digestion and isobaric labeling.....	38
2.3.4 Two-dimensional liquid chromatography coupled to mass spectrometry.....	38

2.3.5	Data processing.....	40
2.4	Results and discussion	
2.4.1	Cultivation of <i>C. elegans</i> and sample preparation for proteomics.....	41
2.4.1.1	Measurement of <i>C. elegans</i> dimensions by microscopy.....	43
2.4.1.2	Validation of nylon net filter assisted sorting.....	44
2.4.1.3	Sample preparation for <i>C. elegans</i> -pathogen interaction proteomics....	45
2.4.2	Quantitative proteome analysis by isobaric labeling and 2D-LC MS	47
2.4.3	Proteins identified from pathogenic <i>B. thuringiensis</i>	53
2.4.4	Effects of pathogenic challenge on the <i>C. elegans</i> proteome.....	54
2.4.5	Differentially abundant protein families.....	55
2.5	Conclusion.....	62

Chapter 3. Differential quantitative proteome analysis of *Escherichia coli* grown on acetate vs. glucose

3.1	Abstract.....	64
3.2	Introduction.....	66
3.3	Material and methods	
3.3.1	Cultivation of <i>E. coli</i>	68
3.3.2	Proteomics sample preparation.....	68
3.3.3	Peptide fractionation by high pH reversed-phase HPLC.....	69
3.3.4	LC-MS analysis of pooled fractions.....	69
3.3.5	Data analysis.....	71
3.3.6	Data visualization and integration with external data sources.....	73
3.4	Results and discussion	
3.4.1	Differential cultivation.....	74
3.4.2	2D-LC-MS analysis.....	74
3.4.3	Procedures to reduce ratio-compression and underestimation.....	78
3.4.4	Qualitative assessment of the dataset.....	82
3.4.5	Comparison to previous system-wide studies.....	85
3.4.5.1	Correlation with a transcript level expression profile.....	85
3.4.5.2	Correlation with relative protein abundance in strain BW25133.....	87
3.4.6	Functional annotation of differentially abundant proteins.....	89
3.4.6.1	Central carbon metabolism.....	89

3.4.6.2	Fatty acid metabolism.....	93
3.4.6.3	Energy production.....	94
3.4.6.4	Protein synthesis.....	95
3.4.6.5	Bacterial chemotaxis and motility.....	105
3.4.6.6	Periplasmic binding proteins and ABC-transporters.....	107
3.4.6.7	Phosphotransferase systems.....	108
3.4.7	Implications towards competitive environments.....	109
3.5	Conclusion.....	110

Chapter 4. Identification and characterization of N-acyl homoserine lactones by mass spectrometry

4.1	Abstract.....	112
4.2	Introduction	
4.2.1	<i>Quorum</i> sensing.....	114
4.2.2	<i>Quorum</i> quenching.....	116
4.2.3	Bioanalytics of N-acyl homoserine lactones.....	117
4.3	Material and methods	
4.3.1	Analysis of N-acyl homoserine lactones by direct injection ESI MS.....	119
4.3.2	Analysis of the <i>quorum</i> quenching mechanism of <i>Hydra</i> polyps.....	120
4.3.3	Analysis of <i>quorum</i> quenching samples by MALDI-TOF MS.....	120
4.3.4	Analysis of <i>quorum</i> quenching samples by LC-ESI MS.....	120
4.3.5	Heterologous expression of <i>Curvibacter</i> sp. AHL-synthases.....	121
4.3.6	Identification of AHL by LC-ESI MS.....	122
4.4	Results and discussion	
4.4.1	LC-ESI MS of small compounds.....	123
4.4.2	HPLC separation of AHL.....	124
4.4.3	Fragmentation of AHL in MS/MS experiments.....	129
4.4.3.1	Fragmentation of the unsubstituted AHL in MS/MS experiments.....	132
4.4.3.2	Fragmentation of the 3O-AHL in MS/MS experiments.....	135
4.4.3.3	Fragmentation behavior of 3OH-AHL in MS/MS experiments.....	137
4.4.4	Analysis of <i>Hydra quorum</i> quenching activity.....	141
4.4.5	Identification of <i>quorum</i> sensing signals produced by AHL-synthases	146
4.5	Conclusion.....	154

Appendix

Chapter 1. References..... 158

Chapter 2. References..... 161

 Java-Script A 2.1..... 167

 VBA-Script A 2.2..... 168

Chapter 3. References..... 169

Chapter 4. References..... 175

List of abbreviations

2D-GE	Two-dimensional gel electrophoresis
2D-LC	Two-dimensional liquid chromatography
ABC transporter	ATP binding cassette transporter
ACP	Acyl carrier proteins
AGC	Automatic gain control
AHAS	Acetohydroxy acid synthases
AHL	N-acyl-homoserine lactone
BCA assay	Bicinchoninic acid assay
<i>Bt</i>	<i>Bacillus thuringiensis</i>
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CHCA	Alpha-Cyano-4-hydroxycinnamic acid
CID	Collision induced dissociation
CID-IT	CID spectra acquired with the ion trap
Cry toxin	Crystal toxin
CTLD	C-type lectin-like domain containing proteins
DIGE	Differential two-dimensional gel-electrophoresis
EDTA	Ethylenediaminetetraacetic acid
EF	Enrichment factor
ESI	Electrospray ionization
FA	Formic acid
FDR	False discovery rate
FT	Fourier transform
FTMS	Fourier transform mass spectrometry
FWHM	Full width at half maximum
GABA	4-Aminobutanoate
GeLC-MS	Gel electrophoresis followed by LC-MS
GF	Germfree
GO	Gene ontology
HCD	Higher energy C-trap dissociation; higher-energy collisional dissociation
HCD-FT	HCD spectra with the Orbitrap
HILIC	Hydrophilic interaction chromatography
HPLC	High-performance liquid chromatography
HSL	Homoserine lactone
IP-RP	Ion-pair reversed-phase
IT	Ion trap
iTRAQ	Isobaric Tags for Relative and Absolute Quantification
LC	Liquid chromatography
LC-MS	Liquid chromatography coupled to mass spectrometry
LTQ	Liner quadrupole ion trap
<i>m/z</i>	Mass to charge ratio
MALDI	Matrix-assisted laser desorption/ionization

List of abbreviations

MCP	Methyl accepting chemotaxis protein
MS	Mass spectrometry
MS/MS	Tandem mass spectrometry
MTA	Methylthioadenosine
NCE	Normalized collision energy
NGM	Nematode growth medium
NHS	N-hydroxysuccinimide
OD ₆₀₀	Optical density at 600 nm
P2T	Precursor to transmission
PBP	Periplasmic binding proteins
PFM	Peptone-free medium
PI	Prediction interval
PON	Paraoxonase
ppm	Parts per million
PSM	Peptide-spectrum-match
PTM	Posttranslational modification
PTS	Phosphotransferase system
RDBE	Ring and double bond equivalents
RF	Radio frequency
ROI	Region of interest
RP-HPLC	Reversed-phase-HPLC
rpm	Rounds per minute
RP-TLC	Reverse-phase thin layer chromatography
RT	Retention-time
S2I	signal to interference
SAM	S-adenosyl-methionine
SAX	Strong anion exchange
SCX	Strong cation exchange
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SILAC	Stable isotope labeled amino acids in cell culture
SPE	Solid-phase extraction
Tat-system	Twin-arginine translocation
TCA cycle	Tricarboxylic acid cycle
TCEP	Tris 2-carboxyethyl phosphine
TEA	Triethylamine
TEAB	Triethylammonium bicarbonate
TFA	Trifluoroacetic acid
TLC	Thin layer chromatography
TMT	Tandem Mass Tag
TOF	Time of flight
UHPLC	Ultra-high-performance liquid chromatography
v/v	volume/volume
XIC	Extracted ion chromatogram

Zusammenfassung

Das Ziel dieser Arbeit war die Etablierung und Anwendung massenpektrometrischer Analysemethoden zur Untersuchung biologisch relevanter Fragestellungen an Modellorganismen. Im Vordergrund stand dabei die relativ-quantitative Proteomanalyse. Als sekundäre Technologie war die mehrdimensionaler Flüssigkeitschromatographie ein Schwerpunkt der Arbeit. Um eine möglichst umfassende massenspektrometrische Analyse komplexer, aus Vollproteom-Isolaten hergestellter Peptidlösungen zu gewährleisten, ist eine effiziente Fraktionierung und Auftrennung der Probe essentiell. Hochleistungsflüssigkeitschromatographie in Kombination mit Massenspektrometrie (LC-MS) wurde in den drei Teilprojekten dieser These angewandt.

Das erste Projekt umfasste die Untersuchung von Interaktionen zwischen dem Fadenwurm *Caenorhabditis elegans* als Wirt und dem Bakterium *Bacillus thuringiensis* als Pathogen. Um diese Pathogen-Wirt Interaktionen auf Proteinebene zu charakterisieren, wurde ein experimenteller Workflow ausgearbeitet und optimiert. Dazu wurde *C. elegans* in vier biologischen Replikaten jeweils pathogenen und nicht-pathogenen Bakterien ausgesetzt und die unterschiedlich behandelten Würmer wurden mittels Nylonnetz-Filtern geerntet. Nach Extraktion der Proteine und proteolytischer Prozessierung wurden die Peptide isobar markiert, um eine quantitative Analyse mittels Tandem-Massenspektrometrie zu ermöglichen. Zwölf biologische Proben wurden in drei Markierungsexperimenten mittels zweidimensionaler Hochleistungsflüssigkeitschromatographie (2D-HPLC) aufgetrennt und massenspektrometrisch analysiert. In der gesamten Versuchsreihe wurden quantitative Daten für mehr als 3.600 Proteine gewonnen, die einen detaillierten Überblick der physiologischen Auswirkungen des pathogene Bakteriums auf den Nematoden darstellen. Mit 288 statistisch relevanten, differenziell abundanten Proteinen konnten verschiedene Proteinklassen mit der angeborenen Immunantwort von *C. elegans* in Verbindung gesetzt werden.

In einem zweiten Projekt wurde die vorher etablierte Methode angewandt, um die Unterschiede zwischen dem Acetat- und Glukose-Stoffwechsel des gramnegativen Bakteriums *Escherichia coli* auf Proteinebene zu charakterisieren. In einem einzigen Experiment wurden jeweils drei Replikate der unterschiedlich kultivierten bakteriellen Proben mittels isobarer Markierung und 2D-LC-MS analysiert. Mit über 2.000 quantifizierten Proteinen lieferte diese differenzielle Proteomanalyse Daten für fast die Hälfte aller in *E. coli* vorhergesagten Proteine. Der gewonnene Datensatz beschreibt damit die proteomischen Unterschiede zwischen den untersuchten Stoffwechselzuständen in bis *dato* unerreichem Detail. Die Anpassung der meisten Enzyme der

zentralen Stoffwechselwege wurden detektiert und insbesondere die Integration mit früheren Forschungsarbeiten lieferte Erkenntnisse über die metabolische Organisation des Bakteriums.

Das dritte Projekt befasste sich mit einem weiteren Aspekt der Interaktionen zwischen einem Wirt und seinen beherbergten Mikroben. Mittels Massenspektrometrie und Chromatographie wurde die molekulare Grundlage eines *quorum quenching* und *quorum sensing* Netzwerkes zwischen dem Süßwasserpolyt *Hydra vulgaris* und dessen assoziierten Mikroben der Gattung *Curvibacter* aufgeklärt. Mittels HPLC gekoppelter Elektrosprayionisations Massenspektrometrie (LC-ESI MS) und Matrix-unterstützter Laser-Desorption/Ionisation Massenspektrometrie (MALDI MS) konnte das *quorum quenching* von N-Acyl-Homoserin-Lactonen (AHL) durch *Hydra* nachgewiesen werden. Außerdem wurde das Produktspektrum von zwei AHL-Synthasen aus *Curvibacter* mittels LC-ESI MS charakterisiert.

Summary

The aim of this work was both the development and application of mass spectrometry (MS) based analytical methods for the investigation of biological research questions on model organisms. The primary focus was on the relative-quantitative proteome analysis by mass spectrometry. A secondary technological aspect of import was the separation of biomolecules by high-performance liquid chromatography (HPLC). The efficient fractionation and separation of complex peptide mixtures from enzymatic full-proteome digests is essential to facilitate a comprehensive mass spectrometric proteome analysis. The combination of HPLC and MS (LC-MS) was applied in the three projects of this thesis.

The first project focused on the interactions between the nematode *Caenorhabditis elegans* and a nematocidal strain of the bacterium *Bacillus thuringiensis*. An experimental workflow was developed and optimized to analyze these host-pathogen interactions at the protein level. The nematode was cultivated in four biological replicates on three different bacterial lawns, including pathogenic and non-pathogenic *B. thuringiensis*. Bacteria-free worm samples of a defined developmental stage were obtained by nylon-net filter assisted sorting. Full-proteome samples were extracted and enzymatically digested. Isobaric labeling of the produced peptides allowed the subsequent quantification by tandem mass spectrometry. In total, twelve biological samples were analyzed in three labeling experiments using two-dimensional liquid chromatography followed by mass spectrometry (2D-LC-MS). The experimental series allowed the quantification of more than 3,600 proteins, providing a detailed overview of the physiological effects of the pathogenic bacterium on the nematode. Among 288 differentially abundant proteins, several protein families could be associated with the innate immune response of *C. elegans*.

In a second project, the previously established workflow for relative quantitative proteomics was applied to analyze the differences between the acetate- and glucose-metabolism of the gram-negative bacterium *Escherichia coli*. Using isobaric labeling, three replicates of the differently cultivated bacterial samples were analyzed in a single 2D-LC-MS experiment. More than 2,000 proteins were quantified covering almost half of the predicted *E. coli* proteome. The acquired dataset provides a highly detailed view of the proteomic differences between the examined metabolic states of the bacterium. The adaptation of most enzymes associated with central metabolic pathways were characterized and the integration of previous research provided insights into the metabolic organization and resource allocation of the bacterium under these conditions.

The third project addressed a different aspect of the interactions between a host and its associated microbes. MS and HPLC methods were used to investigate the molecular basis of a *quorum* quenching and *quorum* sensing network between the fresh-water polyp *Hydra vulgaris* and its main colonizing bacteria of the genus *Curvibacter*. The *quorum* quenching of N-acyl-homoserine lactone (AHL) autoinducers by *Hydra* was detected using HPLC coupled to electrospray ionization mass spectrometry (LC-ESI MS) and matrix-assisted laser desorption/ionization mass spectrometry (MALDI MS). Additionally, the product spectrum of two AHL-synthase from *Curvibacter* were characterized by LC-ESI MS.

1. General introduction

1.1 Bioanalytical research on model organisms

Model organisms are important tools for the development and validation of analytical approaches and for the study of core principles behind complex biological processes (Tholey, Treitz et al. 2013). Life sciences in general have profited immensely from research on model organisms because these model systems provide a set of experimental advantages, which facilitate biological research, in particular compared to research on human subjects. Classical model organisms are usually easy to maintain and manipulate. Preferred characteristics include a fast cultivation time under strict control of environmental variables, in addition to a homogenous genetic background. In summary, they fulfil the basic requirement to generate highly reproducible biological samples for a variety of research applications.

This is of particular importance in the cell-molecular research fields collectively ending with the affix “-omics”, which describe the system-wide bioanalytical characterization of different cellular aspects under the related suffix “-ome”. Popular examples for these disciplines are genomics, transcriptomics, proteomics or metabolomics, which respectively refer to analyses of the complete genome sequence, all transcripts, proteins or metabolites of a biological sample. While the genome sequence is mostly static, the transcriptome, proteome and metabolome are highly dynamic. In particular, transcripts and proteins are gene products and subject to change in gene expression. A proteome is therefore not only defined by the organism but also by the conditions at the time of sampling. In addition to addressing the primary structure of proteins, which includes the identification of processed isoforms and posttranslational modifications, proteomic research also aims to discover the abundance, localization, functions and interactions of all proteins present in a defined sample.

In functional proteomics, protein quantification has received a lot of attention for more than a decade (Ong and Mann 2005). A number of protein quantification techniques, which allow the system wide quantitative analysis of proteins or peptides, have been developed. Using these techniques, data on thousands of biomolecules can be generated in a single experiment (Sechi 2016). Consequently, it is necessary to correct statistical results for multiple testing, which reduces statistical power significantly. This is an additional factor to the biological, environmental and technical variability and inhomogeneous genetic background, which can reduce the statistical power and therefore necessitate an increased number of replicates to generate consistent and statistically significant results. For these reasons, model organisms are especially attractive for quantitative proteomic research, because homogenous samples are easily acquired. The

community working on model organisms is growing steadily and the number of model organisms for which proteome data are being generated is continuously increasing (Jones, Aebersold et al. 2012).

In addition to the transient nature of the proteome, two important attributes are its complexity and the high dynamic range of protein abundances. Two well-established model organisms, *Caenorhabditis elegans* and *Escherichia coli*, can be used as examples to illustrate this point. A recent *Caenorhabditis elegans* reference proteome set contains 26,721 proteins and 1,293 additional isoforms. The reference proteome of *Escherichia coli* strain K-12 substrain MG1655 consists of 4,306 proteins and 9 additional isoforms (UniProt 21.02.2017). The concentration range of proteins in both organisms is estimated to cover more than six orders of magnitude (Wang, Weiss et al. 2012). In addition to the concentration range, proteins are very heterogenic in chemical parameters like isoelectric point or hydrophobicity. This illustrates the complexity of full-proteome extracts as samples and the analytical challenge they pose, even if not all the protein-coding genes are expressed at a given time.

The biological samples, used in the three studies conducted for this thesis, were collected from established model organisms. In Chapter 3, the differences between the metabolic organizations of the bacterium *E. coli* grown on the carbon sources acetate compared to glucose were analyzed by relative protein quantification. With 4,306 possible proteins, *E. coli* samples are of relatively low complexity compared to full proteome samples of other organisms. The generation of a comprehensive dataset to dissect the complete metabolic network on the protein level is nonetheless an elaborate experimental process.

Although the bacterium was grown in an isolated experimental setup, specific growth conditions may still provide implications towards the behavior of the organism in natural environments. For example, the growth on acetate compared to glucose may show parallels to the slow growth in a competitive environment on poor nutrients such as the intestine of endotherms as compared to nutrient-rich environments like fruit and vegetables. Such basic research is especially rewarding in an organism that is not only well characterized, but also of high importance due to its close association with humans. In particular, since the host-microbe interaction can take shape with the human as a host of commensal strains or potentially afflicted with pathogenic strains of the same species.

The key aspect of variability among the biological samples or replicates under study becomes even more important if the subject of research is not cultivated in an isolated setting. Interspecies interactions depend on variables exceeding the molecular reactions and interactions in a single organism are therefore even more complex. Chapter 2 describes a quantitative proteome analysis

of host-pathogen interactions in *C. elegans*. One exceptional characteristic of this model organism, which makes it well suited for the study of innate immunity, is the ease of providing sterile cultures of the nematode. This aspect improves the control of cultures and facilitates the unbiased setup of host-microbe interaction experiments.

Chapter 4 introduces another example of interspecies interactions in the *Hydra* holobiont. Both the *quorum* quenching abilities of the sweet water polyp as a host and the *quorum* sensing capabilities of its main colonizing bacteria were investigated. Instead of proteins, a class of bacterial metabolites, N-acyl homoserine lactones, were analyzed in this project. Although these *quorum* sensing signals of gram-negative bacteria are different organic molecules, the analytical methods used in the proteomic studies were also successfully applied to the analysis of these small compounds.

1.2 Proteomics sample preparation

The focus of the bioanalytical studies in model organisms conducted for this thesis were on proteomics. In particular, shotgun or bottom-up full-proteome analyses were performed. These analyses involve the enzymatic digestion of proteins into peptides, which are more amenable to the subsequent analysis workflow compared to intact proteins. After suitable biological material is obtained, sample preparation for a “full-proteome analysis” involves cell disruption and ideally the solubilization of all proteins to ensure accessibility for subsequent proteolytic digestion. Conversely, the sample cleanup should involve the removal of cell debris and all other biomolecules like nucleic acids, membrane lipids, *et cetera* (Figure 1.1). Unfortunately, the heterogenic physicochemical properties of proteins ensures that this cannot be entirely achieved in a real scenario. Even during cell lysis, localized heating may lead to protein aggregation. Furthermore, highly basic or acidic proteins can be lost due to incompatibility with a sample cleanup procedure or fractionation method used in the proteomics workflow. Hydrophobic proteins like membrane proteins, in particular, may be depleted together with the membrane fraction of samples. This issue persists when common in-solution workflows are applied for bottom-up full-proteome analyses.

Multitudes of cell lysis, sample preparation and clean-up procedures have been developed to make a comprehensive MS based proteome analysis possible. Chaotropic agents like urea or guanidinium chloride and detergents like SDS or Triton X100 can be used to improve hydration of proteins. However, these substances often denature sample proteins and the enzymes used for digestion alike, or they interfere with the following analytics. To resolve this issue, cleavable

surfactants have been developed (Stjerndahl and Holmberg 2005). These detergents can be added to crude cell lysates to enhance protein denaturation and solubilization and may be removed by a specific condition or treatment of the sample (e.g. acidity or UV light), to ensure compatibility with the subsequent procedures.

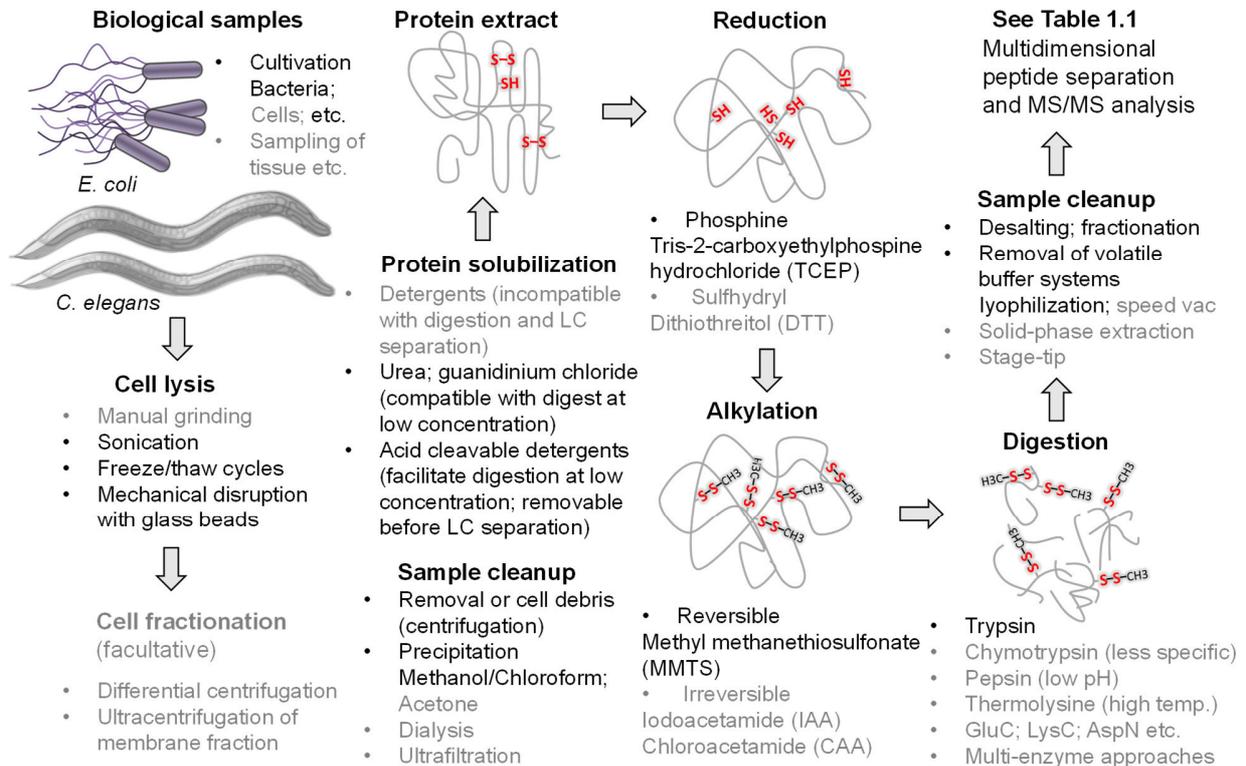


Figure 1.1: Sample preparation scheme for bottom-up proteomics. A brief overview of commonly applied procedures are included for each step in the workflow. Several sample cleanup steps are usually necessary to perform a proteome analysis at the peptide level. The composition of cell lysis buffer has to ensure protein solubilization. Conversely, the composition of protein extracts must be compatible with digestion and ensure the functionality of the used protease. In addition, chemicals interfering with chromatography and MS analysis have to be removed from peptide samples.

Both LC-separation and MS analysis are more difficult to perform on large molecules like intact proteins which have an average mass of ca. 40-50 kDa, but range from less than 5 kDa up to more than 3000 kDa. In contrast the majority of tryptic peptides are in a mass range between 500 Da and 2 kDa, which is better suited for comprehensive profiling of samples by LC-MS. Besides the reasonable mass range of produced peptides, another aspect that makes tryptic digest suitable for MS analysis is the proteolytic site specificity of the endopeptidase. Trypsin cleaves amides with lysine and arginine residues at the non-prime site (Schechter and Berger 1967). The produced peptides can therefore carry positive charges both at the N-terminus and at the C-terminal basic amine side-group of lysine or arginine. This facilitates MS analysis, since the polarity of MS instruments for the analysis of proteins and peptides is generally set to accelerate

positively charged ions. These circumstances have made trypsin the most common peptidase used for bottom-up proteomics.

1.3 Protein and peptide separation

The central technologies used in this thesis were liquid chromatography and mass spectrometry. The development of modern mass spectrometry, which allowed the analysis of large biomolecules by soft ionization techniques, was essential to the emergence of proteomics as a research field. However, due to the complexity and vast concentration range of proteins in biological samples, another central technological branch developed together with proteomics, which addresses the separation or fractionation of proteins and peptides. A subset of the protein and peptide separation methods are summarized in Table 1.1.

Table 1.1: Protein and peptide separation techniques. Several examples of commonly used and combined separation techniques are included. The predominant mode of interaction between stationary phase and analyte or the physicochemical parameters of the analyte, which determine the separation behavior, are noted. The primary method utilized in the following work is emphasized (bold font).

Separation method	Separation mode
Electrophoresis	
○ Tris-Glycine SDS-PAGE (Polyacrylamide gel electrophoresis); (protein)	➤ Molecular weight/size; m/z ratio after binding of SDS proportional to molecular weight
○ Tris-Tricin SDS-PAGE; (peptide)	
○ Gel-eluted liquid fraction entrapment electrophoresis (GELFrEE); (protein)	
○ Isoelectric focusing (IEF); (protein/peptide)	➤ Isoelectric point; charge dependent migration
Reversed phase chromatography	
○ Reversed phase-HPLC (RP-HPLC); C18; (peptide); low/high pH;	➤ Hydrophobicity; hydrophobic effect; Van der Waals or dispersion force
○ Ion-pair RP-HPLC (IP-RP-HPLC) (peptide); C4-;C8-RP-HPLC; (protein)	
Ion exchange chromatography	
○ Strong cation exchange-HPLC (SCX-HPLC); (protein/peptide)	➤ Charge state dependent electrostatic Coulomb interactions
○ Strong anion exchange-HPLC (SAX-HPLC); (protein/peptide)	
Normal phase chromatography	
○ Hydrophilic interaction chromatography HILIC; (protein/peptide)	➤ Elution in order of increasing polarity and solvation; partitioning; hydrogen donor interactions
Size exclusion chromatography SEC	
○ Gel-filtration chromatography (protein/peptide)	➤ Size; hydrodynamic volume
Affinity chromatography	
○ Immunoaffinity chromatography/purification (protein)	➤ Highly specific protein-protein interactions
○ Kinobeads; (kinases)	➤ Specific binding of immobilized kinases inhibitors
○ Titanium-dioxide (TiO ₂); (phospho-peptides)	➤ Coulomb interactions
○ Immobilized ion metal affinity chelate (IMAC); (phospho-peptides)	

In general, both mass spectrometric analysis and separation of proteins can be classified into two methodological approaches. They can either be applied on the intact protein or on peptides after digestion. Mass spectrometric analysis of intact proteins is referred to as top-down proteomics. Conversely, profiling the peptides from proteolytic digests is called bottom-up or shotgun proteomics (Figure 1.2). Both approaches have their advantages and issues. The main drawback of bottom-up approaches is the loss of primary structure information. With MS analysis at the peptide level, it may not be possible to allocate the combinations of multiple modifications on different peptides to a distinct proteoform, or to identify isoforms from alternative translation starts

or splice variants. However, in terms of resolution and sensitivity, the performance of both MS instruments and LC-separation technologies is much higher for analytes in the mass range typically covered for peptide analysis (0.5-6 kDa).

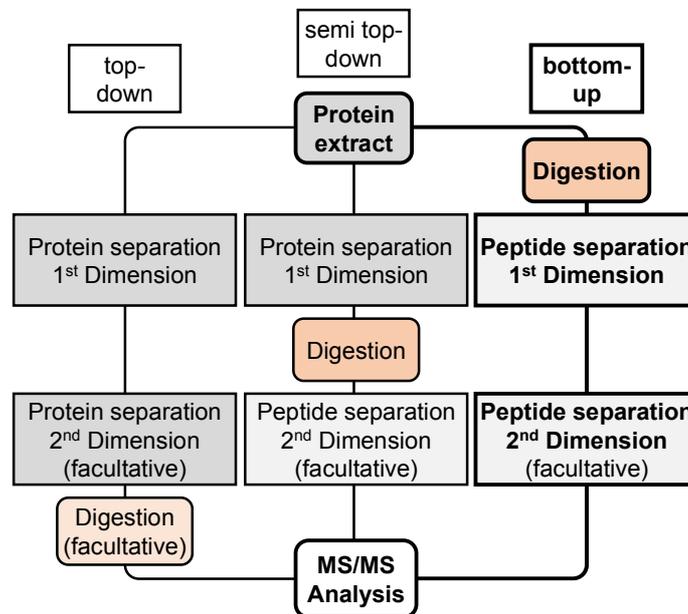


Figure 1.2: Strategies employed in MS based proteome analyses. The different strategies depend on when or if proteins are digested in the course of the workflow. The type of analysis employed in this thesis is emphasized (bold font).

There are several powerful electrophoretic separation techniques for intact proteins, most notably gel electrophoresis, isoelectric focusing or GelFree fractionation systems (Cassidy, Prasse et al. 2016). However, these often involve the introduction of chemicals (e.g. detergents), which are incompatible with downstream analytics and make additional sample cleanup steps necessary. High-performance liquid chromatography (HPLC) is the second major technology for protein and peptide separation used in proteomic research. Although most chromatography methods can be adjusted for intact proteins separation, on the peptide level these methods are usually superior with respect to analysis time, sample throughput, sensitivity and resolution.

Multidimensional separations are often applied in the analysis of complex proteome samples to fully utilize the sensitivity of MS instruments and increase the specificity of quantitative approaches. To this end, fractionation in a first dimension is followed with an orthogonal or semi-orthogonal separation of the produced fractions prior to MS analysis. In particular, the reversed phase (RP-) HPLC separation of peptides features a high resolution and excellent compatibility with subsequent MS analysis. It has become the most commonly applied separation technique directly prior to MS analysis.

The possibilities of different combinations of separation methods in multidimensional approaches are numerous. Due to the exponential increase in samples with each separation dimension, two-dimensional separations are most commonly used. A select few 2D-separations approaches have proven most effective in proteome analyses and have become standard workflows in proteomics.

For several years, two-dimensional gel electrophoresis (2D-GE) was the most frequently used method to separate complex proteome samples (O'Farrell 1975, Rabilloud, Chevallet et al. 2010). In 2D-GE, proteins are first separated on an isoelectric focusing gel strip featuring an immobilized pH gradient to facilitate electrophoretic separation of proteins by their isoelectric point. Afterwards, this gel strip is integrated into a sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) for the orthogonal second dimension of protein separation by molecular weight. 2D-GE can be considered a top-down approach because the processed polypeptides remain intact, retaining primary structure information. However, protein spots embedded in the gel matrix have to be extracted and transferred to a medium amenable with subsequent MS analysis. The most widely adopted procedure for this is in-gel digestion and subsequent analysis of peptide extracts. Following this approach, semi top-down separation have been developed, combining gel electrophoresis, in-gel digestion and a standard RP-HPLC-MS analysis of the digests. Top-down and bottom-up separation approaches or a combination of both can yield different information and analysis depth. The method of choice ultimately depends on what information is of the most scientific interest.

In contrast to the gel-based separations, a different approach to the separation of samples for proteome analysis utilizes two-dimensional liquid chromatography (2D-LC). Here, the classical workflow combines a fractionation of the proteome samples by strong cation exchange (SCX) chromatography, followed by the standard RP-HPLC-MS analysis of these fractions (Opiteck, Lewis et al. 1997). In this approach, the complete separation procedure is usually performed at the peptide level. This scheme provides orthogonality in both dimensions of peptide separations and both LC-separations can be coupled on a single hybrid column or a multicolumn setup (Wolters, Washburn et al. 2001). However, SCX as a first dimension has been shown to provide limited chromatographic efficiency for peptides separation, due to charge state clustering, narrow elution windows of the analytes and generally a lower resolution compared to a reversed phase stationary phase (Delmotte, Lasaosa et al. 2007). Additionally, eluted fractions are inhomogeneous in their peptide load and contain increasing salt contents. This may lead to inconsistent data quality in the following LC-MS analysis or to sample loss, if additional cleanup steps are employed. In contrast, RP chromatography was shown to provide the highest practical peak capacity among commonly used stationary phases (Gilar, Olivova et al. 2005). On the other

hand, two-dimensional combinations of reversed-phase chromatography separations have been shown to provide limited orthogonality (Toll, Oberacher et al. 2005). Even if the mobile phase is modified with different ion-pair reagents, organic modifiers or at different pH, hydrophobic interactions are the predominant cause for peptide retention on the stationary phase. Although a first dimension RP-HPLC separation provides a high resolution and efficient separation, the first dimension fractions show a narrow elution window. This results in an inefficient use of the chromatographic space in the second dimension RP separation. However, fraction concatenation can solve the problem of orthogonality. This simple procedure describes a fraction-pooling scheme in which several equidistant first dimension separation fractions are combined to optimize the use of second dimension chromatographic space. 2D-RP-HPLC separation combined with fraction concatenation was used in both proteomic studies described in this thesis, revealing many advantages of this approach. These include a substantial reduction in analysis time, a simple preparation for subsequent RP-HPLC-MS analysis by vacuum concentration or lyophilization of the volatile organic solvents, the high resolution of separation and homogenous peptide contents of pooled fractions (see Chapter 2 and Chapter 3).

1.4 High-performance liquid chromatography

High-performance liquid chromatography was of central importance to this work and the following chapter describes the technical capabilities of HPLC applications and instrumentation in more detail. In addition to the different column materials (Table 1.1), which permit the separation dependent on different physicochemical parameters of the analyte, the column dimension and other system parameters like pressure limits and flow rate determine the capabilities and possible applications of different systems. Depending on the recommended flow-rate, HPLC systems can be defined as preparative (100 mL - 1000 mL/min) or semi-preparative (5 mL - 50 mL/min), analytical (1 mL - 0.2 mL/min), micro-flow (5 μ L - 50 μ L/min), capillary (1 μ L - 10 μ L/min), and nano-flow systems (1.5 μ L - 50 nL/min). The primary use of preparative systems in proteomics is the collection of high amounts of pure solutes, like purified proteins for research, industrial or pharmaceutical applications. Analytical and microbore column systems are very robust and have a high loading capacity. With a variety of detectors, most commonly UV cells or photodiode arrays, these systems serve for many standard analytical applications. In proteomic research, they are often used to fractionate complex peptide or protein samples in multidimensional separation approaches. Capillary- and nano-flow HPLC systems are often directly coupled to MS instruments. Capillary LC-MS is routinely used to analyze metabolites, lipids and other small organic

compounds. Nano LC-MS platforms are the method of choice to provide the high sensitivity needed to analyze complex protein and peptide mixtures. The HPLC-systems used in this capacity are composed of several functionally distinct compartments. These compartments include a rack to store eluents, degas these solvents and supply the pump system. The pump compartment includes a low-pressure ternary gradient pump and the high-pressure nano-flow pump-system, which can supply a continuous flow of a binary solvent system to the nano-column. A temperature-controlled compartment houses the columns. This device contains a multiport-stator with a flow control valve. Here, capillaries connect and control the flows from the pumps and the sample-inlet to the columns and ultimately towards the MS or an alternative detector. Lastly, an autosampler can be used for cooling, storage and to control the injection of samples.

In addition to the flow-rate, the performance of HPLC instruments used for peptide separation, and in particular the sensitivity of coupled LC-MS platforms, strongly depend on the pressure limits of the instrument and the equipped column. The reason for this is that the efficiency of separation is inversely proportional to particle size and proportional to column length (Horvath, Lukacs et al. 2014). To support pressures needed to separate over longer columns with smaller particle size, HPLC coupled to mass spectrometry has seen a transition from high-performance systems to ultra-high-performance systems (UHPLC, > 400 bar) (Dong 2013). Table 1.2 provides an overview of system parameters including column dimensions (inner diameter and length) and particle size of C₁₈ reversed-phase HPLC and the associated backpressure at defined flow-rates and column temperatures.

Table 1.2: Column dimensions and backpressure. The relationship between different column dimensions and particle size of commercial C₁₈ columns is specified with the associated backpressure observed at defined flow-rates. The column-pressure fluctuates with the composition (and viscosity) of the mobile phase. The listed pressures were observed with 50% (v/v) acetonitrile-water. Columns used in the following work are emphasized (bold font).

LC-mode	Column length	Inner diameter	Particle size	Flow-rate	Pressure at 40°C	
analytical	15 cm	2 mm	3 µm	300 µl/min	ca. 120 bar	HPLC
analytical	25 cm	3 mm	3 µm	300 µl/min	ca. 180 bar	
capillary	25 cm	300 µm	5 µm	4 µl/min	ca. 80 bar	UHPLC
capillary	15 cm	300 µm	3 µm	4 µl/min	ca. 140 bar	
nano	25 cm	75 µm	3 µm	0.3 µl/min	ca. 160 bar	
nano	15 cm	75 µm	2 µm	0.3 µl/min	ca. 180 bar	
nano	25 cm	75 µm	2 µm	0.3 µl/min	ca. 410 bar	
nano	50 cm	75 µm	2 µm	0.3 µl/min	ca 650 bar	

Most common nano-columns used for HPLC are composed of a fused silica capillary (75 µm inner diameter), filled with silica particles of a defined size (Table 1.2). The silica stationary phase can be used for normal phase separation or surface-modified to facilitate the separation dependent on

different physicochemical properties of the analyte. Reversed-phase (RP) separation of peptides is generally performed on HPLC columns modified with octadecyl-chains. Due to steric hindrance between fully hydroxylated silica and the bulky side-chains involved in the reaction, not all of the silanol-groups on the surface react during the bonding process. To reduce the influence of residual silanols on retention and selectivity of RP-LC columns, these groups are end-capped with shorter aliphatic groups (Figure 1.3).

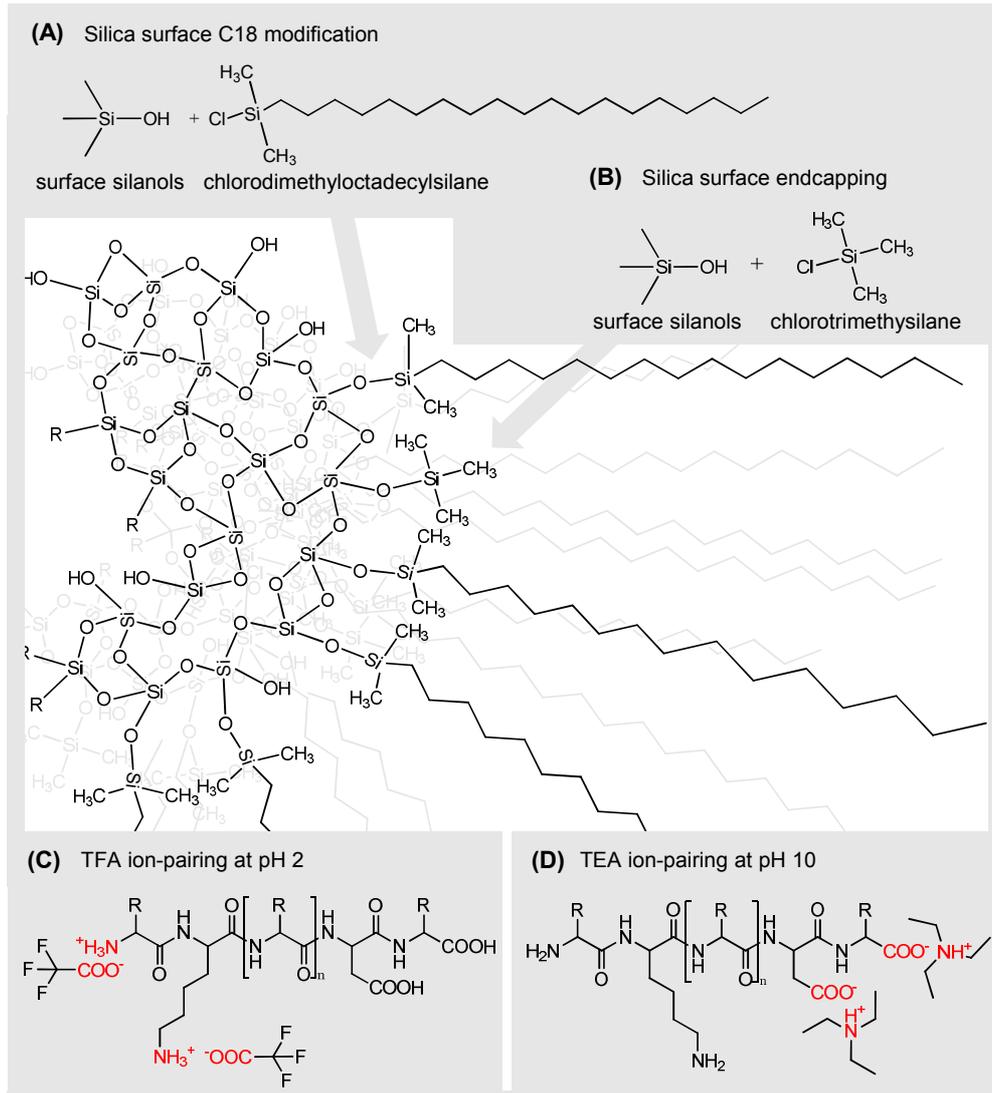


Figure 1.3: Microstructure of C₁₈ RP-columns. (A) The modification of the silica support surface with octadecyl chains is illustrated. (B) Endcapping reduces remaining free silanol-groups and thereby prevents potential hydrophilic interactions between an analyte and the stationary phase. Both endcapping and the use of ion-pairing reagents improve peak shape and resolution in reversed-phase chromatography of peptides. The interaction between two potential ion-pairing reagents and a peptide analyte at pH 2 or pH 10 are shown in (C) and (D), respectively. Ion-pairing reduces the exposure of charged groups of the analyte to the hydrophobic stationary phase. However, examination of the influence of the charge state of peptides on retention in high-pH RP-HPLC revealed that hydrophobicity determines retention and suggested that electrostatic interactions are of less importance (Delmotte, Lasaosa et al. 2007).

Because both the HPLC columns and the following MS instruments are highly sensitive and expensive, HPLC instruments are often equipped with two columns. In this setup, the first column is used to reduce contaminants, concentrate the analyte and protect the following separation-column. To this end, the loading of sample and the separation of the analytes are performed on separate flows. A valve with multiple ports regulates the loading and injection procedure; 6-, 8- or 10-port valves are commonly used. Trapping of peptides and simultaneously cleanup is performed on a preconcentration-, loading- or guard-column of shorter length and large inner diameter. This pre-column allows a high flow-rate (30 $\mu\text{L}/\text{min}$) and the use of trifluoroacetic acid (TFA) to control the pH and act as an ion-pairing reagent in the loading buffer and eluent. As an ion-pairing reagent, the amphiphilic TFA facilitates the trapping of peptides on the pre-column (Figure 1.3). Afterwards a valve switch ensures a seamless transition from loading to separation nano-flow, rerouting the analyte towards the analytical nano-column (Figure 1.4).

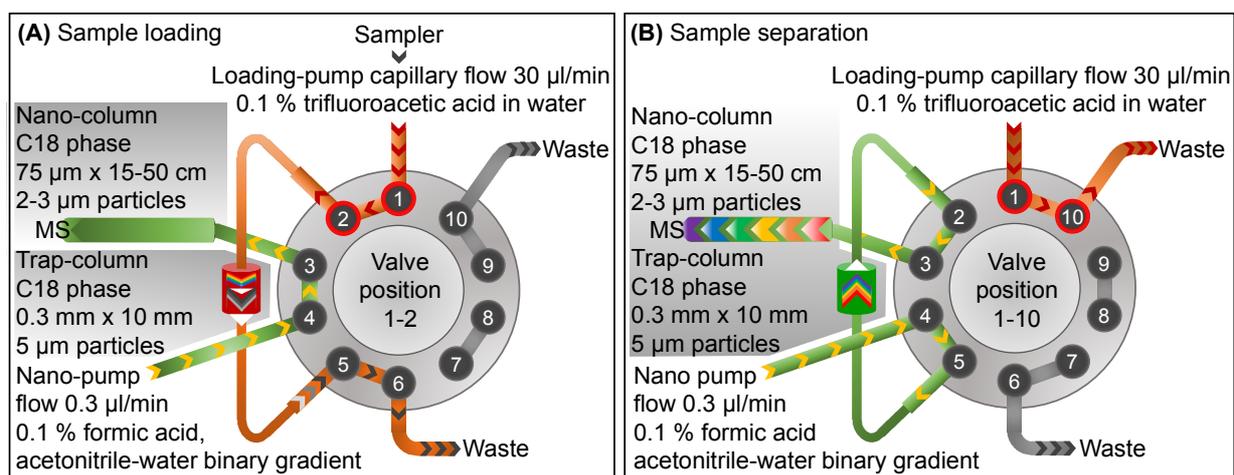


Figure 1.4: Schematic of a two-column setup on a 10-port valve. (A) Samples are loaded in backflush mode onto a pre-column of large internal diameter and particle size, allowing capillary flow-rate for desalting and trapping of peptides. **(B)** After switching the valve, peptides are transferred in the opposite direction from the pre-column to the analytical column. The flow from a binary gradient nano-pump system is used for separation.

On coupled LC-ESI MS platforms, the nano-flow eluents usually contain formic acid (FA) instead of TFA as a source of protons in the mobile phase. Although the FA is not suitable as an ion-pairing reagent, TFA is avoided because the acid can participate in the formation of gas-phase ion-pairs, which can neutralize positively charged analyte ions causing much stronger ion suppression in ESI (Kuhlmann, Apffel et al. 1995). Ion-pair reversed-phase separation (IP-RP) is often distinguished from RP-HPLC because with the former, in addition to solvophobic interactions between the peptides and the stationary phase, the addition of amphiphilic ions to the mobile phase results in a combination of electrostatic and solvophobic retention (Delmotte, Lasaosa et al. 2007).

1.5 Mass spectrometry

The principle of mass spectrometry is based upon the transfer of an analyte into the gas phase, its ionization and the acceleration of these ions by an electric field. Only with the development of soft ionization techniques, in particular electrospray ionization (ESI) (Yamashita and Fenn 1984) and matrix-assisted laser desorption/ionization (MALDI) (Hillenkamp, Karas et al. 1991), has the analysis of proteins and peptides by MS become possible (Figure 1.5). The intact gas phase transfer and ionization of these fragile biopolymers under atmospheric pressure, and the discovery of the characteristic fragmentation behavior upon collision with gas molecules have driven both the rapid development of increasingly sophisticated mass spectrometers and the emergence of proteomics as a dedicated research field. Since then, MS has become the method of choice to determine amino acid sequence information for the identification of proteins and peptides.

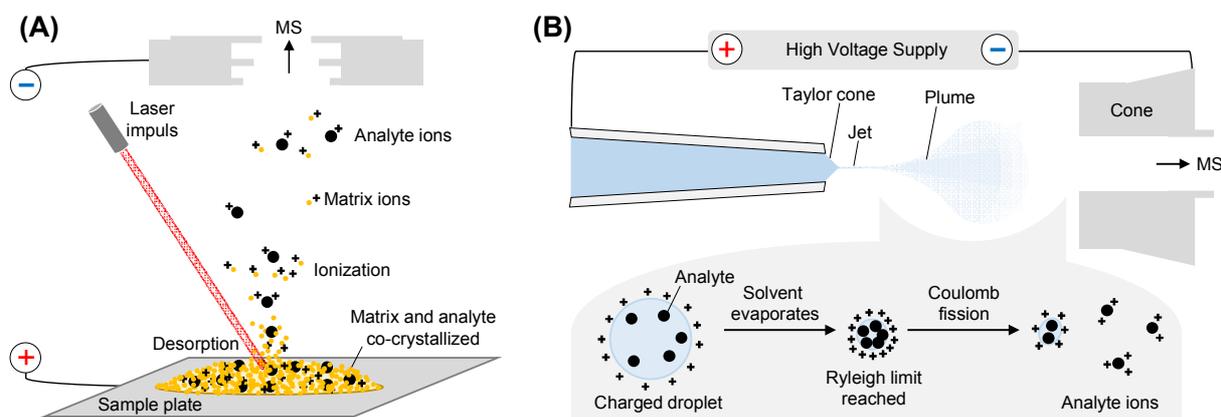


Figure 1.5: Soft ionization techniques. (A) The mechanism of matrix-assisted laser desorption/ionization (MALDI). The analyte is co-crystallized with a chromophore matrix. Pulsed laser irradiation causes excitation of matrix molecules followed by desorption and ionization of analyte and matrix molecules. (B) In electrospray ionization (ESI), the analyte is sprayed from a capillary needle. Solvent droplets containing analyte molecules, with a surface charge of the same polarity as the needle, evaporate until the Ryleigh limit is reached. The droplet undergoes Coulomb fission and dissociates to smaller droplets and charged analyte molecules. The figures are adapted from Lottspeich and Engels (2012).

The development of MS instruments has been focused on three key aspects: accuracy, resolution and speed. Today, peptide samples, separated by high-performance liquid chromatography, couple to MS can generate thousands of high-resolution fragment (MS/MS) spectra per hour.

In addition to the ion source, the main parts of a mass spectrometer include the mass analyzer and a detector. The two MS instruments used in this work are the LTQ Orbitrap Velos (Thermo Fisher Scientific) and the TOF/TOF 5800 system (ABSciex). In addition to the difference in ion source, where the LTQ Orbitrap Velos performs electrospray ionization and the TOF/TOF 5800

system relies on matrix assisted laser desorption ionization, the instruments also utilize different mass analyzers.

The TOF/TOF 5800 system relies on a time of flight (TOF) mass analyzer. TOF analyzers measure the drift time of ions accelerated down a flight tube by an electric field of known strength. The kinetic energy transferred to ions of equal charge state is the same; however, the velocity of the ions depends on their mass. Low mass particles reach higher speeds and impact earlier at the detector to the end of the flight tube than ions of higher mass with the same charge. With the measured drift time, the known flight distance and the specified acceleration voltage the mass to charge (m/z) ratio of analyte ions can be calculated. In addition, the TOF/TOF 5800 system has a collision cell where the ions are slowed by applying a 1-2 kV voltage below the source voltage to enable MS/MS fragmentation with collision gas.

The LTQ Orbitrap Velos is a hybrid mass spectrometer. It contains two different mass analyzers; an ion trap (Paul and Steinwedel 1953) and an Orbitrap Fourier transform (FT) mass analyzer (Makarov 2000). In general, ion traps isolate charged particles using dynamic electric fields. The LTQ Orbitrap Velos implements a linear quadrupole ion trap (LTQ), which consists of four parallel rod electrodes. Opposing electrical current are applied to each of the two opposite electrodes respectively and the polarity is switched with a high frequency. These antiparallel currents produce a radio frequency (RF) electric field, which radially confines ions. Ions can be moved along the linear axis of the LTQ by a current applied to end-electrodes. In addition, the RF field can be manipulated to stabilize and isolate ions of a specific m/z ratio to function as a selective mass filter or for the excitation of ions with a high frequency for collision induced dissociation (CID).

The Orbitrap mass analyzer operates by electrostatic axially orbital trapping of ions between an outer barrel-like electrode and a spindle-shaped inner electrode. The signal is recorded from the image current produced by ion packets which oscillate around the inner electrode of the trap (Makarov, Denisov et al. 2009). The m/z values of trapped analyte ions is calculated by Fourier transformation (FT). In contrast to the LTQ, the Orbitrap is not used as a fragmentation cell.

For CID fragmentation in the LTQ, resonant excitation selectively excites a particular m/z species and induces fragmentation due to collision with background gas in the trap. However, a lower mass cutoff limitation is observed in RF ion trap CID fragmentation due to instability of smaller ions (below approximately 28% of precursor ion m/z) in an RF field that can efficiently trap and activate the precursor (Jonscher and Yates 1997, Rauniyar and Yates 2014). This cutoff prevents the analysis of low mass product ions like mass tags for quantification, immonium ions and other small fragment ions (Griffiths and Unwin 2016).

For the acquisition of MS/MS spectra with the Orbitrap mass analyzer, the fragmentation occurs in a separate trap and the product ions are subsequently injected and monitored in the Orbitrap. In the early iterations of Orbitrap MS instruments, the fragmentation and accumulation of ions was performed in a C-trap and this type of ion activation was named higher energy C-trap dissociation (HCD). However, these early instruments had the same limitations as linear quadrupole ion traps, with a lower m/z cutoff depending on the precursor ion. In newer instruments, fragmentation is initiated in a dedicated octapole collision cell supplied with a radiofrequency voltage of which the direct current offset can be varied. This allows the isolation and transfer of low mass product ions (Olsen, Macek et al. 2007) and the name of the activation mode was changed to higher-energy collisional dissociation (HCD).

The major performance difference between the two mass analyzers is the resolution and the scan rate. While the accumulation, isolation, activation and detection of ions with the dual pressure LTQ system is very fast, the mass accuracy and resolution of acquired mass spectra is relatively low. Conversely, the acquisition of spectra with the Orbitrap mass analyzer is slower but features a higher mass resolution. The combination of both mass analyzers and two collision cells in a single MS instrument provide highly versatile peptide fragmentation modes. In particular, the simultaneous acquisition of CID spectra in the ion trap (CID-IT) with high frequency and sensitivity in the LTQ and high-resolution HCD-FT spectra in the Orbitrap mass analyzer facilitates a comprehensive mass analysis of complex peptide mixtures by online LC-MS runs.

1.6 Peptide identification

With the capability of high throughput sample analysis, automated data processing has become essential. Here, the characterization of the fragmentation behavior of peptides, which allows the manual interpretation of spectra, also facilitated the development of bioinformatic tools for automated assignment of spectra (Figure 1.6).

Although the specific fragmentation behavior permits the *de-novo* sequencing of peptides, experimental data are generally matched against theoretical spectra generated on the basis of sequence databases by search engines such as Mascot (Perkins, Pappin et al. 1999) or Sequest (Yates, Eng et al. 1996). With completely mapped and annotation genomes, information about protein coding sequences and the translation into protein primary structure became available as the fundamental requirement for this process.

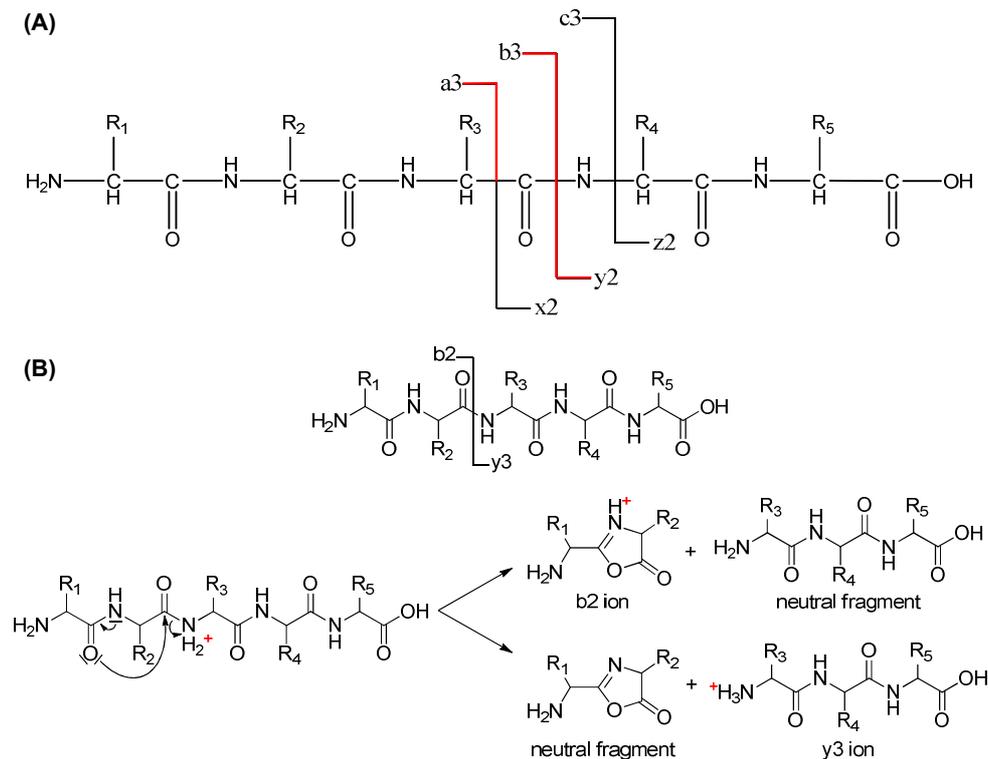


Figure 1.6: Peptide fragmentation in tandem mass spectrometry. (A) The nomenclature for peptide fragments according to Roepstorff and Fohlman (1984) and Biemann (1988) is illustrated. The most common fragment ion series observed upon CID and HCD are a-, b- and y-ions. (B) The mechanism of b- and y-ion formation according to the mobile proton model; b-ions forms a stable protonated oxazolone.

To match spectral data, a sequence database is digested *in-silico* according to the cleavage specificity of a defined protease used in the experiment. This provides the theoretical m/z values for comparison with spectral data. Furthermore, the sequence information together with the characteristic fragmentation behavior, determine the corresponding characteristic fragment ions for comparison with MS/MS spectra. To validate the resulting peptide spectrum matches, search algorithms generally calculate a theoretical false discovery rate (FDR) by matching the same dataset of MS/MS spectra against a randomized or reversed database using the same matching and scoring criteria.

As described above, most proteome analyses focused on providing high coverage are performed at the level of peptides instead of intact proteins. Considering this, the proteome samples, with thousands of proteins theoretically possible for a given organism, is again multiplied in complexity by the number of peptides produced during proteolytic digestion. This collective of possible peptides constitutes the search space used to match experimental data. For the human proteome, the mean number of peptides produced per protein upon tryptic digestion, amount to about 150 (allowing for two missed cleavages and masses from 500 Da to 4000 Da). Additionally, any

variable modifications allowed in the database search increase the search space. Conversely, the most important experimental parameter to reduce the degrees of freedom in a database search is the mass accuracy of the used MS instrument. The search space for MS/MS spectra can be restricted to only those peptide matches within the expected mass tolerance of the observed precursor ion signal. Spectra acquired using a mass analyzer with a high resolution and mass accuracy, like the Orbitrap, can be searched with a mass tolerance below 10 parts per million (ppm) of the experimental precursor m/z .

1.7 Quantitative proteomics

Since the discovery of gene expression as the central dogma of cell biology, the processing of genetic information by transcription to RNA and translation into proteins has become a central focus in life sciences. The two most commonly used methods to assess changes in gene expression upon environmental change, perturbation or experimental manipulation of a biological system are the system wide determination of mRNA or protein levels.

Although such analyses seldom measure the true activity of the transcription or translation process, and also do not take into account the turnover of the produced biomolecules (rate of mRNA or protein synthesis and degradation), changes in the abundance of either mRNA or proteins are frequently equated with an induction or decrease in gene expression. Even with this limitation, these quantitative datasets are valuable resources for systems biology.

In general, quantitative proteomics can be classified into absolute and relative quantitation of proteins. For absolute protein quantification, the concentration of the protein is determined and calculated in relation to a parameter of the analyzed sample. For example, this procedure can be used to calculate the copy number or concentration of the target protein per cell or the fractional amount relative to the total protein amount in the sample.

In contrast, relative quantitative proteomics compares the protein abundance between two or more samples. This approach is useful to determine the difference in protein level between sick and healthy subjects, wild type and mutants or to analyze the effect of different environmental influences. Absolute quantitative analyses usually concentrate on a few specific target proteins for accurate quantification. However, it is also possible to determine less accurate system-wide estimates of absolute protein abundance. Such absolute protein abundance profiles can be used to assess the resource allocation in proteomes of different cell types or even organisms. Both large-scale data sets of absolute and relative protein abundances are also very valuable for

systems biology and biotechnological applications. Together with cell-molecular data from other experiments, mathematical models can be constructed to explain or predict cellular and metabolic processes, simulate the effect of environmental changes and explore the capabilities and limitations of an organism.

Analogous to the qualitative workflows for protein detection and identification, quantitative proteomics techniques can be divided into gel-based approaches and chromatography-based methods. Both relative and absolute quantification methods on 2D-gels have been developed. These methods usually quantify based on image processing of dye stained protein spots of the separated proteins and subsequent identification of excised spots (Maass, Sievers et al. 2011).

The experimental focus of this thesis was on the LC-based and peptide level approaches to proteome analysis and accordingly the MS-based quantifications methods. These approaches rely directly on the spectral data acquired by MS for peptide and protein quantification. Most MS-based quantification strategies utilize the incorporation of labels, in particular stable isotopes (e.g. ^2H , ^{13}C , ^{15}N or ^{18}O). Although differentially labeled analytes can be mixed and show the same retention time during LC-separation, additional signals with a defined mass-shift appear in MS or MS/MS spectra. The comparison of signals originating from labeled and unlabeled samples or the signal from a sample and a labeled internal standard yield quantitative data.

In addition, there are also methods for label free quantification. These approaches rely solely on MS spectral intensity parameters for quantification. These methods have become increasingly accurate due to progress in bioinformatics, in particular by improvements in the alignment of LC-MS features and more sophisticated algorithms to calculate both relative and absolute quantitative data from spectral counts and signal intensities (Schwanhausser, Busse et al. 2011, Cox, Hein et al. 2014).

For the quantification depending on stable isotope labels, numerous strategies have been developed, where the label is incorporated into the sample at different steps of the classical bottom-up proteomics workflow. The incorporation of isotopes during cell cultivation, by supply of isotope labeled nutrients, is called metabolic labeling. Organisms cultivated for several generations with labeled nutrients metabolize the supplied nutrients, leading to the natural incorporation of the isotopes or isotope labeled precursors into the complete proteome during translation. Labeled and unlabeled organisms can be combined before sample preparation and peptides originating from either of the samples can be assigned by the mass-shift between signals observed in MS spectra.

Metabolic labeling refers to the supply of isotope labeled nutrients which are then metabolized by cells or organisms and incorporated during protein synthesis. Two prominent examples of this approach are stable isotope labeled amino acids in cell culture (SILAC) and $^{14}\text{N}/^{15}\text{N}$ labeling. The most widely used method of metabolic labeling is SILAC (Ong, Blagoev et al. 2002). In a SILAC experiment only selected, essential amino acids are labeled, typically ^{13}C -Arginine, ^{15}N -Lysine or ^2H -Lysine. Using these amino acids, tryptic peptides contain at least one labeled amino acid. Today a broad spectrum of model organisms, including bacteria, yeast, numerous cell lines, *C. elegans*, *Drosophila* and even mice have been labeled in this fashion. Metabolic labeling techniques are limited to samples grown under laboratory conditions; and allow only the relative quantification of two samples in a single experiment. However, a great advantage is the introduction of the label early in workflow, which reduces technical variability. The cultivated samples can be mixed before cell disruption and further processing.

In contrast, chemical labeling refers to the modification of specific functional groups of the intact protein or proteolytic peptides. The proteolytic processing in bottom-up proteomic workflows can also be utilized for enzymatic labeling. This approach refers to the enzymatic incorporation of ^{18}O isotopes into peptides during digestions in ^{18}O -water (Schnolzer, Jedrzejewski et al. 1996, Yao, Freas et al. 2001).

Chemical labels can be attached at the level of purified proteins or after proteolysis at the peptide level. Because the incorporation takes place later in the workflow, the method is not limited to cultivated biological material. This facilitates the analysis of samples like primary cell lines and clinical samples, which are not accessible to metabolic labeling. Chemical labeling strategies rely on reagents that attack specific functional groups. Previously derivatized reactive sites and functional groups include the cysteine sulfhydryl-groups, carboxylic-groups of aspartic acid and glutamic acid as well as the C-terminus, the indole ring of tryptophan and the tertiary amine functions of lysine residues and at the N-terminus. Key features of chemical labeling are the efficiency and specificity of the modification reaction. Ideally, the derivatization should be complete and without side reactions (Somasundaram, Koudelka et al. 2016). In general, the labeling of peptides is more efficient than the derivatization of intact proteins. In addition, the number of reactive groups on the analyte can be an issue because labeling less frequent amino acids can limit the number of peptides targeted for quantification (Gregorius, Jakoby et al. 2013).

The derivatization of primary amine functions by reductive dimethylation or by acylation with active esters has proven highly efficient and reasonably specific. Advantages of this approach include the frequency and activity of the nucleophilic amine group and the ideal compatibility with tryptic

peptides, which often feature a C-terminal Lysine residue in addition to the active site at the N-terminus. For these reasons, primary amines have become the preferred site to introduce labels for relative and absolute quantification.

Arguably, the most elegant approach to label based quantification of peptides and proteins is isobaric labeling. The two most common tags for isobaric labeling are commercialized under the acronyms iTRAQ (isobaric Tags for Relative and Absolute Quantification; ABSciex) and TMT (Tandem Mass Tag; Thermo Fisher Scientific). Figure 1.7 shows the structure and working mechanism of TMT 6-plex.

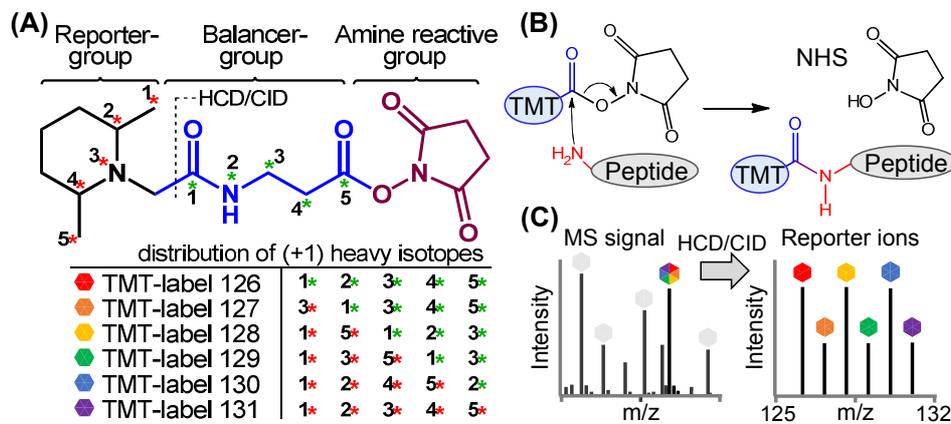


Figure 1.7: TMT 6-plex labeling. (A) The structure of TMT 6-plex reagents. The reagents are composed of a reporter group, a balancer and a peptide reactive group. Reporter- and balancer-group build the isobaric mass tag. By the incorporation of specific combinations of ^{13}C and ^{15}N isotopes in the reporter- and balancer-groups, the six tags have the same mass. The amine reactive group is an N-hydroxysuccinimide (NHS) ester. (B) The nucleophilic amine attacks the active ester; NHS leaves the reaction under formation of an amide bond between the peptide and the isobaric tag. (C) Labeled peptides coelute from LC and produce a single signal in MS spectra; fragmentation results in reporter ions suitable for relative quantification of the different labelled peptides by intensity.

Isobaric tags are very convenient because they allow the analysis of several samples in a single analysis so several conditions and replicates can be processed together. iTRAQ 4-plex reagents accommodates four channels (Chapter 2) and the TMT 6-plex kit can be used to label and co-analyze six samples in a single experiment (Chapter 3). In contrast to other differential labels, where the signature mass tags produce distinct signals for the labeled peptide in MS spectra, peptides labeled with isobaric tags produce a single signal in MS spectra. Labeled peptide precursor ions are then isolated and fragment spectra show a series of reporter ions that provide the relative quantitative data for the different channels.

2. Quantitative proteome analysis of *Caenorhabditis elegans* upon exposure to nematicidal *Bacillus thuringiensis*

2.1 Abstract

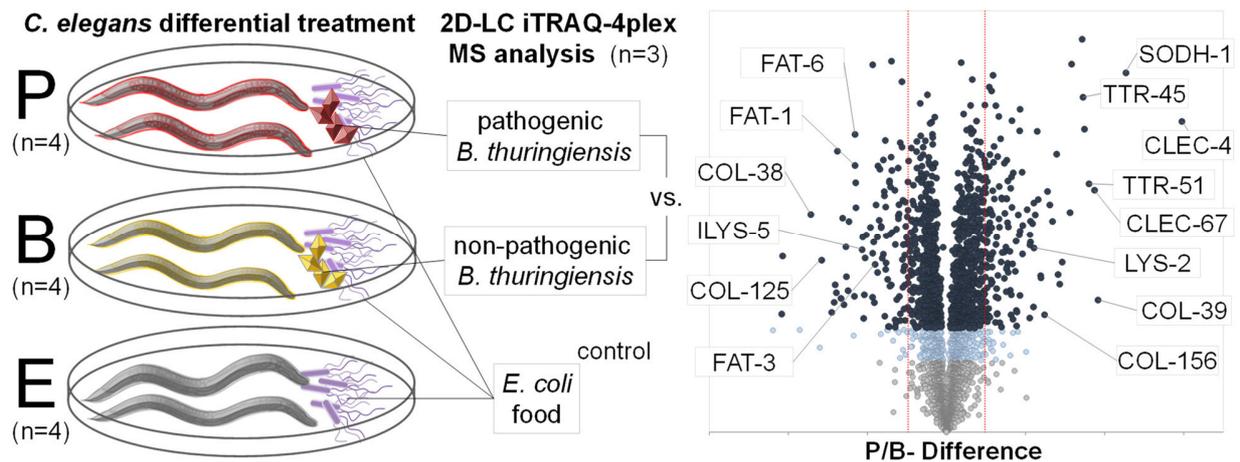


Figure 2.1: Graphical abstract. *Caenorhabditis elegans* were grown on petri dishes to compare three cultivation conditions. Differential treatments included a xenic culture of the food source *Escherichia coli* (E) and composite cultures of the food bacterium with spore crystal mixtures of either a pathogenic (P) or a non-pathogenic (B) strain of *Bacillus thuringiensis*. Cultivations were performed in four independent biological replicates. The collected samples were analyzed in three bottom-up proteome analyses using two-dimensional liquid chromatography for peptide separation and iTRAQ 4-plex labeling for relative quantification to compare the protein levels in nematodes grown under pathogenic and non-pathogenic conditions (P/B).

Caenorhabditis elegans can be infected by a plethora of pathogens, some of which are also pathogenic for humans. Consequently, the nematode has emerged as a powerful surrogate host to model microbial human infectious diseases in a non-vertebrate, for the study of innate immunity and host-pathogen interactions. The signaling cascades *C. elegans* employs in response to bacterial or fungal pathogens are well investigated. In this study, the downstream processes of these cascades, i.e. the differential expression of effector and regulatory molecules due to a microbial challenge with the pathogenic bacterium *Bacillus thuringiensis* (*Bt*) in comparison to a non-pathogenic *Bt* strain were analyzed. The protein abundance profile of the nematode was studied by quantitative proteomics using iTRAQ labeling and 2D-LC-MS analysis. To this end, a novel method for the preparation of defined *C. elegans* samples from host-pathogen experiments was developed. The collected samples were combined using an isobaric labeling scheme to

reduce the number of necessary LC-MS experiments and a pooling strategy for fractions in 2D-LC separation schemes was employed to optimize measurement time and data output.

More than 3,600 proteins were quantified, 288 of which showed altered abundances. These results implicate protein classes such as lectins, lysozymes, and transthyretin-like proteins in the nematode innate immune defense. A number of gene products previously identified by transcriptomic profiling could be verified at the protein level. Moreover, several other protein classes such as proteases, proteins related to autophagy and apoptosis, structural proteins, and proteins involved in chromatin organization were identified at differential abundance and may account for the damage caused by *Bt* infection.

The results of this study provide an overview of the physiological response towards a pathogen at protein level in the important model organism *C. elegans*, giving insights into the complexity of host-pathogen interactions.

2.2 Introduction

Understanding the molecular mechanisms involved in the interaction of higher organisms with their microbiota is an important key for understanding host-parasite coevolution (Schulte, Makus et al. 2013) and for the identification of processes related to both the maintenance of homeostasis and the development of diseases. The nematode *Caenorhabditis elegans* is an important model organism that has been extensively used for the study of host-pathogen interactions and the elucidation of host innate immunity (Marsh and May 2012, Pukkila-Worley, Feinbaum et al. 2012). Besides general aspects making it a suitable model organism, e.g. its short generation time, a well curated genomic repository (Harris, Antoshechkin et al. 2010) and a wide repertoire of methods for manipulation and assessment of genetic information, a major benefit provided by this organism is the ability to generate sterile and synchronous cultures by hypochlorite treatment (Rehman Khan and McFadden 1980). Such precise control over the presence of microbes enabled the development of a number of robust and simple infection assays which facilitate the unbiased investigation of interactions between *C. elegans* and a range of microbial pathogens (Couillault and Ewbank 2002, Mylonakis, Ausubel et al. 2002, Couillault, Pujol et al. 2004).

In its natural environment, *C. elegans* interacts with a diverse range of microorganisms, including bacteria, fungi, and protists, that can serve as food sources or act as pathogens. Among these microorganisms, the gram-positive soil bacterium *Bacillus thuringiensis* (*Bt*) coexists with *C. elegans* in its natural habitat (Borgonie, Van Driessche et al. 1995). *Bt* strains can produce several classes of proteinogenic crystal (Cry) toxins that are highly toxic to specific susceptible insect larvae, while showing negligible biological or environmental toxicities (Betz, Hammond et al. 2000). In this study a *Bt* isolate (B-18247) showing nematicidal activity in both its spore crystal mixtures and purified endotoxin crystals (Edwards, Payne et al. 1989) was utilized to challenge the nematode. The interactions of *Bt* and Cry toxins with several different host organisms, e.g. insects and nematodes including *C. elegans*, have been studied previously (van Frankenhuyzen 2009, Schulte, Makus et al. 2010). However, many aspects of these interactions are still under investigation, not least due to growing concerns about the emergence of *Bt*-toxin resistant strains (Griffitts, Whitacre et al. 2001).

Several classes of proteins were identified as immune-effectors involved in the interaction of *C. elegans* with microorganisms. These include lysozymes (Schulenburg and Boehnisch 2008), neuropeptide-like proteins (McVeigh, Alexander-Bowman et al. 2008, Pujol, Davis et al. 2012), caenacins (Zugasti and Ewbank 2009), thaumatins (Shatters, Boykin et al. 2006), lipases (O'Rourke, Baban et al. 2006), saposin-like proteins (caenopores) (Roeder, Stanisak et al. 2010, Hoeckendorf and Leippe 2012, Hoeckendorf, Stanisak et al. 2012), medritin-like Shk toxins

(Pukkila-Worley, Feinbaum et al. 2012), proteins containing CUB-like domains (Shapira, Hamlin et al. 2006, Nandakumar and Tan 2008), C-type lectins and collagens (Schulenburg, Hoepfner et al. 2008, Simonsen, Gallego et al. 2012). By bioinformatic annotation of the *C. elegans* coding sequences, more than 750 genes can be assigned to one of these protein families (Hunter, Jones et al. 2012). This shows the diversity of potential effectors involved in the innate immune response of the nematode.

Aside from quantitative changes of these particular protein families, it can be expected that the host organism will undergo global changes while interacting with microorganisms. The majority of system wide studies of host-pathogen interactions in *C. elegans* have analyzed gene expression at the transcript level (Simonsen, Gallego et al. 2012). However, the study of the host proteome can deliver complementary data and provide deeper insights into the underlying systemic mechanisms at the molecular and hence the functional level (Schrimpf and Hengartner 2010). Proteomic analyses of *C. elegans* host-microbe interactions are limited in number, however, they represent an important foundation towards the understanding of far more complex host-microbiome interactions (Cassidy and Tholey 2014).

The interactions of *C. elegans* with the gram-positive bacterium *Staphylococcus aureus* (Bogaerts, Beets et al. 2010) and the gram-negative bacterium *Aeromonas hydrophila* (Bogaerts, Temmerman et al. 2010) have been studied by differential two-dimensional gel-electrophoresis based proteomics (DIGE). In these two studies, 108 and 65 proteins, respectively, were identified as differentially abundant following infection. In both cases, the effect of the pathogenic bacteria and the non-pathogenic *E. coli* OP50 on the host was studied independently. In another study, the interaction of *C. elegans* with the pathogenic *E. coli* strain LF82 in comparison to the *E. coli* standard laboratory food strain OP50 was investigated using ¹⁴N/¹⁵N-metabolic labeling followed by first dimension gel-electrophoresis separation and second dimension LC-MS analysis (GeLC-MS) (Simonsen, Moller-Jensen et al. 2011). Between the two growth conditions, 117 proteins were identified at differential abundance.

In the present study, the protein abundance profile of *C. elegans* exposed to a pathogenic in comparison to a non-pathogenic *Bt* strain was investigated. In both treatments, *E. coli* OP50 was present as a food source to prevent starvation in order to keep experimental conditions as close as possible to the situation in natural environments. For the same reason, the reaction of the host to the presence of the whole bacterium was analyzed, rather the influence of isolated crystal toxins, which has been studied earlier (Griffitts, Whitacre et al. 2001, Wei, Hale et al. 2003, Huffman, Abrami et al. 2004, Kho, Bellier et al. 2011). To differentiate changes to the *C. elegans* proteome

induced by the challenge with *Bt*, a xenic culture maintained on *E. coli* OP50 was also analyzed as a negative control.

To ensure the consistency of relative quantitative proteomics studies, the compared samples have to be strictly controlled to differ only in the condition which is of interest. The culture of the biological material is the first crucial step in this regard. Especially for an organism with a diversity of developmental stages like *C. elegans* this is not a trivial matter as four larval stages as well as the adults, gravid hermaphrodites, male worms and dauer larvae can be differentiated in mixed cultures. Consequently, the proportions of these developmental stages in cultures using different worm strains or culture conditions can diverge significantly. For this reason homogenous samples of different strains or under different growth conditions for the quantitative analysis of *C. elegans* on a molecular level have to be focused on one specific developmental stage of the worm. The developmental stage of interest has to be enriched by synchronization or stage specific sorting. To this end, a method for the preparation of defined *C. elegans* samples was developed in this study. In addition to improvement of the developmental stage resolution, the newly developed nylon-net filter sorting device simplified the removal bacterial contaminants and worm harvest.

Along with the identification of global proteome changes, a major goal of the presented work was to establish an analytical workflow enabling a straightforward quantitative analysis with minimized requirements in respect to overall measurement times. Multidimensional chromatography coupled online or offline to ESI- or MALDI-MS in combination with isobaric labeling, e.g. iTRAQ (Ross, Huang et al. 2004), provides a number of advantages for quantitative proteomics. A major benefit is the multiplexing capacity of isobaric labeling, which allows comparison of several biological samples in a single experiment. However, depending on the number of biological replicates and samples to be analyzed, the overall measurement time is still a serious issue. Another critical parameter is the chromatographic separation scheme applied; collecting many fractions in the first dimension separation results in significantly increased instrument times for the second dimension separation and MS analysis. One of the innovations of this study includes the application of an isobaric labeling scheme, permitting the combination of different biological and technical replicates with a low number of LC-MS experiments. Additionally, a suitable pooling strategy for fractions of first dimension LC separation in a powerful two dimensional separation scheme employing gradients at high/low-pH was applied.

The results of this study demonstrate that the developed methodological approach allows a comprehensive analysis of the change in proteome composition in *C. elegans* upon challenge with microbes.

2.3 Material and methods

Cultivation and microscopy of *C. elegans* were performed in cooperation with Dr. Aylin Höckendorf, (Zoological Institute, Zoophysiology, Christian-Albrechts-Universität zu Kiel; chair: Prof. Dr. Matthias Leippe). *Bacillus thuringiensis* preparations were provided by the group of Prof. Dr. Hinrich Schulenburg (Zoological Institute, Evolutionary Ecology Genetics, Christian-Albrechts-Universität zu Kiel)

2.3.1 Materials

Deionized water (18.2 MΩ/cm), used throughout this study was prepared with an Arium611 VF system from Sartorius (<https://www.sartorius.de>). Nylon net filters were from Millipore (<http://www.merckmillipore.com>), LoBind reaction tubes were from Eppendorf (<https://www.eppendorf.com>) and the BCA assay was from Thermo Fisher Scientific (<https://www.thermofisher.com>). RapiGest SF was from Waters (<http://www.waters.com>) and cOMplete protease inhibitor cocktail was from Roche Diagnostics (<http://www.roche.de>). Sequencing grade modified trypsin was purchased from Promega (<https://www.promega.de>) and the iTRAQ 4-plex reagent kit from ABSciex (<https://sciex.com>). All other chemicals were acquired from Sigma-Aldrich (<https://www.sigmaaldrich.com>).

2.3.2 *C. elegans* cultivation and sample preparation

Standard worm maintenance was performed as described earlier (Brenner 1974). To validate the stage specific sorting procedure by nylon net filter for large scale proteomics sample preparation, *C. elegans* N2 Bristol and two variant strains *adm-4* and *sup-17ts* were used. The quantitative proteome analysis was performed with *C. elegans* N2 Bristol.

2.3.2.1 Synchronization by hypochlorite treatment

Worms were washed off 5 petri dishes (Ø 9 cm) with 3 mL M9 buffer and pipetted to a conical 15 mL tube. After worms settled at the bottom (10 min) the buffer was reduced to 3.5 mL, bleaching solution was added (1 mL; 2.5 M NaOH, 6.5 % NaOCl) and the tube was gently mixed. Worms were treated with hypochlorite for 5 min. The tube was centrifuged (4,000 g, 3 min, 20 °C), the pellet containing living eggs was washed with 15 mL M9 buffer, centrifuged again (4,000 g, 3 min, 20 °C) and the buffer was replaced by 10 mL of fresh M9 buffer. Eggs were hatched in M9 buffer overnight (16 °C) without food and an average of 1000 worms were transferred to petri dishes seeded with 1 mL of an overnight culture of *E. coli* OP-50.

2.3.2.2 Saccharose floating

Bacteria, dead worms and other debris can be removed by sucrose floating. Worms were harvested by washing of five petri dishes with 3 mL M9 buffer and collected in a conical 15 mL tube. After settling for 10 min on ice the buffer was reduced to 5 mL and an equal amount of ice cooled 70 % sucrose in M9 buffer was added. The solution was mixed thoroughly and centrifuged (4000 g; 10 min; 4 °C) to float the worms. The worms were collected from the top of the solution, washed with 10 mL of M9 buffer and pelleted by centrifugation (4,000 g; 5 min; 4 °C). Washing was repeated with water and worms were transferred to 2 mL reaction tubes.

2.3.2.3 Worms sifting using nylon net filters

A worm sifting device was built using nylon net filters (\varnothing 4.7 mm) with mesh sizes of 10 μ m, 20 μ m, 30 μ m and 40 μ m, depending on the sizes of worms to be filtered. Filters were fixed between two GUKO natural rubber funnels and racks were mounted on a tube with an outlet valve. The column was belted by a metal brace. To facilitate the sorting of the worms and increase the flow rate of the buffer 7 small imbalance motors with a 3 V power supply were fastened to the brace (see Figure 2.2).

Worms were washed of 15 petri dishes with 25 mL M9 buffer and the solution was collected in a sterile glass beaker. The worm solution was sifted through nylon net filters in the aforementioned device and washed with an additional 200 - 300 mL of fresh M9 buffer. In the end 100 mL of water were used to wash off buffer, the filtered worms were recovered from the nylon nets with 2 mL water and transferred to 2 mL reaction tubes. Worms were pelleted by centrifugation (4000 g, 5 min, 20 °C), water was reduced to cover pellet and worms were directly used in subsequent analyses or shock frozen in liquid nitrogen until further use.

2.3.2.4 Validation of sample preparation procedures by microscopic survey

Worms were collected in M9 buffer and anesthetized in 0.01 % sodium azide. About 40 worms were transferred per glass slides and DIC microscopic images (50 x magnification) were recorded with a Zeiss Axio imager Z1 (<https://www.zeiss.de>). The images were processed and analyzed using the Java based software ImageJ (<https://imagej.nih.gov>). The Java programming language was used to write a script for fully automated image processing (Appendix Script A 2.1 and A 2.2). To measure worm length, width and area, the “Measure ROI Curve” plugin by Bob Dougherty (<http://www.optinav.com/Measure-Roi.htm>) was used. The plugin allows the curve length of a

region of interest (ROI) to be calculated as the average centerline between the outline arc curves connecting the points of largest separation of an indicated particle. The width of a centerline point is calculated as the straight line distance between opposite edge points with an equal average arc length as the length of the center line point. The resulting minimum distance is taken as the width corresponding to that centerline point, and the largest width over all of the centerline points is reported as the width of the object (<http://www.optinav.com/Measure-Roi.htm>). The distribution of worm sizes in samples after synchronization by bleaching, sifting through nylon net filters or applying both procedures in sequence was determined by surveying 500 – 700 worms.

2.3.2.5 Cultivation of *C. elegans* and incubation with *Bacillus thuringiensis*

For proteome analysis, synchronized L1 larvae of the *C. elegans* laboratory strain N2 Bristol were grown at 16 °C for 48 h on nematode growth medium (NGM) agarose petri dishes, seeded with *E. coli* strain OP-50 until larval stage L3 - L4 and collected from plates with M9 buffer. The worms were harvested, washed with M9 buffer and recovered with the worm sifting device (Figure 2.2) enabling stage specific sorting on Millipore nylon net filters. L3 - L4 larval stage worms were recovered between two nylon net filters of mesh sizes 11 µm and 30 µm.

Recovered worms were transferred to peptone-free medium (PFM) agarose plates. On 9 cm petri dish plates, about 1500 (\pm 300) worms were exposed to the different treatments. One sample for each replicate (and condition) was collected from 12 to 15 petri dishes, each seeded with 1 mL of three different mixtures of bacteria in M9 buffer. The cultures for differential treatment were grown on a bacterial lawn of 6×10^9 *E. coli* cells per petri dish either as a monoculture or with the addition of 2×10^8 spores of the pathogenic *Bt* strain B-18247 or the non-pathogenic *Bt* strain DSM350. These three conditions are described as either **E** (*E. coli*), **P** (pathogenic *Bt*), or **B** (non-pathogenic *Bt*), throughout the remainder of the text. The nematicidal *B. thuringiensis* strain PS69D1 NRRL B-18247 was used as a pathogen (Edwards, Payne et al. 1989). The strain was provided by the Agricultural Research Service Patent Culture Collection (United States Department of Agriculture). The host-control treatment included the non-pathogenic *Bt* strain DSM-350 (German Collection of Microorganisms and Cell Cultures). *Bt* spore and crystal toxin mixtures were prepared as described elsewhere (Borgonie, Van Driessche et al. 1995, Schulte, Makus et al. 2010).

Worms were exposed to each condition at 20 °C for 12 h, before harvesting L4 larvae between nylon net filters with mesh-sizes of 20 µm and 30 µm in the sifting device. Worms were washed on filter with 100 mL of M9 buffer followed by 100 mL of water to remove contaminants before being recovered from the filters with 2 mL of water. Nematodes were sedimented by centrifugation (1,000 *g*, 1 min, 20 °C) and the supernatant was discarded leaving approximately 200 µL of

residual water. The samples were then snap-frozen in liquid nitrogen and stored at -80 °C until further use. Each treatment condition was analyzed in four independent biological replicates. The survival of 50 (\pm 5) L3 - L4 worms was monitored on petri dishes (\varnothing 3.5 cm) seeded with 200 μ L of the three bacterial suspensions for 14 h in five replicates.

2.3.3 Protein extraction, digestion and isobaric labeling

Proteins were extracted in 400 μ L of extraction buffer (100 mM KCl, 100 mM triethylammonium bicarbonate (TEAB), 2 mM tris(2-carboxyethyl) phosphine (TCEP), 2 mM ethylenediamine tetraacetic acid (EDTA), 0.5 % 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS), and 2 x cOmplete protease inhibitor (Sigma-Aldrich). Acid-washed glass beads (\varnothing 0.1 - 0.25 mm; Sigma-Aldrich) were added (approximately 200 μ L) and cells were disrupted in ten cycles of freezing (90 s, -80 °C ethanol bath) and partial thawing in a sonication bath (30 s), followed by disruption in a mixer mill for 60 s at 30 Hz (MM 400; Retsch; <http://www.retsch.de>). The cell debris and glass beads were removed by centrifugation (1,000 g, 10 min, 4 °C), the protein concentration of the supernatant was determined using the bicinchoninic acid (BCA) protein assay and aliquots of 120 μ g total protein were cleaned by methanol-chloroform precipitation. The precipitate was dissolved in 20 μ L of 0.4 % RapiGest in 100 mM TEAB-buffer, heated at 99 °C for 5 min, and reduced and alkylated as described in the vendor's iTRAQ protocol. RapiGest was diluted to 0.2 % using 100 mM TEAB to give a final volume of 40 μ L before digestion was carried out with sequencing-grade modified trypsin in a protease to protein ratio of 1:40. Subsequently, the samples were labeled with the iTRAQ 4-plex reagents following the manufacturer's protocol. The volume of the samples was increased with TEAB buffer to 150 μ L, and 70 μ L aliquots of each sample, labeled with one of the four respective iTRAQ reagents, were combined and incubated with 0.2 % formic acid (FA) to remove the acid-cleavable detergent. The samples were centrifuged (55,000 g, 4 °C, 1 h), the supernatant was transferred to a fresh 1.5 mL LoBind reaction tube, lyophilized to dryness and redissolved in 60 μ L of eluent A for HPLC separation at pH 10.

2.3.4 Two-dimensional liquid chromatography coupled to mass spectrometry

In the first dimension, an Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific) equipped with an analytical Gemini-C₁₈ column (250 mm x 3 mm, 3 μ m; Phenomenex; <http://www.phenomenex.com>) was used to separate ca. 200 μ g of iTRAQ labeled peptides at a flow rate of 300 μ L/min. Eluent A: 72 mM triethylamine in deionized water, adjusted to pH 10 with acetic acid; eluent B: ACN, 72 mM triethylamine, 0.35 % v/v acetic acid. After injection, the sample

was washed with 2 % eluent B for 10 min and then fractionated over a linear gradient from 2 % B to 55 % B in 50 min collecting 50 fractions of 300 μ L each. Separation was monitored with a UV detector at 280 nm. The fractions were lyophilized to dryness and redissolved in 30 μ L of 0.1 % aqueous TFA. The fifty fractions obtained were pooled to ten fractions according to a previously described fractionation scheme (Stephanowitz, Lange et al. 2012) and used for second dimension separation.

Second dimension separation was performed on an Ultimate 3000 HPLC system, equipped with an Dionex Acclaim PepMap100 nano-column (75 μ m x 15 cm, 3 μ m, 100 Å), coupled online to a LTQ Orbitrap Velos MS instrument (Thermo Fisher Scientific) via a nanospray ion source. Each of the ten pooled fractions were injected independently (6 μ L). The labeled peptides were washed for 10 min on a PepMap C₁₈ guard column (300 μ m x 10 mm; Thermo Fisher Scientific) with 0.1 % aqueous TFA at a flow rate of 30 μ L/min. Reversed-phase (RP) HPLC separation was performed using a 180 min linear gradient from 95 % eluent A (water with 0.1 % FA) to 60 % eluent B (80 % ACN, 0.1 % FA) at a flow rate of 250 nL/min, followed by a sharp increase to 95 % eluent B for 10 min and an isocratic 5 min washing step. Afterwards, the column was equilibrated with 5 % B for 15 min. UV detection was performed at 214 nm.

MS conditions were: 1.3 kV spray voltage and 197 °C capillary temperature with a 30 μ m PicoTip emitter (New Objective; www.newobjective.com). After a delay time of 35 min, full scans of 300 – 2,000 m/z range were recorded in profile mode with a resolution of 30,000 (AGC target 1E6; maximum inject time 500 ms, preview mode for FTMS master scans and Wideband activation were used). The top 5 precursors (minimum signal intensity 500 and rejecting charge state 1) were selected for both CID fragmentation and HCD fragmentation with a repeat count of 2, a repeat duration of 20 s and subsequent dynamic exclusion of the selected m/z values (\pm 5 ppm) for 60 s (Xcalibur software v2.1.0.1140). The CID isolation window was set to 3 Da, the AGC target was 1E4 with a maximum inject time of 400 ms, activation time of 10 ms, and an Activation Q of 0.25. Normalized collision energy (NCE) for CID fragmentation was set to 36 %. The isolation window for HCD precursors was set to 3 Da, AGC target was 1E5 with a maximum inject time of 500 ms and activation time was 0.1 ms. NCE of 45 % was used for fragmentation and HCD spectra were acquired with a resolution of 15,000. All MS/MS spectra were recorded in centroid mode.

Each pooled fraction was analyzed at least three times by LC-MS. The first two runs, true technical replicates, were processed and after database search (see below), used to compile an exclusion list of m/z values representing peptides identified with high confidence (false discovery rate (FDR) \leq 0.01). These m/z values were excluded from data dependent MS/MS acquisition over a retention time window of 3 min and in a mass range of 0.5 Da for the third measurement of each sample.

2.3.5 Data processing

The complete dataset consisted of 96 raw files. After one point recalibration with Recal-offline (Xcalibur software v2.1.0.1140) using the polysiloxane contaminant peak at m/z 445.12003, raw files were subjected to database search and protein quantification with Proteome Discoverer Software (v1.4.0.288) using a database of WormPep238 (26,248 entries) extended by the UniProt complete proteome set of *E. coli* B strain REL606 (4,142 entries, release: 01.Sept.2013), the closest genetic match to strain OP50, as well as the UniRef 90 clusters of *Bt* (30,801 entries, release: 01.Sept.2013) and common contaminants (cRAP, contact-dust and laboratory contaminants; 42 entries). Mascot (v 2.2.07), SequestHT and MS Amanda (v1.4.4.1909) search algorithms were used to annotate HCD-FT and CID-IT spectra with a fragment mass accuracy of 0.04 Da and 0.4 Da, respectively. Parent-mass accuracy for all spectra was set to 7 ppm, with semitrypsin specified as protease, allowing for two missed cleavages. Carbamidomethylation (C) and iTRAQ 4-plex (N-term and K) were set as static modifications, oxidation of methionine and iTRAQ 4-plex (Y) as variable modifications. Percolator (v 2.04) was used for posterior error calculation (Kall, Storey et al. 2008) combining the results of the database searches restricted by q-value to $FDR \leq 0.01$. Single peptide identifications were excluded and protein groups were assigned using strict parsimony rules. For quantification only unique peptides were used and peptide-ratios were normalized by the median of protein-ratios.

Only proteins identified by at least two peptides with quantitative data acquired in all three independent experiments were considered for statistical evaluation with the Perseus software (v1.4.1.3; www.maxquant.org). Quantitative results of \log_2 transformed protein-ratios describing the condition pathogen/non-pathogen *Bt* (P/B) were tested against all \log_2 transformed control-ratios of biological replicates in the same iTRAQ experiment using two sided Welch's t-test. To adjust for multiple testing, p-values were corrected by permutation based FDR approach with 1000 randomizations. Protein groups with \log_2 ratios of at least ± 0.485 (which corresponds to iTRAQ-ratios ≥ 1.4 or ≤ 0.71) with FDR-corrected p-value = $q \leq 0.05$ were considered differentially abundant.

Perseus software was further used for functional annotation of the quantified proteins with GO (gene ontology) terms and InterPro (Hunter, Jones et al. 2012) annotations from UniProt database. Enrichment analyses were performed with Perseus on the selections of (i) differentially, (ii) higher, and (iii) lower abundant proteins using Fisher exact test. One-dimensional (1D) annotation enrichment test (Cox and Mann 2012) was performed on the complete quantitative data. Test results were FDR corrected (Benjamini and Hochberg 1995) and limited to $q \leq 0.05$.

2.4 Results and discussion

2.4.1 Cultivation of *C. elegans* and sample preparation for proteomics

The culture of homogenous samples of *C. elegans* is a prerequisite of relative quantitative proteomics. For this reason, comparative analyses of different growth conditions are usually focused on one specific developmental stage of the worm, which has to be enriched by synchronization or stage specific sorting.

Hypochlorite treatment is the standard method to synchronize a large number of worms (Khan and McFadden 1980, Lewis and Fleming 1995). In this procedure, all adult worms and larval stages are killed and dissolved by bleaching, leaving only embryos protected by the chitinous eggshell. After hatching overnight with no food, worms can be arrested in larval stage L1 and the following differentiated larval stages or adult worms can be harvested after cultivation for the appropriate time. With subsequent hypochlorite treatment over two generations, the temporal resolution of the resulting population can be improved at the expense of yield (Schauer and Wood 1990, Baugh 2009, Baugh, Demodena et al. 2009). Other procedures to stage worms rely on microfluidic devices (Rohde, Zeng et al. 2007) or flow cytometers (Pulak 2006) which use the physical parameters of animal length, optical density, or the intensity of fluorescent markers. However, because the methods have a lower throughput and consequently limit the sample size, they are more suited for microscopic applications or expression analyses at the transcript level.

Conversely, a considerably higher amount of sample is needed for proteomics studies. With a protein content of ca. 250 ng per adult worm (personal communication, F. Döring 2012), an amount of 4,000 worms should suffice to achieve sample sizes of 1mg of soluble protein. However, focusing on L4 larval stage worms, ca. 12,000 worms were harvested for 1mg soluble protein from 10-15 (9 cm) petri dishes. Sorting these amounts of worms with the aforementioned methods for proteomics sample preparation is often not possible or impractical due to time constraints and reproducibility issues. For this reason, a sifting device was developed to facilitate the sorting of large numbers of worms using nylon net filters with mesh sizes of 11, 20, 30 and 40 μm , depending on the sizes of worms to be filtered (Figure 2.2). The functional capabilities of the device were tested by filtering different worm populations.

The flow rate of the device varied with the number of worms, the mesh sizes and number of filters used in a sorting procedure. Commercially available nylon net filters were available with mesh sizes of 11 μm , 20 μm , 30 μm and 40 μm . With a 40 μm and 30 μm filter setup, flow rates of more than 5 mL/s were achieved. Using 30, 20 and 10 μm filters in tandem, the flow rate can decrease to 0.4 mL/s.

The performance of the sifting device while employing different nylon nets was assessed by microscopy of the collected worms. Recorded microscopic images were processed and analyzed using the Java based software ImageJ.

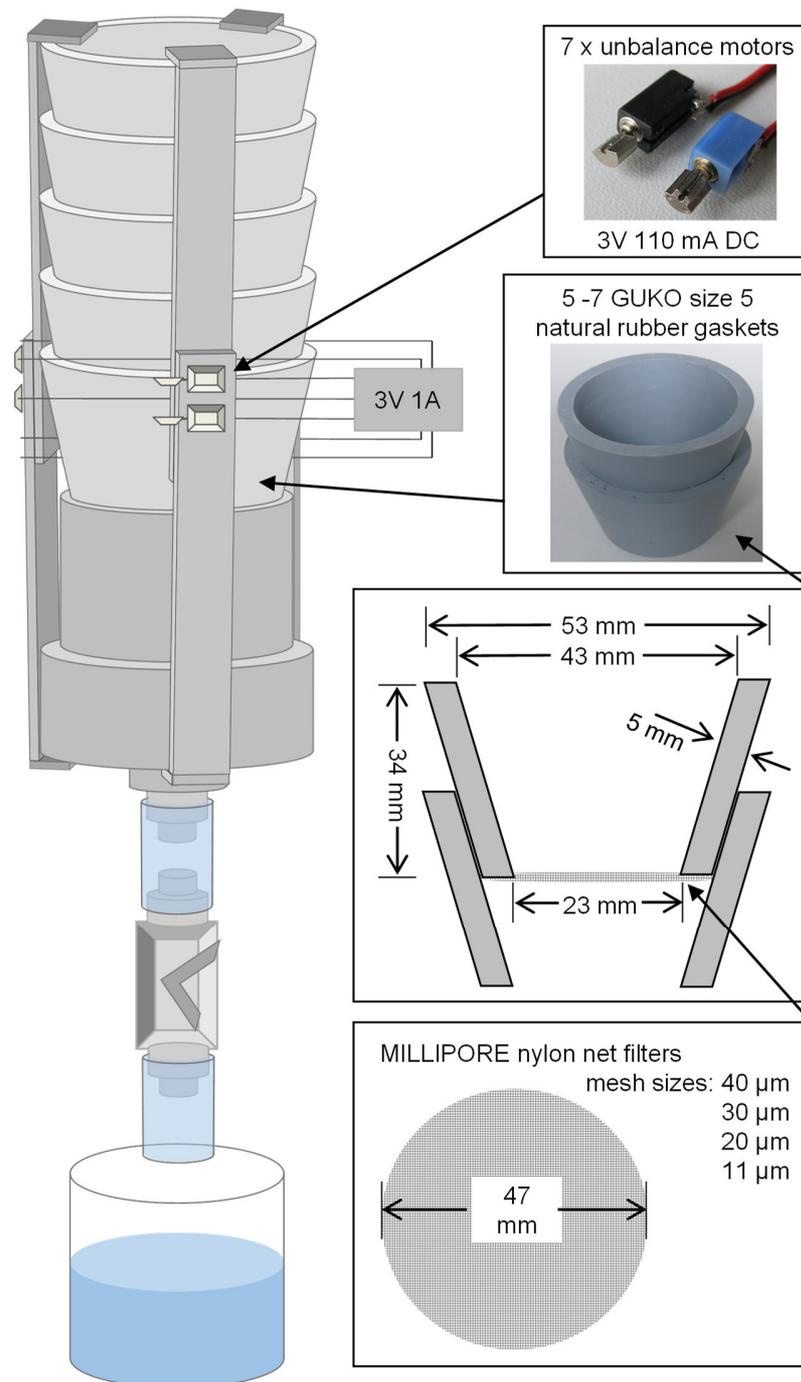


Figure 2.2: Schematic of the sifting device used for the harvest and cleanup of *C. elegans* cultures. The basic module consists of one nylon net filter fixed between two GUKO natural rubber funnels. Stacked modules are secured in place via a metal brace equipped with seven unbalance motors which prevents clogging and facilitate sorting. Worms from petri dishes were collected in solution. The solution was filtered through nylon nets of mesh sizes between 11 µm and 40 µm.

2.4.1.1 Measurement of *C. elegans* dimensions by microscopy

The Java programming language was used to write a script for fully automated image processing. The standard processing steps for all images followed the scheme outlined in Figure 2.3. First, the scale of images was set to translate the width of pixel into real spatial units. Then several steps were necessary for reliable object recognition, to remove shading and differentiate objects from background. In addition to the functions included in the ImageJ software package, the “ROI Curve” plugin by Bob Dougherty (2007) was used to measure worm length, width and area. The plugin can measure the curve length of a region of interest (ROI) as the average centerline between the outline arc curves connecting the points of largest separation of an indicated particle. Therefore, the accurate measurements of worms in a straight or curved orientation is supported.

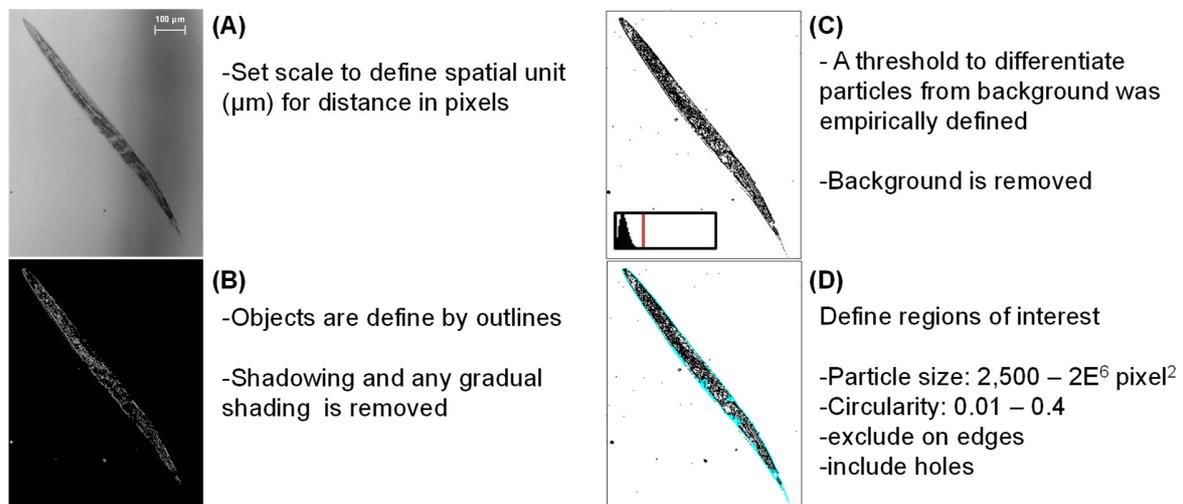


Figure 2.3: Processing of microscopic images. Image processing steps were necessary to ensure reproducible automated measurement of worms on microscopic images: **(A)** Spatial calibration; **(B)** find edges; **(C)** set threshold; **(D)** analyze particles.

For automation, the steps were run in a loop to batch-process microscopic images and a table of the measurement results including the maximum curve width, curved length and particle size was exported for data processing in Excel.

Using the body width for volume calculation was found to introduce exponentially large errors because the body width was often too small to accurately define the borders using an automated image processing script. This issue applies to *C. elegans* in particular, which has a length to width ratio of about 15. The worm volume therefore calculated from microscopic dimensions using only the body length and area of the animal (as the bisectonal area of the cylinder) according to Formula 2.1.

Formula 2.1: $V_w = \frac{A^2 \pi}{4L}$; with A = area of the body

2.4.1.2 Validation of nylon net filter assisted sorting

To test sorting capabilities with different available mesh-sizes used in the sorting device, a continuous mixed culture was separated over the sorting device. Young adults were again filtered by nylon nets (30 μ m mesh size) and ca. 50 young adult worms were cultivated on 9 cm petri dishes for 7 days. The obtained culture containing worms of all stages was harvested and applied to the device for sorting. All available filters were used from a mesh size of 40 μ m at the top, decreasing to 30 μ m, then 20 μ m and 11 μ m nylon net at the bottom. The filters of 40 μ m mesh size contained exclusively old adult worms. Worms of larval stage L4 were specifically loaded on 20 μ m nylon nets. The worm population filtered by 10 μ m mesh sizes consists of ca. 25 % L3 and 75 % L4 larvae. All worms smaller than L3 as well as eggs were completely filtered out (Table 2.0.1).

Table 2.0.1: Worm stages and dimensions retained on nylon net filters of specified mesh sizes. Worms were collected from indicated nylon net filters and larval stages were determined from microscopic images. The percentage of worm stages retained on respective net sizes is indicated. Additionally, the average width and length of worms collected in each filter is added with standard deviations (σ).

μ m	width x length [μ m \pm σ]	Adult [%]	L4 [%]	L3 [%]	L2	dauer	L1	eggs
40	65 \pm 7 x 1039 \pm 76	100	0	0	0	0	0	0
30	59 \pm 6 x 906 \pm 76	73	27	0	0	0	0	0
20	38 \pm 6 x 647 \pm 81	0	100	0	0	0	0	0
11	26 \pm 5 x 518 \pm 87	0	56	44	0	0	0	0

The proportions of developmental stages in cultures using different worm strains or culture conditions can diverge significantly. The capabilities of the sorting device for large scale sample preparation of different *C. elegans* strains was tested. Three strains, which showed differences in their growth rate, including *C. elegans* N2 Bristol and two variant strains *adam-4* and *sup-17ts*, were used for this purpose (see Figure 2.4).

The different strains were synchronized by hypochlorite treatment and then grown for appropriate times so that most worms reached larval stage L4 (Figure 2.4 A). Worms were harvested either by sucrose floating or between 20 μ m and 30 μ m nylon net filters using the sorting device. The distribution of worms from different strains clearly show that after growth for 60 – 70 h, the resulting samples show an inhomogeneous size distribution. In particular the variant strain *sup17ts* showed a much broader range of worm sizes in the collected samples. In contrast the differences were

minimized after filtering samples through the sorting device, collecting only worms trapped between 20 μm and 30 μm mesh sized nylon nets.

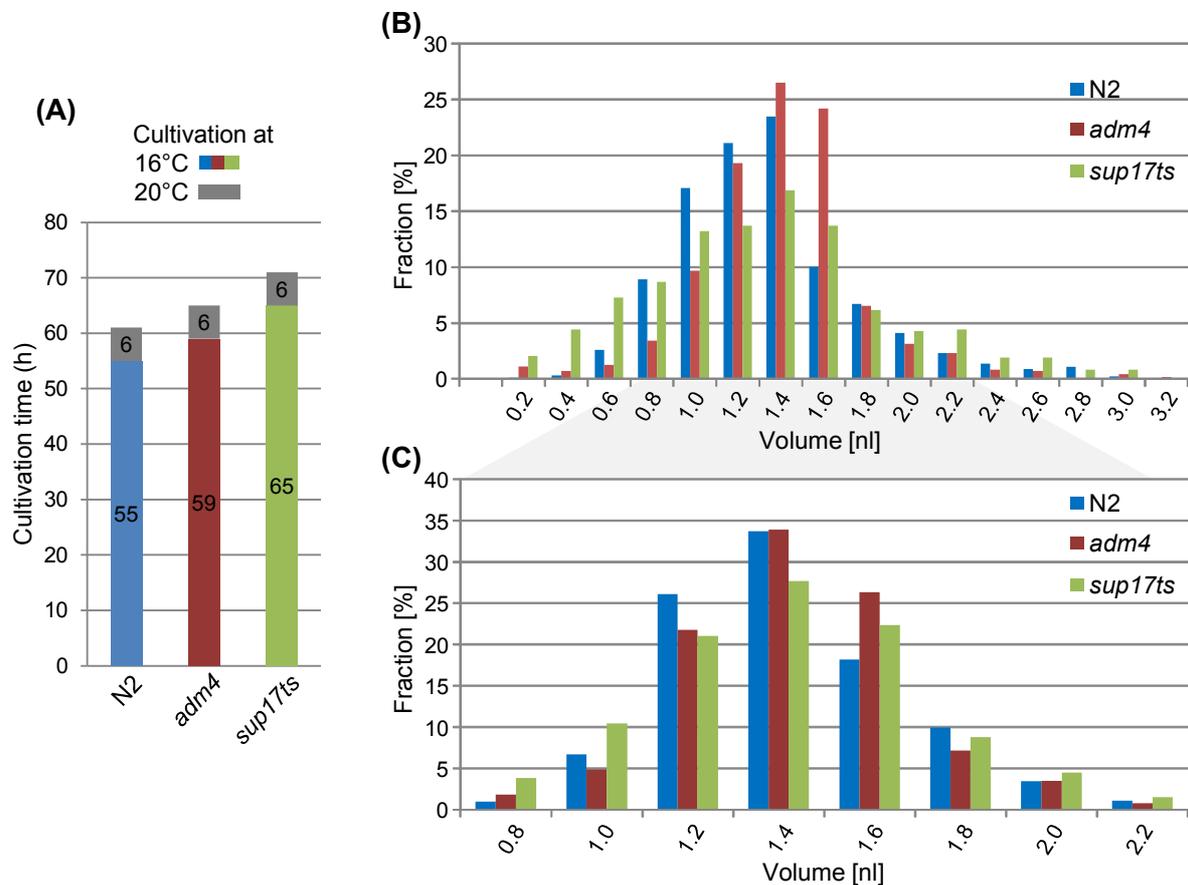


Figure 2.4: Hypochlorite treatment vs. nylon net filtered sorting. Three different *C. elegans* strains were grown for different times after synchronization (A). The histograms show the worm size distributions of populations collected from cultures synchronized by bleaching (B) or additionally sorted with nylon net filters (C) ($n = 500 \pm 100$). Worms were filtered through a device containing two filters with mesh sizes 30 μm followed by 20 μm , and harvested from the 20 μm nylon net.

2.4.1.3 Sample preparation for *C. elegans*-pathogen interaction proteomics

With the results of previous sorting experiments using nylon net filters, the following protocol was designed to obtain samples for the analysis of *C. elegans* host-pathogen interactions by relative protein quantification. First, mixed cultures were synchronized by hypochlorite treatment and grown to larval stage L3 - L4. These worms could be efficiently collected on 11 μm nylon net filters using the sifting device.

C. elegans were harvested in larval stage L3 - L4 and bacteria from previous cultivation were removed. Subsequently, ca. 1200 (± 200) worms were transferred to petri dishes seeded with

different bacterial lawns for differential treatment. After treatment on the different plates for 12 h, worms were again collected from plates and sorted on three filters of 10, 20 and 30 μm mesh size. Late L4 and young adults are collected in the 30 μm filter and remaining L3 as well as some small L4 larvae are collected in the 10 μm filter. Only L4 worms could be detected in the 20 μm filters and were used as the sample for proteome analyses.

In addition to the reduction of the dynamic range of worm sizes, the sample preparation time was also significantly shorter using the net filter sorting procedure. As an additional advantage of the sorting procedure using nylon nets, sucrose floating could be avoided entirely. Reliable quantitative proteomic analysis requires removal of contaminating bacteria, deceased worms, and other debris. Sucrose floating is the conventional method to remove these undesirables. To process worms collected from 12 to 15 petri dishes using the conventional method takes approximately 1 h and exposes the cultures to both cold and hypotonic environments which may artificially change the proteome. In contrast, harvesting an equivalent number of cultures by sifting on nylon net filters and on-filter washing at 20 °C is a more gentle procedure that can be performed in half the time. The removal of bacterial cells could be verified by sucrose floating of sifted worm populations with the result that no bacterial pellet was formed upon centrifugation. Additionally, proteomic data sets acquired from samples prepared by sucrose floating or by filtered on nylon nets were searched against a database taxonomically restricted to *C. elegans* and *E. coli* protein sequences. No significant difference in the proportion of identified *E. coli* proteins was observed, which accounted in both preparations for less than 0.6 % of the overall identifications.

In summary, nylon net filter assisted stage specific worms sorting is a quicker and much gentler way to attain homogenous sample of sufficient size for quantitative proteomics studies than the classical synchronization of cultures by bleaching with subsequent sucrose floating to remove contaminants. However, although the approach provides a useful alternative to sucrose floating it does not remove the necessity for frequent synchronization by bleaching, especially to remove contaminants and sterilize *C. elegans* cultures.

The selection of synchronized cultures of *C. elegans* was of considerable importance to the accuracy and reproducibility of this study. The harvesting procedure utilized herein, aided by an in-house developed shaking apparatus, allowed the selective isolation of target populations of *C. elegans*. In addition, the harvesting procedure facilitated a gentle clean and fast transfer of the cultures between routine growth plates and treatment plates.

2.4.2 Quantitative proteome analysis by isobaric labeling and 2D-LC MS

In order to achieve robust, statistically significant results in quantitative proteomics, the analysis of biological and technical replicates is indispensable. As a consequence, the overall instrument times increase significantly. The isobaric labeling approach addresses both of these issues. Although the iTRAQ 4-plex isobaric labeling strategy is still limited to four channels, the multiplexing capability reduces the number of necessary experiments and minimizes technical variability. The complete workflow is summarized in Figure 2.5.

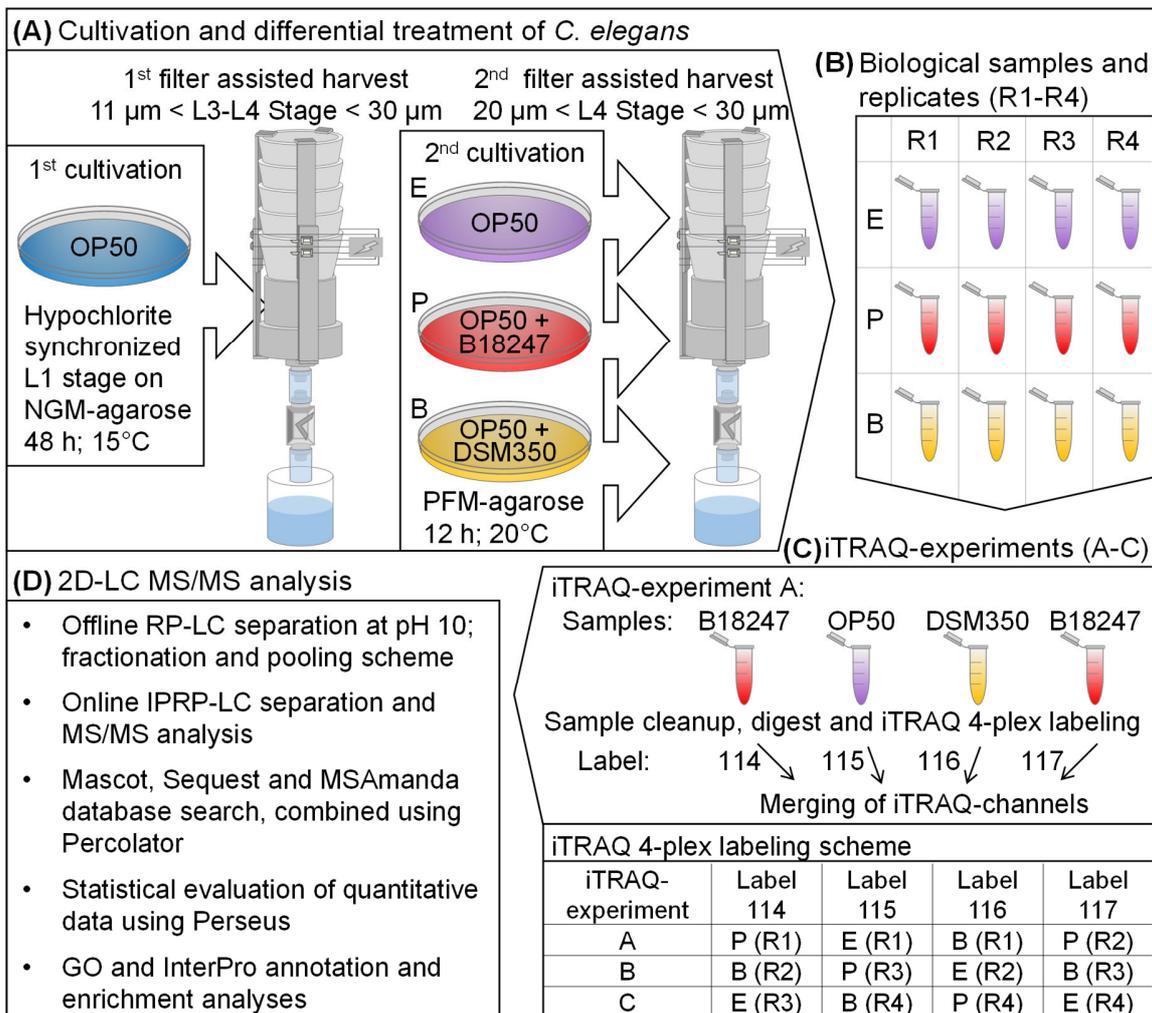


Figure 2.5: Workflow and experimental design. (A) Cultivation and treatment of *C. elegans* including the filtering steps employed to facilitate clean-up and transfer. (B) Illustrates the number of samples and replicates. (C) Shows the experimental design after sample preparation, iTRAQ labeling, twelve individual samples and combination in three LC-MS experiments, Exp-A, Exp-B and Exp-C. (D) Summarizes key steps of 2D-LC-MS/MS analysis and data evaluation.

After labeling the digested proteomes with iTRAQ 4-plex reagents, a mixing scheme was used to combine the twelve proteome samples to three 2D-LCMS experiments, containing all biological replicates (Exp-A, -B and -C).

For each of these three experiments an offline two-dimensional separation scheme was applied to separate the iTRAQ labeled peptides. The first dimension of separation was achieved by reversed-phase chromatography, using a water/acetonitrile gradient at pH 10. To make better use of the chromatographic separation space, a staggered pooling system was used prior to the second dimension separation (Figure 2.6).

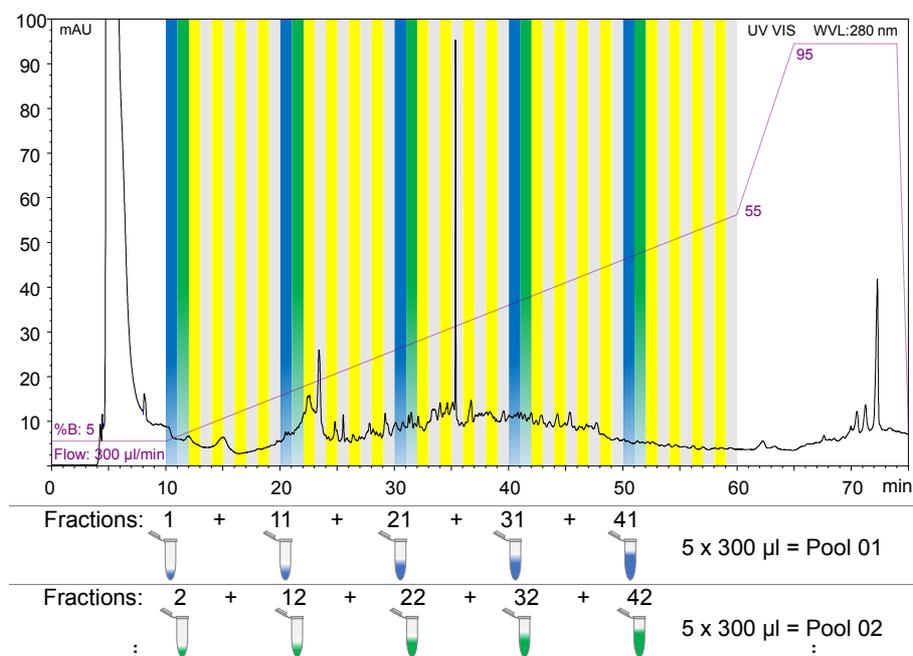


Figure 2.6: Diagram of the pooling scheme applied to concatenate first dimensional fractions. The chromatogram of a first dimension reversed-phase separation (Experiment A) shows the fractionation and pooling strategy utilized prior to 2nd dimension separations. First dimension fractions were collected in 1 min intervals and pooled as exemplarily shown for pool 01 and pool 02 in blue and green, respectively.

Separation in the second dimension was performed by classical reversed-phase liquid chromatography (RP-LC) using a water-acetonitrile gradient at pH 2 with formic acid. Each of the ten pooled samples within the three experiments, were measured in triplicate on independent second dimension LC-MS runs (technical replicates). Additionally, six of the pooled fractions were re-measured in a fourth LC-MS run in order to balance out technical issues within individual runs, e.g. poor spray stability.

Analysis of the number of peptides shared across adjacent pooled fractions revealed that less than 10 % of the peptides are present in the preceding or following fractions (Figure 2.7 A). This demonstrates the high separation capacity of the chosen fractionation/pooling approach which minimizes repetitive measurements leading to deeper coverage of the proteome.

Comparing the peptide identifications in the first two technical replicates revealed a very high degree of redundancy (85 %) in both runs. In order to further increase proteome coverage an exclusion list, compiled from the retention times and m/z values of peptides identified with high

confidence (FDR <0.01) in the first two measurements, was generated and applied for the third replicate LC-MS run. Redundant identifications in the third technical replicate decreased to an average of 47 % compared to each of the first two runs. However, it should be noted that due to the exclusion list, the number of identified peptides was reduced to approximately 59 % compared to the two preceding replicates. In summary, the absolute gain of non-redundant peptide identifications increases from an average of 349 per run (for the first two replicates) to 992 for each pooled fraction measured while employing exclusion lists (Figure 2.7 B).

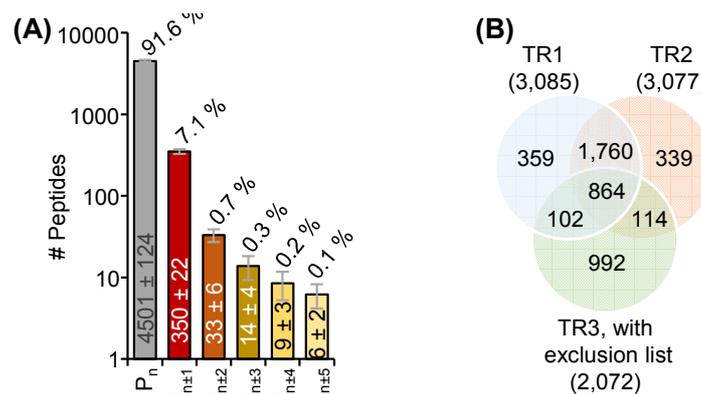


Figure 2.7: Assessment of the 2D-LC separation and technical replicates of MS/MS analysis. (A) Columns illustrate the overlap of peptides identified in the pooled fractions of first dimension LC with adjacent fractions, exemplarily shown for the 2D-LC-MS/MS experiment Exp-A. The average number of peptides identified across the ten fractions (P_n), and the average overlap with neighboring fractions ($P_{n\pm 1-5}$) are represented as total numbers (\pm standard deviations) inside columns. **(B)** Venn-diagram compares the average overlap of peptide identifications from three technical replicates across all fractions of the three iTRAQ experiments. High confidence identifications from the first two technical replicates (TR1, TR2) were excluded from data dependent acquisition in the third run (TR3) to increase overall proteome coverage.

The use of an efficient two-dimensional separation scheme significantly contributed to the success of the quantitative study. Coupling a first dimension reversed-phase liquid chromatography separation at pH 10 to a staggered fraction pooling design (Stephanowitz, Lange et al. 2012), followed by second dimension separation using classical ion pairing reversed-phase liquid chromatography (IP-RP-LC) at pH 2, resulted in a highly efficient use of the chromatographic separation space. The quasi-orthogonal separation scheme offers a higher level of robustness compared to traditional strong cation exchange (SCX) x IPRP-LC approaches as it omits increased salt load from the first dimension, and charge state clustering (Delmotte, Lasaosa et al. 2007). Altogether, the strategy employed both reduced the time requirements for the second dimension separation and the ESI-MS instrument by 80 % while at the same time making more efficient use of the available chromatographic space.

Across the three experiments, 61,326 grouped peptides (by sequence) were identified in a database that combined proteins from each of the three species under study, namely *C. elegans*,

E. coli, and *Bt*. These peptides could be assigned to 20,086 database entries (WormBase: 12,349, *B. thuringiensis*: 6,725, *E. coli*: 984, contaminants: 28). Conservative maximum parsimony rules were employed which assigned the peptides to a total of 7,366 protein groups; 5,371 grouped proteins were identified by multiple peptide hits, while 1,995 protein groups were single peptide identifications. However, due to the relatively low proportion of *Bt* compared the host cells in the analyzed sample, only 28 *Bt* proteins were identified by multiple peptides (Figure 2.8 A).

The overlap of peptides and proteins identified in the three different 2D-LC MS experiments emphasizes the high level of reproducibility for the chosen analytical workflow (Figure 2.8 B and C). Of the 7,366 identified protein groups, 75.9 % were found in at least two of the three experiments, while 60.7 % were found in all three experiments.

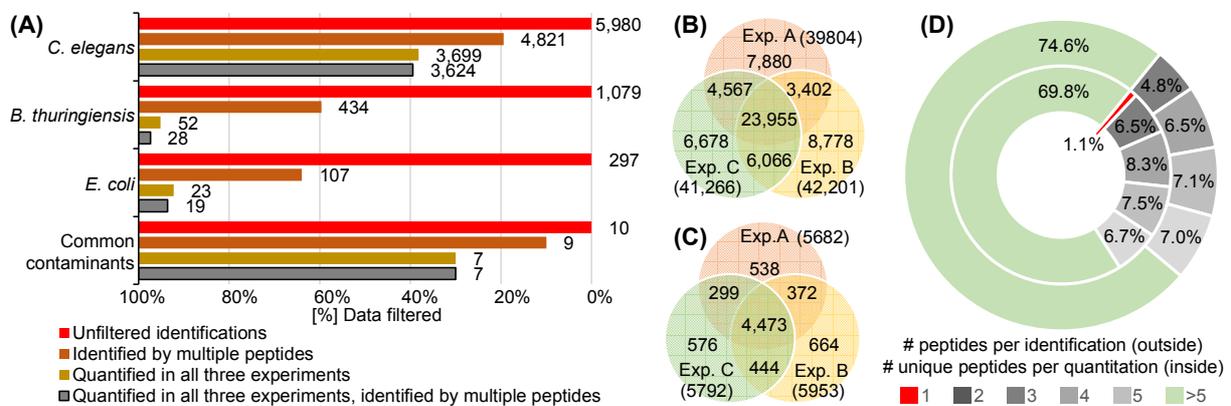


Figure 2.8: Assessment of identifications and quantitative data. (A) Shows the distribution of identified protein groups by taxonomy and the results of data filtering with the applied quality criteria. The total number of protein groups in each category is shown next to columns. (B) Peptides identified in the three independent iTRAQ experiments (Exp. A-C). (C) Identified protein groups in Exp. A-C. (D) Number of peptides supporting protein identifications vs. unique peptides supporting quantitation of the 3,624 filtered *C. elegans* protein groups.

To ensure reliable relative quantification, only proteotypic peptides were used for quantification. In addition, protein groups were only considered if the identification covered at least two peptides and if quantitative data were acquired in all three independent 2D-LC MS experiments. Employing these criteria 96 % of the contaminant protein groups and 39 % of the *C. elegans* proteins were filtered out (Figure 2.8 A). Overall, 3,678 protein groups met the criteria, 3,624 of which were taxonomically assigned to *C. elegans*. Approximately 70 % of these proteins were quantified with more than five peptides and an additional 29 % were quantified with two to five peptides (Figure 2.8 D).

The use of iTRAQ 4-plex allowed robust validation of the data. Because in each of the three experiment one of the three biological replicates was represented twice, the labeling scheme enabled the generation of control-ratios that allow both inter- and intra-experimental validation.

The distribution of control-ratios between these replicates takes into account both, the biological and the technical variations associated with the chosen experimental design (Figure 2.9 A).

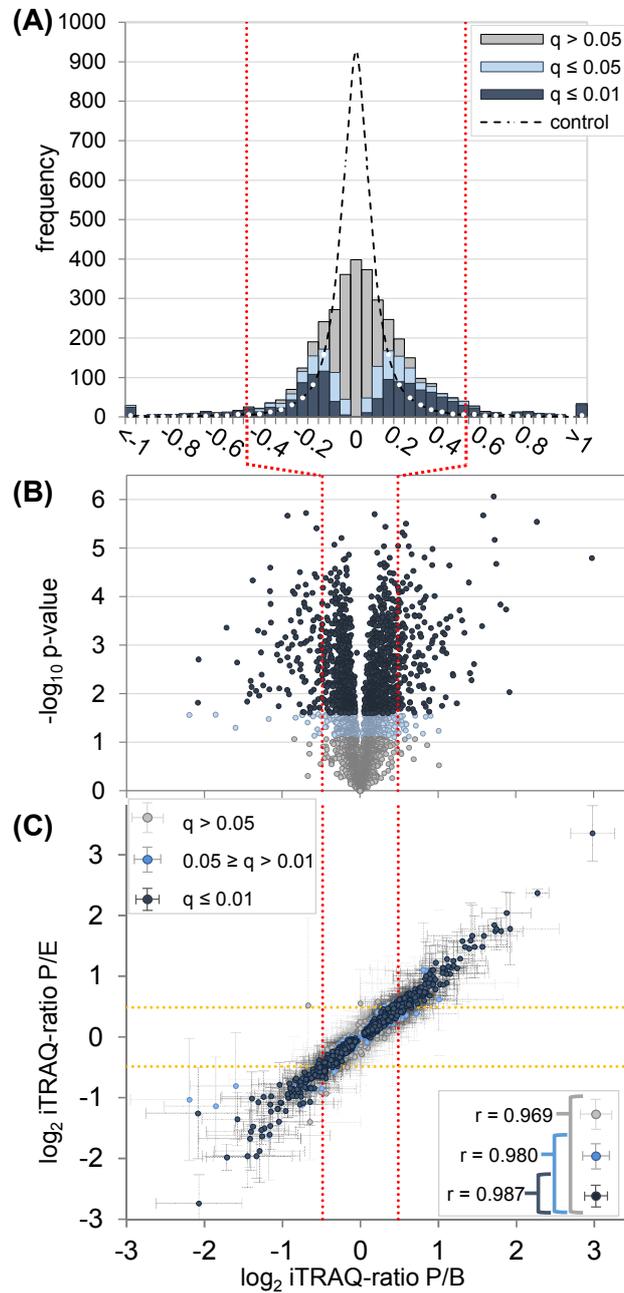


Figure 2.9: Quantitative data acquired for 3,624 *C. elegans* protein groups. (P = pathogen *Bt*; B = non-pathogen *Bt*; E = *E. coli*). **(A)** Histogram (bin size 0.05) of mean \log_2 iTRAQ protein-ratios for P/B vs. control ratios of biological replicates of the same treatment (stacked columns indicate proportions of indicated q-value bins for the P/B protein-ratios). **(B)** Volcano plots showing the median \log_2 iTRAQ-ratios and $-\log_{10}$ p-values from Welch's t-test of the quantified proteins; the q-value cut-off for FDR ≤ 0.01 and ≤ 0.05 are indicated by grey, light blue and dark blue colors, respectively. Dashed lines indicate the 1.4 fold-change cut-off. **(C)** Correlation between P/B and P/E ratios; error bars indicate geometric standard deviation. The Pearson's correlation coefficients for the subgroups of proteins with $q \leq 0.01$; $q \leq 0.05$ (P/B vs. control) and the complete dataset are shown in the lower right corner

On FDR filtered data obtained by statistical evaluation of the quantitative results (q -value ≤ 0.05), an additional fold-change cut-off ≥ 1.4 (\log_2 ratio cut-off ± 0.485) was applied to determine differential abundance in this study. This corresponds to a value which is outside the boundaries set by three times the standard deviation observed between \log_2 control-ratios ($\sigma = 0.149$) (Figure 2.9 B). The ratios between the worms challenged with pathogenic *Bt* (P) and those challenge with non-pathogenic strain *Bt* (B) were reported. The protein-ratios describing the conditions P/E vs. P/B showed a strong correlation ($r = 0.965$, see Figure 2.9 C). With the defined cut-off values, 288 proteins showed a significant difference in abundance in *C. elegans* challenged with pathogenic *Bt* compared to those cultivated with non-pathogenic *Bt*. Within these, 171 were increased in abundance, while 117 were of lower abundance.

Fisher exact test was used to assess the differentially abundant proteins for either enrichment or depletion of InterPro protein families and functional domains. Additionally, all quantitative data were analyzed by 1D annotation enrichment (Cox and Mann 2012) to detect systematic trends of the proteins within categories. Significant results, after removal of redundant categories, are summarized in Table 2.0.2.

Table 2.0.2: Results of enrichment analyses for protein functional domains showing the number of selected proteins in each category; the corresponding Benjamini-Hochberg corrected p -values as critical FDR values (q) are parenthesized for significant test results. Significant hits of 1D-enrichment analysis are included with respective mean- and median \log_2 ratio of all proteins in the category with q -values in parentheses.

Term	Identified	differential (q)	higher (q)	lower (q)	mean ; median \log_2 ratio (q)
Total number:	3624	288	171	117	3624
Galectin	12	7 (6.4E-3)	7 (1.1E-3)	0	0.57 ; 0.56 (1.7E-4)
C-type lectin	18	10 (2.6E-4)	7 (6.9E-3)	3	0.44 ; 0.29
Glycoprotein (DUF274)	7	5 (2.7E-2)	5 (4.1E-3)	0	0.92 ; 0.96 (1.3E-02)
Fatty acid desaturase	6	5 (9.4E-3)	0	5 (2.4E-4)	-0.92 ; -0.89 (9.4E-3)
Transthyretin-like (DUF290)	22	6	6	0	0.5 ; 0.37 (6.0E-6)
Intermediate Filament	11	6 (2.6E-2)	6 (4.8E-3)	0	0.56 ; 0.50 (2.2E-4)
Myosin tail	9	3	3	0	0.39 ; 0.46 (1.4E-2)
Calponin homology domain	22	2	1	1	0.21 ; 0.20 (1.5E-2)
Collagen	37	22 (1.7E-12)	4	18 (7.7E-15)	-0.42 ; -0.46 (4.4E-4)
SH3 domain	26	1	1	0	0.26 ; 0.25 (1.1E-4)
Ig-like fold	32	0	0	0	0.22 ; 0.29 (7.9E-5)
Rib L2 domain 2	9	0	0	0	-0.22 ; -0.24 (2.4E-2)

The enrichment analyses highlight several protein classes as being strongly affected by the presence of the pathogenic strain, e.g. a number of galectins, C-type lectins, transthyretin-like proteins, and several structural proteins.

Enrichment analysis for gene ontology annotations (Ashburner, Ball et al. 2000) of the 171 proteins increased in abundance following *Bt* challenge revealed the involvement of structural proteins of the cytoskeleton ($q < 5.2E-3$; mean $\log_2(P/B) = 0.29$), proteins located in the extracellular space ($q < 4.6E-3$; EF = 5.1; mean $\log_2(P/B) = 0.3$) and nucleosomes ($q < 4.1E-3$; EF = 8.2; mean $\log_2(P/B) = 0.38$) as well as proteins associated with cellular component assembly ($q < 7.1E-3$; EF = 2.96; mean $\log_2(P/B) = 0.17$). In addition, proteins with peptidase regulator activity ($q < 1.4E-2$; EF = 6.3; mean $\log_2(P/B) = 0.29$) and proteins with carbohydrate binding activity ($q < 8.3E-7$; EF = 7.1; mean $\log_2(P/B) = 0.32$) were significantly affected. 1D annotation enrichment revealed systematic trends towards higher relative abundance of several additional characteristic protein groups including those involved in immune defense ($q < 4.4E-2$; mean $\log_2(P/B) = 0.22$), the response to external stimuli ($q < 7.5E-4$; mean $\log_2(P/B) = 0.24$), neuron projection ($q < 4.6E-6$; mean $\log_2(P/B) = 0.23$), and the binding of calcium ions ($q < 1.6E-2$; mean $\log_2(P/B) = 0.17$) and phospholipids ($q < 6.1E-5$; mean $\log_2(P/B) = 0.22$).

In contrast, the 117 proteins with decreased abundance in worms challenged by *Bt* were involved in organic acid ($q < 3.8E-2$; EF = 2.5; mean $\log_2(P/B) = -0.11$) and lipid metabolism ($q < 2.2E-3$; EF = 4.1; mean $\log_2(P/B) = -0.11$) or structural constituents of collagen ($q < 7.7E-15$; EF = 15.1; mean $\log_2(P/B) = -0.42$) and of ribosomes ($q < 8.2E-32$; mean $\log_2(P/B) = -0.19$).

GO and InterPro annotations allowed the classification of many of the identified proteins and differentially abundant proteins were identified with associations to particular physiological processes or molecular functions, such as the innate immune response, proteolysis, signal cascades, apoptosis, or gene regulation. However, it must be noted that a more than 25 % of the 288 differentially abundant proteins have not been functionally characterized yet.

2.4.3 Proteins identified from pathogenic *B. thuringiensis*

The *Bt* strain B-18247 used in this study was reported to express three crystal toxins, only one of which (Cry6Ba) has been characterized. However, *E. coli* transformants expressing this toxin did not show nematicidal activity, possibly due to non-native folding or the absence of additional factors necessary for activation of the toxin (Wei, Hale et al. 2003). Due to the relatively low proportion of *Bt* compared to the host cells in the analyzed sample, comparatively few *Bt* proteins

were identified. Although unfiltered data included more than 1,000 possible protein identifications, the low quality of these data is apparent by filtering for multiple peptide identifications which result in only 434 proteins. Additionally filtering for proteins quantified in all three experiments leaves only 28 proteins (see Figure 2.8 A). Interpretation of such a limited set of data in a biologically meaningful way is not straightforward; however, several proteins of potential interest were identified. Of the *Bt* derived proteins identified, Cry6Ba was detected with high confidence and sequence coverage (72 %), which may be a potential virulence factor. The presence of at least one other toxin was confirmed by the identification of Cry1Ba (44 % sequence coverage). However, the identified peptides may be derived from a different toxin, not represented in the database, but showing sufficient sequence similarity with Cry1Ba for positive identification.

One protein potentially involved in the activation of toxins is the metalloprotease camelysin which has been shown to process cytosolic endotoxins (Nisnevitch, Sigawi et al. 2010). The protein was detected in the UniProt Reference Clusters (UniRef, see Materials and Methods) with 49 % sequence coverage of the *Bacillus cereus* (strain ATCC 14579/DSM 31) camelysin and 17 % sequence coverage of the camelysin from *Bt* strain Al Hakam. Another identified protein, which may contribute to pathogenicity, was the Immune inhibitor A (24 % sequence coverage).

2.4.4 Effects of pathogenic challenge on the *C. elegans* proteome

In this study *C. elegans* was challenged either with a pathogenic or with a non-pathogenic strain of *B. thuringiensis*. This allows differentiating between effects induced by pathogenicity of *Bt* and those caused by the pure presence of *Bt* as a species in xenic cultivation. A special aspect in the applied experimental setup, not taken into account in many earlier studies, is the constant presence of *E. coli* as food source even during challenge with pathogenic or non-pathogenic *Bt*. With this, artificial effects in the proteome composition due to starvation of the worm can be excluded. Furthermore, the worms solely grown in axenic culture on the laboratory food bacterium *E. coli* (E) served as an additional control.

Of the 3,624 quantified *C. elegans* proteins, a strong correlation was observed between the protein-ratios describing the conditions pathogenic *Bt* vs. non-pathogenic *Bt* (P/B) compared to pathogenic *Bt* vs. *E. coli* (P/E) (Figure 2.10). This demonstrates that the non-pathogenic gram-positive bacterium does not alter the *C. elegans* proteome significantly compared to the food source *E. coli*. This observation is supported by comparing the distribution of control-ratios (between biological replicates in the same iTRAQ experiments) against protein-ratios comparing the two controls, the non-pathogenic *Bt* vs *E. coli* B/E (B/E). The observed differences fit inside

the expected biological and experimental variability. Furthermore, no significant changes were observed comparing the protein abundance profiles of *C. elegans* grown either exclusively on *E. coli* or after addition of non-pathogenic *Bt*. With these controls, the differential abundances of 288 proteins identified in this study could be clearly attributed to the interaction with the pathogenic *Bt* strain.

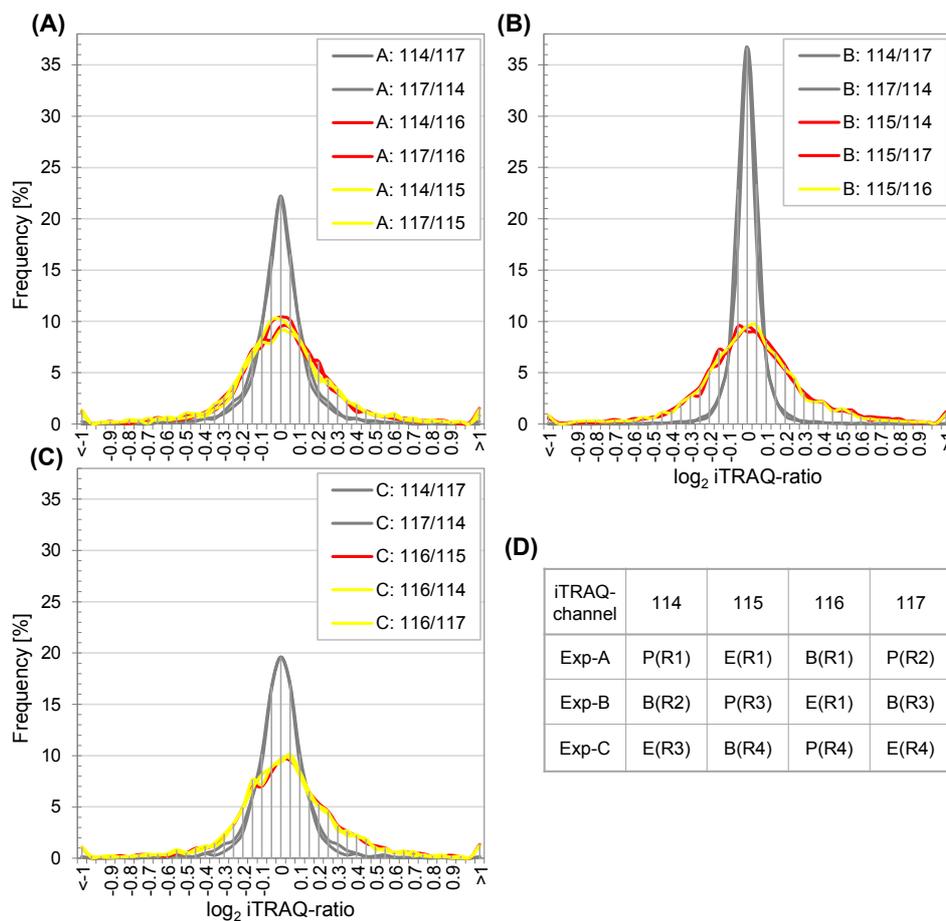


Figure 2.10: Frequency distribution of all iTRAQ protein-ratios. *C. elegans* proteome after treatment with pathogenic *Bacillus thuringiensis* (P), non-pathogenic *Bt* (B) or the standard laboratory food strain of *E. coli* (E). Histograms with a bin size of 0.05 (± 0.025 around the value indicated on horizontal axes) show the frequency of P/B (red), P/E (yellow), and the control-ratios between biological replicates of the same treatment (grey), for the tree iTRAQ 4-plex experiments in (A), (B) and (C). The allocation of iTRAQ channels is summarized in (D).

2.4.5 Differentially abundant protein families

From the results of the enrichment analysis (Table 2.0.2) and previous host-pathogen experiments with *C. elegans*, several protein families were of particular interest. These included the lectins, the lysozymes and the transthyretin-like proteins.

Lectins are proteins that bind carbohydrates. The interaction of lectins with its target carbohydrate can be highly specific and allows this class of protein to participate in a vast range of recognition and interaction phenomena in nature (Ghazarian, Idoni et al. 2011). The biological role of many members of this diverse class of proteins is yet undefined. Nevertheless, there is a growing body of evidence supporting their role in cellular recognition, cell signaling and differentiation, cell adhesion and migration, and most importantly, with regard to this study, host-pathogen interactions and innate immunity (Ghazarian, Idoni et al. 2011). Two families of lectins, the galectins and the C-type lectins, were significantly enriched in the differentially abundant proteins (Table 2.0.2).

The galectins LEC-1, LEC-2, LEC-5, LEC-6, LEC-8, LEC-9, and LEC-10 were identified at higher abundances in response to pathogenic *Bt* (Figure 2.11 A); of these, LEC-6, LEC-9 and LEC-10 are known to be localized in the intestine (Maduzia, Yu et al. 2011). It has been shown that bacterial crystal toxins bind to glycolipids and glycoproteins using two different lectin domains (Ma, Rahman et al. 2012). Moreover, pre-feeding of insect larvae (*Helicoverpa armigera*) with recombinant LEC-8 enhanced the tolerance to a crystal toxin in this organism (Ma, Schmidt et al. 2012). LEC-8, LEC-9, and LEC-10 all can bind glycolipids and LEC-8 has been shown to prevent Cry5B binding in a dose-dependent competitive manner (Ideo, Fukushima et al. 2009). Thus, these findings support the hypothesis that not only LEC-8 but also several other galectins identified as over-abundant may be involved in the defense against *Bt* toxins.

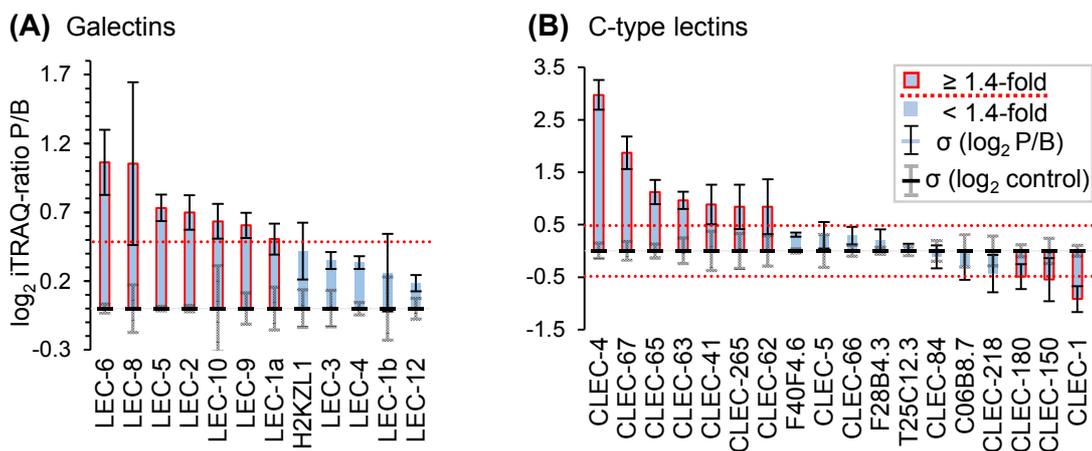


Figure 2.11: Relative abundance changes of carbohydrate binding proteins upon pathogen challenge. (A) Galectins; (B) C-type lectins. Columns show the mean \log_2 protein-ratios in *C. elegans* grown with pathogenic vs. non-pathogenic *Bacillus thuringiensis* (P/B). Black error bars indicate geometric standard deviations of \log_2 (P/B), grey error bars show standard deviations of transformed control-ratios between biological replicates of the same condition (P/P; B/B and E/E) in the three iTRAQ experiments (around the default mean of \log_2 transformed ratios = 0) for the indicated proteins. The dotted lines indicates the ± 1.4 fold-change cut-off values.

In contrast to the galectins, which showed a general increase upon challenge with pathogenic *Bt*, out of the eighteen C-type lectins identified in this study, seven increased significantly in abundance while three were observed to decrease (Figure 2.11 B). C-type lectin-like domain containing proteins (CTLD) have been described to play a role in cell adhesion, glycoprotein binding, and binding of specific pathogen-derived molecules by Ca^{2+} dependent binding of carbohydrates (Ghazarian, Idoni et al. 2011). The first evidence supporting a direct role of *C. elegans* CTLD in pathogen pattern recognition was reported recently, however, binding was found to be Ca^{2+} independent (Miltsch, Seeberger et al. 2014).

In the complete dataset, the highest protein-ratio was observed for the protein Clec-4, which contains two C-type lectin domains and one CUB domain. The Clec-4 gene in *C. elegans* has previously been identified as transcriptionally induced both in response to *E. coli* heterologously expressing the nematocidal *Bt* toxin Cry5B (Griffitts, Whitacre et al. 2001) and to *Bt* spore crystals of the same strain used in this study (Boehnisch, Wong et al. 2011).

The abundance of the proteins Clec-62, Clec-63, Clec-65, and Clec-67 was also observed to be significantly increased in response to the pathogenic *Bt* in this study. Transcriptional upregulation of genes coding for these four C-type lectins was observed previously in response to both gram-positive (e.g. *Enterococcus faecalis* (Wong, Bazopoulou et al. 2007); *Microbacterium nematophilum* (O'Rourke, Baban et al. 2006)) and gram-negative bacteria (e.g. *Pseudomonas aeruginosa*) (Shapira, Hamlin et al. 2006). At the protein level, a similar response was observed for Clec-63 and Clec-65 following incubation with pathogenic *E. coli* LF82 (Simonsen, Moller-Jensen et al. 2011). The gene of Clec-41, the protein of which was also found in higher abundance here, has been previously described to be transcriptionally induced as a reaction against *P. aeruginosa* (Shapira, Hamlin et al. 2006).

Clec-1, Clec-150, and Clec-180 appeared to be less abundant in the presence of pathogenic *Bt*, the coding genes of the latter two have been reported to be induced in response to the gram-negative pathogens *Photobacterium luminescens* and *Erwinia carotovora* (Mallo, Kurz et al. 2002, Huffman, Abrami et al. 2004, O'Rourke, Baban et al. 2006, Shapira, Hamlin et al. 2006, Wong, Bazopoulou et al. 2007).

More than 270 C-type lectin-like domain containing proteins are encoded in the *C. elegans* genome. The observed differential expression of CTLDs following *Bt* toxin exposure supports their role as part of a specific response mechanism. The diverse nature of this protein family would make them a versatile tool for the recognition of broad range of microbes to initiate distinct immune responses (Schulenburg, Hoepfner et al. 2008, Miltsch, Seeberger et al. 2014).

The overabundant proteins were enriched for two additional protein groups (Table 2.0.2; Figure 2.12): transmembrane glycoproteins, which have been reported to interact with lectins, and CUB-like domain proteins. Both of these were implicated in the *C. elegans* immune defense by transcriptional responses in various infection assays (Schulenburg, Hoepfner et al. 2008). However, both of these protein groups are poorly characterized.

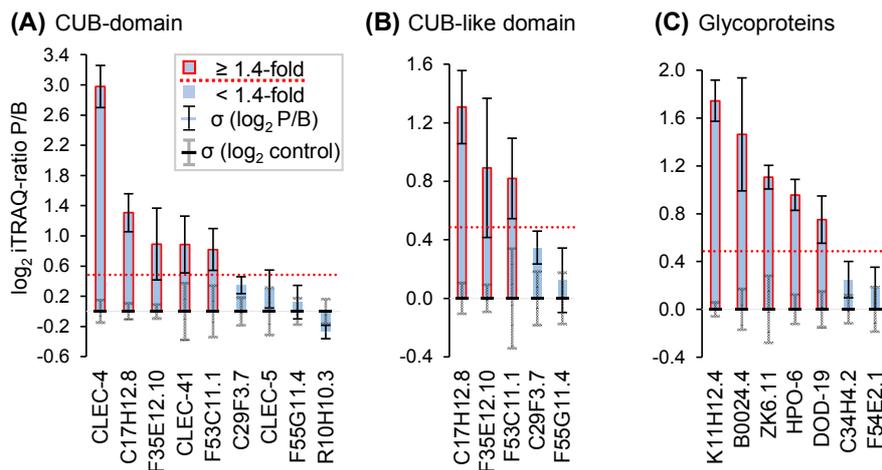


Figure 2.12: Relative abundance changes of potential immune effector proteins upon pathogen challenge. (A) CUB-domain proteins; (B) CUB-like domain proteins; (C) transmembrane glycoproteins of unknown function. Columns show the mean \log_2 protein-ratios in *C. elegans* grown with pathogenic vs. non-pathogenic *Bacillus thuringiensis* (P/B). Black error bars indicate geometric standard deviations of \log_2 P/B ratios, grey error bars show standard deviations of transformed control-ratios between biological replicates of the same condition (P/P; B/B and E/E) in the three iTRAQ experiments (around the default mean of \log_2 transformed ratios = 0) for the indicated proteins. The dotted lines indicates the ± 1.4 fold-change cut-off values.

Lysozymes were the second protein family to show a distinct pattern of differential abundance between the monocultures of *E. coli* and the culture including the nematocidal spore-crystals mixture (Figure 2.13 A).

Lysozymes catalyze the cleavage of peptidoglycan, a major component of bacterial cell walls. The disruption of bacterial cells is considered as an important function employed in both defense against infectious bacteria and in the digestion of bacteria as a nutrient source (Leippe 1999, Schulenburg and Boehnisch 2008). *C. elegans* potentially synthesizes 15 lysozymes belonging to the two different classes, namely of the invertebrate-type and the protist-type. A transcript level analysis of the protist-type lysozyme genes in *C. elegans* after *Bt* infection has previously been performed (Boehnisch, Wong et al. 2011). The four protist-type lysozymes quantified here show changes in agreement with qPCR results of the previous study, with LYS-1 (1.9 fold) and LYS-2 (2.2 fold) strongly increased, LYS-8 (1.2 fold) slightly increased and LYS-4 (0.6 fold) strongly decreased.

Additionally, one of the invertebrate type lysozymes, ILYS-5 was more than two-fold lower abundant in pathogenic *Bt* challenged worms compared to worms grown on the non-pathogenic strain or *E. coli*. The lysozymes showing the most extreme ratios in this study, LYS-1, LYS-2, and ILYS-5, are assumed to be primarily active in the intestine (Mallo, Kurz et al. 2002, McGhee, Sleumer et al. 2007), while the tissue localization for LYS-4 and LYS-8 is less clear.

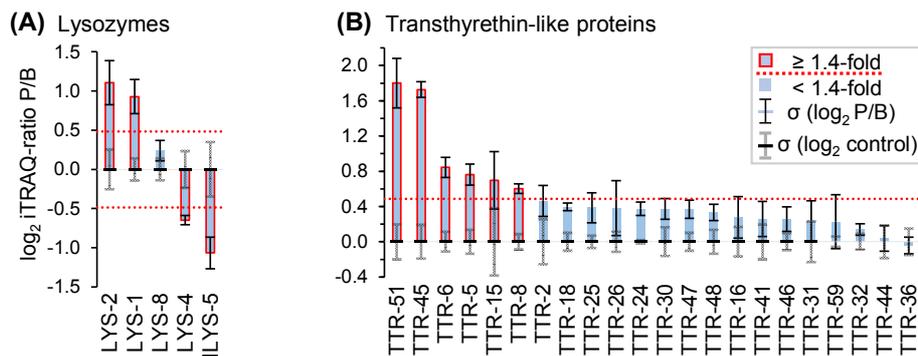


Figure 2.13: Relative abundance changes of selected *C. elegans* protein families upon challenge with pathogenic *B. thuringiensis*. (A) Lysozymes; (B) Transthyretin-like proteins. Columns show the mean \log_2 protein-ratios. Geometric standard deviations (σ) of \log_2 P/B ratios are indicated by black error bars. Grey error bars show standard deviations of transformed control-ratios between biological replicates of the same condition (P/P; B/B and E/E) in the three iTRAQ experiments (around the default mean of \log_2 transformed ratios = 0) for the indicated proteins. The dotted lines indicates the ± 1.4 fold-change cut-off values.

A series of transthyretin-like proteins was identified, of which six were found to be significantly increased in abundance in *C. elegans* upon *Bt* challenge (Figure 2.13 B). This elusive protein group has not been implicated in the response to infection up to now and the function of transthyretin-like proteins is largely unknown. One exception to this is the protein TTR-52 which was characterized as a bridging factor involved in cell corpse engulfment and apoptosis (Wang, Li et al. 2010), and in addition was shown to bind phosphatidylserine (Kang, Zhao et al. 2012). Phosphatidylserine is an important factor in membrane recruitment and signaling function of phosphoinositide-dependent kinase-1, which is a protein with a central role in phosphoinositol-3-kinase signaling pathways (Lucas and Cho 2011). On these grounds one may speculate that the six overabundant TTR proteins are part of the damage control mechanisms in compromised enterocytes.

Recently, compelling evidence has been presented for the involvement of the linker histone HIS-24 in silencing of immune response genes, and possibly as a direct effector in the response to *Bt* (Studencka, Konzer et al. 2012). In addition to the increased abundance of HIS-24 which substantiates these previous findings, significant increases in the abundances of a second histone linker protein HIL-3, a histone H2B (HIS-52), two H3 proteins (HIS-45 and HIS-72) and F40F12.7,

a zinc finger protein, possibly having a histone acetyltransferase-activity were found. Remarkably, two other quantified Histone H1 members (HIL-4 and HIL-5) did not follow this trend, which reduces the possibility of a systematic experimental bias. Nuclear proteins of higher abundance further included PME-5 an orthologue of human tankyrase which is induced in response to DNA damage by ionized radiation (White, Gagnon et al. 2009), and negatively regulates programmed cell death (Stergiou, Eberhard et al. 2011).

Major physiological changes occurring in infected worms are reflected in the higher abundance of a range of structural proteins. This included proteins specifically located in the cytoplasm of pharyngeal or intestinal cells (IFA-1, IFA-2 and IFA-4), along with one localized to the terminal web, the intermediate-filament rich layer of the cytoplasm below the apical surface of epithelial microvilli (IFB-2) (Bossinger, Fukushige et al. 2004), and one located in intestinal cell cytoplasm (IFD-2).

Proteins capable of forming microtubules, such as the tubulin α -subunit (TBA-7) and β -tubulin (MEC-7), were also augmented. Actin-based microfilaments failed to show increases in abundance above the applied threshold, however, several proteins required for their assembly and organization, namely α - and β - subunits of the F-actin end capping protein CAP-2 as well as TTH-1 were significantly increased. The higher abundance of cytoskeletal proteins may be directly linked to the onset of cell-death as the phenomenon of cytoskeletal rearrangement is a well-documented occurrence in cells undergoing apoptosis (Elmore 2007).

The role of proteins involved in autophagy, apoptosis, and necrosis in the course of bacterial infections of *C. elegans* is not obvious. Induced cell death by apoptosis or necrotic events have been proposed as a virulence strategy of pathogenic bacteria (Wong, Bazopoulou et al. 2007, Engelmann and Pujol 2010), while intestinal autophagy was shown to confer higher resistance to *S. typhimurium* infection (Jia, Thomas et al. 2009, Curt, Zhang et al. 2014).

The protein SCRM-4, which is known to be involved in the externalization of phosphatidylserine during apoptosis or cell injury (Wang, Wang et al. 2007), was increased in abundance, as were annexin (NEX-1), a protein required for Ca^{2+} dependent phospholipid binding and engulfment of apoptotic cell corpses in the pharynx (Arur, Uche et al. 2003), and the protein T28F4.5, a predicted homolog of death associated protein DAP-1 (Aravind, Dixit et al. 2001). The higher level of LGG-1, which is localized to autophagic vesicles that are linked to TOR-associated autophagic events (Hansen, Chandra et al. 2008), and TAX-6 a pro-autophagic related protein (Kang and Avery 2010), suggest the involvement of autophagy in response to pathogenic *Bt*.

Although several proteins involved in the major signaling pathways of the *C. elegans* immune system were identified, the majority of these proteins did not show significant differences between

treatments. This may indicate that signal transduction is regulated predominately on the basis of protein modifications rather than on changes in protein abundance. Exceptions were proteins linked to the mitogen-activated protein kinase (MAPK) pathways. Although PMK-1 itself did not show significant ratios, the proteins C17H12.8 and F56D6.2, which depend on PMK-1 and the G-protein coupled receptor FSHR-1 for their full induction (Powell, Kim et al. 2009), were increased more than two-fold. Along with this, the heavy metal and starvation induced MAPK, MEK-1 and Tag-196 which is predicted to interact with Pmk-1 (Zhong and Sternberg 2006) were increased in abundance upon infection.

The abundance of several proteolytic enzymes were affected by the presence of the pathogenic strain; furthermore, particular regulators of proteolysis were significantly enriched in the higher abundant proteins ($q < 1.4E-2$; $EF = 6.3$; $\text{mean log}_2(P/B) = 0.29$). Some proteases, such as the prolylcarboxypeptidase PCP-1 and the cysteine peptidases CPR-5 and F57F5.1, were of lower abundance upon challenge. GO annotation also implicates a possible role of F57F5.1 in receptor-mediated endocytosis. In contrast, the intestinal expressed cysteine peptidases, T28H10.3 along with cathepsin-B like cysteine proteases CPR-4 and the bifunctional Tag-196 cysteine protease and protease inhibitor known to interact with PMK-1, were all significantly increased in abundance. Furthermore, a second cystatin (inhibitor of cysteine proteases), CPI-2 was found at increased abundance after challenge with pathogenic *Bt*. Along with two downregulated serine carboxypeptidases F32A5.3 and F13D12.6, two highly abundant serpine (serine protease inhibitors) SRP-7 and SRP-6 were of particular interest. The first is associated with unfolded protein response and the second was previously described as a regulator of lysosomal-dependent necrotic cell death (Luke, Pak et al. 2007).

Aspartate peptidases are known to provide the major proteolytic function in the gut of *C. elegans* (Geier, Banaj et al. 1999) and have also been linked to neurodegenerative processes in the worm. In total, twelve members of this group were detected and three showed a decrease in abundance following pathogen exposure (ASP-14, ASP-13, ASP-12). The products of *asp-13* and *asp-14* showed the strongest change and both proteases are almost exclusively expressed in the intestine (McGhee, Sleumer et al. 2007). The lower levels of these proteases suggests that the worms digestive capabilities are impaired, possibly due to damage of enterocytes. This would also be an explanation for the increased level of MEK-1, as this kinase is reportedly involved in the intestinal and pharyngeal stress response, in particular to starvation and metal toxins (Koga, Zwaal et al. 2000).

2.5 Conclusion

The availability of defined biological materials and the application of sample processing steps that minimize the introduction of artifacts are major prerequisites for quantitative proteomics. When working with *C. elegans* it is important to start with synchronized cultures in which the developmental stage is homogenous. The previously established hypochlorite synchronizing procedure is effective for large scale sample preparation. However, the addition of sorting steps utilizing nylon net filters improved selective isolation of the target L4 population for a highly reproducible final sample. Reliable quantitative proteomic analysis further requires removal of contaminating bacteria, deceased worms, and other debris. For this purpose, the nylon net filter assisted sorting procedure proved to be gentler and less time consuming than the conventional sucrose floating method. Taken together, these modifications allow the dissection of *C. elegans* developmental stages while at the same time simplifying assay procedures and thereby improving the quality of large scale samples.

In order to achieve robust, statistically significant results in quantitative proteomics, it is also indispensable to analyze multiple biological and technical replicates. As a consequence, the overall instrument times increase significantly. Multiplex labeling schemes for quantitative proteomics address both of these issues. The use of iTRAQ 4-plex allowed robust validation of the data. The applied labeling scheme enabled the generation of control-ratios to account for biological and technical variation that allow both inter- and intra-experimental validation.

The use of an efficient two-dimensional separation scheme significantly contributed to the success of the quantitative study. Coupling two high-resolution reversed-phase liquid chromatography separations resulted in a highly separation of peptides and efficient use of the chromatographic separation space.

Conclusively, this study provides a robust workflow for the analysis of *C. elegans* host-pathogen interactions. Specific needs of proteomic experiments have been taken into account, addressing major issues, such as sample size, culture cleanup and the proper selection of developmental stages of the worm. By an advanced high resolution 2D-LC MS workflow, robust relative quantitative data of more than 3,500 proteins was acquired. With two control groups, the responses to contact with pathogenic relative to non-pathogenic gram-positive bacteria or to standard laboratory cultures could be clearly distinguish. Using very stringent criteria, 288 differentially abundant proteins were defined. Besides carbohydrate binding lectins and lysozymes, significant differences were observed in cytoskeletal proteins. Changes upon infections included a dysregulation of the balance between proteolytic enzymes and protease

inhibitors, including a significant decrease of serine- and aspartate-type proteases while cysteine-type proteases showed positive and negative changes in abundances. Evidence for the involvement of the transthyretin-like protein family in response to infection or the induced damage was found, which share a possible link to the onset of cell death in enterocytes, a dramatic rearrangement of cytoskeletal proteins and a significant up-regulation of proteins involved in chromatin organization.

The effects on *C. elegans* upon exposure to *Bt* spore crystal mixture are primarily attributed to the cry-toxins. However the germination of *Bt* in the changed milieu. Secondary effects of the damage caused to intestines which may include starvation and dehydration. The timeline of spore germination and effect of vegetative cells in compromised intestinal environment.

In conclusion, these findings show the system-wide effects of *Bt* intoxication on *C. elegans*, expanding the catalogue of immune effectors potentially acting towards the pathogen, and providing verification for numerous gene products implicated in previous transcriptomic studies. Collectively, this study presents evidence in support of both a general defense response as well as a specific reaction against *Bt* within the nematode *C. elegans*. It may contribute to a deeper understanding of the host-microbe interaction in the natural environment of the nematode.

3. Differential quantitative proteome analysis of *Escherichia coli* grown on acetate vs. glucose

3.1 Abstract

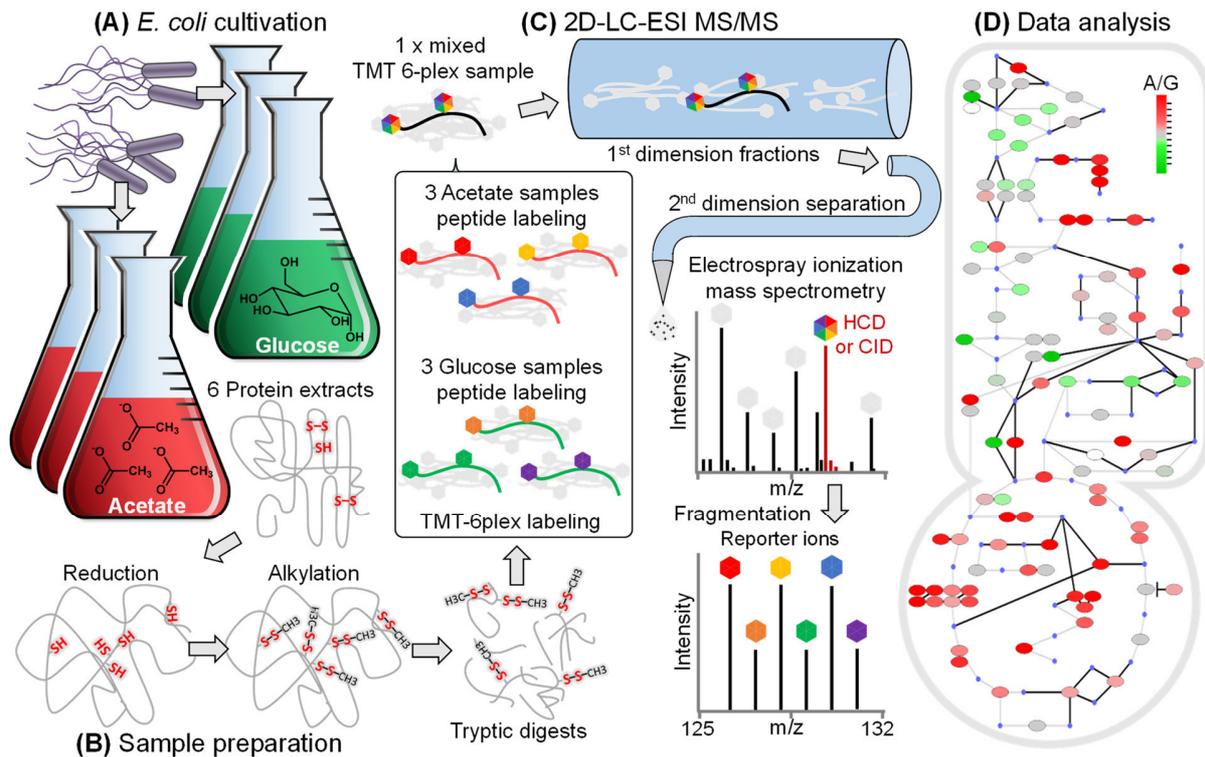


Figure 3.1: Graphical abstract. (A) *E. coli* was cultivated on acetate or glucose as carbon source. Both conditions were analyzed in three biological replicates. (B) Full proteome extracts were digested and peptides were labeled with TMT 6-plex isobaric tags. (C) The combined TMT labeled sample was analyzed in a single 2D-LC-ESI MS/MS experiment. (D) Acquired relative quantitative data were integrated with other system-wide datasets and metabolic maps of *E. coli*.

Relative protein abundances of *E. coli* strain K-12 MG1655, in exponential growth phase on minimal medium with either acetate or glucose as the sole carbon source, were determined by applying a quantitative shotgun approach with TMT 6-plex isobaric labels. With a focus on proteome coverage, peptides were separated using the high-resolution high/low pH 2D-LC method that was previously established with an optimized fraction-pooling scheme (Chapter 2). Mass spectrometric analysis was performed in four replicates, and exclusion lists were employed

to improve the coverage and acquire a high-quality dataset. In addition, to improve the accuracy, quantitative data were corrected for underestimation and ratio compression.

Using this workflow, robust relative quantification data was acquired for 2099 proteins, covering 49% of the predicted *E. coli* proteome. This analysis revealed system wide effects, due to growth on either glucose or acetate, at the protein level, with a total of 507 proteins showing a fold change of at least 1.5 and 205 proteins changing more than 2-fold.

The majority of proteins involved in central carbon metabolism showed significant abundance changes. In particular, proteins of the TCA cycle and glyoxylate shunt, as well as those involved in fatty acid metabolism and glycolysis/glycogenesis pathways.

Proteins involved in amino acid- and protein-synthesis were decrease in abundance, while proteins involved in the processing of environmental information were increased. For example, periplasmic binding proteins were more abundant on acetate, particularly those involved in scavenging specific resources (e.g. sugars).

The comprehensive nature of the data further allowed for comparative analysis against previously published transcriptomic and proteome datasets. This highlighted the differences and similarities between glucose and acetate metabolism at the transcript and proteome level, as well as providing insight into the relative protein abundances in different strains of *E. coli*.

Integration of genomic sequence information and transcript level data with the results of this proteome analysis also allowed an evaluation of the resource allocation in the process of gene expression between the different growth conditions. Evidence of a codon usage bias on the genes encoding differentially abundant proteins was identified, and in addition, transcriptome data showed that tRNA levels are also adjusted in accordance to this bias.

3.2 Introduction

The gram-negative bacterium *Escherichia coli* is one of the most widely studied model organisms. Well curated information for more than 130 *E. coli* genomes, in addition to functional and locational annotations of the encoded proteins have been collected over decades. Consequently, extensive metabolic reconstructions supported by high throughput data from hundreds of system-wide RNA expression studies are available. Most importantly, these data are freely accessible on scientific databases such as EcoCyc (Keseler, Mackie et al. 2013), PortEco (Hu, Sherlock et al. 2014) or Colibri (Médigue, Viari et al. 1993), to name just a few.

Due to the comparatively low complexity of the *E. coli* proteome, which is estimated to encompass 4,305 proteins distributed over only four well-defined cellular compartments, the bacterium has become a preferred biological sample for proteomics method development. This allows comprehensive and reproducible qualitative and quantitative proteome analysis of *E. coli* to be performed with less complicated analytical setups. Conversely, the dynamic range of the proteome still allows the comparative evaluation of methodological developments and experimental designs. In addition, reproducible samples of the bacterium can be prepared easily. *E. coli* protein extracts have helped in the development and validation of numerous proteomic technologies. These include MS instruments, MS acquisition methods, gel- and LC-based separation techniques and sample prefractionation methods, (Han and Lee 2006).

The highly efficiency transformation of competent *E. coli* and the rapid growth and ability to produce proteins at high levels made the bacterium also of immense importance for biotechnological and biomedical research (Huang, Lin et al. 2012, Chen, Zhou et al. 2013). Knowledge of the influence of environmental factors, e.g. nutrients and other growth conditions is paramount to understand the regulation of the organism's behavior at the molecular level.

The growth of *E. coli* on either acetate or glucose as the sole carbon source results in two distinct metabolic states. The classic substrate for biotechnological applications, glucose, is a high yield carbon source that promotes fast growth and is primarily metabolized through the glycolytic and TCA pathways. Conversely, acetate is a byproduct of glucose-based growth and a less efficient carbon source. The utilization of acetate requires gluconeogenesis (the antagonist pathway to glycolysis) and the redirection of the TCA cycle through the glyoxylate shunt.

Two distinct pathways of acetate utilization are known in *E. coli*. In the Pta-Ack pathway, acetate can be converted to acetyl-CoA by successive action of constitutively expressed acetate kinase and phosphotransacetylase (Kim, Ahn et al. 2015). The second pathway involves the activation to acetyl-CoA by acetyl-CoA synthetase (Acs), which is then metabolized through the TCA cycle and

glyoxylate shunt to produce energy and serve as a carbon source (Clark and Cronan 2005). The *Acs* gene is inducible and the enzyme is primarily utilized to scavenge extracellular acetate in the absence of other substrates (Kumari, Beatty et al. 2000).

Acetate overflow from glucose metabolism severely limits the growth rate and final culture density and consequently has a negative effect on primary metabolite production, which is of biotechnological interest (Eiteman and Altman 2006). However, the ability of *E. coli* survive on acetate is important for the survival of the bacterium in the gastrointestinal tract of endotherms (Wei, Shin et al. 2000).

The metabolic organization of *E. coli* on either glucose or acetate has been investigated in a number of system wide studies. These include a fluxome analysis (Zhao and Shimizu 2003) in addition to metabolome, proteome and transcriptome analyses under various culture conditions.

Comprehensive transcriptome datasets are available for growth on acetate relative to growth on glucose on defined minimal medium (Oh, Rohlin et al. 2002, Liu, Durfee et al. 2005), as well as for the long- and short-term adaptive responses to acetate and propionate on nutrient rich complex medium (Polen, Rittmann et al. 2003). Metabolome and transcriptome data were used to describe the dynamics of the molecular processes involved in glucose-acetate transition (Enjalbert, Letisse et al. 2013).

E. coli strain W3110 grown on rich LBK medium (potassium-modified Luria broth) or M63 salt medium in the presence or absence of either acetate, were previously examined by gel based differential proteomics (Kirkpatrick, Maurer et al. 2001). By this method, the level of 37 proteins was discovered to increase in the presence of acetate and 17 proteins were found to be of lower abundance. With an emphasis on the analytical approach, tryptic peptides derived from the *E. coli* K-12 strain ATCC10798 grown on M9 media with either lactose, glucose or acetate were used as reference samples for LC-MS method development (Silva, Denny et al. 2006). Recently, a comprehensive label-free quantitative *E. coli* proteome dataset across 22 conditions, including minimal growth medium with glucose and acetate as carbon source, was published for *E. coli* strain BW25113, incidentally including data for the K-12 strain MG1655 on glucose but not on acetate (Schmidt, Kochanowski et al. 2015).

In the present study, a 2D-LC quantitative proteome analysis using isobaric labeling was performed to provide a reference dataset showing the protein level differences of *E. coli* MG1655 grown on acetate vs. glucose minimal medium with a focus on high proteome coverage.

3.3 Material and methods

3.3.1 Cultivation of *E. coli*

The cultivation of *E. coli* was performed by Dr. Brice Enjalbert and Dr. Fabien Letisse (LISBP, Université de Toulouse, CNRS, INRA, INSA; chair: Prof. Dr. Jean-Charles Portais).

Escherichia coli K-12 strain MG1655 was cultured to mid-exponential phase in triplicate in M9 medium (3.37 mM Na₂HPO₄, 2.2 mM KH₂PO₄, 0.855 mM NaCl, 0.935 mM NH₄Cl, 1mM MgSO₄, 0.3 mM CaCl₂, 1µg/l thiamin, 1µg/l biotin and 1 x trace elements solution). The 100 x trace elements stock solution contained 13.4 mM EDTA (ethylenediaminetetraacetic acid), 3.1 mM FeCl₃-6H₂O, 0.62 mM ZnCl₂, 76 µM CuCl₂-2H₂O, 42 µM CoCl₂-2H₂O, 162 µM H₃BO₃ and 8.1 µM MnCl₂-4H₂O. As a carbon source, M9 minimal medium was supplemented with either glucose (15 mM) or acetate (45 mM). Batch cultures of 50 mL were inoculated at an optical density at 600 nm (OD₆₀₀) of 0.1 on glucose and 0.2 on acetate. Bacteria were harvested at OD₆₀₀ of 1 for proteome analysis after 5 h of exponential growth on glucose ($\mu = 0.67 \pm 0.01$) or 7 h on acetate ($\mu = 0.24 \pm 0.01$). The specific growth rate μ was calculated according to Formula 3.1 (Monod 1949).

$$\text{Formula 3.1: } \mu = \frac{\ln(\text{OD}_2) - \ln(\text{OD}_1)}{(t_2 - t_1)}$$

The culture medium was removed by centrifugation (7,000 g, 3 min, 4 °C), replaced with 200 µL of 0.1 M TEAB (triethylammonium bicarbonate; 4 °C) and samples were immediately frozen in liquid nitrogen and stored at -80 °C until further use.

3.3.2 Sample preparation

E. coli were suspended in protein extraction buffer, adding stock solutions to achieve final concentrations of 0.1 M TEAB, 2 mM TCEP (tris 2-carboxyethyl phosphine), 2 mM EDTA, 0.5 % CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), and 1 x cComplete protease inhibitor (Roche; <https://lifescience.roche.com>) in a final volume of 400 µL. Acid washed glass beads (500 µL, Ø 0.1 - 0.25) were added and bacteria were disrupted in ten cycles of freezing (90 s, -80 °C ethanol-bath), slightly thawing in a sonication bath (30 s) followed by 60 s disruption at 30 Hz in a mixer mill (MM 400, Retsch; <http://www.retsch.de>). Debris and glass beads were removed by centrifugation (1,000 g, 10 min, 4 °C).

The protein concentrations of the supernatants were determined using the BCA protein assay (Thermo Fisher Scientific; <https://www.thermofisher.com>) and aliquots of 120 µg total protein were

cleaned by methanol-chloroform precipitation. The precipitate was dissolved in 20 μL , 0.4 % RapiGest (Waters; <http://www.waters.com>) in 100 mM TEAB-buffer, heated for 5 min at 99 °C and reduced and alkylated as described in the TMT 6-plex protocol (Thermo Fisher Scientific). The acid cleavable detergent RapiGest was diluted to 0.2 % using 100 mM TEAB. Digestion was performed at a protease to protein ratio of 1:40 with sequencing grade modified trypsin (Promega; <https://www.promega.de>).

After digestion, 40 μL aliquots were labeled using the TMT 6-plex reagents following the manufacturers' protocol. The six samples, each labeled with one TMT reagent, were combined and incubated with 0.2 % formic acid to remove the acid cleavable detergent. The sample was centrifuged (55,000 g , 1 h, 4 °C) and the supernatant was transferred to a new 1.5 mL LoBind reaction tubing (Eppendorf; <https://www.eppendorf.com>), lyophilized to dryness and resuspended in 60 μL of eluent A for C_{18} reversed-phase- HPLC separation at pH 10.

3.3.3 Peptide fractionation by high pH reversed-phase HPLC

Labeled peptides were separated by two-dimensional liquid chromatography as described earlier (Delmotte, Lasaosa et al. 2007). The first dimension of separation was performed on a Dionex Ultimate 3000 HPLC system (Thermo Fisher Scientific), equipped with an analytical Gemini- C_{18} column (150 mm x 3 mm, 3 μm ; Phenomenex; <http://www.phenomenex.com>), over a gradient of eluents A (milli-Q water, 72 mM triethylamine adjusted to pH 10 with 3.5 mL acetic acid) and B (acetonitrile, 72 mM triethylamine, 3.5 mL acetic acid) at a flow rate of 300 $\mu\text{L}/\text{min}$. The elution was recorded on a UV detector at 254 nm and 280 nm.

After injection, the sample was washed with 5 % eluent B for 10 min and then fractionated over a gradient from 5 % to 55 % B in 50 min, collecting fifty 300 μL fractions. The fractions were lyophilized to dryness and redissolved in 30 μL of 0.1 % aqueous trifluoroacetic acid (TFA). To maximize usage of the chromatographic space, and minimize the measurement time in the second dimension, a fraction pooling scheme was used (Figure 3.2). By this procedure the 50 first dimensional fractions are reduced to 10 pooled fractions, each with a volume of 150 μL .

3.3.4 LC-MS analysis of pooled fractions

For the second dimension separation and on-line mass spectrometric analysis (LC-MS), a Dionex Ultimate 3000 HPLC system, equipped with a Dionex Acclaim PepMap100 nano-column (75 μm

x 15 cm, 3 μ m, 100 \AA ; Thermo Fisher Scientific), was coupled to an LTQ Orbitrap velos mass spectrometer (Thermo Scientific). IP-RP HPLC separation was performed with a gradient of Eluents A (Milli-Q water with 0.1 % formic acid) and B (80 % acetonitrile, 0.1 % FA). From each of the ten pooled fractions, 6 μ L were injected for LC-MS analysis. The labeled peptides were washed for 10 min on a Dionex PepMap C₁₈ guard column (300 μ m x 5 mm, 5 μ m; Thermo Fisher Scientific) with 0.1 % aqueous TFA at a flow rate of 30 μ L/min. Peptides were then separated over a gradient from 5 % to 60 % eluent B in 180 min at a flow rate of 250 nl/min, followed by a sharp increase to 95 % in 10 min and a 5 min washing step. Afterwards the column was equilibrated with 5 % eluent B for 15 min. UV detection was performed at 214 nm.

The HPLC was coupled to an LTQ Orbitrap velos mass spectrometer (Thermo Fisher Scientific) using a nanospray ion-source with 1.3 kV spray voltage and 197 °C capillary temperature with a 30 μ m PicoTip emitter (New Objective; <http://www.newobjective.com>). After a delay time of 35 min, full scans of 300 m/z – 2,000 m/z range were recorded in profile mode with a resolution of 30,000 for 155 min (AGC target 1E6; maximum inject time 500 ms, preview mode for FTMS master scans and Wideband activation were used). Collision induced dissociation spectra were acquired with the ion trap mass analyzer (CID-IT). Higher energy collision induced dissociation spectra were recorded using the Orbitrap Fourier transform mass analyzer (HCD-FT). The five most intense precursors (minimum signal intensity 500 and rejecting charge state 1) were selected for both CID and HCD with a repeat count of 2, a repeat duration of 20 s, and subsequent dynamic exclusion of a 5 ppm mass window around the selected m/z values for 60 s. The CID isolation window was set to 3 Da, the AGC target was 1E4 with a maximum inject time of 400 ms, activation time of 10 ms and an activation Q of 0.25. The normalized collision energy for CID fragmentation was set to 36. The isolation window for HCD precursors was set to 3 Da, AGC target was 1E5 with a maximum inject time of 500 ms, and activation time of 0.1 ms. HCD spectra were acquired with a resolution of 15,000 and normalized collision energy of 50 was used for fragmentation by HCD. All MS/MS spectra were recorded in centroid mode.

While each of the 10 pooled fractions from first dimension HPLC separation were analyzed at least four times, additional runs of two samples were initially used to optimize acquisition parameters. The final dataset contained 42 LC-MS/MS raw files. Raw files and database search result-files can be accessed on the ProteomeXchange Consortium PRIDE partner repository via the dataset identifier PXD003863 (<http://proteomecentral.proteomexchange.org>).

The first two injections were technical replicates with unchanged acquisition parameters. These LC-MS runs were processed and the results of the database search (see Chapter 4.3.5) were used to compile an exclusion list of m/z values representing the peptides identified with high

confidence ($FDR \leq 0.01$). To reduce redundant scan events and increase sample coverage, corresponding m/z values were excluded from data dependent MS/MS acquisition over a retention-time window of 3 min and in a m/z range of 0.5 Da, in the third and fourth LC-MS run of each sample.

3.3.5 Data analysis

After recalibration using the polysiloxane contaminant peak at 445.12003 m/z , raw-files were subjected to database search and protein-quantification using the Proteome Discoverer Software (v1.4; Thermo Scientific). *Escherichia coli* K-12 reference proteome dataset from UniProt (May 2013; 4305 entries; <http://www.uniprot.org>) was supplemented with common contaminants (cRAP list of laboratory and contact dust contaminants; <http://www.thegpm.org/crap>) and used to search spectra with Mascot (v2.1), SequestHT and MS Amanda search algorithms. For HCD-FT and CID-IT spectra, fragment mass accuracy was set to 0.05 and 0.4 Da respectively. Parent-mass accuracy was set to 7 ppm. One search was performed with semi-trypsin specified as protease, allowing two missed cleavages. Methylthiolation of cysteine and TMT 6-plex at the amino-terminus and lysine residues (N-term and K) were set as static modifications. Oxidation of methionine and TMT 6-plex modified tyrosine were allowed as variable modifications.

To identify possible post-translational modifications, a multi pass search was performed, allowing phosphorylation of serine, threonine and tyrosine as variable modifications. In this case, the protease specificity was set to full trypsin to keep the search space more comparable. All other parameters were the same as in the first database search described above. To search for acetylation events, a SequestHT search allowing oxidation of methionine as well as acetylation and TMT 6-plex modification at lysine was performed. The corresponding MS Amanda search allowed acetylation and TMT 6-plex modification at lysine, TMT 6-plex modification at the amino-terminus of peptides and acetylation at the protein amino-terminus as variable modifications. Finally, a Mascot search was performed allowing oxidation of methionine, acetylation and TMT 6-plex modification at lysine and amino-termini as variable modifications. The protease specificity of these searches were set to semi-trypsin, with methylthiolation of cysteine as static modification.

Percolator (v1.2) was used for posterior error calculation combining the results of the database searches with a false discovery rate restricted to 0.01 by q-value (Kall, Canterbury et al. 2007). Single peptide identifications were excluded and protein-groups were assigned using strict parsimony rules. Only unique peptides were used for quantification and the peptide-ratios were normalized by the median of peptide-spectrum-matches (PSM).

To minimize ratio-compression and underestimation, reporter ion signals of PSM were corrected for signal interference and filtered for signal interference and relative precursor intensity as described previously (Savitski, Sweetman et al. 2011). Signal to interference (S2I) and precursor to transmission (P2T) values were calculated using the multiplierz software script provided here:

<https://sourceforge.net/projects/multiplierz/files/PublishedScripts/>.

Briefly, the sum of reporter ion intensities for each PSM was normalized to one. Isolation interference for each channel was assumed to be reflected by the median of corresponding reporter ion intensities. The normalized median intensities for each channel were multiplied by (1-S2I) to calculate the corresponding interference of each TMT channel for each PSM. The result was subtracted from the normalized reporter ion intensities of the corresponding TMT channel in each PSM. Recalculation of the normalized reporter ion intensities resulted in negative values occurring at a frequency below 0.1 %, these values were replaced by the minimum positive value in each channel, respectively. A moderate quality filter was applied by using only spectra with S2I > 0.5 and P2T > 4 for relative quantification of peptide- and protein-groups.

The Perseus software was used for statistical analysis of the dataset (www.maxquant.org). TMT-ratios were \log_2 transformed and the TMT-ratios describing the conditions Acetate/Glucose (A/G) were tested against \log_2 control-ratios (ratios between biological replicates of the same treatment: A/A; G/G) using Welch's t-test. Resulting p -values were adjusted for multiple testing by permutation based FDR calculation. Gene ontology (GO) and KEGG annotations were assigned from UniProt (<http://www.uniprot.org>) and EcoCyc (Keseler, Mackie et al. 2013).

The Perseus software was also used to preform enrichment/depletion analyses of GO and KEGG (Kanehisa, Sato et al. 2016) categories by fisher exact test and 1D annotation enrichment. The results of enrichment tests are noted with FDR-corrected p -values (q-value) and enrichment factors (EF) (Cox and Mann 2012) (see Formula 3.2).

$$\text{Formula 3.2: } EF = \frac{\text{total \# proteins in database}}{\text{quantified proteins}} \times \frac{\text{total \# proteins in category}}{\text{quantified in category}}$$

The 2,099 quantified proteins were tested for enrichment/depletion of categories compared to the complete reference proteome set of *E. coli* used for the database search (4,305 entries). In three additional annotation enrichment analyses the set of (I) differentially, (II) lower and (III) more abundant proteins were tested in relation to all the quantified proteins.

3.3.6 Data visualization and integration with external data sources

The codon usage information for all *E. coli* proteins was retrieved from the Codon Usage Database (Nakamura, Gojobori et al. 2000) and the frequency of codons on genes encoding the proteins quantified was calculated accordingly.

Previously published transcriptome and a proteomics datasets comparing acetate and glucose based growth (Liu, Durfee et al. 2005, Schmidt, Kochanowski et al. 2015) were compared to the relative quantitative data acquired in this study. The prediction intervals (*PI*) for the forecast values \hat{y}_0 for each value of x were calculated according (Formulas 3.3), assuming a normal distribution of the data points around the regression line as described here:

<http://www.real-statistics.com/regression/confidence-and-prediction-intervals/>.

$$\textbf{Formula 3.3: } PI = \hat{y}_0 \pm N_{crit} \times SE$$

N_{crit} is the inverse of normal cumulative distribution with a specified critical confidence level. SE is the standard error calculated according to Formula 3.4.

$$\textbf{Formula 3.4: } SE = SE_{xy} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x_0 - \bar{x})^2}}$$

SE_{xy} is the standard error of the \hat{y}_0 value for each value x in the regression. The standard error is a measure of the amount of error in the prediction of y for an individual x_0 .

Metabolic pathways were manually constructed using the computer software Cytoscape v.3.2 (Su, Morris et al. 2014) based on data from the KEGG (Kanehisa, Sato et al. 2016) and EcoCyc (Keseler, Mackie et al. 2013). Nodes were used to represent gene expression on the transcript and protein level by continuous mapping of expression data. Node borders were colored according to relative protein quantification data acquired in this study. To illustrate transcript level gene expression, node color was adjusted according to a dataset previously published by Liu, Durfee et al. (2005).

3.4 Results and discussion

3.4.1 Differential cultivation

E. coli was cultivated in triplicate in M9 minimal medium containing either glucose or acetate. By providing 15 mM glucose compared to 45 mM of acetate, a proportional amount of carbon source were available under both conditions, to ensure appropriate physiological conditions for a comparative analysis at the protein level.

Batch cultures had to be inoculated at OD 0.1 on glucose and 0.2 on acetate, to reach OD 1 after 5 hours of exponential growth on glucose and after 7 hours on acetate. The calculated specific growth rates were $\mu = 0.67 \pm 0.01$ and $\mu = 0.24 \pm 0.01$ on glucose and acetate, respectively. Consequently, the generation time was 1.0 h on glucose and 2.89 h on acetate. An almost three times faster growth rate on glucose, shows the dramatic difference in energy- and carbon-yield from the two nutrients and indicate major differences in the physiology and metabolism of bacteria growing under both conditions.

3.4.2 2D-LC-MS analysis

For bottom-up relative quantitative proteome analysis by isobaric labeling, high resolution 2D-LC peptide separation and ESI-MS analysis were slightly modified from the previously established workflow (Treitz, Cassidy et al. 2015). TMT 6-plex was used for relative protein quantification of bacteria grown on acetate vs. glucose. The capacity for multiplexing of six channels enabled the analysis of all three biological replicates of protein extracts from *E. coli* cultures grown under the two conditions in one experiment. In addition, MS measurement time is considerably shortened and technical variability is reduced significantly as compared to label-free quantitative proteomics approaches. As established in the previous project (see Chapter 2), two high resolution separations over reversed-phase columns at different pH were combined to achieve an in depth coverage of the analyzed sample.

The first dimension of peptide separation was performed over a reverse phase column at pH 10. Peptides were separated into 50 fractions which were then pooled to ten samples for second dimension separation by reversed-phase chromatography at pH 2 and on-line MS analysis. Fraction concatenation was used as a pooling scheme to optimize the use of the second dimension chromatographic space (Wang, Yang et al. 2011, Stephanowitz, Lange et al. 2012).

The separation parameters of peptides change between reversed-phase separation at pH 10 and pH 2, however, the retention-factors of peptides are still similar. A shift in retention-times and a change in hydrophobicity of the analytes can be attributed to the inversion of the charge state of acidic and basic amino acid functional groups under the different pH conditions. However, the dominant interaction between the peptides and stationary phase are still hydrophobic in nature. In a standard separation with an acetonitrile/water binary solvent system, this would result in a narrow elution-profile of all the first-dimensional fractions in the second dimension of separation. For this reason, equidistant fractions spanning the widest possible first dimension chromatographic space were merged before subsequent RP-LC-ESI MS analysis at pH 2. As a result, MS acquisition time is further reduced, without severely decreasing peptide and proteome coverage because the chromatographic space in second dimension of separation is used more efficiently (Figure 3.2).

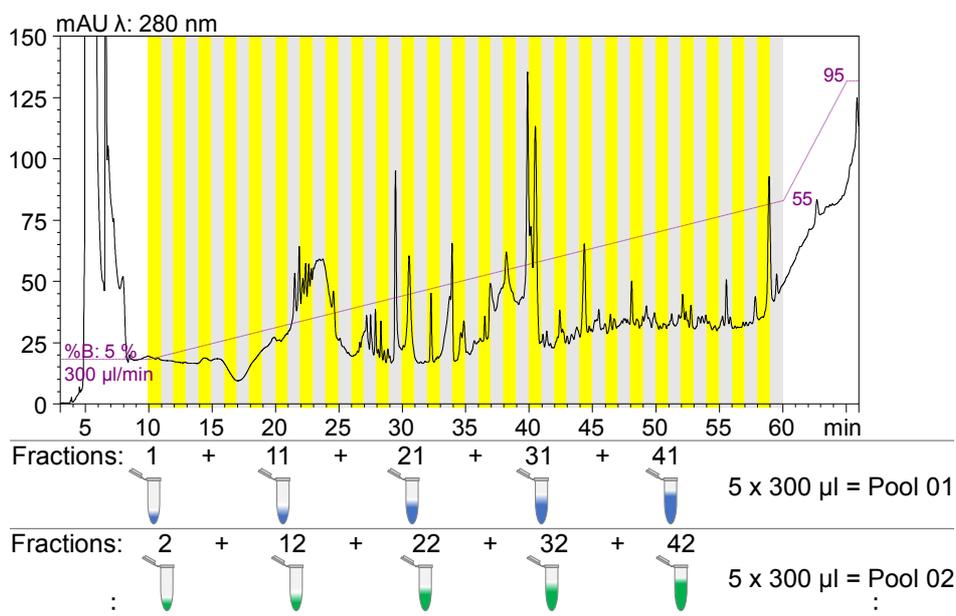


Figure 3.2: First dimension LC separation and sample concatenation scheme. The chromatogram shows the elution of peptides in the first dimension reversed-phase separation at pH 10. A UV detector was used to monitor the absorption at a λ of 280 nm. Flow rate and binary gradient of eluent A (water with 72 mM triethylamine adjusted to pH 10 with acetic acid) and eluent B (acetonitrile with 72 mM triethylamine, 0.3 % acetic acid) are indicated. The collected fractions are denoted by alternating yellow and grey background colors. The applied pooling scheme for the first and second merged fractions is displayed below.

Ten pooled first dimensional fractions were analyzed over 3 h gradient LC-MS runs in at least four replicates. The first two measurements were exact technical replicate runs. However, consecutive repeated injections of the same sample usually result in an exponential increase of redundant scan events. For in-depth analysis of the fractions and to reduce redundancy in technical replicate runs, m/z values corresponding to peptides already identified with high confidence from the in the

first two technical replicates were compiled into exclusion lists. These exclusion lists for each fraction were implemented on the third and fourth measurement.

The assessment of the distribution of redundant and unique peptide identifications over replicate LC-MS measurement of the pooled fractions shows the efficiency of the separation and MS-acquisition scheme. The average number of peptides detected across the four LC-MS analyses of the ten pooled fractions was 3,335 with a standard deviation of only 105 (3 %). This shows that pooling five concatenated first dimensional fractions results in ten very homogenous pooled fractions in regard to peptide content (Figure 3.3).

Searching the complete dataset, 19,105 different peptide sequences and 23,909 different peptide species (counting peptides with the same sequence of amino acids with different modifications as unique) were identified with an FDR below 1 %. Including also the different charge states of the same peptides, the number of identified unique m/z values sum up to 33,876. The number of “unique” peptide identifications increases by approximately 1 % if peptides with different modifications are taken into account and by 40 % if different charge states of the same peptide are considered as unique. Although these identifications can be regarded as redundant in terms of sequence and proteome coverage, the accumulation of 40 % more data represents an improvement for protein quantification, where these spectra are all taken into account. The relative standard deviation of the number of identifications between the ten fractions does not change and accounts for about 3 % of the total either way (Figure 3.3).

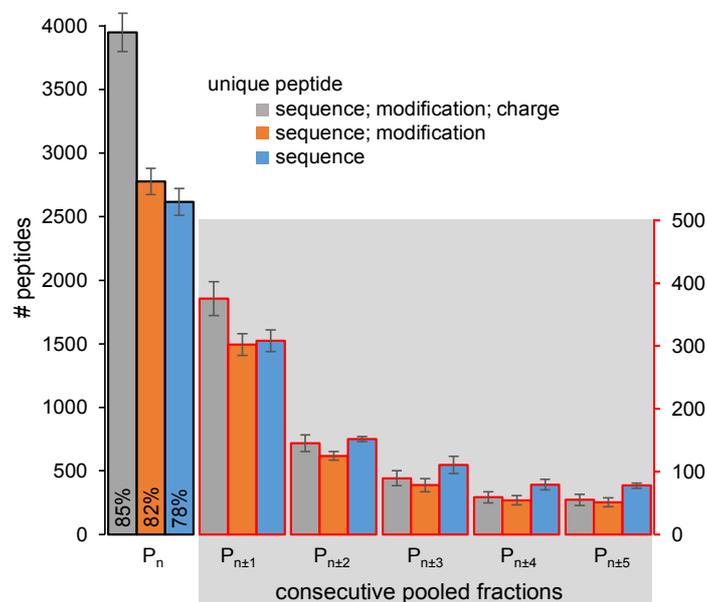


Figure 3.3: Assessment of the first dimension of LC separation. The average number of peptide identifications from the quadruplicate measurement of each of the ten pooled fractions (P_n) are indicated by columns. Error bars show standard deviations between the counts of peptide identifications of the pooled fractions. The number of peptide identifications shared between neighboring pooled LC-fractions ($P_{n±x}$) are shown on a separate axis to the right (outlined in red).

The percentage of peptides shared among adjacent fractions allows an assessment of the resolution of the first dimension separation and the applied pooling scheme. If all peptide species are taken into consideration (including different charge states and modifications), about 15 % were also identified in adjacent fractions. If only the peptides sequence is considered for uniqueness the number of peptides shared among the pooled fractions amounts to ca. 22 % (Figure 3.3). The 2D-LC MS analysis of *C. elegans* (Chapter 2) showed much higher resolution in the first dimension of separation. This discrepancy may be attributed to the use of a much longer analytical column for first dimension of separation in addition to the lower complexity of the *E. coli* sample.

The conventional analysis of peptides by LC-ESI MS/MS is called a data-dependent acquisition. In this mode, a survey MS spectrum is first acquired, to provide an overview of the analytes eluted at a given LC retention-time. This is followed by a fixed number of MS/MS scans of the most intense precursors detected in the survey MS spectrum. The retention-time window of eluting analytes and the sampling rate of ms instruments limit the number of analytes that can be acquired in a LC-ESI MS/MS run. Due to this, not all analytes within a sample of high complexity (e.g. full proteome digests), can be targeted for MS/MS acquisition in a single run. By performing duplicate LC-MS/MS analyses of the ten pooled fractions, a total of 14,963 and 15,181 peptides were identified (Figure 3.4).

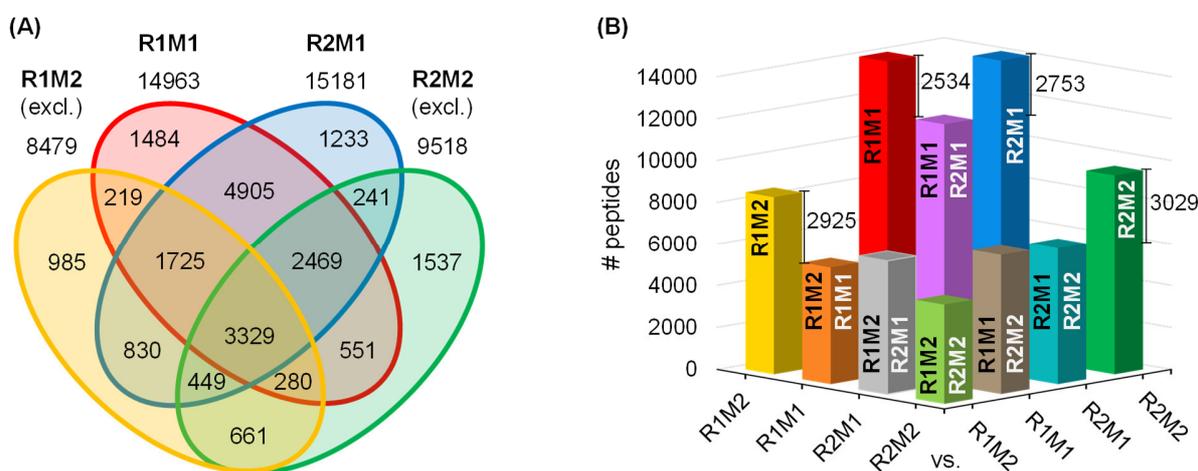


Figure 3.4: Distribution of peptide-spectrum-matches from replicate LC-MS measurements. (A) The Venn diagram shows how peptide identifications overlap between the four measurements summed up from the 10 pooled fractions. The technical replicate LC-MS/MS runs (R1 vs. R2) and repeated injections with or without employing exclusion lists (M1 vs. M2). **(B)** The total number of identified peptides in each measurement is illustrated by the yellow, red, blue and green columns. The six columns intersecting the respective category axes show the number of redundant peptide identifications shared between each of the four measurements (e.g. the purple column shows the 83 % overlap between R1M1 and R2M1). The exact number of unique peptides between true technical replicate runs and those employing exclusion lists from previous injections are noted to the right of the four columns.

The vast majority of acquired fragment spectra result from repeated activations of the same precursor, identifying the same peptides in both consecutive runs. These account for approximately 83 % (12,428) of the identified peptides. Although redundant identifications provide additional quantitative data, proteome coverage is not increased. Redundant scan events were reduced by employing exclusion lists of m/z values corresponding to high confident identifications from the first and second LC-ESI MS runs on the two subsequent measurements of the same pooled fraction. This procedure resulted in a reduced total number of identified peptides (approximately 60 % compared to those replicates without employing exclusion lists), however, the proportion of redundant peptide identifications decreased to only 67 % (compared to 83 % from the first two replicate runs).

Even after employing exclusion lists, a redundancy between runs of 63 %, in terms of unique peptide sequence identifications, is still relatively high. This can be explained due to the charge distribution of peptides in ESI MS. Different charge states of the same peptides are not covered by an exclusion list of unique m/z values. Nonetheless, unique identifications did not further decrease, as would be expected, with repeated measurements of the same sample, but increased slightly instead. This shows that employing exclusion lists for multiple injections of the same sample reduces redundant and increases complementary scan events and consequently increases proteome coverage as shown previously (Chen, Rejtar et al. 2005). Of all peptide identifications in the four replicates measurements of ten pooled fractions, only 25 % were non-redundant while 35 %, 24 % and 16 % were identified in two, three or all four runs, respectively.

3.4.3 Procedures to reduce ratio-compression and underestimation

One serious issue discovered with the isobaric labelling based relative quantification strategy is a general ratio-compression and underestimation of very high or low ratios. In essence, control experiments consistently showed that the calculated protein-ratios and the underlying peptide-ratios collected by MS analysis are usually significantly lower than the nominal ratios specified in the experimental setup (Bantscheff, Boesche et al. 2008, Ow, Salim et al. 2009, Karp, Huber et al. 2010). The more disproportionate the protein amounts are between the compared samples, the higher is the underestimation of the quantitative ratios. This results in a compression of the range of acquired protein- and peptide-ratios. The systematic error was shown to be increased with sample complexity and to be decreased with the efficiency of peptide separation (Ow, Salim et al. 2011, Wang, Alvarez et al. 2012). Ratio-compression and underestimation are

predominantly attributed to the presence of coeluting peptides within the isolation window used for selection and subsequent cofragmentation with the target peptide (Altelaar, Frese et al. 2013).

A premise of meaningful quantitative comparison of protein amounts in biological samples is that the majority of proteins in both samples are equal in abundance, while a small number of proteins are differentially abundant and functionally linked with the examined conditions. In ideal cases the average quantitative ratio is therefore one, and correspondingly, the range of logarithmically transformed data should be normally distributed around zero. Under this assumption, the reporter ion intensities of cofragmented peptides are likewise distributed and the signals shift all quantitative ratios of target peptides towards the mean.

This has a detrimental effect on the accuracy of isobaric label based quantification, particularly in complex samples, and results in a systematic underestimation of the actual difference between differentially abundant peptides. This issue is further aggravated because biologically relevant and highly interesting proteins, which are often low abundant, are highly sensitive to isolation interference (Sandberg, Branca et al. 2014). In contrast, many “housekeeping proteins” are frequently higher in abundance and contribute to interference. Corrective measures have been developed to reduce ratio-compression and underestimation. Among these are both optimizations of the LC-MS analysis method and post-acquisition data correction procedures. LC-MS method developments include the adjustment of the MS/MS isolation window and delayed precursor isolation to the apex of chromatographic peaks (Savitski, Sweetman et al. 2011). Different approaches involve the acquisition of MS/MS spectra at a reduce isolation width. However, these methods often represent a trade-off between the accuracy of quantification and the yield of identifications from a given analysis.

In this study, several analytical and post-acquisition procedures were applied to provide accurate quantitative data. As discussed above, two high resolution separations were combined to reduce the sample complexity as far as possible. Additionally, MS/MS spectra of precursors were acquired with a repeat count of two, prior to dynamic exclusion. This ensured the acquisition of MS/MS spectra near the apex of chromatographic peaks and further minimizes the interference of coeluting analytes. Of the 80,616 PSM from the complete analysis, 87.8 % were isolated at more than half-maximum intensity of the ion chromatogram (Figure 3.5).

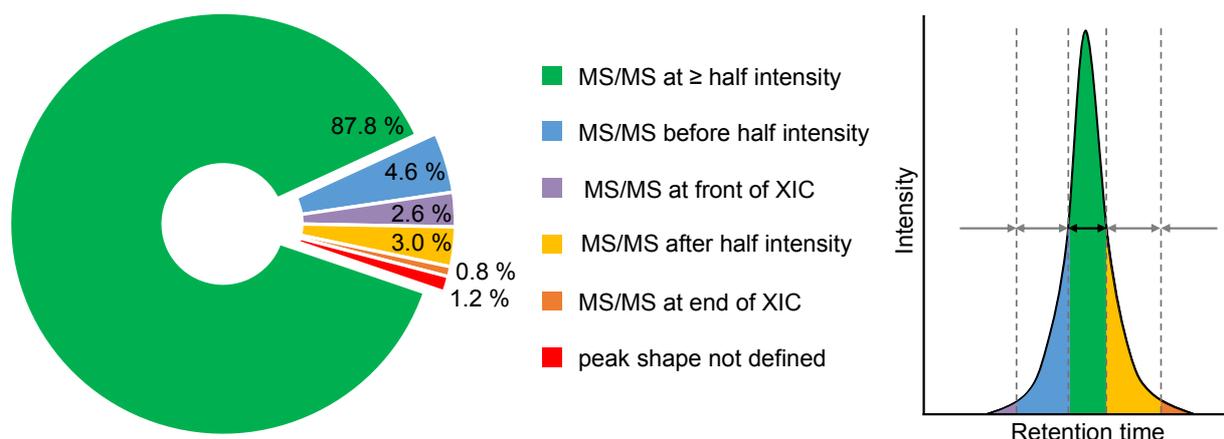


Figure 3.5: MS/MS acquisition at the apex of chromatographic peaks. The pie chart provides an overview of how MS/MS spectra were distributed over their corresponding extracted ion chromatograms (XIC). Isolation mostly occurred within the full width at half maximum (FWHM) of the chromatographic peak (green). More MS/MS spectra were acquired towards the front (7.2 %; blue and purple) than at the end (3.8 %; yellow and orange) of the eluting peak.

Optimization of the normalized collision energy parameters of MS/MS spectra on the LTQ Orbitrap velocity mass spectrometer was performed on additional LC-MS runs for two of the pooled fractions. In these LC-MS runs, normalized collision energies (NCE) of 45, 50 and 55 were tested. The lowest variability of TMT reporter-ions for quantification without deterioration of fragment ions used to identify peptides was achieved with an NCE value of 50. A NCE value of 50 was applied to all following pooled fractions and replicate measurements.

Post-acquisition, moderate filters for isolation interference and relative precursor intensity were applied to remove PSM of low quality quantitative data. To filter spectra with high isolation interference, a signal to interference value (S2I) was calculated (Savitski, Sweetman et al. 2011). From 80,616 PSM initially used for quantification, 10.9 % (8,765 PSM) were removed due to higher than 50 % isolation interference. Subtraction of noise from MS scans is an automatic post-processing step performed by the acquisition software. The relative precursor intensity filter applied on PSM refers to the signal intensity of the precursor compared to the noise threshold of the corresponding MS spectrum, calculated as a precursor to threshold value (P2T) (Savitski, Sweetman et al. 2011). Excluding fragment spectra from precursors with a P2T below four removed 4.5 % (3,656 PSM) of the identified spectra from quantification (Figure 3.6).

Combined, both filters reduced the number of PSM used for quantification by 13.5 % to 69,726. Peptide-groups decreased by 9.0 % to 21,746, while protein groups were reduced by 6.2 %, leaving a total of 2,309 proteins-groups (including single peptide identifications). The impact of filtered out PSM decreases with each identification level from PSM to protein-group. This reveals that the filtered out spectra not only display poor quantitative data but are also of low quality in

terms of the reliability of the underlying peptide- and protein-group identification. Likewise, at the protein-group level, filtered PSM show a noticeable higher impact on single peptide identifications (Figure 3.6).

Lastly, the reporter ion intensities of all remaining PSM were recalculated to account for isolation interference as described earlier (Savitski, Mathieson et al. 2013). Normalized reporter ion intensities for each TMT channel were reduced by the fraction attributed to the coisolated interference for each spectrum (Figure 3.7). Reprocessing of the data significantly increased the spread of \log_2 protein-ratios, decreasing ratio-compression.

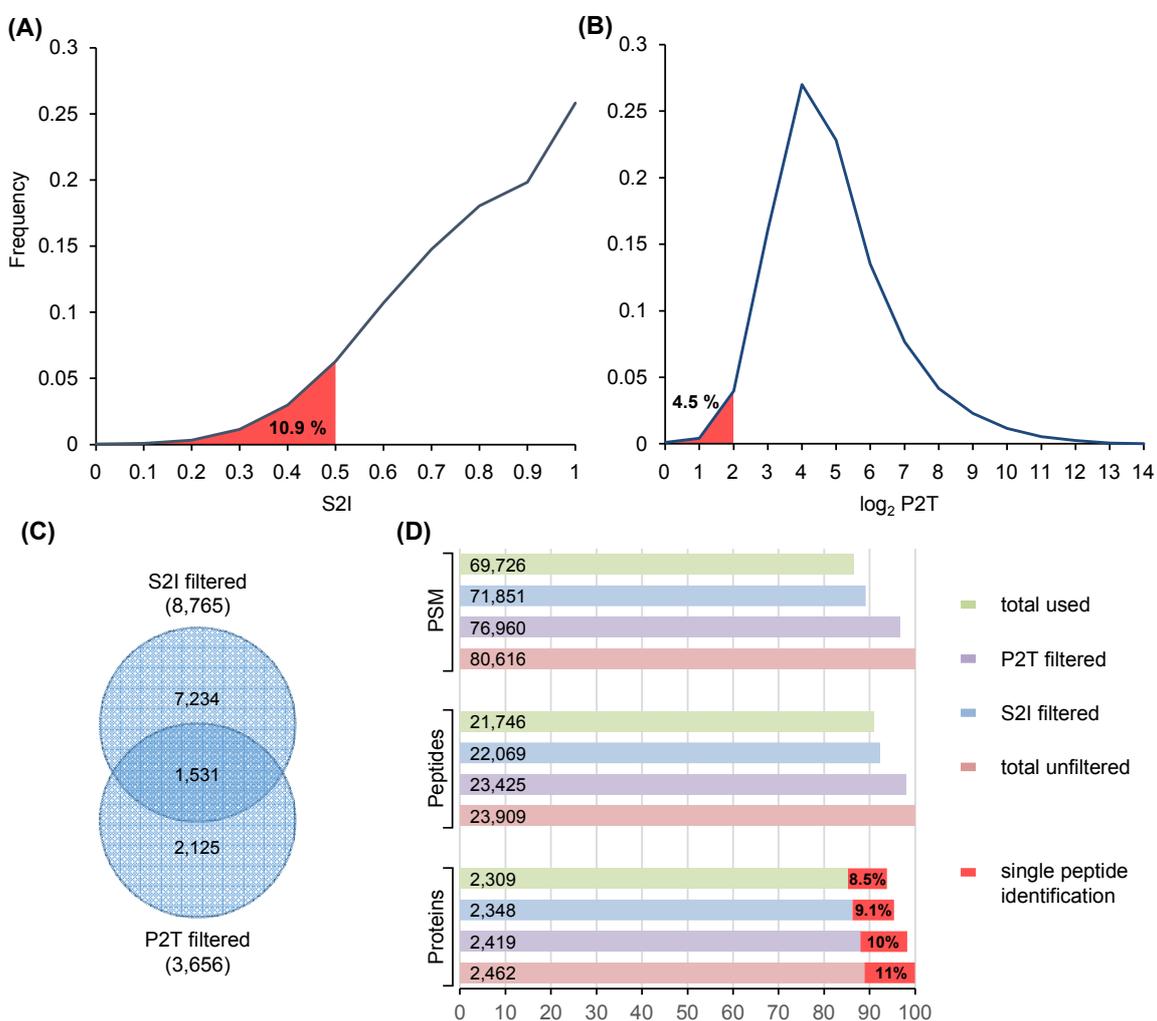


Figure 3.6: PSM filters applied to reduce isolation interference. Density curves show **(A)** the distribution of signal to interference ratios (S2I) and **(B)** \log_2 transformed precursor to transmission values observed for the 80,161 PSM. The area under the curve in red illustrate applied filters. **(C)** The Venn diagram illustrates the overlap of filtered out PSM. **(D)** The bars show the effect of the applied filters at the level of PSM, peptide-groups and protein-groups.

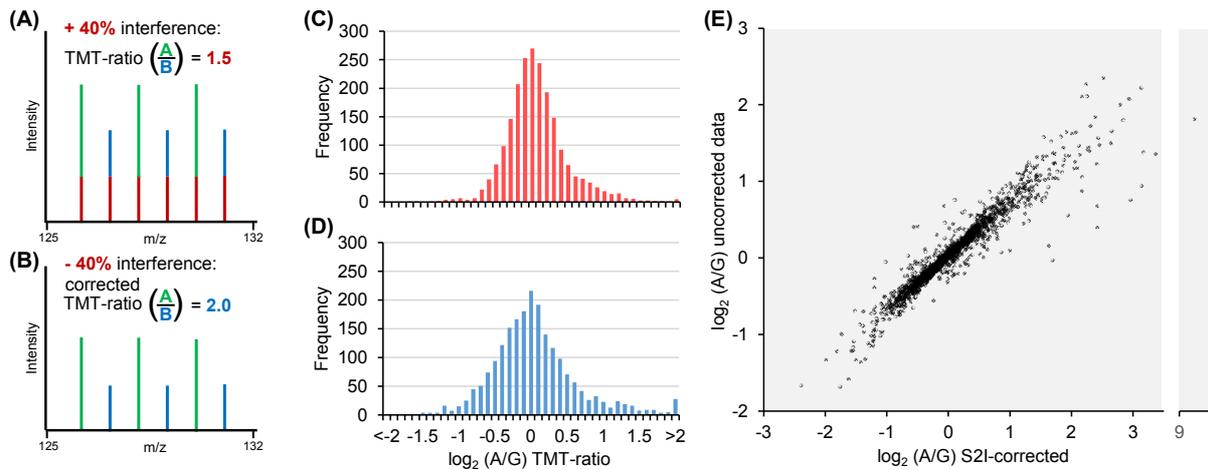


Figure 3.7: Reprocessing of TMT-ratios to reduce ratio compression. An example of the correction procedure is shown on the left. The schematic shows the reporter ion region of a theoretical MS/MS spectrum with six TMT reporter-ions representing two biological conditions (green and blue) **(A)** before and **(B)** after correction for isolation interference (red). The histograms show the spread of $\log_2(A/G)$ for 2099 quantified proteins **(C)** before and **(D)** after data correction. **(E)** The plot shows the correlation between data before and after S2I-correction.

3.4.4 Qualitative assessment of the dataset

The complete analysis consisted of 42 LC-MS runs adding up to 126 h of raw measurement time (excluding customary blank runs between samples and quality control samples). The collective dataset combined a total of 722,764 MS/MS spectra which could be matched to 23,909 peptide sequences with high confidence (FDR < 1 %) and identified 3,123 *E. coli* proteins. Proteins identified by a single peptide were excluded to reduce the possibility of false positive identifications, which left 2,681 proteins. Enforcing protein grouping by strictest parsimony principles further filtered the identifications to 2,590 *E. coli* protein-groups of which 2,246 were identified by multiple peptides (Figure 3.8 A).

Relative quantification was possible for 2,462 protein-groups using only proteotypic (unique) peptides, which were reduced to 2,309 by filtering peptide-spectrum-matches to minimize ratio-compression (Chapter 3.4.5). Also removing identification supported only by a single peptide left 2,099 protein groups. Collectively, 15 % of the protein-groups were excluded with these filter criteria and only on the remaining 2,099 protein-groups statistical analysis was performed. On average, these were identified by 11 peptides (median = 7). Quantitation of these protein-groups is based on an average of 10 unique peptides (median = 6) and about 50 % of the quantifications are supported by the quantitation data of at least five unique peptides (Figure 3.8 A).

Relative protein quantification is usually performed comparing samples of equal protein concentrations. This is ensured both by using equal amounts of proteins after measuring the protein concentration of the different samples and by median normalization of the produced data. Although these procedures are necessary for data interpretation and statistical analysis, systematic errors may be introduced because structural differences between the compared biological materials, e.g. the size of cells or organelles were not considered.

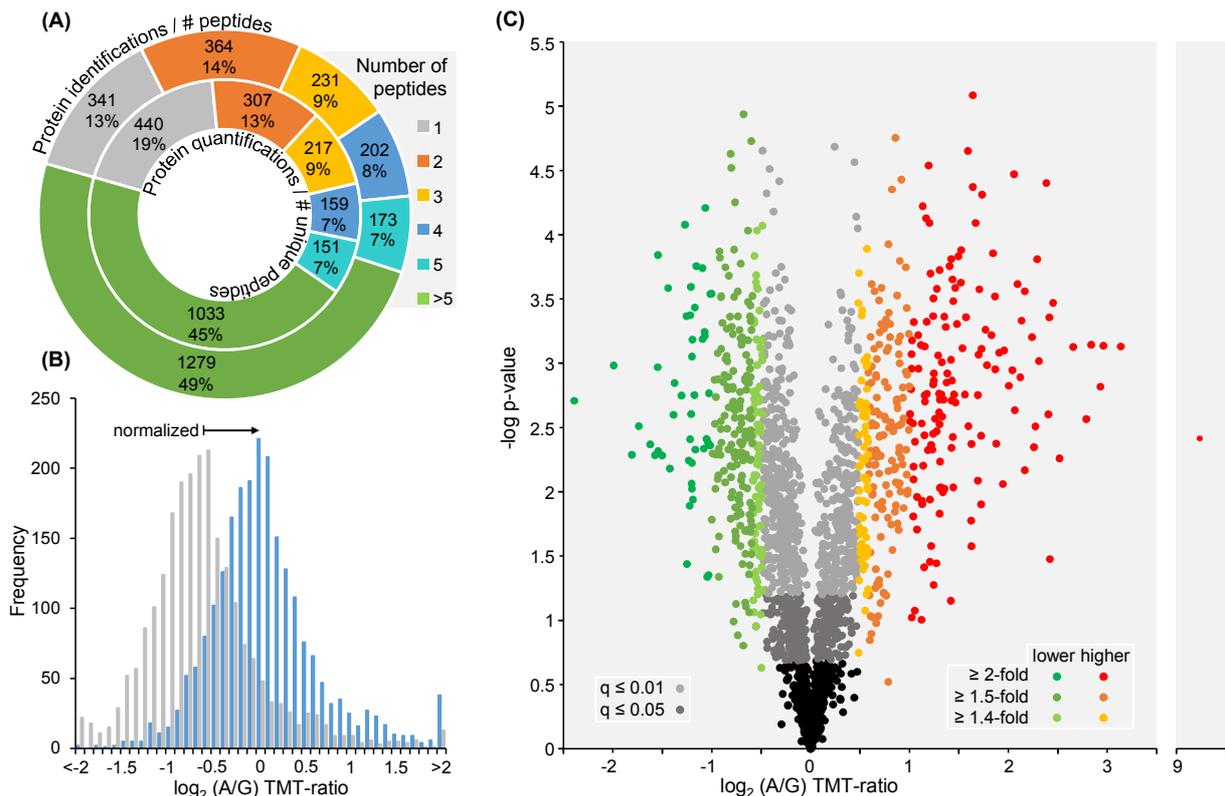


Figure 3.8: Overview of the proteomic dataset. (A) The pie-chart summarizes the peptide evidence per protein-group identification (outer circle) and the number of unique peptides supporting protein-group quantification (inner circle). (B) The overlaid histogram shows the spread of data before (grey) and after normalization (blue). Observed fold-changes are shifted by the median from 0.61 to 1. (C) The volcano-plot shows the relative abundances and statistical significance from Welch's t-test of the selected 2,099 *E. coli* proteins between bacteria grown on acetate or glucose. *P*-values were calculated using two-tailed Welch's t-test of (A/G) protein-ratios compared to control-ratios between biological replicates of equal treatments, corrected for multiple testing by permutation based Benjamini Hochberg FDR procedure. *P*-value and fold-change thresholds are indicated by colors (see legend).

A noticeable difference between *E. coli* grown on acetate compared to glucose was the overall protein content. The protein concentration was determined directly after cell disruption. According to BCA assay results of the triplicate cultures, bacteria grown on glucose yielded approximately 25 % more protein (mean A/G ratio = 0.8). These differences were balanced before protein purification, digestion and labeling. Furthermore, post-processing of quantitative data included

median normalization, shifting the median of TMT-ratios (A/G ratio = 0.766) to one. Altogether, an (A/G) protein-ratio of 0.61 can be calculated, relating to equal amounts of OD₆₀₀ = 1 culture. The combination of different volumes of the protein extracts and subsequent data normalization consequently lead to an accumulated shift of the complete dataset (Figure 3.8 B).

Surprisingly, this discrepancy was even more severe if calculated in relation to the dry weight of biomass from acetate or glucose cultures, which were 0.51 g/L and 0.4 g/L, respectively, at OD₆₀₀ = 1. The overall (A/G) protein-ratio calculated in relation to equal amount of dry biomass would be 0.48.

This may be one reason, why the quantitative data are skewed towards higher (A/G) protein levels (Figure 3.8 C) and suggests that ratios towards higher abundance are collectively overestimated, while ratios indicating lower abundance are generally underestimated.

To evaluate the proteome coverage of the dataset, the 2,099 quantified proteins were tested for enrichment/depletion of gene ontology (GO) and KEGG categories compared to the complete reference proteome set of *E. coli* used for the database search (4,305 entries). The results of enrichment tests are noted with FDR-corrected *p*-values (q-value) for statistical significance and enrichment factors to indicate enrichment or depletion. Compared to the *E. coli* complete proteome set, quantified proteins were enriched in several categories of cellular compartment GO annotations, including proteins of the cytoplasm (q < 1E-45; EF = 1.53), the cytosol in particular (q < 1E-30; EF = 1.86) as well as proteins located in the periplasmic space (q < 1E-7; EF = 1.38). A significant depletion was observed for membrane proteins (q < 1E-12; EF = 0.82). This bias is also reflected in the GO annotations of biological processes and molecular functions. Consequently, the dataset showed remarkable coverage of metabolic pathways that are primarily located in the cytosol and periplasm, whereas significantly fewer proteins responsible for membrane transport or of those involved in other distinct cellular processes including transposition, cell adhesion and DNA recombination were represented.

3.4.5 Comparison to previous quantitative mRNA and protein datasets

To verify and evaluate relative quantitative data, the results from this study were compared to two previously published system wide quantitative datasets (Liu, Durfee et al. 2005, Schmidt, Kochanowski et al. 2015). The proteomic data was compared to a transcript level expression profile of *E. coli* K12 substrain MG1566 on defined minimal medium with either glucose or acetate (Liu, Durfee et al. 2005). A comparable protein level dataset for the analyzed conditions was not available for *E. coli* strain MG1566, however, *E. coli* strain BW25113 was previously analyzed under these conditions by intensity based label free quantification (Schmidt, Kochanowski et al. 2015).

3.4.5.1 Correlation with a transcript level expression profile

In general, system wide transcriptomic analyses yield more comprehensive datasets, compared to quantitative proteomics, due to the ease of signal amplification by polymerase chain reaction. Conversely, protein level data arguably reflect cellular functions more directly. The consulted transcriptome dataset encompassed 4,124 relative mRNA levels of protein coding genes. In total, transcript levels for 2,085 of the 2,099 quantified proteins were found in both datasets. Considering that the experiments were conducted in completely independent laboratories, proteomic and transcriptome data showed a moderate linear relationship, comparable to previously published results for such system wide data sets (Pearson's $r = 0.51$; Spearman's $\rho = 0.54$). The correlations between analyses of gene expression at the mRNA level compared to the protein level have been consistently reported as modest, both for absolute- and relative quantitative data (Griffin, Gygi et al. 2002, Ishihama, Oda et al. 2005, Maier, Guell et al. 2009). This highlights how post-transcriptional and post-translational regulatory processes and the stability of the respective gene products can influence these data. However, decoupling of transcription and translation should be reflected in a deviation from the linear association between data at the transcript or protein level. The combination of results at both levels may provide complementary insight into the process of gene expression.

The correlation of relative quantitative data for both gene products from *E. coli* strain MG1655 grown either on acetate or glucose showed 107 data points outside the calculated 95 % prediction interval, with 88 entries showing a significantly higher ratio and 19 with a significantly lower (A/G) ratio at the transcript level as compared to protein-ratios. Corresponding results for a less stringent 90 % confidence level for predicted values gave 170 deviating data points, with 119 above and 51

below the interval. These outliers were further categorized and tested for enrichment and depletion of functional annotations.

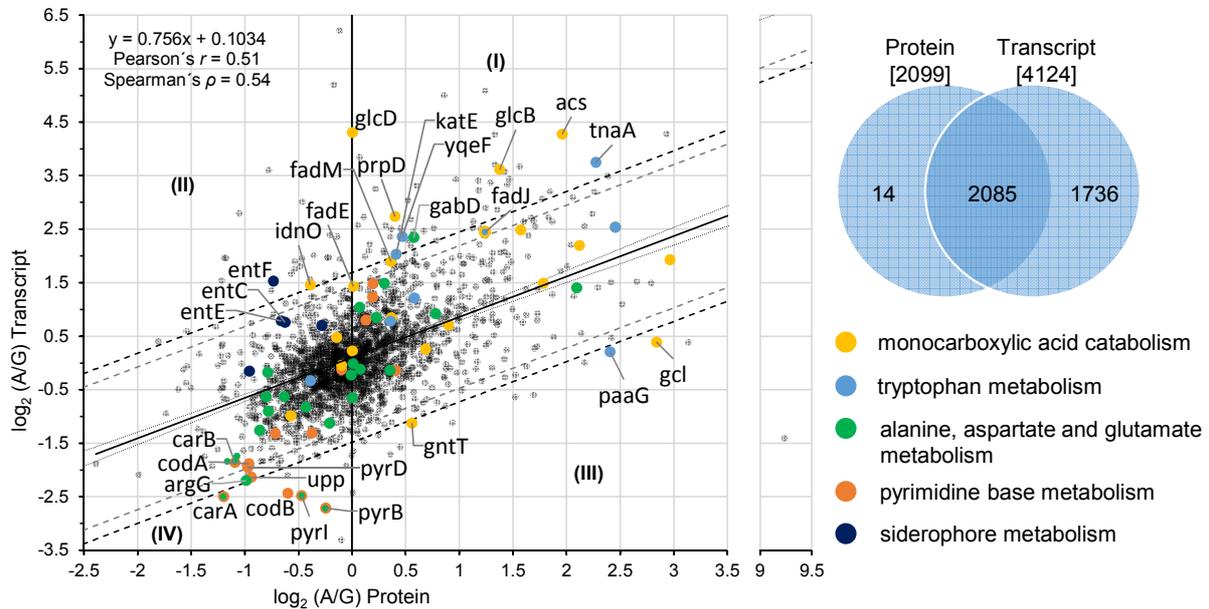


Figure 3.9: Correlation of relative gene expression data at the protein and transcript level. Proteomic data generated in this study was compared to a transcriptome expression profile of *E. coli* on defined minimal medium with glucose or acetate, previously published by Liu, Durfee et al. (2005). In total, transcript levels for 2,085 entries could be matched. Light grey and dark grey dashed lines show 90 % and 95 % prediction intervals of correlation, respectively. The four areas, separating data points tested for GO and KEGG category enrichment, are indicated (I-IV).

Four categories with significant differences between relative abundance at the transcript level compared to the protein level were partitioned into four areas (Figure 3.9):

(I) observed gene expression data suggests significantly stronger induction at the transcript level among proteins with positive (A/G) ratios (69 proteins); (II) stronger induction at the transcript level with negative (A/G) protein-ratios (36 proteins); (III) reduced transcript levels among positive (A/G) protein-ratios (33 proteins); and (IV) stronger transcript level reduction among the group of proteins with negative (A/G) ratios (25 proteins).

Few functional groups were significantly enriched in these four subsets of outliers. Proteins associated with the siderophore metabolic process showed a trend towards the group in category (II) ($q < 7.2E-3$), suggesting that these proteins are induced at the transcript level (mean \log_2 mRNA (A/G) = 0.5), whereas the corresponding proteins were of lower abundance on acetate (mean \log_2 protein (A/G) = -0.72). Proteins involved in pyridine base biosynthesis ($q < 4.4E-3$; EF = 27.8) and alanine, aspartate and glutamate metabolism ($q < 4.7E-2$; EF = 12.83) were enriched in group (IV) and appear to be strongly down regulated at the transcript level. Gene products of

monocarboxylic acid catabolic process ($q < 3.4E-2$; EF = 8.24) and tryptophan metabolism ($q < 4.4E-2$; EF = 13.43) were clustered in area (I).

3.4.5.2 Correlation with relative protein abundance in strain BW25133

Comparison of relative quantitative data at the protein level for the two different strains, MG1655 and BW25113 (Schmidt, Kochanowski et al. 2015) was possible for a total of 1,734 proteins quantified in both datasets (Figure 3.10).

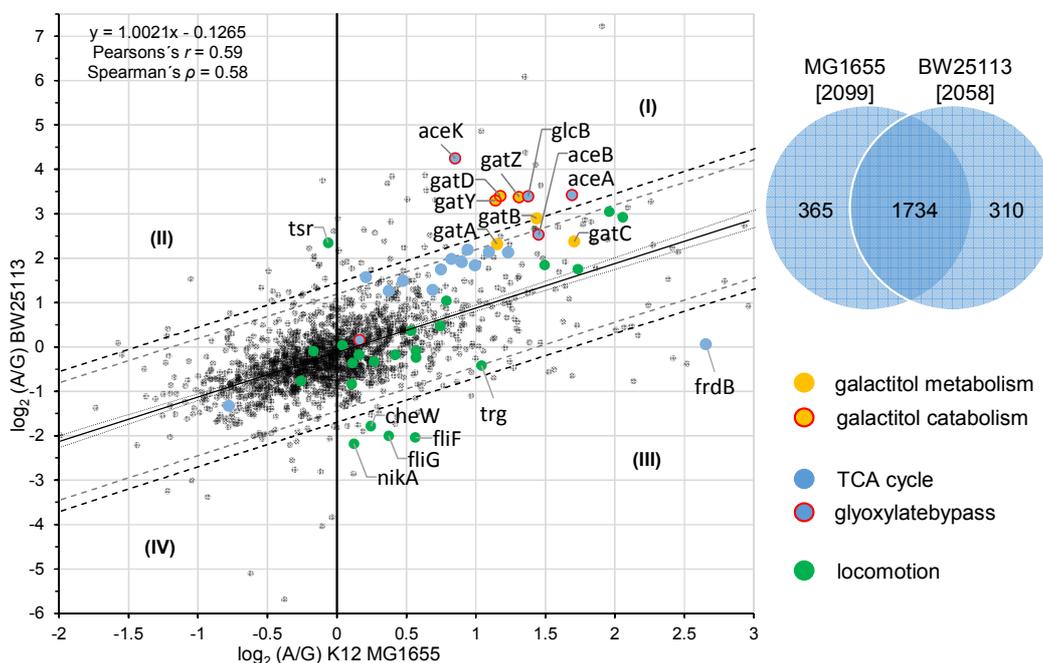


Figure 3.10: Correlation of relative protein abundances in *E. coli* strains MG1655 and BW25113.

Proteomic data generated in this study was compared to a dataset of *E. coli* on defined minimal medium with glucose vs. acetate, previously published by Schmidt, Kochanowski et al. (2015). The consulted dataset encompassed 2,058 relative protein levels for acetate vs. glucose. In total, 1,734 entries could be compared. Light grey and dark grey dashed lines show 90 % and 95 % prediction intervals of correlation. The four areas separating data points tested for GO and KEGG category enrichment are indicated (I-IV).

Overall, the linear relationship between the proteomic and transcriptome datasets was weaker than the correlation between the relative protein abundances in strains MG1655 and BW25113 (Pearson's $r = 0.61$; Spearman's $\rho = 0.62$). Additionally, fewer outliers were calculated from prediction intervals. Data pairs outside the 95 % prediction interval amounted to 101, with 59 entries indicating higher acetate over glucose protein abundances for BW25113 and 42 entries below the prediction interval with significantly lower relative abundances for BW25113 compared

to MG1655. The corresponding sum of data points outside the 90 % prediction level was 129, with 72 entries above and 57 entries below the calculated limits.

The same categories with significant differences between relative protein abundances in the two strains, were defined as described for the transcriptome vs. proteome data comparison: (I) data suggests significantly stronger increase in BW25113 among proteins with positive (A/G) ratios in MG1655 (45 proteins); (II) stronger increased abundance in BW25113 with negative (A/G) protein-ratios (27 proteins); (III) lower protein level change in BW25113 among positive (A/G) protein-ratios (38 proteins); and (IV) a significantly stronger decrease in BW25113 among the group of proteins with negative (A/G) ratios in MG1655 (19 proteins).

Proteins associated with locomotion were enriched in the subsets of data pairs in quadrant (III) ($q < 2.2E-2$; EF = 11.41). This indicates a higher acetate over glucose protein abundance in strain MG1655, whereas proteins responsible for locomotion were lower abundant on acetate compared to glucose in strain BW25113.

Proteins represented in quadrant (I) were enriched in the two GO biological process annotations of galactitol metabolism ($q < 2.6E-2$; EF = 25.69) and catabolism ($q < 2.5E-2$; EF = 38.53). However, the enrichment for both categories can be attributed to proteins involved in the catabolism of galactitol as a subset of the proteins of galactitol metabolism (Figure 3.10). A similar trend is observed for proteins of the TCA cycle ($q > 2.1E-4$). These proteins show a higher fold change in strain BW25113 (mean $\log_2(A/G) = 0.77$) than in strain MG1655 (mean $\log_2(A/G) = 0.72$), and are clustered around the upper 90 % prediction line. However, the most significant difference is observed in the subset of proteins responsible for the glyoxylate bypass (Figure 3.10), which indicates a stronger up regulation of the corresponding genes in strain BW25113 upon growth on acetate.

3.4.6 Functional annotation of differentially abundant proteins

In this study, differentially abundant proteins between *E. coli* grown on acetate compared to glucose were defined using Welch's t-test allowing for a false discovery rate of 1 % (Benjamini and Hochberg 1995), in addition to a fold-change cut-off for TMT-ratios above 1.5. Using these thresholds, 507 differentially abundant proteins were identified, of which 202 proteins changed by more than 2-fold (Figure 3.8).

For the following discussion, quantitative results refer to protein ratios describing relative abundance on acetate over glucose condition (A/G). The differentially abundant proteins include 221 lower abundant proteins and 286 proteins of higher abundance in acetate medium compared to glucose. To categorize these proteins, three annotation enrichment analyses for the set of (I) differentially, (II) lower and (III) higher abundant proteins in relation to the 2,099 quantified proteins were performed using fisher exact test. Additionally, all quantitative data were analyzed by 1D annotation enrichment (Cox and Mann 2012) to assess consistent trends of proteins within categories that could not be detected using the strict fold change cutoff employed here.

3.4.6.1 Central carbon metabolism

Relative quantification was possible for a near complete set of the proteins associated with central carbon metabolism (Figure 3.11). During anaerobic growth on acetate the production of energy and reduction equivalents is dependent on the TCA cycle. Significant enrichment of the pathway was observed among the 507 differentially abundant ($q > 2.1E-4$, EF = 2.93) as well as the 286 higher abundant proteins ($q > 2.1E-4$, EF = 2.93). The 25 TCA cycle proteins show significant differences in the examined conditions. In total, 17 proteins were differentially abundant (≥ 1.5 fold differences and $q > 0.01$). Most were more abundant (15 proteins), and only two proteins, Phosphoenolpyruvate carboxylase (Ppc) and Malate:quinone oxidoreductase (Mqo) were of lower abundance in the bacteria grown on acetate.

The abundances of the two enzymes, which can catalyze the oxidation of malate to oxaloacetate, are of particular interest. Mqo is known to be regulated by the carbon source (van der Rest, Frank et al. 2000), and both transcript and proteome data suggest a down regulation of *mgo* gene expression on acetate. Contrary to this, the NAD⁺-dependent malate dehydrogenase (Mdh) was found to be more abundant at the protein level while equal levels of the *mdh* transcript were detected under both cultivation conditions (Liu, Durfee et al. 2005) (Figure 3.11).

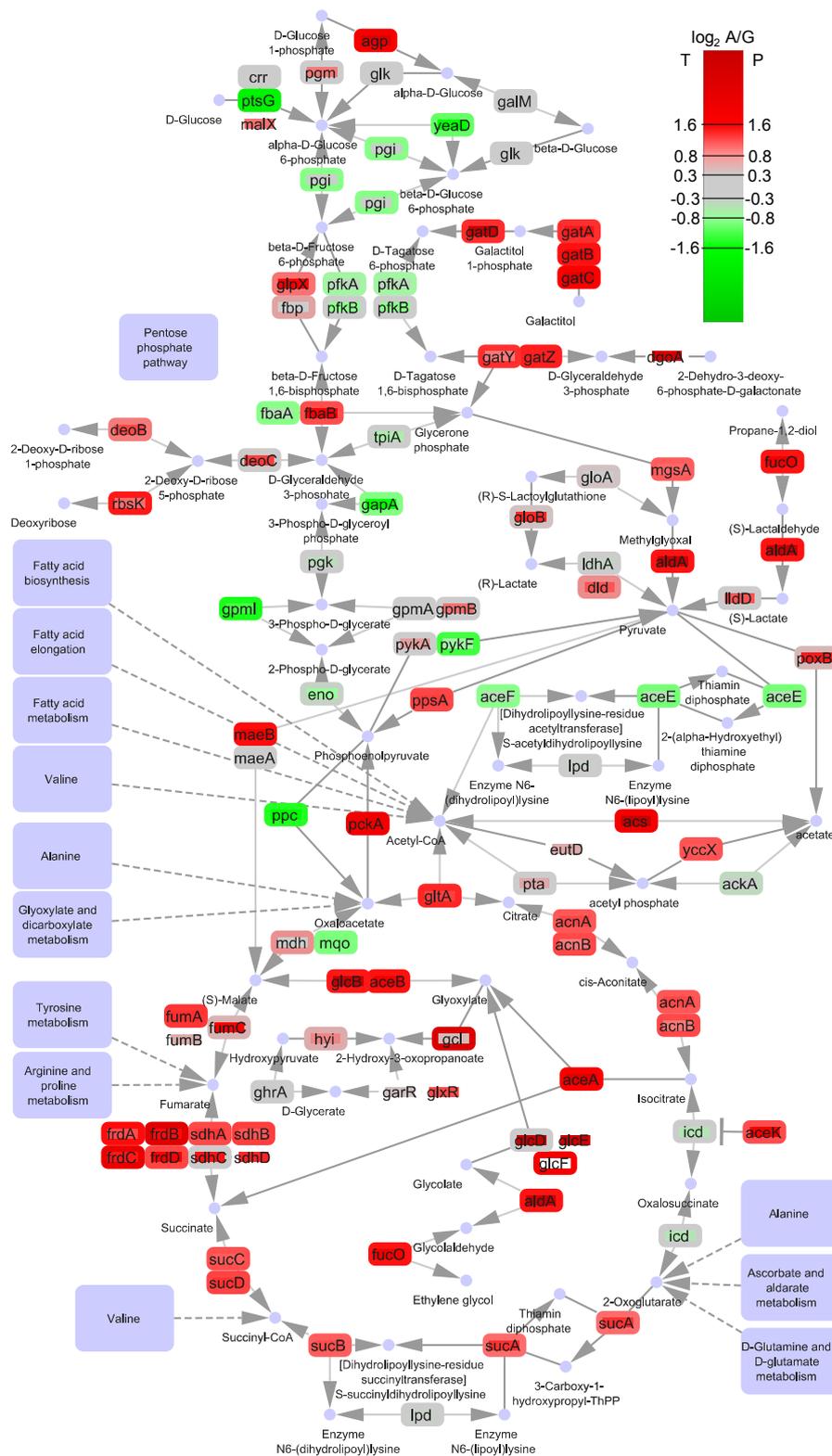


Figure 3.11: *E. coli* central carbon metabolism. Proteins are represented by encoding genes (common gene names from UniProt), node colors indicate \log_2 transformed acetate/glucose relative expression values at the protein (outside) and transcript level (inside) (Liu, Durfee et al. 2005). Proteins or mRNA which missing quantitative data are without color. Sources: KEGG, Ecocyc.

In particular, the products of the glyoxylate shunt operon (*aceBAK*) were more abundant on acetate. The Ace proteins regulate the branch point from TCA cycle to components of the glyoxylate bypass of acetate metabolism. The glyoxylate bypass is essential for growth on acetate because each turn of the TCA cycle involves the loss of two carbon atoms as carbon dioxide. To replenish necessary intermediates for cellular biosynthesis requires the operation of this separate anaplerotic pathway (Hillier and Charnetzky 1981, Nunn 1986).

Isocitrate dehydrogenase kinase/phosphatase (AceK), which regulates the entry from TCA cycle into glyoxylate bypass, was 1.8-fold higher abundance on acetate. The kinase inactivates Isocitrate dehydrogenase (IcdA) by phosphorylation of the enzyme at serine 113 (Cozzone 1998). IcdA was also identified and the regulatory post-translational modification could be quantified on the peptide level. While the protein was not differentially abundant, a significant change of IcdA was detected exclusively for the peptides containing the phosphorylation site. This post-translational modification was detected three distinct peptides with high confidence by a multi-pass database search for phosphorylated peptides. Peptide-ratios of TMT reporter ions showed both the higher abundance of the modified peptide (peptide-ratios between 12.3 and 24.0), and the lower abundance of the unmodified peptide (peptide-ratio of -0.29) in acetate dissimilating bacteria. (Figure 3.12).

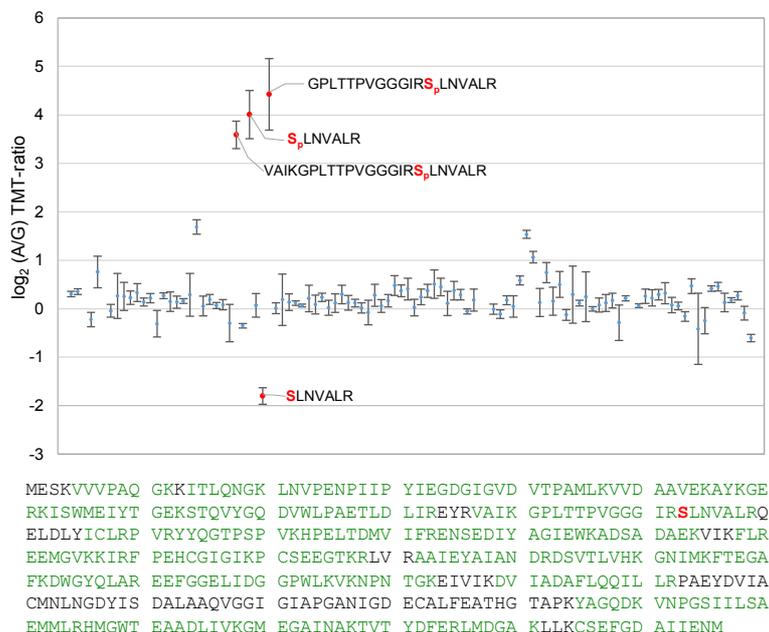


Figure 3.12: Sequence coverage and peptide quantitative data of IcdA protein. IcdA was identified by 64 unique peptides covering 83.2 % of the protein sequence as highlighted in green. Including distinct peptide species counting different modifications on the same sequence, 85 peptide-ratios contribute to a protein-ratio of 1.12 ($\log_2(A/G) = 0.16$) indicating almost equal amounts of the protein under both conditions. In contrast, the four peptides identified containing serine 135 (red) show differential expression for the phosphorylated and unmodified form of the enzyme. Error-bars show the standard deviation of peptide TMT-ratios; phosphoserine is designated as “S_p”.

The two remaining proteins of glyoxylate shunt operon are Isocitrate lyase (AceA), which initiates the glyoxylate bypass and one of two *E. coli* malate synthase isozymes (AceB). Although AceB is reported to be the predominant malate synthase for glyoxylate formed from acetate dissimilation (Molina, Pellicer et al. 1994), the second isozyme (GlcB) showed equally high (A/G) protein-ratios.

Besides the anaplerotic glyoxylate shunt, several other enzymes involved in dicarboxylate and pyruvate metabolism were more abundant. In contrast, glycolytic enzymes were down regulated on acetate compared to glucose ($q < 1.3E-02$, $EF = 3.07$). Of 34 quantified proteins associated with the pathway, 15 were differentially abundant and the 4 proteins of higher abundance in acetate are all associated with gluconeogenesis.

The class II fructose bisphosphate aldolase (FbaA) is among the 10 lower abundant proteins whereas the class I protein (FbaB) which is assumed to be involved in gluconeogenesis was of higher abundance in bacteria grown on acetate, supporting the distinct metabolic roles of these isozymes. Among the three other proteins of higher abundance were the glucogenetic fructose-1,6-bisphosphatase, which is required for growth on acetate (Fraenkel and Horecker 1965). The remaining two proteins were Glucose-1-phosphatase (Agp) a periplasmic glucose 1-phosphate scavenger (Cottrill, Golovan et al. 2002), and phosphoenolpyruvate carboxykinase (PckA), which delivers the main carbon flux from TCA cycle to gluconeogenesis pathways in acetate metabolism (Chao, Patnaik et al. 1993).

In summary, glycolytic proteins, PykF, Ppc, FbaA and PfkA were more abundant on glucose while their gluconeogenic counterparts PpsA, PckA, FbaB and Fbp, were more abundant on acetate.

In previous transcriptional studies (Liu, Durfee et al. 2005, Enjalbert, Letisse et al. 2013), the *fbp* gene was surprisingly reported with the same level of expression on acetate and glucose. This implies regulation of the gene product at the translational or post-translational level.

Acetyl-CoA synthetase (Acs) is of primary importance in the transition from growth on acetogenic carbon sources to scavenging for acetate (Kumari, Beatty et al. 2000). and *acs* gene expression was observed during growth on acetate and at stationary phase (Kumari, Simel et al. 2000, Brock, Maerker et al. 2002, Keseler, Mackie et al. 2013). These findings are consistent with the protein level data acquired in this study.

In addition to Acs, all the other proteins encoded on the transcription unit *acs-yjcHG* were of higher abundances. These findings are also in agreement with the differences in transcript levels between acetate and glucose fed bacteria (Oh, Rohlin et al. 2002, Liu, Durfee et al. 2005). Acs, the cation/acetate symporter ActP and the inner membrane protein YjcH, all belong to the same operon. They are all related to acetate utilization and were among the proteins with the highest relative abundance on acetate compared to glucose fed *E. coli*.

3.4.6.2 Fatty acid metabolism

Significant differences were also observed in proteins associated with proteins of fatty acid metabolism. In particular enzymes involved in fatty acid degradation were elevated on acetate, while proteins necessary for fatty acid biosynthesis were not significantly affected (Figure 3.13).

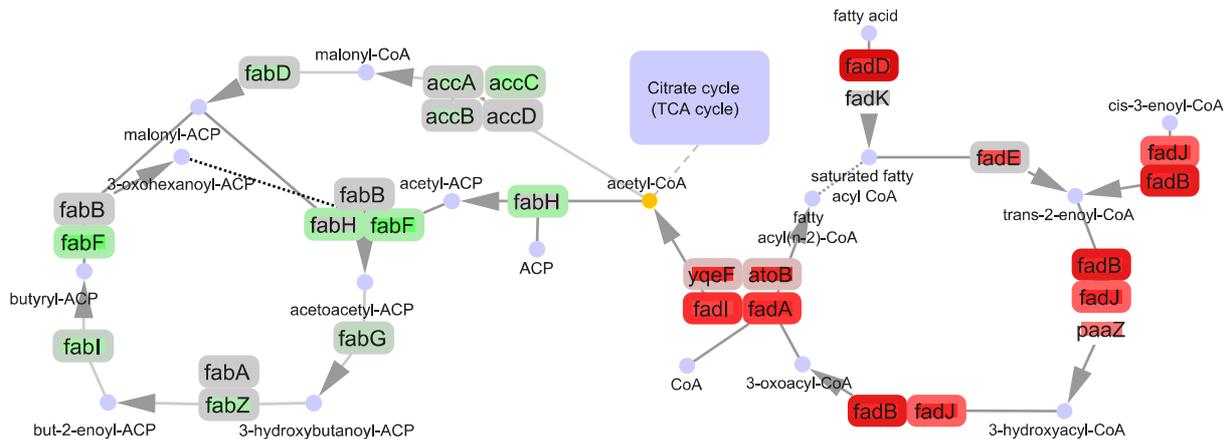


Figure 3.13: *E. coli* fatty acid metabolism. Proteins are represented by encoding genes (common gene names from UniProt), node colors indicate relative expression values \log_2 acetate/glucose at the protein (outside) and transcript level (inside) (Liu, Durfee et al. 2005). Proteins which were not quantified are without color. Sources: KEGG, EcoCyc.

Although acetate can be classified as a short chained fatty acid, it is not a substrate of fatty acid degradation. Already a two-carbon compound, after entering the cell, acetate is converted to the high energy intermediate acetyl-CoA to be metabolized in the TCA cycle. However, the fatty acid degradation regulon is subject to catabolite repression when *E. coli* grows on glucose. The observed differences could therefore be viewed as lower abundance on glucose due to repression of the *fad* operon instead of an increase of these proteins on acetate. Nonetheless, the catabolism of longer fatty acids and acetate show parallels in that both substrates are converted to acetyl-CoA and require the glyoxylate shunt to assimilate carbon.

In previous gene expression studies, the transcript level abundance of *fadE* has been consistently reported as higher on acetate compared to glucose (Oh, Rohlin et al. 2002, Liu, Durfee et al. 2005). However, in contrast to all the other proteins involved in fatty acid degradation, the relative protein abundance determined here indicate the presence of equal amounts of the enzyme on glucose and acetate. This implies post-transcriptional regulation of the gene product. Given that FadE catalyzes the rate determining step of any one cycle of oxidation of acetyl-CoA (O'Brien and Frerman 1977), the equal relative abundance of this protein on acetate and glucose may reflect a regulation of the pathway under these growth conditions. This may how fatty acid degradation is

regulated in the absence of medium- and long-chained fatty acid, the substrates and inducers of the full pathway.

3.4.6.3 Energy production

To generate energy from acetate, the respiratory chain is required to oxidize NADH produced in the TCA cycle. Although the proteins involved in electron transport chain of *E. coli* are relatively well described, many aspects of the changes to the process under conditions of slow growth and the in the absence of preferred nutrients have yet to be fully characterized.

Both succinate dehydrogenase (SdhABCD) and fumarate reductase complex (FrdABCD) were strongly induced on acetate. Succinate dehydrogenase provides a junction from the TCA cycle to the respiratory electron transport chain. The electron supply from oxidation of succinate to fumarate is used for energy production. Fumarate reductase is primarily associated with the reduction of fumarate to succinate under anaerobic conditions and the *fdr* operon is known to be repressed by preferred electron acceptors (like O₂) and induced by fumarate (Jones and Gunsalus 1985). Contrary to these previous findings, both protein and transcript level data suggest that the induction of *fdr* operon on acetate under aerobic conditions is even stronger than that of *sdh*. *Fdr* is structurally similar to succinate dehydrogenase and their functions are partially interchangeable (Guest 1981). As both proteins are more abundant in bacteria growing on acetate, they may play a role in regulating the equilibrium between electron acceptors and donors.

Another component of the electron transport machinery found in significantly elevated levels in bacteria grown on acetate was CydAB, for which both subunits were affected (CydA, log₂ (A/G) = 0.57; CydB, log₂ (A/G) = 1.05). The cytochrome bd-I oxidase (CydAB) is a terminal complex of the electron transport chain. The oxidase catalyzes the two electron oxidation of ubiquinol and the four electron reduction of oxygen to water and contributes to the production of proton motive force by 1e⁻/H⁺ (Borisov, Gennis et al. 2011). Reportedly induced under several stress conditions, including oxygen limitation pH stress and transmembrane gradient uncoupling chemicals (Giuffre, Borisov et al. 2014), it is unclear if or how the change in terminal oxidase benefits *E. coli* under these circumstances. Despite this, the protein level data suggest that energy production is part of the metabolic programs, which are restructure in *E. coli* to suit the growth conditions on these different nutrients.

Also of interest are the maintenance of the balance between reduction equivalents NAD⁺ and NADP⁺. *E. coli* acetate metabolism produces an excess of NADPH which is regenerated by a system of antithetic transhydrogenases. *SthA* uses surplus NADPH to convert NAD⁺ to NADH (Sauer, Canonaco et al. 2004). In accordance, the enzyme was twofold higher in abundance on

acetate ($\log_2 (A/G) = 1.02$). Conversely, both subunits of PntAB the enzyme catalyzing the regeneration of NADH by reduction of NADP^+ were of lower abundance ($\log_2 (A/G) = -1.02$ for α -subunit and -0.63 for β -subunit).

3.4.6.4 Protein synthesis

Although *E. coli* can utilize acetate as a carbon and energy source as described in the previous chapters, the exposure to acetate during glucose-based growth is known to limit growth and inhibit protein biosynthesis (Aristidou, San et al. 1995). This is one of the main reasons why the acetate metabolism and acetate overflow on glucose based growth are of considerable biotechnological interest. In adapted bacteria growing long-term on acetate, the lower energy carbon source results in a much slower growth rate accompanied by a lower rate of protein synthesis compared to glucose based growth. The correlation of protein synthesis and growth rate has also been studied previously (Milne, Mak et al. 1975).

The high proteome coverage in this study allowed a comprehensive analysis of *E. coli* protein synthesis machinery, including the proteins involved in amino acid metabolism, the aminoacyl-tRNA synthetases, ribosomal proteins and the protein transport systems (Table 3.1).

Table 3.1: Enrichment of KEGG terms associated with protein synthesis. The number of selected proteins in categories involved in protein transport, amino acid and protein biosynthesis are stated. The corresponding Benjamini Hochberg corrected p -values as critical FDR values (q) are included for significant test results ($q \leq 0.05$). Significant values from 1D-enrichment analysis are shown with mean \log_2 ratio of all proteins in the respective category and q -values in parenthesis. Amino acids are indicated by common 3-letter symbols.

KEGG Term	total # in database	quantified (q)	differential (q)	higher (q)	lower (q)	mean \log_2 ratio (q)
total	4305	2099	507	286	221	
Val, Leu, Ile degradation	12	9	6	5 (3.4E-02)	1	1.07 (1.6E-02)
Val, Leu, Ile biosynthesis	20	18 (8.7E-04)	8	6	2	0.02
Phe, Tyr and Trp biosynthesis	21	20 (9.5E-05)	6	0	6	-0.32 (1.0E-02)
Trp metabolism	11	9	4	4	0	1.09 (7.7E-03)
His metabolism	9	9 (8.7E-03)	5	0	5 (2.5E-02)	-0.59 (2.9E-03)
Ala, Asp, Glu metabolism	30	26 (1.9E-04)	13 (4.8E-02)	2	11 (1.6E-03)	-0.28 (3.6E-02)
Arg, Pro metabolism	43	33 (8.9E-04)	15 (4.6E-02)	3	12 (2.7E-03)	-0.29 (3.0E-03)
Gly, Ser, Thr metabolism	32	25 (2.9E-03)	8	1	7	-0.26 (4.3E-02)
Cys, Met metabolism	29	27 (1.5E-05)	7	2	5	-0.17
Lys biosynthesis	18	17 (4.0E-04)	6	0	6	-0.44 (2.8E-03)
Lys degradation *	10	8	4	4	0	1.08 (2.8E-03)
Ribosome	57	55 (3.3E-13)	26 (6.5E-03)	1 (2.9E-02)	25 (2.3E-09)	-0.53 (1.8E-17)
tRNA biosynthesis	27	26 (4.2E-06)	2	0	2	-0.33 (3.9E-04)
Protein export	18	15 (1.2E-02)	1	0	1	-0.29 (4.9E-02)

* Lysine degradation as described in the KEGG repository is incomplete in *E. coli*, missing key proteins resulting in a futile pathway which primarily involves the proteins of fatty acid degradation (see Chapter 2.4.6.2)

Ribosomal proteins and aminoacyl-tRNA ligases

With the exception of YkgO and YkgM, two predicated paralogs of ribosomal subunits, all ribosomal polypeptides could be detected and quantified. In general, ribosomes were of significantly lower abundance on acetate ($q < 1.8E-17$; mean $\log_2 (A/G) = -0.53$) with 25 proteins showing ratios below 0.71. Only two ribosome associated proteins were observed to be of higher abundance, namely, ribosome associated inhibitor1 (RaiA; $\log_2 (A/G) = 1.31$) and the stationary phase induced ribosome associated protein (Sra; $\log_2 (A/G) = 1.27$). Incidentally, these proteins are both linked to a modulation of protein synthesis (Maki, Yoshida et al. 2000, Izutsu, Wada et al. 2001), as with many of the proteins associated with stationary growth phase. This suggests that the proteins are not specifically induced by stationary phase but more generally induced with slow growth and the associated decrease of protein synthesis.

1D enrichment analysis indicates that the category of proteins involved in aminoacyl-tRNA biosynthesis, as a whole, was also significant lower in abundance, with moderate but very consistently low TMT-ratios ($q < 1.8E-17$; mean $\log_2 (A/G) = -0.53$).

Integration of transcript level data also shows a biased distribution of the different tRNA isoacceptors on acetate compared to glucose. As with the aminoacyl-tRNA synthetases, most of their tRNA substrates were significantly lower in abundance on acetate. However, the three proline tRNA acceptors were exceptions to this (Figure 3.14).

Additionally, the abundance pattern of several other tRNA were clearly biased for specific anticodons among carriers of the same amino acid. For example, the four UAC anticodon tRNA^{val} species encoded by *valT*, *valXY* and *valZ* were of significantly lower abundance on acetate whereas the two GAC anticodon tRNA^{val}, encoded by *valV* and *valW*, showed equal level at both growth conditions. Equal abundances (against the overall trend) were also observed for the tRNA carrying glycine with anticodon CCC (*glyU*), those for serine with anticodon CGA (*serU*) and anticodon UCU (*argU*) for arginine. For tRNA carrying leucine, no significant abundance change were observed for those with anticodons UAG (*leuU*), GAG (*leuZ*), and UAA (*leuX*) in contrast to the four tRNA with the anticodon CAG (*leuQPV* and *leuT*) (Figure 3.14).

The codon usage information for all *E. coli* K12 coding DNA sequences was retrieved from the Codon Usage Database (Nakamura, Gojobori et al. 2000) and the frequency of codons on genes encoding the quantified proteins was calculated. This showed that the few tRNA, which are not down regulated on acetate, represent codons that occur with low frequency on the *E. coli* K12 coding DNA sequences (Table 3.2). To investigate if these specific anticodons coincide with a codon bias on genes associated with acetate or glucose based growth, the codon usage of genes encoding differentially abundant proteins was calculated.



Figure 3.14: *E. coli* aminoacyl-tRNA synthesis. Proteins and tRNA are represented by encoding genes (common gene names from UniProt), node colors indicate \log_2 (A/G) ratios at the protein (outside) and transcript level (inside) (Liu, Durfee et al. 2005). Nodes representing the amino acid synthesis pathways are colored by the mean of \log_2 ratios of all involved proteins. Sources: KEGG, Ecocyc.

Interestingly, five of the tRNA with abundances not following the overall trend of lower abundance on acetate, match codons which are more frequent in genes encoding more abundant proteins compared to those of lower abundance on acetate. In two cases, this frequency bias also extends to possible wobble-pairing codons. Codons that are recognized by the anticodons of tRNA not down regulated on acetate are less frequently used on those genes encoding proteins of lower abundance on acetate (Table 3.2).

Table 3.2: Codon usage bias among genes encoding higher or lower abundant proteins. Amino acids are represented in three-letter code; the gene names encoding tRNA are included with their respective codon and anticodon. The table is limited to those amino acids with tRNA abundances that do not follow the overall trend of lower abundance on acetate. The tRNA genes not differentially abundant on acetate are in red font, red background denotes tRNA of higher abundance on acetate. Possible wobble-pairing codons for tRNA of interest are shown on green background. In each category, the total number of occurrence for each codon, the fraction of occurrence per thousand codons (#/1000) and the fraction per encoding amino acid are included.

Amino Acid	tRNA	Anticodon 3'-5'	Codon mRNA	Total in Database		Quantified		Higher abundant		Lower abundant		Ratio Higher/Lower					
				Total number	#/1000	Fraction	Total number	#/1000	Fraction	Total number	#/1000	Fraction	Ratio [# /1000]	Ratio [Fraction]			
Gly	glyU	CCC	GGG	17628	10.99	0.15	6885	9.88	0.132	1004	10.60	0.136	699	8.74	0.112	1.212	1.215
	glyT	CCU	GGA	12696	7.92	0.11	4420	6.34	0.085	711	7.50	0.096	429	5.36	0.069	1.399	1.402
	glyVWXY	CCA	GGU	39862	24.85	0.34	18876	27.09	0.362	2550	26.91	0.346	2345	29.33	0.376	0.918	0.920
Val		CCG	GGC	47212	29.44	0.40	21926	31.46	0.421	3108	32.80	0.422	2766	34.59	0.443	0.948	0.951
		CAC	GUG	42097	26.25	0.37	18526	26.58	0.370	2531	26.71	0.375	2102	26.29	0.355	1.016	1.054
	valITUXYZ	CAU	GUA	17443	10.88	0.15	7876	11.30	0.157	1005	10.61	0.149	985	12.32	0.166	0.861	0.893
Arg		CAA	GUU	29487	18.38	0.26	13431	19.27	0.268	1802	19.02	0.267	1680	21.01	0.284	0.905	0.939
	valVW	CAG	GUC	24406	15.22	0.22	10200	14.64	0.204	1420	14.99	0.210	1151	14.39	0.194	1.041	1.080
	argW	UCC	AGG	1949	1.22	0.02	447	0.64	0.011	73	0.77	0.016	40	0.50	0.009	1.540	1.793
Arg	argU	UCU	AGA	3291	2.05	0.04	743	1.07	0.019	122	1.29	0.027	64	0.80	0.014	1.609	1.873
	argX	GCC	CGG	8631	5.38	0.10	2948	4.23	0.076	416	4.39	0.091	246	3.08	0.055	1.427	1.662
	argQVVZ	GCU	CGA	5668	3.53	0.06	1816	2.61	0.047	225	2.37	0.049	184	2.30	0.041	1.032	1.202
Ser		GCA	CGU	33711	21.02	0.38	16918	24.28	0.435	1834	19.36	0.401	2118	26.49	0.471	0.731	0.851
		GCG	CGC	35311	22.02	0.40	16006	22.97	0.412	1902	20.07	0.416	1841	23.02	0.410	0.872	1.015
	serV	UCA	AGU	13976	8.71	0.15	5048	7.24	0.132	727	7.67	0.135	459	5.74	0.109	1.337	1.236
Ser	serU	UCG	AGC	25716	16.03	0.28	11123	15.96	0.290	1550	16.36	0.287	1180	14.76	0.280	1.109	1.025
	serT	AGC	UCG	14305	8.92	0.15	5752	8.25	0.150	829	8.75	0.153	662	8.28	0.157	1.057	0.977
	serWX	AGU	UCA	11438	7.13	0.12	3988	5.72	0.104	613	6.47	0.113	416	5.20	0.099	1.244	1.150
Leu		AGA	UCU	13633	8.50	0.15	6169	8.85	0.161	817	8.62	0.151	739	9.24	0.175	0.933	0.863
	leuW	AGG	UCC	13783	8.59	0.15	6292	9.03	0.164	867	9.15	0.160	760	9.50	0.180	0.963	0.890
	leuX	AAC	UUG	22000	13.72	0.13	8357	11.99	0.118	1045	11.03	0.113	868	10.85	0.109	1.016	1.040
Leu	leuPQTV	AAU	UAU	22279	13.89	0.13	7624	10.94	0.108	1071	11.30	0.116	777	9.72	0.097	1.163	1.190
	leuZ	GAC	CUG	84714	52.82	0.50	38747	55.60	0.549	5007	52.84	0.542	4644	58.08	0.582	0.910	0.931
	leuY	GAU	CUA	6182	3.85	0.04	2054	2.95	0.029	298	3.15	0.032	211	2.64	0.026	1.192	1.220
Pro		GAA	CUU	17707	11.04	0.10	6708	9.63	0.095	898	9.48	0.097	707	8.84	0.089	1.072	1.097
	proK	GAG	CUC	17715	11.04	0.10	7149	10.26	0.101	920	9.71	0.100	771	9.64	0.097	1.007	1.030
	proM	GCC	CCG	37316	23.27	0.52	17804	25.55	0.574	2252	23.77	0.552	2144	26.81	0.604	0.886	0.913
Pro		GGU	CCA	13664	8.52	0.19	5732	8.23	0.185	780	8.23	0.191	647	8.09	0.182	1.017	1.048
	proL	GGA	CCU	11291	7.04	0.16	4492	6.45	0.145	583	6.15	0.143	468	5.85	0.132	1.051	1.083
		GGG	CCC	8861	5.52	0.12	3015	4.33	0.097	466	4.92	0.114	290	3.63	0.082	1.356	1.397

The growth conditions analyzed here represent two extremes with highly disparate metabolic and physiologic organizations of the bacteria. In an evolutionary sense, it would be beneficial for a consolidation of specific codons to occur on genes mostly needed under either of two different physiological states. Furthermore, codon bias as a means to optimize gene expression patterns tailored to specific, highly divergent living conditions would only makes sense to apply to codons which are used at an overall lower frequency across the genome.

A bias in *E. coli* codon usage has been discussed with respect to highly abundant proteins, growth rate dependence and the optimization of heterologous protein synthesis (McHardy, Puhler et al. 2004, Henry and Sharp 2007, Angov, Hillier et al. 2008). In the case of the two growth conditions examined here, genes associated with the fast glucose based growth show a bias towards the more abundant codons. Conversely, a bias towards lower frequency codons on genes important for acetate metabolism could be detected.

A synchronization of the tRNA abundance to match the codon usage bias of genes associated with glucose or acetate based growth would indicate a very precise resource allocation. From the datasets analyzed here, tRNA abundances seem to be aligned in this fashion, however, previous studies agree that the levels of tRNA isoacceptors cognate to less abundant codons remain unchanged with increasing growth rates (Dong, Nilsson et al. 1996, Berg and Kurland 1997). With these previous observations, the codon bias and synchronized tRNA levels on acetate are probably incidental and a function of the reduced growth rate on acetate. Nonetheless, a bias towards the low frequency codons and the comparatively higher abundance of unregulated tRNA would still optimize gene expression in terms of energy consumption and resource allocation on acetate as a carbon source.

Along with the decrease in growth rate, protein synthesis is among the most strongly affected functions under the compared conditions. However, growth on the low energy carbon source also affects other aspects of gene expression. This is reflected in differences in amino acid synthesis and degradation, the abundance of ribosomes, the regulation of tRNA-synthetases and tRNA levels and even the codon frequency on genes associated with the different growth conditions.

Amino acid synthesis and degradation

The molecular basis for the inhibition of protein production includes a reduction of the majority of enzymes involved in amino acid synthesis pathways, whereas those necessary for amino acid degradation were increased (Figure 3.15).

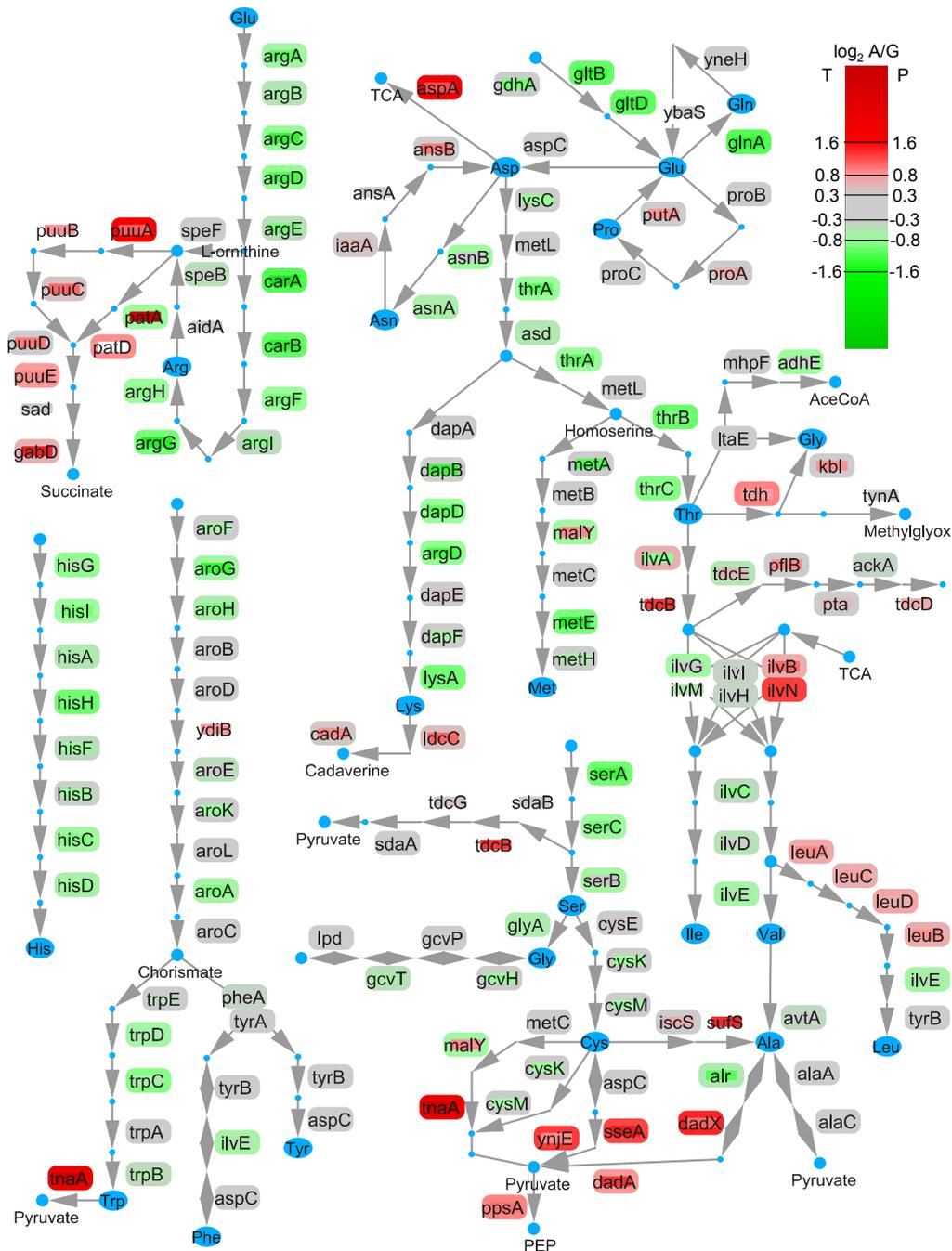


Figure 3.15: *E. coli* amino acid metabolism. Proteins are represented by encoding genes (common gene names from UniProt), node colors indicate \log_2 transformed acetate/glucose relative expression values at the protein (outside) and transcript level (inside) (Liu, Durfee et al. 2005). Proteins or mRNA, which were not quantified, are without color. Sources: KEGG, Ecocyc.

Relative quantitative data were acquired for 26 proteins associated with the pathways of alanine, aspartate and glutamate metabolism. Of thirteen differentially abundant proteins, eleven were less abundant on acetate. Among the two significantly higher abundant proteins involved in the pathways was the aspartate-ammonia lyase (AspA). AspA can supply fumarate to the TCA cycle by deamination of aspartate. It is remarkable, that AspC did not show a difference in abundance even though AspA and AspC together comprises glutamate degradation pathway II which is regarded as the primary pathway for use of L-glutamate as carbon source in *E. coli* (Marcus and Halpern 1969). In contrast elevated levels of PuvE, GabD and GabT suggest the alternative possibility to assimilate glutamate via 4-aminobutanoate (GABA) into the TCA cycle.

The majority of proteins involved in arginine, proline and histidine biosynthesis were also significantly lower in abundance on acetate. This applied to, all proteins encoded by the *argCBH* operon in addition to ArgD, ArgF and ArgG as well as both subunits of carbamoyl phosphate synthetase CarAB. In contrast, the lysine/arginine/ornithine ABC transporter periplasmic protein ArgT, which enables scavenging of arginine, was more abundant on acetate medium. Most proteins of the arginine degradation pathway II (AST pathway) were not detected and the proteins of pathway III were not significantly regulated. However, several proteins of the superpathway of L-arginine, putrescine and GABA degradation (in particular those involved in the degradation of putrescine to succinate and glutamate), were elevated on acetate medium. The highest ratios were observed for glutamate-putrescine ligase (PuvA), and GabD, a dehydrogenase also involved in glutamate degradation. PuvE and PuvA are also part of the enzymes constituting one of the routes available in *E. coli* for the oxidation of putrescine to GABA.

These data account for three pathways, higher abundant on acetate compared to glucose, which allow glutamate, arginine and putrescine to be converted to GABA and subsequently fed into the TCA cycle as succinate.

While the majority of protein synthesis pathways were down regulated at the level of enzyme abundance, those involved in proline synthesis were the exception. None of the proteins necessary to catalyze the conversion of glutamate to proline proteins showed a noticeable difference in abundance between glucose and acetate. This is in agreement with previous research showing that proline synthesis is regulated by feedback inhibition of ProB the Glutamate-5-kinase rather than a change in gene expression (Perez-Arellano, Rubio et al. 2005). Additionally, proline is osmoprotective (Measures 1975) and plays an important role in *E. coli* stress response. Most importantly, the amino acid is easily reconverted to glutamic acid by PutA, a direct link for the oxidation of proline to the electron transport chain, which also did not show a difference in abundance between the two analyzed conditions.

Glutamine utilization is also achieved by initial deamination to glutamate and subsequent degradation of glutamate. As with proline degradation, the necessary enzyme was of equal abundance in both conditions. On the other hand, the glutamine synthase, which catalyzes the reverse reaction, was significantly lower abundant on acetate.

Alanine can be used as carbon and energy source via the alanine degradation pathway. The two proteins involved, degradative alanine racemase (DadX) and D-amino acid dehydrogenase (DadA) were of higher abundance in *E. coli* grown on acetate medium. These proteins can convert alanine into pyruvate, which phosphoenolpyruvate synthase (PpsA) can then channel into carbon metabolism.

The enzymes catalyzing cysteine metabolism were also of equal abundance in glucose and acetate medium. The only exceptions to this were two enzymes necessary for cysteine degradation; the multifunctional hydrogen sulfide synthesis protein (SseA), which metabolizes cysteine to pyruvate and hydrogen sulfide, and the L-cysteine desulfhydrase active multifunctional enzyme TnaA. Also of interest is the high abundant cysteine sulfinate desulfinate, CsdA, which catalyzes the breakdown of cysteine to L-alanine, sulfite and hydrogen, thereby also providing a link to alanine degradation.

Six proteins included in glycine, serine and threonine metabolism were significantly lower in abundance in *E. coli* grown on acetate. One-dimensional enrichment analysis showed the overall down regulation of the pathways ($q > 4.3E-2$; mean $\log_2 (A/G) = -0.26$).

The same trends were observed for the proteins involved in phenylalanine, tyrosine and tryptophan biosynthesis. Of the 21 proteins associated with these pathways (according to KEGG), 20 were quantified. Although only 6 proteins were of significantly lower abundance in acetate medium, with the conservative fold change cut-off, 1D analysis showed the overall trend of the involved proteins towards lower abundance on acetate ($q < 1.0E-2$; mean $\log_2 (A/G) = -0.32$).

Most interesting were the relative abundances of proteins involved in valine, leucine and isoleucine synthesis. The two functional acetohydroxy acid synthases (AHAS), necessary for the first step of valine biosynthesis and the second step of isoleucine biosynthesis showed opposite trends in relative abundance. AHAS III subunits IlvH and IlvI were lower abundant whereas both AHAS I components (IlvBN) were higher abundant on acetate ($\log_2 (A/G) = 1.28$ for IlvN and 0.62 for IlvB). This is also in accordance with previous observations showing that AHAS I is essential for growth on acetate (Dailey and Cronan 1986). Also surprising and not easily explained were the higher abundances of all proteins encoded by the *leuABCD* operon. The congruent relative levels of the *leu* operon gene products on both mRNA and the protein levels show that this observation is not an artifact. It has been noted previously that starvation for leucine triggers derepression of the *leu* operon and increases rates of *leuB* mRNA transcription (Wright, Longacre et al. 1999). However,

the partial induction of these pathways seems counterproductive because *E. coli* does not possess the capability to metabolize the branched-chain amino acids. This means that, once produced, a surplus of these amino acids cannot be easily repurposed for energy production or carbon assimilation.

One-dimensional annotation enrichment also revealed a significant trend towards lower abundance for the 17 proteins involved in lysine biosynthesis ($q < 2.8E-3$; mean $\log_2 (A/G) = -0.44$) whereas the eight proteins annotated to lysine degradation showed a trend towards higher abundance ($q < 2.8E-3$; mean $\log_2 (A/G) = 1.08$). Although the lysine degradation protein lysine decarboxylase 2 (LdcC) was overabundant on acetate medium it has to be noted that the *E. coli* is missing downstream proteins necessary for lysine degradation, namely predicted 5-aminovalerate aminotransferase (Yamanishi, Mihara et al. 2007), glutarate-CoA ligase and glutaryl-CoA dehydrogenase. Consequently, the produced cadaverine is not further utilized and the lysine degradation would end as a futile reaction. However, as with proline synthesis, conversion of lysine to cadaverine is attributed with an important role in stress response instead of nutrient assimilation. It functions as one of the acid resistance systems of *E. coli* active under conditions of anaerobiosis and starvation (Rhee, Kim et al. 2007).

With the exception of amino acid synthesis pathways involved in stress response, the majority of pathways associated with amino acid synthesis were generally of lower abundance on acetate compared to glucose. Conversely, proteins necessary to metabolize amino acids were found to be higher abundant on acetate.

E. coli is known to utilize twenty pathways for the degradation, utilization, or assimilation of proteinogenic amino acids. However, only twelve amino acids can be degraded and, while the surplus pathways are redundant, they may be utilized under specific conditions.

Although amino acids were not supplied with minimal culture conditions, the drastic differences of metabolic proteins may affect the concentration of intracellular metabolites. In this context, additional data about the relative abundance of metabolites and the metabolic fluxes under both conditions would be necessary to draw further conclusions. In recent investigation of the effects starvation has on the intracellular amino acid pool of *E. coli*, two classes of amino acids were defined. Link, Fuhrer et al. (2015) could clearly differentiate amino acids of decreasing or constant concentrations, from those which accumulate during starvation. Essentially, these data showed that amino acids, which *E. coli* cannot metabolize, accumulate while most of the other amino acids are consumed.

With the configuration of enzymes driving amino acid metabolism, observed in this proteome analysis, a similar change in the level of available amino acids can be assumed for the bacterium growing on the low yield carbon source acetate.

In addition, faster adaptation of *E. coli* to favorable changes in living conditions are to be expected if scavenging proteins and degradation pathways for a variety of food sources, including amino acids, are induced. In contrast, the fast growth on glucose, which demands more resources for protein synthesis during glucose based growth, can also account for the differences seen in synthesis pathways. Conversely, the relative lower abundance of pathways necessary for degradation of amino acids can also be attributed to catabolite suppression on glucose.

Protein transport

With the limitation of protein synthesis on acetate, it is reasonable to assume protein export across the cytoplasmic membrane is also reduced. This assumption is supported by the lower abundance of both Sec and twin-arginine translocation (Tat) system proteins. While Sec-translocase is responsible for the transport of unfolded proteins, the Tat-pathway can move folded, cofactor containing macromolecular substrates across the cytoplasmic membrane (Natale, Bruser et al. 2008).

Collectively, the Sec proteins were less abundant on acetate (mean $\log_2 (A/G) = -0.22$). However, reduction of Sec-system components was limited to three proteins which build the protein conducting channel (SecYEG) and only SecY showed a ratio below 0.7 ($\log_2 (A/G) = -0.57$). The ATP-motor protein SecA and ancillary SecDF complex were not noticeably affected. Interestingly, the highest relative abundance was detected for SecB ($\log_2 (A/G) = 0.31$), a chaperone for many periplasmic proteins, which is consistent with the overall increase of periplasmic proteins observed on acetate (Chapter 3.4.6.6).

The Tat system is encoded by *tatABC* and *tatE*. TatA and TatB were not significantly changed in abundance ($\log_2 (A/G) = -0.42$ for TatA and 0.08 for TatB). However, TatE was one of the more strongly affected proteins with a ratio indicating more than 2.3-fold lower protein-abundance in bacteria grown on acetate medium ($\log_2 (A/G) = -1.2$). Besides the much lower abundance of TatE in acetate medium, previous studies implicate TatE as a primary modulator of twin arginine dependent transport: Firstly, TatE is encoded separately, while the other *tat* genes are coexpressed (Weiner, Bilous et al. 1998). Secondly, *tatE* mutants show changes in the spectrum of translocated proteins, while *tatA tatE* double mutants block the system completely. This was previously interpreted as a partial functional redundancy of the two proteins (Sargent, Bogsch et al. 1998). The moderate reduction of TatA may therefore contribute to a general reduction of protein export while the levels of TatE, may control the spectrum of export substrates. In addition, a higher responsiveness of the Tat-system to environmental changes is to be expected if only

modulators of the system are regulated. This is interesting because changes between the conditions examined here included the redox cofactor binding protein complexes associated with energy production (Chapter 3.4.6.3). Furthermore, those proteins involved in energy production would probably be among the first adaptations if a preferred energy source is found.

3.4.6.5 Bacterial chemotaxis and motility

In this analysis, all known proteins associated with bacterial chemotaxis were detected of which a significant number were more abundant in the bacteria metabolizing acetate ($q = 1.56E-4$; mean $\log_2 (A/G) = 0.6$). The most intense differences were observed for proteins involved in galactose/glucose and ribose chemotaxis. In particular, the solute binding subunits of periplasmic galactoside transporter (MglB) and D-ribose-binding periplasmic protein (RbsB), and the corresponding receptor, methyl accepting chemotaxis protein (MCP) Trg.

All four periplasmic binding proteins (PBP) involved in bacterial chemotaxis, which are associated with the ligands maltose, D-ribose, D-galactose and dipeptides (MalE, RbsB, MglB and DppA) were significantly more abundant on acetate medium.

Although the two MCP involved in serine and peptide sensing (Tsr and Tap) did not show significant difference in abundance in acetate relative to glucose medium, the three remaining MCP were of higher abundance on acetate. In addition to Trg these include the aerotaxis receptor Aer and the receptor for attractants L-aspartate and related amino acids, Tar (Figure 3.16).

MCP are known to be methylated by CheR and demethylated by CheB in response to chemical stimuli. Attractants increase the level of methylation, while repellents decrease the level of methylation (Sourjik 2004). A multi pass search for these modifications was performed but no modified peptides were identified. This was not unexpected, considering the lack of possible attractants in the culture media. The relative concentrations of CheR and CheA did not show significant differences between acetate and glucose grown bacteria. Only for the response regulator CheY, TMT-ratio indicated a significantly elevated level on acetate medium ($\log_2 (A/G) = -0.57$). The second response regulator CheB and CheZ showed moderate differences in abundance which contribute to a significant, overall shift of the twenty chemotaxis proteins towards higher abundance on acetate ($q = 1.56E-4$; mean $\log_2 (A/G) = 0.6$).

A significant shift towards higher relative abundance on acetate was also observed for the proteins involved in flagellar assembly ($q = 1.32E-5$; mean $\log_2 (A/G) = 0.43$) and motility ($q = 2.84E-6$; mean $\log_2 (A/G) = 0.6$). This is plausible, since chemotaxis of *E. coli* is functionally dependent on motility (Figure 3.16). Accordingly, the increase of *E. coli* motility and the induction of chemotaxis

gene expression at the transcript level were previously observed as a long-term adaptation to acetate in complex medium (Polen, Rittmann et al. 2003). Furthermore, the products of both acetate metabolism pathways by acetyl CoA synthetase (Acs) and acetate kinase (Ack-Pta) have been shown to modulate the direction of flagellar rotation, possibly by direct modification of the response regulator CheY by phosphorylation or acetylation (Barak, Abouhamad et al. 1998). However, as with the MCP proteins, no post-translational modifications could be detected by multi pass search.

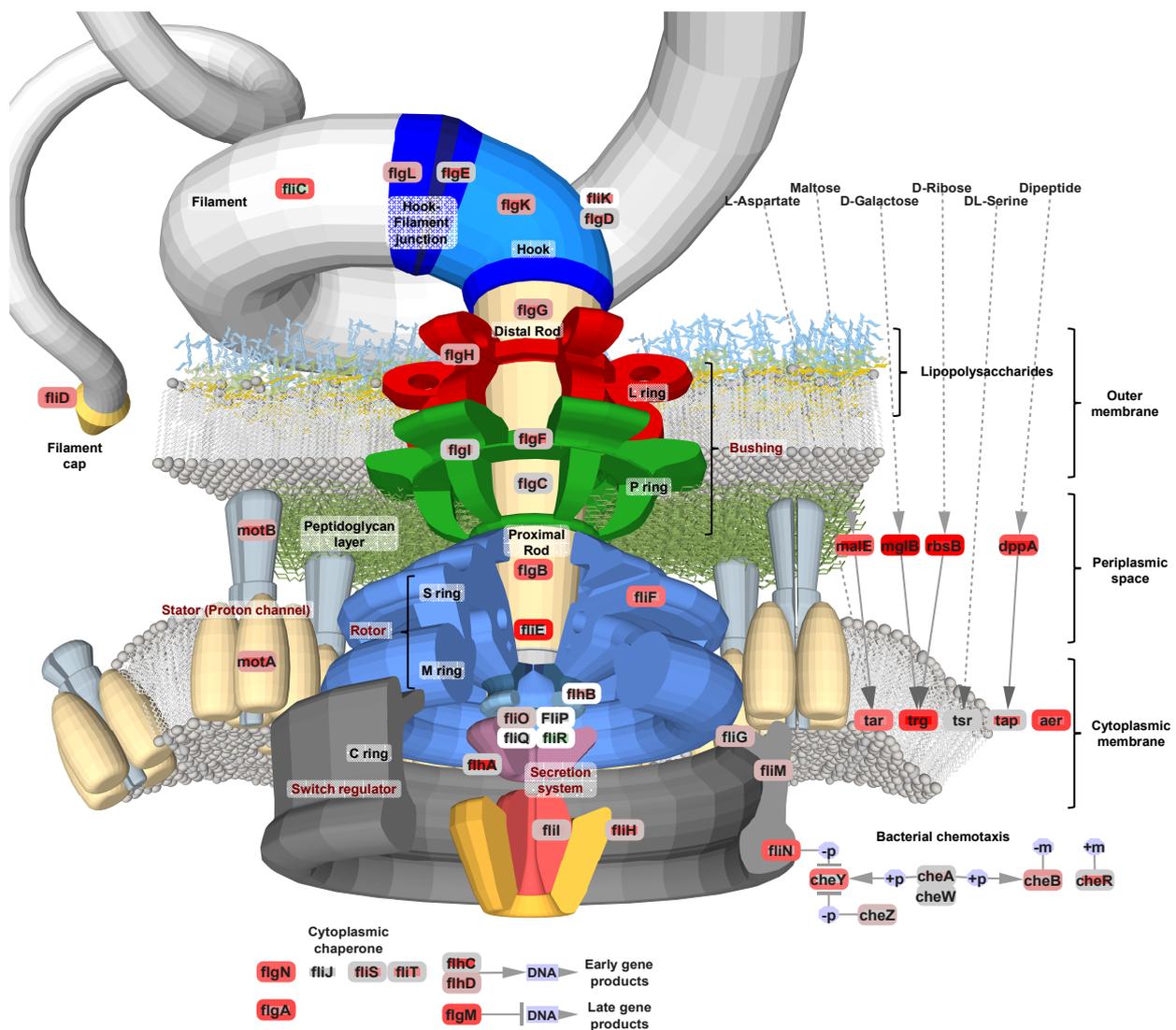


Figure 3.16: Flagellum and bacterial chemotaxis. Proteins are represented by encoding genes (common gene names from UniProt), node colors indicate relative expression values $\log_2 (A/G)$ at the protein (outside) and transcript level inside (Liu, Durfee et al. 2005). The proteins involved in bacterial chemotaxis are shown at the right panel. Periplasmic binding protein (periplasmic space) and methyl accepting chemotaxis proteins (inner membrane) are shown with their respective substrates.

3.4.6.6 Periplasmic binding proteins and ABC-transporters

Of the 199 proteins allocated to the periplasmic space by gene ontology of cellular compartment annotation, 134 were covered by the dataset. All but two of the 44 differentially abundant proteins were higher abundant on acetate, resulting in a significant enrichment of the category in the higher abundant proteins ($q < 3.6E-6$; $EF = 2.3$; $\text{mean } \log_2(A/G) = 0.4$) and depleted in proteins of lower abundance ($q < 9.4E-4$; $EF = 0.14$). Furthermore, a significant number of the elevated periplasmic proteins belong to the superfamily of periplasmic binding domains of ATP-binding cassette (ABC) transporters. In total 37 periplasmic binding proteins (PBP) were quantified (Figure 3.17).

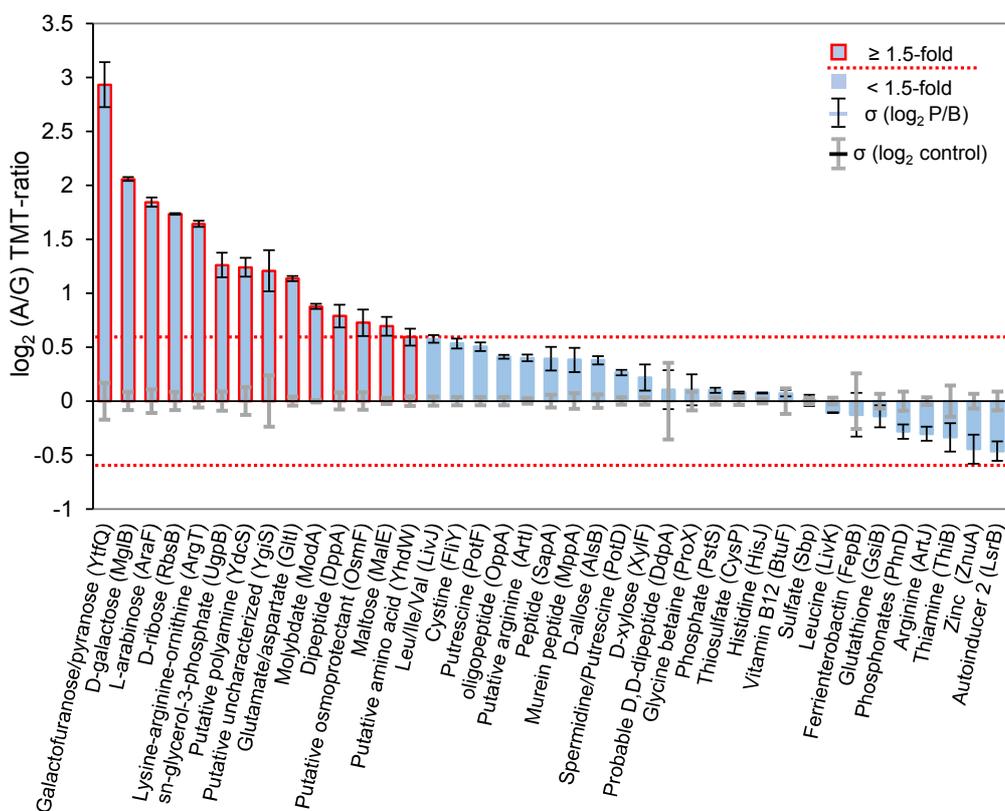


Figure 3.17: Quantification of periplasmic binding proteins. Proteins are indicated by specific ligands and protein names in parenthesis (UniProt). Columns show \log_2 transformed acetate/glucose (A/G) protein-ratios and the corresponding geometric standard deviations (σ). The grey error bars around zero represent the standard deviations of the respective control-ratios (TMT-ratios between replicate treatments, A/A and G/G). The dotted red line shows the ≥ 1.5 -fold change cutoff.

Including those directly involved in chemotaxis mentioned earlier, 15 PBP were significantly more abundant in bacteria growing on acetate and nine displayed TMT-ratios indicating more than two times the concentration observed in bacteria grown on glucose. The PBP showing the highest ratios were all specific for carbohydrate ligands. Two of these were galactose binding proteins, MglB and YtfQ in addition to the L-arabinose binding protein (AraF) and the D-ribose binding

protein RbsB. Other differentially expressed PBP were shown to bind peptides, amino acids or their precursors.

Contrary to the high number of PBP detected, a comprehensive quantification of the ATP binding proteins and in particular the membrane components of the corresponding ABC transporter complexes was not possible. A number of these proteins were not identified at all, or only by a limited number of peptides. Consequently, less evidence to support relative quantitation was acquired and only in two cases did these proteins show abundance ratios in agreement with their respective PBP.

Besides the proteins involved in ribose taxis (RbsB, Trg), the products of the *rbsDACBKR* operon encode the proteins necessary for the uptake and conversion of D-ribose for use in the pentose phosphate pathway and central carbon metabolism. The ribose ABC transporter (RbsABC), ribose pyranase (RbsD) and ribokinase (RbsK) were all strongly elevated; the sole exception was the dual regulator RbsR, which was not detected.

The second case was for galactose transport system. In agreement with the ratio observed for MglB, the level of ATP binding subunit MglA of galactose/methyl-galactoside ABC transporter was increased on acetate, while the third, transmembrane component of the MglABC transporter complex was not covered by the dataset.

The relative excess of PBP for a wide range of substrates in spite of the general limitation on protein synthesis suggests that *E. coli*, growing slowly on acetate, cannels a considerable amount of resources to scavenging proteins in preparation for fast adaptation towards a favorable change in living conditions and the appearance of alternative food sources.

3.4.6.7 Phosphotransferase systems

The protein level difference observed between *E. coli* grown on acetate compared to glucose also includes several components of Phosphotransferase systems (PTS). Of the 44 proteins annotated as PTS components, a total of 22 were encompassed in the quantitative dataset. A significant number were differentially abundant ($q = 9.08E-3$; $EF = 2.45$; $\text{mean } \log_2(A/G) = -0.14$). However, in contrast to most of the other enriched functional groups, PTS components did not show a clear trend towards higher (5 proteins) or lower abundance (8 proteins).

All proteins encoded by galactitol metabolism operon *gatYZABCD* were significantly higher in abundance. This includes the Galactitol PTS permease subunits (GatA, GatB and GatC), the

galactitol-1-phosphate dehydrogenase GatD, in addition to the subunits of tagatose-1,6-bisphosphate aldolase 2 (GatZ and GatY). Two additional proteins with a significantly higher abundance in acetate were the cellobiose/chitobiose PTS permease subunit ChbB ($\log_2 (A/G) = 0.9$) and, to a lesser extent, the N-Acetyl-D-glucosamine PTS protein NagE ($\log_2 (A/G) = 0.3$).

Most other identified components of PTS were of lower abundance on acetate, including phosphoenolpyruvate-protein phosphotransferase (PtsI) and proteins of Glucose- (PtsG), Mannose- (ManZ, ManX) and Fructose- (FruA, FruB) specific PTS. In this regard, relative quantitative results for PTS clearly contrast with those observed for ABC transporters in general and in particular their respective periplasmic binding proteins.

3.4.7 Implications towards competitive environments

Several proteins, which were of higher abundance on acetate, can be associated with a response to natural competitive environments (e.g. the mammalian intestine; where available nutrients include short-chained fatty acids). This supports the hypothesis of an adaptive program modulated to such growth conditions with possible implications towards pathogenic *E. coli* strains. One example is the higher abundant periplasmic protein ecotin (Eco). Eco is an inhibitor of trypsin, chymotrypsin, elastase and other pancreatic serine-type endopeptidases, which does not affect *E. coli* peptidases. Due to this functional specificity, it has been suggested that the inhibitor plays a role in the protection against proteases in environments such as the mammalian gastrointestinal tract (Chung, Ives et al. 1983). The study of interactions between commensal or harmful *E. coli* strains and host organisms, and also the differences between beneficial and pathogenic strains, would be a worthwhile avenue of further research. In the same context, structural and functional differences on the cell surface of *E. coli* would be of interest, as these represent the interface with possible host organisms. The metalloprotease YebA, which is involved in cell-wall biogenesis, and UspF, a stress protein which is reported to promote adhesion at the expense of motility (Nachin, Nannmark et al. 2005), were of higher abundance on acetate. However, as these observations do not give insights into changes in the composition of the outer membrane, further analysis of the lipopolysaccharide of the outer leaflet would be necessary to draw meaningful conclusions. Such analyses would also profit from the inclusion of other *E. coli* serotypes as the cell wall of K12 strain MG1655 is restricted to core polysaccharides and does not produce O-antigens attached to the LPS core (Stevenson, Neal et al. 1994).

3.5 Conclusion

The growth of *E. coli* on acetate compared to glucose as the sole carbon source results in extensive differences in the bacterial proteome. Overall, the low energy yield of acetate leads to a significant decrease in the growth rate accompanied by an overall decrease in protein content of the bacteria. In this context, a reduction of components of protein synthesis and processing could be observed. Due to the overall reduced protein content, the vast majority of proteins are of lower abundance in bacteria on acetate compared to glucose and only a very small number of proteins are of higher abundance. Although this discrepancy is remarkable, for the purpose of system wide proteome or transcriptome analyses, quantitative data is generally median normalized. This process balances the overall protein or mRNA content and introduces a systematic error, leading to an inconsistency that has to be kept in mind about results of comparable analyses.

The correlation between gene expression at the transcript and protein level data showed a determinant role of transcriptional control in programs for acetate or glucose consumption. A comparison of our dataset with a recently published proteomic analysis (Schmidt, Kochanowski et al. 2015), using a different *E. coli* strain showed a high correlation for the majority of proteins. However, there were also significant differences. For example, the protein level difference between acetate compared to glucose-grown bacteria was contrasting for proteins associated with locomotion and the increased abundance of glyoxylate bypass was significantly stronger in one strain.

In addition to the slow growth rate and reduction of protein synthesis, multiple changes can be attributed to different metabolic alignment of *E. coli* on acetate. On glucose, catabolite repression prevents the use of many alternative substrates while on acetate, the bacteria crave any superior nutrient. This generalization can account for a number of the observed protein level differences between glucose and acetate fed bacteria. However, some of the observed changes could also be interpreted as a metabolic shift of *E. coli* towards nutrients associated with one of its natural niche environments e.g. the mammalian colon, and carbon sources such as short chain fatty acids (or acetate) and sugar residues like N-acetyl glucosamine or D-galactose. A number of differences between glucose metabolizing *E. coli* and those growing on acetate apply to transporters or the enzymes required to initiate uptake and utilization of metabolites. Examples of this are the first protein of fucose and mannose metabolism (FucI), enzymes initiating hexuronate and D-galacturonate degradations (UxaC and UxuB), and the galactose ABC-transporter.

ABC transporters were generally upregulated on acetate. Even transporters and periplasmic binding proteins for nutrients not provided under the tested conditions were affected, pointing to their role in scavenging and response to environmental changes. This includes proteins involved in scavenging extracellular resources such as sugars (galactose, ribose, maltose, xylose and arabinose), amino acids and peptides or vitamins and ions.

In contrast to ABC transporters, in particular the respective periplasmic binding proteins, PTS were not collectively up-regulated. Instead, PTS for high yield carbon sources like glucose or fructose were lower abundant on acetate, whereas PTS for cellobiose and galactitol were of higher abundance.

It also has to be noted, that the use of the common laboratory strains of *E. coli* has several limitations as they may differ considerably in their response to artificial stimuli compared to strains interacting directly with the environment. To exemplify, all the K-12 substrains are derive form an isolate that has been cultivated under laboratory conditions since 1922. In the course of subsequent passages it has not only lost its O-antigen but probably accumulate a range of mutations which improved adaptation to a laboratory environment (Stevenson, Neal et al. 1994).

To conclude, this work provides a comprehensive quantitative dataset of *E. coli* proteins on different media delivering important information in terms of metabolic organization, physiology and regulatory processes under the specified conditions. However, as with the procured data, the conclusions drawn from such information alone are unidimensional. The integration of additional data is necessary to deepen the understanding of the cellular processes. At the protein level, information about the absolute abundance, localization, function, post-transcriptional regulation and activity of said proteins is necessary to fully explain the processes that they drive. The same information about the underlying genetic material and transcripts as well as the pool and distribution of available metabolites and metabolic fluxes would give a more multifaceted picture of the workings of the organism.

4. Identification and characterization of N-acyl homoserine lactones by mass spectrometry

4.1 Abstract

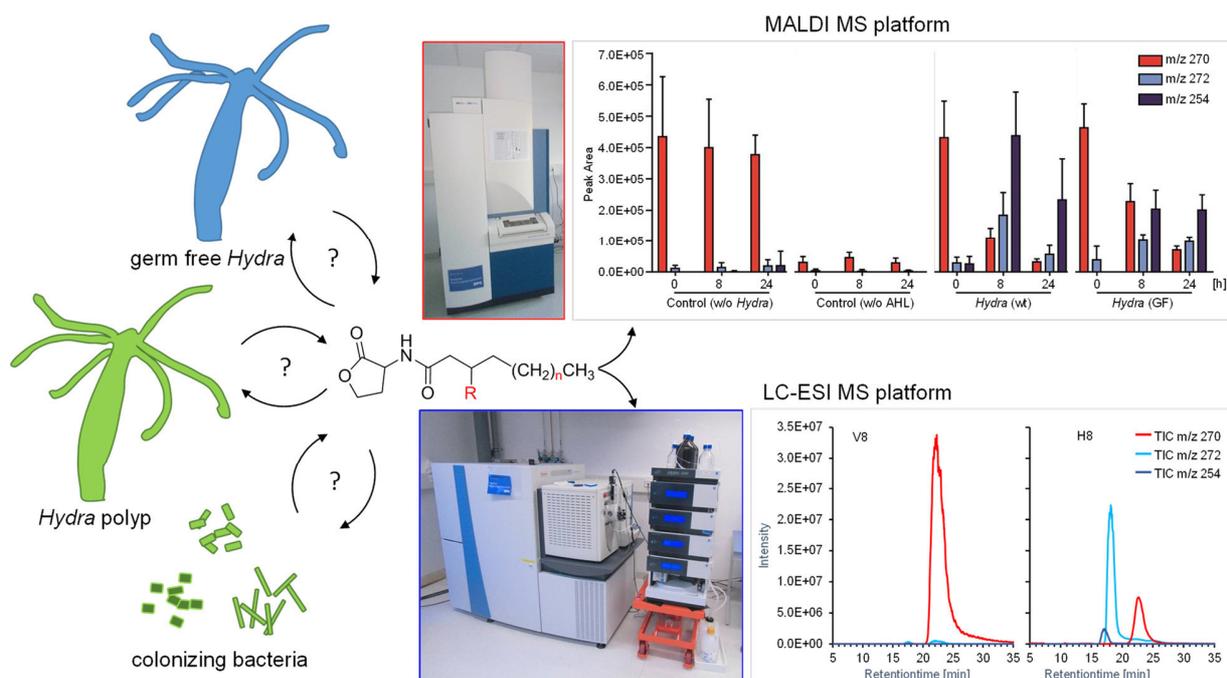


Figure 4.1: Graphical Abstract. Analysis of the N-acyl homoserine lactone based *quorum* modulation network of the *Hydra vulgaris* metaorganism by mass spectrometry.

In natural environments, no organism lives in isolation and the processes that govern the communal behavior of humans, animals, plants and microbes have become a new frontier for life sciences in recent years. In particular, the idea of the metaorganism or holobiont as the sum of all organisms, including the microbes and viruses in interdependent cohabitation with a higher order host organism, is of increasing interest.

Our collaboration partners Dr. Cleo Pietschke and Dr. Sebastian Fraune have studied the freshwater polyp *Hydra* as a model metaorganism (Zoologisches Institut der Christian Albrechts Universität zu Kiel; principal: Prof. Dr. Dr. h.c. Thomas Bosch). The goal of this project was to investigate a *quorum* sensing and *quorum* quenching network between *Hydra* and its communal bacteria as one aspect of interspecies interactions. MALDI MS and ESI MS based bioanalytical methods were established focusing on the class of N-acyl homoserine lactone (AHL)

autoinducers. To facilitate the identification of AHL in biological samples, a spectral library of the AHL standards was created by LC-ESI MS, which includes the spectral data and retention-time information of 14 commercial AHL compounds. The fragmentation of these AHL in MS/MS experiments was extensively characterized to further allow the analysis of fragmentation patterns of unknown AHL derivatives from biological samples.

Using these data, the *quorum* quenching activity of *Hydra vulgaris* polyps was analyzed by LC-ESI MS and MALDI MS and the resulting *quorum* quenching product were identified. Furthermore, the product spectrum of two AHL-synthases isolated from *Curvibacter*, the major colonizing bacterium of the *Hydra* holobiont were analyzed by LC-ESI MS, upon heterologous expression of the enzymes in *E. coli*. (Figure 4.1).

4.2 Introduction

4.2.1 Quorum sensing

Quorum sensing describes the intercellular signaling ability, employed by bacteria, to collectively adapt to changes in their environmental population density. The density of the culture is conveyed by the accumulation of a class of constitutively, but very slowly produced and freely cell-permeating signal molecules, which are called autoinducers. *Quorum* sensing systems include the proteins for production of a restricted range of autoinducers and receptors or sensors. The latter selectively detect the autoinducer and respond to the stimulus by threshold dependent changes in bacterial gene expression.

Controlled processes generally determine the communal behavior of the microbes. These often include genes associated with pathogenicity, e.g. those involved in motility vs. adhesion and biofilm formation, as well as genes influencing the growth rate and metabolic states of the microbe population.

N-acyl homoserine lactones (AHL) are a class of autoinducers employed by many gram-negative bacteria for *quorum* sensing. The general chemical structure of AHL consists of a polar homoserine lactone and a hydrophobic acyl-chain attached via a carboxamide bond to the homoserine lactone nitrogen. AHL differ in the length of the side-chain and the substituent at the third carbon (C₃-position) in the side-chain (Figure 4.2). This provides a spectrum of signal molecules, which can be used by different bacterial strains and species for independent and interdependent *quorum* sensing networks in xenic cultures.

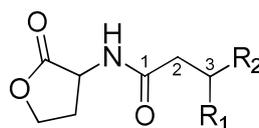


Figure 4.2: General structure of N-acyl-homoserine lactones. The homoserine lactone (α -amino- γ -butyrolactone) connects to the acyl side-chain by a carboxamide bond. Numbers indicates side-chain carbons C₁-C₃. The substituent at C₃ (R₁) can be a hydrogen or a hydroxy- or carbonyl-function. R₂ is a saturated or singly unsaturated and unbranched alkyl-chain. Total chain lengths of 4 to 18 carbons have been described (Lade, Paul et al. 2014).

The first *quorum* sensing system identified, and thoroughly characterized, is responsible for the cell density dependent control of bioluminescence in the marine bacterium *Vibrio fischeri* (Nealson, Platt et al. 1970, Eberhard 1972, Eberhard, Burlingame et al. 1981). On the basis of the mechanism discovered in *V. fischeri*, AHL autoinducer *quorum* sensing is classically defined as a LuxI/LuxR type *quorum* sensing system (Fuqua, Winans et al. 1996). The proteins involved belong

to the LuxI and LuxR family. LuxI enzymes are AHL-synthases and the LuxR family proteins function as AHL-receptors and transcriptional activators. In *V. fischeri*, the LuxR-AHL complex binds to the *lux* promoter and induces transcription of the *luxICDABEG* operon which encodes the proteins necessary for bioluminescence (Engebrecht and Silverman 1984). Thereby, *luxI* is further expressed in a positive feedback loop and in this context AHL signal molecules are named as autoinducers.

AHL-synthases catalyze the production of AHL from S-adenosyl-methionine (SAM), the precursor for the lactone moiety, and specific acyl-acyl carrier proteins (acyl-ACP) which determine the side-chain (Sitnikov, Schineller et al. 1995, More, Finger et al. 1996) (Figure 4.3).

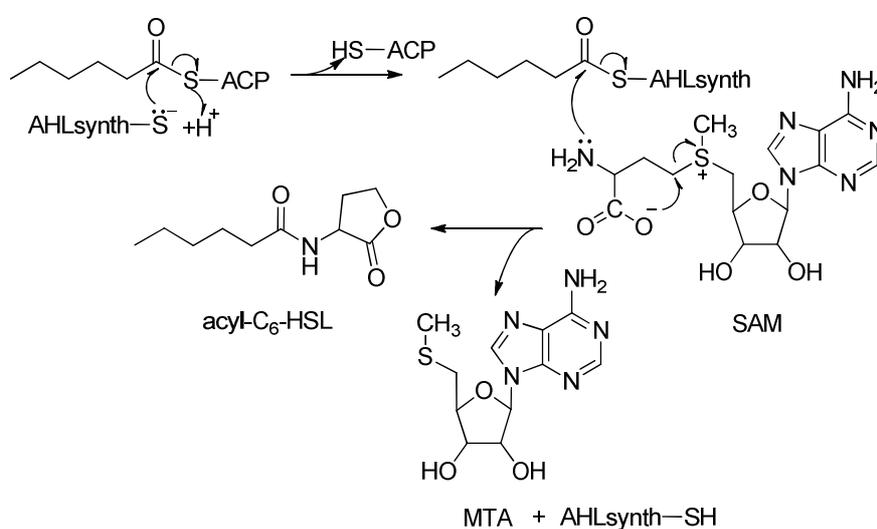


Figure 4.3: The mechanism for AHL synthesis proposed by Sitnikov et al. 1995. The hexanoyl-group is transferred from an acyl-carrier protein to another cysteine residue in the active site of the AHL-synthase by transthioesterification. The acyl-carrying AHL-synthase then catalyzes formation of the carboxamide bond between the amino-function of S-adenosyl-methionine (SAM) and the first carbon of the activated fatty acid, followed by lactone ring-formation. Acyl-C₆-HSL and methylthioadenosine (MTA) are produced and the enzyme is released. Modified from Sitnikov, Schineller et al. (1995) and More, Finger et al. (1996).

SAM is a common cosubstrate for a number of metabolic processes. Acyl-ACPs originate from fatty acid metabolism. The synthesis of fatty acids by successive addition of the C₂-groups bound by malonyl-coenzyme A, is regarded as the reason why the majority of AHL compounds are restricted to side-chain lengths with an even numbers of carbons. However, while not extensively documented, AHL with odd acyl-chain lengths have been isolated as well (Horng, Deng et al. 2002, Morohoshi, Inaba et al. 2004, Brader, Sjoblom et al. 2005, Sio, Otten et al. 2006, Ortori, Atkinson et al. 2007).

AHL acylases have been isolated from many bacteria and eukaryotes. These enzymes destroy AHL signals by hydrolyzing the amide bond between the acyl side-chain and homoserine moiety. Bacteria can utilize the resulting homoserine as a nitrogen source and the fatty acid can be metabolized (Figure 4.4 C).

AHL lactonases hydrolyze the homoserine lactone ring. In contrast to the amide bond cleavage catalyzed by the AHL-acylases, the lactone hydrolysis reaction is reversible under acidic conditions (Dong, Xu et al. 2000). Several enzyme families with AHL lactonase activity by different catalytic mechanisms have been identified in both bacteria and eukaryotes (Figure 4.4 B).

AHL oxidoreductases, which catalyze the reduction of carbonyl-groups substituent at the C₃-position to hydroxy-function, have been identified in bacteria (Uroz, Chhabra et al. 2005) and could be isolated from a soil metagenome sample (Bijtenhoorn, Mayerhofer et al. 2011) (Figure 4.4 D). A cytochrome P450 monooxygenase from *Bacillus megaterium* has been described to hydroxylate AHL at the terminal (ω -1, ω -2 or ω -3) side-chain positions (Chowdhary, Keshavan et al. 2007) (Figure 4.4 E).

Overall, the vast majority of *quorum* quenching enzymes discovered are of microbial origin, with the exception of the paraoxonase (PON) family of AHL lactonases found in human epithelial cells (Chun, Ozer et al. 2004) and an acylase found in the porcine kidney (Xu, Byun et al. 2003).

4.2.3 Bioanalytics of N-acyl homoserine lactones

The two most successful methods for the analysis of AHL autoinducers are bioassays using bacterial reporter strains for AHL detection and mass spectrometry for AHL identification. AHL sensitive bioassays employ genetically modified bacteria. These reporter strains harbor easily observed, chromogenic or bioluminescence cytochemical systems under gene expression control of AHL receptors. The receptors introduced in such strains can be relatively unspecific, detecting a range of several different AHL signals (Ravn, Christensen et al. 2001, Zhu, Chai et al. 2003). Although reporter strains are extremely sensitive, a limitation to this method is that specific AHL may not generate a response with the used receptor. Therefore, multiple strains may be needed for complex AHL mixtures. This substantially increases the workload. Furthermore, the quantification capabilities using this method are limited as the response strength may also be biased depending on the autoinducer-receptor combination. Nonetheless, biosensor strains have been used to successfully detect and characterize autoinducer activity in the extracts of bacterial cultures supernatants, in particular combined with reversed-phase thin layer chromatography (RP-TLC). In this approach, developed by Shaw et al. (1997), TLC is used to separate the AHL in

biological samples based on the length and saturation of their acyl side-chain and the substituent at C₃. The TLC sheet is then coated with a solid culture medium, seeded with a reporter strain biosensor for visual detection. Poor resolution of AHL separation on TLC makes the unambiguous identification of spots by comparison with AHL standards difficult. However, the reporter strain showed a linear response to declining amounts of specific AHL over concentration ranges of at least two orders of magnitude. This behavior enables AHL quantification, provided a dilution series for each AHL of interest is used for calibration (Shaw, Ping et al. 1997).

To identify AHL, complementary techniques for structural analysis are often necessary. Using mass spectrometry, most AHL can be identified by the mass/charge ratio (m/z) of their protonated molecular ions and characteristic fragmentation patterns (Gould, Herman et al. 2006). However, in addition to suitable separation techniques, more extensive sample cleanup is necessary prior to analysis (Frommberger, Schmitt-Kopplin et al. 2004). Among the numerous methods available, chromatography has dominated the separation techniques utilized in combination with mass spectrometric analysis. Both gas chromatography-MS (Charlton, de Nys et al. 2000, Cataldi, Bianco et al. 2004) and LC-MS (Pearson, Gray et al. 1994, Marketon, Gronquist et al. 2002, Orsatti, Di Marco et al. 2002, Morin, Grasland et al. 2003, Teplitski, Eberhard et al. 2003) have been successfully used for the analysis of AHL. For HPLC separation of homoserine lactones similar conditions as used for fatty acids and peptides are utilized: a reversed-phase (RP) column with stationary phase of C₈- to C₁₈- and binary solvent systems containing water and a polar organic solvent, most frequently acetonitrile or methanol to elute the compounds (Spencer, Bertrand et al. 2007, Masuo, Terabayashi et al. 2010, Ruiz-Rodriguez, Reglero et al. 2010). The LC-MS systems most extensively applied to the analysis of AHL mixtures rely on a HPLC system with C₁₈ stationary phase and methanol/water mobile phase, coupled to ion-trap mass spectrometry in positive ion mode. However, wide varieties of both positive and negative ion MS techniques and MS instruments have been used successfully. These also include high resolution MS to achieve mass accuracies within 2 ppm (Spencer, Bertrand et al. 2007) and multiple reaction monitoring experiments with triple-quadrupole MS instruments to provide at least semi quantitative data for several AHL with a single isotope labelled internal standard (Gould, Herman et al. 2006).

4.3 Material and methods

Dr. Cleo Pietschke and Dr. Sebastian Fraune (Zoologisches Institut der Christian Albrechts Universität zu Kiel) performed cultivations and transformation of bacteria, the extraction of AHL from cultures, as well as *Hydra quorum* quenching assays.

The chemicals used in LC and MS experiments and the commercial AHL were acquired from Sigma-Aldrich (<http://www.sigmaaldrich.com>).

The names of AHL were adjusted from conventional IUPAC (International Union of Pure and Applied Chemistry) nomenclature: For the two substitutions at the third carbon (C₃-position), the following abbreviations were used: “3O-“ for the carbonyl-function; “3OH-“ for the hydroxy-function. This was followed by the acyl-chain length “C_n” (with n = number of carbons). In case of unsaturated acyl-chains, “C_{n:1}” denotes a single double bond in the acyl-chain. The location and configuration of double bonds were disregarded due to the limitations of the structural information gained by the methodology applied. Finally, “homoserine lactone” is abbreviated as “HSL” at the end (see Figure 4.5).

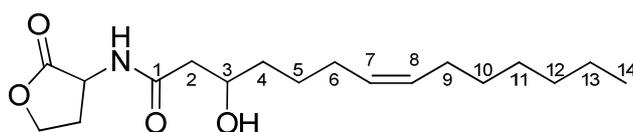


Figure 4.5: Structure named 3OH-C_{14:1}-HSL. The structure of the molecule, (7Z)-3-Hydroxy-N-[(3S)-tetrahydro-2-oxo-3-furanoyl]-7-tetradecenamide by IUPAC convention is shown, which is hereafter abbreviated as “3OH-C_{14:1}-HSL”. Acyl side-chain carbons are numbered.

4.3.1 Analysis of N-acyl homoserine lactones by direct injection ESI MS

Commercial AHL were first used to establish suitable MS-parameters for ESI MS by direct (off-line) injection of samples into an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific; <https://www.thermofisher.com>). AHL samples were lyophilized and then dissolved in methanol to a concentration of 10 nmol/μL. C₄-; C₆-; C₈-; C₁₀-; and C₁₂-HSL samples were diluted 1:1 with deionized water and directly used for off-line measurement in negative or positive ion mode. Each sample was loaded into a PicoTip emitter and injected via a nanospray ion source into the mass spectrometer with a spray voltage of 1.1 kV and a capillary temperature of 250 °C. Automatic gain control (AGC) target for the Orbitrap mass analyzer was set to 1E6; the microscan number and maximum ion inject time were set to 2 and 500 ms, respectively. Full FTMS (Fourier transform mass spectrometry) scans of different mass ranges from lower boundary of 50 - 100 *m/z* to upper boundary of 300 - 500 *m/z* were acquired with a resolution of 30,000. Intense signals corresponding to AHL *m/z* values were manually selected for fragmentation using collision induced

dissociation (CID) and higher energy collisional dissociation (HCD). MS/MS spectra were acquired using the Fourier transform (Orbitrap) and ion trap (radio frequency trap) mass analyzers, testing a range of normalized collision energies (NCE) from 0-95.

4.3.2 Analysis of the *quorum* quenching mechanism of *Hydra* polyyps

To test *Hydra* for *quorum* quenching activity, 500 μL of a 100 μM 3O-C₁₀-HSL solution was incubated with ca. 50 germfree or wild type *H. vulgaris* (AEP) polyyps. The assay was performed in ten replicates with sterile and non-sterile *Hydra* polyyps. As negative controls, AHL solutions were also incubated with (non-sterile) Volvic table water or sterile S-medium and *Hydra* polyyps were incubated without AHL addition. Incubation times of 0, 8 and 24 h were tested. For AHL extraction, *Hydra* polyyps were homogenized with a pestle and extracted with 1:1 (v/v) acidified ethyl acetate (0.01 % acetic acid). The samples were shaken for 10 min at room temperature and centrifuged for 10 min at 3,600 g (18 °C). The organic layers were combined and stored at -80 °C until further use.

4.3.3 Analysis of *quorum* quenching samples by MALDI-TOF MS

AHL extracts were lyophilized, resuspended in 50 μL acetonitrile and 1 μL was spotted onto on a 384-well Opti-TOF MALDI Insert (ABSciex; <https://sciex.com>) with 1 μL CHCA matrix solution (10 mg/mL α -cyano-4-hydroxycinnamic acid in 60 % ACN, H₂O). MS was performed on a TOF/TOF 5800 mass spectrometer (ABSciex), accumulating 2,000 shots over an m/z -range from 100 m/z to 500 m/z . The instrument was calibrated with an AHL standard solution (acyl-C₄-; acyl-C₆-; acyl-C₈-; acyl-C₁₀- and acyl-C₁₂-HSL) to calibrate the prospective m/z range. MS spectra were manually evaluated for potential AHL degradation products and an inclusion list containing targets for MS/MS acquisition was assembled. Fragment spectra were acquired with a medium CID pressure of 2E-6 Torr, minimum signal-to-noise filter of 10 and a precursor mass tolerance ± 200 ppm.

4.3.4 Analysis of *quorum* quenching samples by LC-ESI MS

LC-MS analysis was performed on the control samples (in Volvic medium V0 and V8) as well as *Hydra* incubated with AHL (H0 and H8) using a Dionex Ultimate 3000 HPLC system, equipped with a Dionex Acclaim PepMap100 nano-column (75 μm x 15 cm, 3 μm , 100 Å; Thermo Fisher Scientific; <https://www.thermofisher.com>). The LC was coupled to an LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific). Reversed-phase HPLC separation was performed with

a gradient of Eluents A (Milli-Q water with 0.1 % formic acid) and B (80 % acetonitrile, 0.1 % FA). Five μL of each sample were injected and washed for 5 minutes on a Dionex PepMap C₁₈ guard column (300 μm x 10 mm; Thermo Fisher Scientific) with 0.1 % aqueous TFA at a flow-rate of 30 $\mu\text{L}/\text{min}$. The samples were then separated by isocratic elution with 55 % B for 50 min at a flow-rate of 250 nL/min, followed by a sharp increase to 95 % eluent B in 1 min and a 5 min washing step. The column was then equilibrated with 5 % B for 15 min. UV detection was performed at 214 nm. HPLC-flow was coupled to the LTQ Orbitrap Velos (Thermo Fisher Scientific) using a nanospray ion source with 1.3 kV capillary spray voltage and 210 °C capillary temperature with a 30 μm PicoTip emitter (New Objective; <http://www.newobjective.com>). After a delay time of 5 min, full scans of 50 m/z to 500 m/z range were recorded in positive ion mode with a resolution of 30,000 (AGC target 1E6; maximum inject time 500 ms, preview mode for FTMS master scans and Wideband activation were used). Top 5 precursors with a minimum signal intensity of 400 and default charge state of 1, were selected for MS/MS acquisition from targets on a parent mass list. The parent mass list included 85 m/z values for 3OH-C_n-HSL, 3O-C_n-HSL, acyl-C_n-HSL and their respective singly unsaturated AHL with side-chain lengths from C₄ to C₂₀. The CID isolation window for IT-MS spectra was set to 3 Da, the AGC target was 1E4, with a maximum inject time of 400 ms, activation time of 10 ms. The normalized collision energy (NCE) for CID was set to 28. HCD-FT MS/MS scans were acquired for parent masses on the inclusion list with a minimum signal intensity of 400 with an AGC target of 1E5 and maximum inject time of 500 ms. The isolation width was 1 Da and NCE values between 35 and 37 activation time of 0.1 ms. The mass range was fixed at first m/z of 50, the highest m/z varies depending on precursor mass.

4.3.5 Heterologous expression of *Curvibacter* sp. AHL-synthases

Dr. Cleo Pietschke and Dr. Sebastian Fraune (Zoologisches Institut, Christian Albrechts Universität zu Kiel) performed these experiments.

The AHL-synthases from *Curvibacter* sp. (Curl1 and Curl2) were cloned into the pET22b-vector and expressed in *E. coli* Rosetta 2 (DE3) pLysS, as described previously (Gould, Herman et al. 2006). *E. coli* transformed with either the expression constructs or the control vector, were grown in 5 mL Luria-Bertani (LB) broth supplemented with 100 $\mu\text{g}/\text{mL}$ ampicillin and 34 $\mu\text{g}/\text{mL}$ chloramphenicol at 37 °C while shaking (250 rpm) for 12 to 16 h. Each culture was diluted 1:50 into 50 mL fresh LB broth with 100 $\mu\text{g}/\text{mL}$ ampicillin and 34 $\mu\text{g}/\text{mL}$ chloramphenicol. After incubation at 37 °C and (250 rpm) to an optical density (OD₆₀₀) of 0.6 to 0.8 was reached, afterwards the temperature was lowered to 25 °C. Cultures were induced by addition of 0.5 mM IPTG (isopropyl- β -D-thiogalactopyranoside) and incubated for 6 h at 25 °C.

For mass spectrometric sample preparation, bacterial cultures were pelleted by centrifugation at 3,600 *g* for 10 min at 18 °C. The supernatants were aspirated by 24 mL syringes and passed through 0.2 µm filters. Samples were extracted with one volume of acidified ethyl acetate (0.01 % acetic acid). The samples were shaken for 10 min at room temperature, centrifuged at 3,600 *g* for 10 min at 18 °C. The organic layer was collected and completely dried with a rotating evaporator in a 50 mL glass flasks, then stored at -80 °C until further use.

4.3.6 Identification of AHL by LC-ESI MS

Ethyl acetate extracts from bacterial culture supernatants were further purified by solid-phase extraction (SPE) as described by Gould et al. (Gould, Herman et al. 2006). Each sample was dissolved in 500 µL methanol, 5 mL isooctane-ethyl ether (1:1 v/v) was added and then applied to an activated Sep-Pak silica cartridge (3 cc Vac RC Cartridge with 500 mg Sorbent; Waters; <http://www.waters.com>), attached to a vacuum manifold. The sorbent material was conditioned and equilibrated by successive washings with 6 mL of the solvents isooctane-ethyl ether (1:1 v/v), then with acidified ethyl acetate (0.01 % acetic acid) and again with isooctane-ethyl ether. SPE cartridges were washed twice with 5 mL of isooctane-ether and then 5 mL of acidified ethyl acetate were used to elute AHL into a 10 mL glass reaction vial. The purified samples were lyophilized to dryness, dissolved in 100 µL of 0.1 % aqueous TFA and transferred to a HPLC vial. Finally, the samples were analyzed by LC-ESI MS using an Ultimate 3000 HPLC coupled to the LTQ Orbitrap Velos mass spectrometer.

The LC-MS conditions were adjusted from Chapter 4.3.4 to accommodate more complex mixtures and improve on LC-resolution. Separation was performed over a gradient of eluent A (Milli-Q water with 0.1 % FA) and eluent B (80 % Acetonitrile, 0.1 % FA). Five µL of each sample were injected and washed for 5 minutes on a PepMap C₁₈ guard column (300 µm x 10 mm; Thermo Fisher Scientific) with 0.1 % aqueous TFA at a flow-rate of 30 µL/min and then separated on a PepMap C₁₈ nano column (75 µm x 15 cm, 3 µm, 100 Å; Thermo Fisher Scientific) over a gradient from 5 % eluent B to 90 % eluent B in 90 min, at a flow-rate of 250 nL/min, followed by isocratic elution with 95 % eluent B for 10 min. After decrease to 5 % eluent B in 1 min, the column was equilibrated for 15 min. As described before, the HPLC was coupled to the LTQ Orbitrap Velos (Thermo Fisher Scientific). Both HCD and CID spectra were acquired by FTMS. CID-FT MS/MS scans were acquired for precursors with a minimum signal intensity of 500, other parameters were the same as for HCD-FT MS/MS scans, except for the normalized collision energy which was set to 28. HCD-FT MS/MS scan parameters were the same as described in Chapter 4.3.4.

4.4 Results and discussion

As a complementary method to biosensor reporter strains for AHL detection, mass spectrometry is the most convenient method for AHL identification. Despite this, AHL compounds are not well represented in the numerous comprehensive metabolic web libraries for GC-MS and LC-MS spectral data. As of October 2015, no reference spectra of AHL were found in any of the following databases:

- METLIN (Smith, O'Maille et al. 2005)
- *m/z* cloud (<https://www.mzcloud.org/>)
- GOLM metabolome database (Kopka, Schauer et al. 2005)
- MassBank (Horai, Arita et al. 2010)
- NIST Data Gateway (National Institute of Standards and Technology; <http://srdata.nist.gov/gateway/>)

For this reason, reference spectra for a number of commercially available AHL compounds were acquired by LC-ESI MS. This spectral library was used to characterize commonalities and differences in the MS/MS fragmentation behavior of AHL. In addition, these data were useful for the identification of AHL in biological samples by comparison of their MS/MS fragmentation and the chromatographic retention-times observed in LC-ESI MS analyses.

4.4.1 LC-ESI MS of small compounds

In a common shotgun or bottom-up proteome analysis by positive ion mode LC-ESI MS, multiply charged peptides are the default analytes and ubiquitous singly charged ions are routinely excluded from MS/MS analysis. With the exception of several polyethers and a few laboratory contaminant detergents, most interfering signals can be distinguished from peptide by their net charge, charge distribution or mass (Keller, Sui et al. 2008). Contaminants with a higher net charge occur less frequently in the *m/z* range targeted for peptide analysis. Therefore, precursors can be automatically chosen from the most intense, multiply charged signals observed in a survey MS scan.

In contrast to this, singly charged ions are the targets of ESI MS analyses of smaller compounds, i.e. most non-polymerized metabolites including hormones, drugs, and also AHL. For such analytes it is necessary to manage the high number of contaminants with a single net charge observed in ESI MS. Contrary to the data dependent selection of precursors for fragmentation

(selection based on most intense signals from preceding survey MS scans), an inclusion list has to be prepared to cover m/z values of interest. MS/MS analysis is then limited to these precursors. A combination of both approaches, choosing targets by intensity from the MS survey scan in case no targets from an inclusion list are found, is possible as well. To increase the specificity of the analysis and decrease false positives in precursor selection, inclusion list entries can be further limited to a specific HPLC retention-time window, if the retention-time of the analyte in question has been determined.

4.4.2 HPLC separation of AHL

The homoserine lactone (HSL) moiety is the same all AHL derivatives analyzed in this study. Therefore, the behavior of these AHL in reversed-phase chromatography is determined by differences in the acyl side-chain in accordance with the corresponding fatty acids. For fatty acids, the retention-time in reversed-phase chromatography increases with the length of the alkyl-chain and is reduced if double bonds are present (Cazes 2010). In this study, the same behavior was observed for AHL compounds with different acyl-chain lengths. Among AHL compounds of the same side-chain length, with differences in the oxidation state at C₃ position, the following sequence was observed for retention-times in reversed-phase chromatography: 3O-AHL derivatives (with a carbonyl-group at C₃) elute first, followed by 3OH-AHL compounds (with a hydroxy-group at C₃), while AHL with aliphatic side-chains show the longest retention-times (Figure 4.6). This is due to the strong hydrophobic interactions of unbranched aliphatic chains with the C₁₈-RP column material. The polar carbonyl and hydroxy-groups undergo weakened interactions with the stationary phase but may strongly interact with the mobile phase.

Among the commercial standards used in this study, all three possible C₃-substituent derivatives were represented for AHL compounds with the chain lengths C₈, C₁₀ and C₁₂. The retention-times of 3O-AHL and 3OH-AHL of equal chain length were similar. In particular, the respective C₈ and C₁₀ compounds were not baseline separated, although the corresponding peaks did not overlap at half maximum. With increasing chain length, resolution of 3O- and 3OH-derivatives improved. The acyl-AHL compounds showed significantly higher retention-times, however C₆-, C₈- and C₁₀-compounds co-eluted with corresponding 3OH-C_{x+2}-HSL or 3O-C_{x+2}-HSL.

Overall, peak width decreased with increasing acyl-chain length and retention-times, while the distance between peaks decreased with each additional C₂-unit and longer retention-time. Because resolution is proportional to the distance between peaks and inverse proportional to peak width, this results in an overall increasing resolution for compounds differing in a single C₂-unit.

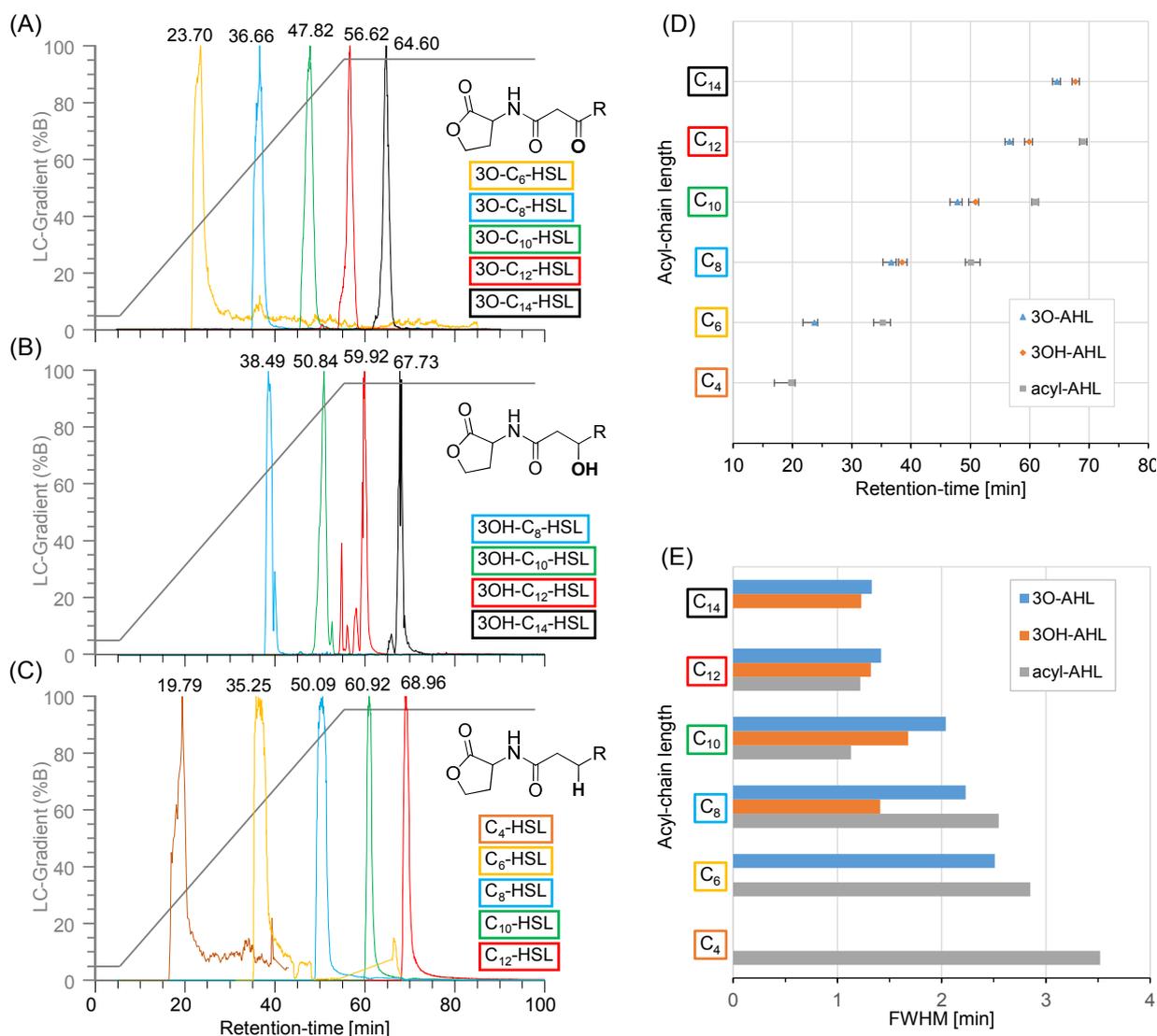


Figure 4.6: C₁₈-RP-HPLC separation of AHL standards. Extracted ion chromatograms for the 3O-AHL, 3OH-AHL, and acyl-AHL standard compounds included in the spectral library are shown in (A), (B) and (C), respectively. The AHL mixtures were separated over a C₁₈-nano-column (75 μm x 15 cm, 3 μm, 100 Å). The vertical axis shows the gradient used (eluent A: H₂O, 0.1 % FA; eluent B: 80 % ACN, 0.1 % FA; flow-rate 250 nL/min). The chromatographic distribution of analyzed AHL compounds is summarized in (D), with the full width at half maximum (FWHM) of peaks indicated by bars. Bar chart (E) shows FWHM values separately.

In chromatography, the resolution of two peaks is defined by the quotient of the difference of the two retention-times and the mean of the peak-widths (Formula 4.1).

$$\text{Formula 4.1: } R = 2 \times \frac{t_{R2} - t_{R1}}{(w_{R1} + w_{R2})} \quad (t_{R2} > t_{R1})$$

If the peak-width cannot be accurately determined, the calculation can be simplified by assuming a peak with Gaussian distribution. With this, the peak width $W = 4\sigma$ (where σ is the standard

deviation) and the peak FWHM (full width at half maximum) is $W_{0.5} = 2.354\sigma$. Using these relationships the peak width in Formula 4.1 can be substitute for the peak width at half maximum (Formula 4.2) (European Pharmacopoeia Commission 2010).

$$\text{Formula 4.2:} \quad R = 1.18 \times \frac{t_{R2} - t_{R1}}{(W_{0.5R1} + W_{0.5R2})} \quad (t_{R2} > t_{R1})$$

Resolution of two peaks is needed for their identification based on their retention-times or retention-volumes. Assuming that both peak profiles are similar, a complete (baseline separation) is achieved at $R = 1.5$. Baseline separation is optimal for quantitative analysis based on peak height or peak area measurements (Fraissard 1997). The distance between two peak maxima with a resolution of 1.5 is 6σ (6σ -separation), with $R = 1$ the distance of both peak maxima is 4σ . A 4σ -separation is still sufficient for quantitation because the peak areas overlap to only 2.3 % (Fraissard 1997, Engelhardt, Steiner et al. 2005).

A selectivity term of the chromatographic separation is the separation factor (α), calculated as the quotient of the corrected retention-times (t'_R) of two peaks (t_R after subtraction of dead time, the migration-time of non-retained compounds t_0) (Formula 4.3). A high separation factor for two substances indicates a high selectivity of the chromatographic system (Cazes and Ewing 2005).

$$\text{Formula 4.3:} \quad \alpha = \frac{k_2}{k_1} = \frac{t_{R2} - t_0}{t_{R1} - t_0} = \frac{t'_{R2}}{t'_{R1}} \quad (t_{R2} > t_{R1})$$

The separation factor is identical to the quotient of the retention-factors (k) of the two eluting compounds (Formula 4.3). The retention-factor is independent of instrument parameters like the column dimensions or flow-rate and is calculated as the quotient of the corrected retention-time (t'_R) to the dead time (t_0) (Formula 4.4) (Cazes and Ewing 2005, Engelhardt, Steiner et al. 2005).

$$\text{Formula 4.4:} \quad k = \frac{t_R - t_0}{t_0} = \frac{t'_R}{t_0}$$

Chromatographic parameters for the analyzed AHL compounds are summarized in Table 4.1. The resolutions calculated pairwise for all compounds confirm the observations made from retention times and the overlap of corresponding peaks. In particular, the pairs of compounds 3OH-C₁₀-HSL and C₈-HSL as well as 3O-C₈-HSL and C₆-HSL are not resolved on the HPLC-system used here. However, these coeluting compounds are easily distinguished in systems were the LC is coupled to a mass spectrometer, by differences is the detected m/z ratios and characteristic MS/MS fragment spectra (see Chapter 4.4.3 Fragmentation of AHL in MS/MS experiments).

Table 4.1: Chromatographic properties of AHL compounds. The chromatographic properties of AHL compounds on a C₁₈-nano column (see Chapter 4.3.6 for analysis parameters) are summarized: retention-times (t_R), full peak-width at half-maximum (FWHM) and retention-factors (k) for each observed AHL peak. The resolution (calculated using Formula 4.2) and the separation factors (α) (see Formula 4.3) for each pair of AHL compounds are included.

	t _R [min]	FWHM [min]	Retention-factor (k)	Resolution														
				C ₄ -HSL	3O-C ₆ -HSL	C ₆ -HSL	3O-C ₈ -HSL	3OH-C ₈ -HSL	3O-C ₁₀ -HSL	C ₈ -HSL	3OH-C ₁₀ -HSL	3O-C ₁₂ -HSL	3OH-C ₁₂ -HSL	C ₁₀ -HSL	3O-C ₁₄ -HSL	3OH-C ₁₄ -HSL	C ₁₂ -HSL	
C ₄ -HSL	19.79	3.52	0.741		1.53	5.73	6.92	8.95	11.90	11.78	14.09	17.59	19.57	20.87	21.80	23.82	24.48	
3O-C ₆ -HSL	23.7	2.51	1.084	1.464		5.09	6.45	8.90	12.51	12.31	15.29	19.77	22.32	24.13	25.14	27.78	28.64	
C ₆ -HSL	35.25	2.85	2.100	2.836	1.937		0.66	1.79	6.07	6.49	8.12	11.81	13.96	15.22	16.57	18.79	19.55	
3O-C ₈ -HSL	36.66	2.23	2.224	3.004	2.051	1.059		1.19	6.17	6.63	8.56	12.91	15.46	17.04	18.52	21.19	22.10	
3OH-C ₈ -HSL	38.49	1.41	2.385	3.221	2.200	1.136	1.072		6.38	6.91	9.43	15.12	18.53	20.84	22.49	26.14	27.34	
3O-C ₁₀ -HSL	47.82	2.04	3.206	4.329	2.956	1.526	1.441	1.344		1.17	1.92	6.00	8.50	9.75	11.75	14.37	15.30	
C ₈ -HSL	50.09	2.55	3.405	4.599	3.140	1.621	1.531	1.428	1.062		0.42	3.88	5.99	6.95	8.83	11.01	11.81	
3OH-C ₁₀ -HSL	50.84	1.68	3.471	4.688	3.201	1.653	1.561	1.455	1.083	1.019		4.40	7.14	8.47	10.79	13.70	14.75	
3O-C ₁₂ -HSL	56.62	1.42	3.980	5.374	3.670	1.895	1.789	1.669	1.241	1.169	1.146		2.84	3.98	6.85	9.89	11.03	
3OH-C ₁₂ -HSL	59.92	1.32	4.270	5.766	3.938	2.033	1.920	1.790	1.332	1.254	1.230	1.073		0.96	4.17	7.23	8.40	
C ₁₀ -HSL	60.92	1.13	4.358	5.885	4.019	2.075	1.959	1.827	1.359	1.280	1.255	1.095	1.021		3.53	6.81	8.07	
3O-C ₁₄ -HSL	64.6	1.33	4.682	6.322	4.317	2.229	2.105	1.963	1.460	1.375	1.349	1.176	1.096	1.074		2.89	4.04	
3OH-C ₁₄ -HSL	67.73	1.23	4.957	6.694	4.571	2.360	2.229	2.078	1.546	1.456	1.428	1.246	1.161	1.137	1.059		1.18	
C ₁₂ -HSL	68.96	1.22	5.065	6.840	4.671	2.412	2.277	2.124	1.580	1.487	1.459	1.273	1.186	1.162	1.082	1.022		
				Separation Factor (α)														

To screen bacterial cultures isolated from the *Hydra* holobiont for AHL, the approach developed by Shaw, Ping et al. (1997) was used, performing RP-TLC separation followed by reporter strain biosensor detection. The same commercial AHL standards analyzed by LC-MS were also separated by TLC (Pietschke, Treitz et al. 2017). Interestingly, the migration of AHL with OH- and O- substituents at the third carbon was inverted when separated over C₁₈-RP thin-layer chromatography with a methanol/water mobile phase compared to C₁₈-RP-HPLC separation using an acetonitrile/water gradient. This difference in migration behavior may occur due to differences in the interactions with the mobile phase.

Acetonitrile is aprotic whereas methanol as a protic solvent can act as a donor and acceptor of hydrogen bonds and is also more polar than acetonitrile. The interactions of 3O-AHL compounds with the stationary phase may also change due to a shift in keto-enol equilibrium. In general, 1,3-dicarbonyl compounds can exist in three tautomeric forms (Reichardt and Welton 2011). The cis-enole form and the diketo form are predominant in most cases, due to the stabilization of the cis-enolic form by an intramolecular hydrogen bond, which cannot form in the trans-enol.

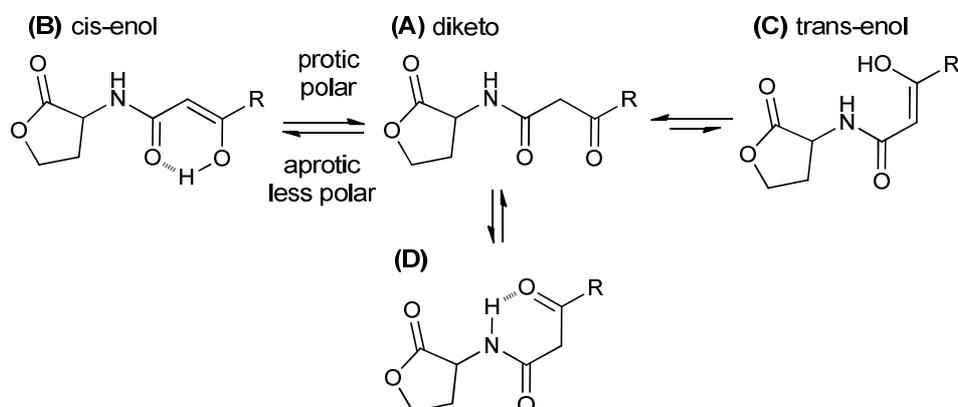


Figure 4.7: Keto-enol tautomerism in 3-oxo AHL compounds. In 1,3-dicarbonyl compounds the hydrogen bond (dashed bond) donor and acceptor in 1,3 position can stabilize the cis-enol (**B**). Hydrogen bonding solvent will generally make lone pairs less available for bonding and shift the equilibrium to the diketo form (**A**). For many 1,3-dicarbonyl compounds, the trans-enol (**C**) is not observed. In case of AHL compounds, the carboxamide group may reduce or prevent formation of (**B**) by forming an intramolecular hydrogen bond with the 3-carbonyl group (**D**). However, the OH:O hydrogen bond should be significantly stronger than NH:O.

Although the effects described for 1,3-dicarbonyl may be different in compounds like AHL, where the 3-carbonyl group is next to the resonance stabilized amide function, similarities cannot be ruled out (Figure 4.7). Therefore, solvent effects on keto-enol tautomerism may influence the polarity of 3O-AHL compounds, containing a second keto-function. The effects of organic solvents on a keto-enol equilibrium have been studied extensively (Burdett and Rogers 1964, Anatolii and Kheifets 1971, Ishida, Hirata et al. 1998, Sandusky 2014). In principal, a higher enol content of cis-enolizing 1,3-dicarbonyl compounds is observed in apolar aprotic than in polar protic solvents (Figure 4.7). On dissolution in solvents of low polarity the percentage of cis-enolic form increases, whereas polar solvents displace the equilibrium towards the diketo tautomer (Reichardt and Welton 2011). However, the amide function next to the 3-dicarbonyl group may stabilize the diketo form with an intramolecular hydrogen bond with the carbonyl-group at C₃ and prevent the formation of the enol tautomer.

Retention-times in general and in particular solvent effects in the microenvironment around the stationary phase are not easy to predict. With these reservations, the aprotic, less polar solvent, acetonitrile, may shift the keto-enol equilibrium to the stabilized cis-enol tautomer, whereas in the protic, polar methanol, the diketo form would be preferred.

If this is the case, the shorter retention-time of 3O-AHL in acetonitrile could be explained because the tautomer 3OH-C_x-1-HSL would show weaker interaction with the RP stationary phase than 3OH-C_x-HSL, due to the additional double bond. In general, the carbonyl-function is less polar

than the hydroxy-function and should therefore show a higher retention as was observed in TLC with methanol.

The amphiphilic character of AHL is fundamental for the biological function of these molecules, enabling the high permeability through biological membranes. However, similar to phospholipids and detergents, AHL show a strong adhesion to reversed-phase column material and contact surfaces like the injector needle or the sample loop. As a general caution for the analysis of AHL by HPLC, it should be mentioned that persistent sample to sample carryover in subsequent LC-runs was frequently observed. This caused a need for extensive column wash- and equilibration procedures after each separation. In an attempt to reduce carryover, different stationary phases were tested with samples containing the three (acyl, 3O- and 3OH-) C₁₀-HSL compounds. C₄- and C₈- nano columns as well as a hydrophilic interaction chromatography (HILIC)-nano column were insufficient for separation of even C₁₀-HSL and 3O-C₁₀-HSL, showing increasingly poor resolution over column materials of shorter chain-length (data not shown).

4.4.3 Fragmentation of AHL in MS/MS experiments

One of the parameters which determine the quality of MS/MS spectra and the yield of structural information, is the collision energy used for fragmentation. depending on the stability of the analyte upon activation, the collision energy to be adapted to acquire MS/MS spectra with intense fragment signals under full or near complete decomposition of the isolated precursor ion. The optimal normalized collision energies (NCE) for HCD and CID applied on the LTQ Orbitrap Velos mass spectrometer (Thermo Fisher Scientific) were experimentally determined by fragmentation of the AHL standard compounds and was found to depend on the length and character of the side-chain. Near complete decomposition for all tested compounds was observed with a NCE of 32 for HCD and 28 for CID. These parameters apply to all MS/MS spectra described in the following discussion.

Independent of the ion activation, all AHL compounds produced a strong characteristic signal at m/z 102.055, corresponding to the singly charged homoserine lactone ion. This signal results from AHL fragmentation at the amide bond and can be used as a first diagnostic ion to screen for AHL derivatives in biological samples. The corresponding signal derived from the acyl side-chain moiety is a resonance stabilized acylium ion. Figure 4.8 shows the accepted scheme for the positive ion MS fragmentation of AHL into two major products (Shaw, Ping et al. 1997, Morin, Grasland et al. 2003, Gould, Herman et al. 2006).

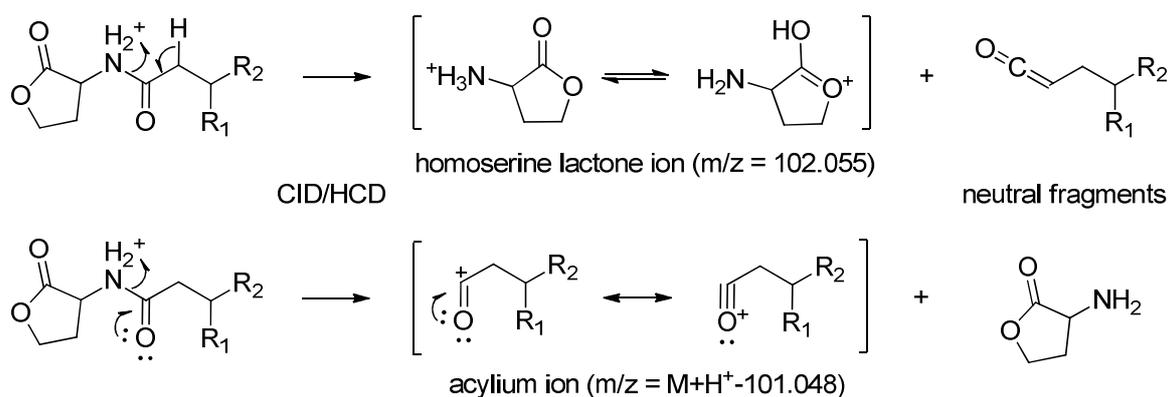


Figure 4.8: Scheme for the fragmentation of AHL into two major fragments according to (Morin, Grasland et al. 2003). Two major signals result from fragmentation of the amide bond, either of which can accept the proton while the other part escapes detection as a neutral fragment. The HSL moiety fragment is protonated α -amino- γ -butyrolactone and generates a signal at m/z 102.055; the other fragment can form a resonance stabilized acylium ion with a m/z corresponding to $M+H^+-101.048$.

The two fragmentation mechanisms at the amide bond are described as the products of the protonated molecular ion carrying the charge at the amine. Two additional structures for the $[M+H]^+$ ion, with protonation of either the carbonyl-function in the acyl-chain, or of the ethoxy-group of the lactone, have been proposed. The fragmentation of the former ion leads to the loss of a water molecule, the latter produces the loss of carbon monoxide from the molecule ion (Morin, Grasland et al. 2003). The respective m/z signals for these proposed fragments were observed in all acquired AHL fragment spectra, however, generally with lower intensity.

The calibration mix recommended by the manufacturer for use with the LTQ Orbitrap Velos MS covers a mass range from m/z 74 to m/z 1922. The lowest molecular weight calibrant is *n*-butylamine (m/z 74.097), followed a caffeine fragment (m/z 138.066), caffeine (m/z 195.088) and the peptide MRFA (m/z 524.265) to cover the lower m/z range. However, the signal at m/z 74.097 is used only for calibration of the Ion trap mass analyzer. Within the range between 50 m/z and 500 m/z , default FT-calibration utilizes only the two caffeine signal. Therefore, the standard calibration mix is not perfectly suited to low mass calibration.

FTMS/MS spectra of the low mass range between 50 m/z and 500 m/z , acquired with an LTQ Orbitrap Velos MS calibrated to the manufacturer specification limits, showed mass deviations of up to ± 15 ppm. The number of possible elemental compositions for a given signal decreases as the mass accuracy of the measurement improves. A 30 ppm window around a given m/z in the range between 200 m/z and 500 m/z is often insufficient for the unambiguous assignment of a chemical composition. Using only the most common elements in organic molecules (H, C, O, N), a 30 ppm range can include theoretical m/z values for more than five chemical formulas. However, with the Orbitrap FTMS, the predominant part of mass deviation can generally be attributed to a

global mass shift that applies to all signals in a given spectrum. This means that the mass accuracy between peaks in the same spectrum is much higher and stays within the limits of about 2 ppm, where the assignment of chemical composition is usually unambiguous.

Since the exact mass is known for the ubiquitous protonated homoserine lactone fragment at m/z 102.05495, this signal is not only useful as a diagnostic ion for AHL compounds, but also for the internal calibration of fragment-spectra to a very high mass accuracy. Concerning the calculation of chemical composition of ions from MS signals, as mentioned above, these assignments are usually unambiguous. Nonetheless, in the following chapter, elemental compositions of signals are assigned by m/z values alone, these only represent highly probable, and the most likely empirical formula of the detected ions to a point where the false assignment is highly unlikely.

If the elemental composition is known, the ring and double bond equivalents (RDBE), also known as the degree of unsaturation, can be calculated. The RDBE value provides insights into the number of rings, double bonds or triple bonds to expect in an unknown structure. The number of RDBE can be calculated using Formula 4.5 (Dayringer and McLafferty 1977).

$$\text{Formula 4.5: } RDBE = C + Si - \frac{1}{2(H+F+Cl+Br+I)} + \frac{1}{2(N+P)} + 1$$

Each element symbol here represents the count of this element in the molecular formula (Kind and Fiehn 2007). It has to be noted that the protonated molecular- or fragment ions detected by mass spectrometry in positive ion mode are often $[M+H]^+$. If this is the case, the calculation of RDBE by Formula 4.5 results in a non-integer, which is by 0.5 lower than the actual number of rings or double bonds present in the structure. Additionally, the formula has its limitations as it only accounts for the lowest valence state of each element. However, this issue does not apply to RDBE calculations with the composition of AHL compounds analyzed here.

The low mass range of AHL fragment spectra (between 50 m/z and 100 m/z) shows signals at m/z of 74.060; 84.044 and 85.028. These m/z values correspond to the respective losses of carbon monoxide, water or ammonia from the protonated HSL moiety at m/z 102.055. Among these three signals, the carbon monoxide-loss was comparatively the strongest, accounting for about 10 % of the total ion current, while the product of water elimination showed very weak signals below 5 % of relative abundance. The ammonia-loss was not detected in all spectra. It could be confirmed that these signals are derived from the HSL moiety by direct fragmentation of the m/z 102.055 ion, which accompanied all AHL signals in MS survey scans probably formed by in-source fragmentation. This is also in accordance with the ESI MS fragmentation behavior, following the loss of small neutrals, previously observed for several different five membered lactones by Crotti

et al., although the specific compound, α -amino- γ -butyrolactone, was not represented in that study (Crotti, Fonseca et al. 2004).

As the HSL moiety is the same in all AHL derivatives, the differences between AHL spectra upon CID or HCD are generally due to distinct chain lengths and substitutions at the C₃-position of the acyl side-chain. To characterize these differences, AHL standards for the three commonly described substituents at C₃, which are -O, -OH and -H respectively, with various acyl side-chain lengths, were fragmented using both CID and HCD.

4.4.3.1 Fragmentation of the unsubstituted AHL in MS/MS experiments

The AHL standards with no substituent at the C₃-position included compounds with acyl-chain lengths of C₄-, C₆-, C₈-, C₁₀-, and C₁₂. HCD and CID MS/MS-spectra of these compounds are shown in Figure 4.9 - Figure 4.13. HCD commonly fragments at the amide bond generating the [M+H]⁺ ions of the HSL moiety and the corresponding acylium ion [M-100]⁺, as described in the fragmentation scheme in Figure 4.8. However, one notable exception to this rule was observed. The shortest AHL compound, namely C₄-HSL, did not fragment according to this scheme (see Figure 4.9). Although the base peak at m/z 102.055 was present, no signal for the corresponding acylium ion was detected at m/z 71.050. Signals for further decomposition of the [C₄H₇O]⁺ ion were out of the detected m/z range from 50 m/z to 500 m/z recorded for HCD spectra.

In fragment spectra of the longer chained AHL, the acylium ion [M-100.040]⁺ is accompanied by a weaker signal of the successive water-loss [M-100.040-18.011]⁺. Both signals decrease in intensity with increasing length of the acyl-chains. For the acylium ions, relative signal intensities range from approximately 90 % in C₆-HSL to 30 % in C₁₂-HSL fragment spectra. The water-loss from the acylium ion was not detected in C₆-HSL HCD spectra, which instead shows a 15 % signal most probably due to carbon monoxide-loss from the acylium ion [M-100.040-27.995]⁺, but was observed in all other spectra with intensities decreasing from ca. 25 % for C₈-HSL to 10 % for C₁₂-HSL. Additional characteristic signals which could be assigned were the neutral loss of carbon monoxide [M+H-27.995]⁺ and water [M+H-18.011]⁺ from the protonated molecular ion with relative intensities of about 10 % and 5 %, respectively.

Compared to CID, HCD has been demonstrated to display fragment ions resulting from multiple steps of activation. Having access to both dissociation methods can be a significant advantage for the structural investigation of small molecule as the two dissociation methods can provide different energy pathways to access fragmentation fingerprints due to the intrinsic mechanistic differences of ion dissociation (Huang 2009). As a result, prominent neutral loss fragments and larger

fragment signals are often produced by CID, while HCD in many cases leads to richer spectra and a higher degree of fragmentation (Jedrychowski, Huttlin et al. 2011).

This is also observed in HCD spectra of AHL. While the major acylium ion current decreased with increasing chain length, an increasing number of small aliphatic ions with 0.5 and 2.5 RDBE could be assigned. Figure 4.31 shows this in detail for C₁₂-HSL.

In contrast to HCD, which generally dissociated unsubstituted AHL derivatives in accordance with the fragmentation scheme described in Figure 4.8, CID produced much weaker signals for the acylium ion [M-100]⁺ and the successive water-loss. Instead, the major fragment besides the HSL-moiety corresponds to the water-loss of the protonated molecular ion [M+H-18.011]⁺. The relative intensity of this ion increases with the AHL chain length. From ca. 60 % relative intensity for C₄-HSL, the water-loss signal even succeeds the HSL moiety ion as the base peak (strongest peak in a spectrum, i.e. relative signal intensity 100 %) in CID spectra of C₁₂-HSL. The carbon monoxide-loss [M+H-27.995]⁺ shows the same behavior with relative intensities between approximately 10 % and 25 %.

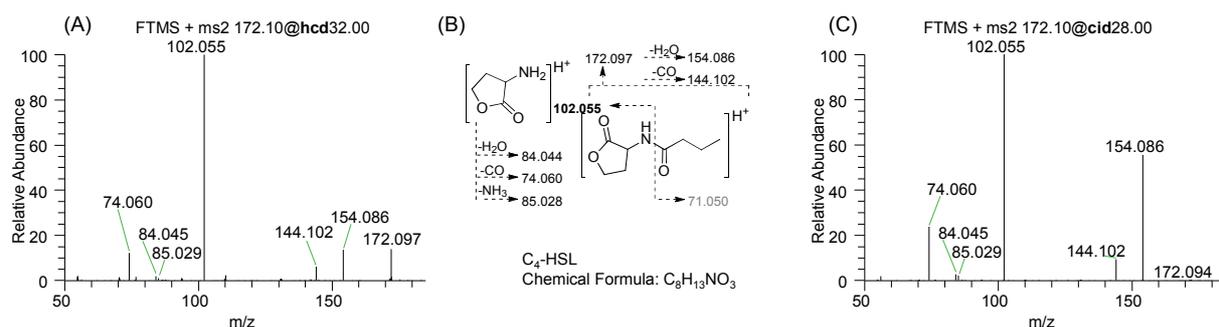


Figure 4.9: Reference spectra of C₄-HSL. Fragmentation by HCD (A) and CID (C). A fragmentation scheme is shown in (B). The signal corresponding to the acylium ion from fragmentation of the amide bond (*m/z* 71.050) is not observed.

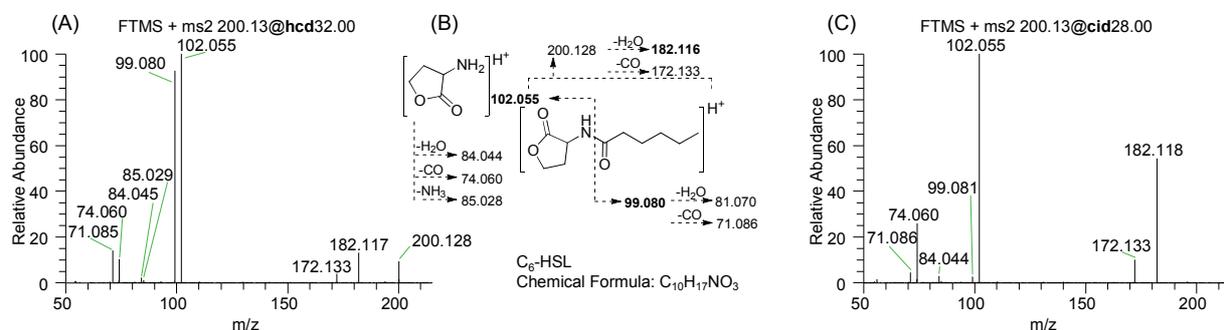


Figure 4.10: Reference spectra of C₆-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. HCD produces two major signals from fragmentation at the amide bond. In contrast, CID results in formation of the water loss from the protonated molecular ion (*m/z* 182.116).

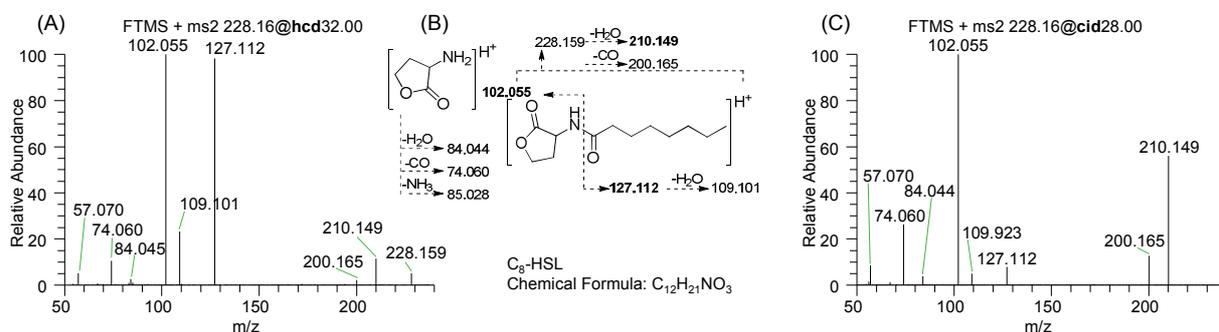


Figure 4.11: Reference spectra of C₈-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. HCD of unsubstituted AHL with acyl chain length of C₈ and longer produces a signal corresponding to the loss of water from the acylium ion (m/z 109.101).

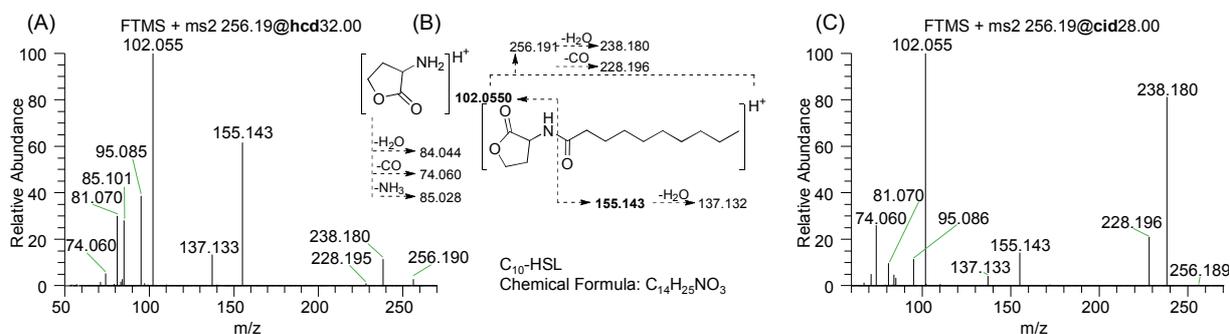


Figure 4.12: Reference spectra of C₁₀-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. The signal pair corresponding to the acylium ion (155.143 and the loss of water from the acylium ion (m/z 127.132) decrease with increasing chain length upon HCD.

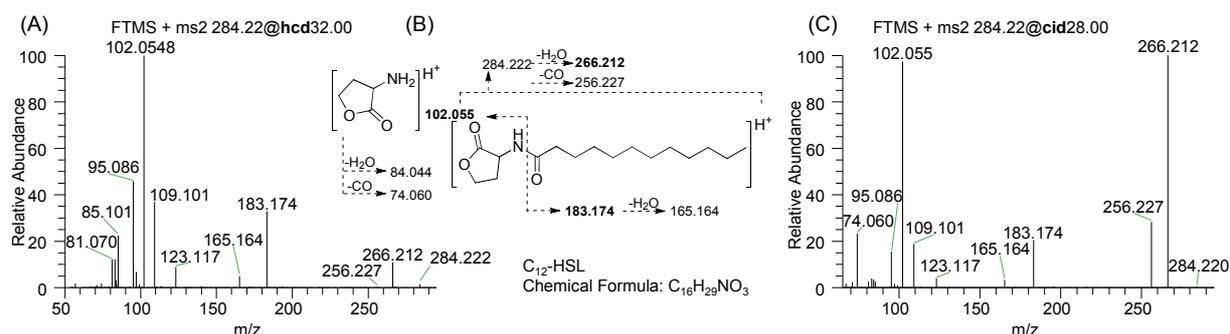


Figure 4.13: Reference spectra of C₁₂-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. The relative intensity of the signal corresponding to the acylium ion and the corresponding water loss generally observed in HCD spectra of acyl AHL compounds decreases with the acyl chain length. In CID spectra, the signals resulting from loss of water or carbon monoxide from the protonated molecular ion increase with the acyl-chain length.

4.4.3.2 Fragmentation of the 3O-AHL in MS/MS experiments

Five AHL standards carrying a carbonyl-group at C₃ and different acyl-chain lengths, 3O-C₆-, 3O-C₈-, 3O-C₁₀-, 3O-C₁₂- and 3O-C₁₄-HSL, were analyzed. The corresponding spectra are shown in Figure 4.14 - Figure 4.18. HCD generally produced ions in accordance with the fragmentation scheme in Figure 4.8. While the HSL moiety at *m/z* 102.055 was the base peak, the corresponding acyl-chain ions showed with increasing relative abundances from 50 % for the smaller 3O-C₆-HSL to approximately 95 % for the longer chained 3O-C₁₄-HSL. In addition to the complete side-chain acylium ion, HCD spectra of 3O-AHL also show a characteristic ion from the fragmentation between the α-carbon and the β-carbonyl-group, which results in a shorter acylium ion [M-100.040-C₂H₂O]⁺. For compounds of the general formula 3O-C_x-HSL, these signals have the same empirical formula as the complete acylium ions of a C_{x-2}-HSL. Among the 3O-AHL compounds tested, the sole exception from this fragmentation pattern was found for 3O-C₆-AHL. Potentially due to the short side-chain, fragmentation of 3O-C₆-HSL did not result in a signal of the short acylium ion at *m/z* 71.050 (Figure 4.14). Interestingly, this signal was also missing in fragment spectra of C₄-HSL as discussed previously (Figure 4.9). In HCD spectra of longer chained 3O-AHL, this secondary acylium ion accounted for 20 % to 30 % of the base peak.

CID of 3O-AHL was markedly different from HCD concerning the formation of acylium ions. Firstly, the signal of the secondary acylium ions observed upon HCD between the α-carbon and the β-carbonyl-group was not detected in CID spectra. Secondly, the intensity of the complete acyl-chain ion was very low for the shorter chained and increased more strongly for longer chained AHL. In CID, 3O-C₆-HSL formed an acylium ion at *m/z* 113.060 of only 10 % intensity relative to the 102.055 *m/z* base peak (Figure 4.14). In contrast, 3O-C₁₄-HSL showed a relative abundance of 20 % for the 102.055 *m/z* signal with the acylium ion at 225.185 *m/z* as the base peak (see Figure 4.18).

Spectra from both ion activation mechanisms have in common that no dominant signals for water- or carbon monoxide-loss from the protonated molecular ion were observed. However, compared to the previously described unsubstituted AHL derivatives, the signals produced due to loss of small neutrals from the protonated molecular ion were more numerous. In addition to the generally observed water [M+H-18.011]⁺ and carbon monoxide [M+H-27.995]⁺ losses, these included below 10 % signals corresponding to the consecutive loss of carbon monoxide in addition to water [M+H-46.005]⁺, carbon monoxide [M+H-55.990]⁺ or formaldehyde [M+H-58.005]⁺. Finally, an increasing number of peaks for the ion series of alkene hydrocarbon fragments of the general composition C_nH_{2n-3} (with RDBE 2.5) were observed in the MS/MS spectra of AHL compounds with longer acyl-chain length. Figure 4.28 shows this in detail for the fragmentation of 3O-C₁₂-HSL.

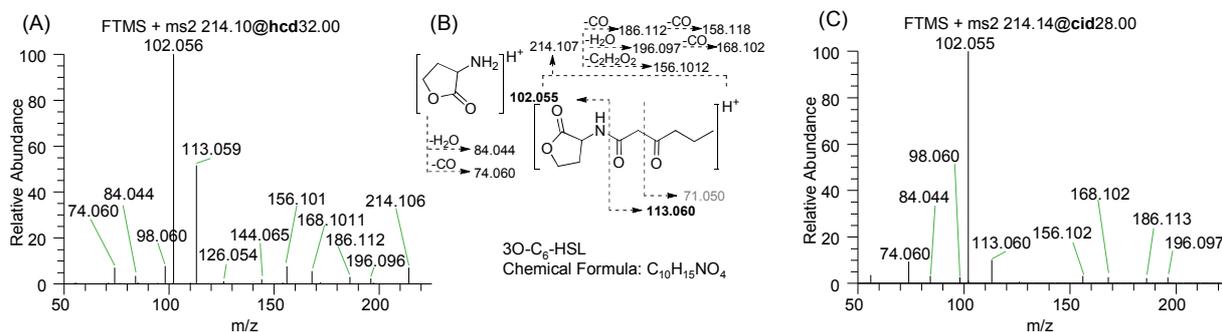


Figure 4.14: Reference spectra of 3O-C₆-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. Fragmentation of 3O-C₆-HSL did not result in a signal corresponding to the short acylium ion (m/z 71.050), which is generally observed in HCD spectra of longer chained 3O-AHL compounds.

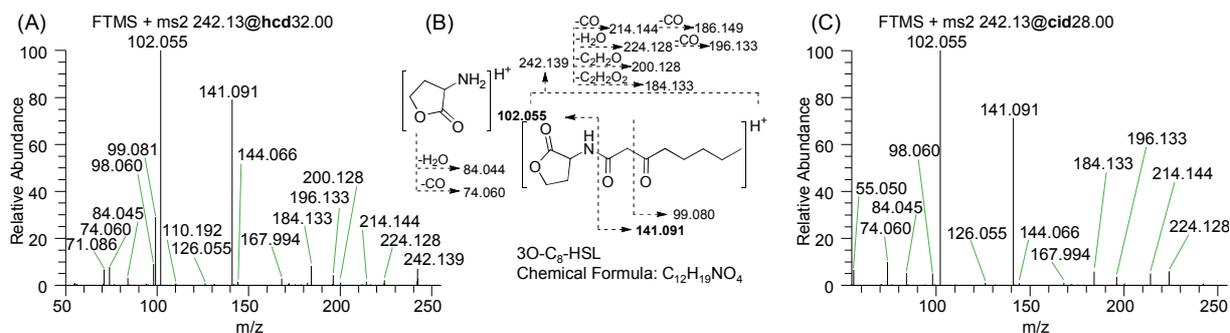


Figure 4.15: Reference spectra of 3O-C₈-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. In particular, CID spectra of 3O-AHL compounds show a number of low intensity signals corresponding to the loss of small neutrals from the protonated molecular ion.

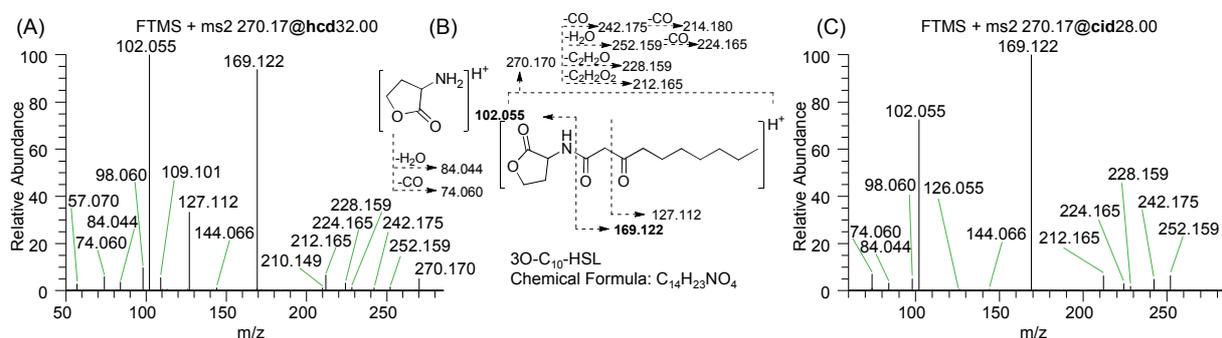


Figure 4.16: Reference spectra of 3O-C₁₀-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between.

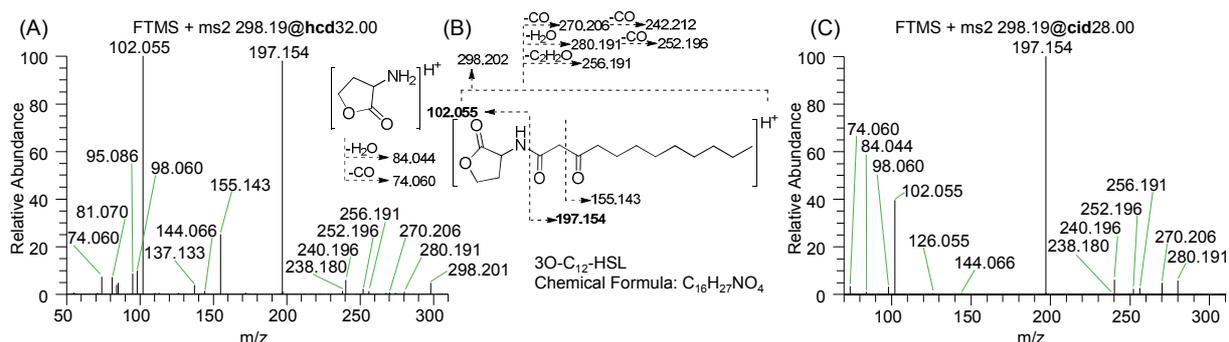


Figure 4.17: Reference spectra of 3O-C₁₂-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. In CID spectra, the relative signal intensities of the acylium ion (m/z 197.154) strongly increase with the chain length and replace the homoserine lactone ion as the base peak.

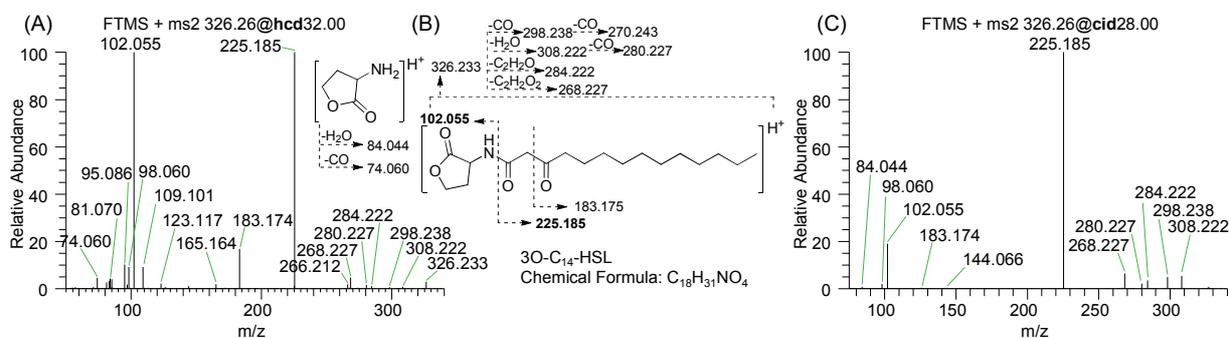


Figure 4.18: Reference spectra of 3O-C₁₄-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between.

4.4.3.3 Fragmentation behavior of 3OH-AHL in MS/MS experiments

The 3OH-AHL standards included 3OH-C₈-HSL, 3OH-C₁₀-HSL, 3OH-C₁₂-HSL and 3OH-C₁₄-HSL (Figure 4.19 - Figure 4.22). HCD of these compounds produced very dominant signals for the protonated HSL moiety at m/z 102.055. The corresponding acyl-chain ion was generally not detected or below 1 % relative signal intensity. The second- and third-strongest signals account for less than 15 % of the ion current and correspond to two consecutive neutral losses of water $[M-100.040-18.011]^+$ and $[M-100.040-36.021]^+$ or water and carbon monoxide $[M-100.040-46.005]^+$ from the acyl-chain. The same neutral loss signals are observed for the protonated molecular ion with relative intensities below 5 %.

In contrast, CID results in very intense signals for the water-loss, which replaces the HSL moiety at m/z 102.005 as the base peak in fragment spectra of longer chained 3OH-ASL compounds. As in HCD, no signal was observed for the acyl-chain. However, a low peak for the water-loss product

of the latter is also generally present with relative intensities below 5 %. With increasing acyl-chain length, fragmentation of 3OH-AHL results in more numerous low intensity signals for the two hydrocarbon ion series with 2.5 RDBE and 3.5 RDBE (see Figure 4.30 for details).

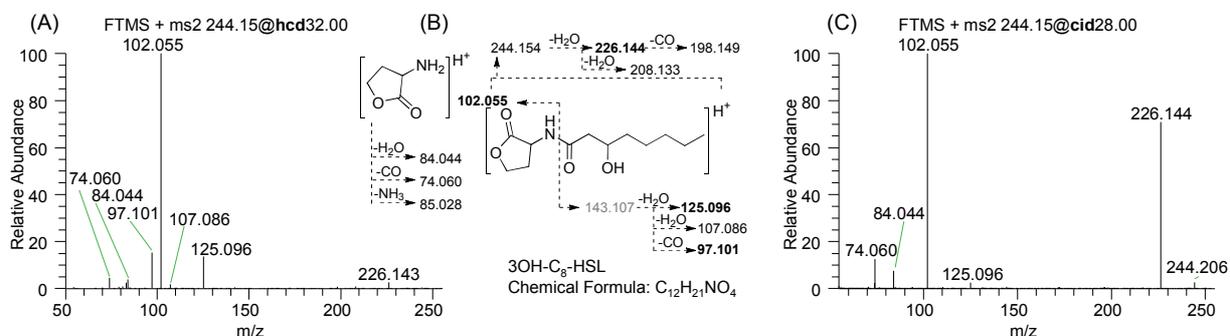


Figure 4.19: Reference spectra of 3OH-C₈-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. HCD of 3OH-AHL compounds produce very dominant signals for the protonated HSL moiety at m/z 102.055. CID spectra show an additional signal corresponding to the loss of water from the molecular ion.

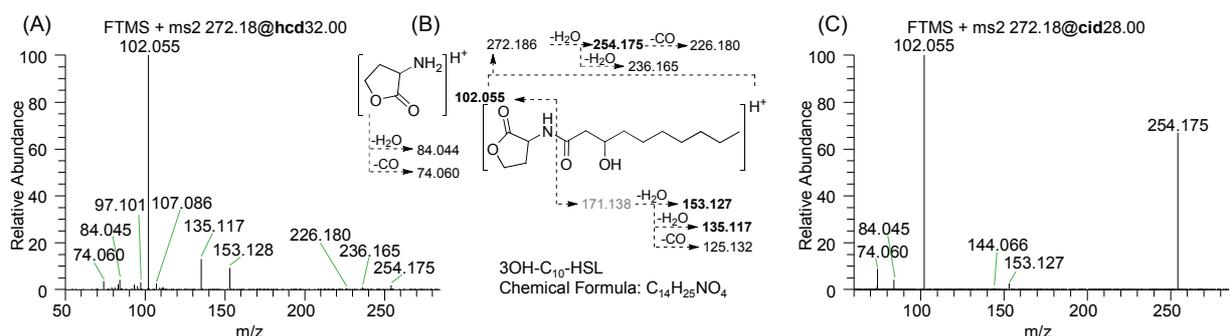


Figure 4.20: Reference spectra of 3OH-C₁₀-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. Low intensity signals corresponding to the consecutive water loss from the acylium ion can be detected in HCD spectra of 3OH-AHL compounds.

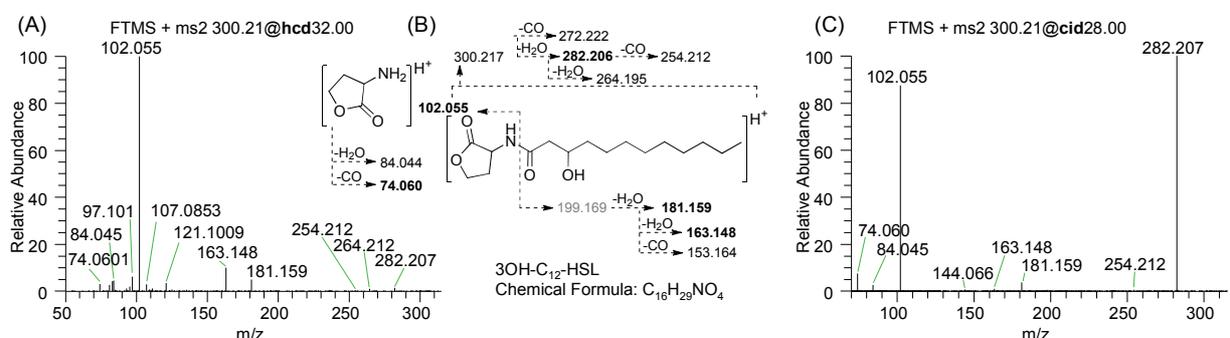


Figure 4.21: Reference spectra of 3OH-C₁₂-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between.

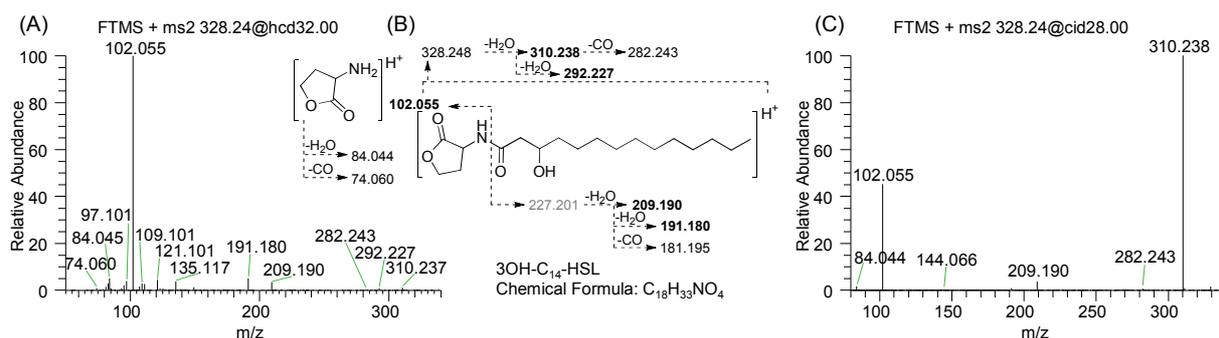


Figure 4.22: Reference spectra of 3OH-C₁₄-HSL. Fragment spectra produced by (A) HCD and (C) CID. (B) A fragmentation scheme is shown in between. In CID spectra, the intensity of the signals corresponding to the water loss from the protonated molecular ion increase with the acyl chain length and replace the homoserine lactone ion as the base peak.

In summary, the fragmentation behavior of AHL derivatives that only differ in the acyl-chain length, are overall very similar for a specific ion activation method. The main difference being minor changes in the intensity of major fragments, rather than different fragmentation patterns. However, the different ion activation methods and different substituents at the C₃-position have a significant impact on fragmentation fingerprints (see Table 4.2). Only HCD of 3O-AHL and acyl-AHL and CID of 3O-AHL followed the commonly proposed fragmentation scheme (Figure 4.8), from which in particular 3OH-AHL deviate. This is not too surprising because the commonly accepted scheme is based on data of acyl- and 3O-AHL. 3OH-AHL derivatives are overall underrepresented both in analytical studies towards method development and in the majority of biological systems analyzed to date. Especially for the purpose of AHL quantification, this lack of data can have adverse effects. Gould, Herman et al. (2006) noted that the intensities of the two generally proposed major CID fragments, the AHL moiety and the acyl side-chain, correlate with the AHL quantity and can therefore be used to estimate the composition of AHL mixtures. Furthermore, to infer the absolute quantity of several AHL with reasonable accuracy, a single isotope labeled acyl-AHL internal standard was used (Gould, Herman et al. 2006). However, due to the differences in fragmentation behavior of AHL with different C₃-substituents, at least one standard for each of these AHL derivatives would be necessary for reliable quantification. In addition it has to be considered, that the fragmentation behavior of AHL strongly depends on the MS instrument and activation method used. Furthermore, signal intensities may be biased due to in-source fragmentation of less stable AHL or differences in ionization efficiency, depending on acyl-chain length. For these reasons, reliable absolute quantification of all AHL in complex mixtures ideally would require internal standards for each individual compound

Table 4.2: Summary of LC-ESI MS/MS characteristics of AHL standard compounds. The table shows chromatographic retention-times (see chapter 4.4.2) and nominal m/z values of the protonated molecular ions $[M+H]^+$, observed for AHL standard compounds in LC-MS experiments. The molecular identity and m/z values of the six strongest signals observed upon HCD or CID are listed and colored according to the relative signal intensities observed. Abbreviations: HSL for the homoserine lactone fragment; Ac for the Acyl side-chain.

AHL	RT [min]	$[M+H]^+$ m/z	HCD					CID					[%] Relative abundance
			Base peak m/z	2 nd	3 rd	4 th	5 th	Base peak m/z	2 nd	3 rd	4 th	5 th	
30-C ₆ -HSL	23.70	m/z 214	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 113	$[C_5H_8ON]^+$ m/z 98	$[M+H-C_2H_2O_2]^+$ m/z 156	$[HSL+H-CO]^+$ m/z 74	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 113	$[M-100-CO]^+$ m/z 74	$[M+H-C_2H_2O_2]^+$ m/z 156	$[M+H-H_2O-CO]^+$ m/z 168	0 - 5
30-C ₈ -HSL	36.66	m/z 242	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 141	$[Ac-C_2H_2O]^+$ m/z 99	$[C_5H_8ON]^+$ m/z 98	$[M+H-C_2H_2O_2]^+$ m/z 184	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 141	$[M-100-CO]^+$ m/z 74	$[M+H-C_2H_2O_2]^+$ m/z 184	$[M-H_2O]^+$ m/z 224	5 - 10
30-C ₁₀ -HSL	47.82	m/z 270	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 169	$[Ac-C_2H_2O]^+$ m/z 127	$[C_5H_8ON]^+$ m/z 98	$[M+H-C_2H_2O_2]^+$ m/z 212	$[HSL+H]^+$ m/z 102	$[HSL+H]^+$ m/z 102	$[M-100-CO]^+$ m/z 74	$[M+H-C_2H_2O_2]^+$ m/z 212	$[M-H_2O]^+$ m/z 252	10 - 15
30-C ₁₂ -HSL	56.62	m/z 298	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 197	$[Ac-C_2H_2O]^+$ m/z 155	$[C_5H_8ON]^+$ m/z 98	$[C_7H_{11}]^+$ m/z 95	$[HSL+H]^+$ m/z 102	$[HSL+H]^+$ m/z 102	$[M+H-C_2H_2O_2]^+$ m/z 240	$[M-H_2O]^+$ m/z 280	$[M-CO]^+$ m/z 270	15 - 20
30-C ₁₄ -HSL	64.60	m/z 326	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 225	$[Ac-C_2H_2O]^+$ m/z 183	$[C_7H_{11}]^+$ m/z 95	$[C_8H_{13}]^+$ m/z 109	$[HSL+H]^+$ m/z 102	$[HSL+H]^+$ m/z 102	$[M+H-C_2H_2O_2]^+$ m/z 268	$[M-H_2O]^+$ m/z 308	$[M-CO]^+$ m/z 298	20 - 25
30H-C ₈ -HSL	38.49	m/z 244	$[HSL+H]^+$ m/z 102	$[Ac-H_2O-CO]^+$ m/z 97	$[Ac-H_2O]^+$ m/z 125	$[HSL+H-CO]^+$ m/z 74	$[HSL+H-H_2O]^+$ m/z 84	$[HSL+H]^+$ m/z 102	$[M+H-H_2O]^+$ m/z 226	$[HSL+H-CO]^+$ m/z 74	$[HSL+H-H_2O]^+$ m/z 84	$[M+H]^+$ m/z 244	25 - 30
30H-C ₁₀ -HSL	50.84	m/z 272	$[HSL+H]^+$ m/z 102	$[Ac-2xH_2O]^+$ m/z 135	$[Ac-H_2O]^+$ m/z 153	$[HSL+H-H_2O]^+$ m/z 74	$[HSL+H-CO]^+$ m/z 74	$[HSL+H]^+$ m/z 102	$[M+H-H_2O]^+$ m/z 254	$[HSL+H-CO]^+$ m/z 74	$[HSL+H-H_2O]^+$ m/z 84	$[Ac-H_2O]^+$ m/z 153	30 - 35
30H-C ₁₂ -HSL	59.92	m/z 300	$[HSL+H]^+$ m/z 102	$[Ac-2xH_2O]^+$ m/z 163	$[C_7H_{13}]^+$ m/z 97	$[Ac-H_2O]^+$ m/z 181	$[HSL+H-H_2O]^+$ m/z 84	$[HSL+H]^+$ m/z 102	$[HSL+H]^+$ m/z 102	$[HSL+H-CO]^+$ m/z 74	$[Ac-H_2O]^+$ m/z 181	$[HSL+H-H_2O]^+$ m/z 84	35 - 40
30H-C ₁₄ -HSL	67.73	m/z 328	$[HSL+H]^+$ m/z 102	$[Ac-2xH_2O]^+$ m/z 191	$[HSL+H-H_2O]^+$ m/z 84	$[C_9H_{13}]^+$ m/z 121	$[C_7H_{13}]^+$ m/z 97	$[HSL+H]^+$ m/z 102	$[HSL+H]^+$ m/z 102	$[Ac-H_2O]^+$ m/z 209	$[HSL+H-H_2O]^+$ m/z 84	$[M-H_2O-CO]^+$ m/z 282	40 - 45
C ₄ -HSL	19.79	m/z 172	$[HSL+H]^+$ m/z 102	$[M+H]^+$ m/z 172	$[M+H-H_2O]^+$ m/z 154	$[HSL+H-CO]^+$ m/z 74	$[M+H-CO]^+$ m/z 144	$[HSL+H]^+$ m/z 102	$[M+H-H_2O]^+$ m/z 154	$[HSL+H-CO]^+$ m/z 74	$[M+H-CO]^+$ m/z 144	$[HSL+H-H_2O]^+$ m/z 84	45 - 50
C ₆ -HSL	35.25	m/z 200	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 99	$[Ac-CO]^+$ m/z 71	$[M+H-H_2O]^+$ m/z 182	$[HSL+H-CO]^+$ m/z 74	$[HSL+H]^+$ m/z 102	$[M+H-H_2O]^+$ m/z 182	$[HSL+H-CO]^+$ m/z 74	$[M+H-CO]^+$ m/z 172	$[Ac-CO]^+$ m/z 71	55 - 60
C ₈ -HSL	50.09	m/z 228	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 127	$[Ac-H_2O]^+$ m/z 109	$[M+H-H_2O]^+$ m/z 210	$[HSL+H-CO]^+$ m/z 74	$[HSL+H]^+$ m/z 102	$[M+H-H_2O]^+$ m/z 210	$[HSL+H-CO]^+$ m/z 74	$[M+H-CO]^+$ m/z 200	$[Ac]^+$ m/z 127	60 - 65
C ₁₀ -HSL	60.92	m/z 256	$[HSL+H]^+$ m/z 102	$[Ac]^+$ m/z 155	$[C_7H_{11}]^+$ m/z 95	$[C_6H_9]^+$ m/z 81	$[C_6H_{13}]^+$ m/z 85	$[HSL+H]^+$ m/z 102	$[M+H-H_2O]^+$ m/z 238	$[HSL+H-CO]^+$ m/z 74	$[M+H-CO]^+$ m/z 228	$[Ac]^+$ m/z 155	65 - 70
C ₁₂ -HSL	68.96	m/z 284	$[HSL+H]^+$ m/z 102	$[C_7H_{11}]^+$ m/z 95	$[C_8H_{13}]^+$ m/z 109	$[Ac]^+$ m/z 183	$[C_6H_{13}]^+$ m/z 85	$[M+H-H_2O]^+$ m/z 266	$[HSL+H]^+$ m/z 102	$[M+H-CO]^+$ m/z 256	$[HSL+H-CO]^+$ m/z 74	$[Ac]^+$ m/z 183	70 - 75

4.4.4 Analysis of *Hydra quorum* quenching activity

With the use of bacterial reporter strains, our collaboration partners discovered *quorum* quenching of several AHL signals by different *Hydra* spp. So far, the only enzymes described to be capable of quenching AHL signals found in eukaryotic systems are the paraoxonase (PON) family of AHL lactonases (Chun, Ozer et al. 2004, Draganov, Teiber et al. 2005). *Hydra* possesses one homologue to the human paraoxonases, however, a knockdown experiment of the encoding gene did not reduce the *quorum* quenching activity (Pietschke, Treitz et al. 2017). For this reason the *Hydra* induced quenching of one AHL was analyzed by MS to characterize the underlying reaction mechanism and to identify the *quorum* quenching product. To this end, *Hydra vulgaris* polyps were incubated for 24 hours with 100 μ M of the long-chain AHL, 3O-C₁₀-HSL and remaining AHL was extracted with ethyl acetate. As positive and negative controls, the experiments were conducted in parallel with the same AHL-concentration incubated in sterile medium and with *Hydra* polyps without supplement. The extracts of *quorum* quenching experiments were analyzed by MALDI-TOF MS and LC-ESI MS. First, MALDI-TOF MS was used for high-throughput screening of *Hydra quorum* quenching assays (between 8 and 14 replicates). Possible *quorum* quenching product candidates identified by this screening process were then further characterized by LC-ESI MS. The results of the MALDI-TOF MS screening are summarized in Figure 4.23. The substrate 3O-C₁₀-HSL with an [M+H]⁺ ion at m/z 270.17 was stable over 24 hours in control, while in the presence of *Hydra* polyps the respective signal decreased significantly over time. In parallel, two [M+H]⁺ ions at an m/z of 272.19 and m/z 254.18 were detected as potential products increasing in intensity with longer incubation times.

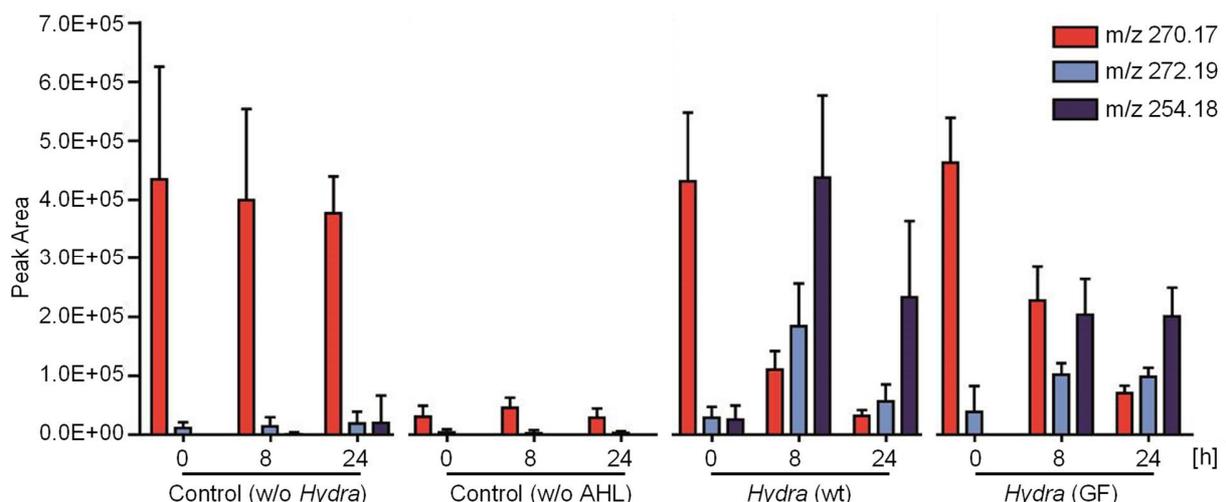


Figure 4.23: Screening of *Hydra quorum* quenching experiments by MALDI-TOF MS. Columns represent the peak areas for the signals observed at m/z 270.17, m/z 254.18 and m/z 272.19. *Hydra* (wt) polyps were with intact microbiota, *Hydra* (GF) were germfree animals. Control experiments without *Hydra* and without AHL addition were analyzed and data for incubation times of 0 h, 8 h and 24 h are shown.

For MALDI-TOF MS analysis of the *Hydra quorum* quenching activity, the substrate 3O-C₁₀-HSL protonated molecular ion at m/z 270.17 and the two potential products at m/z of 272.19 and m/z 254.18 were monitored. Both AHL samples incubated with *Hydra vulgaris* (with intact microbiota) and germfree *Hydra vulgaris* show the decrease of the AHL signals. These results attribute the degradation of 3O-C₁₀-HSL to *Hydra* rather than its communal microorganisms. The degradation clearly advances with the incubation time from 0 to 8 and 24 h.

3O-C₁₀-HSL was incubated in Volvic medium and sterile medium without *Hydra*, as respective controls to test the stability of the compound during incubation. No significant reduction of the AHL substrate was apparent in these samples. As a negative control for possible inherent AHL production by *Hydra* itself, samples without additional AHL were extracted and analyzed. Low intensity signals were detected at m/z 270.17 in these negative controls; and signals of comparably low intensity were observed at m/z values of the supposed *quorum* quenching products (m/z 254.18 and m/z 272.19) in control samples incubated without *Hydra*. There are several possible explanations for this. These signals could be attributed to the high noise level observed in the low m/z range of MALDI-TOF MS spectra or to chemical noise due to contaminants from *Hydra* extracts directly screened by MALDI-TOF MS without sample cleanup. In the control samples without *Hydra*, the second isotope peak of the 3O-AHL substrate can also contribute to the signal at m/z 272.19. A high noise-level and the expected contaminant signals were also observed in the pure AHL standard mixture used for MS calibration (Figure 4.24).

MALDI-TOF MS generally displayed a low sensitivity for the detection of AHL and high variability in signal intensities (Figure 4.23 and Figure 4.24). Although no contaminant was detected at m/z 270.17, the mass range between 100 m/z and 500 m/z is populated by numerous CHCA matrix signals. These signals are of higher intensities than the signal generated by 500 pmol of 3OH-C₁₀-HSL (Figure 4.24), which shows the low ionization efficiency for this class of analytes.

The low ionization efficiency of AHL in CHCA matrix makes relative quantification of the signals between MALDI-TOF MS spectra highly error prone. This is shown by the high standard deviations between spots of the same sample (Figure 4.23). However, despite these inherent drawbacks of MALDI-TOF MS, the short processing time and high sample throughput of the method facilitates the analysis of high numbers of technical and biological replicates. For example, the target plates for manual sample spotting used in the MALDI TOF/TOF 5800 MS instrument accommodate 384 samples. Compared to LC-ESI MS with the instrumentation described here (Chapter 4.3.4), the acquisition time to process a complete 384-well target is about equal to that of a single LC-MS analysis, in particular if only survey scans are needed. Even direct injection offline ESI MS analysis, omitting LC-separation, is more time consuming.

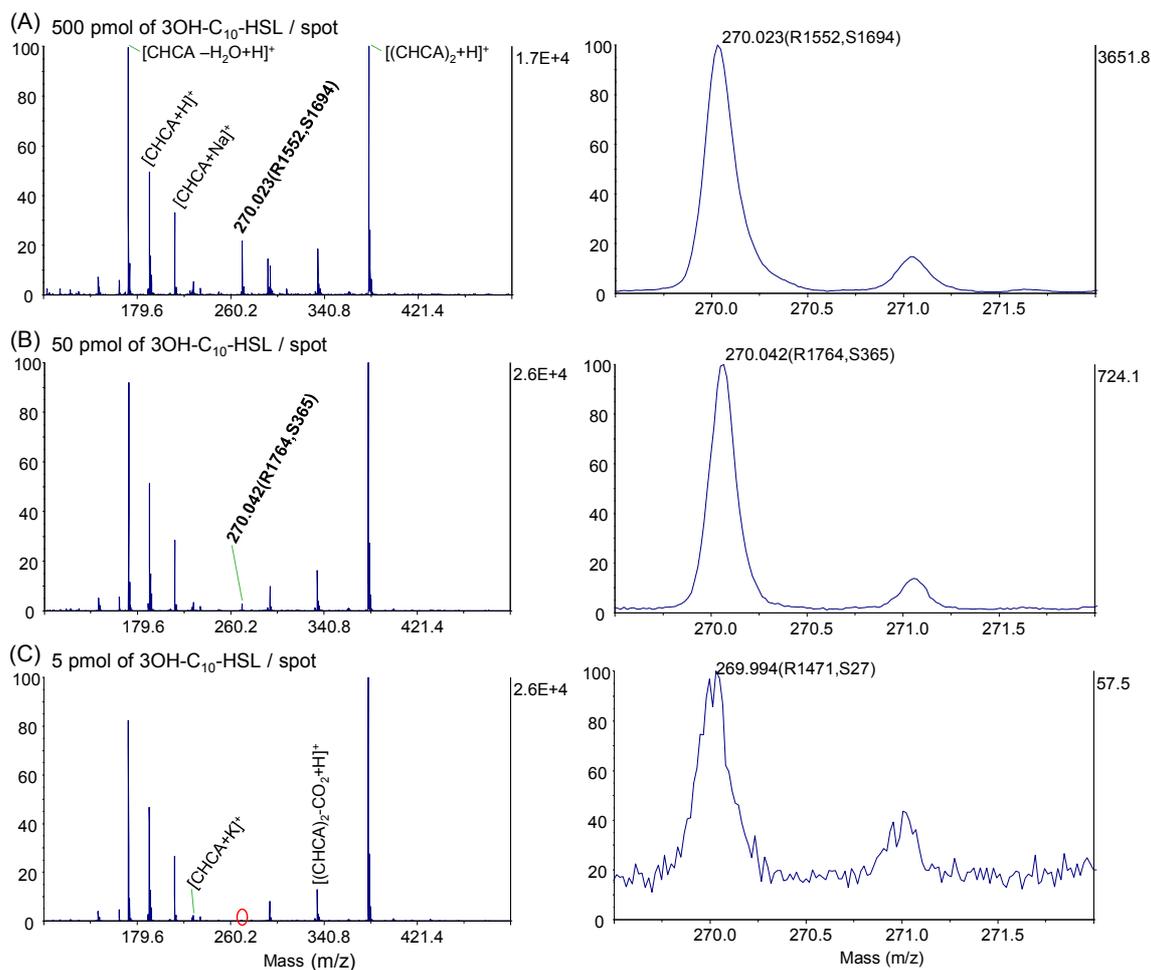


Figure 4.24: MALDI-TOF MS spectra of a dilution series of a 3OH-C₁₀-HSL standard. Spectra accumulating 2,000 shots of a narrow mass range from 100 *m/z* to 500 *m/z* were acquired from spots containing (A) 500 pmol, (B) 50 pmol and (C) 5 pmol of 3OH-C₁₀-HSL, respectively. The spectra on the left pane show the high intensity of interfering signals. Several signals observed in all spectra are labelled with the corresponding CHCA matrix ions in (A) and (C). The magnified excerpt of the *m/z* range around the 3OH-C₁₀-HSL protonated molecular ion (*m/z* 270.17) to the right illustrates the low ionization efficiency of AHL. At 5 pmol on target, 3OH-C₁₀-HSL shows a signal-to-noise ratio (S) of 27 (calculated using the software of the instrument manufacturer). A signal-to-noise ratio of 3 is generally accepted estimate for the limit of detection, commonly applied to analytical methods that exhibit baseline noise (ICH-Q2B 1996).

For ESI MS analysis by offline injection, each sample has to be manually injected into a single-use, coated glass needle emitter. To ensure spray stability, the approximately 1 mm diameter capillary has to be manually loaded without air bubbles, followed by manual reconciliation of several spray parameters such as spray voltage, temperature and needle position. On the other hand, the mass resolution and accuracy of MALDI-TOF MS are usually not sufficient to determine the elemental composition of unknown compounds, which is another major shortcoming compared

to high resolution LC-ESI MS. Altogether, these differences make the two instruments complementary and differently well suited for particular research applications.

After establishing the degradation of 3O-C₁₀-HSL by MALDI-TOF MS screening, a subset of the samples were chosen for high-resolution LC-ESI MS analysis. To identify the AHL and the *quorum* quenching derivatives, samples of *Hydra vulgaris* and control samples in Volvic medium incubated with 100 μM of 3O-C₁₀-HSL for 8 h were analyzed (Figure 4.25).

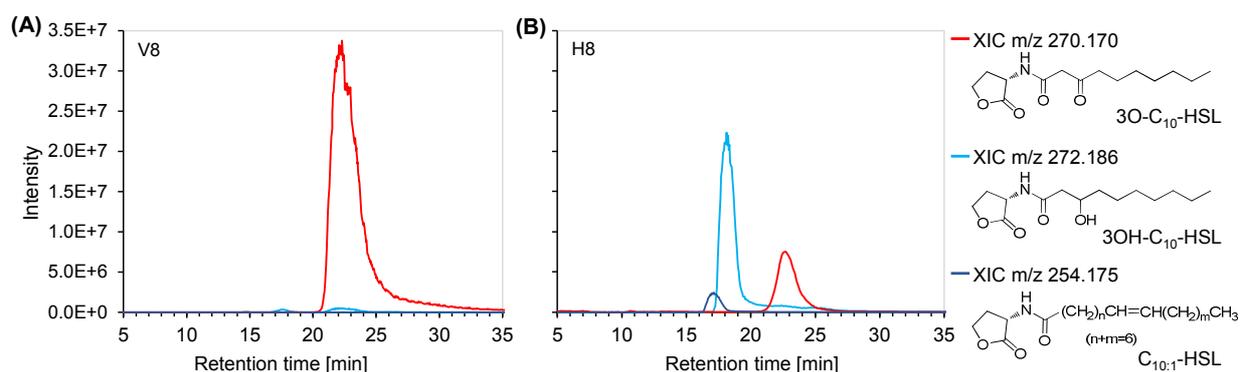


Figure 4.25: LC-ESI MS analysis of *Hydra quorum* quenching activity. (A) Analysis of the negative control with 3O-C₁₀-HSL in Volvic medium and (B) the *quorum* quenching assay in which the AHL was incubated with *Hydra*. The extracted ion chromatograms (XIC) of 3O-C₁₀-HSL (m/z 270.170), and the two possible *quorum* quenching products at m/z 272.186 and m/z 254.175 are shown in red, light blue and dark blue, respectively.

Consistent with the results obtained by MALDI-TOF MS analysis, extracted ion chromatograms of the protonated 3O-C₁₀-HSL from LC-ESI MS analysis showed lower signal of the AHL-substrate after incubation with *Hydra*, compared to the control sample in medium without animals. However, the two ions of interest previously detected by MALDI-TOF MS (m/z 272.186 and m/z 254.175) showed inverse proportions, with LC-ESI MS analysis suggesting the ion at m/z 272.186 as the main *quorum* quenching product, whereas the signal at m/z 254.175 was much stronger in MALDI-TOF MS analysis. MS fragmentation of the signal at 270.170 m/z confirmed the unmodified substrate 3O-C₁₀-HSL. Fragmentations leading to the formation of the protonated homoserine lactone fragment (m/z 102.055) and the acylium ion of the side-chain (m/z 169.122) were consistent with the reference spectra of the 3O-C₁₀-HSL (see Figure 4.16). The two candidates for *quorum* quenching products were also chosen as precursors for fragmentation. The identity of both signals could be determined by characteristics of their fragmentation (Figure 4.26).

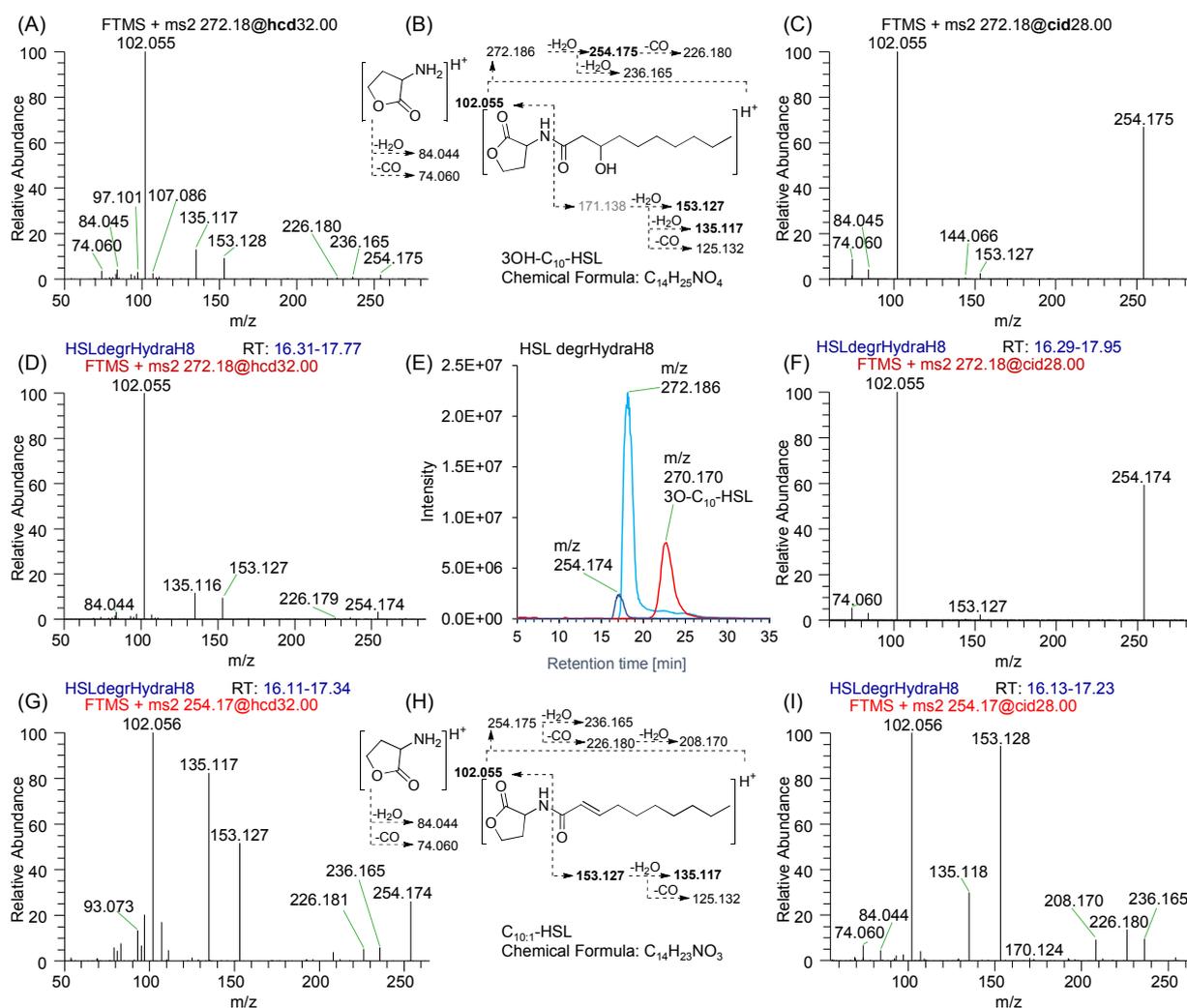


Figure 4.26: MS/MS analysis of *Hydra quorum* quenching products. (A-C) The HCD and CID spectra of the 3OH-C₁₀-HSL standard match the corresponding spectra of the precursor at *m/z* 272.186 (D) and (F). (E) MS/MS precursor ions were selected from the LC-ESI MS analysis of the *quorum* quenching activity with *Hydra* after incubation for 8 h. (G-I) In contrast, fragmentation of the signal at *m/z* 254.175 corresponds to the water-loss signal from 3OH-C₁₀-HSL resulting in the unsaturated C_{10:1}-HSL. The fragmentation scheme (H) shows the identified fragment signals, however the location of the double bond could not be verified.

HCD of the potential *quorum* quenching product at *m/z* 272.186 resulted in three main signals at *m/z* 102.055, *m/z* 135.117 and *m/z* 153.127. By comparison with HCD spectra of AHL-standards the *quorum* quenching product could be unambiguously identified as 3OH-C₁₀-HSL, with the main fragments matching the homoserine lactone fragment ion and the acylium fragment ion of the side-chain after consecutive neutral losses of water (Figure 4.26 A and D). The identity was also confirmed by the characteristic CID spectra, conform to reference spectra, with a dominant signal for the water-loss of the protonated molecular ion at *m/z* 254.175, in addition to the signals at *m/z* 102.055 (Figure 4.26 C and F). The second ion of interest at *m/z* 254.175, initially also considered

as a *quorum* quenching product, was found to be an analytical artifact. The signal corresponds to an unsaturated C_{10:1}-HSL, same as the major CID product in spectra of 3OH-C₁₀-HSL due to the neutral loss of water from the protonated molecular ion (Figure 4.26 F and G). Both signals at *m/z* 102.055 and *m/z* 254.175 were commonly observed in the same MS spectra together with the AHL parent signal. This suggests that they are artifacts, potentially produced by in-source fragmentation of 3OH-C₁₀-HSL. This is further confirmed by the co-elution of *m/z* 254.175 and *m/z* 272.186 (Figure 4.26 E). The same signals were also found in the 3OH-AHL standard samples. These observations highlight that even the so called “soft ionization methods”, routinely used for the analysis of biomolecules, may generate artifacts in certain compounds. 3OH-AHL in particular seem to be less stable during the MALDI than the ESI process.

MS data showed conclusively, that *Hydra* reduces the long-chain AHL 3O-C₁₀-HSL (*m/z* 270) to 3OH-C₁₀-HSL (*m/z* 272). Oxidoreductase enzymes discovered in bacteria (Uroz, Chhabra et al. 2005, Chan, Atkinson et al. 2011) and a short-chain dehydrogenase isolated from a soil metagenome (Bijtenhoorn, Schipper et al. 2011) have been previously shown to catalyze this conversion. Although the *quorum* quenching mechanism in *Hydra* has yet to be characterized at the molecular level, the results of this study provide the first example of an equivalent *quorum* quenching mechanism in a eukaryote system.

4.4.5 Identification of *quorum* sensing signals produced by AHL-synthases

The discovery of *quorum* quenching in *Hydra* simultaneously raised the question how this host mechanism affects *Hydra* as a metaorganism. For this reason, the study was directed towards the *quorum* sensing systems of *Hydras* communal bacteria. By cDNA sequence analysis, two AHL-synthases were discovered in *Curvibacter*, the main commensal colonizing microbe of the *Hydra* holobiont. The coding genes for these two enzymes were expressed in *E. coli* and the transformants were cultivated under conditions stimulating AHL production as described by Gould, Herman et al. (2006). Culture media were then extracted with ethyl acetate. After sample cleanup by solid phase extraction, the product spectrum of the two enzymes was determined by LC-ESI MS. In the LC-ESI MS runs of the samples from *Hydra quorum* quenching experiments, extensive peak tailing and broad signal peaks were observed (see Figure 4.25).

To accommodate a more complex AHL mixture and to test if chromatographic resolution could be increased, the HPLC eluent flow was changed from isocratic to a gradient and the runtime was increased from 40 min to 90 min. The same LC-MS acquisition parameters were used for the construction of the reference spectra library, to acquire correlating retention-times and fragment spectra for the commercial AHL standards (see Table 4.2). Although the overall performance of

HPLC separation significantly improved with the new parameters, peak tailing was still observed for several signal peaks. Furthermore, isobaric contaminants eluted prior to 3OH-C₁₁-HSL and 3OH-C₁₃-HSL signals (Figure 4.27). These issues did not prevent AHL identification.

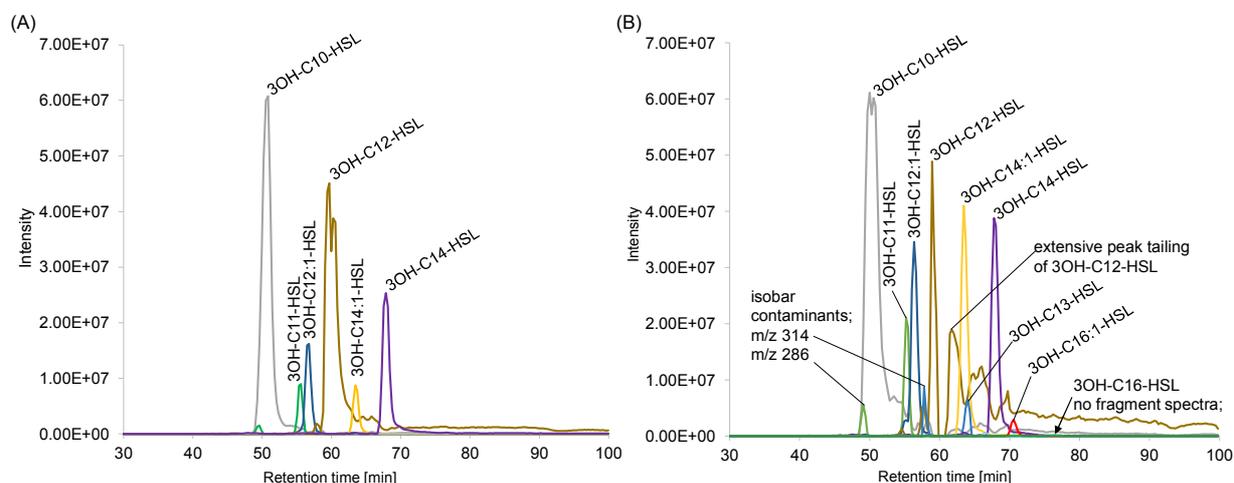


Figure 4.27: LC-MS data of AHL produced by *Curvibacter*. (A) AHL-synthase-1 and (B) AHL-synthase-2 expressed in *E. coli*. The extracted ion chromatograms show that while the proportion of the produced AHL compounds differ, the product range of both synthases is very similar. AHL 3OH-C₁₀-HSL (m/z 272.186) shows a peak at retention-time (RT) of 50.47 min (grey). AHL 3OH-C₁₂-HSL (m/z 300.217) elutes at 59.33 min (ochre) and 3OH-C₁₄-HSL (m/z 328.248) elutes at 67.74 min (violet). Fragment spectra and retention-times were in accordance with LC-MS data acquired using AHL standards. Five additional AHL could be assigned by analyzing fragment spectra. Two AHL with unusual side-chain lengths were detected with lower intensities. 3OH-C₁₁-HSL (m/z 286.201) which a RT of 55.21 min and 3OH-C₁₃-HSL (m/z 314.232) with a RT of 64.03 min both fit into the elution series of 3OH-AHL derivatives and show the corresponding fragmentation behavior. The three remaining AHL signals were at m/z 298.202, m/z 326.233 and m/z 354.264. Fragment spectra for these precursors were completely different from the reference spectra of 3O-C₁₂- and 3O-C₁₄-HSL and could be identified as the isobaric singly unsaturated 3OH-AHL compounds, 3OH-C_{12:1}-HSL, 3OH-C_{14:1}-HSL and 3OH-C_{16:1}-HSL, with retention times of 56.34 min, 63.45 min and 70.59 min, respectively.

Both *Curvibacter* AHL-synthases produced a spectrum of relatively long-chained 3OH-AHL signal molecules in *E. coli*. By comparison with the retention-times and the MS/MS spectra obtained from commercial 3OH-AHL compounds with chain length of C₁₀, C₁₂, and C₁₄, corresponding signals could be unambiguously identified in both *E. coli* supernatant extracts (see Table 4.2).

Two AHL derivatives with unusual chain lengths were identified. The 3OH-C₁₁-HSL (m/z 286.201) was detected in both *Curvibacter* AHL-synthase samples and 3OH-C₁₃-HSL (m/z 314.232) was detected exclusively in the extract of supernatants from the culture of *E. coli* transformed with AHL-synthase-2. The corresponding fragment spectra followed the fragmentation behavior of even chain length 3OH-AHL and fit into the elution series with successive retention-times observed for 3OH-AHL standards.

The expression system for AHL-synthases in *E. coli* used here was described by Gould, Herman et al. (2006). In the original study, the system was used to analyze AHL-synthase LasI from *P. aeruginosa*. Transformants were found to mainly produce even-chained acyl-AHL and 3O-AHL from C₁₀ to C₁₄. However, AHL with odd chain length, in particular 3O-C₁₁-HSL, were also detected. These were discussed as probable artifacts from the heterologous expression system, e.g. due to the use of propionyl-ACP as a precursor in the acyl-ACP elongation instead of acyl-ACP.

Signals corresponding to the *m/z* values and retention times of 3O-C₁₂-HSL (*m/z* 298.202) and 3O-C₁₄-HSL (*m/z* 326.233) were detected in both AHL-synthase samples. AHL-synthase-2 showed an additional signal corresponding to 3O-C₁₆-HSL (*m/z* 354.264). However, comparison of the acquired MS/MS spectra showed differences to the fragmentation behavior of 3O-AHL compounds and revealed the isobaric singly unsaturated 3OH-AHL compounds. Detailed fragment spectra of the reference 3O-C₁₂-HSL and the isobaric AHL produced by heterologously expressed *Curvibacter* AHL-synthase-2 are shown in Figure 4.28 and Figure 4.29, respectively.

As discussed previously, the two most intense signals observed in HCD spectra of 3O-AHL resulted from fragmentation at the amide bond, while the third strongest signal represented the second acylium ion from fragmentation between the α -carbon and the C₃-carbonyl position. For 3O-C₁₂-HSL these fragments generate signals at *m/z* 102.055; *m/z* 197.154 and *m/z* 155.143 (Figure 4.28).

These characteristic signals for 3O-C₁₂-HSL were also observed in HCD spectra of the isobaric compound detected in the supernatant extracts of transformed *E. coli*. However, the fragment ion at *m/z* 197.154 was of much lower intensity and dominated by the corresponding water-loss at *m/z* 179.143.

The most intense signals, besides the base peak of the homoserine lactone moiety at *m/z* 102.055, were from fragmentation between the α -carbon and the C₃-carbon at *m/z* 155.143 and the corresponding water-loss at *m/z* 137.133. HCD spectra of precursor 298.191 from AHL-synthase-2 samples also differed from HCD spectra of the 3O-C₁₂-HSL standard by a more extensive and higher intensity ion series with 2.5 RDBE (*m/z* 81.069; 95.084; 109.100; 137.131; 151.146) and the occurrence of highly unsaturated hydrocarbon ion series with 3.5 RDBE and even 4.5 RDBE (Figure 4.29). The differences between these two spectra support the identification of the unknown AHL as 3OH-C_{12:1}-HSL rather than 3O-C₁₂-HSL. The strong signal for the water-loss from both the protonated molecular ion and from the side-chain acylium ion, similarly observed in fragment spectra of saturated 3OH-AHL standards, support the presence of a hydroxy-group. For a direct comparison, the annotated fragment spectra of 3OH-C₁₂-HSL are shown in Figure 4.30.

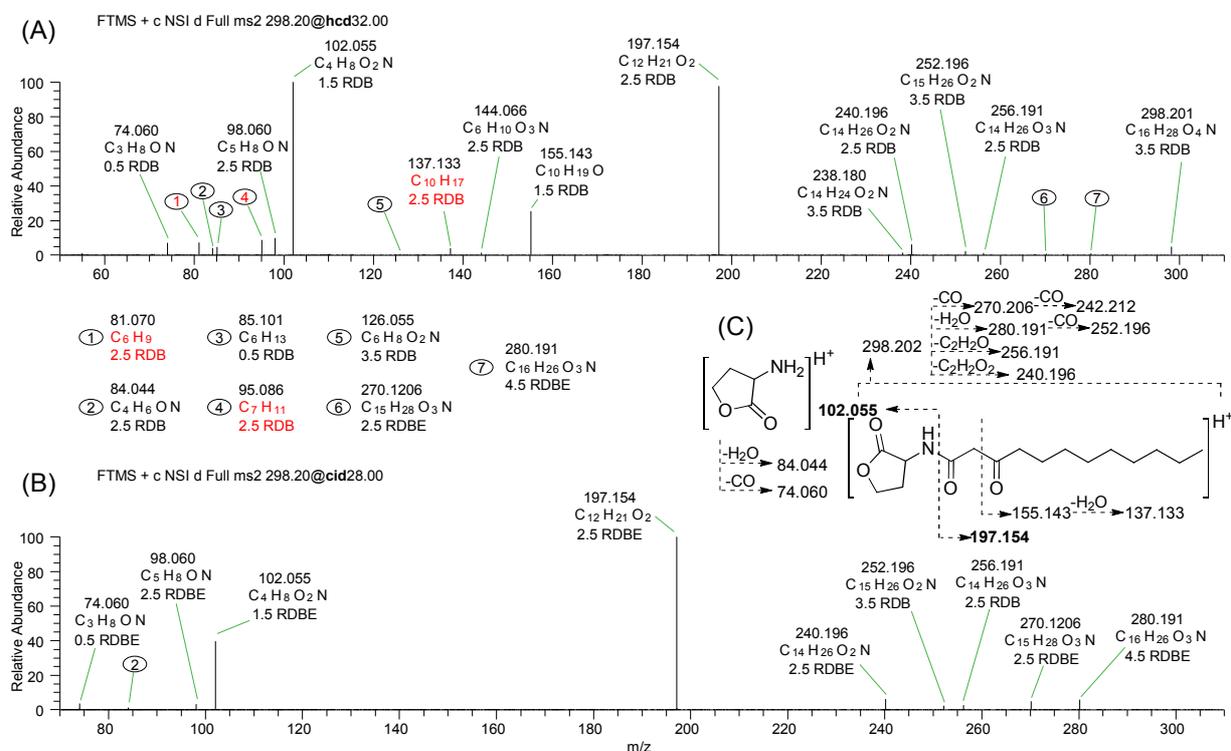


Figure 4.28: Annotated fragment spectra of 3O-C₁₂-HSL. (A) FTMS/MS spectra of the 298.202 *m/z* precursor ion at a retention-time of 61.06 min fragmented by HCD with NCE of 32; and (B) by CID with NCE of 28 are shown. (C) A fragmentation scheme of 3O-C₁₂-HSL is shown in-between. The elemental composition and ring double bond equivalents (RDBE) for all peaks down to a relative abundance of 0.5 % are annotated. Fragmentation at the amide bond creates the strongest signals, with the homoserine moiety as the base peak upon HCD and the acylium ion representing the base peak in the CID spectrum. HCD fragment spectra are more populated in the lower *m/z* range and show a higher overall number of identifiable signals. CID fragment spectra are complementary; fewer signals are observed, with stronger signal intensities in the higher *m/z* range resulting from the characteristic consecutive eliminations of water and carbon monoxide. Few weak ions of the series corresponding to alkene hydrocarbon fragments of the composition C_nH_{2n-3} (red) were detected.

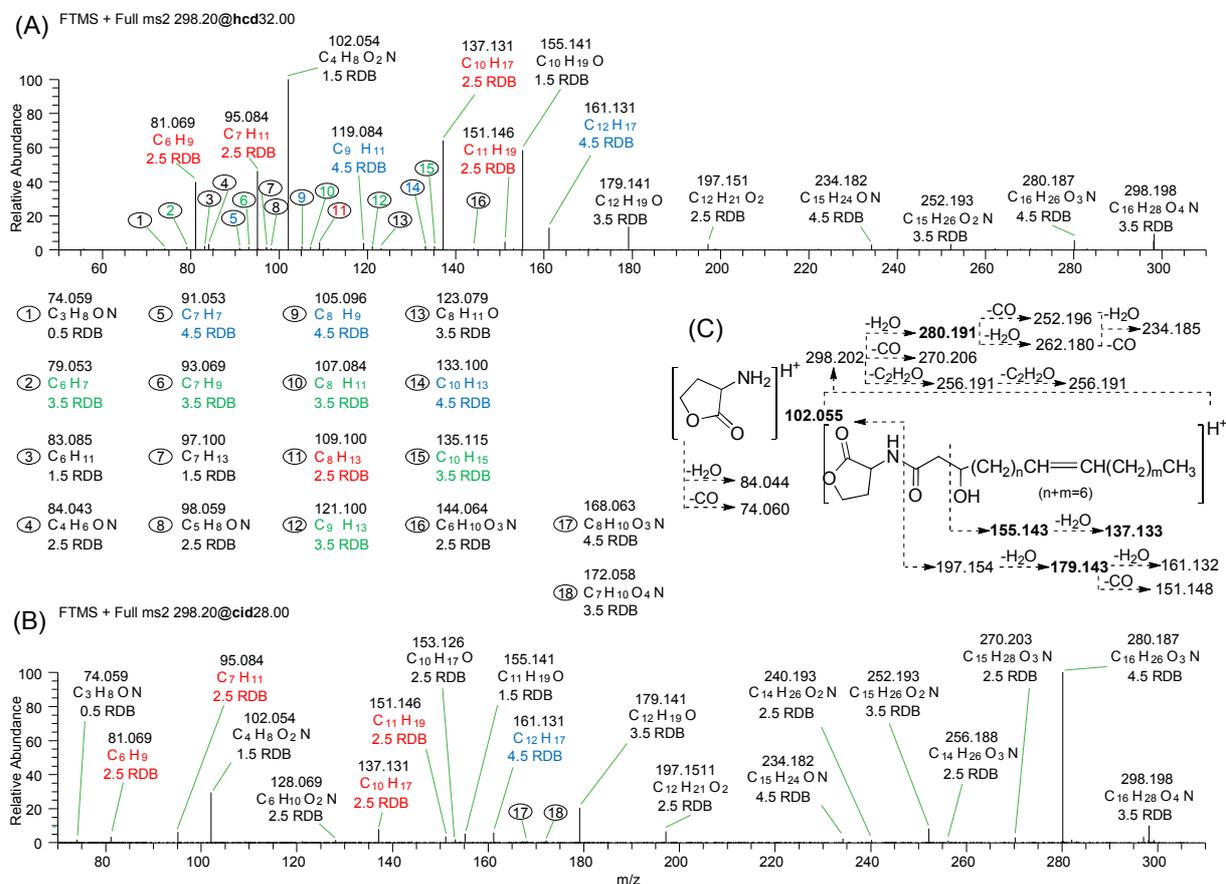


Figure 4.29: Annotated fragment spectra identifying 3OH-C_{12:1}-HSL. (A) FTMS/MS spectra of the 298.202 *m/z* precursor ion at a retention-time of 56.72 min acquired by HCD with normalized collision energy (NCE) of 32; and (B) by CID with NCE of 28 are shown. (C) A fragmentation scheme of the identified 3OH-C_{12:1}-HSL is shown in-between. The elemental composition and ring double bond equivalents (RDBE) for all peaks down to a relative abundance of 1 % are annotated. HCD spectra are more populated in the lower *m/z* range and show a higher overall number of identifiable signals. CID spectra are complementary; fewer signals are observed, with stronger signal intensities in the higher *m/z* range resulting from the characteristic consecutive losses of water and carbon monoxide. An extensive ion series correlating to alkene hydrocarbon fragments of the composition C_nH_{2n-3} (red), is accompanied by two weaker, highly unsaturated series with 3.5 RDBE (green) and 4.5 RDBE (blue).

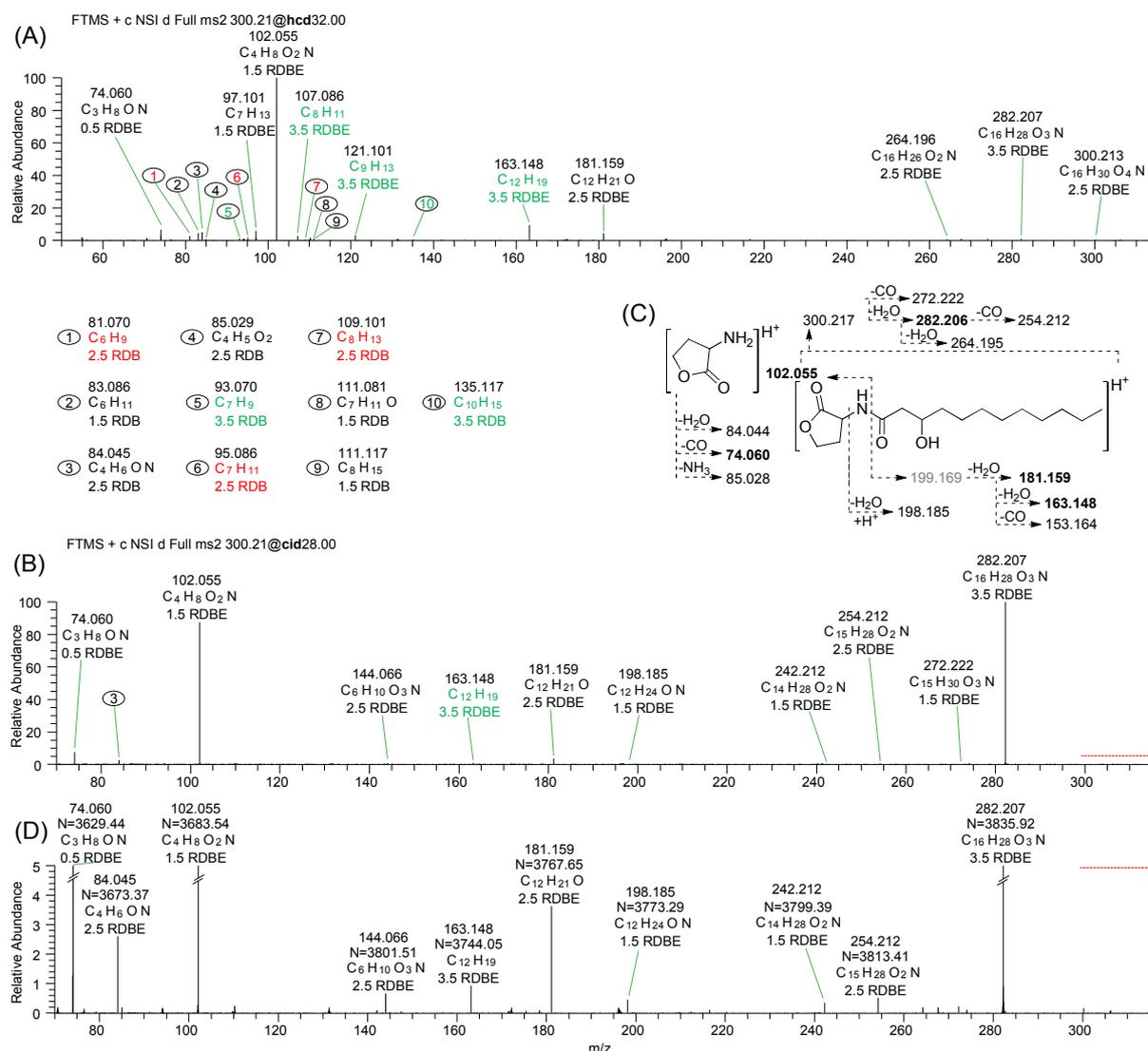


Figure 4.30: Annotated fragment spectra of 3OH-C₁₂-HSL. (A) FTMS/MS spectra of the 300.22 *m/z* precursor ion at a retention-time of 57.67 min fragmented by HCD with NCE of 32; and (B) by CID with NCE of 28 are shown. (C) A fragmentation scheme of 3OH-C₁₂-HSL is shown in-between. The elemental composition and ring double bond equivalents (RDBE) for all peaks down to a relative abundance of 0.5 % are annotated. Fragmentation at the amide bond and the loss of water from the protonated molecular ion generate the strongest signals. The protonated homoserine lactone fragment is very dominant after HCD and the ion resulting from water-loss represents the base peak upon CID. HCD spectra show two extensive ion series with 2.5 RDBE and 3.5 RDBE with low intensity. (D) Shows an excerpt of the low intensity range up to 5 % relative abundance of the CID spectrum in (B), to illustrate the difference between low intensity signals from noise. The estimated noise level (N) is indicated at each signal and amounts to ca. 3,700, whereas the smallest annotated peak (at *m/z* 198.185) has an absolute intensity of over 10,000. This amounts to a relative intensity of around 0.5 % of the base peak and a signal to noise ratio of ca. 2.7. For reference, the base peak at *m/z* 282.207 has an absolute intensity of 2.2 E7. The noise was calculated using the proprietary algorithm implemented in the Xcalibur software.

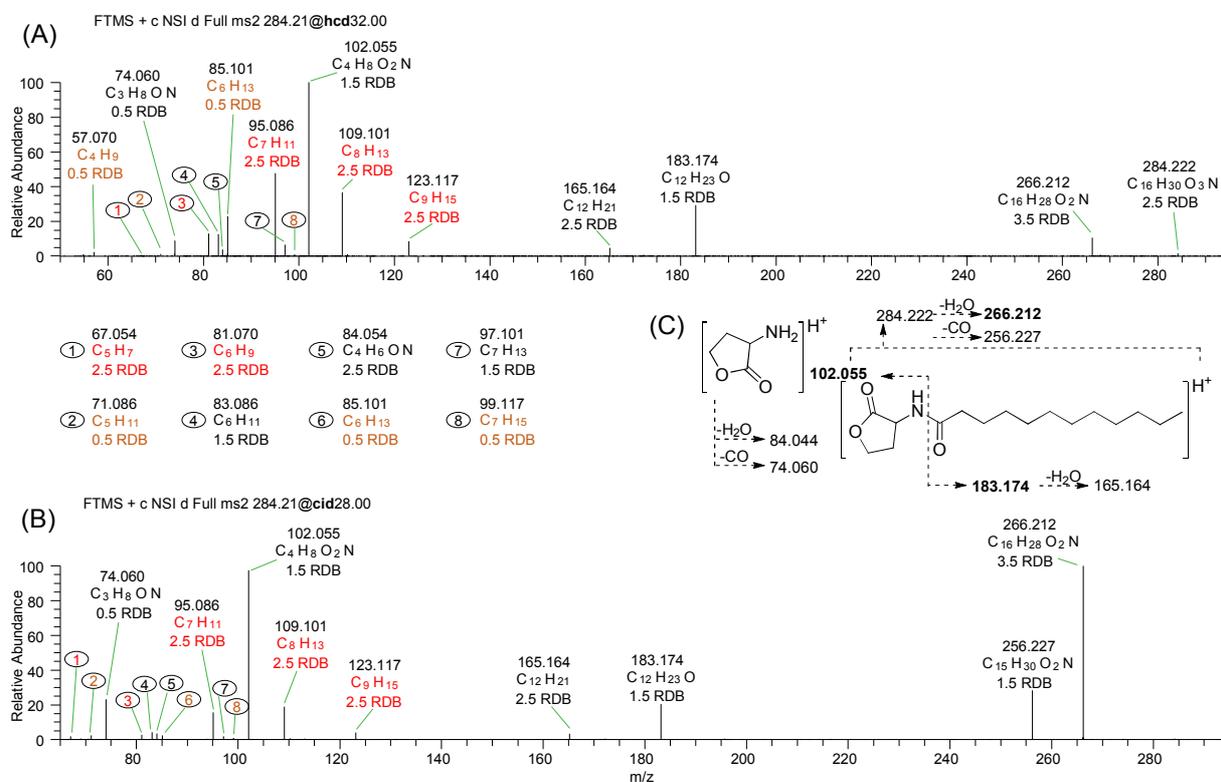


Figure 4.31: Annotated fragment spectra of C₁₂-HSL. FTMS/MS spectra of the 284.21 *m/z* precursor ion at a retention-time of 68.45 min fragmented by HCD with NCE of 32 (A); and by CID with NCE of 28 (B) are shown. A fragmentation scheme of C₁₂-HSL is shown in-between (C). The elemental composition and ring double bond equivalents (RDBE) for all peaks down to a relative abundance of 1 % are annotated. Fragmentation at the amide bond and the loss of water produce the strongest signals, with the homoserine lactone moiety as the base peak in HCD and again the product of water elimination from the protonated molecular ion as the base peak upon CID. The most prominent ion series corresponds to alkene hydrocarbon fragments of the composition C_nH_{2n-3} from C₅ to C₉ with 2.5 RDBE (red). A weaker series with 0.5 RDBE from C₄ to C₇ (orange) was identified as well.

The fragmentation behavior of 3OH-C₁₂-HSL (RDBE 3.5) with a saturated side-chain further corroborates this identification (see Figure 4.30). The observed hydrocarbon ion series are less dominant and show a lesser degree of unsaturation up to 3.5 RDBE as compared to an ion series with 4.5 RDBE detected in fragment spectra of the supposed singly unsaturated 3OH-C_{12:1}-HSL (RDBE 4.5). The signals at *m/z* 326.233 detected in both AHL-synthase samples and *m/z* 354.264 detected only in AHL-synthase-2 were accordingly identified as 3OH-C_{14:1}-HSL and 3OH-C_{16:1}-HSL rather than the isobaric 3O-AHL.

Although LC-MS data support the presence of an additional double bond in these compounds, the location of the double bond could not be determined based on HCD or CID spectra.

The wide consensus from literature is that double bond assignment in long alkenes or fatty acids by mass spectrometry is not trivial. Reliable location of double bonds often requires more elaborate analysis methods such as chemical derivatization (Francis 1981, Reaney, Liu et al. 2001) or special MS techniques, e.g. low energy ionization and charge-remote fragmentation of dilithiated fatty acid adducts (Hsu and Turk 1999, Zhang, Tao et al. 2011). These methods were primarily developed for lipidomics studies to locate double bonds in unsaturated fatty acids or lipids and have yet to be applied to, or optimized for, AHL compounds. One method for double bond location, the fixation of the double bond by electrophilic methylthiolation with dimethyl disulfide (Shibahara, Yamamoto et al. 2008), was unsuccessfully tested and further method development was outside the scope of this study. Definite assignment of the double bond location and conformation of well purified, unsaturated AHL has been achieved by nuclear magnetic resonance spectroscopy (Schripsema, de Rudder et al. 1996, Puskas, Greenberg et al. 1997).

Generally, HCD of long-chained AHL produced low intensity signals of extended ion series. Interestingly, these hydrocarbon ions seem to be characteristic for both the substituent at C₃-position and the saturation of the acyl-chain. HCD of 3O-C₁₂-AHL created the 2.5 RDBE ion series of C₆, C₇ and C₁₀ (*m/z* 81.070; 95.086; 137.133). HCD of 3OH-C₁₂-HSL produced the series of C₆, C₇ and C₈ with 2.5 RDBE (*m/z* 81.070; 95.086; 109.101) and C₇, C₈, C₉, C₁₀ and C₁₂ with 3.5 RDBE (*m/z* 93.070; 107.086; 121.101; 135.117; 163.148).

As discussed earlier, 3OH-C₁₂-AHL is prone to the neutral loss of water, resulting in C_{12:1}-AHL (3.5 RDBE). Accordingly, the more unsaturated 3.5 RDBE hydrocarbon ions are detected. In the fragment spectra identifying 3OH-C_{12:1}-HSL, hydrocarbon ions are most extensive and included 2.5 RDBE ions of C₆, C₇, C₈, C₁₀ and C₁₁ (*m/z* 81.070; 95.086; 109.101; 137.133; 151.146) and the ions from C₆ to C₉ with 3.5 RDBE (*m/z* 79.053; 93.069; 107.084; 121.100; 135.115). Additionally, and congruent with a neutral loss of water from 3OH-C_{12:1}-HSL (3.5 RDBE) to 3O-C_{12:2}-HSL at *m/z* 280.191 (4.5 RDBE) as the main product of CID, 4.5 RDBE ions of C₇, C₈, C₉, C₁₀ and C₁₂ (*m/z* 91.053; 105.096; 119.084; 133.100; 161.131) were produced (see Figure 4.29). Finally, the annotated spectrum of the unsubstituted and saturated C₁₂-HSL is singular in displaying the C₄ to C₇ ions with only 0.5 RDBE (*m/z* 57.070, 71.086, 85.101, 99.117) in addition to the C₅ to C₉ members of the 2.5 RDBE series (*m/z* 67.054; 81.070; 95.086; 109.101, 123.117) (see Figure 4.31).

4.5 Conclusion

The high mass resolution of FTMS instruments enables unambiguous identification of most AHL based on m/z ratios of the protonated molecular ion alone. The mass accuracy of FTMS makes the mass difference between e.g. 3O- C_x -HSL and C_{x+1} -HSL (0.0364 Da) easily sufficient to distinguish the respective signals.

In general, MS/MS fragmentation of AHL compounds with the different substituents at C_3 position leads to the formation of distinct patterns for the two strongest peaks. HCD with a collision energy high enough for complete or near complete decomposition of the here represented AHL compounds consistently produced the base peak at m/z 102.055, corresponding to the protonated homoserine lactone fragment. 3O-AHL derivatives showed the most consistent HCD fragmentation pattern. Following the accepted scheme (see Figure 4.8), the acylium ion from fragmentation of the amide bond generates the second strongest signal. The same was generally observed for acyl-AHL derivatives. However, fragmenting the short-chained C_4 -AHL did not produce a signal corresponding to the side-chain acylium ion. Furthermore, for long-chained AHL, the side-chain acylium ion is less stable and subject to further decomposition resulting in more intense signals for several hydrocarbon ions. Most distinct is the difference for 3OH-AHL. Due to the presence of a hydroxy function, the acyl-chain ion is replaced by a dominant signal of the corresponding water-loss.

For CID, the homoserine lactone fragment did not always represent the base peak signal. Instead, all spectra showed a significant bias towards larger ions. For the 3O-AHL derivatives, this results in more dominant acylium ions with expanding chain length. Both 3OH-AHL and the acyl-AHL compounds showed signals of increasing intensity for the water loss from the protonated molecular ion instead. CID spectra generally showed fewer signals and the low intensity hydrocarbon ion series were detected in particular in the HCD spectra. The detected ion series are also determined by characteristics of the acyl-chain and change with the degree of saturation and the substituent at C_3 .

The few isobaric AHL which cannot be distinguished on the basis of m/z ratios alone are 3O- C_x -HSL with saturated side-chains and the corresponding unsaturated 3OH- C_{x-1} -HSL. 3O-AHL compounds with C_{12} - and C_{14} -chains were represented in the analyzed AHL standards and unsaturated 3OH-AHL were discovered in AHL-synthase extracts. HPLC separations with the setup did not result in a significant shift in retention-times of these isobaric compounds, leaving only fragment spectra as a means to identify these compounds. Characteristic fragmentation patterns were indeed observed for the saturated 3O-AHL and isobaric unsaturated 3OH-AHL

counterparts. Most distinctive, unsaturated 3OH-AHL show very strong signals for the water-loss from the protonated molecular ion, a common observation for organic compounds containing a hydroxy-group (McLafferty and Tureček 1993). For CID of unsaturated 3OH-AHL, this neutral loss even replaces the homoserine lactone fragment at m/z 102.055 as the base peaks. Strong signals for the neutral loss of water are observed for both the protonated molecular ion as well as the acylium ion created by fragmentation at the amide bond. Although AHL derivatives with unsaturated side-chains could be detected, the location and configuration of the double bond could not be determined based on HCD and CID spectra. This confirms that for confident assignment of double bond locations and conformation, well purified and highly concentrated AHL samples have to be generated and analyzed by nuclear magnetic resonance spectroscopy as described earlier (Schripsema, de Rudder et al. 1996, Puskas, Greenberg et al. 1997).

The three analysis techniques mostly employed for AHL detection, identification and quantification have been discussed in the introduction. AHL receptor strains can detect low femtomole amounts of AHL in unpurified extracts, however, the receptors usually detect a range of AHL which makes unambiguous identification more difficult (Shaw, Ping et al. 1997). Well-purified samples are needed for definite structure determination by NMR spectroscopy, ideally containing micromole amounts of a single AHL. Compared to these techniques, LC-ESI MS is intermediate in terms of sensitivity and sample amount requirements, as complex AHL mixtures can be analyzed. In very pure samples of reference compounds, LC-ESI MS and MALDI MS could detect low picomole amounts of AHL. The detection of such low amounts of AHL in biological extracts is less certain, due to ion-suppression and interference by contaminants. However, the unambiguous identification of most AHL and the possibility of AHL quantification make LC-MS highly valuable for the bioanalytical exploration of AHL based *quorum* sensing.

The analysis of *Hydra quorum* quenching experiments by MALDI-TOF MS confirmed the results obtained with bacterial reporter strains, showing that *Hydra* polyps destroy specific *quorum* sensing signals. ESI MS analysis revealed that this *quorum* quenching ability is achieved by a mechanism to date not discovered in eukaryotes. The identification of the *quorum* quenching product indicates that *Hydra* reduces the long-chained 3O-C₁₀-AHL to 3OH-C₁₀-AHL. Although this catalytic activity has previously been described for a short-chained dehydrogenase of metagenomic origin (Bijtenhoorn, Schipper et al. 2011), this mechanism is hardly characterized as compared to other AHL-degrading enzymes. The disruption of specifically 3O-C₁₀-HSL and more generally long-chained 3O-AHL signals is accomplished by a wide spectrum of AHL acylases and AHL lactonases of a diverse bacterial origin as reviewed elsewhere (Lade, Paul et al. 2014). For several bacteria, including *P. aeruginosa*, *Er. carotovora*, *V. anguillarum* and *A.*

hydrophila, the phenotypes resulting from 3O-AHL signal disruption have also been analyzed previously (Chen, Zhou et al. 2010, Chan, Atkinson et al. 2011, Torres, Romero et al. 2013). Affected processes frequently include an inhibition or attenuation of pathogenicity or virulence, the reduction of biofilm formation or of the microbes swarming ability.

The discovery of *Hydras* ability to quench specific autoinducer signals raises the question as to what implication this mechanism has on the metaorganism level. To gain insights on its communal bacteria, the *quorum* sensing systems of the major colonizer *Curvibacter* were characterized. The catalytic capabilities of the encoded AHL-synthases of two luxI family genes isolated from *Curvibacter* were analyzed after heterologous expression in *E. coli*. A very similar product spectrum, consisting exclusively of 3OH-AHL, was detected for both enzymes. Although several AHL with different chain lengths were detected, including 3OH-C₁₀-HSL, 3OH-C₁₂-HSL and 3OH-C₁₄-HSL, the enzymes seem to specifically process acyl-ACP substrates with an acyl-chain length of at least C₁₀.

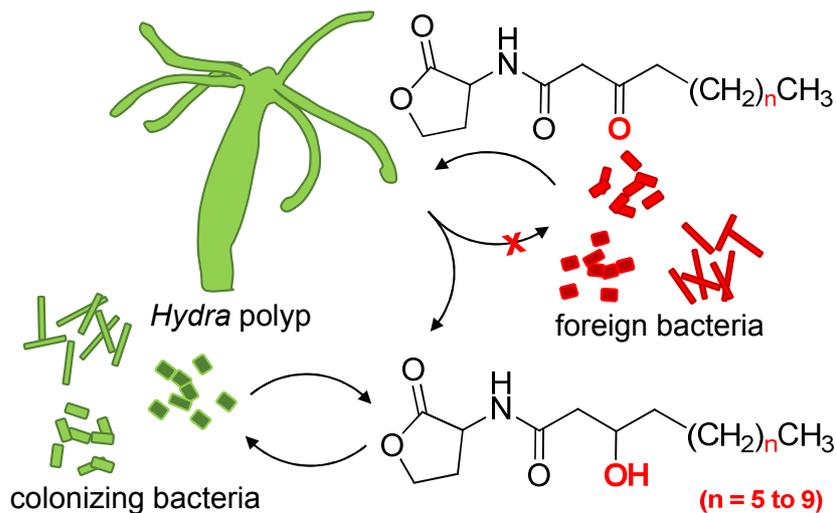


Figure 4.32: Proposed interdependent *quorum* regulation network of the *Hydra* metaorganism. *Hydra* selectively subverts a frequently used AHL signal, generating the less common AHL signal the indigenous microbe population utilizes. This mechanism may exert negative selective pressure on bacteria foreign to the system.

The discovery of a highly customized and interconnected *quorum* quenching and *quorum* sensing network in a host organism and its bacterial colonizer is very intriguing (Figure 4.32). The microenvironment *Hydra* provides for commensals appears to be specifically shaped by *quorum* quenching of long-chained 3O-AHL signals. In fact, the high turnover rate observed for single *Hydra* polyps implies microecological effects, which bear further investigation. This interference with widely used AHL signals may be a determining factor for the makeup and assembly of the

hydra microbiota. Furthermore, the coordinated maintenance of a 3OH-AHL based *quorum* sensing homeostasis, which relies both on the host and its bacterial colonizers, may provide these colonizers with a considerable selective advantage over the numerous bacteria in possession of the 3O-AHL type *quorum* sensing system. This selective pressure could be a factor contributing to the high temporal stability observed of the *Hydra* microbiota or may support the remarkable regenerative ability of germ free animals to re-establish holobiont communities with a very similar bacterial composition, as described earlier (Fraune and Bosch 2007).

To conclude, in this study, MALDI MS and LC-ESI MS methods for the analysis of N-acyl-homoserine lactones were established and applied to biological samples. However, it also became apparent, that further improvements in terms of separation and MS analysis methods may be possible. MALDI MS optimization could entail testing the efficiency of different matrices, including ionic liquids (Tholey and Heinzle 2006, Leipert, Treitz et al. 2017), to improve ionization and reduce in-source fragmentation of unstable AHL derivatives such as the 3OH-AHL. For LC-ESI MS, the application of different organic solvents could be tested to optimized separation performance and reduce sample carryover in subsequent measurements, if possible. In particular for MALDI-MS analysis, compatible TLC sheets are commercially available and represent a promising alternative to HPLC for the separation of AHL mixtures prior to MS analysis. Finally, the synthesis of isotope labelled standards for absolute quantification of 3O-, 3OH and acyl-AHL compounds in biological samples should be established and applied to analyze the commensal and competing gram negative bacteria encountered in the natural habitat of *Hydra*.

Appendix

References Chapter 1.

Biemann, K. (1988). "Contributions of mass spectrometry to peptide and protein structure." Biomed Environ Mass Spectrom **16**(1-12): 99-111.

Cassidy, L., D. Prasse, D. Linke, R. A. Schmitz and A. Tholey (2016). "Combination of Bottom-up 2D-LC-MS and Semi-top-down GelFree-LC-MS Enhances Coverage of Proteome and Low Molecular Weight Short Open Reading Frame Encoded Peptides of the Archaeon *Methanosarcina mazei*." J Proteome Res **15**(10): 3773-3783.

Cox, J., M. Y. Hein, C. A. Lubner, I. Paron, N. Nagaraj and M. Mann (2014). "Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ." Mol Cell Proteomics **13**(9): 2513-2526.

Delmotte, N., M. Lasoosa, A. Tholey, E. Heinzle and C. G. Huber (2007). "Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis." J Proteome Res **6**(11): 4363-4373.

Dong, M. W. (2013). "Myths in Ultrahigh-Pressure Liquid Chromatography." Lc Gc North America **31**(10): 868-+.

Gilar, M., P. Olivova, A. E. Daly and J. C. Gebler (2005). "Orthogonality of separation in two-dimensional liquid chromatography." Anal Chem **77**(19): 6426-6434.

Gregorius, B., T. Jakoby, D. Schaumlöffel and A. Tholey (2013). "Metal labeling for accurate multiplexed peptide quantification via matrix-assisted laser desorption/ionization mass spectrometry." Analytical and Bioanalytical Chemistry **405**(8): 2735-2741.

Griffiths, J. R. and R. D. Unwin (2016). Analysis of Protein Post-Translational Modifications by Mass Spectrometry, Wiley.

Hillenkamp, F., M. Karas, R. C. Beavis and B. T. Chait (1991). "Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry of Biopolymers." Analytical Chemistry **63**(24): 1193A-1203A.

Horvath, K., D. Lukacs, A. Sepsey and A. Felinger (2014). "Effect of particle size distribution on the separation efficiency in liquid chromatography." J Chromatogr A **1361**: 203-208.

Jones, A. M., R. Aebersold, C. H. Ahrens, R. Apweiler, K. Baerenfaller, M. Baker, E. Bendixen, S. Briggs, P. Brownridge, E. Brunner, M. Daube, E. W. Deutsch, U. Grossniklaus, J. Heazlewood, M. O. Hengartner, H. Hermjakob, M. Jovanovic, C. Lawless, G. Lochnit, L. Martens, C. Ravensborg, S. P. Schrimpf, Y. H. Shim, D. Subasic, A. Tholey, K. Wijk, C. Mering, M. Weiss and X. Zheng (2012). "The HUPO initiative on Model Organism Proteomes, iMOP." Proteomics **12**(3): 340-345.

Jonscher, K. R. and J. R. Yates, 3rd (1997). "The quadrupole ion trap mass spectrometer--a small solution to a big challenge." Anal Biochem **244**(1): 1-15.

Kuhlmann, F. E., A. Apffel, S. M. Fischer, G. Goldberg and P. C. Goodley (1995). "Signal enhancement for gradient reverse-phase high-performance liquid chromatography-electrospray ionization mass spectrometry analysis with trifluoroacetic and other strong acid modifiers by postcolumn addition of propionic acid and isopropanol." J Am Soc Mass Spectrom **6**(12): 1221-1225.

Lottspeich, F. and J. W. Engels (2012). Bioanalytik, Springer Spektrum.

- Maass, S., S. Sievers, D. Zuhlke, J. Kuzinski, P. K. Sappa, J. Muntel, B. Hessling, J. Bernhardt, R. Sietmann, U. Volker, M. Hecker and D. Becher (2011). "Efficient, global-scale quantification of absolute protein amounts by integration of targeted mass spectrometry and two-dimensional gel-based proteomics." Anal Chem **83**(7): 2677-2684.
- Makarov, A. (2000). "Electrostatic axially harmonic orbital trapping: a high-performance technique of mass analysis." Anal Chem **72**(6): 1156-1162.
- Makarov, A., E. Denisov and O. Lange (2009). "Performance evaluation of a high-field Orbitrap mass analyzer." J Am Soc Mass Spectrom **20**(8): 1391-1396.
- O'Farrell, P. H. (1975). "High resolution two-dimensional electrophoresis of proteins." J Biol Chem **250**(10): 4007-4021.
- Olsen, J. V., B. Macek, O. Lange, A. Makarov, S. Horning and M. Mann (2007). "Higher-energy C-trap dissociation for peptide modification analysis." Nat Methods **4**(9): 709-712.
- Ong, S. E., B. Blagoev, I. Kratchmarova, D. B. Kristensen, H. Steen, A. Pandey and M. Mann (2002). "Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics." Mol Cell Proteomics **1**(5): 376-386.
- Ong, S. E. and M. Mann (2005). "Mass spectrometry-based proteomics turns quantitative." Nat Chem Biol **1**(5): 252-262.
- Opiteck, G. J., K. C. Lewis, J. W. Jorgenson and R. J. Andereg (1997). "Comprehensive on-line LC/LC/MS of proteins." Anal Chem **69**(8): 1518-1524.
- Paul, W. and H. Steinwedel (1953). "'Ein Neues Massenspektrometer Ohne Magnetfeld." Zeitschrift Fur Naturforschung Section a-a Journal of Physical Sciences **8**(7): 448-450.
- Perkins, D. N., D. J. Pappin, D. M. Creasy and J. S. Cottrell (1999). "Probability-based protein identification by searching sequence databases using mass spectrometry data." Electrophoresis **20**(18): 3551-3567.
- Rabilloud, T., M. Chevallet, S. Luche and C. Lelong (2010). "Two-dimensional gel electrophoresis in proteomics: Past, present and future." J Proteomics **73**(11): 2064-2077.
- Rauniyar, N. and J. R. Yates, 3rd (2014). "Isobaric labeling-based relative quantification in shotgun proteomics." J Proteome Res **13**(12): 5293-5309.
- Roepstorff, P. and J. Fohlman (1984). "Proposal for a common nomenclature for sequence ions in mass spectra of peptides." Biomed Mass Spectrom **11**(11): 601.
- Schechter, I. and A. Berger (1967). "On the size of the active site in proteases. I. Papain." Biochem Biophys Res Commun **27**(2): 157-162.
- Schnolzer, M., P. Jedrzejewski and W. D. Lehmann (1996). "Protease-catalyzed incorporation of ¹⁸O into peptide fragments and its application for protein sequencing by electrospray and matrix-assisted laser desorption/ionization mass spectrometry." Electrophoresis **17**(5): 945-953.
- Schwanhausser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen and M. Selbach (2011). "Global quantification of mammalian gene expression control." Nature **473**(7347): 337-342.
- Sechi, S. (2016). Quantitative proteomics by mass spectrometry, Springer.
- Somasundaram, P., T. Koudelka, D. Linke and A. Tholey (2016). "C-Terminal Charge-Reversal Derivatization and Parallel Use of Multiple Proteases Facilitates Identification of Protein C-Termini by C-Terminomics." Journal of Proteome Research **15**(4): 1369-1378.

- Stjern Dahl, M. and K. Holmberg (2005). "Hydrolyzable nonionic surfactants: stability and physicochemical properties of surfactants containing carbonate, ester, and amide bonds." J Colloid Interface Sci **291**(2): 570-576.
- Tholey, A., C. Treitz, M. Kussmann, E. Bendixen, S. P. Schrimpf and M. O. Hengartner (2013). "Model organisms proteomics--from holobionts to human nutrition." Proteomics **13**(17): 2537-2541.
- Toll, H., H. Oberacher, R. Swart and C. G. Huber (2005). "Separation, detection, and identification of peptides by ion-pair reversed-phase high-performance liquid chromatography-electrospray ionization mass spectrometry at high and low pH." J Chromatogr A **1079**(1-2): 274-286.
- Wang, M., M. Weiss, M. Simonovic, G. Haertinger, S. P. Schrimpf, M. O. Hengartner and C. von Mering (2012). "PaxDb, a database of protein abundance averages across all three domains of life." Mol Cell Proteomics **11**(8): 492-500.
- Wolters, D. A., M. P. Washburn and J. R. Yates, 3rd (2001). "An automated multidimensional protein identification technology for shotgun proteomics." Anal Chem **73**(23): 5683-5690.
- Yamashita, M. and J. B. Fenn (1984). "Electrospray ion source. Another variation on the free-jet theme." The Journal of Physical Chemistry **88**(20): 4451-4459.
- Yao, X., A. Freas, J. Ramirez, P. A. Demirev and C. Fenselau (2001). "Proteolytic ¹⁸O labeling for comparative proteomics: model studies with two serotypes of adenovirus." Anal Chem **73**(13): 2836-2842.
- Yates, J. R., J. K. Eng, K. R. Clauser and A. L. Burlingame (1996). "Search of sequence databases with uninterpreted high-energy collision-induced dissociation spectra of peptides." J Am Soc Mass Spectrom **7**(11): 1089-1098.

Chapter 2. References

- Aravind, L., V. M. Dixit and E. V. Koonin (2001). "Apoptotic molecular machinery: vastly increased complexity in vertebrates revealed by genome comparisons." Science **291**(5507): 1279-1284.
- Arur, S., U. E. Uche, K. Rezaul, M. Fong, V. Scranton, A. E. Cowan, W. Mohler and D. K. Han (2003). "Annexin I is an endogenous ligand that mediates apoptotic cell engulfment." Dev Cell **4**(4): 587-598.
- Ashburner, M., C. A. Ball, J. A. Blake, D. Botstein, H. Butler, J. M. Cherry, A. P. Davis, K. Dolinski, S. S. Dwight, J. T. Eppig, M. A. Harris, D. P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J. C. Matese, J. E. Richardson, M. Ringwald, G. M. Rubin and G. Sherlock (2000). "Gene ontology: tool for the unification of biology. The Gene Ontology Consortium." Nat Genet **25**(1): 25-29.
- Baugh, L. R. (2009). "Staging worms for next-generation analysis." Nat Methods **6**(10): 727-728.
- Baugh, L. R., J. Demodena and P. W. Sternberg (2009). "RNA Pol II accumulates at promoters of growth genes during developmental arrest." Science **324**(5923): 92-94.
- Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) **57**(1): 289-300.
- Betz, F. S., B. G. Hammond and R. L. Fuchs (2000). "Safety and advantages of Bacillus thuringiensis-protected plants to control insect pests." Regul Toxicol Pharmacol **32**(2): 156-173.
- Boehnisch, C., D. Wong, M. Habig, K. Isermann, N. K. Michiels, T. Roeder, R. C. May and H. Schulenburg (2011). "Protist-type lysozymes of the nematode *Caenorhabditis elegans* contribute to resistance against pathogenic *Bacillus thuringiensis*." PLoS One **6**(9): e24619.
- Bogaerts, A., I. Beets, L. Temmerman, L. Schoofs and P. Verleyen (2010). "Proteome changes of *Caenorhabditis elegans* upon a *Staphylococcus aureus* infection." Biol Direct **5**: 11.
- Bogaerts, A., L. Temmerman, B. Boerjan, S. J. Husson, L. Schoofs and P. Verleyen (2010). "A differential proteomics study of *Caenorhabditis elegans* infected with *Aeromonas hydrophila*." Dev Comp Immunol **34**(6): 690-698.
- Borgonie, G., R. Van Driessche, F. Leyns, G. Arnaut, D. De Waele and A. Coomans (1995). "Germination of *Bacillus thuringiensis* spores in bacteriophagous nematodes (Nematoda: Rhabditida)." J Invertebr Pathol **65**(1): 61-67.
- Bossinger, O., T. Fukushige, M. Claeys, G. Borgonie and J. D. McGhee (2004). "The apical disposition of the *Caenorhabditis elegans* intestinal terminal web is maintained by LET-413." Dev Biol **268**(2): 448-456.
- Brenner, S. (1974). "THE GENETICS OF CAENORHABDITIS ELEGANS." Genetics **77**(1): 71-94.
- Cassidy, L. and A. Tholey (2014). "Model organism proteomics as a tool for the study of host-microbiome interactions." Proteomics Clinical Applications **8**(9-10): 665-676.
- Couillault, C. and J. J. Ewbank (2002). "Diverse bacteria are pathogens of *Caenorhabditis elegans*." Infect Immun **70**(8): 4705-4707.
- Couillault, C., N. Pujol, J. Reboul, L. Sabatier, J. F. Guichou, Y. Kohara and J. J. Ewbank (2004). "TLR-independent control of innate immunity in *Caenorhabditis elegans* by the TIR domain adaptor protein TIR-1, an ortholog of human SARM." Nat Immunol **5**(5): 488-494.

- Cox, J. and M. Mann (2012). "1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data." BMC Bioinformatics **13 Suppl 16**: S12.
- Curt, A., J. Zhang, J. Minnerly and K. Jia (2014). "Intestinal autophagy activity is essential for host defense against *Salmonella typhimurium* infection in *Caenorhabditis elegans*." Developmental & Comparative Immunology **45**(2): 214-218.
- Delmotte, N., M. Lasaosa, A. Tholey, E. Heinzle and C. G. Huber (2007). "Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis." J Proteome Res **6**(11): 4363-4373.
- Edwards, D. L., J. Payne and G. G. Soares (1989). "Novel isolates of *Bacillus thuringiensis* having activity against nematodes." European Patent **EP 0303426 A2**.
- Elmore, S. (2007). "Apoptosis: A review of programmed cell death." Toxicologic Pathology **35**(4): 495-516.
- Engelmann, I. and N. Pujol (2010). "Innate Immunity in *C. Elegans*." Invertebrate Immunology **708**: 105-121.
- Geier, G., H. J. Banaj, H. Heid, L. Bini, V. Pallini and R. Zwillig (1999). "Aspartyl proteases in *Caenorhabditis elegans*. Isolation, identification and characterization by a combined use of affinity chromatography, two-dimensional gel electrophoresis, microsequencing and databank analysis." Eur J Biochem **264**(3): 872-879.
- Ghazarian, H., B. Itoni and S. B. Oppenheimer (2011). "A glycochemistry review: carbohydrates, lectins and implications in cancer therapeutics." Acta Histochem **113**(3): 236-247.
- Griffitts, J. S., J. L. Whitacre, D. E. Stevens and R. V. Aroian (2001). "Bt toxin resistance from loss of a putative carbohydrate-modifying enzyme." Science **293**(5531): 860-864.
- Hansen, M., A. Chandra, L. L. Mitic, B. Onken, M. Driscoll and C. Kenyon (2008). "A role for autophagy in the extension of lifespan by dietary restriction in *C. elegans*." PLoS Genet **4**(2): e24.
- Harris, T. W., I. Antoshechkin, T. Bieri, D. Blasiar, J. Chan, W. J. Chen, N. De La Cruz, P. Davis, M. Duesbury, R. Fang, J. Fernandes, M. Han, R. Kishore, R. Lee, H. M. Muller, C. Nakamura, P. Ozersky, A. Petcherski, A. Rangarajan, A. Rogers, G. Schindelman, E. M. Schwarz, M. A. Tuli, K. Van Auken, D. Wang, X. Wang, G. Williams, K. Yook, R. Durbin, L. D. Stein, J. Spieth and P. W. Sternberg (2010). "WormBase: a comprehensive resource for nematode research." Nucleic Acids Res **38**(Database issue): D463-467.
- Hoeckendorf, A. and M. Leippe (2012). "SPP-3, a saposin-like protein of *Caenorhabditis elegans*, displays antimicrobial and pore-forming activity and is located in the intestine and in one head neuron." Dev Comp Immunol **38**(1): 181-186.
- Hoeckendorf, A., M. Stanisak and M. Leippe (2012). "The saposin-like protein SPP-12 is an antimicrobial polypeptide in the pharyngeal neurons of *Caenorhabditis elegans* and participates in defence against a natural bacterial pathogen." Biochem J **445**(2): 205-212.
- Huffman, D. L., L. Abrami, R. Sasik, J. Corbeil, F. G. van der Goot and R. V. Aroian (2004). "Mitogen-activated protein kinase pathways defend against bacterial pore-forming toxins." Proc Natl Acad Sci U S A **101**(30): 10995-11000.
- Hunter, S., P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A. F. Quinn, C. Rivoire, A. Sangrador-Vegas, J. D. Selengut, C. J. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajananathan, P. D. Thomas, C. H. Wu, C. Yeats and S. Y. Yong (2012).

- "InterPro in 2011: new developments in the family and domain prediction database." Nucleic Acids Res **40**(Database issue): D306-312.
- Ideo, H., K. Fukushima, K. Gengyo-Ando, S. Mitani, K. Dejima, K. Nomura and K. Yamashita (2009). "A *Caenorhabditis elegans* glycolipid-binding galectin functions in host defense against bacterial infection." J Biol Chem **284**(39): 26493-26501.
- Jia, K. L., C. Thomas, M. Akbar, Q. H. Sun, B. Adams-Huet, C. Gilpin and B. Levine (2009). "Autophagy genes protect against *Salmonella typhimurium* infection and mediate insulin signaling-regulated pathogen resistance." Proceedings of the National Academy of Sciences of the United States of America **106**(34): 14564-14569.
- Kall, L., J. D. Storey and W. S. Noble (2008). "Non-parametric estimation of posterior error probabilities associated with peptides identified by tandem mass spectrometry." Bioinformatics **24**(16): i42-48.
- Kang, C. and L. Avery (2010). "Death-associated protein kinase (DAPK) and signal transduction: fine-tuning of autophagy in *Caenorhabditis elegans* homeostasis." FEBS J **277**(1): 66-73.
- Kang, Y., D. Zhao, H. Liang, B. Liu, Y. Zhang, Q. Liu, X. Wang and Y. Liu (2012). "Structural study of TTR-52 reveals the mechanism by which a bridging molecule mediates apoptotic cell engulfment." Genes Dev **26**(12): 1339-1350.
- Khan, F. R. and B. A. McFadden (1980). "A Rapid Method of Synchronizing Developmental Stages of *Caenorhabditis-Elegans*." Nematologica **26**(2): 280-282.
- Kho, M. F., A. Bellier, V. Balasubramani, Y. Hu, W. Hsu, C. Nielsen-LeRoux, S. M. McGillivray, V. Nizet and R. V. Aroian (2011). "The pore-forming protein Cry5B elicits the pathogenicity of *Bacillus* sp. against *Caenorhabditis elegans*." PLoS One **6**(12): e29122.
- Koga, M., R. Zwaal, K. L. Guan, L. Avery and Y. Ohshima (2000). "A *Caenorhabditis elegans* MAP kinase kinase, MEK-1, is involved in stress responses." EMBO J **19**(19): 5148-5156.
- Leippe, M. (1999). "Antimicrobial and cytolytic polypeptides of amoeboid protozoa--effector molecules of primitive phagocytes." Dev Comp Immunol **23**(4-5): 267-279.
- Lewis, J. A. and J. T. Fleming (1995). "Basic culture methods." Methods Cell Biol **48**: 3-29.
- Lucas, N. and W. Cho (2011). "Phosphatidylserine binding is essential for plasma membrane recruitment and signaling function of 3-phosphoinositide-dependent kinase-1." J Biol Chem **286**(48): 41265-41272.
- Luke, C. J., S. C. Pak, Y. S. Askew, T. L. Naviglia, D. J. Askew, S. M. Nobar, A. C. Vetica, O. S. Long, S. C. Watkins, D. B. Stolz, R. J. Barstead, G. L. Moulder, D. Bromme and G. A. Silverman (2007). "An intracellular serpin regulates necrosis by inhibiting the induction and sequelae of lysosomal injury." Cell **130**(6): 1108-1119.
- Ma, G., M. M. Rahman, W. Grant, O. Schmidt and S. Asgari (2012). "Insect tolerance to the crystal toxins Cry1Ac and Cry2Ab is mediated by the binding of monomeric toxin to lipophorin glycolipids causing oligomerization and sequestration reactions." Dev Comp Immunol **37**(1): 184-192.
- Ma, G., O. Schmidt and M. Keller (2012). "Pre-feeding of a glycolipid binding protein LEC-8 from *Caenorhabditis elegans* revealed enhanced tolerance to Cry1Ac toxin in *Helicoverpa armigera*." Results Immunol **2**: 97-103.
- Maduzia, L. L., E. Yu and Y. Zhang (2011). "*Caenorhabditis elegans* galectins LEC-6 and LEC-10 interact with similar glycoconjugates in the intestine." J Biol Chem **286**(6): 4371-4381.

- Mallo, G. V., C. L. Kurz, C. Couillault, N. Pujol, S. Granjeaud, Y. Kohara and J. J. Ewbank (2002). "Inducible antibacterial defense system in *C. elegans*." Curr Biol **12**(14): 1209-1214.
- Marsh, E. K. and R. C. May (2012). "Caenorhabditis elegans, a model organism for investigating immunity." Appl Environ Microbiol **78**(7): 2075-2081.
- McGhee, J. D., M. C. Sleumer, M. Bilenky, K. Wong, S. J. McKay, B. Goszczynski, H. Tian, N. D. Krich, J. Khattri, R. A. Holt, D. L. Baillie, Y. Kohara, M. A. Marra, S. J. Jones, D. G. Moerman and A. G. Robertson (2007). "The ELT-2 GATA-factor and the global regulation of transcription in the *C. elegans* intestine." Dev Biol **302**(2): 627-645.
- McVeigh, P., S. Alexander-Bowman, E. Veal, A. Mousley, N. J. Marks and A. G. Maule (2008). "Neuropeptide-like protein diversity in phylum Nematoda." Int J Parasitol **38**(13): 1493-1503.
- Miltsch, S. M., P. H. Seeberger and B. Lepenies (2014). "The C-type lectin-like domain containing proteins Clec-39 and Clec-49 are crucial for Caenorhabditis elegans immunity against Serratia marcescens infection." Dev Comp Immunol **45**(1): 67-73.
- Mylonakis, E., F. M. Ausubel, J. R. Perfect, J. Heitman and S. B. Calderwood (2002). "Killing of Caenorhabditis elegans by Cryptococcus neoformans as a model of yeast pathogenesis." Proc Natl Acad Sci U S A **99**(24): 15675-15680.
- Nandakumar, M. and M. W. Tan (2008). "Gamma-linolenic and stearidonic acids are required for basal immunity in Caenorhabditis elegans through their effects on p38 MAP kinase activity." PLoS Genet **4**(11): e1000273.
- Nisnevitch, M., S. Sigawi, R. Cahan and Y. Nitzan (2010). "Isolation, characterization and biological role of camelysin from Bacillus thuringiensis subsp. israelensis." Curr Microbiol **61**(3): 176-183.
- O'Rourke, D., D. Baban, M. Demidova, R. Mott and J. Hodgkin (2006). "Genomic clusters, putative pathogen recognition molecules, and antimicrobial genes are induced by infection of *C. elegans* with *M. nematophilum*." Genome Res **16**(8): 1005-1016.
- Powell, J. R., D. H. Kim and F. M. Ausubel (2009). "The G protein-coupled receptor FSHR-1 is required for the Caenorhabditis elegans innate immune response." Proc Natl Acad Sci U S A **106**(8): 2782-2787.
- Pujol, N., P. A. Davis and J. J. Ewbank (2012). "The Origin and Function of Anti-Fungal Peptides in *C. elegans*: Open Questions." Front Immunol **3**: 237.
- Pukkila-Worley, R., R. Feinbaum, N. V. Kirienko, J. Larkins-Ford, A. L. Conery and F. M. Ausubel (2012). "Stimulation of host immune defenses by a small molecule protects *C. elegans* from bacterial infection." PLoS Genet **8**(6): e1002733.
- Pulak, R. (2006). "Techniques for analysis, sorting, and dispensing of *C. elegans* on the COPAS flow-sorting system." Methods Mol Biol **351**: 275-286.
- Rehman Khan, F. and B. A. McFadden (1980). "A rapid method of synchronizing developmental stages of Caenorhabditis elegans." Nematologica **26**(2): 280-282.
- Roeder, T., M. Stanisak, C. Gelhaus, I. Bruchhaus, J. Grotzinger and M. Leippe (2010). "Caenopores are antimicrobial peptides in the nematode Caenorhabditis elegans instrumental in nutrition and immunity." Dev Comp Immunol **34**(2): 203-209.
- Rohde, C. B., F. Zeng, R. Gonzalez-Rubio, M. Angel and M. F. Yanik (2007). "Microfluidic system for on-chip high-throughput whole-animal sorting and screening at subcellular resolution." Proc Natl Acad Sci U S A **104**(35): 13891-13895.

- Ross, P. L., Y. N. Huang, J. N. Marchese, B. Williamson, K. Parker, S. Hattan, N. Khainovski, S. Pillai, S. Dey, S. Daniels, S. Purkayastha, P. Juhasz, S. Martin, M. Bartlet-Jones, F. He, A. Jacobson and D. J. Pappin (2004). "Multiplexed protein quantitation in *Saccharomyces cerevisiae* using amine-reactive isobaric tagging reagents." *Mol Cell Proteomics* **3**(12): 1154-1169.
- Schauer, I. E. and W. B. Wood (1990). "Early *C. elegans* embryos are transcriptionally active." *Development* **110**(4): 1303-1317.
- Schrimpf, S. P. and M. O. Hengartner (2010). "A worm rich in protein: Quantitative, differential, and global proteomics in *Caenorhabditis elegans*." *J Proteomics* **73**(11): 2186-2197.
- Schulenburg, H. and C. Boehnisch (2008). "Diversification and adaptive sequence evolution of *Caenorhabditis* lysozymes (Nematoda: Rhabditidae)." *BMC Evol Biol* **8**: 114.
- Schulenburg, H., M. P. Hoepfner, J. Weiner, 3rd and E. Bornberg-Bauer (2008). "Specificity of the innate immune system and diversity of C-type lectin domain (CTL) proteins in the nematode *Caenorhabditis elegans*." *Immunobiology* **213**(3-4): 237-250.
- Schulte, R. D., C. Makus, B. Hasert, N. K. Michiels and H. Schulenburg (2010). "Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite." *Proc Natl Acad Sci U S A* **107**(16): 7359-7364.
- Schulte, R. D., C. Makus and H. Schulenburg (2013). "Host-parasite coevolution favours parasite genetic diversity and horizontal gene transfer." *J Evol Biol* **26**(8): 1836-1840.
- Shapira, M., B. J. Hamlin, J. Rong, K. Chen, M. Ronen and M. W. Tan (2006). "A conserved role for a GATA transcription factor in regulating epithelial innate immune responses." *Proc Natl Acad Sci U S A* **103**(38): 14086-14091.
- Shatters, R. G., Jr., L. M. Boykin, S. L. Lapointe, W. B. Hunter and A. A. Weathersbee, 3rd (2006). "Phylogenetic and structural relationships of the PR5 gene family reveal an ancient multigene family conserved in plants and select animal taxa." *J Mol Evol* **63**(1): 12-29.
- Simonsen, K. T., S. F. Gallego, N. J. Faergeman and B. H. Kallipolitis (2012). "Strength in numbers: "Omics" studies of *C. elegans* innate immunity." *Virulence* **3**(6): 477-484.
- Simonsen, K. T., J. Moller-Jensen, A. R. Kristensen, J. S. Andersen, D. L. Riddle and B. H. Kallipolitis (2011). "Quantitative proteomics identifies ferritin in the innate immune response of *C. elegans*." *Virulence* **2**(2): 120-130.
- Stephanowitz, H., S. Lange, D. Lang, C. Freund and E. Krause (2012). "Improved two-dimensional reversed phase-reversed phase LC-MS/MS approach for identification of peptide-protein interactions." *J Proteome Res* **11**(2): 1175-1183.
- Stergiou, L., R. Eberhard, K. Doukoumetzidis and M. O. Hengartner (2011). "NER and HR pathways act sequentially to promote UV-C-induced germ cell apoptosis in *Caenorhabditis elegans*." *Cell Death Differ* **18**(5): 897-906.
- Studencka, M., A. Konzer, G. Moneron, D. Wenzel, L. Opitz, G. Salinas-Riester, C. Bedet, M. Kruger, S. W. Hell, J. R. Wisniewski, H. Schmidt, F. Palladino, E. Schulze and M. Jedrusik-Bode (2012). "Novel roles of *Caenorhabditis elegans* heterochromatin protein HP1 and linker histone in the regulation of innate immune gene expression." *Mol Cell Biol* **32**(2): 251-265.
- van Frankenhuyzen, K. (2009). "Insecticidal activity of *Bacillus thuringiensis* crystal proteins." *J Invertebr Pathol* **101**(1): 1-16.

Wang, X., W. Li, D. Zhao, B. Liu, Y. Shi, B. Chen, H. Yang, P. Guo, X. Geng, Z. Shang, E. Peden, E. Kage-Nakadai, S. Mitani and D. Xue (2010). "Caenorhabditis elegans transthyretin-like protein TTR-52 mediates recognition of apoptotic cells by the CED-1 phagocyte receptor." Nat Cell Biol **12**(7): 655-664.

Wang, X., J. Wang, K. Gengyo-Ando, L. Gu, C. L. Sun, C. Yang, Y. Shi, T. Kobayashi, Y. Shi, S. Mitani, X. S. Xie and D. Xue (2007). "C. elegans mitochondrial factor WAH-1 promotes phosphatidylserine externalization in apoptotic cells through phospholipid scramblase SCRM-1." Nat Cell Biol **9**(5): 541-549.

Wei, J. Z., K. Hale, L. Carta, E. Platzer, C. Wong, S. C. Fang and R. V. Aroian (2003). "Bacillus thuringiensis crystal proteins that target nematodes." Proc Natl Acad Sci U S A **100**(5): 2760-2765.

White, C., S. N. Gagnon, J. F. St-Laurent, C. Gravel, L. I. Proulx and S. Desnoyers (2009). "The DNA damage-inducible C. elegans tankyrase is a nuclear protein closely linked to chromosomes." Mol Cell Biochem **324**(1-2): 73-83.

Wong, D., D. Bazopoulou, N. Pujol, N. Tavernarakis and J. J. Ewbank (2007). "Genome-wide investigation reveals pathogen-specific and shared signatures in the response of Caenorhabditis elegans to infection." Genome Biol **8**(9): R194.

Zhong, W. and P. W. Sternberg (2006). "Genome-wide prediction of C. elegans genetic interactions." Science **311**(5766): 1481-1484.

Zugasti, O. and J. J. Ewbank (2009). "Neuroimmune regulation of antimicrobial peptide expression by a noncanonical TGF-beta signaling pathway in Caenorhabditis elegans epidermis." Nat Immunol **10**(3): 249-256.

Java-Script A 2.1: Image processing with ImageJ

The batch processing commands were adjusted from instructions found here:
https://imagej.net/How_to_apply_a_common_operation_to_a_complete_directory

```
// define folders for Images and Output
Images = " J:\\Imageprocessing\\Images\\";
Processed = " J:\\Imageprocessing\\Processed\\" ";
Data = " J:\\Imageprocessing\\Data\\" ";

setBatchMode(true);
List = getFilesList(Images);
for (i = 0; i < list.length; i++) {
    if (endsWith(list[i], ","))
        ImageProcessing (Images, Processed, list[i]);
setBatchMode(false);

Function ImageProcessing(Images, Processed, filename) {

// spatial calibration for 5 x microscopic images
open("Images + filename");
run("Scale...");
makeLine(1198, 97, 1318, 98);
run("Scale...");
makeLine(1130, 158, 1221, 160);
run("Scale...");
run("Set Scale...", "distance=93 known=100 unit=µm");

// find regions of interest by outline
run("Find Edges");

// remove background; threshold may have to be adjusted depending on exposure
run("Subtract Background...");
setAutoThreshold("Default");
//run("Threshold...");
setAutoThreshold("Default dark");
setAutoThreshold("Default");
//setThreshold(0, 71);
run("Convert to Mask");
run("Close");
run("Invert");

// measure region of interest (using ROI Curve script by Bob Dougherty)
run("Analyze Particles...", "size=2890.51-2312406.06 circularity=0.10-0.40
show=Outlines display exclude include record");
close();

// save data to Excel table
saveAs("Results", "J:\\Imageprocessing\\Data\\Results.xls");
close();
}
```

VBA-Script A 2.2

Individual Excel files were combined using a visual basic application script adjusted from “Merge multiple excel files into a single spreadsheet – VBA” instructions by Raghu R: <http://quadexcel.com/>

```
Sub simpleXlsMerger()  
Dim bookList As Workbook  
Dim mergeObj As Object, dirObj As Object, filesObj As Object, everyObj As Object  
Application.ScreenUpdating = False  
Set mergeObj = CreateObject("Scripting.FileSystemObject")  
  
'change folder path of excel files here  
Set dirObj = mergeObj.Getfolder("J:\\Imageprocessing\Data\\")  
Set filesObj = dirObj.Files  
For Each everyObj In filesObj  
Set bookList = Workbooks.Open(everyObj)  
  
'change "A2" with cell reference of start point for every files here  
'for example "B3:IV" to merge all files start from columns B and rows 3  
'If you're files using more than IV column, change it to the latest column  
'Also change "A" column on "A65536" to the same column as start point  
Range("A2:IV" & Range("A65536").End(xlUp).Row).Copy  
ThisWorkbook.Worksheets(1).Activate  
  
'Do not change the following column. It's not the same column as above  
Range("A65536").End(xlUp).Offset(1, 0).PasteSpecial  
Application.CutCopyMode = False  
bookList.Close  
Next  
End Sub
```

Chapter 3. References

- Altelaar, A. F., C. K. Frese, C. Preisinger, M. L. Hennrich, A. W. Schram, H. T. Timmers, A. J. Heck and S. Mohammed (2013). "Benchmarking stable isotope labeling based quantitative proteomics." J Proteomics **88**: 14-26.
- Angov, E., C. J. Hillier, R. L. Kincaid and J. A. Lyon (2008). "Heterologous protein expression is enhanced by harmonizing the codon usage frequencies of the target gene with those of the expression host." PLoS One **3**(5): e2189.
- Aristidou, A. A., K. Y. San and G. N. Bennett (1995). "Metabolic engineering of Escherichia coli to enhance recombinant protein production through acetate reduction." Biotechnol Prog **11**(4): 475-478.
- Bantscheff, M., M. Boesche, D. Eberhard, T. Matthieson, G. Sweetman and B. Kuster (2008). "Robust and sensitive iTRAQ quantification on an LTQ Orbitrap mass spectrometer." Mol Cell Proteomics **7**(9): 1702-1713.
- Barak, R., W. N. Abouhamad and M. Eisenbach (1998). "Both acetate kinase and acetyl coenzyme A synthetase are involved in acetate-stimulated change in the direction of flagellar rotation in Escherichia coli." J Bacteriol **180**(4): 985-988.
- Benjamini, Y. and Y. Hochberg (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." Journal of the Royal Statistical Society. Series B (Methodological) **57**(1): 289-300.
- Berg, O. G. and C. G. Kurland (1997). "Growth rate-optimised tRNA abundance and codon usage." J Mol Biol **270**(4): 544-550.
- Borisov, V. B., R. B. Gennis, J. Hemp and M. I. Verkhovsky (2011). "The cytochrome bd respiratory oxygen reductases." Biochim Biophys Acta **1807**(11): 1398-1413.
- Brock, M., C. Maerker, A. Schutz, U. Volker and W. Buckel (2002). "Oxidation of propionate to pyruvate in Escherichia coli. Involvement of methylcitrate dehydratase and aconitase." Eur J Biochem **269**(24): 6184-6194.
- Chao, Y. P., R. Patnaik, W. D. Roof, R. F. Young and J. C. Liao (1993). "Control of gluconeogenic growth by pps and pck in Escherichia coli." Journal of Bacteriology **175**(21): 6939-6944.
- Chen, H.-s., T. Rejtar, V. Andreev, E. Moskovets and B. L. Karger (2005). "Enhanced Characterization of Complex Proteomic Samples Using LC-MALDI MS/MS: Exclusion of Redundant Peptides from MS/MS Analysis in Replicate Runs." Analytical Chemistry **77**(23): 7816-7825.
- Chen, X., L. Zhou, K. Tian, A. Kumar, S. Singh, B. A. Prior and Z. Wang (2013). "Metabolic engineering of Escherichia coli: a sustainable industrial platform for bio-based chemical production." Biotechnol Adv **31**(8): 1200-1223.
- Chung, C. H., H. E. Ives, S. Almeda and A. L. Goldberg (1983). "Purification from Escherichia coli of a periplasmic protein that is a potent inhibitor of pancreatic proteases." J Biol Chem **258**(18): 11032-11038.
- Clark, D. P. and J. E. Cronan (2005). "Two-Carbon Compounds and Fatty Acids as Carbon Sources." EcoSal Plus **1**(2).
- Cottrill, M. A., S. P. Golovan, J. P. Phillips and C. W. Forsberg (2002). "Inositol phosphatase activity of the Escherichia coli agp-encoded acid glucose-1-phosphatase." Can J Microbiol **48**(9): 801-809.

- Cox, J. and M. Mann (2012). "1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data." BMC Bioinformatics **13 Suppl 16**: S12.
- Cozzone, A. J. (1998). "Regulation of acetate metabolism by protein phosphorylation in enteric bacteria." Annual Review of Microbiology **52**: 127-164.
- Dailey, F. E. and J. E. Cronan (1986). "Acetohydroxy Acid Synthase-I, a Required Enzyme for Isoleucine and Valine Biosynthesis in Escherichia-Coli K-12 during Growth on Acetate as the Sole Carbon Source." Journal of Bacteriology **165**(2): 453-460.
- Delmotte, N., M. Lasaosa, A. Tholey, E. Heinzle and C. G. Huber (2007). "Two-dimensional reversed-phase x ion-pair reversed-phase HPLC: an alternative approach to high-resolution peptide separation for shotgun proteome analysis." J Proteome Res **6**(11): 4363-4373.
- Dong, H., L. Nilsson and C. G. Kurland (1996). "Co-variation of tRNA abundance and codon usage in Escherichia coli at different growth rates." J Mol Biol **260**(5): 649-663.
- Eiteman, M. A. and E. Altman (2006). "Overcoming acetate in Escherichia coli recombinant protein fermentations." Trends Biotechnol **24**(11): 530-536.
- Enjalbert, B., F. Letisse and J. C. Portais (2013). "Physiological and Molecular Timing of the Glucose to Acetate Transition in Escherichia coli." Metabolites **3**(3): 820-837.
- Fraenkel, D. G. and B. L. Horecker (1965). "Fructose-1, 6-diphosphatase and acid hexose phosphatase of Escherichia coli." J Bacteriol **90**(4): 837-842.
- Giuffre, A., V. B. Borisov, M. Arese, P. Sarti and E. Forte (2014). "Cytochrome bd oxidase and bacterial tolerance to oxidative and nitrosative stress." Biochim Biophys Acta **1837**(7): 1178-1187.
- Griffin, T. J., S. P. Gygi, T. Ideker, B. Rist, J. Eng, L. Hood and R. Aebersold (2002). "Complementary profiling of gene expression at the transcriptome and proteome levels in Saccharomyces cerevisiae." Mol Cell Proteomics **1**(4): 323-333.
- Guest, J. R. (1981). "Partial replacement of succinate dehydrogenase function by phage- and plasmid-specified fumarate reductase in Escherichia coli." J Gen Microbiol **122**(2): 171-179.
- Han, M. J. and S. Y. Lee (2006). "The Escherichia coli proteome: past, present, and future prospects." Microbiol Mol Biol Rev **70**(2): 362-439.
- Henry, I. and P. M. Sharp (2007). "Predicting gene expression level from codon usage bias." Mol Biol Evol **24**(1): 10-12.
- Hillier, S. and W. T. Charnetzky (1981). "Glyoxylate bypass enzymes in Yersinia species and multiple forms of isocitrate lyase in Yersinia pestis." J Bacteriol **145**(1): 452-458.
- Hu, J. C., G. Sherlock, D. A. Siegele, S. A. Aleksander, C. A. Ball, J. Demeter, S. Gouni, T. A. Holland, P. D. Karp, J. E. Lewis, N. M. Liles, B. K. McIntosh, H. Mi, A. Muruganujan, F. Wymore and P. D. Thomas (2014). "PortEco: a resource for exploring bacterial biology through high-throughput data and analysis tools." Nucleic Acids Research **42**(Database issue): D677-D684.
- Huang, C. J., H. Lin and X. Yang (2012). "Industrial production of recombinant therapeutics in Escherichia coli and its recent advancements." J Ind Microbiol Biotechnol **39**(3): 383-399.
- Ishihama, Y., Y. Oda, T. Tabata, T. Sato, T. Nagasu, J. Rappsilber and M. Mann (2005). "Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein." Mol Cell Proteomics **4**(9): 1265-1272.

- Izutsu, K., C. Wada, Y. Komine, T. Sako, C. Ueguchi, S. Nakura and A. Wada (2001). "Escherichia coli ribosome-associated protein SRA, whose copy number increases during stationary phase." J Bacteriol **183**(9): 2765-2773.
- Jones, H. M. and R. P. Gunsalus (1985). "Transcription of the Escherichia coli fumarate reductase genes (frdABCD) and their coordinate regulation by oxygen, nitrate, and fumarate." J Bacteriol **164**(3): 1100-1109.
- Kall, L., J. D. Canterbury, J. Weston, W. S. Noble and M. J. MacCoss (2007). "Semi-supervised learning for peptide identification from shotgun proteomics datasets." Nat Methods **4**(11): 923-925.
- Kanehisa, M., Y. Sato, M. Kawashima, M. Furumichi and M. Tanabe (2016). "KEGG as a reference resource for gene and protein annotation." Nucleic Acids Research **44**(Database issue): D457-D462.
- Karp, N. A., W. Huber, P. G. Sadowski, P. D. Charles, S. V. Hester and K. S. Lilley (2010). "Addressing accuracy and precision issues in iTRAQ quantitation." Mol Cell Proteomics **9**(9): 1885-1897.
- Keseler, I. M., A. Mackie, M. Peralta-Gil, A. Santos-Zavaleta, S. Gama-Castro, C. Bonavides-Martinez, C. Fulcher, A. M. Huerta, A. Kothari, M. Krummenacker, M. Latendresse, L. Muniz-Rascado, Q. Ong, S. Paley, I. Schroder, A. G. Shearer, P. Subhraveti, M. Travers, D. Weerasinghe, V. Weiss, J. Collado-Vides, R. P. Gunsalus, I. Paulsen and P. D. Karp (2013). "EcoCyc: fusing model organism databases with systems biology." Nucleic Acids Res **41**(Database issue): D605-612.
- Kim, J. N., S. J. Ahn and R. A. Burne (2015). "Genetics and Physiology of Acetate Metabolism by the Pta-Ack Pathway of Streptococcus mutans." Appl Environ Microbiol **81**(15): 5015-5025.
- Kirkpatrick, C., L. M. Maurer, N. E. Oyelakin, Y. N. Yoncheva, R. Maurer and J. L. Slonczewski (2001). "Acetate and formate stress: Opposite responses in the proteome of Escherichia coli." Journal of Bacteriology **183**(21): 6466-6477.
- Kumari, S., C. M. Beatty, D. F. Browning, S. J. Busby, E. J. Simel, G. Hovel-Miner and A. J. Wolfe (2000). "Regulation of acetyl coenzyme A synthetase in Escherichia coli." J Bacteriol **182**(15): 4173-4179.
- Kumari, S., C. M. Beatty, D. F. Browning, S. J. W. Busby, E. J. Simel, G. Hovel-Miner and A. J. Wolfe (2000). "Regulation of acetyl coenzyme A synthetase in Escherichia coli." Journal of Bacteriology **182**(15): 4173-4179.
- Kumari, S., E. J. Simel and A. J. Wolfe (2000). "sigma(70) is the principal sigma factor responsible for transcription of acs, which encodes acetyl coenzyme A synthetase in Escherichia coli." J Bacteriol **182**(2): 551-554.
- Link, H., T. Fuhrer, L. Gerosa, N. Zamboni and U. Sauer (2015). "Real-time metabolome profiling of the metabolic switch between starvation and growth." Nat Methods **12**(11): 1091-1097.
- Liu, M., T. Durfee, J. E. Cabrera, K. Zhao, D. J. Jin and F. R. Blattner (2005). "Global Transcriptional Programs Reveal a Carbon Source Foraging Strategy by Escherichia coli." Journal of Biological Chemistry **280**(16): 15921-15927.
- Maier, T., M. Guell and L. Serrano (2009). "Correlation of mRNA and protein in complex biological samples." FEBS Lett **583**(24): 3966-3973.
- Maki, Y., H. Yoshida and A. Wada (2000). "Two proteins, YfiA and YhbH, associated with resting ribosomes in stationary phase Escherichia coli." Genes Cells **5**(12): 965-974.
- Marcus, M. and Y. S. Halpern (1969). "The metabolic pathway of glutamate in Escherichia coli K-12." Biochim Biophys Acta **177**(2): 314-320.

- McHardy, A. C., A. Puhler, J. Kalinowski and F. Meyer (2004). "Comparing expression level-dependent features in codon usage with protein abundance: an analysis of 'predictive proteomics'." Proteomics **4**(1): 46-58.
- Measures, J. C. (1975). "Role of amino acids in osmoregulation of non-halophilic bacteria." Nature **257**(5525): 398-400.
- Médigue, C., A. Viari, A. Hénaut and A. Danchin (1993). "Colibri: a functional data base for the Escherichia coli genome." Microbiological Reviews **57**(3): 623-654.
- Milne, A. N., W. W. Mak and J. T. Wong (1975). "Variation of ribosomal proteins with bacterial growth rate." J Bacteriol **122**(1): 89-92.
- Molina, I., M. T. Pellicer, J. Badia, J. Aguilar and L. Baldoma (1994). "Molecular characterization of Escherichia coli malate synthase G. Differentiation with the malate synthase A isoenzyme." Eur J Biochem **224**(2): 541-548.
- Monod, J. (1949). "The Growth of Bacterial Cultures." Annual Review of Microbiology **3**: 371-394.
- Nachin, L., U. Nannmark and T. Nystrom (2005). "Differential roles of the universal stress proteins of Escherichia coli in oxidative stress resistance, adhesion, and motility." J Bacteriol **187**(18): 6265-6272.
- Nakamura, Y., T. Gojobori and T. Ikemura (2000). "Codon usage tabulated from international DNA sequence databases: status for the year 2000." Nucleic Acids Res **28**(1): 292.
- Natale, P., T. Bruser and A. J. Driessen (2008). "Sec- and Tat-mediated protein secretion across the bacterial cytoplasmic membrane--distinct translocases and mechanisms." Biochim Biophys Acta **1778**(9): 1735-1756.
- Nunn, W. D. (1986). "A molecular view of fatty acid catabolism in Escherichia coli." Microbiol Rev **50**(2): 179-192.
- O'Brien, W. J. and F. E. Frerman (1977). "Evidence for a complex of three beta-oxidation enzymes in Escherichia coli: induction and localization." Journal of Bacteriology **132**(2): 532-540.
- Oh, M. K., L. Rohlin, K. C. Kao and J. C. Liao (2002). "Global expression profiling of acetate-grown Escherichia coli." J Biol Chem **277**(15): 13175-13183.
- Ow, S. Y., M. Salim, J. Noirel, C. Evans, I. Rehman and P. C. Wright (2009). "iTRAQ underestimation in simple and complex mixtures: 'the good, the bad and the ugly'." J Proteome Res **8**(11): 5347-5355.
- Ow, S. Y., M. Salim, J. Noirel, C. Evans and P. C. Wright (2011). "Minimising iTRAQ ratio compression through understanding LC-MS elution dependence and high-resolution HILIC fractionation." Proteomics **11**(11): 2341-2346.
- Perez-Arellano, I., V. Rubio and J. Cervera (2005). "Dissection of Escherichia coli glutamate 5-kinase: functional impact of the deletion of the PUA domain." FEBS Lett **579**(30): 6903-6908.
- Polen, T., D. Rittmann, V. F. Wendisch and H. Sahm (2003). "DNA microarray analyses of the long-term adaptive response of Escherichia coli to acetate and propionate." Appl Environ Microbiol **69**(3): 1759-1774.
- Rhee, H. J., E. J. Kim and J. K. Lee (2007). "Physiological polyamines: simple primordial stress molecules." J Cell Mol Med **11**(4): 685-703.
- Sandberg, A., R. M. Branca, J. Lehtio and J. Forshed (2014). "Quantitative accuracy in mass spectrometry based proteomics of complex samples: the impact of labeling and precursor interference." J Proteomics **96**: 133-144.

- Sargent, F., E. G. Bogsch, N. R. Stanley, M. Wexler, C. Robinson, B. C. Berks and T. Palmer (1998). "Overlapping functions of components of a bacterial Sec-independent protein export pathway." EMBO J **17**(13): 3640-3650.
- Sauer, U., F. Canonaco, S. Heri, A. Perrenoud and E. Fischer (2004). "The Soluble and Membrane-bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of Escherichia coli." Journal of Biological Chemistry **279**(8): 6613-6619.
- Savitski, M. M., T. Mathieson, N. Zinn, G. Sweetman, C. Doce, I. Becher, F. Pachi, B. Kuster and M. Bantscheff (2013). "Measuring and managing ratio compression for accurate iTRAQ/TMT quantification." J Proteome Res **12**(8): 3586-3598.
- Savitski, M. M., G. Sweetman, M. Askenazi, J. A. Marto, M. Lang, N. Zinn and M. Bantscheff (2011). "Delayed fragmentation and optimized isolation width settings for improvement of protein identification and accuracy of isobaric mass tag quantification on Orbitrap-type mass spectrometers." Anal Chem **83**(23): 8959-8967.
- Schmidt, A., K. Kochanowski, S. Vedelaar, E. Ahrne, B. Volkmer, L. Callipo, K. Knoops, M. Bauer, R. Aebersold and M. Heinemann (2015). "The quantitative and condition-dependent Escherichia coli proteome." Nat Biotechnol.
- Silva, J. C., R. Denny, C. Dorschel, M. V. Gorenstein, G. Z. Li, K. Richardson, D. Wall and S. J. Geromanos (2006). "Simultaneous qualitative and quantitative analysis of the Escherichia coli Proteome - A sweet tale." Molecular & Cellular Proteomics **5**(4): 589-607.
- Sourjik, V. (2004). "Receptor clustering and signal processing in E. coli chemotaxis." Trends Microbiol **12**(12): 569-576.
- Stephanowitz, H., S. Lange, D. Lang, C. Freund and E. Krause (2012). "Improved two-dimensional reversed phase-reversed phase LC-MS/MS approach for identification of peptide-protein interactions." J Proteome Res **11**(2): 1175-1183.
- Stevenson, G., B. Neal, D. Liu, M. Hobbs, N. H. Packer, M. Batley, J. W. Redmond, L. Lindquist and P. Reeves (1994). "Structure of the O antigen of Escherichia coli K-12 and the sequence of its rfb gene cluster." J Bacteriol **176**(13): 4144-4156.
- Su, G., J. H. Morris, B. Demchak and G. D. Bader (2014). "BIOLOGICAL NETWORK EXPLORATION WITH CYTOSCAPE 3." Current protocols in bioinformatics / editorial board, Andreas D. Baxeavanis ... [et al.] **47**: 8.13.11-18.13.24.
- Treitz, C., L. Cassidy, A. Hockendorf, M. Leippe and A. Tholey (2015). "Quantitative proteome analysis of Caenorhabditis elegans upon exposure to nematicidal Bacillus thuringiensis." J Proteomics **113**: 337-350.
- van der Rest, M. E., C. Frank and D. Molenaar (2000). "Functions of the membrane-associated and cytoplasmic malate dehydrogenases in the citric acid cycle of Escherichia coli." J Bacteriol **182**(24): 6892-6899.
- Wang, H., S. Alvarez and L. M. Hicks (2012). "Comprehensive comparison of iTRAQ and label-free LC-based quantitative proteomics approaches using two Chlamydomonas reinhardtii strains of interest for biofuels engineering." J Proteome Res **11**(1): 487-501.
- Wang, Y., F. Yang, M. A. Gritsenko, Y. Wang, T. Clauss, T. Liu, Y. Shen, M. E. Monroe, D. Lopez-Ferrer, T. Reno, R. J. Moore, R. L. Klemke, D. G. Camp and R. D. Smith (2011). "Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells." Proteomics **11**(10): 2019-2026.

Wei, B., S. Shin, D. LaPorte, A. J. Wolfe and T. Romeo (2000). "Global regulatory mutations in *csrA* and *rpoS* cause severe central carbon stress in *Escherichia coli* in the presence of acetate." J Bacteriol **182**(6): 1632-1640.

Weiner, J. H., P. T. Bilous, G. M. Shaw, S. P. Lubitz, L. Frost, G. H. Thomas, J. A. Cole and R. J. Turner (1998). "A novel and ubiquitous system for membrane targeting and secretion of cofactor-containing proteins." Cell **93**(1): 93-101.

Wright, B. E., A. Longacre and J. M. Reimers (1999). "Hypermutation in derepressed operons of *Escherichia coli* K12." Proceedings of the National Academy of Sciences **96**(9): 5089-5094.

Yamanishi, Y., H. Mihara, M. Osaki, H. Muramatsu, N. Esaki, T. Sato, Y. Hizukuri, S. Goto and M. Kanehisa (2007). "Prediction of missing enzyme genes in a bacterial metabolic network. Reconstruction of the lysine-degradation pathway of *Pseudomonas aeruginosa*." FEBS J **274**(9): 2262-2273.

Zhao, J. and K. Shimizu (2003). "Metabolic flux analysis of *Escherichia coli* K12 grown on ¹³C-labeled acetate and glucose using GC-MS and powerful flux calculation method." J Biotechnol **101**(2): 101-117.

References Chapter 4.

Anatolii, I. K. t. and G. M. Kheifets (1971). "Investigation of Keto–Enol Tautomerism by Nuclear Magnetic Resonance Spectroscopy." Russian Chemical Reviews **40**(9): 773.

Bijtenhoorn, P., H. Mayerhofer, J. Muller-Dieckmann, C. Utpatel, C. Schipper, C. Hornung, M. Szesny, S. Grond, A. Thurmer, E. Brzuszkiewicz, R. Daniel, K. Dierking, H. Schulenburg and W. R. Streit (2011). "A novel metagenomic short-chain dehydrogenase/reductase attenuates *Pseudomonas aeruginosa* biofilm formation and virulence on *Caenorhabditis elegans*." PLoS One **6**(10): e26278.

Bijtenhoorn, P., C. Schipper, C. Hornung, M. Quitschau, S. Grond, N. Weiland and W. R. Streit (2011). "BpiB05, a novel metagenome-derived hydrolase acting on N-acylhomoserine lactones." J Biotechnol **155**(1): 86-94.

Brader, G., S. Sjoblom, H. Hyttiainen, K. Sims-Huopaniemi and E. T. Palva (2005). "Altering substrate chain length specificity of an acylhomoserine lactone synthase in bacterial communication." J Biol Chem **280**(11): 10403-10409.

Burdett, J. L. and M. T. Rogers (1964). "Keto-Enol Tautomerism in Beta Dicarboxyls Studied by Nuclear Magnetic Resonance Spectroscopy .1. Proton Chemical Shifts + Equilibrium Constants of Pure Compounds." Journal of the American Chemical Society **86**(11): 2105-&.

Cataldi, T. R., G. Bianco, M. Frommberger and P. Schmitt-Kopplin (2004). "Direct analysis of selected N-acyl-L-homoserine lactones by gas chromatography/mass spectrometry." Rapid Commun Mass Spectrom **18**(12): 1341-1344.

Cazes, J. (2010). Encyclopedia of chromatography. Boca Raton, FL, CRC Press.

Cazes, J. and G. W. Ewing (2005). Ewing's analytical instrumentation handbook. New York, Marcel Dekker.

Chan, K. G., S. Atkinson, K. Mathee, C. K. Sam, S. R. Chhabra, M. Camara, C. L. Koh and P. Williams (2011). "Characterization of N-acylhomoserine lactone-degrading bacteria associated with the *Zingiber officinale* (ginger) rhizosphere: co-existence of quorum quenching and quorum sensing in *Acinetobacter* and *Burkholderia*." BMC Microbiol **11**: 51.

Charlton, T. S., R. de Nys, A. Netting, N. Kumar, M. Hentzer, M. Givskov and S. Kjelleberg (2000). "A novel and sensitive method for the quantification of N-3-oxoacyl homoserine lactones using gas chromatography-mass spectrometry: application to a model bacterial biofilm." Environ Microbiol **2**(5): 530-541.

Chen, R., Z. Zhou, Y. Cao, Y. Bai and B. Yao (2010). "High yield expression of an AHL-lactonase from *Bacillus* sp. B546 in *Pichia pastoris* and its application to reduce *Aeromonas hydrophila* mortality in aquaculture." Microb Cell Fact **9**: 39.

Chowdhary, P. K., N. Keshavan, H. Q. Nguyen, J. A. Peterson, J. E. Gonzalez and D. C. Haines (2007). "Bacillus megaterium CYP102A1 oxidation of acyl homoserine lactones and acyl homoserines." Biochemistry **46**(50): 14429-14437.

Chun, C. K., E. A. Ozer, M. J. Welsh, J. Zabner and E. P. Greenberg (2004). "Inactivation of a *Pseudomonas aeruginosa* quorum-sensing signal by human airway epithelia." Proc Natl Acad Sci U S A **101**(10): 3587-3590.

Czajkowski, R. and S. Jafra (2009). "Quenching of acyl-homoserine lactone-dependent quorum sensing by enzymatic disruption of signal molecules." Acta Biochim Pol **56**(1): 1-16.

Dayringer, H. E. and F. W. McLafferty (1977). "Computer-Aided Interpretation of Mass-Spectra .14. Stirrs Prediction of Rings-Plus-Double-Bonds Values." Organic Mass Spectrometry **12**(1): 53-54.

- Dong, Y.-H. and L.-H. Zhang (2005). "Quorum sensing and quorum-quenching enzymes." J. Microbiol **43**(Spec. No.): 101-109.
- Dong, Y. H., J. L. Xu, X. Z. Li and L. H. Zhang (2000). "AiiA, an enzyme that inactivates the acylhomoserine lactone quorum-sensing signal and attenuates the virulence of *Erwinia carotovora*." Proc Natl Acad Sci U S A **97**(7): 3526-3531.
- Draganov, D. I., J. F. Teiber, A. Speelman, Y. Osawa, R. Sunahara and B. N. La Du (2005). "Human paraoxonases (PON1, PON2, and PON3) are lactonases with overlapping and distinct substrate specificities." J Lipid Res **46**(6): 1239-1247.
- Eberhard, A. (1972). "Inhibition and activation of bacterial luciferase synthesis." J Bacteriol **109**(3): 1101-1105.
- Eberhard, A., A. L. Burlingame, C. Eberhard, G. L. Kenyon, K. H. Nealson and N. J. Oppenheimer (1981). "Structural identification of autoinducer of *Photobacterium fischeri* luciferase." Biochemistry **20**(9): 2444-2449.
- Engbrecht, J. and M. Silverman (1984). "Identification of genes and gene products necessary for bacterial bioluminescence." Proc Natl Acad Sci U S A **81**(13): 4154-4158.
- Engelhardt, H., F. Steiner, C. Huber, R. Wintringer and e. al. (2005). "Fortgeschrittene Techniken der HPLC und HPLC-MS."
- European Pharmacopoeia Commission, C. o. E. (2010). European pharmacopoeia. Strasbourg, Council Of Europe : European Directorate for the Quality of Medicines and Healthcare.
- Fetzner, S. (2015). "Quorum quenching enzymes." Journal of Biotechnology **201**: 2-14.
- Fraissard, J. P. (1997). Physical adsorption : experiment, theory, and applications. Dordrecht ; Boston, Kluwer Academic Publishers.
- Francis, G. W. (1981). "Alkylthiolation for the determination of double-bond position in unsaturated fatty acid esters." Chemistry and Physics of Lipids **29**(4): 369-374.
- Fraune, S. and T. C. Bosch (2007). "Long-term maintenance of species-specific bacterial microbiota in the basal metazoan Hydra." Proc Natl Acad Sci U S A **104**(32): 13146-13151.
- Frommberger, M., P. Schmitt-Kopplin, G. Ping, H. Frisch, M. Schmid, Y. Zhang, A. Hartmann and A. Kettrup (2004). "A simple and robust set-up for on-column sample preconcentration--nano-liquid chromatography--electrospray ionization mass spectrometry for the analysis of N-acylhomoserine lactones." Anal Bioanal Chem **378**(4): 1014-1020.
- Fuqua, C., S. C. Winans and E. P. Greenberg (1996). "Census and consensus in bacterial ecosystems: the LuxR-LuxI family of quorum-sensing transcriptional regulators." Annu Rev Microbiol **50**: 727-751.
- Gould, T. A., J. Herman, J. Krank, R. C. Murphy and M. E. Churchill (2006). "Specificity of acyl-homoserine lactone synthases examined by mass spectrometry." J Bacteriol **188**(2): 773-783.
- Grandclement, C., M. Tannieres, S. Morera, Y. Dessaux and D. D. Faure (2015). "Quorum quenching: role in nature and applied developments." FEMS Microbiol Rev.
- Horai, H., M. Arita, S. Kanaya, Y. Nihei, T. Ikeda, K. Suwa, Y. Ojima, K. Tanaka, S. Tanaka, K. Aoshima, Y. Oda, Y. Kakazu, M. Kusano, T. Tohge, F. Matsuda, Y. Sawada, M. Y. Hirai, H. Nakanishi, K. Ikeda, N. Akimoto, T. Maoka, H. Takahashi, T. Ara, N. Sakurai, H. Suzuki, D. Shibata, S. Neumann, T. Iida, K. Tanaka, K. Funatsu, F. Matsuura, T. Soga, R. Taguchi, K. Saito and T. Nishioka (2010). "MassBank: a public repository for sharing mass spectral data for life sciences." J Mass Spectrom **45**(7): 703-714.

- Hong, Y. T., S. C. Deng, M. Daykin, P. C. Soo, J. R. Wei, K. T. Luh, S. W. Ho, S. Swift, H. C. Lai and P. Williams (2002). "The LuxR family protein SpnR functions as a negative regulator of N-acylhomoserine lactone-dependent quorum sensing in *Serratia marcescens*." Mol Microbiol **45**(6): 1655-1671.
- Hsu, F.-F. and J. Turk (1999). "Distinction among isomeric unsaturated fatty acids as lithiated adducts by electrospray ionization mass spectrometry using low energy collisionally activated dissociation on a triple stage quadrupole instrument." Journal of the American Society for Mass Spectrometry **10**(7): 600-612.
- Huang, Y. C., K. (2009). Multiple Fragmentation Methods for Small Molecule Characterization on a Dual Pressure Linear Ion Trap Orbitrap Hybrid Mass Spectrometer. Proceedings 18th International Mass Spectrometry Conference, Bremen, Germany.
- ICH-Q2B (1996). Validation of Analytical Methods: Methodology. International Conference on Harmonization (ICH) of Technical Requirements for the Registration of Pharmaceuticals for Human Use.
- Ishida, T., F. Hirata, H. Sato and S. Kato (1998). "Molecular Theory of Solvent Effect on Keto-Enol Tautomers of Formamide in Aprotic Solvents: RISM-SCF Approach." The Journal of Physical Chemistry B **102**(11): 2045-2050.
- Jedrychowski, M. P., E. L. Huttlin, W. Haas, M. E. Sowa, R. Rad and S. P. Gygi (2011). "Evaluation of HCD- and CID-type fragmentation within their respective detection platforms for murine phosphoproteomics." Mol Cell Proteomics **10**(12): M111 009910.
- Keller, B. O., J. Sui, A. B. Young and R. M. Whittall (2008). "Interferences and contaminants encountered in modern mass spectrometry." Anal Chim Acta **627**(1): 71-81.
- Kind, T. and O. Fiehn (2007). "Seven Golden Rules for heuristic filtering of molecular formulas obtained by accurate mass spectrometry." BMC Bioinformatics **8**: 105-105.
- Kopka, J., N. Schauer, S. Krueger, C. Birkemeyer, B. Usadel, E. Bergmuller, P. Dormann, W. Weckwerth, Y. Gibon, M. Stitt, L. Willmitzer, A. R. Fernie and D. Steinhauser (2005). "GMD@CSB.DB: the Golm Metabolome Database." Bioinformatics **21**(8): 1635-1638.
- Lade, H., D. Paul and J. H. Kweon (2014). "N-acyl homoserine lactone-mediated quorum sensing with special reference to use of quorum quenching bacteria in membrane biofouling control." Biomed Res Int **2014**: 162584.
- Leipert, J., C. Treitz, M. Leippe and A. Tholey (2017). "Identification and Quantification of N-Acyl Homoserine Lactones Involved in Bacterial Communication by Small-Scale Synthesis of Internal Standards and Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry." J Am Soc Mass Spectrom **28**(12): 2538-2547.
- Marketon, M. M., M. R. Gronquist, A. Eberhard and J. E. Gonzalez (2002). "Characterization of the *Sinorhizobium meliloti* sinR/sinI locus and the production of novel N-acyl homoserine lactones." J Bacteriol **184**(20): 5686-5695.
- Masuo, S., Y. Terabayashi, M. Shimizu, T. Fujii, T. Kitazume and N. Takaya (2010). "Global gene expression analysis of *Aspergillus nidulans* reveals metabolic shift and transcription suppression under hypoxia." Mol Genet Genomics **284**(6): 415-424.
- McLafferty, F. W. and F. e. Tureček (1993). Interpretation of mass spectra. Mill Valley, Calif., University Science Books.
- More, M. I., L. D. Finger, J. L. Stryker, C. Fuqua, A. Eberhard and S. C. Winans (1996). "Enzymatic synthesis of a quorum-sensing autoinducer through use of defined substrates." Science **272**(5268): 1655-1658.

- Morin, D., B. Grasland, K. Vallee-Rehel, C. Dufau and D. Haras (2003). "On-line high-performance liquid chromatography-mass spectrometric detection and quantification of N-acylhomoserine lactones, quorum sensing signal molecules, in the presence of biological matrices." J Chromatogr A **1002**(1-2): 79-92.
- Morohoshi, T., T. Inaba, N. Kato, K. Kanai and T. Ikeda (2004). "Identification of quorum-sensing signal molecules and the LuxRI homologs in fish pathogen *Edwardsiella tarda*." J Biosci Bioeng **98**(4): 274-281.
- Nealson, K. H., T. Platt and J. W. Hastings (1970). "Cellular control of the synthesis and activity of the bacterial luminescent system." J Bacteriol **104**(1): 313-322.
- Orsatti, L., S. Di Marco, C. Volpari, A. Vannini, P. Neddermann and F. Bonelli (2002). "Determination of the stoichiometry of noncovalent complexes using reverse-phase high-performance liquid chromatography coupled with electrospray ion trap mass spectrometry." Anal Biochem **309**(1): 11-18.
- Ortori, C. A., S. Atkinson, S. R. Chhabra, M. Camara, P. Williams and D. A. Barrett (2007). "Comprehensive profiling of N-acylhomoserine lactones produced by *Yersinia pseudotuberculosis* using liquid chromatography coupled to hybrid quadrupole-linear ion trap mass spectrometry." Anal Bioanal Chem **387**(2): 497-511.
- Pearson, J. P., K. M. Gray, L. Passador, K. D. Tucker, A. Eberhard, B. H. Iglewski and E. P. Greenberg (1994). "Structure of the autoinducer required for expression of *Pseudomonas aeruginosa* virulence genes." Proc Natl Acad Sci U S A **91**(1): 197-201.
- Pietschke, C., C. Treitz, S. Foret, A. Schultze, S. Kunzel, A. Tholey, T. C. G. Bosch and S. Fraune (2017). "Host modification of a bacterial quorum-sensing signal induces a phenotypic switch in bacterial symbionts." Proc Natl Acad Sci U S A **114**(40): E8488-E8497.
- Puskas, A., E. P. Greenberg, S. Kaplan and A. L. Schaefer (1997). "A quorum-sensing system in the free-living photosynthetic bacterium *Rhodobacter sphaeroides*." J Bacteriol **179**(23): 7530-7537.
- Ravn, L., A. B. Christensen, S. Molin, M. Givskov and L. Gram (2001). "Methods for detecting acylated homoserine lactones produced by Gram-negative bacteria and their application in studies of AHL-production kinetics." J Microbiol Methods **44**(3): 239-251.
- Reaney, M. J. T., Y.-D. Liu and W. G. Taylor (2001). "Gas chromatographic analysis of diels-alder adducts of geometrical and positional isomers of conjugated linoleic acid." Journal of the American Oil Chemists' Society **78**(11): 1083-1086.
- Reichardt, C. and T. Welton (2011). Solvents and solvent effects in organic chemistry. Weinheim, Germany, Wiley-VCH.
- Ruiz-Rodriguez, A., G. Reglero and E. Ibanez (2010). "Recent trends in the advanced analysis of bioactive fatty acids." J Pharm Biomed Anal **51**(2): 305-326.
- Sandusky, P. O. (2014). "Expansion of the Classic Acetylacetone Physical Chemistry Laboratory NMR Experiment: Correlation of the Enol-Keto Equilibrium Position with the Solvent Dipole Moment." Journal of Chemical Education **91**(5): 739-742.
- Schripsema, J., K. E. de Rudder, T. B. van Vliet, P. P. Lankhorst, E. de Vroom, J. W. Kijne and A. A. van Brussel (1996). "Bacteriocin small of *Rhizobium leguminosarum* belongs to the class of N-acyl-L-homoserine lactone molecules, known as autoinducers and as quorum sensing co-transcription factors." J Bacteriol **178**(2): 366-371.
- Shaw, P. D., G. Ping, S. L. Daly, C. Cha, J. E. Cronan, Jr., K. L. Rinehart and S. K. Farrand (1997). "Detecting and characterizing N-acyl-homoserine lactone signal molecules by thin-layer chromatography." Proc Natl Acad Sci U S A **94**(12): 6036-6041.

- Shibahara, A., K. Yamamoto, A. Kinoshita and B. L. Anderson (2008). "An Improved Method for Preparing Dimethyl Disulfide Adducts for GC/MS Analysis." Journal of the American Oil Chemists' Society **85**(1): 93-94.
- Sio, C. F., L. G. Otten, R. H. Cool, S. P. Diggle, P. G. Braun, R. Bos, M. Daykin, M. Camara, P. Williams and W. J. Quax (2006). "Quorum quenching by an N-acyl-homoserine lactone acylase from *Pseudomonas aeruginosa* PAO1." Infect Immun **74**(3): 1673-1682.
- Sitnikov, D. M., J. B. Schineller and T. O. Baldwin (1995). "Transcriptional regulation of bioluminescence genes from *Vibrio fischeri*." Mol Microbiol **17**(5): 801-812.
- Smith, C. A., G. O'Maille, E. J. Want, C. Qin, S. A. Trauger, T. R. Brandon, D. E. Custodio, R. Abagyan and G. Siuzdak (2005). "METLIN: a metabolite mass spectral database." Ther Drug Monit **27**(6): 747-751.
- Spencer, C. C., M. Bertrand, M. Travisano and M. Doebeli (2007). "Adaptive diversification in genes that regulate resource use in *Escherichia coli*." PLoS Genet **3**(1): e15.
- Teplitski, M., A. Eberhard, M. R. Gronquist, M. Gao, J. B. Robinson and W. D. Bauer (2003). "Chemical identification of N-acyl homoserine lactone quorum-sensing signals produced by *Sinorhizobium meliloti* strains in defined medium." Arch Microbiol **180**(6): 494-497.
- Tholey, A. and E. Heinzle (2006). "Ionic (liquid) matrices for matrix-assisted laser desorption/ionization mass spectrometry-applications and perspectives." Anal Bioanal Chem **386**(1): 24-37.
- Torres, M., M. Romero, S. Prado, J. Dubert, A. Tahrioui, A. Otero and I. Llamas (2013). "N-acylhomoserine lactone-degrading bacteria isolated from hatchery bivalve larval cultures." Microbiol Res **168**(9): 547-554.
- Uroz, S., S. R. Chhabra, M. Camara, P. Williams, P. Oger and Y. Dessaux (2005). "N-Acylhomoserine lactone quorum-sensing molecules are modified and degraded by *Rhodococcus erythropolis* W2 by both amidolytic and novel oxidoreductase activities." Microbiology **151**(Pt 10): 3313-3322.
- Xu, F., T. Byun, H. J. Deussen and K. R. Duke (2003). "Degradation of N-acylhomoserine lactones, the bacterial quorum-sensing molecules, by acylase." J Biotechnol **101**(1): 89-96.
- Zhang, J. I., W. A. Tao and R. G. Cooks (2011). "Facile determination of double bond position in unsaturated fatty acids and esters by low temperature plasma ionization mass spectrometry." Anal Chem **83**(12): 4738-4744.
- Zhu, J., Y. Chai, Z. Zhong, S. Li and S. C. Winans (2003). "Agrobacterium bioassay strain for ultrasensitive detection of N-acylhomoserine lactone-type quorum-sensing molecules: detection of autoinducers in *Mesorhizobium huakuii*." Appl Environ Microbiol **69**(11): 6949-6953.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und nur mithilfe der von mir angegebenen Quellen und Hilfsmittel unter Anleitung meines Akademischen Betreuers verfasst habe. Alle wörtlich oder inhaltlich übernommenen Stellen habe ich als solche gekennzeichnet. Die vorliegende Arbeit entstand unter Einhaltung der Regeln guter wissenschaftlicher Praxis der Deutschen Forschungsgemeinschaft und wurde keiner anderen Universität als der Christian-Albrechts-Universität zu Kiel zur Begutachtung vorgelegt.

Christian Treitz

Kiel, den 10. Dezember 2017