TUM School of Life Sciences

Technische Universität München

Advancing Metabolomics in *Caenorhabditis elegans* - Exploring Innovative Approaches for Metabolite Analysis and Annotation

Liesa Salzer

Vollständiger Abdruck der von der TUM School of Life Sciences der Technischen Universität München zur Erlangung des akademischen Grades einer

Doktorin der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitz:
Prof. Dr. Wilfried Schwab

Prüfende der Dissertation:
1. apl. Prof. Dr. Philippe Schmitt-Kopplin
2. Prof. Dr. Corinna Dawid

Die Dissertation wurde am 28.11.2023 bei der Technischen Universität München eingereicht und durch die TUM School of Life Sciences am 29.05.2024 angenommen.

"In the middle of difficulty lies opportunity."
- Albert Einstein

Acknowledgement

The research in this thesis was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) — Project number 431572533 (MetClassNet).

I would like to express my heartfelt gratitude to all those who have supported me throughout this remarkable journey of completing my dissertation. Their unwavering guidance, encouragement, and support have been invaluable, and I am deeply thankful for their contributions.

First and foremost, I would like to thank my doctoral advisor, Prof. Dr. Philippe Schmitt-Kopplin, for giving me the opportunity to work in this wonderful research group. I am grateful for your understanding and empathy, especially during the challenging times of Covid. Your valuable input and guidance for my thesis, as well as our engaging discussions about capillary electrophoresis, have played a crucial role in shaping my work.

I extend my deepest appreciation to my supervisor, Michael Witting, for bringing me into this fantastic project and sharing his expertise in metabolomics, mass spectrometry, bioinformatics, R coding, and many more. I consider myself fortunate to have had you as my supervisor, always providing valuable advice and helping me bring out the best in my work.

Furthermore, I would like to thank Steffen Neumann for mentoring me and providing valuable ideas and hints that have significantly enhanced the quality of my research. Your assistance with coding issues in R and GitHub, as well as our insightful discussions within the project, have been immensely helpful.

I would also like to express my gratitude to all remaining members of the MetClassNet project, Clément Frainay, Fabien Jourdan, Elva María Novoa-Del-Toro, Adam Amara, Reza Salek, and Sarah Scharfenberg. It has been an honor to work with such highly skilled experts in a collaborative and stimulating environment. I have gained a wealth of knowledge about metabolomics, graphs, and networks, far beyond what I could have imagined. I will truly miss our discussions and monthly meetings.
A special thank you goes to Elva, who has provided me with scientific and personal advice, delightful chats, and Zoom meetings that have made me laugh so hard that I cried. From the dance party at the Metabolomics 2022 conference to always lending an open ear to my concerns, you have become a dear friend.

I want to acknowledge my colleagues and fellow researchers at the research group Analytical BioGeoChemistry (BGC). Your support and encouragement have made this academic journey all the more rewarding and enjoyable. In particular, I want to express my gratitude to Michelle Berger for generously sharing your tips and tricks, especially regarding LC-MS troubleshooting and R coding. Our conversations have revealed that we are all impostors in a sinking boat, having the same concerns and issues. You have been a crucial anchor of mental support here at BGC. I would also like to thank Dr. Marco Matzka, who has been a constant source of inspiration. Our shared thoughts and opinions have illuminated the path toward our common goal, making it seem more achievable than ever. Additionally, I appreciate Bastian Blume for teaching me *C. elegans* cultivation and also for his company during the weekends of worm bleaching in lab 27, making it much more enjoyable. Many thanks to Alesia Walker for her support regarding LC-MS troubleshooting, as well as Jenny Uhl and Astrid Bösl for their assistance with all administrative matters. I would also like to thank the members of the BGC "Frühstücksclub" (Tina Daubmeier, Brigitte Look, Marco, Jenny, and Astrid) for the breakfasts and coffee that have helped me get through the day. Special thanks to Leopold Weidner and Philippe Diederich for the enjoyable discussions, laughs, and lunch breaks we have shared. Furthermore, I want to express my gratitude to remaining fellow doctoral researchers, Eva-Maria Harrider, Yingfei Yan, Erwin Kupczyk, Habiba Selmi, Xin Zhang, and all those who have already completed their journey. To all the other (present and former) members of the BGC whom I haven't mentioned by name, you are also cherished and have contributed to my wonderful experience.

I consider myself incredibly fortunate to have a circle of friends who have always been patient and understanding when I had to cancel plans due to work commitments. Your open ears and support in helping me release stress have been invaluable.

Last but certainly not least, my deepest gratitude goes to my family. Words cannot express how grateful I am for your unconditional love, unwavering support, and endless encouragement. You have believed in me wholeheartedly, and your sacrifices and understanding have made this journey possible.

Abstract

Metabolomics is a rapidly growing field of research that focuses on the analysis of small molecules (metabolites) that are produced by cells and organisms as part of their metabolism. It has the potential to provide insights into their metabolic processes and therefore revolutionize our understanding of health and disease. *Caenorhabditis elegans* (*C. elegans*) is a model organism used in biological research due to its genetic tractability, but analyzing its metabolome is challenging due to the complexity and diversity of its metabolites, the majority of which remain unknown. To address this challenge, this thesis employed two innovative approaches in *C. elegans* metabolomics.

Polar metabolite analysis in *C. elegans* has traditionally relied on hydrophilic interaction liquid chromatography coupled to mass spectrometry (HILIC-MS) or nuclear magnetic resonance (NMR) spectroscopy methods. To enhance our understanding of *C. elegans* metabolism, we[1] explored capillary electrophoresis coupled to mass spectrometry (CE-MS) as an alternative approach for analyzing polar and charged compounds in the nematode.
One crucial aspect we addressed in this thesis was effective mobility transformation of migration time scale in CE-MS, which led to more reproducible data. To streamline this process and seamlessly integrate with other metabolomics processing and analysis tools in R, we developed the *MobilityTransformR* R package. A direct comparison was conducted with HILIC-MS using 60 reference standards known to be present in *C. elegans* to validate the applicability of CE-MS. Our results demonstrated that CE-MS offers several advantages, including narrower peak widths, reduced saturation of MS signals, and distinct selectivity. CE-MS analysis of *daf-2*, a well-described long-lived mutant of *C. elegans*, revealed both shared and distinct metabolites compared to HILIC-MS. Consequently, not only known biomarkers of longevity were recovered but we also discovered novel putative biomarkers, thus advancing our comprehension of the molecular mechanisms that govern lifespan regulation in this model organism.

The second approach used in this thesis focuses on developing bioinformatics tools to enhance annotation of novel metabolites in non-targeted metabolomics data, particularly for those lacking reference spectra. The **A**nnotation **P**ropagation through multiple **EX**perimental networks (APEX) workflow was developed, which combines spectral similarity networks with mass difference networks and homologous series. The APEX workflow was applied

[1] In the context of this thesis, the pronoun "we" is employed to refer to the author of the work. It is a convention in scientific writing to use "we" to convey authorship, even when an individual is the sole contributor.

to automate the annotation of glycerophospho *N*-acyl ethanolamides (GPNAEs), a metabolite class recently identified in *C. elegans*. APEX successfully annotated GPNAEs through network propagation of manually annotated seed nodes. Furthermore, its most significant achievement was the annotation of GlycoGPNAEs, which are glycovariants disconnected in the spectral similarity network because of slight changes in fragmentation. Furthermore, APEX proved valuable in annotating metabolite features with missing MS^2 data. To validate the efficacy of APEX, we employed different sets of seed nodes, conducted leave-one-out cross-validations, utilized diverse datasets, and compared the results with molecular formulas provided by other software. By applying APEX to the annotation of ascarosides and related variants we demonstrated its versatility in annotating various types of metabolites, further reinforcing its utility as a powerful tool in metabolomics research.

Overall, this thesis has significant implications for metabolomics and our understanding of *C. elegans*' metabolism. The use of CE-MS and bioinformatics methods in non-targeted metabolomics offered valuable tools for identifying and characterizing known and unknown metabolites in *C. elegans*, providing insights into biochemical pathways and metabolic processes.

Zusammenfassung

Die Metabolomik ist ein sich schnell entwickelndes Forschungsgebiet, das sich auf die Analyse von kleinen Molekülen (Metabolite) konzentriert, welche von Zellen und Organismen als Teil ihres Stoffwechsels produziert werden. Sie hat das Potenzial Einblicke in komplexe Stoffwechselprozesse zu liefern und unser Verständnis von Gesundheit und Krankheit zu revolutionieren. *Caenorhabditis elegans* (*C. elegans*) ist ein Modellorganismus, der aufgrund seiner genetischen Zugänglichkeit häufig in der biologischen Forschung verwendet wird. Jedoch stellt die Analyse seines Metaboloms aufgrund der Komplexität und Vielfalt der Metaboliten, von denen der Großteil noch unbekannt ist, eine besondere Herausforderung dar. Um diese Herausforderung zu addressieren, wurden in dieser Arbeit zwei innovative Ansätze angewendet und entwickelt.

Die Analyse polarer Metaboliten in *C. elegans* stützt sich traditionell auf die Hydrophile Interaktionsflüssigchromatographie gekoppelt mit Massenspektrometrie (HILIC-MS) oder Kernspinresonanz (NMR)-Spektroskopie. Um unser Verständnis des *C. elegans*-Stoffwechsels zu verbessern, haben wir[1] die Kapillarelektrophorese gekoppelt mit Massenspektrometrie (CE-MS) als alternative Methode zur Analyse polarer und geladener Verbindungen im Fadenwurm verwendet.

Ein wichtiger Punkt den wir in dieser Arbeit demonstriert haben, war die Umwandlung der Migrationszeit in eine effektive Mobilitätsskala der CE-MS Daten. Dies führte zu reproduzierbareren Daten. Um diesen Prozess zu vereinfachen und nahtlos in andere Metabolomik-Verarbeitungs- und Analysetools in R zu integrieren, wurde das R-Paket *MobilityTransformR* entwickelt. Ein direkter Vergleich von CE-MS und HILIC-MS wurde durchgeführt, unter Verwendung von 60 Referenzstandards, die bekanntermaßen in *C. elegans* vorkommen. Unsere Ergebnisse zeigten, dass CE-MS mehrere Vorteile bietet, darunter schmalere Peakbreiten, verringerte Sättigung von MS-Signalen und eine unterschiedliche Selektivität. Die CE-MS-Analyse von *daf-2*, einer gut beschriebenen langlebigen Mutante von *C. elegans*, zeigte sowohl gemeinsame als auch zusäzliche Metaboliten im Vergleich zur HILIC-MS. Dadurch konnten wir nicht nur bekannte Biomarker für Langlebigkeit wiederfinden, sondern auch neue potenzielle Biomarker entdecken. Diese Ergebnisse verbessern unser Verständnis für molekularen Mechanismen, die die Regulation der Lebensdauer in dem Modellorganismus steuern.

[1] In dieser Arbeit bezieht sich das Pronomen "wir" auf den Autor der Arbeit. Es ist in wissenschaftlichen Texten üblich, "wir" zu verwenden, um die Urheberschaft auszudrücken, selbst wenn eine Einzelperson der alleinige Verfasser ist.

Der zweite Ansatz, der in dieser Arbeit verwendet wurde, konzentriert sich auf die Entwicklung von Bioinformatik-Tools, um die Annotation neuer Metabolite in nicht-zielgerichteten Metabolomik-Daten zu verbessern, insbesondere für solche ohne Referenzspektren. Wir haben den ***A**nnotation **P**ropagation through multiple **EX**perimental networks* (APEX)-Workflow entwickelt, der spektrale Ähnlichkeitsnetzwerke mit Massendifferenznetzwerken und homologen Reihen kombiniert. Wir verwendeten APEX zur automatisierten Annotation von Glycerophospho *N*-Acyl-Ethanolamiden (GPNAEs), einer kürzlich identifizierten Metabolitklasse in *C. elegans*. Mit APEX konnten erfolgreich GPNAE durch Netzwerk-Propagation von manuell annotierten Startknoten annotiert werden. Jedoch war sein bedeutendster Nutzen die Annotation von GlycoGPNAEs, Glucose Varianten der GPNAE, die aufgrund geringfügiger Änderung der Fragmentierung im spektralen Ähnlichkeitsnetzwerk getrennt von GPNAE vorliegen. Außerdem erwies sich APEX als wertvoll bei der Annotation von Substanzen mit fehlenden MS^2-Daten. Um die Effektivität von APEX zu validieren, wurden verschiedene Sets von Startknoten getestet, *leave-one-out cross-validation* durchgeführt, verschiedene Datensätze genutzt und die Ergebnisse mit den molekularen Formeln aus anderer Software verglichen. Durch die Anwendung von APEX für die Annotation von Ascarosiden und verwandten Varianten wurde seine Vielseitigkeit für der Annotation verschiedener Arten von Metaboliten demonstriert und damit seine Nützlichkeit als leistungsstarkes Werkzeug in der Metabolomik-Forschung untermauert.

Insgesamt hat diese Arbeit bedeutende Auswirkungen auf die Metabolomik und unser Verständnis des Stoffwechsels von *C. elegans*. Die Verwendung von CE-MS und bioinformatischen Methoden in der nicht-zielgerichteten Metabolomik bietet wertvolle Werkzeuge zur Identifizierung und Charakterisierung bekannter und unbekannter Metaboliten in *C. elegans* und ermöglicht neue Einblicke in biochemische Wege und Stoffwechselprozesse.

Contents

Scientific Communications

Publications

Publications directly addressed in this thesis

Chapter 1: Salzer, L. & Witting, M. (2021).*Quo Vadis Caenorhabditis elegans Metabolomics-A Review of Current Methods and Applications to Explore Metabolism in the Nematode.* Metabolites, 11(5), 284. https://doi.org/10.3390/metabo11050284

Chapter 2: Salzer, L., Witting, M., & Schmitt-Kopplin, P. (2022). *MobilityTransformR: an R package for effective mobility transformation of CE-MS data.* Bioinformatics (Oxford, England), 38(16), 4044–4045. https://doi.org/10.1093/bioinformatics/btac441

Chapter 3: Salzer, L., Schmitt-Kopplin, P., & Witting, M. (2023). *Capillary electrophoresis-mass spectrometry as a tool for Caenorhabditis elegans metabolomics research.* Metabolomics : Official journal of the Metabolomic Society, 19(7), 61. https://doi.org/10.1007/s11306-023-02025-7

Chapter 4: Salzer, L., Novoa-del-Toro, EM., Frainay, C., Kissoyan, K., Jourdan, F., Dierking, K., Witting, M., (2023). *APEX – an annotation propagation workflow through multiple experimental networks to improve the annotation of new metabolite classes in Caenorhabditis elegans.* Analytical Chemistry. https://doi.org/10.1021/acs.analchem.3c02797

Further publications within the framework of the doctoral studies

Amara A, Frainay C, Jourdan F, Naake T, Neumann S, Novoa-Del-Toro EM, Salek RM, Salzer, L., Scharfenberg S, Witting M. *Networks and Graphs Discovery in Metabolomics Data Analysis and Interpretation.* Frontiers in molecular biosciences, 9, 841373. https://doi.org/10.3389/fmolb.2022.841373

Rainer, J.; Vicini, A.; Salzer, L..; Stanstrup, J.; Badia, J.M.; Neumann, S.; Stravs, M.A.; Verri Hernandes, V.; Gatto, L.; Gibb, S.; Witting, M. *A Modular and Expandable Ecosystem for Metabolomics Data Annotation in R.* Metabolites 2022, 12, 173. https://doi.org/10.3390/metabo12020173

Oral presentations

Metabolomics 2020 online conference, Metabolomics Society, Title: *Spectra consolidation of high resolution tandem mass spectrometry data in LC-MS based non-targeted metabolomics and lipidomics*

3rd de.NBI/ELIXIR-DE metaRbolomics Hackathon, 2021, Lutherstadt Wittenberg, Germany, Title: *Why we need an R package with effective mobility transformation for CE-MS in Metabolomics*

32. Doktorandenseminar Hohenroda, 2022, online. Title: *CE-MS in Caenorhabditis elegans - A complementary platform to HILIC-MS?*

Metabolomics 2022 conference, Valencia Spain, Metabolomics Society, Title: *Multi-network integration to analyze non-targeted LC-MS metabolomics data from Caenorhabditis elegans*

Poster presentations

Metabolomics 2020 online conference, Metabolomics Society, Title: *Spectra consolidation of high resolution tandem mass spectrometry data in LC-MS based non-targeted metabolomics and lipidomics*

Metabolomics 2021 online conference, Metabolomics Society, Title: *CE-MS in Caenorhabditis elegans Metabolomics – A Complementary Platform to HILIC-MS?*

Abbreviations

ANP	Aqueous normal phase
APCI	Atmospheric pressure chemical ionization
APEX	Annotation through multiple experimental networks
ATP	Adenosine triphosphate
BGE	Background electrolyte
C. elegans	*Caenorhabditis elegans*
CE	Capillary electrophoresis
CI	Chemical ionization
CID	Collision induced dissociation
CL	Cardiolipin
CoA	Coenzyme A
COSY	^{1}H-^{1}H correlated spectroscopy
CSS	Collision cross-section
CZE	Capillary zone electrophoresis
DANS	Differential analysis by 2D-NMR spectroscopy
DDA	Data-dependent acquisition
DI	Direct infusion
DIA	Data-independent acquisition
DT	Drift tube
E. coli	*Echerichia coli*
EI	Electron ionization
EIC	Extracted ion chromatogram
EIE	Extracted ion electropherogram
ESI	Electrospray ionization
EOF	Electroosmotic flow
FA	Fatty acid
FBA	Flux balance analysis
FID	Flame ionization detector
GC	Gas chromatography
GGM	Gaussian graphical model
GSM	Genome scale metabolic model
GSMN	Genome-scale metabolic networks
GPNAE	Glycerophospho N-acyl ethylamide
HCA	Hierarchical cluster analysis

HEPT	Height equivalent to a theoretical plate
HILIC	Hydrophilic interaction liquid chromatography
HMDB	Human metabolome database
HMN	^{13}C-Heteronuclear Multidimensional NMR
HPLC	High performance liquid chromatography
HR-MACS	High-resolution magic-angle coil spinning
HR-MAS	High-resolution magic-angle spinning
HR-MS	High resolution-mass spectrometry
HSQC	^{1}H-^{13}C Heteronuclear single quantum coherence
IGF	Insulin/insulin-like growth factor
IIS	Insulin/insulin-like signaling
IM	Ion mobility
KEGG	Kyoto Encyclopedia of Genes and Genomes
LPE	Lysophosphatidylethanolamine
m/z	Mass-to-charge ratio
MeOH	Methanol
MMBA	Monomethyl branched chain fatty acid
MRM	Multiple reaction monitoring
mRNA	Messenger RNA
MS	Mass spectrometry
MT	Migration time
MUFA	Monounsaturated fatty acid
MW	Molecular weight
NAE	N-acyl ethylamide
NMR	Nuclear magnetic resonance spectroscopy
OPLS-DA	Orthogonal projection to latent structure with discriminant analysis
PC	Phosphatidylcholine
PCA	Principal component analysis
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
PI	Phosphatidylinositol
PL	Phospholipids
PUFA	Polyunsaturated fatty acid
qToF-MS	Quadrupole time-of-flight mass spectrometry
RNA	Ribonucleic acid
RNAi	RNA interference
RP(LC)	Reversed phase (liquid chromatography)
RSD	Relative standard deviation
RT	Retention time
SFC	Supercritical fluid chromatography
SM	Sphingomyelin

SMID DB	Small Molecule Identifier Database
SNS	Seed node set
SPE	Solid phase extractions
SWATH	Sequential window acquisition of all theoretical fragment ion spectra
TCA	Tricarboxylic acid cycle
TIC	Total ion chromatogram
TIE	Total ion electropherogram
TLC	Thin layer chromatography
TOCSY	^{1}H-^{1}H Total correlation spectroscopy
ToF	Time-of-flight
UPLC	Ultra-high-performance liquid chromatography
YA	Young adult

1 General Introduction and Methods

1.1 Metabolomics

1.1.1 Metabolomics and other -omics technologies

Understanding health and disease in humans and animals is one of the major goals of modern science. The analysis of biomolecules is crucial for this purpose. To obtain a complete picture of the current state of a biological system, so-called -omics technologies have been developed. They involve the collective analysis of as many biomolecules simultaneously as possible, looking at the system as a whole.
The most prominent members of the -omics field are genomics, transcriptomics, proteomics and metabolomics (Figure 1.1). Today, the term of -omics technologies has even extended to different fields, such as the microbiome (microbiomics), the epigenome (epigenomics), interactions of biomolecules (interactomics), sugars and carbohydrates (glycomics), or food (foodomics) [1].

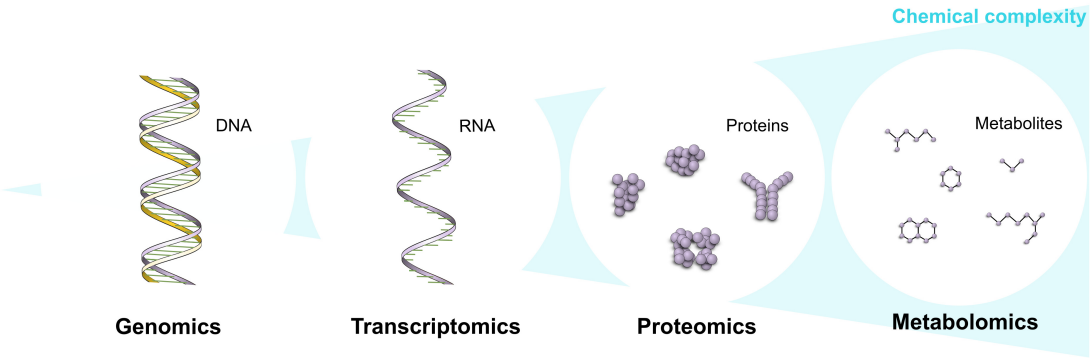


Figure 1.1 Scheme of -omics technologies. With increasing chemical complexity: Genomics, i.e. the analysis of all genes (DNA). Transcriptomics, i.e. the analysis of all RNA molecules. Proteomics, i.e. the analysis of all proteins. Metabolomics, i.e. the analysis of all metabolites.

The first of the four main -omics technologies was genomics, which is also probably the best known. Genomics analyses the genome, i.e. the complete set of genes of a biological system. The term genome was introduced more than a hundred years ago by Hans Winkler, referring to "the haploid chromosome set, which, together with the pertinent protoplasm, specifies the material foundations of the species" (rough translation) [2]. 1986 the term genomics was used for the first time by Thomas H. Roderick to name the journal *Genomics*,

with the idea to "discuss the feasibility of mapping the entire human genome" [3]. Genomics is an indispensable tool in today's medicine and used to predict, diagnose and treat diseases, demonstrating the direct impact of the genome on the phenotype [4]. However, in addition, various factors such as diet and environment have been shown to influence the phenotype, which is why other -omics technologies are finding their way into science.
The measurement of gene expression is more detailed and used to assess the gene activity in different cells and species. The first transcriptome, i.e. the entirety of all ribonucleic acid (RNA) molecules in a system, was elucidated in *Saccharomyces cerevisiae* [5]. Transcriptomics is one step closer to the actual phenotype compared to genomics, and find various applications in medicine such as classification of cancer types and identification of biomarkers for cancer in early stages [6].
Another step closer towards the phenotype is represented by an organisms proteome, i.e. the completeness of proteins in a cell or organism. Proteomics technologies are being developed since 1996 and have shown great significance in today's medicine [7]. Proteomics is the characterization of proteins, including their abundance, structure and post-translational modifications, which cannot be accessed by genomics or transcriptomics. Compared to deoxyribonucleic acid (DNA) and RNA molecules, proteins are much more diverse regarding their size, expression, hydrophobicity and (post-translational) modifications. Therefore, their analysis is more challenging and more bioanalytical tools exist [8]. Proteomics can help to understand disease mechanisms or help to discover biomarkers and is applied in fields such es food, health and disease, or microbiology. Another advantage of proteomics is that biomarkers and proteins can be found in all types of tissues in biofluids enabling the analysis of less invasive samples [9]. For example, proteomics analysis of urine samples was successfully applied to discover biomarkers for kidney disease [10].
However, closest to the phenotype is the metabolome, being defined as all metabolites present in a biological system. Metabolites are small molecules up to 1500 Da with varying concentrations in several orders of magnitude [11]. The metabolome is organism-specific and can vary greatly throughout different species. For example the *E. coli* metabolome have been estimated at over 3700 metabolites (`https://ecmdb.ca/`), where 3261 of are associated with experimental data (NMR or MS spectra) [12]. For the human metabolome, the prediction of metabolites reaches over 100.000 [13]. However, we can only detect a small portion (< 10%) of those metabolites experimentally [14]. The majority of predicted metabolites in human categorize into lipid classes. Predictions of lipids are relatively easy due to ontologies but their detection is limited due to the exact estimation of the position of double bonds or the location of alkyl chains. The plant metabolome is even bigger. There are more than 200.000 known secondary metabolites and estimations even reach numbers over 1.000.000 [15, 16].
Metabolites are connected in a complex network of biochemical reactions, resulting in high co-dependency on each other as they are products and substrates of those reactions. This makes the metabolome highly dynamic and sensitive to external factors such as environment

and diet [17]. Metabolism is the key to all life and includes the main functions of converting higher molecular weight compounds into energy (catabolism), the building of tissues, proteins, and energy stores (anabolism), or detoxification processes. The energy released by catabolic pathways will be used by anabolic pathways, making them complementary and dependent on one another. Even more, there are pathways that will be either anabolic or catabolic depending on the needs and availability of energy to maintain homeostasis in living systems. One particular example is glycolysis which is the conversion of glucose into pyruvate in order to generate energy in the form of adenosine triphosphate (ATP). The reversed reaction, i.e. building glucose from pyruvate intermediates is called gluconeogenesis and takes place when glucose levels in the organism are too low (e.g. after fasting, starvation, or exercise). The product of glycolysis, pyruvate, is then metabolized in the TCA (tricarboxylic acid) cycle in order to release more energy, highlighting the high connectivity and dependency of all metabolic pathways and metabolites. Information on the metabolic pathways for several organisms can be accessed in various public databases such as the Kyoto Encyclopaedia of Genes and Genomes (KEGG; `http://www.genome.jp/kegg/`), which are obtained from sequencing the organism's genome.

Term	**Definition**
non-targeted metabolomics	analysis of as many metabolites as possible in an unbiased manner
targeted metabolomics	limited to a pre-defined set of molecules, a hypothesis on the sample is required, often referred to as metabolite profiling
metabolite fingerprinting	"fingerprint" metabolite patterns, without prior knowledge comparative analysis of two or more phenotypes
metabolite footprinting	exometabolomics, analysis of all excreted metabolites

Table 1.1 Terms and definitions of metabolomics analysis according to Oliver Fiehn [17–19]

In principle, a distinction is made between targeted and non-targeted metabolomics (Table 1.1) [17, 19]. Non-targeted metabolomics aims to detect as many metabolites as possible, whereas targeted metabolomics (often called metabolite profiling) is limited to selected compounds of interest, already requiring a hypothesis of the samples. Even more, the term metabolic fingerprinting exists, which is a non-targeted approach to "fingerprint" metabolite patterns, without prior knowledge. Metabolic fingerprinting is mostly used in the context of the comparison of two phenotypes such as normal vs. disease [18]. Metabolic footprinting, or exometabolomics, describes the analysis of all metabolites that are excreted from the organism [20].
Moreover, the term lipidomics exist, which is a sub-branch of metabolomics and is defined as the analysis of all lipids in a biological system. The differentiation between metabolites and lipids merges into one another and is not always straightforward. However, the main

criterion for lipids is their hydrophobicity and solubility in organic solvents.
Lipids are important biomolecules since they are involved in different biological processes, such as lipid-protein interactions, energy storage, signaling pathways, and serve as scaffold for the cell membrane [21]. Moreover, lipids are known to play a crucial role in diseases including Alzheimer's disease, diabetes, and cancer [22–24].
The lipidome is very complex and contains many isomeric structures. To date 47825 unique lipid structures have been curated in the LIPID MAPS® Structure Database (LMSD, `www.lipidmaps.org/data/structure/`) [25]. Based on their chemical structures (e.g. the position of bonds and chains, branching and the addition of polar head groups), the Lipid Metabolites and Pathways Strategy (LIPID MAPS®) consortium has proposed a lipid classification system that allows lipids to be grouped into different categories, classes and subclasses [26]. According to an ontology, the structures of lipids within the same (sub)class are very similar, often differing only in the chain length or the number and position of double bonds, making classification straightforward. Due to their different physicochemical properties, lipids are often considered separately from metabolites and require different analytical approaches.

1.1.2 Challenges in metabolomics

The major bottleneck in non-targeted metabolomics is metabolite identification. Compared to other -omics technologies, metabolomics is challenging because of increased chemical diversity of small molecules (Figure 1.1). Metabolites vary greatly in their physicochemical properties, such as molecular weight, structure, pKa, stability, polarity, or charge. This makes it impossible to analyze the complete metabolome with a single analytical platform. Even more and in contrast to DNA and RNA analysis, there is no amplification of metabolites possible (same applies to protein analysis) [27]. This is especially challenging for the detection of low abundant metabolites.
Moreover, identification relies on comparing experimental data, such as mass spectra and retention times, with reference databases of known metabolites. These databases are not exhaustive and are biased towards well-studied compounds, leaving numerous novel or less-studied metabolites without reliable matches. Also a lack authentic reference standards for specific metabolites is an issue. Without these standards, it becomes challenging to confidently identify and quantify compounds.
Other issues in metabolite identification are related to analyte instability, such as degradation during extraction, storage and processing, high chemical reactivity, enzymatic degradation, or fragmentation of molecules in the ionization source of mass spectrometry analysis. In addition, biological samples are complex mixtures containing a wide range of compounds. Co-elution of metabolites, ion suppression, and other matrix effects can interfere with the accurate detection and identification of metabolites, leading to reduced sensitivity and reli-

ability. The high variability of metabolite concentration levels is also critical to the correct detection and quantification of all metabolites.

Metabolite nomenclature can be inconsistent, leading to challenges in mapping compounds with various names. This is exemplified by 2180 human metabolites having over 27700 synonyms [13]. To address this, uniform molecular descriptors like SMILES, InChI, or InChI Keys are employed, aiding in organizing and comparing molecules [28]. Tools like ClassyFire group compounds into chemical classes automatically [29].
Additionally, metabolite reporting can pose issues. Detailed structural reporting might overlook critical distinctions; for instance, amino acids are typically reported in their L-form due to human prevalence. However, differentiating D- and L-enantiomers requires specialized chiral analytics.

1.1.3 Applications of metabolomics

Metabolomics technologies have been used since 1998 with the first application on *Saccharomyces cerevisiae* cells by Fourier-transform infrared spectroscopy, where metabolite profiles of two phenotypes could be clearly separated in principal-component analysis (PCA) [30]. Vibrational spectroscopy, such as Raman or infrared spectroscopy can be used for metabolite fingerprinting, but with limited sensitivity and identification [31]. Therefore, nuclear magnetic resonance (NMR) spectroscopy and mass spectrometry (MS) based methods are more commonly used in metabolomics today. Usually, a separation technique is hyphenated to MS such as gas chromatography (GC), capillary electrophoresis (CE) or liquid chromatography (LC) and applied in different fields such as plant science, health and medicine, and also to study metabolism in model organisms.

Metabolomics in plant science

In plant science, metabolomics is used to study transgenic plants or environmental influences [32–35], where mostly GC-MS, LC-MS, or DI-MS was used. For example, Fiehn *et al.* compared four *Arabidopsis* genotypes and found distinct metabolic profiles in each. Using GC-MS analysis, 326 compounds have been quantified, such as fatty acids, sterols, amino acids, sugars, hydroxy acids, or organic monophosphates [33]. In another study, LC-MS analysis of *Arabidopsis* yield in 2000 mass features, which of, many were found to originate from secondary metabolites [36]. More than 200.000 secondary metabolites are found in plants [15, 37]. They are involved in pathogen defenses, deterioration of herbivores, protection against biotic and abiotic stress, and serve as signaling molecules. Primary metabolites on the other hand are found in all plants providing crucial metabolism for life and are necessary for development [38]. The great size of the plant metabolome, understanding secondary metabolism or environment creates the necessity to apply metabolomics technologies in plants.

Metabolomics in health and medicine

Small molecules and their analysis are of great importance in human medicine. For example, the majority of all drugs are small molecules < 1500 Da, such as the prominent cyclooxygenase (COX) inhibitors Ibuprofen, Aspirin, or Diclofenac. Therefore, a lot more potential small metabolites could serve as drugs but have not yet been identified. Metabolomics has proven to be a useful tool in drug development, in order to find biomarkers for drug efficacy [39].
Even more, metabolites are already used to indicate different diseases. More than 80% diagnostic tests in medicine rely on the measurements of small molecules. For example, renal failure is diagnosed by increased urea and creatinine levels in plasma [40], or diabetes can be indicated by high blood glucose levels. There are already 5498 metabolite-disease associations documented in HMDB [13]. This makes metabolomics an important tool to find novel biomarkers of diseases and was already successful to find markers of Parkinson's disease [41], Alzheimer's disease [42], heart diseases [43, 44], or cancer [45–47].

Metabolomics in model organisms

Model organisms are used to understand biological processes in other organisms that are more complex, such as humans, and therefore play an important role to understand human health and disease. Models are usually easy to use, they are cheap, accessible, can be manipulated in a controlled laboratory environment, and often show fast development. Model organisms can be vertebrates and invertebrates, plants, and microbes [48]. The most popular model organisms are mice, *Arabidopsis*, *E. coli*, *Saccharomyces cerevisiae*, *Drosophila melanogaster* and *Caenorhabditis elegans* and have been applied in different fields.
Metabolomics has proven to be a powerful tool to study biological mechanisms in various model organisms [49]. For example, metabolomics studies in zebrafish help to understand the metabolism of early vertebrate development [50]. Metabolomics was also applied in the fruit fly *D. melanogaster* to investigate inbreeding, heat stress or hypoxia [51–53]. Also, transgenic mice were used in metabolomics to understand toxicity, drugs, and cancer and find relevant biomarkers that can be applied to human [54].
Another example is the nematode *Caenorhabditis elegans*, which will be introduced in the following. The small roundworm has been introduced 50 years ago as a versatile model organism and is today used in thousands of laboratories in different fields, such as metabolomics.

1.2 Quo Vadis *Caenorhabditis elegans* Metabolomics - A Review of Current Methods and Applications to Explore Metabolism in the Nematode

Metabolomics and lipidomics recently gained interest in the model organism Caenorhabditis elegans (C. elegans). The fast development, easy cultivation and existing forward and reverse genetic tools make the small nematode an ideal organism for metabolic investigations in development, aging, different disease models, infection, or toxicology research. The conducted type of analysis is strongly depending on the biological question and requires different analytical approaches. Metabolomic analyses in C. elegans have been performed using nuclear magnetic resonance (NMR) spectroscopy, direct infusion mass spectrometry (DI-MS), gas-chromatography mass spectrometry (GC-MS) and liquid chromatography mass spectrometry (LC-MS) or combinations of them. In this review we provide general information on the employed techniques and their advantages and disadvantages in regard to C. elegans metabolomics. Additionally, we reviewed different fields of application, e.g., longevity, starvation, aging, development or metabolism of secondary metabolites such as ascarosides or maradolipids. We also summarised applied bioinformatic tools that recently have been used for the evaluation of metabolomics or lipidomics data from C. elegans. Lastly, we curated metabolites and lipids from the reviewed literature, enabling a prototypic collection which serves as basis for a future C. elegans specific metabolome database.

This review chapter has been published under CC BY 4.0 license as Salzer, L. & Witting, M. (2021).*Quo Vadis Caenorhabditis elegans Metabolomics-A Review of Current Methods and Applications to Explore Metabolism in the Nematode.* Metabolites, 11(5), 284. https://doi.org/10.3390/metabo11050284

Candidate's contributions: L.S. performed the the literature research, curation and analysis of data. L.S. prepared the figures and wrote and revised the review.

1.2.1 Introduction

Caenorhabditis elegans - A Versatile Model Organism

The small nematode *Caenorhabditis elegans* (*C. elegans*) is a frequently used model organism in biomedical research. It was introduced in 1973 by Sidney Brenner as promising model for developmental genetics [55]. The nematode is of microscopic nature with adults having an approximate length of 1 mm and a diameter of 80 µm. *C. elegans* can be found all over the world in moderate climates, mostly in garden soil and compost, but it has been isolated also from snails, woodlice, and other invertebrates [56]. It shows a bacteriovorus feeding on bacteria growing on rotting biomaterial.
The main benefits and reasons of using *C. elegans* as a (genetic) model are its simplicity, and speed of cultivation. Under laboratory conditions it has a short generation time (3 days from fertilized egg to sexually mature adult), and a short lifespan of about 2-3 weeks at 20 °C. Cultivation in the laboratory takes place on a solid support such as agar plates or in liquid by feeding with bacteria, normally with *Echerichia coli* (*E. coli*), which is cheap, easy and can be scaled up to highthroughput screening approaches [57]. *C. elegans* has two sexes and predominantly exists as self-fertilizing hermaphrodites, typically only 0.1% appear to be males. Hermaphroditic reproduction also leads to low genetic variation within a culture. Another characteristic is the invariant development, leading to exactly the same number of somatic cells in each individual (959 and 1031 in hermaphrodites and males respectively). *C. elegans* develops through four larval stages (L1–4) into reproductive adults after hatching. When no food is available or populations are overcrowded, normal development is arrested and L1 larvae enter an alternative life cycle and become so-called "dauer" larvae (“dauer” german for enduring). The cuticles of dauer larvae are highly stable and protect the worm against environmental factors, such as desiccation. In dauer-state, worms can survive several months without food. As soon as food is available again normal development continues through the L4 stage into reproductive adults [58].
About 60–80% of the nematode genes are homologues to human genes [59]. Therefore, investigations on *C. elegans* can have high relevance to studies of human health and disease. Different forward and reverse genetic tools are available in *C. elegans*. RNA interference (RNAi) has become one of the most common used methods for systematic gene inhibition in *C. elegans* [60–62]. RNAi inhibits gene activity by introducing double-stranded RNA (dsRNA) with a sequence specific for the target gene. This leads to degradation of the homologous messenger RNA (mRNA) [59, 63, 64]. Also, clustered regularly interspersed short palindromic repeats (CRISPR)–Cas9 has been applied in *C. elegans* for genome engineering, which has been reviewed elsewhere [65–72].
One of the most recent additions to the *C. elegans* toolbox are metabolomics and lipidomics, enabling new and deeper investigations into the metabolism of the nematode. The combination of a genetically tractable model organism such as *C. elegans* with the functional readout of metabolomics and/or lipidomics holds great promise to advance our knowledge on metabolism and metabolic regulation.

Metabolomics and Lipidomics - Systematic Measurements of Metabolites and Lipids

Metabolomics and lipidomics are defined as the systematic measurement and quantification of metabolites or lipids in a given system. They probably represent the most complex of all “-omics” approaches due to the large chemical complexity and concentration range underlying the metabolome and lipidome. Metabolism is of great interest, since it is closest to the observed phenotype and typically is one of the first things to react upon a stimulus. Metabolite concentrations are directly linked to biochemical activity, and many biological processes depend on metabolism [73, 74]. Metabolomics

also enables the investigation of interactions of an organism with its environment or between organisms. The goal of metabolomics is to describe all metabolites in a biological system in a defined state [75].
One can distinguish between targeted and non-targeted metabolomics. In targeted approaches, a predefined set of metabolites is studied. Since these targets are known beforehand, methods are optimized on existing chemical reference standards. Therefore, their analysis is highly precise, accurate and often allows absolute quantification. Metabolites investigated in targeted analysis typically belong to a single class of compounds or a few related compound classes or biochemical pathways, whereby the use of a single analytical technique might be sufficient.
However, targeted metabolomics requires a specific hypothesis on changes in the metabolism, which needs to be tested. In contrast, non-targeted metabolomics characterizes a large number of metabolites by their simultaneous measurement without prior selection and free of any hypothesis. Therefore, the data of non-targeted metabolomic approaches is much more complex and often requires more elaborate statistical and bioinformatic evaluation. A major issue in non-targeted metabolomics in general is the lack of comprehensive measurements of the whole metabolome by a single technique due to its large complexity. At the current state it is not possible to measure the whole metabolome of an organism in a single experiment. Beside several known compounds a large part of the metabolome has not even been identified yet ("metabolic dark matter") [76, 77]. To deal with the large chemical complexity, complementary analytical platforms are required. Metabolomics typically employs high end analytical methods, such as mass spectrometry (MS)—either direct infusion (DI) or hyphenated to different types of separation (gas chromatography (GC), liquid chromatography (LC), capillary electrophoresis (CE) or ion mobility (IM))—or nuclear magnetic resonance spectroscopy (NMR).
C. elegans possess a complex metabolome and several new molecules are described on a routine basis, e.g., novel ascarosides, dafachronic acids or others [78–84]. Different types of analytical methods have been used to analyze the *C. elegans* metabolome and lipidome.

1.2.2 Analytical Methods for *C. elegans* Metabolomics

The following paragraphs give an overview on different extraction and analysis methods that have been used to analyze the *C. elegans* metabolome and lipidome. This overview shall help scientists new to the field to search for their most fitting combination of extraction and measurement methods for their specific application.

Extraction Methods

The first step towards analysis of the metabolome and lipidome is the extraction of compounds of interest from the nematodes. In case of non-targeted analysis extraction of as many substances as possible is required. It has to be mentioned that a truly non-targeted extraction does not exist, since different solvent systems will always favor specific metabolites and metabolite classes over others. Furthermore, extraction depends on the biological question and on the analytical method that shall be used for analysis. In order to analyze the metabolome as accurately as possible, it is also necessary to avoid degradation of the metabolites during extraction. Therefore, extractions shall be carried out at the lowest possible temperatures. Extraction of metabolites from *C. elegans* is challenging since the nematode possess a hard cuticle, which first needs to be broken before extraction. However, the type of tissue disruption seems to be a less critical step in sample preparation as shown by Geier *et al.* who

investigated different disruption techniques in *C. elegans* including manual grinding, homogenization and different grinding media in tissue mills [85]. However, care must be taken to avoid excessive heating of samples e.g., when using bead beating systems, since it may cause sample degradation. In contrast, the extraction procedure itself is much more important. Different extraction methods have been applied to *C. elegans*. Most commonly (cold) methanol, ethanol, and chloroform, either individually or in combination (in diverse rations) have been applied [83, 86–89]. For lipid extractions on the other hand, protocols from Folch and Bligh & Dyer are often used [90, 91]. Thereby, lipids are extracted and separated into the hydrophobic chloroform phase and other compounds into the methanol/water phase. Since chloroform is suspected to cause cancer its reduction to minimal amounts or complete replacement is desirable. Matyash *et al.* developed an extraction based on methyl-tert-butyl-ether (MTBE) using also *C. elegans* as a model system. Extraction yields were similar to other lipid extraction methods. A further advantage is that MTBE forms the upper phase in a two-phase extraction with water, making recovery of lipids easier and even automatable[92].
Analysis of specific metabolites or lipids often requires their enrichment using further sample preparation steps. Thin layer chromatography (TLC) is an approach used either in analytical purposes to determine lipid composition, or as additional step for preparation of specific substance classes. In TLC major lipid classes as glycerophospholipids, ceramides, glycosphingolipids, fatty acids, or sterols can be separated and quantified. Chromatographic resolution may be enhanced through its extension to two-dimensional TLC using a second solvent system [86, 93].
Another challenge in metabolomics is the detection of low abundant compounds, since the detection limit of the analysis is restricted depending on the sensitivity of the employed technique. Metabolites are enriched, e.g., by using solid phase extractions (SPE) or preparative chromatography [76, 79, 83]. For example, Hänel *et al.* used fractionation based on aminopropyl and weak anion exchange SPE of lipid extracts in order to analyze sphingolipids from *C. elegans* [83].

Nuclear Magnetic Resonance (NMR)

Nuclear magnetic resonance (NMR) spectroscopy is based on the interaction between the magnetic moment of atomic nuclei and a magnetic field. Only nuclei that have a nuclear spin do have a magnetic moment and thus are NMR active. In an NMR experiment the free induction decay (FID) is obtained, which is the magnetic resonance response in a unit of time. Via fourier-transformation (FT) the NMR spectra are generated, consisting of signal intensities of NMR active compounds depending on their resonance frequency. The most important nuclei at the analysis of biomolecules are ^{1}H, ^{13}C, ^{15}N, ^{19}F and ^{31}P.
In metabolomics mostly one dimensional ^{1}H NMR spectroscopy is used, since it is fast and very convenient for universal detection of organic compounds. NMR is quantitative and nondestructive with minimal sample preparation and interference, resulting thus to lowest analytical variation compared to other techniques [94]. Its high precision allows the detection of even small changes in metabolite abundances. NMR delivers qualitative (structure) and quantitative information in a single run. Its strengths are particularly evident in substances that are difficult to ionize in MS or require derivatization. However, a major disadvantage of NMR compared to MS is the lower sensitivity. Therefore, only few tens of high abundant compounds are covered by 1D-^{1}H NMR [95].
In order to increase metabolite coverage in non-targeted metabolomics, 1D-^{1}H NMR is often combined with other analytical platforms such as DI-MS, GC-FID, GC-MS, or LC-MS [87, 96–98].
In *C. elegans* metabolomics 1D-^{1}H NMR has been used to investigate changes is central carbon metabo-

lites such as amino acids, organic acids, choline, sugars, nucleotides or cofactors [87, 98–101]. A limitation in 1D-^{1}H NMR is overlapping signals, resulting in ambiguity in metabolite identification and quantification. Extension to 2D-NMR is often used to confirm elucidated structures and thus improve identification rates. 2D-NMR approaches such as ^{1}H-^{1}H correlated spectroscopy (COSY), ^{1}H-^{1}H total correlation spectroscopy (TOCSY), and ^{1}H-^{13}C heteronuclear single quantum coherence (HSQC) have been applied in *C. elegans* metabolomics [84, 85, 102, 103].
Another approach to enhance NMR structural resolution is performing ^{13}C-Heteronuclear Multidimensional NMR (HMN) instead of 1D-^{1}H NMR. Due to the low natural abundance of ^{13}C, ^{13}C HMN suffers from low sensitivity. Using in vivo labeling with ^{13}C can increase sensitivity of ^{13}C HMN. ^{13}C-labeling of *C. elegans* is possible by first feeding ^{13}C-glucose to *E. coli*, followed by feeding labeled bacteria to the nematodes. Such labeling and ^{13}C HMN was applied to metabotyping of *C. elegans* mutants by An *et al.* [104]. Labeling increased sensitivity two orders of magnitude compared to unlabeled samples. It was also possible to perform high resolution two-dimensional HSQC and three-dimensional HCCH-TOCSY NMR experiments [105]. This enabled the detection of much more metabolites than in one-dimensional NMR metabolomics. The major drawback of performing 2D-^{13}C HSQC experiments using ^{13}C-labeled metabolites is additional structures arising from ^{13}C/^{13}C couplings, which are non-existent at natural ^{13}C abundance. These additional structures can reduce the possible sensitivity gain of ^{13}C-labeling and increase chances of peak overlap. Therefore, Geier *et al.* investigated different HSQC pulse programs for fully ^{13}C-labeled tissue extracts from *C. elegans*. They found at constant time HSQC (ct-HSQC) improved peak shape and better peak detection of metabolites, which enabled matching of 300 records from Human Metabolome Database (HMDB) [13, 14, 106–109].
Another promising approach developed using *C. elegans* is differential analysis by 2D-NMR spectroscopy (DANS). DANS is a method based on overlaying and subtracting 2D-NMR spectra from two different conditions. Therefore, detailed structural information of differentially regulated compounds can be directly extracted, even for minor components in complex matrices [80, 110].
To circumvent extraction, nematodes can be directly analyzed using high-resolution magic-angle spinning (HR-MAS) NMR spectroscopy [111, 112]. At HR-MAS the sample is spinned at a distinct angle to the magnetic field to overcome magnetic field heterogeneities within solid samples that are responsible for broadening of resonance lines. The main advantage over other platforms such as in-solution NMR, or MS is that no metabolite extraction is required and *C. elegans* can be directly filled into a HR-MAS NMR rotor. Because of the general insensitivity of NMR, around 1000 nematodes are required for a standard HR-MAS analysis. Wong *et al.* introduced high-resolution magic-angle coil spinning (HR-MACS) combined with a ^{1}H NMR microprobe for the metabolic phenotyping of low number of *C. elegans* [113]. Sensitivity is improved by miniaturizing the rf receiver coil (µcoil). Moreover, an inductively coupled µcoil resonator is spinning together with the sample, which suppresses magnetic susceptibility broadening from both the µcoil and the sample. HR-MACS NMR allowed metabolic profiling from small numbers of nematodes ranging from 10 to 100. It was even possible to acquire a NMR spectrum of a single worm on a 1000 MHz (23.5 T) spectrometer [114].

Mass Spectrometry (MS)

Mass Spectrometry (MS) is beside NMR one of the major analytical technologies used in metabolomics. A mass spectrometer is an instrument which generates ions and separates them according to their respective mass-to-charge ratio (m/z). The ions are separated due to different physical approaches, de-

pending on the type of mass spectrometer being used. In MS *m/z* ratios of ions are measured together with their corresponding intensities, combining both qualitative (*m/z*) and quantitative (intensity) information. However, metabolites with same molecular formula have the same mass and therefore the same *m/z* value (isomers, e.g., Leucine and Isoleucine, $C_6H_{13}NO_2$). Likewise, several sum formulae might have very similar but not identical masses and *m/z* values (isobars, e.g., $[M+Na]^+$ adduct of PC(34:1) and $[M+H]^+$ adduct of PC(36:4)). Such compounds may be further analyzed using tandem MS (sometimes referred to as MS^2 or MS/MS) experiments. Here, the ions of interest are fragmented, mostly using collision induced dissociation (CID), and fragment ion *m/z* values are analyzed yielding information on potential substructures. In non-targeted metabolomics, mostly data dependent acquisition (DDA) is used to generate and acquire MS/MS data. In DDA the most intense ions, meeting certain user defined thresholds, are selected and fragmented. Another approach that is increasingly used in non-targeted metabolomics is data independent acquisition (DIA), such as Sequential Window Acquisition of all Theoretical fragment ion spectra (SWATH) MS or All ions. Benefit of DIA is that every precursor ion will be fragmented, allowing retrospective data analysis without the need of reacquisition.

MS is either employed without or with prior chromatographic or electrophoretic separation. The employed ionization source depends on the upfront sample introduction system, e.g., electron ionization (EI) or chemical ionization (CI) for GC, electrospray ionization (ESI) and atmospheric-pressure chemical ionization (APCI) for direct infusion, LC or CE.

In direct infusion (DI) MS the sample is directly injected into the ionization source without prior separation. In order to differentiate between isobaric structures, the use of high-resolution (HR) MS, such as time-of-flight (TOF) MS, fourier-transform ion cyclotron resonance (FT-ICR) MS, or Orbitrap MS, is a prerequisite at DI-MS.

Separation is used to reduce ion suppression and separate isomeric and isobaric structures, which cannot be separated by MS or MS/MS alone. Furthermore, additional information about physicochemical properties of detected metabolites based on the respective separation characteristics is provided.

GC has been adopted early for the analysis of endogenous metabolites, even before the term metabolomics existed, e.g., Pauling *et al.* used GC for the analysis of urine and breath [115]. GC-MS offers highly efficient and highly reproducible separation of volatile metabolites or metabolites that can be made volatile by derivatization. GC is typically combined with EI, which produces highly reproducible fragmentation rich spectra. Therefore, GC-MS analysis allows ready dereplication of known and previously measured metabolites with many spectral libraries available. However, if unknown substances are analyzed CI might be preferred, since the parent ion is mostly preserved, as for instance shown by Jaeger *et al.*, who were using GC-APCI-MS for metabotyping in *C. elegans* [116].

If substances are not volatile or can be made volatile by derivatization, LC-MS might be used for analysis. While selection of GC stationary phases is rather limited, LC offers a large selection of different column chemistries to optimize metabolite separations. In LC metabolites are retained depending on the stationary and mobile phase and their physicochemical properties. Today, Ultra-high-performance liquid chromatography (UPLC), using particle sizes of below 2 µm, is widely used in non-targeted metabolomics, due to fast and efficient separation [117].

Despite the large number of possible separation chemistry available, reversed-phase (RP) columns are used most. They allow the separation of mid- to non-polar metabolites and lipids. However, very polar metabolites are not retained on RPLC columns. Therefore, hydrophilic interaction liquid chromatography (HILIC) is frequently used. HILIC columns are either silica or derivatized silica columns

[89, 118]. However, polar, charged metabolites often show bad peak shape in LC. Capillary zone electrophoresis (CZE), normally referred to as CE, coupled to MS is another rising powerful analytical technique in metabolomics [119, 120]. In CE charged molecules migrate in a liquid background electrolyte along an electrical field. Separation occurs due to different velocities of the metabolites, depending on their mobility and the electric field strength. CE requires only a small sample amount, only a few nl per injection. It is well suited for highly polar and charged metabolites but has not been applied to *C. elegans* so far. Its usability on metabolomics in general has already been reviewed elsewhere [120].
Peak capacity, spectral clarity and fragmentation specificity of metabolomics LC-MS data can be further increased by adding ion mobility (IM) separation as a second separation dimension post ionization. Moreover, IM provides an additional parameter that may be useful for metabolite identification - the collision cross-section (CSS) value. At IM separation takes place in the gas phase due to different ion mobilities, that depend on charge, shape and size of the metabolites. Its applicability on metabolomics and lipidomics in general have already been confirmed [121].
Most of the mentioned analytical techniques have also been used for the analysis of the *C. elegans* metabolome. Table 1.2. summarizes the analytical methods used in *C. elegans* metabolomics and lipidomics, including their advantages and disadvantages. Mass spectrometry has already been demonstrated to be an effective tool for metabolomic profiling in *C. elegans*. For instance, DI-MS is frequently used for the analysis of the *C. elegans* lipidome [84, 122]. Identification of lipids is thereby accomplished on their accurate mass and/or MS/MS characterization. GC-MS has been used for metabolomics in *C. elegans* to identify fatty acids, amino acids, or organic acids [86, 87, 96, 103, 110, 117, 123–126]. Not only coupled to MS, but also GC using a flame ionization detector (FID) has been used as a complementary analytical platform in *C. elegans* to determine the fatty acid composition [86, 96, 98]. UPLC has also frequently been applied to separate lipids and metabolites of *C. elegans* prior to mass spectrometric analysis [85, 97, 125, 127]. RPLC columns are used for lipids and mid- to non-polar metabolites. Lipidomic studies were mostly accomplished by C18 [86, 93, 96, 127–129]but also C8 stationary phases [86]. For polar metabolites on the other hand HILIC columns were predominantly used [89, 118]. Moreover, van Assche *et al.* used aqueous normal phase (ANP) columns for the separation of polar metabolites in *C. elegans* [130]. Besides various column selectivity, separation is also influenced by additives, as shown by Wang *et al.* who used ion-pairing reversed phase in negative ion ESI mode, where tributylamine was added to the aqueous mobile phase [126]. Analytical platforms have been used individually and in combination used to enhance the coverage of the metabolome.

Bioinformatic Tools for the Analysis of the *C. elegans* Metabolome / Lipidome

Beside the actual chemical analysis of the metabolome and/or lipidome, extensive bioinformatic and statistical analysis is required to interpret large data sets produced using the aforementioned analytical approaches. Evaluations on all types of data, either NMR or MS based, were most commonly accomplished using multivariate data analysis methods. Multivariate methods can identify relationship patterns between metabolites by clustering and correlation analysis. Widely used chemometric methods such as principal component analysis (PCA), hierarchical cluster analysis (HCA), or orthogonal projection to latent structure with discriminant analysis (OPLS-DA) have been applied [139, 152, 153]. Moreover, metabolite and spectral databases have become indispensable for the analysis of metabolomics data and the identification of metabolites. Through searching against in-house and public databases, metabolites can be annotated and identified at different levels according to the Metabolomics Standard

	Method	Avantage	Disadvantage	Remarks	References
NMR	^{1}H NMR	Quantitative, non-destructive, minimal sample preparation	Only aqueous, high abundant metabolites	Overlapping signals result in ambiguity of metabolite identification	[74, 75, 84, 85, 87, 94–103, 131]
	DANS	Simple linking of metabolites with biological function	Only high abundant metabolites		[80, 110]
	HR-MAS	No metabolite extraction needed, intact worms	Large populations of *C. elegans* needed, only high abundant metabolites		[111, 112]
	HR-MACS + ^{1}H NMR microprobe	Small number of worms can be analyzed	Only high abundant metabolites		[113, 114]
	^{13}C HMN + ^{13}C-labeling	Much more metabolites detected than in ^{1}H-1D NMR	Reduced sensitivity due to ^{13}C-^{13}C coupling, proper pulse program required (ct-HSQC)	Higher spectral range than at ^{1}H NMR	[104, 105]
MS	DI-MS	Fast and high throughput	Isomers cannot be differentiated	Frequently used in lipidomics	[84, 122, 132–137]
	GC-MS	High resolution, Absolute quantification possible	Derivatization necessary	Analysis of lipids and metabolites possible	[86–88, 96, 103, 110, 116, 122–126, 138–141]
	LC-MS	Absolute quantification possible, separation of isomers, a lot of metabolites may be identified	Lower resolution than GC	Selectivity depending on the stationary phase	[76, 78, 85–88, 93, 95–97, 118, 125–130, 133, 142–151]

Table 1.2 Overview on analytical methods used in *C. elegans* metabolomics/ lipidomics and their advantages, disadvantages and corresponding references.

Initiative [154]. Most commonly HMDB (http://www.hmdb.ca) [107], METLIN (http://metlin.scripps.edu/index.php) [155], KEGG
(http://www.genome.jp/kegg/), BMRB (http://www.bmrb.wisc.edu) [156], or LipidMaps
(http://www.lipidmaps.org) [157] are used for this purpose. However, they are only partially useful for *C. elegans* metabolomics. At the current stage, no *C. elegans* metabolome specific database exists. The Small Molecule Identifier Database (SMID DB, http://www.smiddb.org/) is storing information on *C. elegans* and related species-specific secondary metabolites, such as ascarosides and recently added MS^2 spectra and retention times [76]. Still, the exact number of metabolites in the *C. elegans* metabolome remains elusive.

Genome-scale metabolic models (GSMs) are aiming to close this knowledge gap and to deliver computer readable mathematical models of the nematode's metabolism. For *C. elegans* four different models have been published. Path2Models convert metabolic pathways from KEGG into SBML models, which were then gapfilled [158]. The model reconstructed by Gebauer *et al.* contains 1921 reactions and 2357 metabolic compounds of *C. elegans*[159]. Moreover, Yilmaz *et al.* created a model including 1985 reactions and 1718 compounds based on SACURE curative approach [160]. The fourth model CeCon was developed with the PathwayTools software [131]. At last, the WormJam model has been created with the existing metabolic reconstructions of *C. elegans* have been merged into a single consensus model. WormJam is freely available and continuously improved to cover different aspects of *C. elegans* metabolism [161–163]. Furthermore, first steps towards the development of tissue and cell type specific models of *C. elegans* have been made [68]. Since WormJam is curated by using only metabolites relevant to *C. elegans*, it may be useful to accelerate annotation of metabolomics data and avoid over annotation.

Additionally, pathway and network analysis are increasingly used to evaluate metabolomics experiments. Pathways found in *C. elegans* and other organisms were documented and visualized for instance in the PathBank database (http://www.pathbank.org) [109]. Pathway analysis uses prior biological knowledge to analyze metabolic patterns. Network analysis on the other hand uses correlation of metabolites to build metabolic networks that outline relationships between the metabolites. The basis of the network is therefore the experimental data itself [164]. Krumsiek *et al.* developed gaus-

sian graphical model (GGM), which is based on partial correlation between metabolites, and has also been applied on *C. elegans* metabolomics data [96, 165]. Compared to common correlation networks, GGM's enable differentiation of direct and indirect interactions between metabolites. Lastly, molecular networking based on analysis on MS/MS fragmentation patterns have been used to evaluate *C. elegans* metabolomics data [76, 142].
Flux balance analysis (FBA) is a mathematical method that analyzes the flow of metabolites in a metabolic network and have also been recently used for investigating *C. elegans* metabolism [86, 118, 166]. Using FBA, it is possible to determine pathways that are influenced if concentrations of a certain metabolite changes. FBA can integrate metabolomics data to constraint reaction boundaries, e.g., Yilmaz *et al.* Used FBA to analyze metabolites from tissue specific pathways in *C. elegans* [167].

1.2.3 Applications of *C. elegans* Metabolomics

The previously mentioned metabolomics techniques are used to identify differentially regulated metabolites between different states that are involved in relevant biological pathways or that are specific for the respective state (biomarker). Frequently, studies on development, ageing, longevity, pheromones, lipid metabolism and regulation, effect of genetic perturbations, exposure to different chemicals and effect of different diets were investigated in the model organism *C. elegans* using metabolomics and lipidomics.

Development, Ageing and Longevity

Different studies investigated the relationship between ageing, longevity and metabolism in *C. elegans*. There are several mutants of *C. elegans* that increase its lifespan [168]. Frequently, non-targeted metabolomics has been used to analyze how the metabolism of those longlived worms changed in comparison to controls with normal lifespan.
For instance knockout of genes involved in the Insulin/Insulin-Like Signaling (IIS) pathway tend to increase longevity [168]. One of the most studied mutants in *C. elegans* is *daf-2*. *daf-2* encodes for the orthologue of the insulin/insulin-like growth factor (IGF) receptor.
An altered amino acid profile and changes in the carbohydrate metabolism had been found to be a key feature in *daf-2* mutants. A common trend for different alleles of *daf-2* and *daf-28* (which disrupts a *daf-2* ligand) is the increase of branched chain amino acids leucine, isoleucine and valine [96, 133]. For example, Fuchs *et al.* used NMR based metabolomics to study three different alleles of *daf-2* (*e1370, m21* and *m596*) as well as *daf-28*(*sa191*) and *ife-2*(*ok306*) [100]. Martin *et al.* similarly compared *daf-2*(*e1370*) and *daf-2*(*e1370*); *daf-16* (*m26*) mutants [99]. However, BCAA levels drop at later stages as shown by Davies *et al.* who were performing a time course analysis of metabolites across age using NMR. Another interesting trend was seen for trehalose. This metabolite increases strongly throughout aging in *daf-2*(*m41*) mutants. Trehalose has been previously shown to increase the lifespan of wildtype worms but did not further extend the lifespan of *daf-2* mutant worms [74].
The lifespan of *daf-2* mutants is even further enhanced by additional knockouts, as for example *pept-1(lg601)*, which encodes for an intestinal di- and tripeptide transporter [99] or prohibitin deficient *daf-2*(*e1370*) mutants shown by Lourenço *et al.* However, prohibitin deficiency shortens the lifespan of wild type nematodes. Fatty acid composition, and amino acid and carbohydrate metabolism, analyzed by NMR, was more deeply affected by prohibitin depletion in wild type nematodes compared to *daf-2* mutants. GC-FID analyses showed that prohibitin depletion of *daf-2* results in changes of

polyunsaturated fatty acid contents, which has also found to be a key feature of longevity [98]. Also, lipid analysis showed an increase in triacylglycerols, especially containing branched and monounsaturated fatty acids in *daf-2* mutants [74].
Moreover, Prasain *et al.* performed lipidomic analysis of *daf-1* and *daf-2* mutants using a workflow called MSMSALL. They identified changes in several phospholipid classes as well as di- and triacylglyerols. *daf-2* mutants are described to show a fat phenotype with higher number of lipid droplets. This is mirrored by the lipidomic analysis showing that *daf-2*(*e1370*) have higher levels of triacylglycerols compared to wildtype [169].
Metabolomics is able to produce a static snapshot of the current metabolic state. However, in most cases it remains elusive how this state was reached in regard to which pathway was active. The use of isotopic tracers can help to capture the dynamic nature of metabolism. Tracer analysis in *C. elegans* represents a complicated task since no axenic medium is available and always cometabolism of tracers between *E. coli* and *C. elegans* would be determined. Perez and van Gilst developed an interesting approach that allows to follow lipogenesis in *C. elegans* based on the feeding of a mixture of isotopically non-labeled or labeled *E. coli*. The *E. coli* food source was grown in either non- or fully-labeled growth medium. *C. elegans* was then fed with a 50/50 mixture of non-labeled and labeled *E. coli*. Fatty acids from *C. elegans* have been analyzed with GC-MS. Fatty acids directly derived from *E. coli* are either 100% non-labeled or labeled, while all fatty acids that are produced by *C. elegans* show a mixed labeling pattern. Based on the degree of labeling it can be determined if complete *de novo* synthesis or elongation has been performed. This method was applied to different alleles of *daf-2*: m577, e1368, m596, e1371, e1370 and m41. Interestingly, not all alleles showed the same extent in changes of *de novo* fatty acid synthesis, with highest levels found in *daf-2*(*1370*) and *daf-2*(*m41*) [140].
Recently, a ring trial comparing the WT and the *daf-2*(*e1370*) lipidome has been performed in order to investigate interlaboratory reproducibility. Varying overlap of differentially regulated lipids from different laboratories have been observed. Main cause for these differences appeared to be the fold changes between both variants and the prerequisite of lipids being present in at least 80% per biological group. These ring-trials highlight the variance in lipid profiles that arises due to the small differences in cultivation between different laboratories. However, the most important markers for *daf-2*, as for example higher number of TG's or decrease in PC (20:5/20:5), have been found in the majority of laboratories [170].
Beside *daf-2* different other genetic perturbations lead to increased lifespan. *eat-2* and *slcf-1* mutants, which are often used as genetic models for dietary restriction, also extend lifespan of the nematode. Both mutants show lower phosphocholine levels and therefore correlate with longevity. Additional mutation of *daf-16* suppressed the longevity phenotype and increased phosphocholine and choline kinase levels. Older animals having higher phosphocholine levels seem to have an increased choline kinase (*ckb-2*) expression, probably in adaptation to stress. Moreover, it was found that inactivation of *ckb-2* shortened the lifetime of the nematodes [112].
Shared changes in metabolic pathways between IIS (*daf-2*) and DR (*eat-2*) mutants that promote longevity in *C. elegans* were investigated by Gao *et al.* using metabolomics and transcriptomics. In both mutants an increase of glycerolipid intermediates, purine degradation intermediates and AMP have been observed. On the other hand, amino acid levels and certain fatty acids decreased. Shared longevity signatures in the transcriptome and metabolome were an increase in amino acid metabolism, probably due to lower protein synthesis, and upregulation of purine metabolism [89]. In a targeted GC-MS approach it was shown that also N-acylethanolamines are involved in lifetime extension since

they are decreased at dietary restricted, long lived worms [124].
Dauer larvae live about eight times longer than WT worms, making them also suitable for metabolomic investigations in terms of longevity. The metabolic profiles of dauer larvae were significantly different from WT worms but similar to other longlived mutants. Consistent to other studies the metabolic signature of longevity contains metabolites of distinct pathways as carbohydrate, amino acid, and choline metabolism [100, 133, 171].
It was reported that high levels of SIR-2.1 (sirtuins, which are NAD+ dependent protein deacetylases) extend the lifespan from *C. elegans* up to 50%. Differential ^{13}C HMN metabolomics of *sir-2.1* mutants and WT showed especially differences in branched chained amino acids, triacylglycerol, carnitine and acetyl-CoA, which were elevated in WT. Lactate, alanine, glutamate, fatty acids and AMP on the other hand were upregulated in *sir2.1* mutants. These results point to an increase in glycolysis, nitrogen catabolism and initial lipolysis in *sir2.1* [104].
On the contrary, short-lived worms, such as in mitochondrial mutants (*mev-1(kn1)*), showed upset TCA cycle balance, elevated lactic acid fermentation and increased amino acid catabolism, which has been found by Jaeger *et al.* using GC-EI-MS and GC-APCI-MS [116].
Ablation of germline stem cells (GSCs) leads to infertility and also extends lifespan. The germlineless *glp-1* mutants show an altered lipid metabolism compared to wildtype worms. Moreover, many agerelated metabolites including increased levels of pyrimidine and purine metabolism intermediates and decreased concentration of citric acid cycle metabolites were differentially regulated in *glp-1* mutants. However, *glp-1* mutant worms exhibit the same tendency of metabolic changes as WT during aging, which was measured by both NMR and UPLC-MS. Some age-related metabolites in are dysregulated in the *glp-1* mutants (e.g., valine, GSSG, leucine, malate, serine), although some were not (e.g., cystathionine, glycine, arginine, trehalose). These results indicate that not all aspects of aging are present in long-lived *glp-1* mutants [97]. Future meta-analysis of several aging related studies may pinpoint towards shared principles of metabolism between the different longevity regimes.
Since elevated trehalose have frequently been observed at long-lived mutants, researchers tested if trehalose supplementation increases lifespan of WT worms. Middle aged WT worms showed prolonged lifetime but not early-stage adulthood nematodes. Moreover, at aged worms (10 days adults vs young adults (YA)) levels of glutathione were reduced, while oxidized glutathione was increased. Taurine and hypotaurine, which are antioxidants, were declined with age. Phosphocholine and trehalose have higher concentration with age and decreased levels of purine intermediates were observed. Concentrations of intermediates of pyrimidine and TCA cycle were found to be variable [74].
Pontoizeau *et al.* investigated aging effects on the metabolism by comparing metabolites of YA and adult (7 days adults) worms by means of HR-MAS NMR. Metabolites that increased with age were saturated and unsaturated lipids, glycerophosphocholine, phosphocholines, glutamine and glycine. Metabolites that decreases were a range of amino acids (alanine, arginine, isoleucine, leucine, lysine, phenylalanine, tyrosine, valine, glutamate), acetate, and lactate, glycerol, formate and cystathionine [112].
Age-related metabolites have also been found in adult worms at different time points by Hastings *et al.*. Levels of some amino acids (serine, threonine, leucine, lysine, glutamine/ glutamic acid, methionine, tryptophan, arginine, homoserine, cystathionine), and nucleotides (guanosine, cytidine, uridine, GMP) decreased over time while some other metabolites (betaine, carnitine, leucic acid, pantothenate, kynurenic acid, xanthurenic acid) and degradation products of both nucleotides (hypoxanthine, xanthosine, allantoin) increased. An interesting aspect has been the integration of metabolomics data

with FBA using an approach called MetaboFBA. Differences in fluxes using standard FBA compared to MetaboFBA have been observed, proving the importance of including metabolomics data in silico modeling approaches [118].
Metabolic changes between embryos, larval stages L1–4, and adults have also been investigated. The lowest abundance of most fatty acids was found during larval development and levels increased at the reproductive phase of adults. Most, but not all, amino acids had highest levels at L3 stage and early adulthood and deceased throughout adult phase. Asparagine on the contrary was high in worm eggs and larval stages, but then declined throughout development. Some particular phospholipids, such as many phosphatidylcholines (PCs) and phosphatidylethanolamines (PEs), were present in higher abundance during the larval stage and early adulthood (day 1) and lowered during adult lifespan. Lysophosphatidylethanolamine (LPEs) levels show similar trends, while some cardiolipins (CLs) had low abundance in L2 stage and increased during early adulthood. Other phospholipids such as phosphatidylglycerol (PG) and a sphingomyelin (SM) species showed an opposite pattern; they were less abundant during the larval and early adult stages and accumulated at a later stage of life (day 10) [134].
Adult lifespan and larval development are regulated by *daf-12*, a nuclear hormone receptor that functions as a ligand-dependent switch. 2D-NMR and DANS of WT and *daf-12* mutants combined with activity guided fractionation has been used to identify several potential *daf-12* ligands, including previously known Δ4- and Δ7-dafachronic acid. Thereby it has been found that Δ1,7-dafachronic acid is the most abundant ligand in WT worms [110].
Nicotinamidase deficiency slows reproductive development in *C. elegans*. Wang *et al.* used UPLC-MS and GC-MS to find metabolic changes in *pnc-1* mutants. Using targeted analyses, they found decreases of NAD+ and glycolytic intermediates suggesting that efficient glycolysis seems to be mandatory for reproductive development [126].

Ascarosides and Other Signaling Molecules

Individual *C. elegans* communicate with each other using a blend of small molecules excreted into the environment. This small molecule signaling has high relevance for the development and the behavior of the nematode. These molecules are called ascarosides and are chemically defined as glycosides of ascarylose, a dideoxysugar. They contain a hydrophobic tail derived from long and very long chain fatty acids and can contain several different, additional modifications at various positions, as shown in Figure 1.2. Ascarosides are involved in dauer formation, male attraction and hermaphrodite repulsion and aggregation.

Srinivasan *et al.* used activity guided fractionation, 2D-NMR and targeted LC-MS measurements to identify mating signals, which were excreted by L4, YA and adult hermaphrodites. Three ascarosides called ascr#2, ascr#3 and ascr#4 were identified and the attraction seemed to be concentration dependent. While males were attracted at low concentrations, deterrence occurred at higher concentrations [102]. Further investigations using differential analysis by 2D-NMR in combination with LC-MS led to discovery of previously unknown ascaroside species called ascr#6.1, ascr#6.2, ascr#7 and ascr#8 [80]. It has been shown that DANS is well suitable for the discovery of ascarosides since it can resolve them from chemical background in the spent cultivation media such as amino acids, peptides and other compounds which otherwise dominate the total ion chromatogram (TIC) at LC-MS. On the other hand, LC-MS was more beneficial for the detection of low concentrated ascarosides as ascr#1,

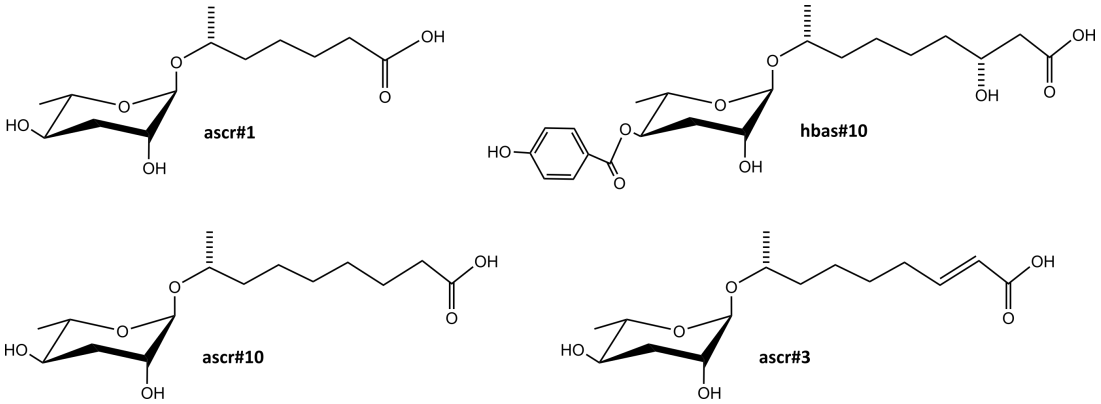


Figure 1.2 Example of ascaroside-structures: ascr#1, hbas#10, ascr#10 and ascr#3.

ascr#6.1 and ascr#6.2.

Targeted analysis using LC-MS was used to profile differences in the ascaroside production between males and hermaphrodites. Hermaphrodites mainly produced ascr#3, containing an α,β-unsaturated fatty acid, while the saturated version ascr#10 was more dominant in males. Interestingly, this slight structural difference, shown in Figure 1.2, significantly affects signaling properties. ascr#3 deters hermaphrodites and attracts males while ascr#10 strongly attracts hermaphrodites [78]. Moreover, Srinivasan *et al.* found that indole ascarosides as icas#1, icas#3 and icas#9 attract both, males and hermaphrodites at high concentrations, whereas at low concentrations males were no longer attracted [172].

Comparative metabolomics by detailed analysis using LC-MS/MS positive and negative ionization of different mutants of peroxisomal β-oxidation detected in total 146 ascarosides of which 124 were previously unreported. This allowed to draft a metabolic pathway of ascaroside biosynthesis and showed that the assembly of ascarosides includes building blocks from carbohydrate, lipid, and amino acid metabolism. The different phenotypic output of signaling of dauer formation, mate attraction, hermaphrodite repulsion or aggregation were linked to the diversity of ascarosides [81].

Lipid Metabolism and Regulation

Regulation of fatty acid and lipid metabolism has been extensively studied in *C. elegans*. Additional techniques different to analysis methods for polar metabolites are necessary due to the very different physicochemical properties of lipids. Different possibilities for analysis of fatty acids and lipids exist and have been applied to *C. elegans*. GC-FID or GC-MS are most widely used for analysis of fatty acids, but also sterols, while intact lipids are analyzed by means of shotgun or LC-MS based lipidomics. Methods for the analysis of lipids and lipid metabolism are reviewed elsewhere [173].

The genetic basis for the biosynthesis of fatty acids have been unraveled by Watts and Browse using GC-FID [174]. The fatty acid composition of *C. elegans* and individual lipid classes has been determined multiple times [87, 123, 135, 141, 143, 175–178]. A comprehensive overview of the *C. elegans* lipidome can be found in the review by Witting *et al.*[173]. Figure 1.3 illustrates the percentage of certain FAs that have been determined in several publications. One of the more recent investigations was performed by Henry *et al.*, which determined fatty acid composition of WT worms using EI and CI at GC-QToF-MS. 28 fatty acids have been identified using their accurate mass and isotopic ratio,

obtained at CI, and accurate mass fragmentation pattern provided by EI [123].

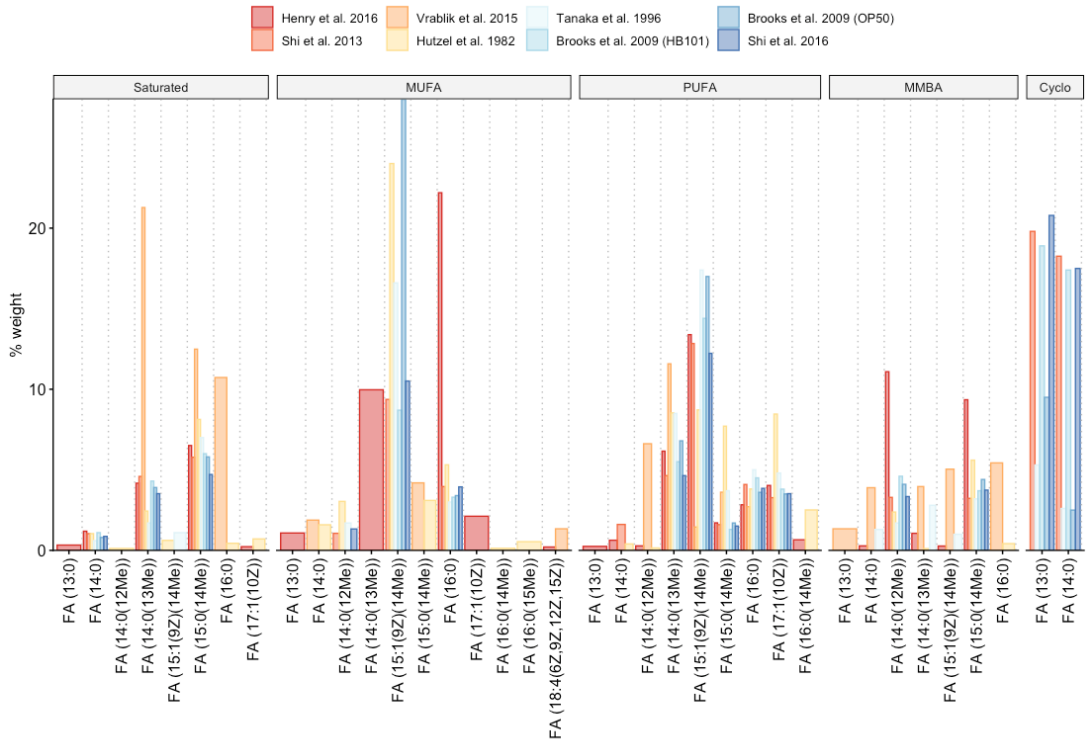


Figure 1.3 FA profiles from different publications. From left to right: saturated, monounsaturated (MUFA), polyunsaturated (PUFA), monomethyl branched chain (MMBA), and Cyclopropane (Cyclo) fatty acids. Publications that were included: Henry *et al.* 2016 [123], Shi *et al.* 2013 [141], Vrablik *et al.* 2015 [143], Hutzel *et al.* 1982 [175], Tanaka *et al.* 1996 [176], Brooks *et al.* 2009 (HB101 & OP50) [177], and Shi *et al.* 2016 [178].

Although the actual fatty acid content is very dependent on the actual bacterial diet [177], several key features are conserved from different analyses. Several fatty acids are directly taken up from the food bacteria, while others are synthesized entirely by *C. elegans*. While the *E. coli* diet of *C. elegans* is rich in the saturated fatty acids lauric acid, myristic acid and palmitic acid, they are only minor species in *C. elegans*. Interestingly, palmitic acid and stearic acid have high percentages in Phosphatidylinositols (PIs) [135]. The nutritional cyclopropane fatty acids cis-9,10-methylene hexadecanoic acid (FA (17:0Δ)) and cis-11,12-methylene octadecanoic acid (FA (19:0Δ)) are enriched in triacylglycerols compared to other lipid classes, but are also found to a certain extent in other lipid classes. Some differences between lipid classes in their fatty acid composition exist. PC and PE are the major building blocks of membranes. Branched chain fatty acids 13-Methylmyristic acid (FA (14:0(13Me)) and 15-Methylpalmitic acid (FA (16:0(15Me)) are exclusively produced by *C. elegans* [140]. However, in most fatty acid profiles they were only found as minor species.

The *C. elegans* lipidome consists of several different lipid classes and shows several parallels to other lipidomes but has also several particularities. A particular example are sphingolipids, which are specifically found in *C. elegans* and makes it different from mammalian lipid metabolic pathways. In *C. elegans* they contain an unusual, branched chain sphingoid base derived from condensation of 13-

methyl myristic acid with serine, resulting in 14-methylhexadecasphinganine [150, 151, 179, 180]. Different sphingolipids were targeted in several investigations. For example, Mosbech *et al.* found that mutation of different ceramide synthases has different effects on lifespan. While loss of HYL-2, required for the synthesis of shorted chain sphingolipids with N-acyls smaller than 22 carbons, decreases lifespan, loss of HYL-1 and LAGR-1 is not affecting the lifespan [128]. Likewise, Menuz *et al.* found that loss of function of HYL-1 leads to increased resistance to anoxia [136]. Recently, an in-depth investigation to define the *C. elegans* sphingolipidome has been performed. Extensive fractionation of lipid extract together with UPLC-UHR-ToF-MS based analysis was used to identify lipids from different classes of sphingolipids [151]. Furthermore, *C. elegans* contains an interesting class of modified glucosyl-ceramides, which have an unusual PE or mono-methyl-PE modification on the sugar residue. These molecules have been shown to be able to rescue phenotypes derived from sterol-depletion [181]. An interesting approach for more detailed structural elucidation of SM species in *C. elegans* was developed by Zhao *et al.* Fragmentation of the [M+HCO3]- adduct of SM led to radical-directed dissociation, yielding a high abundant intact N-acyl-chain fragment allowing to determine the composition of the sphingoid base and N-acyls [182].

The fatty acid profile of sphingolipids shows a different fatty acid profile compared to glycero- and glycerophospholipids. Typically, a mild alkaline saponification is used to generate free fatty acids from glycero- and glycerophospholipids, which are in turn analyzed as their methyl esters by GC-MS. Amide bound fatty acyls in sphingolipids are inert under this condition and require a different strategy. Therefore, total fatty acid profiles normally represent ester bound fatty acids ignoring fatty acyls from sphingolipids. Chitwood *et al.* and Gerdt *et al.* performed analysis of glucosylceramides in *C. elegans* and identified several 2-hydroxy fatty acids, which are not found in glycero- and glycerophospholipids. These fatty acids represent long chain saturated fatty acids ranging from 16 to 26 carbons and also include odd numbered chains. Interestingly, also even numbered iso branched chain fatty acids were reported [183, 184].

Another *C. elegans* specific lipid class has been identified, investigating the lipidome of dauer larvae. The new lipid class, named maradolipids, are chemically 6,6'-di-O-acyltrehaloses with specific fatty acid composition and has been identified by 2D-NMR and analyzed by shotgun lipidomics. Maradolipids might be important for understanding the chemical basis for the resistance of dauer larvae to extreme environmental stress [84]. Further analysis of the *C. elegans* lipidome by shotgun lipidomics led to the discovery of lysomaradolipids [137]. Recently, analysis of maradolipids from dauer larvae using LC-IM-MS has been performed. DIA was combined with IM separation to obtain fragmentation data of maradolipids, which led to the identification of 45 marado- and lysomaradolipids directly from dauer larvae lipid extracts, without the need of further prior purification of glycolipids [185].

Ether lipids are an important part of the lipidome of *C. elegans*. Interestingly, in contrast to mammals, it contains mostly ether-linked phosphatidylethanolamines. Plasmenyl-PE (P-PE) and Plasmanyl-PE (O-PE) represented 5.1% and 4.9% of phospholipid content as determined by LC-MS/MS. These lipids mostly contained a 18:0 side chain at the sn1 position linked as ether or vinyl-ether [186]. Upon mutation of key enzymatic genes of ether lipid biosynthesis an increase in *de novo* fatty acid biosynthesis and reduction in the desaturases *fat-5* and *fat-7* was observed, which indicates a response to altered lipid composition upon ether-lipid deficiency [178].

Dietary restriction (DR) also has profound effects on the lipid metabolism and lipid levels. Klapper *et al.* studied supplementation with choline during DR and has seen that lipid stores are utilized at

a later time point. Interestingly, choline supplementation only changed phenotypes related to lipids, but did not reverse the increase in lifespan [93]. *C. elegans* stores excessive energy in lipid droplets consisting of triacylglycerols. Although believed to be rather "inert" and only utilized open energy demand, they can be indicative of changes in lipid metabolism and serve as a buffer for fatty acids not used. Therefore, studying lipid storage is important to understand lipid metabolism. Schmökel *et al.* performed analysis of genetic of lipid storage in *C. elegans* embryos. Different genes were screened for a large lipid droplet phenotype. An interesting hit was *asm-3*, which is a member of acid sphingomyelinase genes. The closely related genes *asm-1* and *asm-2* did not show the large lipid droplet phenotype, suggesting a specific role for *asm-3*. Using lipidomics by LC-MS/MS difference in several lipid species were observed, while the total lipid content and the relative class distribution did not change too much [129].

The synthesis of lipids requires coordination of different genes in different metabolic pathways. The transcription factor SREBP-1 (sterol regulatory element binding protein) is one of the central transcription factors of lipogenesis. Low levels of PCs stimulate *sbp-1*/SREBP-1. Using a targeted RNAi screen Smulan *et al.* identified *lpin-1* and *arf-1.2* to be important for low-PC activation of *sbp-1*/SREBP-1 [187].

Several lipid related genes are still orphan in regard to their exact function. *lipl-5*, a lipase-like gene, is regulated by nutritional status, with starvation increased expression of *lipl-5*. It has been shown that mostly ceramides and mitochondrial lipids such as cardiolipins are affected by *lipl-5* mutation. Differences in cardiolipins were accompanied by differences in mitochondrial activity [144].

Recent publications show that a strong link between mitochondrial activity, longevity and lipid metabolism exists. Haeussler *et al.* studied mitochondrial dynamics in *C. elegans* and found that autophagy compensates for defects in mitochondrial fusion. Mutation in *fzo-1* led to fragmented mitochondria, decreased mitochondrial membrane potential and induced the mitochondrial unfolded protein response. Comparing the levels of TG species between wildtype N2 and *fzo-1* mutants, species with less carbons and double bonds were down-regulated, while longer and more unsaturated ones were up-regulated [145]. Mitochondrial dynamics and mitochondrial translation are linked to longevity via the transcription factor *hlh-30* [188]. *hlh-30* mutants are susceptible to starvation. WT worms showed increased longchain acyl-carnitines while short-chain acyl-carnitines were mostly unchanged. In the lipidomics analysis cardiolipins were found to show a similar pattern. Upon mutation of *hlh-30* this effect was abolished. Further analysis showed that *hlh-30* mutants shift their metabolism from mitochondrial to peroxisomal beta-oxidation. Further knockdown of *prx-5*, compromising the generation of peroxisomes, further rendered *hlh-30* mutants hypersensitive to starvation [146].

Lipid analysis has also been performed in the context of drugs slowing down ageing. Admasu *et al.* used a combination of different drugs extending the lifespan of *C. elegans*. One of the tested drug combinations required master lipid regulator transcription factor *sbp-1* to be present. Changes in desaturation of lipids has been linked to the extended life span, with a higher monounsaturated fatty acids (MUFA) to polyunsaturated fatty acids (PUFA) ratio. PC, PE and TGs lipids have been detected using LC-MS [147, 148].

Gao *et al.* not only used analysis of polar metabolites as described above, but also performed analysis of lipids over a period of ten days or under different feeding conditions. In contrast to many other studies, which used RP based lipid analysis, they employed a normal phase separation. They determined changes in fatty acids and phospholipids in *mdt-15*mutants. *mdt-15*controls the expression of FA desaturase genes as *fat-7* and is therefore required for the synthesis of PUFAs. They found

increased C18:0 and decreased C18:1 and PUFAs such as C18:2, C20:3, C20:4 and C20:5 levels in *mdt-15*mutants. Also PL having saturated acyl chains showed higher levels and PLs containing PUFAs, such as PS(38:7), PI(40:9), or PC(40:10), were decreased in *mdt-15*deficient worms [134].
Changes in the lipid metabolism of dauer larvae has been studied by Lam *et al.* using HPLC-MS/MS. They found that membrane phospholipids in dauer larvae are enriched in PUFAs. This enrichment is accompanied by an increase of free PUFAs as well as derived oxidative metabolites, such as prostaglandins. Release of PUFAs bound in membranes and higher levels of eicosanoid metabolites are hallmarks of termination of the dauer stage [149].
Metabolomics and lipidomics are also useful to elucidate the function of unknown genes. Schwudke *et al.* used top down lipidomics to elucidate the impact of the knock-down of *pmt-1* and *pmt-2* on the worm lipidome. Analysis of fragmentation data identified monomethyl-phosphatidylethanolamine (MMPE) and dimethyl-phosphatidylethanolamine (DMPE) as intermediates of PC biosynthesis from PE. *pmt-1* and *pmt-2* are homologues to plant methyltransferase. *pmt-1* deficient worms arrest their development at the L4/early adult stage and *pmt-2* worms at L3 stage. Upon silencing of *pmt-1* a decrease in MMPE and DMPE was observed, while RNAi of *pmt-2* lead to an increase of only MMPE [122].
Insights gained into lipid metabolism so far represented snapshots of the current state of lipid metabolism. However, metabolism is a dynamic phenomenon and changes in lipid levels might have been achieved by different means of rerouting fluxes. Isotopic labeling can be used to follow different tracer substances and their metabolism. A first example in *C. elegans* applied to lipid analysis has been described by Perez and Van Gilst as already described above. Fatty acids have been analyzed by GC-MS. If either non-labeled or completely labeled fatty acids were detected, they have been directly derived from the maternal fatty acid pool or the diet. Since *C. elegans* uses dietary metabolites for *de novo* biosynthesis, fatty acids that have been produced by the worm will show a mixed labeling pattern. Using this approach, it was found that >99% of mmBCFAs in *C. elegans* are *de novo* synthesized while saturated fatty acids are mostly taken directly up from the diet. Unsaturated fatty acids showed mixed synthesis rates. However, only fatty acids until 18 carbons were examined, since longer PUFAs show too much fragmentation and overlapping isotopic patterns, which could not be deconvoluted. Furthermore, differences in fatty lipogenesis between different alleles of *daf-2* mutation were studied [140]. The same approach was used to further study membrane dynamics. Comparing the percentage of newly incorporated fatty acids into either phospholipids or neutral lipids it was shown that phospholipids have a higher turnover rate. However, the amount of fatty acids from *de novo* synthesis was only for palmitic acid significantly different. Turnover rates were determined in different mutants and identified *fat-5*, *fat-6* and *fat-8* as membrane maintenance regulators. Furthermore, ^{15}N tracers were used to follow the phospholipid turnover in more detail. This analysis revealed that *fat-7* also influences the turnover of lipids towards slower metabolism [189].
Another sub-category in lipidomics is steroid analysis, or steroidomics. The currently best known steroid pathways are related to dafachronic acids [190]. They play an important role as signaling molecules in dietary restricted worms and dauer formation [191]. There are several potential *daf-12* ligands, as Δ4- and Δ7-dafachronic acid. Due to activity guided fractionation Δ7- and Δ1,7-dafachronic acid have been identified as most abundant ligands in *daf-12* mutants [110]. Different methods for quantification of dafachronic acid have been reviewed elsewhere [83]. Recently, the loss of *sul-2*, the steroid sulfatase has been investigated in *C. elegans*. Steroid hormone sulfatases are involved in development of hormone-regulated cancers, highlighting the importance of their investigation. Re-

searchers found increased longevity in *sul-2* mutants and also sulfated hormones have been increased compared to WT worms [192].

Food Source and Nutrition

Changes in the *C. elegans* metabolome are depending on the food source and nutritional status [94]. Worms fed with different bacteria by show differences in their fatty acid composition. Mainly saturated FAs were directly taken from the diet and unsaturated FA were mixed. For instance, higher abundance of C15:0 and C17:0 fatty acids were observed when fed with *Bacillus subtilis* instead of *E. coli*. Thereby, most fatty acids of the worm reflect the fatty acid composition of the bacteria they eat, while some other fatty acids were absent in all tested bacterial strains. These fatty acids were hence mainly synthesized *de novo* in *C. elegans*. However, changes in lipid metabolism strongly depends on different alleles [87, 140]. On the other hand, very different bacterial amino acid profiles lead to almost identical worm amino acid composition. This suggests that amino acid levels were tightly controlled and regulated in *C. elegans*.

Other Topics

Beside the abovementioned topics, several other investigations have been performed, e.g., toxicity exposure have been accomplished using *C. elegans* as a model. For example, cadmium exposure have been found to decrease cystathionine levels and increase phytochelatin levels due to upregulated methionine transsulfuration pathways, suggesting that cadmium is excreted as a phytochelatin-bound complex as a detoxification mechanism [193]. Exposure to nickel and the pesticide chlorpyrifos affected both mostly branched chain amino acids, lactate, and other energy generation related metabolites [194]. Also, phosphine exposure was investigated in WT and *dld-1* mutants. *dld-1* mutants were less affected by phosphine, whereas WT worms showed an accumulation of branched-chain amino acids, glutamate, and glycine [101]. Lastly, Sudama *et al.* investigate changes of metabolites due to lead exposure. They used HPLC with coulometric array detection to detect and quantify metabolites based on their oxidation-reduction potentials. After lead exposure, changes in tryptophan, tyrosine and purine have been observed in WT worms [195].
The small nematode may also be used as a model for human diseases as shown by Van Assche *et al.*, who were interested in Alzheimer disease (AD). They studied changes in the metabolism after induced expression of human amyloid-beta peptide by GC-MS, RPLC-MS and ANP-LC-MS, which represents an alternative for the separation of polar substances. ANP columns show only small interactions with water, avoiding the major disadvantage of classical NP. Results confirmed previous observations from experiments in AD, also in humans, as for instance increased levels of allantoin [130]. Moreover, Teo *et al.* combined metabolomics data with transcriptomics and computational modeling, and firstly linked amyloid-beta expression to TCA cycle [86].
AD but also Parkinson's disease are neurodegenerative-related diseases believed to be caused by oxidative stress, among other things [196–198]. Helmer *et al.* recently investigated the role of cardiolipins (CLs) as oxidative stress marker in *C. elegans*. Oxidative stress was induced by supplementation of tert-butyl hydroperoxide (tBOOH), resulting in oxidized CL species. CL analysis was performed using 2D-LC-HRMS in order to remove co-eluting lipid species. They identified oxidation products of CLs such as CL 80:14, CL 80:15 and CL 80:16 most abundantly in stressed worms. Based on their high content of poly-unsaturated fatty acids, CLs have been shown to be suitable biomarkers for oxidative

stress [199].
Because of its fast development, *C. elegans* is also an ideal organism for mutation accumulation (MA) experiments. MA experiments give insights about the genetic variation originating from spontaneous mutations. Herby, mutations accumulate in genomes for many generations. The majority of mutations is however expected to be fixed randomly. Therefore, new mutational variances per generation are determined [200]. GC-MS based metabolomic analyses identified 29 metabolites that vary greatly in their vulnerability to mutation [138].
Also, cell death has been investigated in *C. elegans* using 2D-NMR based comparative metabolomics. Researchers identified a blue fluorescent substance that accumulates at *C. elegans* death as anthranilic acid glucosyl esters derived from tryptophan [201].

1.2.4 Publicly Available Datasets

Since *C. elegans* represents an interesting and important model organism, the availability of public data is an important issue. On the one side for further development of bioinformatics approaches and on the other side for the reconstruction of the *C. elegans* metabolome and lipidome. There are also certain publicly available datasets of metabolomics studies of *C. elegans* from different analysis types. Up to date, there have been nine metabolomics studies published on Metabolomics Workbench (https://www.metabolomicsworkbench.org/) and four studies on MetaboLights (https://www.ebi.ac.uk/metabolights/). These contain various information such as sample details, sample-preparation protocols, analyses, post-processing details or annotations. Public deposition of further datasets would facilitate the reconstruction of the *C. elegans* metabolome, allows the extraction of spectral information for secondary metabolites, for which no commercial reference standards are available and extraction of reference values or even concentrations for future use.

1.2.5 Prototyping the *C. elegans* Metabolome

Since currently no *C. elegans* metabolome and/or lipidome database exists, we aimed to collect metabolites and lipids present in *C. elegans*. While reviewing articles on *C. elegans* metabolomics and lipidomics, we have curated metabolites and lipids that have been detected so far in *C. elegans*. Metabolite names were curated from the main text, figures and tables as well as potential supporting information. Corresponding data is collected in a central GitHub repository (https://github.com/wormjam-consortium/wormjam-db) and freely available.
The data is grouped into two libraries. The first library represents metabolites predicted to be present in *C. elegans* based on the WormJam metabolic model. We extracted metabolites from the current version of WormJam v0.1.0 (https://zenodo.org/record/3978712#.X47IqlngpmB,
https://github.com/wormjam-consortium/wormjam/releases). The WormJam metabolite library contains in total 1215 unique metabolites, of which 1205 have a structure associated with it. However, this library is purely in silico.
The second library is based on the literature curation. Since this library might contain different versions of the same metabolite, e.g., full stereochemistry vs. generic structures (e.g., L-glycine vs. glycine) or ions vs. neutral molecules (e.g., lactate vs. lactic acid), we tried harmonizing all names. Besides the correct chemical name, we additionally curated the chemical formula, InChI, InChIKey, SMILES, ChEBI, KEGG, BioCyc, HMDB, LipidMaps, SwissLipids, Wikidata, PubChem, Metabolights and ChemSpider IDs for all metabolites in both libraries, where available. Different IDs were retrieved

using BridgeDB [202]. In total 1421 unique metabolites have been curated from literature. We compared predicted and detected metabolites in *C. elegans* based on the first block of their InChIKey (if available) and found that only 417 metabolites were overlapping, while further 1025 have been detected so far but are not found in the WormJam model (Figure 1.4). The model contains 819 further metabolites that have not been detected yet. Metabolites from literature and from the WormJam model have been combined afterwards to a single consensus database, representing a prototype database of the *C. elegans* metabolome.

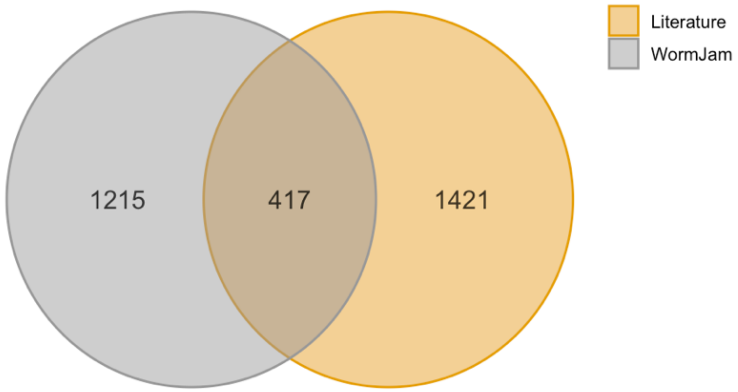


Figure 1.4 Venn diagram of overlap of metabolites between literature (yellow) and WormJam (grey) based on the first block of their InChIKey.

In order to obtain an overview on the properties of the *C. elegans* metabolome and methods that might be required to fully capture it, we combined the list of predicted and curated metabolites into a single database. To estimate a proper analytical platform, we need to gain more information about the molecular properties of the metabolites. For this purpose, we used the rcdk package in R to calculate various physicochemical properties of the metabolites, based on their SMILES [203].
Figure 1.5 shows the molecular weight (MW) and logP distributions of the metabolites combined from literature and WormJam. Both distributions are significantly different from normal distribution (p-value < 2.2e-16). The majority of metabolites (74%) have a lower mass than 500 Da. On the contrary, 17% of the metabolites are above 800 Da. In particular, metabolites in the higher mass-range are originally metabolites from the model, which are mainly coenzyme A (CoA) derivatives. Detected lipids curated from the literature were not counted in, since no definitive structure in form of a SMILES was given for them (see below).
The logP value is a measure for the hydrophobicity/ hydrophilicity of a molecule. It is defined by the partition coefficient P which is basically the solubility of a molecule between an octanol-water phase. The majority (58%) of metabolites from the *C. elegans* prototype metabolome database is very polar, having logP values below 0. And 9% of the metabolites are having logP values below 5. On the other hand, there are 19% of very lipophilic metabolites, having logP values above 5.
In order to cover (more) highly polar metabolites analytically, more techniques specialized in this polarity range needs to be applied. Figure 1.5 also displays the polarity of different analytical separation methods in the logP range. Separation methods as SFC, RPLC, GC, and ANP-LC are more suitable for mid- up to non-polar compounds. Promising separation techniques for highly polar substances are especially HILIC, ion chromatography (IC), and CE. Since very high salt concentrations are used in

IC, it is less suitable for MS coupling. Therefore, we suggest investigating the *C. elegans* metabolome in more detail using HILIC-MS and CE-MS.

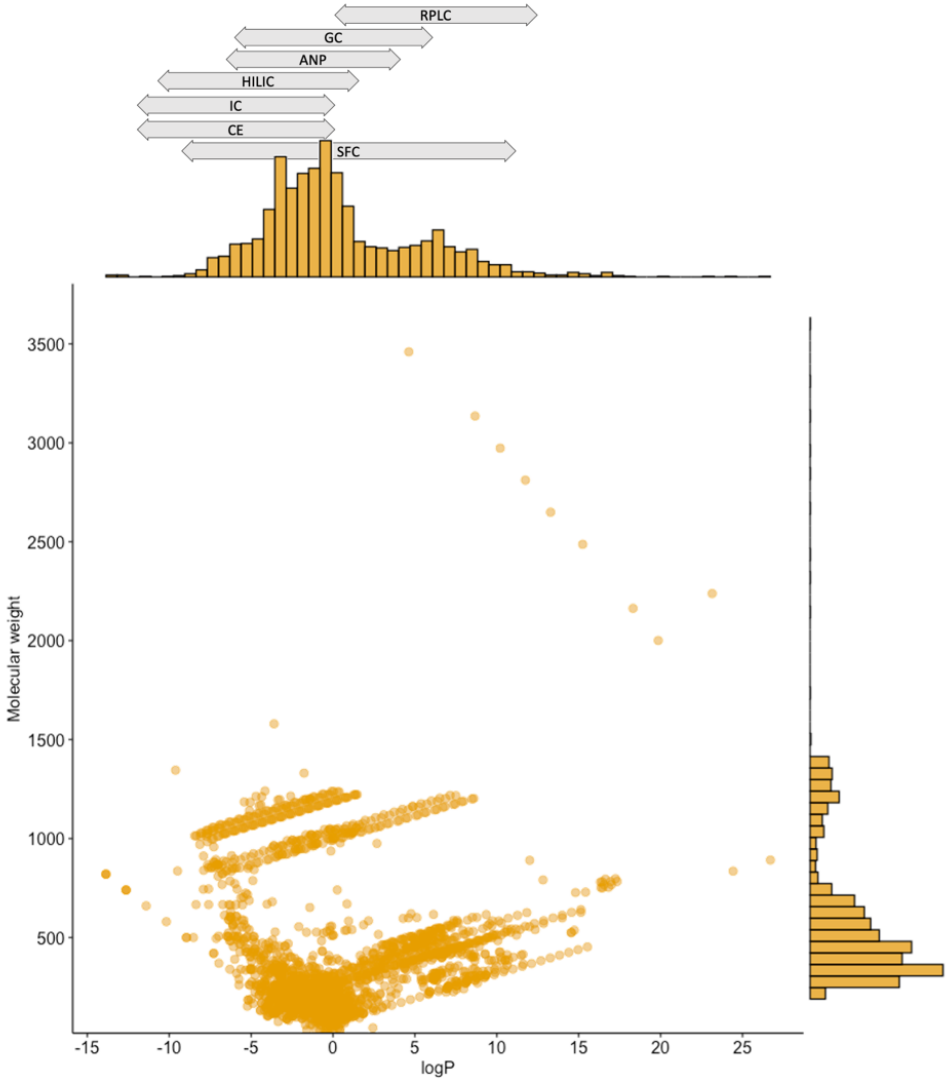


Figure 1.5 Molecular weight and logP distribution of merged molecules from the WormJam model and curated from literature. Different separation methods and their polarity are displayed in the logP range. Reversed phase liquid chromatography (RPLC), gas chromatography (GC), aqueous normal phase (ANP) chromatography, hydrophilic interaction liquid chromatography (HILIC), ion chromatography, capillary electrophoresis (CE) and supercritical fluid chromatography (SFC).

The separation mechanism in HILIC is more based on interactions of the molecules with the stationary phase and the water layer on the stationary phase surface respectively. In contrast, in CE molecules are separated due to different mobilities in an electric field. Besides polarity, the main requirement for separation in CE is charge, which is not necessary in HILIC. Since the separation mechanism between CE and HILIC is different, the two would offer different possibilities and selectivity's for separation of polar metabolites.
To demonstrate whether the *C. elegans* metabolome contains suitable CE candidates we determined the proportion of molecules having a permanent or pH-dependent charge. 157 structures of the

combined database were associated with a permanent charge, based on their SMILES. However, all molecules except of 11 have an ionizable group and thus be made applicable for CE separation. Since CE-MS has not yet been used in *C. elegans* metabolomics, we strongly believe that this technique allows taking further steps in unravelling the *C. elegans* metabolome.
Beside metabolites, several different lipids classes and species have been detected. Identifications were reported on different levels of detail in regard to the structural information. Analysis of lipids has been performed with different approaches. Likewise, lipids were reported on different levels of detail. In order to generate a harmonized list of detected lipids the naming has been standardized on the lipid species level. Since no explicit structure is available for lipids species, no InChI, InChKey or SMILES are available. In order to be able to compile this into a single list of *C. elegans* lipids they have been all normalized to the lipid species level according to Liebisch *et al.* [204]. In total 1402 unique lipid species have been curated together with their molecular formula. The list of unique metabolites and lipids are available from GitHub. We welcome future suggestions for publications to be added to the collection as well as further collaboration in the curation process.

1.2.6 Conclusions and Outlook

In this review, we provided an overview about recently used metabolomics analytical platforms and biological applications from the model organism *Caenorhabditis elegans*. *C. elegans* is an ideal organism for metabolomic investigations because of its invariant cell number, easy and fast cultivation and its short lifespan. One of the most common applications in the field of metabolomics are studies of longlived mutants and ageing in the worm. Thereby, the most studied long-lived mutant is *daf-2*. Similar to other long-lived mutants, as for instance *eat-2*, *slcf-1*, *ife-2*, but also dauer larvae, they share a metabolic signature of longevity as for instance changes in amino acids, carbohydrate, choline or upregulated lipid storage. However, it still remains elusive, which changes in metabolism corroborate long live and which are not related. Metabolomics studies have predominantly been performed using, either individually or in combination, analytical platforms as 1D- and 2D-NMR approaches, and GC and LC with different column selectivity coupled to MS. As the properties are very different to metabolites, lipids are mostly studied individually with their own methods. They are predominantly analyzed by mass-spectrometry, either at direct-infusion or after chromatographic separation.
The *C. elegans* metabolome and lipidome covers a large range of polarity, so multiple analytical methods need to be combined to achieve true comprehensive profiling. Especially, the highly polar region (negative logP values) requires further developments.
Other promising analytical separation techniques prior MS analysis that have not yet been applied to *C. elegans*, but in other metabolomics studies, are ion mobility or CE. IM may further enhance peak capacity of LC-MS data and might probably enhance annotation success of *C. elegans* metabolomics data. The benefits of using IM have been demonstrated in lipidomics, where it enables the detection of new species of maradolipids in the nematode.
CE-MS is another rising powerful technique in metabolomics, which is especially suitable for highly polar and charged metabolites. Therefore, it is a complementary analytical platform to HILIC-MS. Due to its very different separation mechanism to LC, we strongly believe CE will help uncovering the *C. elegans* metabolome. Another characteristic in CE is the low injection volume, which is a few nl. Because of this, we think that CE-MS metabolomics analysis will require fewer amount of worms, compared to LC-MS. This would significantly reduce laboratory efforts and also minimize genetic variation. Given the above reasons, we see a great importance to expand these platforms in *C. elegans*

metabolomics.
Appropriate analytics, sample preparation and evaluations on metabolomics data needs to be adapted on the biological question. As generally valid for metabolomics, there is no analytical platform that provides a comprehensive analysis of the *C. elegans* metabolome. Bioinformatics analysis of (*C. elegans*) metabolomics data is also gaining increased importance. For example, pathway or network analysis as GGM were increasingly applied. Flux balance analysis have also been shown as an effective tool to analyze metabolites from tissue specific pathways in *C. elegans*. There have been first steps in developing tissue and cell type specific models of the nematode.
Lastly, we created a draft metabolomics/ lipidomics database of *C. elegans* by curating all annotated metabolites from reviewed articles into a publicly available repository. Increasing this database and matching it to in silico databases as WormJam helps gaining new insights to the *C. elegans* metabolome and lipidome. Moreover, it may be useful to accelerate annotation of metabolomics data and avoid over annotation.
The remaining question is "quo vadis?", where is *C. elegans* metabolomics and lipidomics heading towards? Current applications are mostly based on populations of *C. elegans* with different absolute number of worms per sample. However, although isogenic worms show an individual response to treatments, growth conditions, etc. Therefore, in future it will be important to go from populations to single worms. In contrast to DNA and RNA, proteins and metabolites cannot be amplified. This requires highly sensitive analytical workflows. A first example of single *C. elegans* proteomics has been demonstrated by Bensadeek *et al.* [205]. Likewise, a first application of single worm metabolomics using μHR-MAS-NMR has shown that metabolites from a single nematode is possible [113].
A further important step required is the creation of *C. elegans*-specific metabolite database following the example of HMDB [13], including metabolites, MS-, MS/MS- and NMR spectra. Potentially this can be realized under the umbrella of WormBase to provide a "one-stop-shop" for *C. elegans* scientists [206].
Since *C. elegans* represents a multicellular organism with differentiated tissues, the location of specific metabolites is an important factor. First tissue specific metabolic models have been developed [167]. The transparent body of *C. elegans* enabled many different microscopy methods to be used for localization of proteins, e.g., using GFP-fusions. Tissue specific metabolomics, e.g., using MS imaging shall be developed for the spatially resolved metabolomics. Likewise, single-cell metabolomics generally gains interest and can be potentially applied to isolated *C. elegans* cells. Although, it is not clear where the next developments will lead to, the future for *C. elegans* metabolomics and lipidomics is bright.

Data availability
Data for Prototyping the *C. elegans* Metabolome in Section 1.2.5 is available at GitHub (https://github.com/wormjam-consortium/wormjam-db)

1.3 Analytical methods in metabolomics

In the following, the separation techniques and mass analyzers used in this thesis are described (specifically Chapter 3).

1.3.1 Liquid chromatography

Not only in metabolomics, but also in other fields like proteomics, high performance liquid chromatography (HPLC, originally referred to high pressure liquid chromatography) is one of the most employed separation techniques, and in most cases coupled to mass spectrometry (MS).

The separation mechanism relies on the interaction of the analyte with a sationary phase (i.e. the material of the column), and the mobile phase (i.e. the solvents). The interaction, or retention of a molecule in such a system can be described by the retention factor k (Equation (1.1)).

$$k = \frac{t_R - t_0}{t_0} \tag{1.1}$$

k: retention factor
t_R: retention time
t_0: void time

The void time t_0 is the retention time for an unretained compound in a given chromatographic system. The selectivity α of the separation of two peaks x and y (where $t_R(x) < t_R(y)$) can be described as follows in Equation (1.2).

$$\alpha = \frac{k_y}{k_x} \tag{1.2}$$

α: selectivity

The resolution R of a chromatographic system consists of the retention factor, the selectivity and efficiency of a chromatographic system as shown in Equation (1.3).

$$R = \frac{k}{k+1} \cdot \frac{\alpha - 1}{\alpha} \cdot \frac{\sqrt{N}}{4} \tag{1.3}$$

R: resolution
N: column efficiency
(number of theoretical plates)

Whereas the column efficiency can be described by the number of theoretical plates N as in Equation (1.4).

$$N = 5.54 \cdot \left(\frac{t_R}{w_{1/2}}\right)^2 = \frac{L}{H} \tag{1.4}$$

$w_{1/2}$: peak width at half height
L: column length
H: height equivalent to a theoretical plate (HETP)

By taken Equation (1.1), Equation (1.2), and Equation (1.4), the resolution R (Equation (1.3)) of two peaks x and y can be simplified as:

$$R = \frac{2 \cdot (t_R(y) - t_R(x))}{w(x) + w(y)} = 1.18 \cdot \frac{(t_R(y) - t_R(x))}{w_{1/2}(x) + w_{1/2}(y)} \tag{1.5}$$

w: peak width at baseline

To enhance the separation efficiency and effectively resolve two peaks, the utilization of selectivity α, as indicated in Equation (1.3), has been a common approach. However, when dealing with non-targeted metabolomics, the application of α becomes impractical due to the unbiased nature of the analysis. In this context, the initial knowledge of analytes is absent.
Consequently, an alternative strategy is employed, involving the reduction of the height equivalent to a theoretical plate (HETP or H) to augment peak resolution, as outlined in Equation (1.4). This objective is attained by optimizing the flow rate (υ) within the chromatographic system. The theoretical basis for this approach is encapsulated in the van Deemter equation (Equation 1.6) [207]. This equation serves as a fundamental guideline for achieving optimal chromatographic separation by balancing factors that influence the broadening of peaks.

$$H = A + \frac{B}{\upsilon} + C \cdot \upsilon \tag{1.6}$$

A: eddy diffusion coefficient
B: longitudinal diffusion coefficient
C: mass transfer coefficient
υ: flow rate

Eddy diffusion A describes the various possible flow paths of an analyte through the column. Hence, it is influenced by the column packing. The longitudinal diffusion B describes the diffusion of molecules in longitudinal direction and decrease with increasing flow rate. Lastly, the mass transfer C describes the interaction effects of the analyte with the column chemistry, i.e. ad- and desorption from the mobile and to the stationary phase. Therefore, a high flow rate also increases the mass transfer and hence lowers separation efficiency.

Modern analytical practices have shifted towards employing Ultra Performance Liquid Chromatography (UPLC) instead of High Performance Liquid Chromatography (HPLC). UPLC offers distinct advantages due to its finer particle size columns and reduced flow rates, which collectively contribute to lower HETP values. This improvement in column efficiency directly translates to enhanced peak resolution. In contrast to traditional HPLC, UPLC's finer particles allow for increased surface area interactions and quicker mass transfer kinetics. As a result, the van Deemter equation is favorably impacted, enabling higher separation efficiency with narrower peaks. Consequently, UPLC has become the preferred choice in non-targeted metabolomics and other high-resolution analytical methodologies, as it better aligns with the need for improved resolution without the constraints posed by the limitations of HETP in HPLC.

Hydrophilic interaction liquid chromatography

Until now, the predominant approach for metabolomics analyses has revolved around reversed-phase liquid chromatography coupled to mass spectrometry (RPLC-MS). However, this methodology exhibits limitations in capturing a significant portion of the polar fraction within the metabolome [208]. Recognizing this gap, hydrophilic interaction liquid chromatography (HILIC) emerged in 1990 to specifically address the separation of polar compounds [209]. This pioneering development marked a crucial turning point in analytical techniques.
Compared to RPLC, separation in HILIC is very complex and is mostly based on the partitioning of polar metabolites between a water-rich layer immobilized on the surface of the column material and

the organic mobile phase. Different stationary phases for HILIC exist and contain different polar or ionic groups to bind water molecules. Due to these groups, different secondary interactions are observed including ionic and electrostatic interactions [210]. These secondary interactions generally lead to broader peaks and lower efficiency in HILIC compared to RPLC.
The mobile phase in HILIC is most commonly acetonitrile (ACN) and water, often with the addition of salts [211]. HILIC separations start at high ACN content and elution proceeds with increasing water composition of the mobile phase. However, these starting conditions are problematic for the reconstitution of polar samples, i.e. the incompatibility with polar metabolites with the injection solvent (high organic), which often lead to analyte precipitation of peak distortion or major RT shifts [212].

1.3.2 Capillary electrophoresis

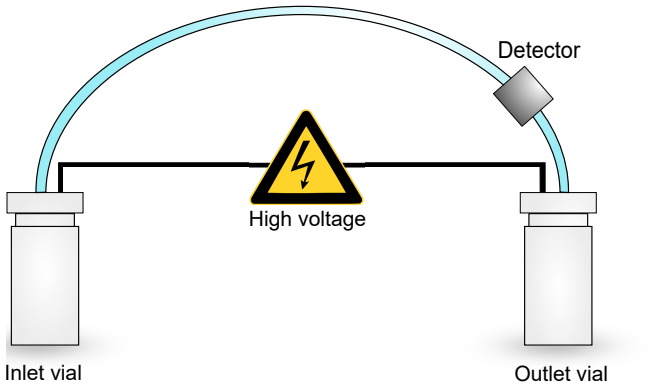


Figure 1.6 Scheme of capillary (zone) electrophoresis systems. Fused silica capillary is filled with background electrolyte (BGE) and stays in contact with two solvent reservoirs, one inlet and one outlet. Separation is performed by applying a high voltage.

Electrophoresis is the migration of charged molecules in a liquid background electrolyte (BGE) by applying an electric field. Molecules separate due to their different velocities, relying on their mobility and the electrical field being applied (Equation (1.7)). The mobility is proportional to the number of charges and inverse proportional to the molecule size as shown in Equation (1.8) [213].

$$v = \mu_e \, E \tag{1.7}$$

v: ion velocity
μ_e: electrophoretic mobility
E: electric field

$$\mu_e = \frac{q}{6 \, \pi \, \eta \, r} \tag{1.8}$$

q: charge
η: viscosity
r: hydrodynamic ion radius

In capillary electrophoresis (CE) a fused silica capillary is used, usually with internal diameters of about 50 µm and an outer diameter of about 375 µm. Most frequently capillary zone electrophoresis (CZE) is being used, where the capillary is filled with a homogeneous BGE and stays in contact

with two solvent reservoirs, one inlet and one outlet (Figure 1.6). Sample injection is conducted either hydrodynamically by applying pressure or electrokinetically by using high voltage. Separation is performed by applying a high voltage, usually within +30 kV and -30 kV. When a positive voltage is applied, cations start migrating in different velocities and in discrete zones towards the cathode (outlet vial), where they are detected. At a negative voltage, anions start migrating towards the outlet and cations are counter-migrating. The detection is usually conducted by UV-detection, or in metabolomics most commonly by hyphenation to MS.
One crucial effect that occurs in CE is the electroosmotic flow (EOF). The inner wall of the capillary consists of silanol groups which are partially deprotonated, depending on the pH of the BGE. The negatively charged capillary wall leads to an accumulation of cations, such as protons, at the wall's proximity ("Stern layer"). If now an electric field is applied, the proton layer migrates towards the cathode which results in a bulk flow of the entire capillary liquid, which is called EOF. The EOF enables the detection of neutral and counter migrating molecules. The EOF is strongly influenced by the pH of the BGE; the higher the pH, the more silanol groups are deprotonated, and the stronger the EOF. Moreover, the EOF can be influenced by coatings of the inner capillary wall, the ionic strength and the viscosity of the BGE [214].
The observed mobility of a molecule is composed of its effective mobility μ_{eff} and the mobility of the EOF μ_{EOF} in the current electrophoretic system (Equation (1.9)). The observed mobility is accessed experimentally by the capillary length, the observed migration time and the applied voltage (Equation (1.10)) [213]. However, a major drawback in CE compared to other separation methods as HPLC, is poor reproducibility of migration times (MT) due to fluctuations in the strength of the EOF between different runs. These shifts in MT can be circumvented by transforming the MT scale into the effective mobility (μ_{eff}) scale using migration markers. Only recently, Drouin *et al.* have shown how mobility transformation can increase reproducibility and improves metabolite annotation and identification in CE-MS based metabolomics, since the effective mobility can be regarded as physicochemical property of the molecule and enables comparisons within the same electrophoretic system [215].

$$\mu_{obs} = \mu_{eff} + \mu_{EOF} \tag{1.9}$$

μ_{obs}: observed mobility
μ_{eff}: effective mobility
μ_{EOF}: mobility of EOF

$$\mu_{obs} = \frac{l_{eff}\, l_{tot}}{t_{obs}\, U} \tag{1.10}$$

l_{eff}: effective length of the capillary
l_{tot}: total length of the capillary
t_{obs}: observed migration time
U: applied Voltage

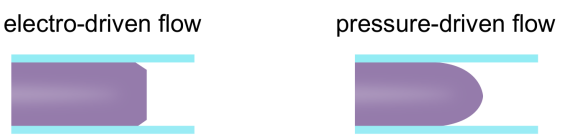


Figure 1.7 Scheme of velocity flow profiles electro-driven and pressure-driven [213].

One of the most remarkable characteristics of Capillary Zone Electrophoresis (CZE) is its remarkably efficient separation capability. The uniform flow profile of the Electroosmotic Flow (EOF), generated by the electro-driven force, is evenly distributed along the capillary, leading to exceptionally narrow peaks (Fig. 1.7). In contrast, pressure-driven flows induce parabolic velocity flow profiles, contributing to peak broadening.
Compared to LC, the issue of peak broadening is less significant in CE, as it is primarily attributed to longitudinal diffusion. Terms A and C from the van Deemter equation (Equation (1.6)) do not apply under ideal conditions, making CE this highly efficient separation method [213]. The separation efficiency can be accessed by the number of theoretical plates N divided by the capillary length (N/l_{tot}), with N being calculated according to Equation (1.4).

1.3.3 Mass spectrometry

Mass spectrometry (MS) stands as one of the foremost analytical techniques extensively employed in metabolomics. In the ensuing sections, the specifics of the mass spectrometer and ionization technique utilized within the context of this thesis will be described.

Time-of-flight mass spectrometry

Time-of-flight (ToF) MS are grouped together with Orbitrap MS and Fourier-transform ion cyclotron resonance (FT-ICR) MS to high-resolution MS (HR-MS). Mass resolution in mass spectrometry refers to the ability of the instrument to distinguish between ions with different mass-to-charge ratios (m/z) in a sample. It is a measure of how effectively the mass spectrometer can separate ions that have slight differences in their masses or mass-to-charge ratios. Mathematically, mass resolution R is expressed as the ratio of the difference in mass between two adjacent ions Δm to the average width of their peaks W at a specific location:

$$R = \frac{\Delta m}{W} \tag{1.11}$$

Mass resolutions up to 60.000 can be reached in ToF-MS, dependent on the flight time of a molecule. High resolutions can be achieved by increasing the flight path. The different m/z values are calculated by their arrival time at the detector by the relation summarized in Equation (1.12).

$$t^2 = \frac{m}{z} \cdot \left(\frac{L^2}{2 \cdot e \cdot U} \right) \tag{1.12}$$

t: traverse time
m: mass of the ion
z: charge number
L: length of the traverse path
e: elementary charge
U: acceleration potential

The main benefit of ToF-MS compared to other HR-MS is its fast scan speed, enabling mass analysis in milliseconds range [216]. The rapid acquisition pace of ToF-MS renders it amenable to seamless integration with diverse supplementary separation techniques like LC or CE, accomplished within a matter of minutes. Remarkably, ToF-MS can even accommodate an additional third dimension of separation, namely ion mobility, executed in mere seconds. These attributes collectively position ToF-MS as an exceedingly enticing and fitting analytical technique across various domains, notably including metabolomics.

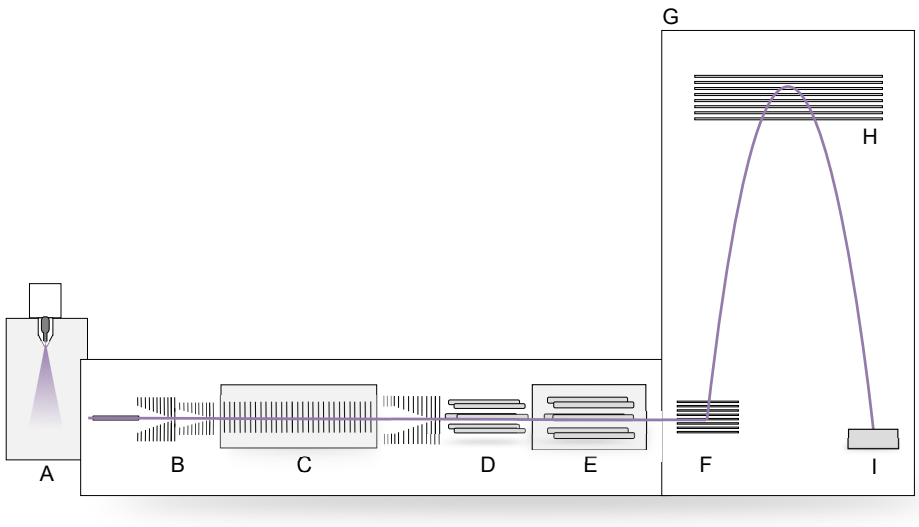


Figure 1.8 Scheme of qToF-MS based on Agilent 6560 Ion Mobility Q-ToF (Agilent Technologies, Waldbronn, Germany). Consists of ionization source (A), ion funnels (B), drift tube (C), quadrupole mass filter (D), collision cell (E), ion pulser (F), ToF tube (G), reflectron (H) and detector (I).

In this research, an Agilent 6560 Ion Mobility (IM) Q-ToF was used (Figure 1.8). However, in this thesis IM separations were not conducted. Notably, the configuration of alternative Q-ToF-MS instruments bears resemblance to this setup. The system is composed of distinct components, each contributing to the mass analysis of incoming molecules. The ionization source (A) is used for the hyphenation of different separation techniques (such as CE, or LC), and ionization of the compounds. Subsequently, generated ions traverse via a transfer capillary to enter ion funnels (B). Functioning as high-pressure focusing optics, these ion funnels adeptly concentrate incoming ions. The ions are then transferred to a trapping funnel. In the context of the Agilent 6560 Ion Mobility-Q-ToF, a drift tube segment (C) follows, facilitating supplementary ion mobility separation. Within the drift tube, ions gain acceleration due to an applied electrical field. Separation is realized through distinct ion mobilities, resulting from ion collisions with counter-flushing gas molecules. Successively, ions are re-focused within the rear funnel and then directed to the quadrupole mass filter (D). The quadrupole operates either in radio frequency (RF) mode, granting passage to all ions, or in direct current (DC) mode to exclusively select precursor ions for subsequent fragmentation within the collision cell (E). The collision cell is typically another quadrupole and used to perform collision-induced dissociation (CID) of those precursor ions. Fragmentation can be orchestrated through several modes, including Data-Dependent Acquisition (DDA), Data-Independent Acquisition (DIA), or targeted Multiple-Reaction Monitoring (MRM). In the realm of non-targeted metabolomics, DDA predominates, involving the selection of the most intense precursor ions for fragmentation. To acquire comprehensive insights, various collision energies are often applied during fragmentation. To circumvent the repetitive selection of the same precursor ion, a dynamic exclusion list is recommended. This list temporarily excludes certain ions for a defined number of cycles, while the second and third most intense ions undergo fragmentation. This strategy aims to maximize the fragmentation of a multitude of molecular features. Subsequent to fragmentation, both fragments and precursor ions proceed to the Time-of-Flight (ToF) analyzer. Here, an ion pulser (F) propels them orthogonally into the ToF tube (G). Upon reaching the tube's terminus, ions encounter a reflectron (H), which compensates for kinetic energy

disparities and facilitates ion reflection. At the end of the qToF-MS, ions are detected e.g. using a micro channel plate (I).

MS hyphenation to LC and CE

Electrospray ionization (ESI) is a pivotal technique in mass spectrometry that enables the efficient transfer of ions from solution into the gas phase for analysis. ESI involves the generation of charged droplets from a liquid sample, followed by the evaporation of solvent molecules from these droplets, resulting in the formation of ions that are then introduced into the mass spectrometer for analysis.

While ESI is a common technique, it can be adapted for different separation methods, such as liquid chromatography (LC) and capillary electrophoresis (CE) (Figure 1.9). The primary difference lies in the nature of the sprayer used in each setup. The LC-ESI sprayer is positioned at the exit of the LC column. It comprises a metal needle connected to a high-voltage power supply. As analytes elute from the chromatographic column, they are mixed with a nebulizing gas, often nitrogen, and delivered to the ESI source.

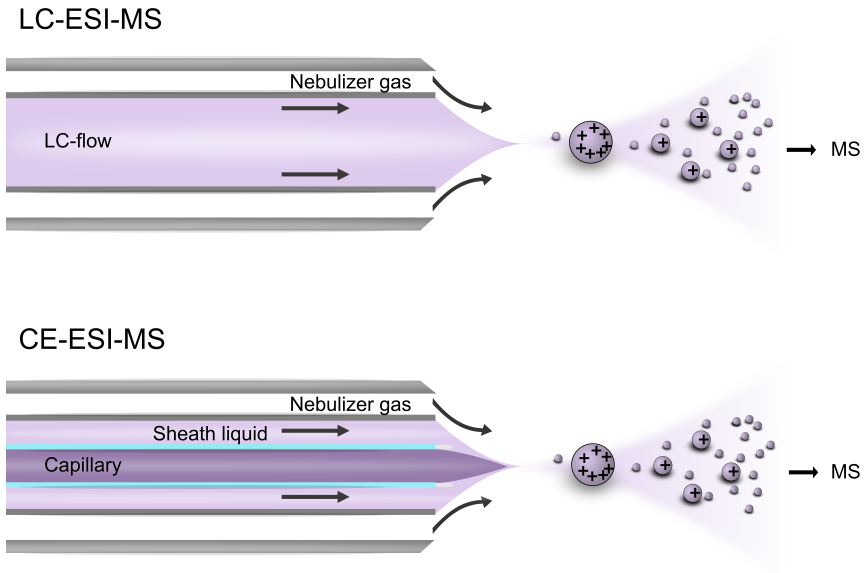


Figure 1.9 Scheme of electrospray ionization on a common LC-ESI-MS interface (top) and co-axial sheath–liquid interface in CE-MS hyphenation (bottom).

Coupling CE to the mass spectrometer introduces distinctive challenges compared to LC. In CE-ESI-MS, the outlet position is located in the ESI sprayer and directed into the mass spectrometer. To apply both the CE voltage for separations and the ESI voltage for spray formation, a closed circuit must be formed using a grounded sprayer [217]. The most frequently used interface is the co-axial sheath–liquid interface developed by Smith *et al.*, schematically shown in Figure 1.9 [218]. This interface consists of three tubes, hence it is often referred to as triple tube sprayer. The CE capillary is fixed in the inner tubing which is covered by a constant flow of sheath liquid (SL). In the outer tubing, a nitrogen nebulizer gas is introduced, similarly to in the conventional (LC-)ESI sprayer. Various parameters, such as composition and flow of the sheath-liquid, flow of the nebulizer gas, and positioning of the capillary in the sprayer can influence electrospray stability. In this thesis, a conventional co-axial sheath–liquid sprayer was used for CE-MS hyphenation. Notably, alternative interfaces have been

developed for CE-MS coupling, displaying augmented sensitivity. Examples include the porous-tip sheathless interface, the flow-through micro vial interface, and the nanoflow sheath liquid interfaces [217, 219].

Analytical challenges in MS-based metabolomics

Mass spectrometry is a highly advantageous tool in non-targeted metabolomics due to its high sensitivity and ability to detect hundreds or even thousands of different metabolites in a single run. Nonetheless, this method also presents certain challenges. In the realm of non-targeted metabolomics, the preference leans toward the utilization of high-resolution mass spectrometry (HR-MS) to ascertain the precise mass of a given compound, which can help to predict its molecular composition. It is also important to acknowledge that within complex molecular compositions, the existence of numerous isomers might persist that cannot be resolved by MS^1 data alone. MS^2 spectra prove insightful in potentially distinguishing isomers under certain circumstances; however, their effectiveness can be variable. For instance, lipids that diverge solely in the positioning of their double bonds remain undifferentiated when assessed solely through MS^2 spectra. Nonetheless, the utility of MS^2 spectra persists in metabolite identification, primarily through matching with publicly available databases. However, MS^2 spectra can diverge significantly for the same compound due to disparities in instrument specifications, gas environment, pressure conditions, ionization methods, or collision energies. Furthermore, the availability of an MS^2 spectrum isn't universal; its presence hinges on compound abundance and the specific type of mass spectrometry employed during acquisition. Finally, the quality of the MS^2 spectrum varies greatly depending on factors such as compound abundance and collision energy [220]. Coupling MS with separation techniques such as LC or CE can help to further distinguish isomers in some instances, but it's not always feasible to entirely overcome co-elution issues for all metabolites. Another challenge in MS-based metabolomics is that in certain instances, metabolites may not be detected at all due to factors such as their low abundance, suboptimal ionization properties, limited stability for ionization, ion suppression by co-eluting compounds, short lifetime, poor solublility in the solvent, or no retention at all in LC (e.g., small molecules). Analytical variation can also be introduced by changing column conditions, resulting in RT shifts and column bleeding. Contamination of the ion source is another issue, as well as changes in detector response over time. Consequently, it is a challenge to distinguishing biological molecules from other *m/z* features, stemming from both the MS process and potential contamination. One effective avenue for such discrimination is the adoption of stable isotope labeling approaches [221]. Moreover, a single metabolite may appear as several *m/z* features due to the formation of different adducts, such as $[M+H]^+$, $[M+Na]^+$, and $[M+NH_4]^+$ in positive ionization mode, or $[M-H]^-$, $[M+CHO_2]^-$, and $[M+Cl]^-$ in negative ionization mode. Double charges or in-source fragmentation can also contribute to the formation of multiple *m/z* features for the same metabolite.

Furthermore, MS-based metabolomics analysis yields vast amounts of data, often spanning gigabytes in size. These data contain information on hundreds to thousands of metabolites. In order to extract all important information, these data need proper bioinformatic evaluations tools, which will be discussed in the next section.

1.4 Bioinformatics for metabolomics

Effective interpretation of non-targeted metabolomics data requires proper data processing and analysis. The bioinformatics pipeline for metabolomics comprises several steps following data acquisition, including raw data pre-processing, normalization and corrections, feature filtering using quality control (QC) samples, metabolite identification, statistical analysis, data mining, and integration with metabolic networks to gain insights into underlying biological processes (Figure 1.10). Each of these steps is critical for accurately interpreting metabolomics data and obtaining reliable results [222].

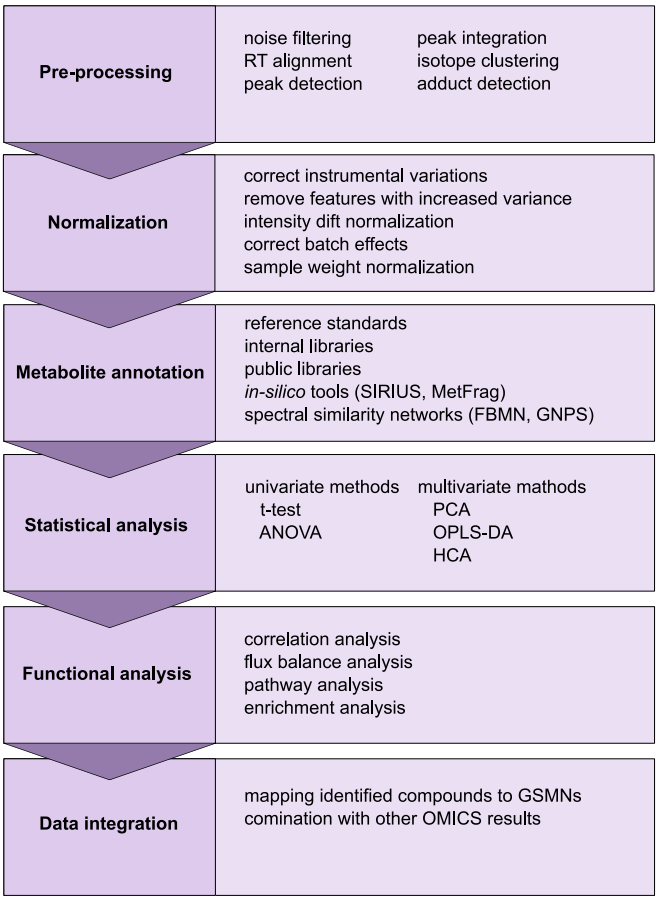


Figure 1.10 Scheme of bioinformatics workflow in non-targeted metabolomics

1.4.1 Data processing

Raw data pre-processing is a critical step in generating a peak or feature table from the raw MS data. It involves a range of processes, including noise filtering, RT alignments, peak detection and integration algorithms, isotope clustering, and adduct detection [223]. Several open-source software are available for processing MS-based metabolomics data, such as XCMS [224], MZmine [225], Skyline [226], and OpenMS [227]. However, these software tools require the conversion of the raw data vendor formats into open-source formats such as mzML or mzXML, which can be done using software such as ProteoWizard's MsConvert [228].
In this thesis primarily Genedata Expressionist for MS was used, which is available under license and

compatible with all vendor formats. The next step in data processing is normalization, which corrects instrumental variations. QC samples can be used to filter features with increased variance, normalize samples for intensity drift that may occur during acquisition, or correct batch effects. Furthermore, the peak table can be normalized by the protein content of each sample to account for different sample weights [223, 229].

1.4.2 Metabolite annotation

The next crucial step after generating a peak or feature table is compound identification, which remains the primary bottleneck in non-targeted metabolomics. The Metabolomics Standards Initiative (MSI) proposed different levels in compound annotation to indicate the reliability of identifications. MSI level 1 represents the highest confidence level where compounds are identified by matching with authentic reference standards acquired with the same experimental setup using orthogonal data such as retention time and mass spectrum. MSI level 2 refers to putative annotations based on physicochemical properties or spectral similarity matching with public databases such as Massbank of North America (MoNA; `https://mona.fiehnlab.ucdavis.edu`), National Institute of Standards and Technology (NIST), METLIN [230], or Global Natural Product Social Molecular Networking (GNPS) [231]. MSI level 3 annotations are putatively characterized compounds and are also based on physicochemical properties or spectral similarity matching with public databases. MSI level 4 represent unknown compounds. To compare, share, and reuse data, it is recommended to report structural compound identifiers such as InChI-key or SMILES in addition to common names [232].
Lipid annotations require a slightly different reporting system. LIPID MAPS proposed a system for lipid nomenclature, classification, structural representation, and initial shorthand nomenclature based on lipid hierarchy. The minimum requirement for shorthand annotation is the acquired precursor *m/z* value, reporting the adduct and the observed fragment ions [233]. It is recommended to report lipids in different levels, i.e. in their lipid class, lipid species, bond type level, fatty acyl/ alkyl level, and fatty acyl/alkyl position level, depending on how confident the annotations are [204]. However, often no explicit structure is available for annotated lipids species, hence no InChI, InChKey or SMILES are available.
The availability of reference standards and spectral libraries is a limitation in the annotation of MS-based non-targeted metabolomics data. To address this issue, various *in-silico* tools have been developed. For example, SIRIUS predicts molecular formula based on the isotopic pattern and includes fragmentation trees as an alternative tool to analyze MS^2 spectra without the need for a spectral database [234, 235]. CSI:FingerID combines molecular structure databases, fragmentation tree computation, and machine learning algorithms to annotate compounds based on their MS^2 spectra [236]. MetFrag retrieves potential candidates from compound databases and uses a bond dissociation approach to fragment them. The observed fragments are compared with the experimentally observed fragmentation peaks and can be used to putatively annotate features based on the best scoring [237, 238].

1.4.3 Network analysis of metabolomics data

Metabolism can be represented as a complex network of interconnected metabolites and enzymatic reactions. Network-based approaches have been widely used to visualize and analyze metabolomics

data, with pathway and network analysis being valuable tools for annotation and interpretation [239–242].

Genome-scale metabolic networks (GSMNs) are constructed using complete genome sequences to reveal enzymes that can be linked to known biochemical pathways. These networks are purely knowledge-based, where metabolites represent the nodes and enzymatic reactions represent the edges. GSMNs can be utilized to analyze metabolomics data by mapping identified metabolites, for instance using MetExplore [243]. They are commonly used in metabolomics to analyze metabolic fluxes and perform centrality analysis to identify key metabolites in a graph [239, 244, 245]. Flux-balance analysis (FBA) can be used to predict changes in the phenotype, such as simulating maximal growth in cells [246]. FBA can differentiate between active metabolic pathways, the level of activity, and how other pathways are influenced [247]. Analysis of metabolic fluxes is useful for identifying drug targets, metabolic engineering, or refining metabolic networks [248]. However, GSMNs have limitations as they are restricted to metabolites linked to enzymes present in the genome annotation, which is often incomplete, such as secondary plant metabolism being frequently missing [244].
Experimental networks, on the other hand, are independent of genome annotations and enable the analysis of compounds beyond what is predicted by genome annotation [249]. Experimental networks are directly derived from acquired mass spectrometry data. Different types of experimental networks exist, including mass difference, spectral similarity, or correlation networks, which will be discussed in more detail.

Mass difference networks

Enzymatic and non-enzymatic bio-reactions lead to gain or loss of elements leading to mass differences between substrate and product. In MS-based metabolomics, mass difference networks are often used to identify potentially biologically important features and reactions. These networks represent detected *m/z* as nodes, and the edges represent the mass difference between pairs of *m/z*. By using mass difference networks, researchers can identify new compounds or infer their molecular formulae by propagating formula differences (corresponding to the mass differences) from known to unknown compounds. The frequency of certain mass differences under different conditions can also provide valuable biological insights [250].
Metabolic reaction databases, such as KEGG, MetaNetX, and MINEs, can be used to extract biotransformations and filter mass differences to generate the network [251–255]. However, using a predefined transformation list may limit the network to certain mass differences, missing other potentially biologically significant ones. Additionally, compounds involved in the same pathway may not be detected by MS due to ion suppression, concentrations below the limit of detection, or short intermediates' lifetime. To connect compounds separated by multiple biochemical reactions, combinations of mass differences are often included in the transformation list. For instance, the addition and subtraction of alkyl chains of different lengths are commonly used [249].
Chemical sub-units shared by molecules tend to show a retention time trend in chromatography. For instance, an increase in the alkyl chain between lipid species is represented by an increase in CH_2 units and increase in RT, forming a homologous series. Therefore, homologous series can aid in annotation, particularly for lipids [241, 249, 256].
Fully connected mass difference networks have also been used to analyze non-targeted metabolomics data. However, interpreting these networks can be challenging as they may contain connections with non-biological features. Therefore, researchers often filter them and integrate them with other infor-

mation, such as retention time trends and high correlations [257].
Various bioinformatic tools, including *mzGroupAnalyzer*, *MetaNetter*, and *MetNet*, exist for constructing mass difference networks [251, 258, 259]. In this thesis, the R package *MetNet* was used to generate mass difference networks based on a predefined list of transformations, enabling easy integration with different types of experimental networks, such as correlation or spectral similarity.

Spectral similarity networks

Metabolites in the same biological pathway often differ by the addition or subtraction of a few atoms, making them chemically and structurally very similar. Chemically similar compounds, such as metabolites and lipids with the same core structure or class, tend to exhibit similar fragmentation patterns. For this reason, spectral similarity networks based on MS^2 data are one of the most prominent tools for analyzing non-targeted metabolomics data.
MS^2 spectra can be similar not only because the same *m/z* peaks appear in two spectra. Compounds of the same class can have the same neutral losses, resulting in different *m/z* peaks. The wide variety of metabolites and their different fragmentation behavior necessitate different approaches to computing spectral similarity. There are different algorithms to compare two spectra and evaluate the similarity score, where the higher the score, the more similar the spectra are.
The standard cosine similarity score considers exact *m/z* matches between two MS^2 spectra (Figure 1.11), making it easy to understand, but does not work for most metabolite classes. Modifications to the cosine that take into account the mass differences between the precursor masses of two MS^2 spectra improve the quality of the spectral similarity score. This mass difference is taken into account when comparing fragments, in order to find peaks with the same neutral loss [260].

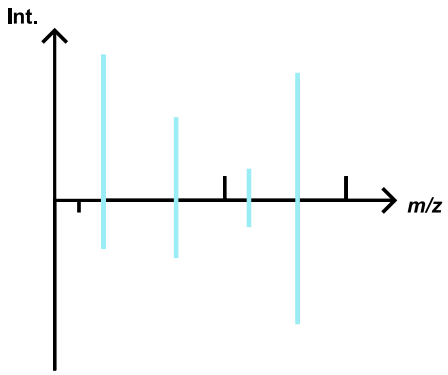


Figure 1.11 Scheme of cosine similarity matching: exact *m/z* matches (displayed in blue) and intensities contribute to a high cosine similarity score.

Molecular Networking (MN) is a bioinformatic tool used in metabolomics that makes use of spectral similarity networks to identify and annotate unknown metabolites. The fundamental principle behind this tool is that structurally similar compounds tend to cluster together in a network based on the similarity of their mass spectra. This makes it possible to visualize complex metabolite profiles and identify structurally related metabolites even if they have not been previously characterized or do not have a corresponding reference standard. Additionally, MN can detect minor changes in the metabolome that may be indicative of a specific biological condition or phenotype.
The most prominent tool being used today is Feature based molecular networking (FBMN) developed

by the Global Natural Product Social Molecular Networking (GNPS; `http://gnps.ucsd.edu`) community [231]. Compared to classic MN, where all MS^2 data are used independent of e.g. their retention time, FBMN incorporates information from MS^1 data, such as peak intensity or retention time (RT), to filter out isomeric structures. Intensity differences between features can be used to interpret the data biologically [261]. In FBMN, adducts of the same species tend to cluster in separate subnetworks. However, they can be detected and combined by approaches such as ion identity networking, which takes into account shape correlation of MS^1 data [242].
There are several algorithms other than cosine and modified cosine that can be used to estimate spectral similarity. For example, *Spec2Vec* uses fragmental relationships established by machine learning to compute spectral similarity and has been shown to perform better than cosine scores [262]. The hypothetical neutral loss spectra algorithm has also been developed to improve spectral similarity evaluation. Core Structure Based Search (CSS) calculates the mass difference between pairs of fragments and uses it for spectral similarity matching. It has shown improved results [263].
The main limitation of spectral similarity networks is the requirement for features with available MS^2 spectra. DDA is mostly used in non-targeted metabolomics, where not all molecular features are selected for fragmentation. This means that not all features appear in the spectral similarity network, highlighting the need for combination with other types of networks, such as mass difference or correlation, to assess all metabolite features.
In conclusion, spectral similarity networks and molecular networking are important tools in metabolomics research that can aid in the identification and annotation of unknown metabolites. By incorporating these tools into bioinformatic pipelines, researchers can improve their chances of successfully identifying novel metabolites in *C. elegans* and other organisms.

Correlation networks

Metabolites along the same biological pathway are codependent and influence each other's concentration. Therefore, correlation networks are a popular tool for analysing non-targeted metabolomics data as they can capture these interdependencies. Correlations are calculated between the intensities of pairs of all features within the acquired samples. When a sufficient correlation is reached between two molecular features, an edge is drawn.
There are several algorithms for calculating the correlation between two features, with Pearson's correlation being the most prominent. However, in metabolomics, full Pearson correlation leads to dense networks due to the high connectivity of pathways, which are not easy to analyze and interpret. Therefore, partial Pearson correlation is often used to correct for indirect correlations. Gaussian graphical modeling, as used by Krumsiek *et al.*, is one example of this approach, which helps to create biologically meaningful networks based on direct interactions [165].
While correlation networks are particularly suitable for large sample sizes, such as cohort studies [165], relationships may appear where none exist for experiments with smaller sample sets. In such cases, the number of features (dimension p) is much larger than the number of samples (observations n), and the correlation is likely to be overestimated because the sample covariance matrix is ill-conditioned. To address this issue, different estimators of the correlation coefficient can be used for the mathematical solution of the covariance matrix [264].
Several packages are available in the R environment to compute correlations, including *corr*, *hmisc*, and *psych* [265]. For the direct analysis of non-targeted metabolomics data, different tools such as

DiffCorr or weighted correlation network analysis (WGCNA) are available [266, 267]. In this study, the *MetNet* R package was used to calculate correlations [259].

1.5 Motivation and aim of the thesis

Understanding the metabolic processes in different organisms is essential for developing new treatments for various diseases. The aim of this dissertation is to enhance our understanding of the *C. elegans* metabolome by identifying both, known and unknown metabolites using innovative approaches in metabolomics. Specifically, this research focuses on the use of capillary electrophoresis-mass spectrometry (CE-MS) to enhance the coverage of the *C. elegans* metabolome and the development of an automated annotation workflow, specifically for new metabolites classes, such as glycerophospho N-acyl ethanolamides (GPNAEs). An overview of the aim and structure of the thesis is given in Figure 1.12.

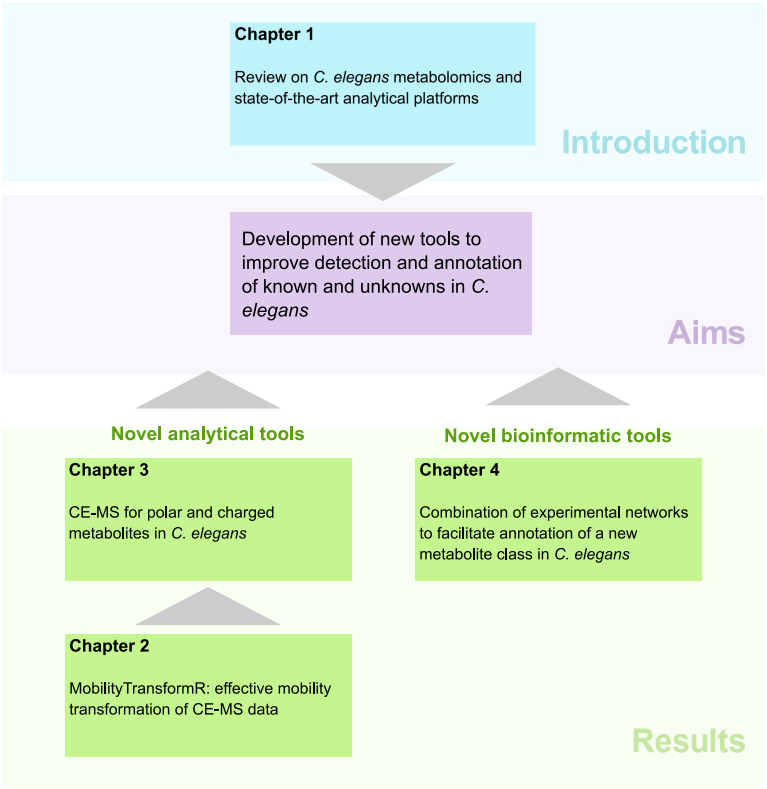


Figure 1.12 Aims and overview of the thesis

The objective of this research is to investigate the effectiveness of CE-MS as a tool for metabolomics analysis in *C. elegans*, particularly for highly polar and charged metabolites that are challenging to analyze using traditional techniques like NMR or HILIC-MS. We developed the *MobilityTransformR*, which is an R package that integrates with existing metabolomics pipelines in the R environment and leverages EOF markers to perform effective mobility transformation and applied it to CE-MS metabolomics data of *C. elegans*. We will compare the effectiveness of CE-MS to HILIC-MS in order to determine its strengths and limitations. Through CE-MS analysis, it is aimed to identify novel metabolites associated with the long-lived *daf-2* mutant in *C. elegans*.
The second objective of this research is to enhance the annotation of unknown compounds in *C. elegans* using the Annotation Propagation through multiple Experimental networks (APEX) workflow. APEX combines different experimental networks such as spectral similarity and mass difference to-

gether with homologous series to improve the accuracy and comprehensiveness of the annotation process, particularly for complex metabolite classes such as GPNAEs. This objective is significant because it contributes to the development of new analytical tools and workflows for identifying novel metabolite classes, leading to critical insights into underlying biological mechanisms.

The motivation behind this research is to advance our understanding of the complex metabolism of *C. elegans* by applying novel approaches in metabolomics. The findings of this research have the potential to open new areas of exploration in *C. elegans* by investigating metabolic changes in disease models, infection, or toxic conditions.

2 MobilityTransformR: An R package for effective mobility transformation of CE-MS data

Summary: *We present* `MobilityTransformR`*, an R/Bioconductor package for the effective mobility scaling of capillary zone electrophoresis-mass spectrometry (CE-MS) data. It uses functionality from different R packages that are frequently used for data processing and analysis in MS-based metabolomics workflows, allowing the subsequent use of reproducible transformed CE-MS data in existing workflows.* ***Availability and Implementation:*** `MobilityTransformR` *is implemented in R (Version >= 4.2) and can be downloaded directly from the Bioconductor database (https://bioconductor.org/packages/MobilityTransformR) or GitHub (https://github.com/LiesaSalzer/MobilityTransformR).*

This chapter has been published under CC BY 4.0 license as Salzer, L., Witting, M., & Schmitt-Kopplin, P. (2022). *MobilityTransformR: an R package for effective mobility transformation of CE-MS data.* Bioinformatics (Oxford, England), 38(16), 4044–4045. https://doi.org/10.1093/bioinformatics/btac441

Candidate's contributions: L.S. programmed the code, analysed the data, prepared the figures, wrote and revised this application note.

2.1 Introduction

Capillary Zone Electrophoresis coupled to Mass Spectrometry (CZE-MS, often simply referred to as CE-MS) is a powerful analytical method for the analysis of charged and polar compounds. Despite its many strengths, it is less common in metabolomics, compared to liquid chromatography-mass spectrometry (LC-MS). One potential reason might be the lower reproducibility of migration times (MT) compared to retention times (RT), leading to erroneous peak alignment and matching between samples, especially for highly charged analytes [268]. The reason for lower reproducibility of MTs is the fluctuation of the electroosmotic flow (EOF) between runs affecting their dwell time in the capillary and thus the time of their detection. The effective mobility μ_{eff} of an ion being its field strength reduced velocity in the capillary, remains stable in the same electrophoretic system (the same background electrolyte) since it can be regarded as a physicochemical property [269]. This makes the use of an effective mobility scale instead of a MT scale an optimal solution to circumvent the drawback of MT shifts. It has been already shown that the high reproducibility in electrophoretic mobility enabled a better match when identifying unknown compounds in complex mixtures [270]. μ_{eff}-values are thus much more suitable for the annotation of metabolites because they remain stable, even if different capillary lengths and voltages are applied. The effective mobility μ_{eff} can be accessed by determining the velocity of the EOF in each individual measurement represented by the MT of e.g. a neutral charged marker-compound.
Recently, ROMANCE has been developed, a Scala-based software that transforms the MT scale of CE-MS data into an μ_{eff} scale, generating two output `.mzML` files, one for positive and one for negative mobilities [271]. Lately the R environment has gained increased popularity in processing and analyzing MS and metabolomics data with different packages available [272, 273].
In order to work with CE-MS data directly in R and to avoid the need for multiple software, we developed the `MobilityTransformR` R package. This package performs qualitative transformation of MTs to μ_{eff} directly on numeric vectors, `Spectra`-objects and `Spectra`-objects within R for improved analysis of CE-MS-based metabolomics data. However, quantitative conversions by correction of peak areas are not included in this package [274].

2.2 Description

`MobilityTransformR` performs the transformation of migration times using functionalities from the packages `MetaboCoreUtils`, `xcms`, `MSnbase` and `Spectra`, which are commonly used in metabolomics workflows [273]. Supported objects for transformation are numeric vectors, `Spectra`-objects, and `Spectra`-objects from `MSnbase` [275]. The base functionality working on base R objects have been implemented in `MetaboCoreUtils`.

2.2.1 Main functions

The effective mobility transformation is performed on each analysis individually using one or two mobility markers added to the sample. Therefore, `MobilityTransformR` incorporates two separate steps:

1. Determination of MTs of EOF marker(s) - `MobilityTransformR::getMtime()`
2. Effective mobility scale transformation - `MobilityTransformR::mobilityTransform()`

A detailed description of the package usage and R code can be found in the SI. Briefly, the CE-MS data is loaded into R as standard `.mzML` format using functionality from `MSnbase` [276]. `MobilityTransformR::getMtime()` uses this raw data as input in order to determine the MTs of mobility markers. For this purpose, it uses the peak picking algorithm from `xcms` to search for peaks corresponding to EOF markers within a user defined m/z and MT window (specified depending on the selected marker) and generates a data frame storing the MTs of markers and the respective file indices [277].
The functionality to convert the MT into μ_{eff} was recently implemented in `MetaboCoreUtils` but is working on base R objects. The function `MobilityTransformR::mobilityTransform()` however uses more advanced data structures, such as `OnDiskMsnExp`. Depending on the input type, different functions will be used to create outputs in the same format as the input. Besides that, `MobilityTransformR::mobilityTransform()` also requires the previously generated data frame containing the marker MT and also the information if the marker is neutral (mobility = 0) or charged. Origin of the formulas can be found in the publication by Codesido *et al.* [278]. If a single marker is used, following equation is called:

$$\mu_{eff} = \mu_A + \frac{l}{E}\left(\frac{1}{t_M - (t_R/2)} - \frac{1}{t_A - (t_R/2)}\right)$$

where μ_A is the effective mobility (0 if neutral) of the marker and t_A its MT. l is the effective capillary length (in mm), E the applied field strength (in kV/ mm), t_M the migration time that will be transformed and optionally t_R, which is the time of the electrical field ramp.
If two mobility markers are used for the corrections, there is no need to provide information about the capillary length and the applied voltage only t_{EOF}, the MT of a neutral charged marker, as shown in:

$$\mu_{eff} = \mu_A\left(\frac{(t_M - t_{EOF})(t_A - t_R/2)}{(t_A - t_{EOF})(t_M - t_R/2)}\right)$$

2.2.2 Application

In order to showcase the functionality and benefit of `MobilityTransformR`, we applied the transformation process on a set of chemical reference standards acquired on a CE-MS system (detailed methods in SI). Figure 2.1 shows overlaid Extracted ion Electropherograms (EIEs) from a leucine standard in different concentrations (1- 25 ppm) acquired in positive mode as example. The migration times fluctuate a lot (RSD 5.18 %), which makes automatic peak alignment across several samples impractical. To normalize for this large drift we spiked paracetamol as neutral EOF marker and procaine as positively charged marker to each sample in order to perform the transformation using `MobilityTransformR`. Transformation from MT to μ_{eff} reduced the RSD from 5.18 % to 0.65 %, highlighting the importance of effective mobility scale transformation of CE-MS data.

2.3 Conclusion

We introduced `MobilityTransformR`, an R package to perform effective mobility scale transformation of CZE-MS data to circumvent migration time shifts. The functions integrate different functionalities from existing R packages that are commonly used in (LC-)MS metabolomics data analysis

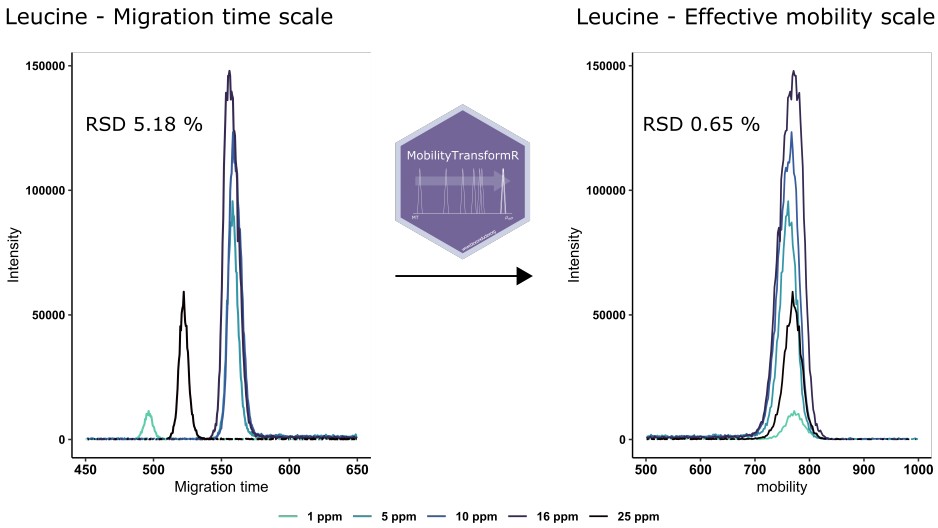


Figure 2.1 Extracted ion electropherogram (EIE) of 5 different CE-MS runs of Leucine at different concentrations (1 - 25 ppm). The left EIE is in migration time scale (seconds) and the right EIE in effective mobility scale ($mm^2/kV*min$)

in order to easily incorporate transformed CZE-MS data to the existing pipelines, which increases reproducibility and metabolite identification rates.

Data availability
R code is available at Bioconductor database (https://bioconductor.org/packages/MobilityTransformR) or GitHub (https://github.com/LiesaSalzer/MobilityTransformR).
CE-MS test data is available at Bioconductor database (https://bioconductor.org/packages/msdata)

3 Capillary electrophoresis-mass spectrometry as a tool for *Caenorhabditis elegans* metabolomics research

Introduction *Polar metabolites in Caenorhabditis elegans (C. elegans) have predominantly been analyzed using hydrophilic interaction liquid chromatography coupled to mass spectrometry (HILIC-MS). Capillary electrophoresis coupled to mass spectrometry (CE-MS) represents another complementary analytical platform suitable for polar and charged analytes.*

Objective *We compared CE-MS and HILIC-MS for the analysis of a set of 60 reference standards relevant for C. elegans and specifically investigated the strengths of CE separation. Furthermore, we employed CE-MS as a complementary analytical approach to study polar metabolites in C. elegans samples, particularly in the context of longevity, in order to address a different part of its metabolome.*

Method *We analyzed 60 reference standards as well as metabolite extracts from C. elegans daf-2 loss-of-function mutants and wild-type (WT) samples using HILIC-MS and CE-MS employing a Q-ToF-MS instrument.*

Results *CE separations showed narrower peak widths and a better linearity of the estimated response function across different concentrations which is linked to less saturation of the MS signals. Additionally, CE exhibited a distinct selectivity in the separation of compounds compared to HILIC-MS, providing complementary information for the analysis of the target compounds. Analysis of C. elegans metabolites of daf-2 mutants and WT samples revealed significant alterations in shared metabolites identified through HILIC-MS, as well as the presence of distinct metabolites.*

Conclusion *CE-MS was successfully applied in C. elegans metabolomics, being able to recover known as well as identify novel putative biomarkers of longevity.*

Candidate's contributions: L.S. performed experimental analysis, analysed the data, prepared the figures, wrote and revised this research article.

3.1 Introduction

The soil dwelling nematode *Caenorhabditis elegans* (*C. elegans*) is one of the premier model organisms for biomedical research introduced in 1973 by Sidney Brenner [55]. Recently, *C. elegans* scientists gained interest in metabolomics to study changes in metabolism during development, aging, or different disease models [96, 112, 118, 130, 199]. Metabolomics is defined as the systematic measurement and (semi)quantification of metabolites in an organism, biofluid, ecosystem or others at a defined state and time point. The metabolome covers a wide range of metabolites with different polarities, molecular weights, and concentrations, many of which are still unknown [76]. Until now most metabolomics analyses are performed using reversed-phase liquid chromatography coupled to mass spectrometry (RPLC-MS), which is missing much of the polar fraction of the metabolome [208]. Hydrophilic liquid interaction chromatography (HILIC) can be employed for the separation of hydrophilic metabolites, but often suffers from long analysis time and low efficiency if not optimized and used correctly (broad, non-gaussian peaks compared to RPLC). However, charged molecules are often still difficult to analyze due to strong secondary interactions with the stationary phase or undesirable interactions with the column, fittings, and metal tubing. Polar and charged metabolites can be accessed by MS hyphenation of capillary zone electrophoresis (CZE, or often simply referred to as CE), which was already applied for the metabolomics analysis of human plasma and urine, HepG2 cells, or mouse tissue [279–281]. However, up until today and to our knowledge CE-MS has not been used in *C. elegans* metabolomics. Most of the currently known metabolites in *C. elegans* are polar and/or also contain one or more ionizable groups [282], which makes them amenable to CE-MS analysis. However, we note that polar metabolites in *C. elegans* have been analyzed mainly with HILIC-MS [118, 283–285].

In order to enhance our understanding of metabolism in the nematode and delve deeper into unraveling its metabolome and metabolic regulation, we employed CE-MS analysis in *C. elegans* metabolomics. We compared HILIC-MS and CE-MS analysis directly on the separation of a set of 60 polar model metabolites known to be present in *C. elegans*. Our objective was to assess the orthogonality and suitability of both methods for metabolomics studies in *C. elegans*. As proof of concept, we used *daf-2*(*e1370*), one of the most studied mutants in *C. elegans*, showing a prolonged lifetime compared to wild-type (WT) worms. We performed a comparative analysis using CE-MS and HILIC-MS techniques to investigate the metabolic profiles of *daf-2* mutants in comparison to WT worms. Our findings demonstrate that CE-MS represents a valuable method for *C. elegans* metabolomics because of the distinct selectivity, narrower peak widths, reduced required sample volume and increased annotation success of metabolites based on their effective mobility. Through CE-MS analysis, we were able to confirm known biomarkers and identify novel biomarkers associated with longevity in *C. elegans*, thus expanding our understanding of the molecular mechanisms underlying lifespan regulation in this model organism.

3.2 Materials and methods

3.2.1 Chemicals

Methanol (MeOH), chloroform ($CHCl_3$), acetonitrile (ACN), 2-Propanol were of LC-MS grade and has been purchased from Sigma-Aldrich (Darmstadt, Germany), same as Methyl tertiary-butyl ether (MTBE, LC grade).

Ammonium acetate (5 M), ammonium formate (10 M), Bicinchoninic Acid Protein assay Kit were from Sigma-Aldrich (Steinheim, Germany), Ethylsulfate (1 mg/mL), Paracetamol (pharmaceutical primary standard) and Procaine (98%) from Sigma-Aldrich (Taufkirchen, Germany). Acetic acid and formic acid (LC-MS grade) were purchased from Fluka® Analytical (Munich, Germany). Water has been purified using a Merck Millipore Integral 3 system (18.2 MΩ; <3 ppm TOC, Millipore, Germany). Chemical reference standards (Mass Spectrometry Metabolite Library of Standards MSMLS; Bile Acid/Carnitine/Sterol Metabolite Library of Standards BACMLS) have been purchased from Sigma-Aldrich (Taufkirchen, Germany) and have been used according to their respective guidelines.

3.2.2 Preparation of standard solutions

Chemical reference standards of different metabolites were dissolved in appropriate solvents at a concentration of 100 ppm (Appendix B: Table B.1, Table B.2). Mixes of metabolite reference standards were prepared containing in total 60 metabolites, each mixture containing metabolites with unique monoisotopic mass and different concentrations (50, 25, 16, 10, 5, and 1 ppm).

3.2.3 *C. elegans* culturing and metabolite extraction

For metabolomics analysis, wild-type N2 Bistrol and *daf-2(e1370)* mutant worms were used and cultivated according to previously used conditions (culturing details in Appendix B.1.2). Metabolite extraction was performed using H_2O/MeOH/$CHCl_3$ (1/3/1, v/v/v). Briefly, around 2000 worms were suspended in 500 µL extraction solvent and homogenized in a Precellys Bead Beating system (Bertin Technologies, Montigny-le-Bretonneux, France), followed by 15 min incubation in an ice-cold ultrasonic bath. The supernatant was collected and separated in two equal aliquots, in order to have the equal sample for HILIC-MS and CE-MS analysis and evaporated to dryness using a SpeedVac Savant centrifugal evaporator (Thermo Scientific, Dreieich, Germany). Dried extracts were redissolved before CE-MS analysis using 50 µL Paracetamol/Procaine/Ethylsulfate (50 ppm/10ppm/10 ppm) in H_2O and for HILIC-MS analysis using 50 µL H_2O/MeOH/$CHCl_3$ (1/3/1, v/v/v).

3.2.4 CE-MS analysis

CE-MS methods were adapted from Drouin *et al.* [286]. Briefly, CE separations were performed using a 7100 Capillary Electrophoresis System (Agilent Technologies, Waldbronn, Germany) equipped with an 80 cm fused silica capillary (Polymicro Technologies, Phoenix, AZ, U.S.A.) with an internal diameter of 50 µm and external diameter of 365 µM. Before initial use, the capillary was conditioned as follows: flush 5' MeOH, 5' water, 5' NaOH (1 M), 5' water, 25' HCl (1 M), 5' water, 5' HCl (0.1 M), 5' water, and 5' 10% acetic acid in H_2O (v/v), serving as background electrolyte (BGE). Between runs the capillary was flushed 5' with BGE. Metabolite mixes were injected hydrodynamically at 50 mbar for 12 s (~ 13nL). *C. elegans* samples were injected at 50 mbar for 30 s (~ 32nL). Separation was performed at 25 °C and +30 kV for cationic and -30 kV for anionic profiling and a constant pressure of 50 mbar applied.

The CE system was hyphenated to an Agilent 6560 IM-QToF-MS (Agilent Technologies, Waldbronn, Germany) with a Dual Agilent Jet Stream ESI source via a coaxial sheath flow ESI interface using a commercial triple tube sprayer from Agilent Technologies. The sheath liquid was delivered using an Agilent 1260 Infinity Isocratic Pump with 1:100 splitter and composed of H_2O/iPrOH/ formic acid

(50/50/0.5, v/v/v) and 10 µM purine (*m/z* 121.051, [M+H]+ and *m/z* 119.035 [M-H]-) and 2 µM Hexakis(1 H, 1 H, 3 H-tetrafluoropropoxy)phosphazine (HP-0921; *m/z* 922.010, [M+H]+ and *m/z* 919.995 [M-H]-) as lock masses. The sheath liquid was delivered at 3 µL/min and 10 µL/min for cationic and anionic profiling, respectively. Detailed MS parameters can be found in Appendix B.1.3. The MS was operated in QToF only mode and additional MS/MS spectra were acquired for *C. elegans* samples.

3.2.5 HILIC-MS analysis

HILIC separation was performed on an Agilent InfinityLab Poroshell 120 HILIC-Z column PEEK-lined (150 mm x 2.1 mm, 2.7 µm, 100 Å) [287]. Anionic profiling (negative ionization mode): A: H_2O+10 mM ammonium acetate + 2.5 µM InfinityLab Deactivator Additive, pH = 9. B: 10% H_2O + 90% ACN + 10 mM ammonium acetate + 2.5 µM InfinityLab Deactivator Additive, pH = 9. A nonlinear gradient was applied (see details in Appendix B.1.4). Column temperature was 50 °C at a flow rate of 0.25 mL/min. Cationic profiling (positive ionization mode): A: H_2O + 10 mM ammonium formate + 0.1% formic acid. B: 10% H_2O + 90% ACN + 10 mM ammonium formate + 0.1% formic acid. Column temperature was 25 °C at a flow rate of 0.25 mL/min. Sample injection was 3 µL in both ionization modes. Hyphenation to MS was performed on the same instrument as for CE-MS. Exact MS parameters can be found in Appendix B.1.4. The MS was operated in QToF only mode and additional MS/MS spectra were acquired for *C. elegans* samples.

3.2.6 Data processing and statistical analysis

MS data was centroided and converted into .mzML format using MSConvert (3.0.20342) from ProteoWizard (`http://proteowizard.sourceforge.net`). In case of CE-MS data, the migration time scale was transformed into an effective mobility scale, using three EOF markers (Table B.4) spiked into each sample and the R package MobilityTransformR [288]. Peak picking and evaluation of CE-MS and HILIC-MS separation of the model metabolites was performed in R using xcms (3.12.0, `https://github.com/sneumann/xcms`, see details in Table B.5) [277].

Metabolomics data of *C. elegans* samples was processed with Genedata Expressionist for MSMS 13.5.4 (Genedata AG, Basel, Switzerland). Processing included chemical noise subtraction, migration time alignment, isotope clustering, peak detection, and grouping. Data was annotated in R using a workflow utilizing the MetaboAnnotation package [289] as described below and statistical analysis was performed in Genedata Expressionist for MS Analyst module, which included normalization on the protein content and intensity drift normalization. Significantly up- and downregulated metabolites in *daf-2*(*e1370*) were determined using a Welch-test with a p-value < 0.05. Features were putatively annotated using an in-house developed annotation workflow based on MS^1 and MS^2 matching (`https://github.com/michaelwitting/MetaboAnnotationGenedata`). MS^2 matching was performed against Fiehn-HILIC, MassBank, MetaboBASE, GNPS and HMDB, which were downloaded from the Massbank of North America (`https://mona.fiehnlab.ucdavis.edu/`) with a cosine > 0.6. The remaining features were annotated using Sirius, CSI:FingerID using the COSMIC score as additional measure of confidence [290, 291].

3.3 Results and discussion

3.3.1 Comparison of CE-MS and HILIC-MS using targeted metabolites

Comparison of sensitivity, relationship of peak intensities across different concentrations, variability, and peak width

We first compared both analytical setups based on a selection of 60 polar metabolites relevant for *C. elegans*. In case of CE-MS the original migration time (MT) based as well as the effective mobility (μ_{eff}) scaled data were included in this comparison. In CE-MS we observed 45 in positive but only 14 metabolites in negative ionization mode (Figure 3.1A). Potential reasons for low coverage in negative ionization mode was a poor ionization efficiency of the analytes due to a low quality of the electro-spray compared to positive ionization mode, also resulting in higher signal-to-noise ratios. The difference in the settings of the peak picking algorithm explains why we observe variations in the number of detected compounds between the MT scale and mobility scale of the CE-MS data [292]. In HILIC-MS we detected much more compounds (50 in positive and 44 in negative mode), even in the lowest concentration of 1 ppm (Figure 3.1A and B). There are multiple potential reasons for the increased sensitivity observed in HILIC-MS. For instance, the dilution of metabolites caused by the sheath liquid-based interface, or the use of larger sample injection volumes in HILIC-MS [293, 294].

Next, we compared the relationship of the peak-intensity within the measured concentration range. We built a linear model of metabolites that were detected at a minimum of 5 different concentration points. In the measured concentration range, CE-MS showed a much better peak-intensity response of the linear model compared to HILIC-MS (Figure 3.1C). A reason for the curvature of the peak-intensity relationship function in HILIC-MS could be saturation of the MS at higher concentrations. Moreover, peak shapes were often not good in HILIC, especially in the higher concentration range (Figure 3.1D), leading to erroneous peak integration. The better response of the peak intensity indicates that we are working in CE within the linear range of the MS for most metabolites, which is necessary for proper quantification and determination of statistically significant features. If metabolite concentrations are out of the linear range of the MS, features might fall below the statistical threshold and are regarded as non-significant.

We then compared the variability of RT/ MT/ mobility of the model metabolites. The relative standard deviation (RSD) has been calculated between detected metabolites in different concentrations. As previously reported, HILIC-MS shows much more reproducible peaks then CE, having a median RSD of 0.7% across detected RT's (Figure 3.1E) [293]. MT conversion into μ_{eff} shows great improvement of precision. The median RSD decreased from 5.4 to 2.4% after conversion (Fig. B.1). The effective mobility is constant for a substance in the same electrophoretic system (i.e. the same background electrolyte). MTs on the other hand fluctuate much more because of variations in the electroosmotic flow between runs based on difference in sample salt concentrations for example. It has also been shown that effective mobility scale transformation enhances annotation of non-targeted metabolomics data [215, 288, 295]. Lastly, we compared the full peak widths at half maximum (FWHM or $w_{1/2}$) between CE and HILIC (Figure 3.1F). Our findings confirm that CE separations are highly efficient, showing very narrow and sharp peaks, due to decreased analyte diffusion compared to HILIC [296, 297]. A high separation efficiency brings a high resolution which is crucial in non-targeted metabolomics analyses where we have thousands of compounds that are desired to be separated.

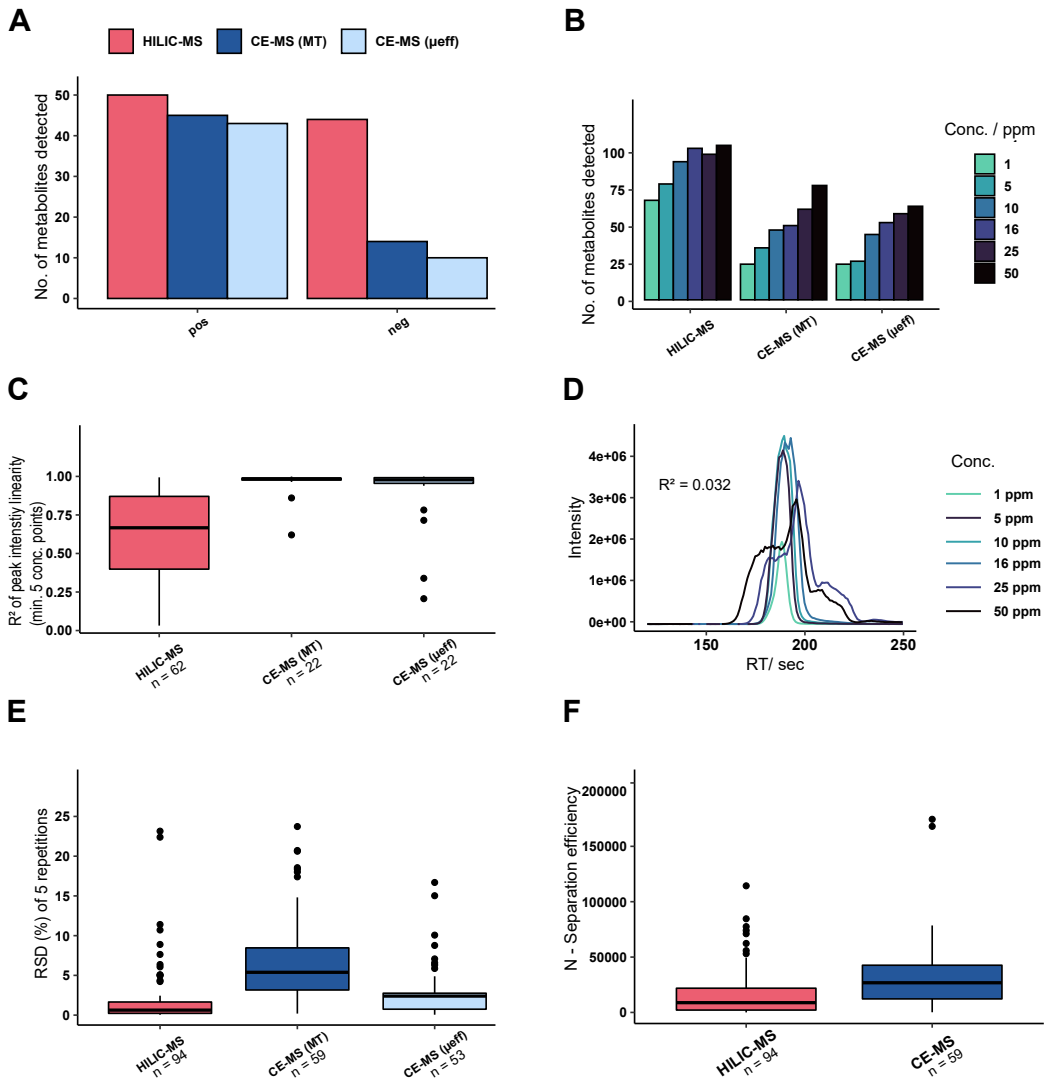


Figure 3.1 Comparison of different descriptors from CE-MS and HILIC-MS. **A**: Number of detected metabolites **B**: Minimum concentration of targeted metabolite that was detected. **C**: R^2 of linear model build on peak-intensity relationship across different concentrations **D**: Extracted ion chromatogram (EIC) of Adenosine of HILIC-(+)-MS. **E**: Relative standard deviation (RSD) in percentage of detected metabolites. F: FWHM: full peak width at half maximum

Orthogonality and complementarity of both methods

40 metabolites have been commonly detected in HILIC and CE positive and 10 in negative ionization mode (Table B.1). We compared the RT and MT order of those metabolites (Fig. B.2) to show the complementarity of CE and HILIC. Since all the data points were scattered over the entire plot, it confirms orthogonal separation principles leading to no correlation between HILIC and CE. Moreover, the distinct separation principles of CE and HILIC (Fig. B.4) lead to the observation of different matrix effects and ion suppression during the analysis [298], which might be advantageous for complementary detection, enhancing the discovery of statistically significant features in metabolomics studies. Moreover, reduction of the matrix effects makes it possible to load less sample compared to HILIC (13 nL vs. 3 µL).

Differences in the separation are due to the different separation mechanism, which is in HILIC more complex and includes different interactions from electrostatic, hydrophilic, and ionic effects. The separation in CE is based on the mobility of a molecule which is defined by their size and the charge as the electric field strength normalized velocity of the ions [293, 298]. The complementary of CE to HILIC is a crucial assumption for the following experiments in *C. elegans* to find novel metabolites.

In summary, we observed a different selectivity, better peak-intensity response across different concentrations and narrower peak widths compared to HILIC. This is important to find potential novel biomarkers which we attempted in the next step where we applied CE-MS analysis in *C. elegans* metabolomics.

3.3.2 CE-MS vs. HILIC-MS in *C. elegans* metabolomics

Comparison of CE-MS and HILIC-MS

In order to evaluate the usability of CE-MS for *C. elegans* metabolomics, we compared HILIC-MS and CE-MS analysis of *C. elegans* extracts of *daf-2* loss-of-function mutants and WT worms. In HILIC-MS we observed more features with a higher intensity (Table 3.1). In HILIC 3 µL of the sample with an approximate calculated concentration of ~20 worms/ µL was injected, which is equivalent to the amount of 60 worms per injection. In CE on the contrary, only 32.2 nL of the sample is injected, corresponding to a calculated value of only 0.6 worms/ injection.

	HILIC neg	**HILIC pos**	**CE neg**	**CE pos**
Features	5684	5667	3534	3616
Significants	1069	1659	615	1480
Exact mass matches in the other method	69 (6.5%)	182 (11.0%)	42 (6.8%)	275 (17.4%)
msi 1	8	66	15	60
msi 2	3	18	17	39
msi 3	23	37	5	14

Table 3.1 Number of detected and significant features and number of annotations at different msi levels

Comparing the *m/z* distribution of the feature tables (Figure 3.2), we observe slightly more features in the lower mass range (below 250 Da) in CE-MS, which could be related to the higher electrophoretic mobility of smaller molecules. Even more, we observe an increase in the number of features in the higher *m/z* range 900–1100 Da in CE negative separation and ionization mode. We then compared significant features in CE and HILIC by exact mass matching in order to find possibly identical metabolites in both methods. Table 3.1 shows that more than 80% or 90% of the significant molecular features are uniquely present for each method. This number is probably even higher, since exact mass matching does not consider isomeric structures as for example the amino acids leucine, isoleucine and norleucine, having the same molecular formula ($C_6H_{13}NO_2$) and therefore the identical exact mass.

Annotation is the major bottleneck in non-targeted metabolomics [299, 300]. We annotated the metabolites based on the rules of the Metabolomics Standards Initiative (MSI) in different annota-

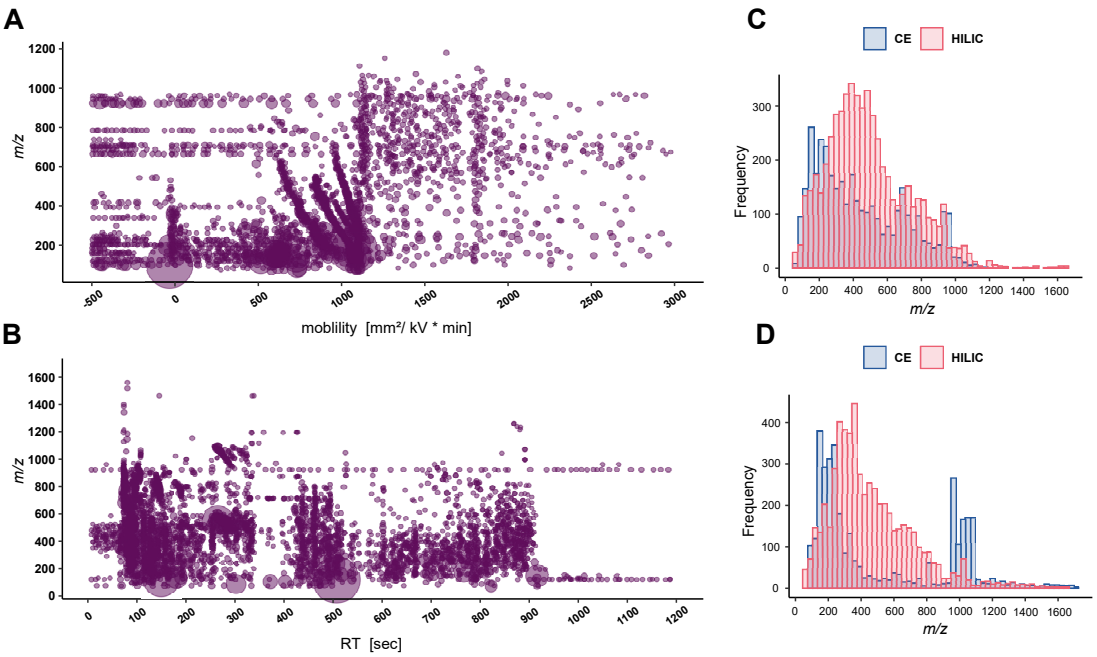


Figure 3.2 *m/z* vs. mobility/RT distribution of molecular features of a pooled quality control (QC) sample in positive ionization and separation mode at **A** CE-MS analysis and **B** HILIC-MS analysis, size represents the relative intensity. *m/z* distribution of features of CE and HILIC in **C** positive and **D** negative ionization (and separation) mode

tion levels (Table 3.1) [154]. In CE-(+)-MS we had more (MSI level 1) identifications, probably due to the high separation efficiency, reduced matrix effects, and reduced ion suppression of (in HILIC) co-eluting compounds. Even with the poor spray in CE-(-)-MS we were able to identify a similar number of metabolites compared to HILIC-(-)-MS. Lower confidence annotations (MSI level 2) are based on physicochemical properties or spectral similarity matching with public MS^2 libraries [154]. Because the effective mobility of a molecule can be regarded as a physicochemical property in the same electrophoretic system [301], CE gives us an increased chance of MSI level 2 annotations in combination with publicly available effective mobility libraries, such as CEUMASS (`http://ceumass.eps.uspceu.es/mediator/cems_effmob_cesearch.xhtml`) [215]. Indeed, our CE-MS MSI level 2 annotations were mainly based on the effective mobility library matching, highlighting the importance of effective mobility scale transformation to normalize for differences between runs and in the analytical setups.

In CE-MS we uniquely annotate 41 compounds at high certainty (MSI level 1 and 2) such as amino acid and related molecules, purines and pyrimidines, and metabolites involved in the nicotinamide pathway (i.e., nicotinamide, nicotinic acid, NAD+, tryptophan) [302, 303]. In HILIC-MS on the other hand we uniquely annotated 20 features i.e., different carnitines, molecules related to sugar metabolism and amino acid metabolism, purines and pyrimidines. Compounds detected exclusively in CE-MS were smaller and more polar (with lower LogP values) in comparison to HILIC-MS. This observation aligns with the overall trend of smaller *m/z* values, as depicted in Figure 3.1 (Table B.3).

CE-MS in *C. elegans* metabolomics to find new biomarkers of longevity

One of the most studied mutants in *C. elegans* is *daf-2*, which encodes for the orthologue of the insulin/insulin-like growth factor (IGF) receptor. *daf-2* worms show prolonged lifetime compared to wild-type (WT) worms and are therefore often used to study longevity and metabolic alterations [168]. Mostly NMR but also HILIC-MS has been used to analyze polar metabolites in *daf-2* [96, 99, 100, 284, 304], showing changes in the amino acid profile (general decrease but increase in BCAAs) and carbohydrate metabolism such as an increase in trehalose, which has been directly linked to longevity [305].

Both methods, CE-MS and HILIC-MS were able to separate the two genotypes from each other in an unsupervised principal component analysis (PCA), meaning both methods are suitable to detect molecular differences between *daf-2* and WT worms (Figure 3.3).

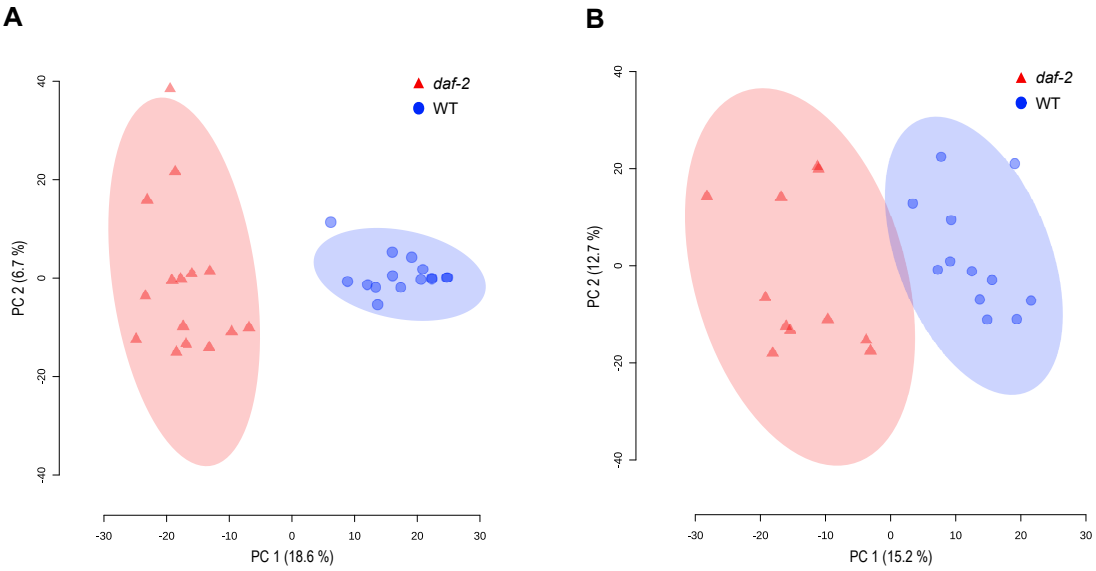


Figure 3.3 PCA scores plot **A** CE-MS positive separation and ionization mode and **B** CE-MS negative separation and ionization mode; Data was normalized on intensity drift based on QC samples, protein content, and log10 transformed

Similarly, as previous results, our HILIC-MS analysis showed changes in the amino acid profile (upregulation in arginine, betaine, histidine, cystathionine and but decrease in tyrosine, isoleucine, alanine, kynurenine, aminoadipic acid). Beyond that, we found changes in purine metabolism intermediates as increased levels in guanine and deoxyadenosine and decreases in guanosine, uracil, and inosine. These changes suggest an alteration in energy metabolism and nucleotide salvage pathways in long lived mutants [143, 284].

However, compared to literature our HILIC-MS analysis showed opposing outcomes in the levels of xanthine and adenine, which were increased in *daf-2* compared to the control group. And even more isoleucine was decreased, whereas it was found to be upregulated in *daf-2* using NMR. Divergent results of our analysis compared to the literature could be explained by differences in the *C. elegans* cultivation, such as differences in the media, food supply (strain, type and amount of bacteria), and actual age of the worm as it was shown previously for lipids [170].

CE-MS analysis revealed similarly as in our HILIC-MS analysis an increase in uric acid, adenine, adenosine, betaine, cystathionine, phenylalanine and spermidine and a decrease in guanosine. However, different from HILIC-MS analysis, CE-MS revealed downregulation in alanine and propionyl

carnitine. As already described before, 41 metabolites have been uniquely detected in CE-MS, e.g. we found changes in metabolites related in amino acid metabolism as increased levels of glycyl-proline, valine, serine, proline, N-methyl aspartic acid, trans-aconitic acid, glutamic acid, aspartic acid, taurine, N-amidino-aspartic acid and urocanic acid and downregulation in tryptophan. Moreover, CE-MS analysis showed alterations in purine and pyrimidine metabolites as hypoxanthine, paraxanthine, purine, cytidine and adenosine monophosphate. Our results in CE-MS strengthen the hypothesis of reorganization of the amino acid metabolism in long-lived worms. But different from our HILIC-MS results and from literature, most amino acids were found to be downregulated. Next, we compared metabolites that were commonly detected but were found to be significantly changed in *daf-2* in CE-MS but not in HILIC-MS. The reason why we observe discrepancies in the significant features might be the potential saturation of the MS signals. As discussed above, we observed a better peak-intensity response in CE of reference standards at different concentrations, compared to HILIC. This observation led us to postulate that there might be a potential for missing significant biological differences in a direct comparison. This hypothesis was subsequently confirmed. For example, the metabolites AMP, glutamic acid, guanidinosuccinic acid, N-acetylgalactosamine / N-acetylmannosamine, N-acetylserine, proline, serine and valine were significantly changed in CE but not in HILIC.

Ion suppression due to matrix effects can also influence the significance of features. Proline for example had a retention time of 791 s. In order to get a first impression of the matrix effect or co-eluting compounds, we determined the number of compounds eluting in the RT window 786–796 s. Indeed, our feature table revealed that within this RT range 87 features have been detected. The determined effective mobility of proline was 480 $mm^2/kV*min$ and in the range of 450–510 $mm^2/kV*min$ only 49 further features have been detected, strengthening the hypothesis of reduced ionization efficiency of e.g. proline due to ion suppression of co-eluting compounds. This effect could explain why it is not significant in our HILIC-MS analysis, whereas it was in CE-MS.

The BCAAs (leucine, isoleucine, and valine) have been reported as key metabolites significantly upregulated in *daf-2* mutants [96, 133]. Our HILIC-MS separation was able to detect and separate valine, leucine and isoleucine. However, changes in valine and leucine were not significant and isoleucine was found to be decreased in *daf-2*, questioning the reliability of the method. In our CE-MS analysis, we failed to detect leucine, but isoleucine and valine were detected and upregulated in *daf-2* mutants, which is consistent with previous findings [96, 133]. Nonetheless, these previous findings rely on NMR, which, compared to MS methods, do not suffer from ion suppression. This example shows the benefit of CE-MS and that it was possible to find the important *daf-2* longevity markers, whereas our HILIC-MS method was not capable of.

Using CE-MS, we observed different metabolite features that have been significantly changed in the long-lived mutant, including increase in most amino acids and related metabolites such as glutamic acid, guanidinosuccinic acid, aspartic acid, proline, serine, or urocanic acid, which is consistent with the literature indicating increased protein catabolism or decrease in protein anabolism and decrease in amino acid dependent energy expenditure [133, 306]. Also, purine salvage intermediates were upregulated in *daf-2*, i.e., AMP, which already has been reported and uridine, reported for the first time. For the first time also a decrease in hypoxanthine has been detected using CE-MS, strengthening the hypothesis of downregulation of purine degradation intermediates of long lived worms, which means altered regulation of nucleotide metabolism [284]. Lastly, we also found an increase in nicotinamide, which is the precursor for the coenzymes NAD+ and NADP+ which are required in many biological pathways and decrease in tryptophan in *daf-2* pointing to an altered tryptophan-nicotinamide path-

way. Interestingly, higher levels of nicotinamide and related molecules (nicotinic acid, NAD+) have already been linked to longevity in *C. elegans* but here, for the first time in *daf-2* [307].

3.4 Conclusions

In this study we systematically evaluated the usability of CE-MS as tool in *C. elegans* metabolomics as a complementary platform to HILIC-MS to analyze polar metabolites. In conclusion, CE separations showed narrower peak widths, a better peak-intensity response of different concentrations pointing to less saturation of MS signals, and a different selectivity compared to HILIC-MS, making it amenable to a different part of the polar metabolome. Researchers have successfully enhanced the sensitivity of CE-MS through the utilization of innovative techniques such as nanoflow sheath liquid or sheath-less coupling approaches [219, 308, 309].

Metabolomics analysis of *C. elegans daf-2* mutants revealed differences in detected metabolites between CE-MS and HILIC-MS. Even more, for some metabolites that were commonly detected the significance changed due to the poor peak-intensity response of those compounds in HILIC-MS. We annotated 41 metabolites uniquely in CE-MS, among others novel metabolite features that have been significantly changed in *daf-2* such as upregulation of purine salvage intermediates such as uridine and downregulation of purine degradation intermediates such as hypoxanthine. Lastly, we also found an altered tryptophan-nicotinamide pathway in *daf-2* which gives us new insights into understanding the mechanism of longevity in *C. elegans*.

Data availability
Data are currently in curation in the open-repository Metabolights under MTBLS6440.

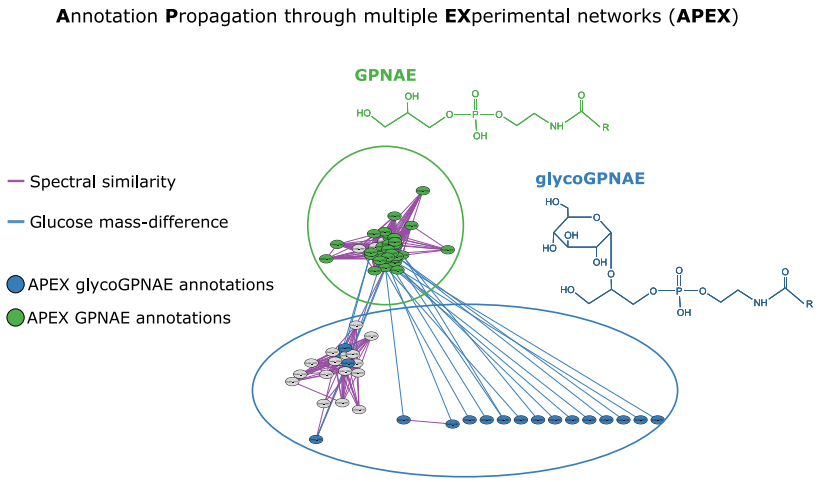
4 APEX – an annotation propagation workflow through multiple experimental networks to improve the annotation of new metabolite classes in *Caenorhabditis elegans*

Spectral similarity networks, also known as molecular networks, are crucial in non-targeted metabolomics to aid identification of unknowns aiming to establish potential structural relation between different metabolite features. However, too extensive differences in compound structures can lead to separate clusters, complicating annotation. To address this challenge, we developed an automated Annotation Propagation through multiple EXperimental networks (APEX) workflow, which integrates spectral similarity networks with mass difference networks and homologous series. The incorporation of multiple network tools improved annotation quality, as evidenced by high matching rates of molecular formula derived by SIRIUS. The selection of manual annotations as the Seed Nodes Set (SNS) significantly influenced APEX annotations, with a higher number of seed nodes enhancing the annotation process.

We applied APEX to different Caenorhabditis elegans metabolomics datasets as a proof-of-principle for the effective and comprehensive annotation of glycerophospho N-acyl ethanolamides (GPNAEs) and their glyco-variants. Furthermore, we demonstrated the workflow's applicability to two other, well-described metabolite classes in C. elegans, specifically ascarosides and modular glycosides (MOGLs), using an additional publicly available dataset.

In summary, the APEX workflow presents a powerful approach for metabolite annotation and identification by leveraging multiple experimental networks. By refining the SNS selection and integrating diverse networks, APEX holds promise for comprehensive annotation in metabolomics research, enabling a deeper understanding of the metabolome.

Candidate's contributions: L.S. developed the APEX algorithm, analysed the data, prepared the figures, wrote and revised this research article.



4.1 Introduction

Networks have emerged as a powerful formalism for modeling and analyzing complex systems of interacting elements. A network is a collection of nodes connected by edges that represent interactions or relationships between them. When a particular phenomenon (such as metabolism) can be modeled as a network, the topology of such a network can be used to study the phenomenon. There are different ways to build metabolism-related networks, but they can be broadly divided into knowledge-based and experimental networks [249].

Knowledge-based networks, such as the Genome-Scale Metabolic Network (GSMN) aggregate knowledge about the metabolism of a specific organism, e.g., Human or *Caenorhabditis elegans* [161, 310, 311]. GSMN are constructed on the basis of annotated genomes. In contrast, experimental networks can be generated from metabolomics data, i.e., holistic measurements that systematically measure and (semi)quantify all metabolites present in a given sample, aiming to correlate changes in metabolite intensities or concentrations with physiological phenotypes. Known and unknown metabolic reactions and pathways might be reconstructed from metabolomics data. Here, we are focusing on liquid chromatography-tandem mass spectrometry (LC-MS/MS) based non-targeted metabolomics, with which different experimental networks can be (re)constructed, such as mass difference and spectral similarity networks; all of which can be used to interpret the obtained metabolomics data.

Mass difference networks use exact *m/z* values and the corresponding pairwise mass differences (represented as edges) [312–314]. These mass differences are compared against a list of known mass differences corresponding to biochemical transformations of interest. For instance, a mass difference of 15.9949 could indicate the gain or loss of an oxygen atom, and if such biochemical transformation is of interest, a connection between the corresponding nodes is added. However, although the mass difference between a pair of metabolite features could be explained by the biochemical transformation link between them, it is not necessarily the case. Two metabolite features may have the same mass difference as a biochemical transformation of interest only by chance or because of unrelated biochemical transformations. Different techniques could be used to improve the quality of a mass difference network (i.e., to reduce the false positive edges), for instance, homologous series. Homologous series are a group of compounds that differ from each other by a specific repeating unit, such as a CH_2 group in a homologous series of fatty acids [241]. Retention time is used to identify those

homologous series based on a consistent trend observed in liquid chromatography (LC) separations, for example in reversed phase separations, wherein the retention time tends to increase as the number of e.g., CH_2 units increase. If unknown metabolites are connected in the mass difference network and are part of a homologous series, it provides strong evidence for the identities of the unknown metabolites. It is to note that, although mass difference networks can be a useful tool for metabolite annotation and identification, they don't consider any structural relation between metabolites.
In contrast, spectral similarity networks, also known as molecular networks, are based on spectral patterns from fragmentation experiments, which can incorporate some aspects of chemical structural information and can provide more accurate metabolite annotation. The nodes represent corresponding MS^2 spectra of metabolite features, which are compared by their spectral similarity, with different scoring metrics available [231, 260]. An edge between two nodes is drawn if the spectral similarity between the corresponding fragmentation spectra is above a specific threshold. It is to note that the topology of a spectral similarity network is dependent on the metric used for comparing the MS^2 spectra and the threshold. Thus, if spectra are too dissimilar, no connection can be added, even if potential biochemical connections exist.
Correct annotation and identification of metabolites in non-targeted metabolomics remains as one of the primary challenges in the field and different types of networks, covering different aspects of the biology serve as valuable tools to aid in this task; especially when combined. In the present work, we introduce an automated **A**nnotation **P**ropagation through multiple **EX**perimental networks (APEX) workflow. Our workflow combines spectral similarity networks with mass difference networks and application of homologous series. The aim of APEX is to bridge between different types of networks and to allow propagating the annotations beyond a single network to uncover new potential biological links useful for metabolite annotation and identification as well as biological interpretation.
In this study, we utilize the APEX workflow to aid in the identification of glycerophospho *N*-acyl ethanolamides (GPNAEs), a recently discovered compound class in *Caenorhabditis elegans* (*C. elegans*), that has been identified in starved larvae and peroxisomal α-oxidation mutants [315, 316]. GPNAEs are intermediates in the synthesis of *N*-acyl ethanolamines (NAEs), which are linked to lifespan extension in the nematode [124]. Due to their recent discovery, no deposited reference spectra, and no chemical reference standards for GPNAEs are yet commercially available. Notably, a characteristic fragmentation pattern of GPNAEs and its acyl chain variants connect the corresponding nodes in spectral similarity networks as performed by Helf *et al.*. However, changes in their structure upon specific biochemical transformation (glycosylation) changes abundance of common fragments and introduces new fragment peaks as well, which results in separated clusters in molecular networks, limiting the ability to propagate annotations between them [316]. Nevertheless, the combination of different experimental networks allows to bridge between such seemingly unrelated clusters.
Using the APEX workflow, we improved species identification within the GPNAE compound class in different *C. elegans* datasets. We also evaluated its effectiveness for annotating species from two other *C. elegans* metabolite classes, namely ascarosides, and modular glycosides (MOGLs). Our results highlight the potential of APEX for annotating GPNAEs and its implications for future metabolomics research.

4.2 Material and Methods

4.2.1 Chemicals

Methanol (MeOH), iso-propanol (iPrOH), acetonitrile (ACN) and formic acid have been of LC-MS grade and purchased from Sigma-Aldrich (Sigma-Aldrich, Taufkirchen, Germany). Water was purified from a Millipore Integral 3 water purification system with a TOC < 3 ppb and > 18.2 MOhm.

4.2.2 *C. elegans* culture

The *C. elegans* N2 strain was maintained on nematode growth medium at 20°C according to the routine protocol [317]. *Pseudomonas lurida* MYb11, *Pseudomonas fluorescens* MYb115, and *Escherichia coli* OP50 were grown on Tryptic Soy Agar (TSA) at 25°C. Worms were grown on 9 cm Peptone Free Nematode Growth Medium (PFM) plates with a bacterial lawn (OD600=10) of either MYb11, Myb115, or OP50 at 20°C for at least two generations. Four biological replicates were used for each treatment group. Each replicate consisted of 1000 to 1500 synchronized hermaphrodites at the first larval stage (L1) pipetted onto the bacterial lawns. Two days later, the worms were transferred to plates containing OP50. Worms were harvested after 24 h by thoroughly washing each plate with chilled M9, followed by centrifugation at 3500 rpm for 1 min. The pellet was collected and washed four more times. Finally, the pellets were transferred into 1 ml of H_2O/MeOH (50/50, v/v) and flash-frozen in liquid N_2.

4.2.3 Metabolite extraction

After worm samples have been thawed on ice, they were transferred to bead beating tubes and homogenized using a Precellys beat beating system with a Cryolys cooling module (Bertin Technologies). After homogenization, samples were centrifuged for 15 minutes at 15,000 rpm at 4°C. The supernatant was transferred to a fresh reaction tube and evaporated to dryness using a Speedvac (Thermo Savant). Samples were stored dry at -80°C until analysis. From the residue protein quantities were determined using a bicinchoninic acid (BCA) kit (Sigma). Prior to analysis samples were redissolved in 50 µL 80% H_2O / 20% ACN. 40 µL were transferred to an autosampler vial and 10 µL from each sample mixed for a pooled quality control (QC) sample.

4.2.4 UPLC-UHR-TOF-MS analysis of *C. elegans* microbiota samples

Metabolite extracts were analyzed on a Waters Acquity UPLC (Waters, Eschborn, Germany) coupled to a Bruker maxis UHR-TOF-MS (Bruker Daltonics, Bremen, Germany). Separation was achieved on a Waters Acquity BEH C18 column (100 mm x 2.1 mm ID, 1.7 µm particle size). Eluent A consisted of 100% H_2O / 0.1% formic acid and Eluent B of 100% ACN / 0.1% formic acid. Gradient conditions were as follows: 95/5 at 0.0 min, 95/5 at 1.12 min, 0.5/99.5 at 6.41 min, 0.5/99.5 at 10.01 min, 95/5 at 10.1 min, 95/5 at 15.0 min. Detection was carried out in positive and negative ionization modes, using data-dependent acquisition. MS parameters were as follows: End plate Offset = -500 V, Capillary = -4500 V (positive mode)/4000 V (negative mode), Nebulizer pressure = 2.0 bar, Dry Gas = 8.0 ml/min, Dry Temperature = 200 °C. MS^2 spectra were acquired with data-dependent acquisition using Bruker AutoMSn with default parameters for isolation window and collision energy ramping. For individual recalibration of each chromatogram, 1:4 diluted Low Concentration Tune Mix (Agilent, Waldbronn, Germany) was injected via a six-port valve before each run between 0.1 and 0.3 min.

4.2.5 Data pre-processing

All datasets (*C. elegans* microbiota, MSV000087885 and MSV000086293) were processed the same way, using Genedata Expressionist for MSMS 13.5.4 (Genedata AG, Basel, Switzerland). Processing included chemical noise subtraction, retention time alignment, isotope clustering, peak detection, and grouping. The resulting feature table and corresponding MS^2 spectra were exported and used to build the experimental networks and to manually annotate the metabolite features that were used as seeds for the APEX workflow (i.e, the GPNAE, the ascaroside, and MOGL compounds), as described in the following sections.

4.2.6 Construction of mass difference networks, homologous series, and spectral similarity networks

Mass difference networks were created using the *MetNet* R package [259] and upon mass matching of 10 ppm and 5 ppm tolerance for the qToF and Orbitrap data, respectively using a list of 27 mass difference of biotransformations that might be relevant GPNAE metabolism (Table C.1) and 21 mass differences relevant to ascaroside/MOGL metabolism (Table C.2).
Homologous series have been calculated using the *nontarget* R package (`https://github.com/blosloos/nontarget`), considering only C, H, and O for the mass difference. Even more, the minimum *m/z* difference was 5 Da, the maximum *m/z* difference 60 Da with a tolerance of 5 Da, the minimum RT shift was 12 seconds, the maximum was 60 seconds with a tolerance of 5 seconds, and a minimum of 4 features per homologous series cluster.
The spectral similarity networks were generated using Feature Based Molecular Networking (FBMN) in GNPS [261]. The feature table and MS^2 spectra were formatted to be compatible with XCMS input format for FBMN. Settings were as follows: mass tolerance of 0.02 Da, minimum cosine of 0.8, maximum 1000 neighbor nodes, minimum 3 matched fragment ions, and unlimited component size.
The experimental networks spectral similarity G_s = (V_s, E_s) and mass difference G_m = (V_m, E_m) and homologous series (G_h = V_h, E_h) are then merged into a single network G_apex = (V, E), merging the duplicated vertices (of V_s + V_m + V_h) and corresponding edges E so that there is a single edge between any pair of nodes and saving the number and type of experimental networks merged as an edge attribute.

4.2.7 Overview of the APEX workflow

The automated **A**nnotation **P**ropagation through multiple **EX**perimental networks (APEX) workflow is schematically shown in Figure 4.1 and depicted as pseudocode in Appendix C.1.1.

Overall, the APEX workflow is designed to propagate annotations from manually annotated seed nodes to their first neighbors, using a combination of mass difference, spectral similarity, and homologous series. The resulting annotations are simplified (maximum one annotation per node) to facilitate downstream analysis.
APEX is implemented in R and uses *igraph*, *Spectra*, *MsBackendMgf* and *MetaboCoreUtils* package and is available on GitHub (`https://github.com/michaelwitting/APEX`) together with all relevant data from the example datasets used.

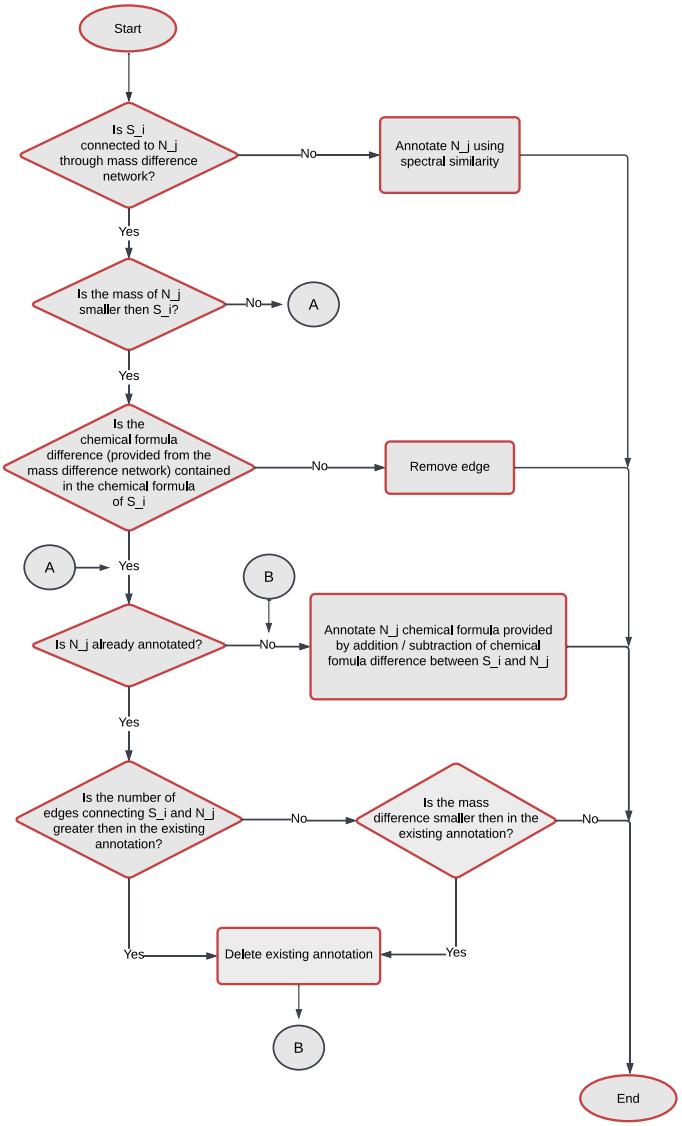


Figure 4.1 Scheme of APEX workflow. A pseudo-code explaining the algorithm can be found in the Appendix C.1.1. The workflow starts by iteration through the manually annotated seed nodes S_i to get their set of neighbors N(S_i). Then the iteration continues through all the neighbors N_j, where the scheme starts. If a neighbor N_j is annotated, new attributes are added to N_j containing the networks connecting N_j and S_i, the seed node S_i, and other available attributes (i.e. values of spectral similarity, mass-difference, and homologous series)

4.3 Results and Discussion

4.3.1 Datasets

To test the effectiveness of the APEX workflow, we used three *C. elegans* metabolomics datasets. The first data set was generated in-house on an UPLC-UHR-ToF-MS using reversed phase (RPLC) separation (*C. elegans* microbiota). The second dataset was taken from Helf *et al.* [316], downloaded from

MassIVE (MSV000087885), and also used to annotate GPNAEs. The third and last dataset was taken from Le *et al.* [318], downloaded from MassIVE (MSV000086293), and used to annotate ascarosides and modular glycosides (MOGLs) in order to test the versatility of APEX. For each dataset, we performed data pre-processing, generated mass difference networks, homologous series, and spectral similarity networks as described above. We would like to note that APEX is agnostic of the LC-MS/MS pre-processing software and only requires a feature table and related MS^2 spectra. Metabolites have been manually annotated by interpretation of fragmentation spectra and/or curated from the respective publications.

4.3.2 Manual annotation of GPNAEs, ascarosides and MOGLs

The first two datasets (*C. elegans* microbiota and MSV000087885) were screened for different GPNAE variants by exact mass matching in negative ionization mode (because of its characteristic fragmentation in negative mode), using an in-house MS^1 library containing GPNAEs with different acyl chain lengths. In order to confirm those GPNAE variants, the corresponding MS^2 spectra were inspected to contain several fragments: *m/z* 79.9668 (metaphosphoric acid, $[H_2PO_3]^-$), *m/z* 171.0064 (glycerol 3-phosphate, $[C_3H_8O_6P]^-$), *m/z* 152.9958 (glycerol 3-phosphate minus water, $[C_3H_6O_5P]^-$), and the neutral loss of 74.0367 ($-C_3H_6O_2$ resulting in diagnostic NAE-phosphate fragment) [319].
In total we manually annotated 19, and 10 features as GPNAEs in the *C. elegans* microbiota and MSV000087885 dataset, respectively. In the third dataset, ascarosides and MOGLs have been annotated. We annotated 9 and 16 candidates (ascarosides and MOGLs, respectively) based on exact mass matching, and MS^2 fragmentation spectra. Even more we used retention time (RT) matching, if available from the respective publication yielding high confidence annotations, which can be used as seeds [318].

4.3.3 Development of the APEX workflow

Each type of network, mass difference or spectral similarity, covers different aspects of metabolic transformations. Mass differences often can be spurious and spectral similarity can be used to establish a potential structural similarity. However, certain structural modifications might change the fragmentation in such a way that potential structural similarity can no longer be established. For example, comparing fragmentation of GPNAE 13:0 and GlycoGPNAE 13:0 (Figure 4.2), we see that those glucose variants exhibit changes in the fragmentation behavior. Even though there are several matching peaks (*m/z* 410.2313 $[C_{18}H_{37}NO_7P]^-$, 336.1940 $[C_{15}H_{31}NO_5P]^-$, 171.0064 $[C_3H_8O_6P]^-$, 152.9958 $[C_3H_6O_5P]^-$, and 79.9668 $[H_2PO_3]^-$), two of which match the neutral loss of hexose (i.e., 572.2835 - 162.0528 = 410.2313 and 333.0592 - 162.0528 = 171.0064), they vary greatly in their intensity. In addition, glucose variants show additional fragments corresponding to internal glucose fragments (*m/z* 101.0244, 119.0708, 89.0239, and 59.0133) [320]. The resulting (modified) GNPS cosine score [231, 260] comparing the spectra of GPNAE and GlycoGPNAE is equal to 0.69 (0.64 without considering the precursor *m/z*), which results in the generation of separate clusters in the spectral similarity network (with a threshold for the modified cosine > 0.8). Nevertheless, both features can be associated by a meaningful mass difference between the precursor *m/z* of 162.0528 Da corresponding to the addition of a hexose moiety.

To establish new connections between features not connected in spectral similarity network, we used the mass difference network. However, some connections in the mass difference network could

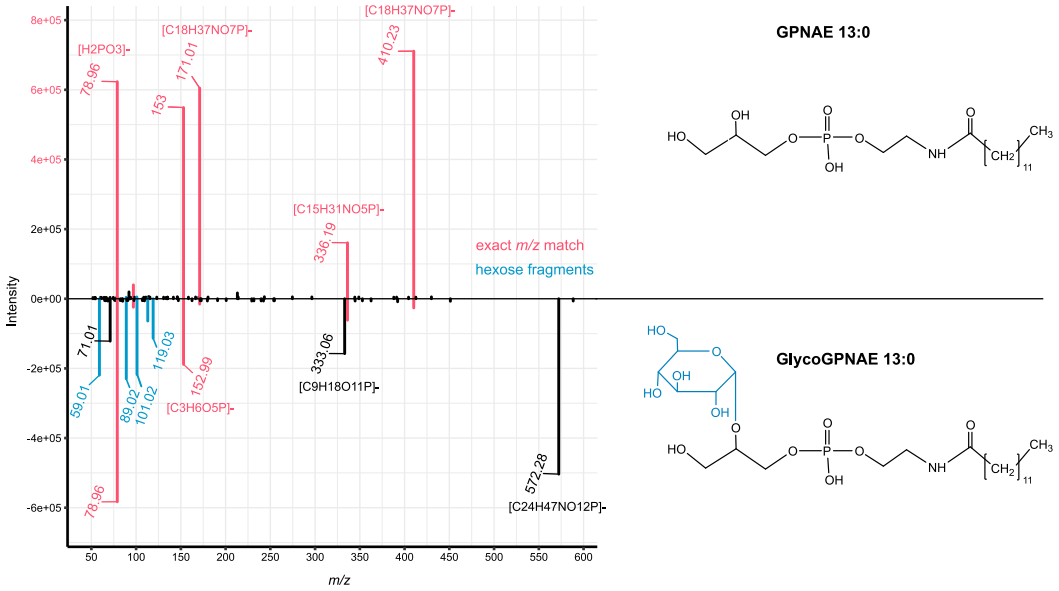


Figure 4.2 Mirror plot of GPNAE 13:0 (top spectrum) and GlycoGPNAE 13:0 (bottom spectrum) and molecular structure of (Glyco)GPNAE. Exact *m/z* matches are displayed in pink, internal glucose fragments of GlycoGPNAE 13:0 are displayed in blue.

lead to incorrect conclusions; for example, connections to isomeric compounds or matches to random features that have no biological meaning. In the case of LC-MS/MS, retention time can be used as an additional level of information.

Certain metabolite classes from homologous series, e.g., lipid-like molecules, show differences in acyl-chain length. For example, fatty acids form well-known homologous series, where each member of the series differs from the previous member by a repeating methylene (CH_2) unit. This can be used for metabolite identification since a distinct pattern in the chromatographic separation will be found for the homologous series. In the case of reversed-phase-based separation, an increase in chain length leads to an increased retention. By grouping features that belong to the same homologous series, we can increase the reliability of some mass difference annotations linked to lipid-like compounds. Hence, we used homologous series as additional information in our APEX workflow.

In APEX, we take advantage of the different topologies of the various experimental networks to propagate and hence predict accurate annotations of specific metabolite classes, such as GPNAE. Starting from a set of seed nodes, the APEX workflow iteratively annotates the first neighbors of each seed, varying the annotation based on the types of connections between the nodes.

If multiple annotations of the same node exist based on different seed nodes (i.e., manual annotations) and leading to the same predicted molecular formula, APEX prioritizes the one that considers the highest number of experimental network connections, and, if equal, the simplest one with the smallest mass difference is preferred. This approach helps to reduce ambiguity in the annotation process and ensures the selection of the most reliable annotation. However, if annotations of the same node differ in their predicted molecular formula, APEX keeps both annotations. Although the APEX workflow focuses mainly on the annotation of first neighbors, it can also annotate those second neighbors (i.e., the nodes that are at a distance equal to two from the seed nodes, where the distance is equivalent

to the minimum number of edges in the path between any pair of nodes) that are connected in all the experimental networks.

4.3.4 Application and validation of the APEX workflow to identify GPNAE

We first tested our developed APEX workflow to annotate GPNAE in our in-house dataset. In total, we manually annotated 19 GPNAE of different chain length and degree of saturation that are all part of the same cluster in the spectral similarity network (cluster 1 in Figure 4.3A). We used all 19 manually annotated nodes as Seed Node Set (SNS) for the APEX workflow. As a result, we were able to annotate most nodes from cluster 1, mainly using the spectral similarity network in combination with the mass difference network, as shown on Figure 4.3B. All annotations, manually and predicted by APEX, reveal that cluster 1 corresponds to (unmodified) GPNAE varying in their number of alkyl chains and saturation.
The APEX workflow was able to connect the GPNAEs (cluster 1) to the separated GlycoGPNAEs (cluster 2) using the mass difference network; with the mass difference of 162.0528 (blue edges), corresponding to a hexose moiety. Figure 4.3 exclusively visualizes the mass differences of hexose molecules, highlighting the interconnectedness facilitated by APEX between the clusters and enabling their annotation (Figure 4.3B).

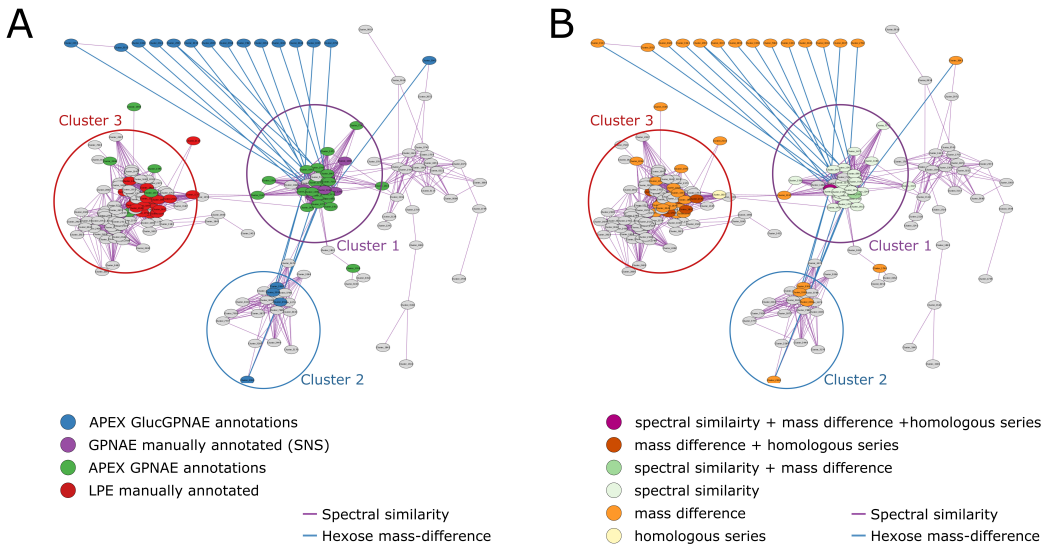


Figure 4.3 Spectral similarity network (threshold > 0.8) of dataset 1 and APEX results using 10 manual annotations as seed node set (SNS). Spectral similarity edges visualized in purple and hexose mass-difference (162.0528) in blue. There are different clusters formed in the network representing GPNAEs (cluster 1), GlycoGPNAEs (cluster 2), and Lysophosphatidylethanolamines (LPEs; cluster 3) **A** node coloring: purple: manual annotations/ SNS; blue: APEX GlycoGPNAE annotations; green: remaining APEX-based annotations; red: manually annotated LPEs **B** APEX annotation levels i.e. combination of edges; node coloring: purple:spectral similarity + mass difference + homologous series; red: mass difference + homologous series; green: spectral similarity + mass difference; light green: spectral similarity; orange: mass difference; yellow: homologous series

Even more, twelve GlycoGPNAE were separated from cluster 2 (in the merged network) and did not even appear in the spectral similarity network. This is because our in-house dataset showed low MS^2 coverage (30% in negative mode). As a result, most of the features are not present in the spectral

similarity network, but since mass difference networks rely on MS^1 data, those features can be also addressed and annotated using the APEX workflow.
Table 4.1 shows most of the APEX-based annotations of our in-house *C. elegans* microbiota dataset are based on mass differences (specifically, 142 consider mass difference edges between nodes that are not connected via spectral similarity). But as mentioned, mass differences are less reliable than spectral similarity because connections to random features without biological meaning might arise. Even more, mass difference networks do not distinguish different potentially present isomers with different retention times and as a result they are also connected. GPNAEs are isomeric to Lysophosphatidylethanolamines (LPEs). Therefore, using only mass differences led to erroneous annotation of 23 LPEs that were annotated as GPNAEs by the APEX workflow, as shown in Figure 4.3. But since they had associated fragmentation spectra which showed a different fragmentation, corresponding nodes exist in the spectral similarity network and they are disconnected, differentiations can be made. Furthermore, in most cases isomeric GPNAEs and LPEs could be baseline separated in the chromatographic dimension.
In order to further evaluate the APEX workflow, we reprocessed the publicly available dataset from Helf *et al.* [316], which also detected GPNAEs alongside their glycovariants, and similar to the clustering of our in-house *C. elegans* microbiota dataset, GPNAEs and GlycoGPNAEs were found in different, isolated clusters in the spectral similarity network. As we mentioned, this is due to the structural difference caused by the glucose moiety leading to different fragmentation (see above).
We manually annotated 10 GPNAEs (using MS^1 and MS^2 data) that were used as SNS to annotate their neighbors by APEX, which resulted in the annotation of 293 nodes (Table 4.1): The APEX workflow also annotated 26 glycoGPNAEs using connections in the mass difference network. Even more, 12 out of those 26 glycoGPNAE annotations were made without the availability of MS^2 spectra, which highlights the strength of our approach in annotating metabolites in cases where fragmentation data is not available.
The inclusion of homologous series to filter the mass difference network allows for potentially filtering out isomeric features with mismatching retention times (Table 4.1). Moreover, it adds additional confidence for features that have been only annotated on the MS^1 level. Additionally, for the current implementation of the APEX workflow, it is crucial that all species of a homologous series are present in the analysis, which is due to limitations of the *nontarget* R package used for generating the homologous series. This can be an issue for low abundant species that are not detected in MS analysis.
The ability to utilize multiple experimental networks is a key strength of the APEX workflow, which can potentially overcome the limitations of using a single network and increase the accuracy and confidence in metabolite annotation.

4.3.5 Comparison of molecular formula predictions: APEX vs. SIRIUS

To benchmark the proposed APEX workflow, we performed a comparison of the molecular formulas predicted by our approach with those obtained from SIRIUS, a widely used software tool for metabolite annotation in LC-MS/MS-based metabolomics that can be used to predict molecular formulas using isotopic patterns and fragmentation trees [234]. While SIRIUS primarily focuses on calculating the best fitting formulas, APEX goes beyond that by leveraging additional biochemical information to refine and enhance formulae propagation.
To ensure a comprehensive comparison, we considered all candidate molecular formulas provided by SIRIUS, i.e., those with multiple candidates ranked based on their similarity to the observed spectrum.

	Dataset 1 In-house	**Dataset 2 Massive MSV000087885**	**Dataset 3 Massive MSV000086293**	
class	GPNAE	GPNAE	ascr	MOGL
no of seed	19	10	9	16
total annotations	204	293	75	226
mass difference + spectral similarity + homol. series	14	11	0	0
mass difference + spectral similarity	10	11	10	23
mass difference + homol. series	23	46	1	0
spectral similarity + homol. series	2	0	0	0
spectral similarity	8	22	7	135
mass difference	142	192	57	68
homol. series	5	11	0	0
multiple annotations	114	241	32	155
multiple conflicting annotations	3	0	0	0

Table 4.1 Overview on number of APEX annotations through the different datasets and metabolite classes

However, because of potentially missing MS^2 data, not all features in the dataset have molecular formulas available in SIRIUS computations. It is important to note that the APEX workflow only provides molecular formulas for features that are connected in the mass difference network, propagating the formula difference. Despite these limitations, we observed matches at 90.9% (i.e., 76 out of 83) in all APEX-based annotations for features with available molecular formulas and SIRIUS formula results (SIRIUS results available at `https://github.com/michaelwitting/APEX`).

To further validate the results of the APEX workflow, we manually annotated 93 different compounds (beyond the 19 GPNAE used as SNS, i.e. organic acids, amino acids, fatty acids, nucleotides and glycerophospholipids; ids available at GitHub) in our in-house *C. elegans* microbiota dataset and compared the observed molecular features with those obtained using the APEX workflow. Remarkably, we found no mismatches between the manually annotated compounds and the APEX-based annotations, providing compelling evidence for the accuracy and reliability of the APEX workflow.

Among the APEX-based annotations from the publicly available dataset (MassIVE MSV000087885), 88.1% of the predicted formulas (i.e., 37 out of 42) matched those from SIRIUS. Notably, all (i.e., 3 out of 3) of the annotations based on the combination of spectral similarity, mass difference, and homologous series have the same molecular formula than those predicted by SIRIUS, which underscores their high reliability.

Additionally, we assessed the performance of the APEX workflow using leave-one-out cross-validation on our in-house *C. elegans* microbiota dataset. This involved utilizing all but one manual annotation as SNS and keeping the left out manual annotation as a Validation Set (VS). This process is repeated for each of the manual annotations, resulting in n number of evaluations, where n is the number of manual annotations. We also validated the APEX workflow using leave-two-out cross-validation. In both cases (leaving one or two seeds out at a time), all the VS were correctly annotated (i.e., matching molecular formulae).

To further evaluate the influence of the number of seed nodes on the APEX workflow results, we applied the leave-one-out cross-validation on three different approaches using all (i.e., 19), 50% (i.e., 10), and 25% (i.e., 5), of the manual GPNAE annotations as SNS, respectively. The VS for each approach was the set of left-out annotation. Remarkably, for all three approaches, each left-out VS was annotated

correctly. This suggests that the APEX workflow performs well and generates reliable annotations, even with small SNS.

4.3.6 Influence on seed nodes

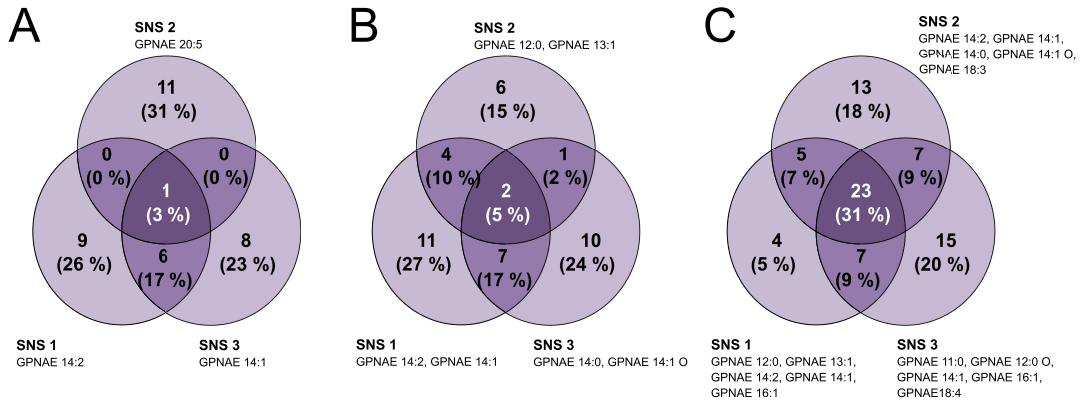


Figure 4.4 Consistency of annotations made by the APEX workflow using different (random) SNS of different sizes. **A**, overlap of annotations using three different (individual) nodes as SNS. **B**, overlap of annotations using three different pairs of nodes as SNS. **C**, overlap of annotations using three different SNS of five nodes each.

We evaluated the impact of the use of different Seed Node Sets (SNS) on the APEX workflow by randomly selecting three SNS of different sizes (1, 2, and 5 nodes) and assessing their influence on the number of observed APEX-based annotations. The SNS used determined the annotation process (see Table C.3, Table C.4, Table C.5). Our results showed that using different SNS resulted in relatively low overlap of the annotated features, especially when only one seed node was used (Figure 4.4A); whereas using more manual annotations as SNS (Figure 4.4C) increased the overlap of APEX-based annotations between different SNS. We recommend using at least five manual annotations as SNS to enhance annotation coverage of GPNAEs. The selection of seed nodes can significantly impact the annotation made by the APEX workflow, so researchers should carefully consider the number and combination of seed nodes used. To assess the overlap of true GPNAE annotations between different SNS, we performed a comparison by only retaining APEX-annotations that were annotated by the combination of spectral similarity, mass difference, and homologous series. We found that the overlap of these annotations increased with the size of the SNS (however, no trend was shown because the APEX-annotation strongly depends on the type of SNS). We compared the molecular formulas predicted by the APEX workflow with other manually annotated features of the dataset and determined that using a minimum of 5 seed nodes as input yielded 36 matching formulas, ensuring high-quality annotations. In general, the larger the SNS, the more APEX-based annotations there will be, but the selection of seed nodes impacts the results. The use of multiple seed nodes is recommended to increase the quality and certainty of the APEX-based annotations.

4.3.7 Validation on independent, publicly available dataset

Finally, we tested the APEX workflow on a different dataset focusing on two other compound classes in *C. elegans*: ascarosides and modular glucosides (MOGLs). Ascarosides are signaling compounds

involved in a wide range of biological processes, including development, reproduction, and behavior [80, 102, 321]. Similar to GPNAEs, no reference spectra exist in public MS^2 databases, and no reference standards are commercially available. MOGLs have been recently described in *C. elegans* and are constructed through various combinations of diverse metabolic building moieties [322], making them an ideal model for evaluating APEX.

We used the dataset from Le *et al.* (MSV000086293), in which we manually annotated 9 ascarosides (also reported in the respective publication, i.e. 8x ascr#, and 1x icas#), and 16 MOGLs (i.e., 4x tyglu#, 11x iglu#, and 2x angl#), that we used as SNS [318].

Starting with the ascarosides as seed nodes, the APEX workflow annotated 7 species (i.e., nodes) using only spectral similarity networks, 10 by a combination of spectral similarity and mass difference, and 57 based only on mass difference (Table 4.1). The annotations obtained solely from the mass difference analysis and those obtained through a combination of mass difference and spectral similarity analysis lead to the annotation of similar variants, i.e., CH_2, C_2H_4, HPO_3, H_2, C_2H_2 and C_9H_5NO variants. These annotations arise from differences in molecular formulas between various compounds or between an ascr# and its corresponding icas# variant related to an indole carboxylic acid residue. Therefore, connections in the mass difference network allow annotating similar variants, even if the corresponding nodes are not connected in the spectral similarity network. This is different from GPNAEs, where we noticed that Hexose mass differences only occurred when there were no spectral similarity edges present. By using APEX, we were able to annotate additional ascarosides, such as bhas#9 by the mass difference 44.0262 (C_2H_4O) to the seed node ascr#5 that resulted in the chemical formula C11H20O7, or phascr#71 by connecting ascr#7 with the mass difference of 79.9663 (corresponding to HPO_3).

The same dataset also included MOGLs, that we aimed to annotate by APEX using a different set of seed nodes and observed 226 annotations (Table 4.1; 135 via spectral similarity, 68 via mass difference, and 23 via combination of both). Interestingly, the main mass differences were C_5H_6O, $C_6H_{10}O_5$ (Hexose), $C_{13}H_{22}O_5$ (corresponding to an ascr#1 block), $CH_2$and HPO_3. Additionally, it is worth noting that the proportion of matching molecular formulas, when compared to the formulas predicted by SIRIUS, is slightly higher for the combination of mass difference and spectral similarity (82.6%, i.e., 19 out of 23) then for the mass difference annotations alone (71.7%, i.e., 30 out of 42).

Using APEX, we found 10 species connected to angl#4 (7 through spectral similarity, 2 through mass difference and 1 through spectral similarity combined with mass difference). A particular example was the annotation of angl#4+C_5H_6O with the molecular formula $C_{25}H_{29}N_2O_{12}P$ which potentially corresponds to angl#26 and was not reported in previous publication. Even more, we annotated one feature with the potential chemical formula $C_{22}H_{30}NO_{10}P$ by the mass difference of C_3H_6O to iglu#8 which potentially corresponds to phicas#11. Another feature was potentially annotated as anglas#2 by APEX and the mass difference of $C_{13}H_{22}O_5$ (which corresponds to an ascr#1 unit) to angl#2.

In conclusion, the APEX workflow was successfully applied to annotate additional species of ascarosides and MOGLs in *C. elegans*. By using spectral similarity and incorporating mass difference, a total of 75, and 226 species (ascarosides and MOGLs, respectively) were annotated. However, since both ascarosides and MOGLs do only rarely form homologous series, the filtering step previously applied to GPNAEs which can be categorized to lipid and lipid-like compounds, could not be used here. Therefore, while the APEX workflow is effective for identifying ascarosides and MOGLs, it has some limitations that must be considered when annotating these compounds.

4.4 Conclusion

Here we introduced an **A**nnotation **P**ropagation through multiple **EX**perimental networks (APEX) workflow and used it for the annotation of glycerophospho *N*-acyl ethanolamides (GPNAEs), a compound class in *C. elegans*. The combination of spectral similarity, mass-difference, and homologous series allowed for accurate and comprehensive annotation of GPNAEs, including automated annotation of Glyco variants that are not connected in the spectral similarity network. The incorporation of different network tools improved the accuracy and comprehensiveness of the annotation process, while the quality of annotations was underscored by their high matching rate with SIRIUS results. Additionally, the homologous series was introduced to filter out non-biological features and improve the identification of compounds with more biological significance. However, it is still necessary to use spectral similarity as a more reliable network since mass-differences tend to be noisier and GPNAEs are isomeric to Lysophosphatidylethanolamines, leading to incorrect annotations using the APEX workflow.
Moreover, the selection of the manual annotations used as Seed Nodes Set (SNS) significantly influences the resulting APEX annotations, and a higher number of seed nodes enhances the annotation process. These results demonstrate the usefulness of the APEX workflow in identifying and characterizing compounds in complex datasets, particularly for glycolipid-related compounds.
In future, different possibilities for the extension of APEX exist. GSMNs capture knowledge on known metabolic pathways and transformations and potentially allows to bridge individual features or cluster using biochemical reactions [239, 243]. Another possibility are correlation networks. Especially, gaussian graph models have been shown to be able to reconstruct biochemical valid links from metabolomics data [165]. Furthermore, it helped to in identifying novel metabolites [323]. By optimizing the selection of SNS and incorporating different complementary networks such as GSMNs, correlation networks, etc., the APEX workflow may provide even more comprehensive annotation results and become an invaluable resource for researchers seeking to decipher the complexities of the metabolome.

Data availability
The APEX workflow, networks and ids are available on GitHub (`https://github.com/michaelwitting/APEX`).

5 Discussion and Outlook

In this thesis, we[1] applied novel analytical and bioinformatic tools in *C. elegans* metabolomics in order to aid the identification of known and unknown metabolites. We demonstrated the need for further research in this area in Section 1.2, where we found a disparity between experimentally detected metabolites and those predicted based on genome knowledge. Even more, unknown metabolites, which have not yet been identified or characterized, present a significant challenge in metabolomics, as they may not be detectable using standard experimental techniques or may not match any known metabolites in databases.

To identify both, unknown and known metabolites, we employed two innovative approaches in *C. elegans* metabolomics. In Chapter 2 we developed *MobilityTransformR* in order to perform effective mobility transformation in the R environment, leading to enhanced reproducibility of CE-MS data, which was subsequently applied in Chapter 3. This Chapter focused on the application of capillary electrophoresis coupled to mass spectrometry (CE-MS) to analyze polar metabolites in *C. elegans*. We investigated differences and strengths compared to hydrophilic interaction liquid chromatography-mass spectrometry (HILIC-MS), which is usually used for the analysis of polar metabolites. The second focus of this thesis was the development of the **A**nnotation **P**ropagation through multiple **EX**perimental networks (APEX) workflow in Chapter 4 to improve and automate annotation of specific metabolite classes in *C. elegans*, by enhancing spectral similarity networks with mass difference networks and homologous series. Specifically, we focused on the annotation of glycerophospho *N*-acyl ethanolamides (GPNAEs), a new compound class in *C. elegans*.

CE-MS has been extensively employed in metabolomics, making it a suitable and powerful tool for analyzing polar and charged compounds [119, 120], which was also confirmed in our study of *C. elegans* metabolites. Our findings, as presented in Chapter 3 and Chapter 2, especially highlighted the importance of transforming migration time into effective mobility of CE-MS data because it significantly enhanced reproducibility. To facilitate this process, we developed *MobilityTransformR*, an R package that leverages EOF markers added to the samples at CE-MS analysis to perform effective mobility scale transformation. We highlighted its usefulness on the effective mobility scale transformation of a leucine reference standard where the relative standard deviation of its migration time was dramatically reduced from 5.18 % to 0.65 %.

Compared to other mobility transformation software tools like ROMANCE [324–326], *MobilityTransformR* does not include quantitative conversions by correcting the peak areas, whereas other conversion tool such as ROMANCE do [274]. Hence no quantitative comparisons between metabolites with varying effective mobility can be performed. Therefore, this package needs further development in order to compete with other existing tools.

However, the main benefit of *MobilityTransformR* compared to other effective mobility transformation tools is the seamless integration with existing metabolomics pipelines in the R environment from

[1] In the context of this thesis, the pronoun "we" is employed to refer to the author of the work. It is a convention in scientific writing to use "we" to convey authorship, even when an individual is the sole contributor.

packages, such as *MetaboCoreUtils*, *xcms*, *MSnbase* and *Spectra* [272, 273]. With our proposed solution, researchers can confidently incorporate CE-MS into their metabolomics workflows, streamlining the analysis process.

Migration time (MT) shifts could be one possible reason for limited use of CE-MS compared to HILIC-MS. However, we have demonstrated that this problem can be avoided by using migration markers and effective mobility transformation. By addressing this issue, CE-MS can become a valuable tool for *C. elegans* metabolomics, offering several advantages over HILIC-MS, such as higher separation efficiency, better peak-intensity linearity, lower saturation of MS signals, and a distinct selectivity for separated compounds, as demonstrated in Chapter 3. Hence, we directly compared the effectiveness of CE-MS and HILIC-MS for the analysis of *C. elegans daf-2* mutants, which are known to have a prolonged lifespan compared to the wild type [96, 133, 168]. The required sample volume was significantly lower in CE-MS (32 nL vs 3 µL, corresponding to approximately 0.6 and 60 worms per injection for CE-MS and HILIC-MS, respectively). Furthermore, our study revealed that CE-MS and HILIC-MS target distinct portions of the *C. elegans* metabolome, as evidenced by the unique *m/z* distribution of each method, with CE-MS exhibiting slightly more features with lower mass than HILIC-MS. Also, CE-MS analysis revealed 41 metabolites that were not detected in HILIC-MS, indicating that it can provide complementary information about the nematode's metabolism. Notably, we observed significant differences in the up- or downregulation of certain metabolites between the two methods, which we attributed to signal saturation in HILIC-MS. Furthermore, CE-MS analysis confirmed previous knowledge on the long-lived mutant, which is associated with the reorganization of amino acid metabolism [96, 99, 100, 133, 168]. We also discovered novel metabolite features associated with *daf-2* using CE-MS, such as an upregulation of uridine and a downregulation of purine degradation intermediates and hypoxanthine. These findings provide additional insights into the metabolic pathways associated with longevity in *C. elegans*. Another major finding of our study was the identification of changes in the tryptophan-nicotinamide pathway in *daf-2* samples using CE-MS, which has not been reported in *daf-2* but in other long-lived mutants in *C. elegans* such as *sir-2.1* [307]. This highlights the importance of using different analytical platforms, such as CE-MS, to explore various components of the complex metabolome and discover novel metabolites and pathways.

Although the study presents compelling evidence regarding the utility of CE-MS for metabolite analysis in *C. elegans*, it is important to recognize its limitations. For example, like all analytical methods in non-targeted metabolomics, CE-MS is limited in analyte coverage, which impact the comprehensiveness of the results. Additionally, we observed limitations in sensitivity due to the applied sheath-liquid based ESI-MS hyphenation. Notably, other hyphenation techniques exists, showing increased sensitivity compared to the standard triple-tube sprayer, such as nanoflow sheath liquid or sheath-less approaches [217, 219, 280]. Using interfaces with increased sensitivity might provide further information on *C. elegans*' metabolism that is worth for further investigation.

Furthermore, in our study we focused on utilizing acetic acid as background electrolyte (BGE) for CE separations, which has shown to be particularly advantageous for the separation of cations [327]. However, it is important to consider alternative approaches to address the limitations associated with metabolite coverage and sensitivity in CE negative ionization mode. For example, Kok *et al.* [328] demonstrated that incorporating triethylamine into the BGE and SL can significantly enhance sensitivity in negative ionization mode. Their findings suggest that this modification can be effective in improving the analysis of metabolites. Moreover, Drouin *et al.* observed increased sensitivity by employing ESI positive ionization mode with reversed CE polarity for the analysis of anions [215]. This

approach provides an alternative strategy to enhance the detection and characterization of anionic metabolites.
Additionally, nonaqueous CE systems, incorporating organic solvents as BGEs along with a small portion of water and volatile salts, have shown promising results in ESI negative mode for the analysis of anions, such as fatty acids [329]. This highlights the benefits of exploring different BGE compositions to improve the identification and quantification of anionic metabolites.
It is worth noting that the choice of sprayer needles used in the electrospray ionization (ESI) process also impacts sensitivity. Soga *et al.* highlighted the importance of using platinum ESI spray needles instead of stainless steel needles. Stainless steel needles can generate metal ions during the ESI process, which may form complexes with analytes and lead to decreased detection sensitivity [330].
Another direction of method development is the application of capillary coatings that influence the strength and direction of the EOF, hence directly affecting the separation. For instance, cationic coatings, which create a reversed EOF, proved beneficial for the analysis of anionic metabolites [331].
Considering these insights, it becomes evident that adjusting analytical setups, such as exploring different BGE, capillary coatings, and optimizing MS coupling, can contribute to a more comprehensive analysis of both, anionic and cationic metabolites in *C. elegans* using CE-MS. These modifications hold the potential to enhance the annotation of novel metabolites and expand our understanding of metabolic signatures in *C. elegans*.
By utilizing CE-MS, we have gained valuable insights into the metabolic profiles of *C. elegans daf-2* mutant, shedding light on the underlying mechanisms of longevity, which has implications for medicine. Understanding the metabolic pathways and processes involved in the longevity of model organisms like *C. elegans* can potentially contribute to the identification of novel therapeutic targets or pathways associated with aging-related diseases, leading to the development of interventions or treatments for age-related conditions.
Practically, our research opens up opportunities for the development of improved analytical techniques and instrumentation in the field of metabolomics. The identification of differences in metabolite profiles between CE-MS and HILIC-MS suggests that combining these methods can provide a more comprehensive understanding of the metabolome. This insight could lead to the development of integrated analytical platforms or workflows that combine the strengths of multiple techniques, enabling more accurate and detailed metabolomic analyses. Additionally, optimizing and fine-tuning CE-MS methodologies could have commercial applications in metabolite profiling, drug discovery, environmental monitoring, and personalized healthcare diagnostics. By building upon our findings, future researchers can advance the field and deepen our understanding of metabolism in *C. elegans* and other organisms.

In addition to the analysis of metabolites using appropriate analytical techniques, metabolite identification remains a major bottleneck in non-targeted metabolomics. The unavailability of reference standards and spectral libraries presents a considerable challenge, particularly when it comes to annotating new metabolites. Therefore, the role of bioinformatics and network analysis in metabolomics research is becoming increasingly crucial in assisting with metabolite identification.
In Chapter 4 we introduced the APEX workflow as a powerful tool to address this challenge, by combining spectral similarity networks with mass difference networks and homologous series. Our proposed annotation workflow holds significant importance in the field of metabolomics, particularly in the automatic annotation of new metabolite classes using network propagation. Specifically, we focused on glycerophospho *N*-acyl ethanolamides (GPNAEs) in *C. elegans* as a first use-case to demon-

strate the benefits of the APEX workflow. One of the key advantages of using APEX was its ability to establish connections between GPNAEs and GlycoGPNAEs, which are glycated variants of GPNAEs. These variants are often separated in the spectral similarity network due to slight differences in fragmentation. However, leveraging the mass difference network, specifically the mass difference 162.0528 (corresponding to hexose), APEX facilitated the automated annotation of these variants.
Furthermore, the mass difference network offers an additional advantage by allowing putative annotation of MS features in instances where MS^2 spectra are unavailable. This advantage becomes particularly prominent when compared to spectral similarity networks, as the latter do not include features lacking MS^2, thereby failing to provide any information about potentially biologically relevant features.
APEX enhanced the accuracy and comprehensiveness of GPNAE annotations, which was confirmed by the high matching rate with the molecular formulas predicted by SIRIUS. Additionally, the APEX workflow demonstrated its robustness and applicability to different datasets acquired with different mass spectrometers, highlighting its versatility and potential for broader use in metabolomics research.
APEX was validated on an independent dataset, successfully annotating ascarosides and modular glycosides (MOGLs) in *C. elegans*. The integration of spectral similarity and mass difference networks showcased APEX's versatility across different compound classes. Accurate annotation was facilitated by leveraging structural similarities and unique mass differences, providing valuable insights into chemical diversity of ascarosides and MOGLs. This reinforces APEX's reliability and broad applicability in metabolomics research, enabling efficient exploration and annotation of diverse metabolite classes.
While our research findings on the APEX workflow provide valuable insights into metabolomics, it is important to acknowledge certain limitations and constraints that may have influenced the validity and generalizability of the results. First, we implemented the use of homologous series in APEX to enhance the identification of "real" metabolites by filtering out non-biological features. However, we observed that relying solely on mass differences and homologous series without considering spectral similarity could lead to false positive annotations, specifically in the case of Lysophosphatidylethanolamines (LPEs) which are isomeric to GPNAEs. Continued advancements in chromatographic separation hold the potential to achieve the separation of homologous series for both LPEs and GPNAEs. By having separate homologous series, we could then filter mass difference connections to LPEs, thereby enhancing the reliability of GPNAE annotations.
Incorporating advanced spectral matching algorithms represents a promising direction for further development of the APEX workflow. Improved scoring algorithms, such as Spec2Vec [262] or the hypothetical neutral loss spectra algorithm [263], have the potential to enhance both, the accuracy and coverage of metabolite identification. Integrating an improved MS^2 scoring algorithm could therefore enable the connection between GPNAEs and glycoGPNAEs in the spectral similarity network, resulting in a clearer distinction from LPEs.
To further enhance the effectiveness of spectral similarity networks, it is crucial to improve the quality of MS^2 data used to construct these networks. Factors such as resolution, sensitivity, and dynamic range play a significant role to enhance metabolite identification, improve spectral matching, and increase metabolite coverage. Therefore, it is essential to focus on improving these aspects during the acquisition process.
Since our research primarily focused on lipid-related compounds, APEX could be easily applied to

other lipid species such as LPEs, Phosphatidylcholines (PCs), Phosphatidylethanolamines (PEs), or Phosphatidylinositols (PIs). The paves the way for expanding into new areas of exploration in *C. elegans* and further facilitating the annotation of lipid species. However, while lipidomics plays a crucial role in metabolomics, it is equally important to explore the application of the APEX workflow to other metabolite classes, including amino acids, carbohydrates, and different secondary metabolites.

In terms of the workflow itself, there are areas that could be further optimized and refined. For example, the selection of the Seed Node Set (SNS) greatly influenced the APEX workflow's performance, with a higher number of seed nodes improving the annotation process. Future studies could investigate alternative strategies for seed node selection and explore the impact of different SNS compositions on the annotation results. Another interesting direction could be investigating the effect of using different compound classes as seed nodes on the annotation success, for example both GPNAEs and LPEs simultaneously.

Regarding the annotation of ascarosides and MOGLs in *C. elegans*, we found that homologous series had limited benefits since they rarely formed such series. Instead, spectral similarity and mass difference networks played a crucial role in accurately annotating both, ascarosides and MOGLs. Incorporating additional experimental networks, such as correlation networks, holds the potential to further enhance the quality of APEX annotations (for all, ascarosides, MOGLs and GPNAEs). Correlation networks address different aspects of metabolism compared to spectral similarity and mass difference networks. They can provide valuable insights into the interdependencies of metabolites within biochemical networks. Especially, Gaussian graphical models have demonstrated the ability to reconstruct biologically valid links from metabolomics data, also aiding in the identification of novel metabolites [165, 323].

However, it is worth noting that correlation networks may not be suitable for datasets with low sample sizes, such as our datasets, as it can result in an ill-conditioned sample covariance matrix and overestimated correlation values. Correlation networks are more applicable to larger sample sizes, such as cohort studies [165]. By incorporating correlation networks in appropriate scenarios, APEX annotations could be further boosted, enabling a comprehensive understanding of metabolite dependencies and their biological significance.

Looking towards the future, APEX holds immense potential for expansion and enhancement. One exciting avenue would be the integration of Genome-Scale Metabolic Networks (GSMNs), which capture extensive knowledge regarding established metabolic pathways and transformations. The integration of GSMNs within APEX could offer two significant benefits. Firstly, it could enable the annotation of metabolites in experimental data by leveraging the existing knowledge within the GSMNs to assign identities to detected features. This integration would enhance the accuracy and completeness of metabolite annotation. Secondly, it could facilitate the identification and filling of gaps in the GSMNs [161, 282] by detecting novel metabolites from the experimental data. This process would contribute to improving the overall quality and comprehensiveness of the GSMN.

Moreover, the creation of a comprehensive metabolome database for *C. elegans* curated from reviewed literature, such as mentioned in Section 1.2.1, and coupled with advanced computational tools could serve as a valuable resource for the research community. Such a database would facilitate the discovery of new biomarkers and metabolic pathways in *C. elegans* and provide a foundation for comparative metabolomics studies across different organisms.

Lastly, future research should focus on integrating metabolomics data with other -omics areas, such as genomics, proteomics, and transcriptomics. It has already been demonstrated that combination of

metabolomics with other -omics technologies boost metabolite annotations and provide a comprehensive understanding of biological systems [332, 333]. By combining APEX-based metabolite annotation with multi-omics integration, researchers can unravel complex molecular interactions and pathways underlying biological processes.
The potential for translating this research into practical applications or commercialization is promising. The APEX workflow, with its ability to enhance metabolite annotation, can be applied to other organisms beyond *C. elegans*, including humans. This expands the scope of its potential applications, such as biomarker discovery, drug development, and precision medicine. In summary, APEX has the potential to make significant contributions to the fields of metabolomics and lipidomics, which are crucial for advancing our understanding of biological systems and improving human health.

In conclusion, this thesis has successfully addressed various challenges associated with metabolite identification in *C. elegans* through the application of novel analytical and bioinformatic tools. By demonstrating the use of CE-MS for the analysis of polar metabolites in Chapter 3 and developing the *MobilityTransformR* tool for reproducible data analysis in Chapter 2, we have contributed to the advancement of *C. elegans* metabolome analyses.
Furthermore, the development of the APEX workflow in Chapter 4 has improved and automated the annotation of specific metabolite classes, focusing on GPNAEs. By integrating spectral similarity networks with mass difference networks and homologous series, we have enhanced the annotation process and expanded our understanding of *C. elegans* metabolism.
Overall, this work has provided valuable insights into the identification of known and unknown metabolites in *C. elegans*, contributing to the broader research goals of the field. The innovative approaches and tools developed in this thesis hold promise for further advancements in metabolomics research and its applications in understanding the intricate biological processes of *C. elegans* and beyond.

A Appendix Chapter 2

A.1 Methods

Metabolite Standard Leucine was dissolved in 25 ppm, 16 ppm, 10 ppm, 5 ppm and 1 ppm in milliQ water containing 10 ppm Procaine, and 50 ppm Paracetamol as EOF markers. Samples were analyzed on a 7100 capillary electrophoresis (CE) system from Agilent Technologies (Waldbronn, Germany) using a 80 cm fused silica capillary (Polymicro Technologies, Phoenix, AZ, U.S.A.) with an internal diameter of 50 µm and external diameter of 365 µM. The capillary was initially preconditioned by flushing 5 min MeOH, 5 min 1 M NaOH, 5 min milliQ, 20 min 1 M HCl, 5 min 0.1 M HCl, 5 min milliQ, and 5 min background electrolyte (BGE), which was 10 % acetic acid (v/v). Samples were injected hydrodynamically at 50 mbar for 12 s (12.9 nL, 0.8 % of total capillary volume) and Separation was performed at 25 °C and +30 kV including a constant pressure of 50 mbar. The CE system was coupled to an Agilent 6560 IM-QToF-MS (Agilent Technologies, Waldbronn, Germany) with a Dual Agilent Jet Stream ESI source via a coaxial sheath flow ESI interface using a commercial triple tube sprayer from Agilent Technologies (p/n G1607). The sheath liquid was delivered using an Agilent 1260 Infinity Isocratic Pump equipped with degasser and 1:100 splitter and composed of 2-propanol/ milliQ/ formic acid 50:50:0.5 (v/v/v) and 10 µM purine and 2 µM HP-0921 as lock masses. The sheath liquid was delivered at 3 µL/min. Source parameters were 10 psi Nebulizer gas, gas flow at 8 L/min at 300°C and 3.5 L/min Sheath Gas at 195 °C. Fragmentor voltage was 400 V, Skimmer 65 V, Octopole 800 V, capillary 2000 V and Nozzle 2000 V. Agilent .d data were transformed to the open-source .mzML format and centroided using MSConvert (3.0.20342) from ProteoWizard (http://proteowizard.sourceforge.net) (Chambers, et al., 2012).

A.2 Mobility Transformation

Here we provide the R script that was used to transform the acquired CE-MS data. For a detailed description of how to use MobilityTransfomR see below

```
library(xcms)
library(tidyverse)
library(MobilityTransformR)


wd_rawdata <- "G:/2021_Agilent_IMS/20210305_CEMS_pos/mzML_centroided"
fl <- list.files(wd_rawdata, pattern = ".mzML")

setwd(wd_rawdata)
raw_data <- readMSData(files = fl, mode = "onDisk")
```

```
# mz tolerance depends on MS mass accuracy
tolerance <- 0.005

# [M+H]+ of paracetamol: mz = 152.071154
mz_paracetamol <- c(152.071154 - tolerance, 152.071154 + tolerance)
mt_paracetamol <- c(650, 1500)

paracetamol <- getMtime(raw_data,
              mz = mz_paracetamol,
              mt = mt_paracetamol
)

mz_procaine <- c(237.160303 - tolerance, 237.160303 + tolerance)
mt_procaine <- c(300, 700)

procaine <- c()


procaine <- getMtime(raw_data,
              mz = mz_procaine,
              mt = mt_procaine)

marker <- paracetamol
marker$markerID = "Paracetamol"
marker$mobility = 0

procaineMobility <- mobilityTransform(x = procaine[i,1],
                                      marker = marker[i,],
                                      tR = 3/60,
                                      U = +30,
                                      L = 800)


procaineMobility <- c()
for (i in 1:nrow(procaine)) {
  m <- mobilityTransform(x = procaine[i,1], marker = marker[i,],
                         tR = 3/60, U = +30, L = 800)
   procaineMobility <- c(procaineMobility, m)

}

procaineMobility_median <- median(procaineMobility)

procaine$markerID = "Procaine"
procaine$mobility = procaineMobility_median
marker <- rbind(marker, procaine)
```

```
mobility_data <- mobilityTransform(x = raw_data, marker = marker)

setwd("G:/2021_Agilent_IMS/20210305_CEMS_pos/")
dir.create("20211210_mzML_centroided_mobility/")

fl_mobility <- "G:/2021_Agilent_IMS/20210305_CEMS_pos/20211210_mzML_
    centroided_mobility/"
fl_mobility_data <- paste0(fl_mobility,fl)

writeMSData(mobility_data, file = fl_mobility_data, copy = FALSE)
```

A.3 References

Chambers, M.C., et al. A cross-platform toolkit for mass spectrometry and proteomics. Nature Biotechnology 2012;30(10):918-920.

A.4 Package Vignette

Description and usage of MobilityTransformR

true

30 März 2022

Introduction

Capillary elecrophoresis coupled to mass spectrometry (CE-MS) in fields as proteomics or metabolomics is less common than for example liquid chromatography-mass spectrometry (LC-MS). A reason might be less reproducible migration times (MT) compared to retention times (RT) because of fluctuations in the Electroosmotic flow (EOF).

However, the effective mobility t_{eff} of a compound remains stable in the same electrophoretic system. The use of an effective mobility scale instead of a migration time scale circumvents the drawback of MT shifts and will result in highly reproducible peaks, which has already been shown in 2001 [1].

Effective mobility transformation for CE-MS data is not as straightforward as in CE-UV and until now and to our knowledge there is no implementation in R that performs effective mobility transformation of CE-MS(/MS) data.

As a model for the `MobilityTransformR` package, we have taken the recently developed open-source ROMANCE software [2], which transforms the MT scale of CE-MS data into an t_{eff} scale. However, the outputs are two separate files, each for positive and negative mobilities. We developed the `MobilityTransformR` package to perform mobility transformation of CE-MS/MS data in the R environment. Also, different to ROMANCE, the output will be a single file containing both, the positive and negative effective mobilities.

The transformation is performed using functionality from the packages *MetaboCoreUtils*, *xcms*, *MSnbase*, and *Spectra*. The transformed data can be exported as `.mzML` file and can be further analyzed in R or other software.

The CE-MS test data are from the *msdata* package.

Setup

Installation

The `MobilityTransformR` package can be installed as follows:

```
if (!requireNamespace("BiocManager", quietly = TRUE)) {
    install.packages("BiocManager")
}

BiocManager::install("MobilityTransformR")
```

To show the different functionality of `MobilityTransformR`, we need also to load the packages `xcms` and `Spectra`. If you have not installed them yet, type

```
if (!requireNamespace("BiocManager", quietly = TRUE)) {
    install.packages("BiocManager")
}

BiocManager::install("xcms")
BiocManager::install("Spectra")
```

into the Console.

Load required libries

`MobilityTransformR` integrates functions from different libraries.

```
# load required libraries
library(MobilityTransformR)
library(xcms)
library(Spectra)
```

Load test data

To showcase the functionality of the `MobilityTransformR` package two CE-MS test sets containing a mixtures of different metabolites at a concentration of 10 ppm and 25 ppm precisely, acquired at positive CE polarity and positive ionization mode is used and imported from the `msdata`package. Paracetamol was added to the analysis as EOF marker in a final concentration of 50 ppm to the sample. Procaine was added as positively charged, secondary marker at a final concentration of 10 ppm. The markers are required for the later effective mobility transformation process. The test data is loaded using the `readMSData()` function from `MSnbase`.

```
fl <- dir(system.file("CE-MS", package = "msdata"),
          full.names = TRUE)

# Load mzXML data with MSnBase
raw_data <- readMSData(files = fl, mode = "onDisk")
```

Get Marker Migration Times

In order to perform the effective mobility transformation, first the migration times (MTs) of the markers needs to be determined. If multiple files are used, the MT of the markers in each file must be determined since they will vary based on EOF fluctuations.

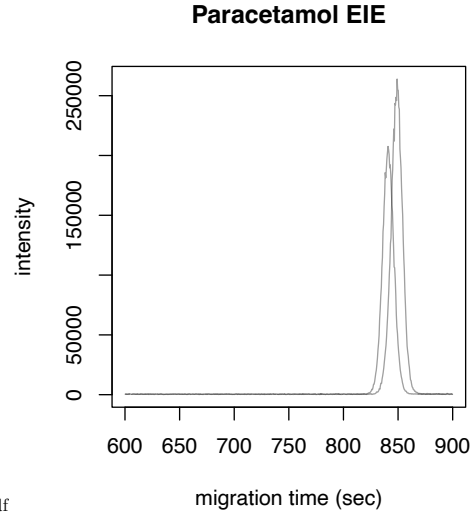
The `getMtime()` function requires an `OnDiskMSnExp` object,`raw_data`, as input and uses an mz-range `mz` and MT-range `mt` to generate an Extracted Ion Electropherogram (EIE). The MT of the peak will be determined by `findChromPeaks` from `xcms`. It is very important to define the `mz`-range and `mt`-range as narrow as possible to ensure that the right peak will be picked. The mz-tolerance defining the `mz`-range depends on the mass accuracy of the mass spectrometer that was used to acquire the data.

Get MT of EOF Marker Paracetamol

In order to check the right `mz`-, and `mt`-window for the exact MT determination, we use the`plot(chromatogram())` function to plot the respective EIE.

Note that the plot functions are adapted from LC-MS package `MSnbase`, which has "retention time" as default x-axis label.

```
# mz tolerance depends on MS mass accuracy
tolerance <- 0.005

# [M+H]+ of paracetamol: mz = 152.071154
mz_paracetamol <- c(152.071154 - tolerance, 152.071154 + tolerance)
mt_paracetamol <- c(600, 900)

marker_EIE <- raw_data |>
    filterMz(mz = mz_paracetamol) |>
    filterRt(rt = mt_paracetamol)

plot(chromatogram(marker_EIE),
    main = "Paracetamol EIE",
    xlab = "migration time (sec)"
)
```



marker-1.pdf

```
# adjust mz and MT windows if necessary
```

Use the adapted values to get the exact MT of Paracetamol. If several files will be used it is important to select a `mt`-range wide enough to get the right Paracetamol-peaks in all files, because single migration times might change significantly due to EOF-fluctuations between measurements.

```
# get the MT of paracetamol
paracetamol <- getMtime(raw_data,
    mz = mz_paracetamol,
    mt = mt_paracetamol
)
paracetamol
```

```
##     rtime fileIdx
## 1 840.796       1
## 2 848.640       2
```

MT of Secondary Marker Procaine

Effective mobilities can be already calculated using the EOF marker Paracetamol. However, it is also possible to calculate the mobility using a secondary marker. This might be of interest for example when the peak shape of the main marker is not sufficient or ambiguity detection.

In order to show how we can use a secondary marker to convert the data, we create in the next step a data.frame containing the MT and file information from another, secondary marker called Procaine.

```
procaine <- data.frame(
    "rtime" = c(450.239, 465.711),
    "fileIdx" = c(1, 2)
)
```

Effective mobility scale transformation

`mobilityTransform` uses different functions to perform the effective mobility transformation depending on the input type. It is possible to convert `numeric`, `Spectra`, and `OnDiskMSnExp`. Here, we show how each input-class can be transformed.

Effective mobility transformation of single migration times

Effective mobility scale transformation is performed either using a single or two mobility markers. Functions were adapted from González-Ruiz et al. [2] . The standard equation to calculate effective mobility μ_{eff} is

$$\mu_{eff} = \frac{l}{E}(\frac{1}{t_M - (t_R/2)} - \frac{1}{t_{EOF} - (t_R/2)})(\#eq : mobility1) \tag{1}$$

with t_M the migration time that is transformed, t_{EOF} the MT of the EOF marker and optional t_R which is the time of the electrical field ramp.

If the peak shape of the EOF marker is bad or it cannot be detected for other reasons, the effective mobility might also be calculated using a secondary mobility marker with known mobility:

$$\mu_{eff} = \mu_A + \frac{l}{E}(\frac{1}{t_M - (t_R/2)} - \frac{1}{t_A - (t_R/2)})(\#eq : mobility2) \tag{2}$$

with t_A, the MT of the secondary marker and its corresponding mobility μ_A.

Last, the mobility might be calculated using both markers. Therefore, there is no need to know the applied electrical field nor the exact capillary length.

$$\mu_{eff} = \mu_A \frac{t_M - t_{EOF}}{t_A - t_{EOF}} * \frac{t_A - (t_R/2)}{t_M - (t_R/2)}(\#eq : mobility3) \tag{3}$$

Calculation of the Secondary Marker Mobility $\mathfrak{t}_A$

First, a `data.frame` is created that stores the marker information, that will be used to transform the data. Either one or two markers can be used for the transformation. The `data.frame` needs at least the columns `rtime` and `mobility`. Additional columns to store marker information might be added as well. Here, we start with the neutral EOF marker Paracetamol, having the effective mobility of 0 $\frac{mm^2}{kV*min}$.

```
# Create a data.frame that stores marker information
marker <- paracetamol
marker$markerID <- "Paracetamol"
marker$mobility <- 0
```

Then the effective mobility of the secondary mobility marker Procaine, as single migration time, is determined according to eq. @ref(eq:mobility1). The transformation is performed using the marker information from Paracetamol. Moreover, if only a single marker is provided for the transformation, also the applied voltage `U` in kV, and the total and effective capillary length `L` in mm is needed. Additionally the field ramping time `tR` (in min) can be included for corrections.

```
procaineMobility <- mobilityTransform(
    x = procaine[1, 1], marker = marker[1, ],
    tR = 3 / 60, U = +30, L = 800
)
procaineMobility
```

```
## [1] 1327.35
```

Note: The unit of the mobility is $\frac{mm^2}{kV*min}$

Effective mobility transformation of whole CE-MS runs

`mobilityTransform` can also be applied to whole CE-MS runs stored either as `Spectra`-object or `OnDiskMSnExp`-object. As shown as before, either a single or two markers can be used for conversion. Here, we add the marker information of Procaine to the marker `data.frame`, so the transformation will be done on both markers according to eq. @ref(eq:mobility3).

```
procaine$markerID <- "Procaine"
procaine$mobility <- procaineMobility
marker <- rbind(marker, procaine)
```

Effective mobility transformation of OnDiskMSnExp objects

The migration time scale of `raw_data2` is transformed, returning the same class.

```
# Conversion of mt in OnDiskMSnExp objects
mobility_data <- mobilityTransform(x = raw_data, marker = marker)

# OnDiskMSnExp can by exported by writeMSData, Note that it is important to
# set copy = FALSE, (otherwise spectrum ordering will be wrong)
fl_mobility_data <- tempfile()
writeMSData(filterFile(mobility_data, 1),
    file = fl_mobility_data, copy = FALSE
)
```

Effective mobility transformation of Spectra objects

`Spectra`-objects can be transformed similarly in `mobilityTransform`. The resulting `Spectra`-objects can then also be exported as `.mzML` for further processing in different software.

```
# load the test data as spectra object
spectra_data <- Spectra(fl[1], backend = MsBackendMzR())

spectra_mobility <- mobilityTransform(
    spectra_data,
    marker[marker$fileIdx == 1, ]
)


# Transformed data can then be exported again as .mzML file to use in xcms or
# other software
fl_mobility <- tempfile()
export(spectra_mobility, MsBackendMzR(), file = fl_mobility)
```

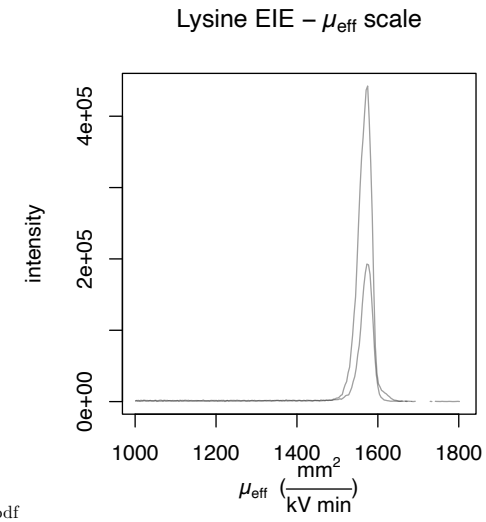
Inspect and Analyze the Transformed Data

The transformed data can displayed and further analyzed with e.g. `xcms`. Note that the effective mobility can be accessed by e.g. `spectra_mobility$rtime`, since functions where adapted from LC-MS based `rformassspectrometry` functions. The resulting unit of the effective mobility is $\frac{mm^2}{kV*min}$.

Here, we inspect them with the `plot(chromatogram())` function from `xcms`.

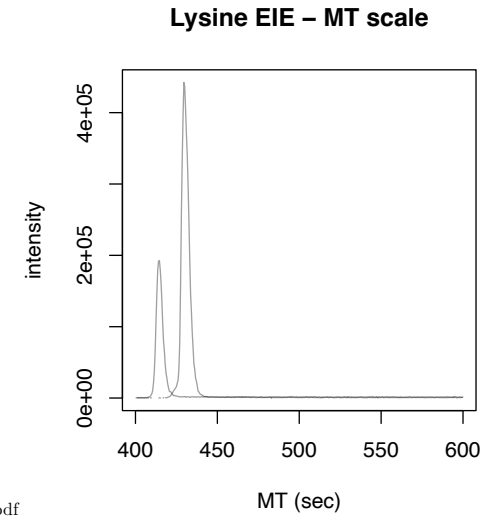
As described before, the test data contain different metabolites in 10 ppm and 25 ppm final concentration, acquired in positive ionization mode. We can check the effective mobilities of single compounds by extracting their EIE as for example Lysine (mz = 147.112806).

```
# Example: Extract ion electropherogram (EIE) from lysine
mz_lysine <- c(147.112806 - tolerance, 147.112806 + tolerance)
mobilityRestriction <- c(1000, 2500)

# Extract ion electropherogram of compound
lysine_EIE <- mobility_data |>
    filterMz(mz = mz_lysine) |>
    filterRt(rt = mobilityRestriction)

plot(chromatogram(lysine_EIE),
    main = expression(paste("Lysine EIE - μ"[eff], " scale")),
    xlab = expression(paste("μ"[eff], "  (", frac("mm"^2, "kV min"), ")"))
)
```



transformed data-1.pdf

```
# compare with extracted ion electropherogram of migration time scale
lysine_mt_EIE <- raw_data |>
    filterMz(mz = mz_lysine) |>
    filterRt(c(400, 600))

plot(chromatogram(lysine_mt_EIE),
    main = "Lysine EIE - MT scale",
    xlab = "MT (sec)"
)
```



transformed data-2.pdf

Session information

```
sessionInfo()
```

```
## R Under development (unstable) (2022-02-28 r81833)
## Platform: x86_64-apple-darwin17.0 (64-bit)
## Running under: macOS Big Sur/Monterey 10.16
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
##
## locale:
## [1] de_DE.UTF-8/de_DE.UTF-8/de_DE.UTF-8/C/de_DE.UTF-8/de_DE.UTF-8
##
## attached base packages:
## [1] stats4    stats     graphics  grDevices utils     datasets  methods
## [8] base
##
## other attached packages:
##  [1] Spectra_1.5.13           xcms_3.17.3
##  [3] BiocParallel_1.29.17     MobilityTransformR_0.99.7
##  [5] MSnbase_2.21.6           ProtGenerics_1.27.2
##  [7] S4Vectors_0.33.11        mzR_2.29.3
##  [9] Rcpp_1.0.8.2             Biobase_2.55.0
## [11] BiocGenerics_0.41.2      BiocStyle_2.23.1
##
```

```
## loaded via a namespace (and not attached):
##  [1] bitops_1.0-7                matrixStats_0.61.0
##  [3] fs_1.5.2                    doParallel_1.0.17
##  [5] RColorBrewer_1.1-2          GenomeInfoDb_1.31.5
##  [7] tools_4.2.0                 utf8_1.2.2
##  [9] R6_2.5.1                    affyio_1.65.0
## [11] DBI_1.1.2                   colorspace_2.0-3
## [13] tidyselect_1.1.2            MassSpecWavelet_1.61.0
## [15] compiler_4.2.0              preprocessCore_1.57.0
## [17] cli_3.2.0                   DelayedArray_0.21.2
## [19] scales_1.1.1                DEoptimR_1.0-10
## [21] robustbase_0.93-9           affy_1.73.0
## [23] stringr_1.4.0               digest_0.6.29
## [25] rmarkdown_2.13              XVector_0.35.0
## [27] pkgconfig_2.0.3             htmltools_0.5.2
## [29] MetaboCoreUtils_1.3.8       MatrixGenerics_1.7.0
## [31] highr_0.9                   fastmap_1.1.0
## [33] limma_3.51.5                rlang_1.0.2
## [35] rstudioapi_0.13             impute_1.69.0
## [37] generics_0.1.2              mzID_1.33.0
## [39] dplyr_1.0.8                 RCurl_1.98-1.6
## [41] magrittr_2.0.2              GenomeInfoDbData_1.2.7
## [43] Matrix_1.4-0                MALDIquant_1.21
## [45] munsell_0.5.0               fansi_1.0.2
## [47] MsCoreUtils_1.7.4           lifecycle_1.0.1
## [49] vsn_3.63.0                  stringi_1.7.6
## [51] yaml_2.3.5                  MASS_7.3-55
## [53] SummarizedExperiment_1.25.3 zlibbioc_1.41.0
## [55] plyr_1.8.6                  grid_4.2.0
## [57] parallel_4.2.0              crayon_1.5.0
## [59] lattice_0.20-45             MsFeatures_1.3.0
## [61] knitr_1.38                  pillar_1.7.0
## [63] GenomicRanges_1.47.6        codetools_0.2-18
## [65] XML_3.99-0.9                glue_1.6.2
## [67] evaluate_0.15               pcaMethods_1.87.0
## [69] BiocManager_1.30.16         vctrs_0.3.8
## [71] foreach_1.5.2               RANN_2.6.1
## [73] gtable_0.3.0                purrr_0.3.4
## [75] clue_0.3-60                 assertthat_0.2.1
## [77] ggplot2_3.3.5               xfun_0.30
## [79] ncdf4_1.19                  tibble_3.1.6
## [81] iterators_1.0.14            IRanges_2.29.1
## [83] cluster_2.1.2               ellipsis_0.3.2
```

References

1. Schmitt-Kopplin P, Garmash AV, Kudryavtsev AV, Menzinger F, Perminova IV, Hertkorn N, Freitag D, Petrosyan VS, Kettrup A: **Quantitative and qualitative precision improvements by effective mobility-scale data transformation in capillary electrophoresis analysis**. *ELECTROPHORESIS* 2001, **22**:77–87.
2. González-Ruiz V, Gagnebin Y, Drouin N, Codesido S, Rudaz S, Schappler J: **ROMANCE: A new software tool to improve data robustness and feature identification in CE-MS metabolomics**. *ELECTROPHORESIS* 2018, **39**:1222–1232.

B Appendix Chapter 3

B.1 Methods

B.1.1 Chemicals

Yeast extract for use in microbial growth medium, Tryptone, NaCl, Bacteriological agar, Cholesterol, $CaCl_2$, $MgSO_4$, K_2HPO_4, KH_2PO_4, Methanol (LC-MS grade), Methyl tertiary-butyl ether (MTBE, LC grade), 2-Propanol (LC-MS grade), NaOH, ACN (LC-MS grade), ammonium acetate, ammonium formiate, Bicinchoninic Acid Protein assay Kit, Ethylsulfate, Paracetamol and Procaine were purchased from Sigma-Aldrich (Darmstadt, Germany). Acetic acid and formic acid were purchased from from Fluka® Analytical (Munich, Germany) and Bacto Peptone from BD Biosciences

B.1.2 *C. elegans* culturing

For metabolomics analysis wild type N2 Bistrol and *daf-2(e1370)* mutant worms were used. Approximately 500 age-synchronized L1 worms were grown on NGM agar seeded with *Escherichia coli (E. coli)* (OP50) at 20 °C and harvested at young adult stage. Overnight cultures of *E. coli* (OP50) were grown at 36 °C in autoclaved liquid LB-medium composed of 5 g/L yeast extract, 10 g/L tryptone and 0.5 g/L NaCl in milliQ at pH 7.0. 100 µL of *E. coli* were spread on 10 cm NGM plates that were prepared using 3g/L NaCl, 2.5 g/L Bacto Peptone, 20g/L Agar,5 µg/mL Cholesterol, 1 mM $CaCl_2$, 1mM $MgSO_2$, 25mM PO_4^{3-} in milliQ water Worms were washed from NMG plates using M9 buffer, which was 22 mM KH_2PO_4, 42 mM Na_2HPO_4, 86 mM $CaCl_2$ and 1mM $MgSO_4$.

B.1.3 CE-MS method

Agilent MassHunter version B.09.00 was used for controlling and monitoring the instrument. Source parameters were optimized in positive and negative ionization mode separately. Briefly, cationic profiling was performed at 10 psi Nebulizer gas, gas flow at 8 L/min at 300°C and 3.5 L/min Sheath Gas at 195 °C. Fragmentor voltage was 400 V, Skimmer 65 V, Octopole 800 V, capillary 2000 V and Nozzle 2000 V. Anions were analyzed using 10 psi Nebulizer gas, gas flow at 5 L/min at 130 °C and 3.5 L/min Sheath Gas at 140 °C. Fragmentor voltage was 400 V, Skimmer 65 V, Octopole 800 V, capillary 2000 V and Nozzle 1800 V. QToF only MS acquisition was performed at m/z 50 - 950 for metabolite mixes and 50 - 1700 for *C. elegans* samples at 2 Hz. Worm extracts were additionally acquired in MS/MS mode with MS/MS m/z range 50 - 1700, MS/MS scan rate of 10 Hz, narrow isolation window and three fixed collision energies of 10, 20, and 40 eV, and max. precursors per cycle were 3 with min of 100 counts intensity.

B.1.4 HILIC-MS method

<table>
<tr><td>Column</td><td colspan="2">Agilent InfinityLab Poroshell 120 HILIC-Z column PEEK-lined (150 mm × 2.1 mm, 2.7 µm, 100 Å)</td></tr>
<tr><td></td><td>Cationic Profiling
(Positive ionization mode)</td><td>Anionic Profiling
(Negative ionization mode)</td></tr>
<tr><td>Solvent</td><td>A: H_2O + 10 mM ammonium formate + 0.1 % formic acid
B: 10 % H_2O + 90 % ACN + 10 mM ammonium formate + 0.1 % formic acid</td><td>A: H_2O + 10 mM ammonium acetate + 2.5 µM InfinityLab Deactivator Additive, pH = 9
B: 10 % H_2O + 90 % ACN + 10 mM ammonium acetate + 2.5 µM InfinityLab Deactivator Additive, pH = 9</td></tr>
<tr><td>Gradient</td><td>
<table>
<tr><th>TIME (MIN)</th><th>% A</th><th>% B</th></tr>
<tr><td>0</td><td>2</td><td>98</td></tr>
<tr><td>3</td><td>2</td><td>98</td></tr>
<tr><td>11</td><td>30</td><td>70</td></tr>
<tr><td>12</td><td>40</td><td>60</td></tr>
<tr><td>16</td><td>95</td><td>5</td></tr>
<tr><td>18</td><td>95</td><td>5</td></tr>
<tr><td>19</td><td>2</td><td>98</td></tr>
<tr><td>20</td><td>2</td><td>98</td></tr>
</table>
</td><td>
<table>
<tr><th>TIME (MIN)</th><th>% A</th><th>% B</th></tr>
<tr><td>0</td><td>4</td><td>96</td></tr>
<tr><td>2</td><td>4</td><td>96</td></tr>
<tr><td>5.5</td><td>12</td><td>88</td></tr>
<tr><td>8.5</td><td>12</td><td>88</td></tr>
<tr><td>9</td><td>14</td><td>86</td></tr>
<tr><td>14</td><td>14</td><td>86</td></tr>
<tr><td>17</td><td>18</td><td>82</td></tr>
<tr><td>23</td><td>35</td><td>65</td></tr>
<tr><td>24</td><td>35</td><td>65</td></tr>
<tr><td>24.5</td><td>4</td><td>96</td></tr>
<tr><td>26</td><td>4</td><td>96</td></tr>
</table>
</td></tr>
<tr><td>Post Time</td><td>4 min</td><td>3 min</td></tr>
<tr><td>Column Temperature</td><td>25 °C</td><td>50 °C</td></tr>
<tr><td>Flow Rate</td><td colspan="2">0.25 mL/min</td></tr>
<tr><td>Injection Volume</td><td colspan="2">3 µL</td></tr>
</table>

QToF only MS acquisition was performed at m/z 50 - 950 at 1 Hz. Worm extracts were acquired in MS/MS mode with MS/MS m/z range 50 - 1700. For positive ionization mode source parameters were 40 psi Nebulizer gas, gas flow at 5 L/min at 225 °C and 10 L/min Sheath Gas at 225 °C. Fragmentor voltage was 400 V, Skimmer 65 V, Octopole 750 V, capillary 3000 V and Nozzle 0 V. Anionic profiling was performed using 35 psi Nebulizer gas, gas flow at 13 L/min at 225 °C and 12 L/min Sheath Gas at 350 °C. Fragmentor voltage was 400 V, Skimmer 65 V, Octopole 750 V, capillary 3500 V and Nozzle 0 V.

B.1.5 CentWave Parameters

	HILIC	CE - MT	CE - mobility
peakwidth	c(10, 200)	c(10, 150)	c(10, 300)
noise	2000	1000	2000 (neg:1600)
snthresh	20	3	20
ppm	30	10	30
mzdiff		0,01	
prefilter	c(4, 1000)	c(5, 1000)	c(4, 500)
integrate	2	1	2

B.1.6 Compare physicochemical properties of *C. elegans* metabolites

In order to investigate the underlying base of the differences in the annotated features, we compared different molecular descriptors such as LogP, number of the longest chain, number of rotational bonds, number of bonds, number of aromatic bonds, number of aromatic atoms, number of hydrogen-bridge donors and acceptors, the topological polar surface area, the sum of the atomic polarizabilities, molecular weight, and number of basic and acidic groups. To calculate the molecular descriptors we used the rcdk R package and the SMILES of the molecules as the respective pH (pH 9 for HILIC neg and pH 2 for CE and HILIC pos; SMILES generated with MarvinSketch). See Table B.3.

B.2 Figures

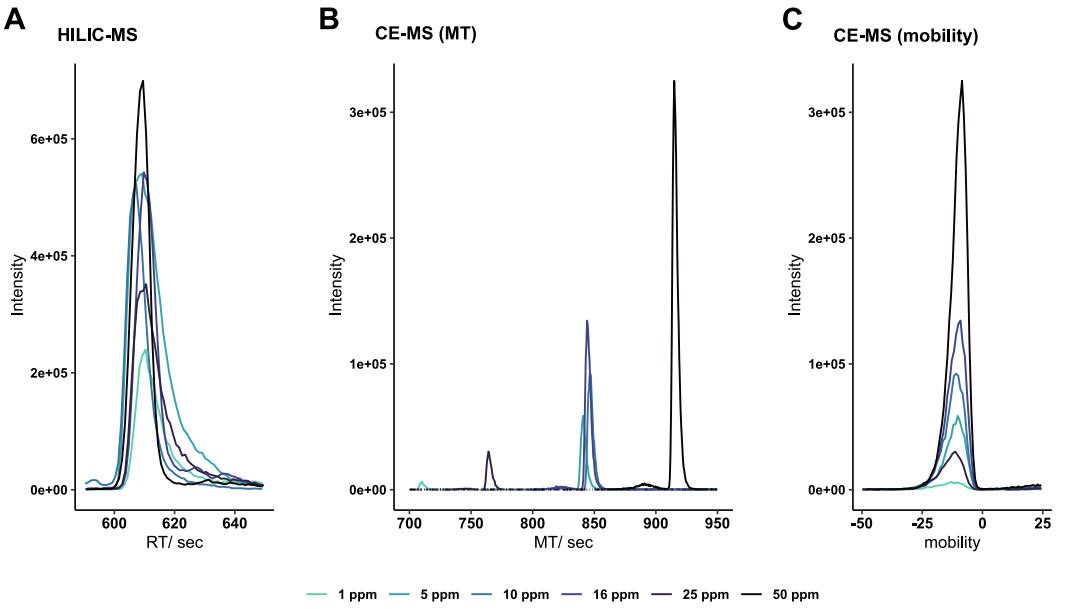


Figure B.1 Extracted ion chromatogram (EIC)/ extracted ion electropherogram (EIE) of adenosine monophosphate (AMP) in different concentrations (1-50 ppm), from left to right: HILIC, CE (MT scale), CE (mobility scale).

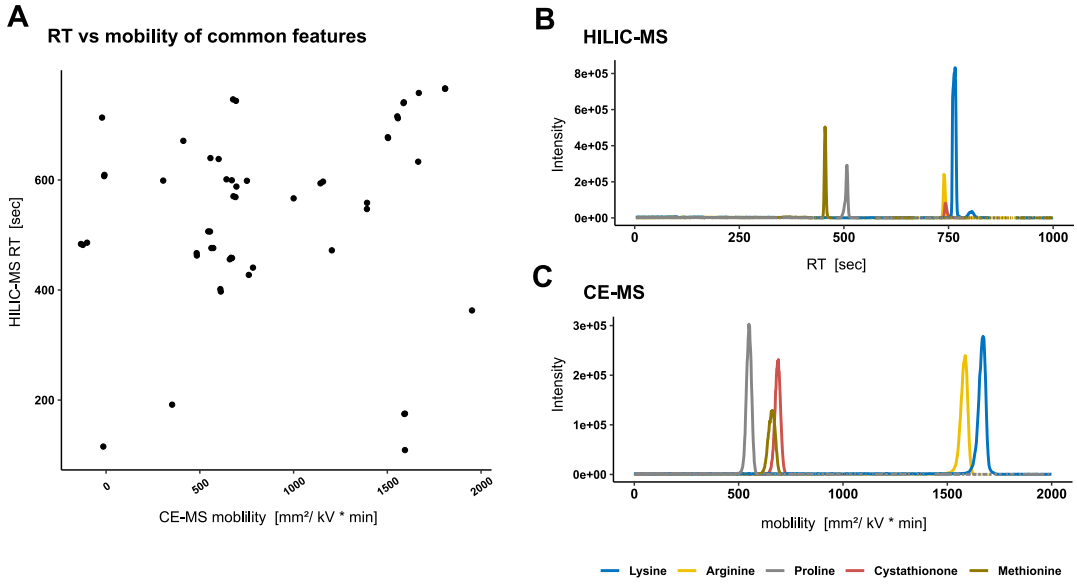


Figure B.2 A: Retention time vs effective mobility of commonly detected model metabolites **B:** EIC of selected model metabolites **C:** EIE of selected model metabolites

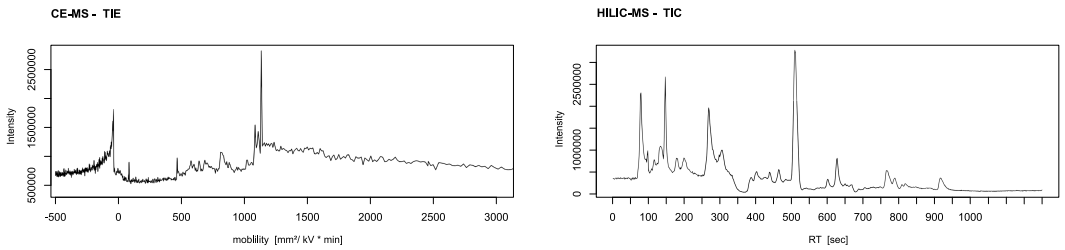


Figure B.3 Total ion electropherogram (TIE) and total ion chromatogram (TIC) of QC sample measures in positive ionization (and separation) mode in CE-MS and HILIC MS

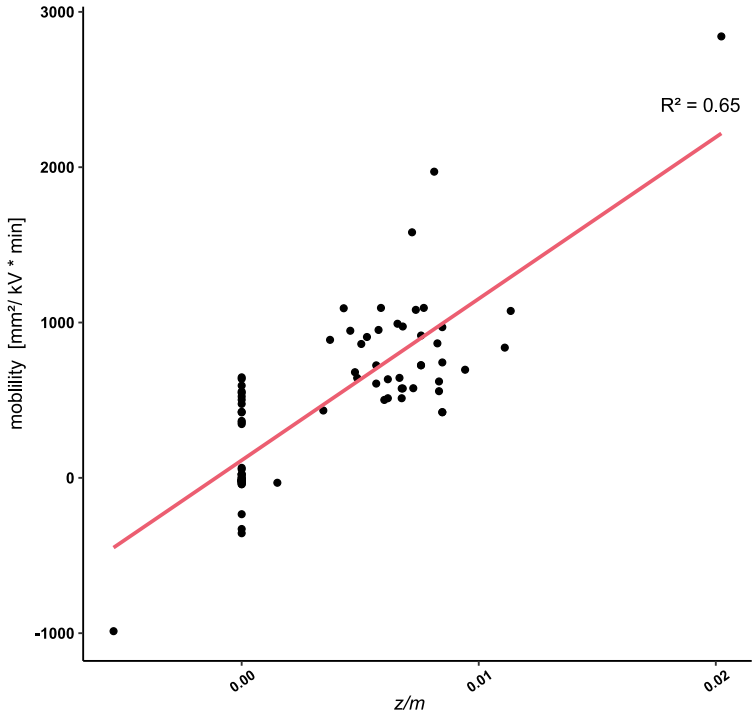


Figure B.4 Separation principle of CE analysis demonstrated by the mobility vs. charge-to-mass ratio (z/m) of the metabolites annotated in *C. elegans*

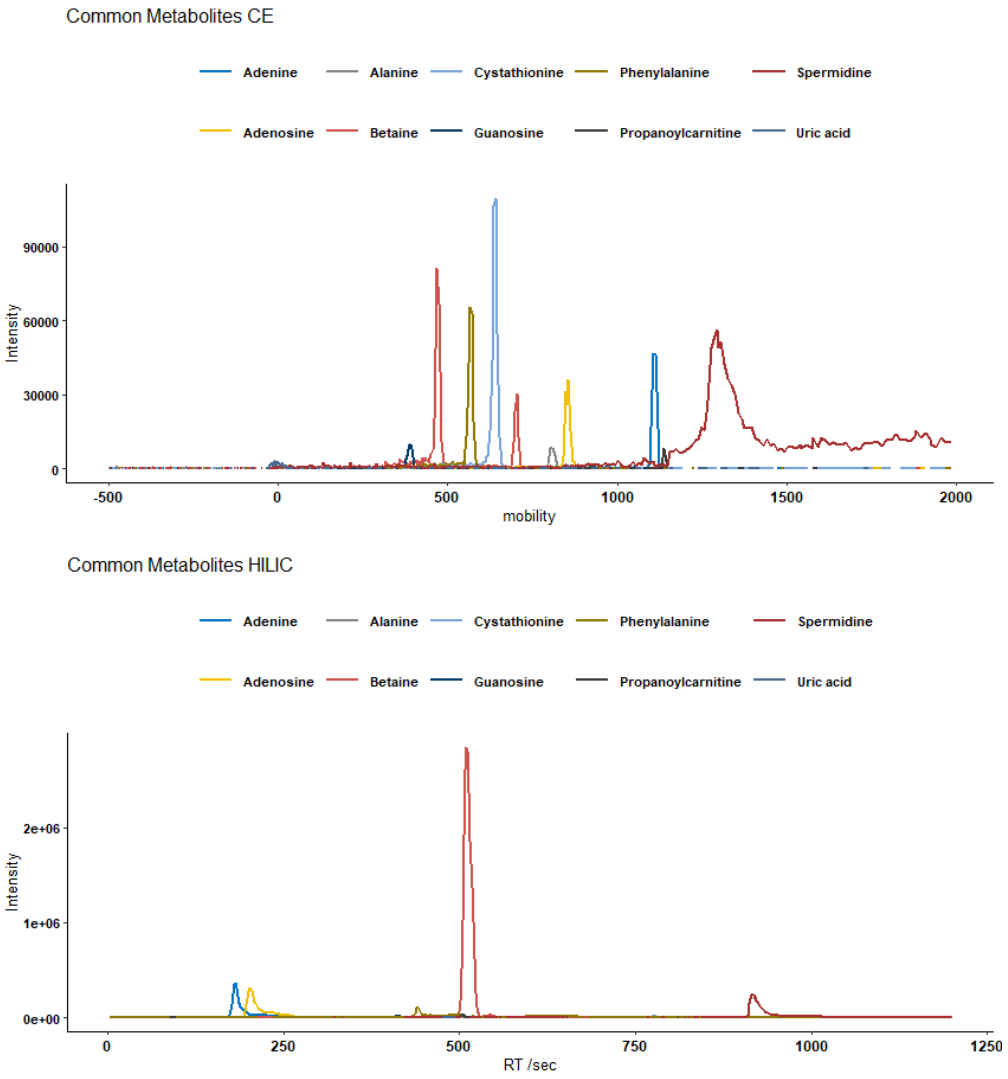


Figure B.5 Separation of a selection of commonly detected metabolites in QC sample

B.3 Tables

			pos													
			CE-MS (μeff-scale)				CE-MS (MT-scale)					HILIC-MS				
Mix*	Mass	Name	μeff (pos)	RSD μeff (pos)	FWHM μeff (pos)	μeff (pos) min conc	MT (pos)	RSD MT (pos)	FWHM MT (pos)	N MT (pos)	MT (pos) min conc	RT (pos)	RSD RT (pos)	FWHM RT (pos)	N RT (pos)	RT (pos) min conc
E	169.0851	1-METHYL-L-HISTIDINE	1503.30	0.15	35.80	1	448.90	6.95	5.11	42700	1	676.26	0.20	12.07	17390	1
F	203.1158	Acetyl-Carnitine	1203.98	5.87	51.63	1	492.27	18.35	43.85	698	1	472.11	0.14	6.82	26562	1
E	135.0545	ADENINE	1594.02		19.94	50	479.94		2.16	273750	50	175.39	5.08	19.99	426	1
E	267.0968	ADENOSINE	352.26		39.76	50	750.32		13.64	16766	50	191.66	1.96	12.66	1270	1
E	89.0477	Alanine	1000.27	1.96	26.43	10	561.04	8.20	6.66	39271	10	566.61	0.34	6.63	40463	10
C / D	347.0631	AMP	-8.75	-125.58	8.99	1	840.12	7.06	4.82	168036	1	609.35	0.28	8.58	27947	1
G	174.1117	Arginine	1586.58	0.28	34.79	1	400.76	3.82	4.45	44920	1	739.71	0.04	6.68	67847	1
D	132.0535	Asparagine	670.46	0.97	27.15	10	577.62	3.95	33.94	1605	10	599.44	0.08	4.82	85630	1
F	133.0375	Aspartic acid	412.06	16.69	38.99	1	685.52	8.71	15.49	10845	10	671.09	0.10	6.14	66219	10
C		ATP														
F	122.0368	BENZOATE	561.67	0.49	130.79	25	612.70	0.19	110.76	170	25	476.30	0.10	25.12	1992	10
G	117.0790	BETAINE	483.58	1.46	31.84	1	610.59	5.93	9.12	24840	1	462.70	0.18	11.97	8279	1
D	244.0882	BIOTIN	-14.68	-10.07	15.27	10	826.99	7.74	8.89	47985	10	115.51	4.87	10.60	658	1
F	161.1052	Carnitine	1391.45	0.86	76.07	1	472.24	20.62	12.10	8433	1	558.47	0.46	34.18	1479	1
G	226.1066	CARNOSINE	1808.05	0.34	42.39		376.27	3.86	4.58	37455	1	765.61	0.04	5.50	107199	1
G	104.1075	CHOLINE	1951.56	0.52	51.49	10	362.70	3.53	5.12	27843	10	362.88	0.60	15.27	3127	1
C / G	174.0164	cis-Aconitate					1036.58		15.05	26293	50	204.97		8.63	3122	50
C	192.0270	Citrate	-130.60	-8.77	13.71	25	969.18	5.44	12.19	35008	50	481.99		2.99	144250	50
F	222.0674	CYSTATHIONINE	690.26	2.55	31.94	1	585.92	10.06	8.54	26079	1	743.80	0.08	7.38	56239	1
D		Cysteine														
E	441.1397	FOLATE	304.74	15.02	41.06	10	727.92	2.02	15.07	12927	10	598.86	7.99	8.41	28091	1
C	339.9961	Fructose-1,6-bis-P										661.77	6.18	26.00	3588	1
C		Fumarate					982.36	4.80	48.10	2311	25					
C	180.0634	Glucose														
C	260.0297	Glucose-6-P										614.43	6.44	39.13	1366	10
G	147.0532	Glutamic acid	595.93	7.07	27.78	1	579.49	6.22	8.50	25756	1	638.08	0.07	5.89	64954	10
C	146.0691	Glutamine	642.17	2.66	44.51	1	589.06	5.39	11.97	13406	1	601.15	0.67	7.19	38711	1
G	132.0423	Glutaric Acid	-23.03	-2.51	15.12	10	816.77	7.79	7.88	59544	10					
C / D	169.9980	Glyceraldehyde-3-P					585.90		55.71	613	50	656.60	4.96	43.92	1238	1
F	106.0266	GLYCERATE	-65.15		13.48	50										
D		Glycine										588.92	0.09	2.87	233050	10
E	155.0695	Histidine	1555.06	0.33	35.19	1	449.50	7.97	5.12	42643	1	712.13	0.60	10.81	24052	1
F	135.0354	HOMOCYSTEINE	689.81	6.58	20.85	1	651.05	18.55	9.23	27555	1	588.01	26.69	9.37	21808	25
D	192.0270	Isocitrate	-101.36		20.30	50	947.25		14.15	24823	50					
D	131.0946	Isoleucine	783.56	0.71	34.62	1	545.18	4.51			1	440.68	0.20	7.00	21960	1
C	90.0317	Lactate														
C	131.0946	Leucine	761.25	2.66	47.39	1	562.32	4.88	13.62	9448	1	427.45	2.45	7.25	19281	1
E	146.1055	Lysine	1666.97	0.66	35.88	1	422.43	18.01	8.30	14359	1	758.12	1.11	7.18	61796	1
C	134.0215	Malate	659.55	2.79	54.26	10	582.32	4.12	12.40	12212	10	458.08	0.71	12.43	7526	16
C	149.0511	Methionine	661.86	2.25	44.38	1	581.44	6.14	12.88	11284	1	455.85	0.84	9.23	13500	1
D	663.1091	NAD+	-21.37	-3.70	19.13	10	858.28	2.92	10.98	33861	10	713.36	0.11	6.46	67492	1
G		NADH										707.20		15.68	11269	50
E	122.0480	NICOTINAMIDE	1594.03	0.21	29.53	16	443.77	5.78	3.98	68823	10	109.12	1.06	14.16	329	1
G	131.0946	NORLEUCINE	771.46	0.93	43.48	10	525.04	9.20	9.39	17305	5					
F	132.0059	Oxaloacetate														
D	165.0790	phe-ala	611.08	0.88	30.70	1	589.42	5.37	8.47	26810	1	397.25	0.14	9.92	8891	10
D	136.0524	PHENYL ACETATE	1590.79	0.21	75.08	16	449.63	25.75			25	174.71	0.25	26.21	246	1
D	164.0473	PHENYLPYRUVATE	592.34	2.35	40.30	25	614.81	2.21	11.70	15308	25					
F	115.0633	Proline	550.80	2.28	28.88	1	626.75	10.57	18.81	6148	1	506.32	0.08	8.64	19020	1
C	88.0160	Pyruvate					564.73	3.88			10					
E	384.1216	S-ADENOSYLHOMOCYSTEINE	1147.42	1.45	36.63	1	554.10	23.73	6.80	36803	1	593.89	0.30	7.78	32314	1
E		S-ADENOSYLMETHIONINE														
C	105.0426	Serine	751.05	2.45	49.58	1	564.73	3.88			10	598.51	0.75			10
E	145.1579	SPERMIDINE					674.51	52.74	18.07	7718	50	860.24	5.87	60.54	1119	1
C	118.0266	Succinate	1596.77	0.05	113.73	25	575.17	6.01	10.58	16367	1					
C	119.0582	Threonine	681.64	2.43	43.74	1	630.85	14.82	8.26	32319	1	570.43	0.53	5.22	66154	1
C	204.0899	Tryptophan										418.81	1.03	13.39	5419	1
F	181.0739	Tyrosine	571.74	2.51	26.82	1						476.30	0.10	7.19	24319	10
G	102.0681	Valeric acid														
D	117.0790	Valine														

Table B.1 60 Reference Standards used for the comparison of CE-MS and HILIC-MS. Table includes detected peaks in each method (for CE data in migration time and mobility/μeff scale), relative standards deviation (RSD) between detected concentrations, full width at half maximum (FWHM) of the peaks, the number of theoretical plates N and the minimum detected concentration of the standard.

			neg													
			CE-MS (μeff-scale)				CE-MS (MT-scale)					HILIC-MS				
Mix*	Mass	Name	μeff (neg)	RSD μeff (neg)	FWHM μeff (neg)	min conc μeff (neg)	MT (neg)	RSD MT (neg)	FWHM MT (neg)	N MT (neg)	min conc MT (neg)	RT (neg)	RSD RT (neg)	FWHM RT (neg)	N RT (neg)	min conc RT (neg)
E	169.0851	1-METHYL-L-HISTIDINE										367.83	0.89	7.85	12164	10
F	203.1158	Acetyl-Carnitine														
E	135.0545	ADENINE										127.86	0.58			1
E	267.0968	ADENOSINE										134.31	0.68	5.45	3366	1
E	89.0477	Alanine										387.23	0.82	7.02	16863	10
C / D	347.0631	AMP	-9.73	-6.31	7.65	10	979.12	3.31	5.52	174423	1	654.35	0.17	40.10	1475	25
G	174.1117	Arginine										922.60	0.08	29.09	5573	10
D	132.0535	Asparagine										442.14	0.25	21.77	2285	1
F	133.0375	Aspartic acid					1499.30		12.59	78554	16	628.83	0.04	11.87	15549	10
C		ATP										1124.23		11.90	49448	50
F	122.0368	BENZOATE										139.09	0.74	5.29	3835	10
G	117.0790	BETAINE														
D	244.0882	BIOTIN					982.93	2.07	10.64	47324	16	254.02	0.81	6.53	8391	1
F	161.1052	Carnitine														
G	226.1066	CARNOSINE										709.15	0.28	12.75	17131	1
G	104.1075	CHOLINE														
C / G	174.0164	cis-Aconitate	-315.22	-54.12	15.29	10	799.85	10.32	7.54	62413	10					
C	192.0270	Citrate	-141.62	-2.67	19.12	10	888.13	3.01	11.43	33450	10					
F	222.0674	CYSTATHIONINE										860.07	0.42	17.97	12690	5
D		Cysteine										377.41		15.77	3174	50
E	441.1397	FOLATE					1271.88	2.80	28.77	10830	10	1101.06	8.89	16.05	26078	10
C	339.9961	Fructose-1,6-bis-P														
C		Fumarate														
C	180.0634	Glucose														
C	260.0297	Glucose-6-P					761.44	17.40	103.21	302	25					
G	147.0532	Glutamic acid										637.31	0.19	11.24	17822	1
C	146.0691	Glutamine										443.54	0.24	9.17	12953	1
G	132.0423	Glutaric Acid	-20.52	-4.89	15.01	10	957.06	1.32	10.77	43759	10	734.86	0.10	13.06	17550	1
C / D	169.9980	Glyceraldehyde-3-P										827.21	27.96			1
F	106.0266	GLYCERATE	-68.50		16.04	50	929.78	1.69	10.70	41860	10	203.86		96.22	25	16
D		Glycine														
E	155.0695	Histidine										455.23	1.42	16.57	4183	1
F	135.0354	HOMOCYSTEINE										290.93	1.06	6.35	11641	16
D	192.0270	Isocitrate	-100.86	-1.08	18.46	10	909.27	2.18	12.08	31392	10	953.30	22.39	67.91	1092	10
D	131.0946	Isoleucine														
C	90.0317	Lactate										321.27	56.08			10
C	131.0946	Leucine														
E	146.1055	Lysine										1010.29	0.74	11.44	43184	10
C	134.0215	Malate	-89.59	-2.59	16.49	10	915.09	2.58			1					
C	149.0511	Methionine									258.765167	0.89	11.61	2751.21	1	
D	663.1091	NAD+										723.81	0.45	8.10	44187	10
G		NADH														
E	122.0480	NICOTINAMIDE										138.70	0.42	7.61	1842	25
G	131.0946	NORLEUCINE														
F	132.0059	Oxaloacetate										184.21		8.93	2357	10
D	165.0790	phe-ala										210.23	0.16	12.61	1541	1
D	136.0524	PHENYL ACETATE										138.73	4.22	7.46	1914	1
D	164.0473	PHENYLPYRUVATE										124.71	25.63	16.07	334	1
F	115.0633	Proline										310.90	0.10	6.57	12412	10
C	88.0160	Pyruvate	-89.59	-2.59	16.49	10						153.30	0.94	9.32	1497	10
E	384.1216	S-ADENOSYLHOMOCYSTEINE										438.00	0.93	9.95	10739	1
E		S-ADENOSYLMETHIONINE										384.72	0.88	10.95	6844	25
C	105.0426	Serine										437.39	0.08	8.69	14051	10
E	145.1579	SPERMIDINE														
C	118.0266	Succinate	-26.94	-3.43	15.67	10	961.55	2.61	10.96	42672	10					
C	119.0582	Threonine										389.55	2.22	8.79	10883	1
C	204.0899	Tryptophan										232.48	1.74	11.34	2329	10
F	181.0739	Tyrosine										293.09	0.11	6.36	11762	10
G	102.0681	Valeric acid										103.78		4.18	3422	1
D	117.0790	Valine										237.68	0.30	8.05	4824	16

Table B.1 (Continued) 60 Reference Standards used for the comparison of CE-MS and HILIC-MS. Table includes detected peaks in each method (for CE data in migration time and mobility/μeff scale), relative standards deviation (RSD) between detected concentrations, full width at half maximum (FWHM) of the peaks, the number of theoretical plates N and the minimum detected concentration of the standard.

Metabolite	SMILES	XLogP	nRotB	nAtomLC	nB	nAtom	nAromBond	naAromAtom	nHBDon	nHBAcc	apol	TopoPSA	MW	nBase	nAcid
1-METHYL-L-HISTIDINE	CN1C=C(N=C1)CC(C(=O)O)N	-2.811	3	8	12	23	0	0	2	5	24.558723	78.92	169.181419	1	1
Acetyl-Carnitine	CC(=O)OC(CC(=O)[O-])C[N+](C)(C)C	-1.001	6	7	13	31	0	0	0	4	31.483481	66.43	203.235939	2	1
ADENINE	Nc1ncnc2[nH]cnc12	-0.337	0	5	11	15	10	9	2	4	17.633965	80.48	135.126899	0	0
ADENOSINE	Nc1ncnc2n(cnc12)[C@@H]1O[C@H](CO)[C@@H](O)[C@H]1O	-1.417	2	10	21	32	10	9	4	8	34.976309	139.54	267.241725	0	0
Alanine	CC(N)C(O)=O	-2.824	1	4	5	13	0	0	2	3	12.651551	63.32	89.093306	1	1
AMP	C1=NC(=C2C(=N1)N(C=N2)C3C(C(C(O3)COP(=O)(O)O)O)O)N	-4.025	4	12	25	37	0	0	5	12	41.679102	192.6	347.221642	0	2
Arginine	NC(CCCNC(N)=N)C(O)=O	-2.924	6	9	11	26	0	0	5	6	25.899102	125.22	174.201209	4	1
Asparagine	NC(CC(N)=O)C(O)=O	-4.435	3	6	8	17	0	0	3	5	16.980344	106.41	132.118091	1	1
Aspartic acid	NC(CC(O)=O)C(O)=O	-3.707	3	6	8	16	0	0	3	5	16.015551	100.62	133.102852	1	2
ATP	Nc1ncnc2n(cnc12)[C@@H]1O[C@H](COP(O)(=O)OP(O)(=O)OP(O)(O)=O)[C@@H](O)[C@H]1O	-5.533	8	16	33	47	10	9	7	17	55.084688	308.56	507.181476	0	4
BENZOATE	OC(=O)c1ccccc1	1.637	1	6	9	15	6	6	1	2	17.924758	37.3	122.121606	0	1
BETAINE	C[N+](C)(C)CC(=O)[O-]	-1.227	2	5	7	19	0	0	0	2	18.838723	40.13	117.146541	2	1
BIOTIN	[H][C@]12CS[C@@H](CCCCC(O)=O)[C@@]1([H])NC(=O)N2	0.463	5	11	17	32	0	0	3	5	35.774688	103.73	244.312117	0	1
Carnitine	C[N+](C)(C)CCC(O)CC([O])=O	-1.26	4	7	10	26	0	0	1	3	25.827895	37.3	161.199181	2	0
CARNOSINE	NCCC(=O)N[C@@H](CC1=CNC=N1)C(O)=O	-3.785	7	10	16	30	0	0	4	7	31.981102	116.81	226.232821	1	1
CHOLINE	C[N+](C)(C)CCO	-0.837	2	5	6	21	0	0	1	1	20.037102	20.23	104.170958	2	0
cis-Aconitate	OC(=O)C\C(=C\C(O)=O)C(O)=O	-1.034	4	7	11	18	0	0	3	6	19.372758	111.9	174.108489	0	3
cis-aconitic acid	OC(=O)C\C(=C\C(O)=O)C(O)=O	-1.034	4	7	11	18	0	0	3	6	19.372758	111.9	174.108489	0	3
Citrate	OC(=O)CC(O)(CC(O)=O)C(O)=O	-2.247	5	7	12	21	0	0	4	7	21.508344	132.13	192.123776	0	3
CYSTATHIONINE	NC(CCSCC(N)C(O)=O)C(O)=O	-3.209	7	10	13	28	0	0	4	6	29.963102	151.94	222.263433	2	2
Cysteine	NC(CS)C(O)=O	-2.575	2	5	6	14	0	0	2	3	15.551551	102.12	121.159391	1	1
FOLATE	Nc1nc2nccC(NCc3ccc(cc3)C(=O)N[C@@H](CCC(O)=O)C(O)=O)nc2c(=O)[nH]1	-3.096	10	20	34	51	17	16	6	12	58.621067	213.28	441.398208	0	2
Fructose-1,6-bis-P	C(C1C(C(CO1)(COP(=O)(O)O)O)O)OP(=O)(O)O	-5.056	6	11	20	34	0	0	7	12	36.779102	223.06	340.115968	0	4
Fumarate	OC(=O)\C=C\C(O)=O	-0.416	2	6	7	12	0	0	2	4	12.915172	74.6	116.072326	0	2
Glucose	C(C1C(C(C(CO1)O)O)O)O	-1.697	1	7	12	24	0	0	5	6	23.373516	110.38	180.156134	0	0
Glucose-6-P	C(C1C(C(C(CO1)O)O)O)OP(=O)(O)O	-3.355	3	9	16	29	0	0	6	9	30.076309	166.72	260.136051	0	2
Glutamic acid	NC(CCC(O)=O)C(O)=O	-3.349	4	7	9	19	0	0	3	5	19.109137	100.62	147.129469	1	2
Glutamic acid	N[C@@H](CCC(O)=O)C(O)=O	-3.349	4	7	9	19	0	0	3	5	19.109137	100.62	147.129469	1	2
Glutamine	NC(CCC(N)=O)C(O)=O	-4.077	4	7	9	20	0	0	3	5	20.07393	106.41	146.14708	1	1
Glutamine	NC(CCC(N)=O)C(O)=O	-4.077	4	7	9	20	0	0	3	5	20.07393	106.41	146.144708	1	1
Glutaric Acid	OC(=O)CCCC(O)=O	-0.352	4	7	8	17	0	0	2	4	17.342344	74.6	132.114825	0	2
GLYCERALDEHYDE	[H]C(=O)C(O)CO	-1.309	2	5	5	12	0	0	2	3	11.686758	57.53	90.078067	0	0
Glyceraldehyde-3-P	[H]C(=O)C(O)COP(O)(O)=O	-2.538	4	7	9	17	0	0	3	6	18.389551	113.87	170.057984	0	2
Glyceraldehyde-3-P	[H]C(=O)C(O)COP(O)(O)=O	-2.538	4	7	9	17	0	0	3	6	18.389551	113.87	170.057984	0	2
GLYCERATE	OCC(O)C(O)=O	-1.723	2	5	6	13	0	0	3	4	12.488758	77.76	106.077472	0	1
Glycine	NCC(O)=O	-3.35	1	4	4	10	0	0	2	3	9.557965	63.32	75.066886	1	1
Histidine	NC(Cc1c[nH]cn1)C(O)=O	-3.115	3	7	11	20	5	5	3	4	21.465137	92	155.154802	1	1
HOMOCYSTEINE	NC(CCS)C(O)=O	-2.217	3	6	7	17	0	0	2	3	18.645137	102.12	135.186008	1	1
Isocitrate	OC(CC(O)=O)C(O)=O)C(O)=O	-2.064	5	7	12	21	0	0	4	7	21.508344	132.13	192.123776	0	3
Isoleucine	CCC(C)C(C(=O)O)N	-1.6	3	6	8	22	0	0	2	3	21.932309	63.32	131.173158	1	1
Lactate	C[C@H](O)C(O)=O	-0.591	1	4	5	12	0	0	2	3	11.686758	57.53	90.078067	0	1
Leucine	CC(C)CC(N)C(O)=O	-1.389	3	6	8	22	0	0	2	3	21.932309	63.32	131.173158	1	1
Lysine	NCCCCC(N)C(O)=O	-2.949	5	8	9	24	0	0	3	4	23.699102	89.34	146.187802	2	1
Malate	OC(CC(O)=O)C(O)=O	-1.474	3	6	8	15	0	0	3	5	15.050758	94.83	134.087613	0	2
Methionine	CSCCC(N)C(O)=O	-1.853	4	7	8	20	0	0	2	3	21.738723	88.62	149.212626	1	1
NAD+	NC(=O)c1ccc[n+](c1)[C@@H]1O[C@H](COP(O)(=O)OP(O)(=O)OC[C@H]2O[C@H]([C@H](O)[C@@H]2O)n2cnc3c(N)ncnc23)[C@@H](O)[C@H]1O	-4.2	11	23	48	72	16	15	8	19	81.818204	337.88	664.43391	1	2
NADH	NC(=O)C1=CN(C=CC1)[C@@H]1O[C@H](COP(O)(=O)OP(O)(=O)OC[C@H]2O[C@H]([C@H](O)[C@@H]2O)n2cnc3c(N)ncnc23)[C@@H](O)[C@H]1O	-4.989	11	23	48	73	10	9	8	20	82.484997	337.24	665.44185	0	2
NICOTINAMIDE	NC(=O)c1cccnc1	-0.343	1	6	9	15	6	6	1	3	17.562758	55.98	122.124871	0	0
NORLEUCINE	CCCCC(C(=O)O)N	-1.328	4	7	8	22	0	0	2	3	21.932309	63.32	131.173158	1	1
Oxaloacetate	OC(=O)CC(=O)C(O)=O	-0.645	3	6	8	13	0	0	2	5	13.717172	91.67	132.071731	0	2
Palmitoyl-CoA	[C@@H]1(N2C3=C(C(=NC=N3)N)N=C2)O[C@H](COP(OP(OCC(C)([C@H](C(NCCC(NCCSC(CCCCCCCCCCCCCC)=O)=O)=O)O)C)(=O)O)(=O)O)O[C@H]([C@H]1O)OP(O)(O)=O	0.629	36	42	67	131	0	0	9	24	144.252338	415.08	1005.9459	0	4
phe-ala	C[C@H](NC(=O)[C@@H](N)Cc1ccccc1)C(O)=O	-1.681	6	11	17	33	6	6	3	5	36.394688	92.42	236.267504	1	1
PHENYL ACETATE	CC(=O)Oc1ccccc1	1.807	2	7	10	18	6	6	0	1	21.018344	26.3	136.148223	0	0
PHENYLPYRUVATE	C1=CC=C(C=C1)CC(=O)C(=O)O	0.907	3	8	12	20	0	0	1	3	23.580344	54.37	164.158364	0	1
Proline	OC(=O)C1CCCN1	-0.185	1	5	8	17	0	0	2	3	17.505137	49.33	115.130659	1	1
Pyruvate	CC(=O)C(O)=O	-0.426	1	4	5	10	0	0	1	3	10.353172	54.37	88.062185	0	1
S-ADENOSYLHOMOCYSTEINE	C1=NC(=C2C(=N1)N(C=N2)C3C(C(C(O3)CSCCC(C(=O)O)N)O)O)N	-4.215	7	15	28	46	0	0	5	11	51.48586	204.65	384.412447	1	1
S-ADENOSYLMETHIONINE	C[S+](CC[C@H](N)C(O)=O)C[C@H]1O[C@H]([C@H](O)[C@@H]1O)n1cnc2c(N)ncnc12	-2.992	7	15	29	50	10	9	5	10	55.246239	182.63	399.447005	2	1
Serine	NC(CO)C(O)=O	-3.956	2	5	6	14	0	0	3	4	13.453551	83.55	105.092711	1	1
SPERMIDINE	NCCCNCCCN	-1.012	6	9	8	26	0	0	3	3	25.195481	64.07	131.219518	3	0
Succinate	OC(=O)CCC(O)=O	-0.71	3	6	7	14	0	0	2	4	14.248758	74.6	118.088208	0	2
Threonine	CC(O[C@H](O)C)N)O	-3.496	2	5	7	17	0	0	3	4	16.547137	83.55	119.119328	1	1
Tryptophan	NC(Cc1c[nH]c2ccccc12)C(O)=O	-1.271	3	9	16	27	10	9	3	3	31.165516	79.11	204.2256	1	1
Tyrosine	NC(Cc1ccc(O)cc1)C(O)=O	-1.785	3	9	13	24	6	6	3	3	26.680723	83.55	181.188889	1	1
Valeric acid	CCCCC(O)=O	1.311	3	6	6	17	0	0	1	2	17.07193	37.3	102.131897	0	1
Valine	CC(C)C(N)C(O)=O	-2.169	2	5	7	19	0	0	2	3	18.838723	63.32	117.146541	1	1

Abbreviations:

XLogP	XLogP
nRotB	RotatableBondsCount
nAtomLC	LargestChain
nB	BondCount
nAtom	AtomCount
naAromBond	AromaticBondsCount
naAromAtom	AromaticAtomsCount
nHBDon	HBondDonorCount
nHBAcc	HBondAcceptorCount
apol	Sum of the atomic polarizabilities
TopoPSA	topological polar surface area (TPSA)
MW	molecularWeight
nBase	BasicGroupCount
nAcid	AcidicGroupCount

Table B.2 60 Reference Standards used for the comparison of CE-MS and HILIC-MS and their physicochemical properties

Name	Formula	SMILES	method	ion	msi	XLogP	nRotB	nAtomLC	nB	nAtom	nAromBond	naAromAtom	nHBDon	nHBAcc	apol	TopoPSA	MW	nBase	nAcid	Charge	mobility / RT
2-Acetamido-2-dexyglucosamine	C8H16N2O5	CC(=O)N[C@H]1[C@H](N)O[C@H](CO)[C@@H](O)[C@@H]1O	HILIC	neg	msi 1	-1.984	3	9	15	31	0	0	5	7	30.95869	125.04	220.2234	1	0	0	222.991
2-Aminoadipic acid	C6H11NO4	[NH3+][C@@H](CCCC([O-])=O)C([O-])=O	HILIC	neg	msi 2	-3.419	5	8	10	21	0	0	1	4	21.53593	107.9	160.1481	1	2	-1	695.1206
2-Phosphoglyceric acid	C3H7O7P	OCC(OP([O-])([O-])=O)C([O-])=O	HILIC	neg	msi 1	-3.433	4	6	10	15	0	0	1	7	17.19117	142.59	183.0336	0	3	-3	1041.279
Adenine	C5H5N5	NC1=C2NC=NC2=NC=N1	HILIC	neg	msi 2	-1.287	0	5	11	15	0	0	2	5	17.63397	75.13	135.1269	0	0	0	127.065
AMP	C10H14N5O7P	NC1=C2N=CN(C3OC(COP([O-])([O-])=O)C(O)C3O)C2=NC=N1	HILIC	neg	msi 1	-4.025	4	12	25	35	0	0	3	12	40.34552	198.26	345.2058	0	2	-2	631.9265
Anthranilic acid	C7H7NO2	Nc1ccccc1C([O-])=O	HILIC	neg	msi 2	0.339	1	6	10	16	6	6	1	3	19.02476	66.15	136.1283	0	1	-1	92.64033
Arginine	C6H14N4O2	NC(N)=[NH+]CCCC([NH3+])C([O-])=O	HILIC	neg	msi 1	-3.651	5	9	11	27	0	0	4	4	26.5659	133.78	175.2091	5	1	1	904.3606
Citric acid	C6H8O7	OC(CC([O-])=O)(CC([O-])=O)C([O-])=O	HILIC	neg	msi 1	-3.69	5	7	12	18	0	0	1	7	19.50797	140.62	189.1	0	3	-3	1046.868
Cystathionine	C7H14N2O4S	NC(CSCCC([NH3+])C([O-])=O)C([O-])=O	HILIC	neg	msi 1	-3.637	7	10	13	27	0	0	2	5	29.29631	159.22	221.2555	2	2	-1	825.6153
Glucaric acid	C6H10O8	OC(C(O)C(O)C([O-])=O)C(O)C([O-])=O	HILIC	neg	msi 1	-4.544	5	8	13	22	0	0	4	8	22.31034	161.18	208.1232	0	2	-2	685.2772
Gluconic acid	C6H12O7	OC[C@@H](O)[C@@H](O)[C@H](O)[C@@H](O)C([O-])=O	HILIC	neg	msi 1	-4.22	5	8	12	24	0	0	5	7	23.50872	141.28	195.1476	0	1	-1	428.7511
Glucose 6-phosphate	C6H13O9P	OC1OC(COP([O-])([O-])=O)C(O)C(O)C1O	HILIC	neg	msi 1	-3.355	3	9	16	27	0	0	4	9	28.74272	172.38	258.1202	0	2	-2	955.2191
Glutamic acid	C5H9NO4	[NH3+]C(CCC([O-])=O)C([O-])=O	HILIC	neg	msi 1	-3.777	4	7	9	18	0	0	1	4	18.44234	107.9	146.1215	1	2	-1	615.7618
Glycerol-3-phosphate	C3H9O6P	OC[C@@H](O)COP([O-])([O-])=O	HILIC	neg	msi 1	-3.109	4	7	9	17	0	0	2	6	18.38955	122.69	170.058	0	2	-2	690.2592
Guanosine	C10H13N5O5	NC1=NC2=C(N=CN2[C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)N1	HILIC	neg	msi 1	-2.272	2	10	22	33	0	0	5	10	35.77831	153	283.2411	0	0	0	223.7971
Inosine	C10H12N4O5	OC[C@H]1O[C@H]([C@H](O)[C@@H]1O)N1C=NC2=C1[N-]C=NC2=O	HILIC	neg	msi 1	-1.178	2	10	21	30	0	0	3	9	33.34472	114.95	267.2185	0	1	-1	197.9324
Isocitric acid	C6H8O7	OC(CC(C([O-])=O)C([O-])=O)C([O-])=O	HILIC	neg	msi 1	-3.507	5	7	12	18	0	0	1	7	19.50797	140.62	189.1	0	3	-3	1133.86
Methionine	C5H11NO2S	CSCCC([NH3+])C([O-])=O	HILIC	neg	msi 1	-1.8	4	7	8	20	0	0	1	2	21.73872	93.07	149.2126	1	1	0	220.4813
N-Acetylserine	C5H9NO4	CC(=O)N[C@@H](CO)C(C[O-])=O	HILIC	neg	msi 1	-4.161	4	6	9	18	0	0	2	5	18.44234	89.46	146.1215	0	1	-1	288.4425
Pantothenate	C9H17NO5	CC(C)(CO)C(O)C(=O)NCCC([O-])=O	HILIC	neg	msi 1	-1.824	7	10	14	31	0	0	3	6	31.61869	109.69	218.2274	0	1	-1	247.8018
Phenylalanine	C9H11NO2	[NH3+]C(CC1=CC=CC=C1)C([O-])=O	HILIC	neg	msi 1	-1.522	3	8	12	23	0	0	1	2	25.87872	67.77	165.1895	1	1	0	195.9783
Serine	C3H7NO3	NC(CO)C([O-])=O	HILIC	neg	msi 1	-4.437	2	5	6	13	0	0	2	4	12.78676	86.38	104.0848	1	1	-1	436.3603
Tryptophyl-Glutamate	C16H19N3O5	N[C@@H](Cc1c[nH]c2ccccc12)C(=O)N[C@@H](CCC([O-])=O)C([O-])=O	HILIC	neg	msi 2	-3.06	9	14	25	41	10	9	3	7	46.80548	151.17	331.3239	1	2	-2	200.0415
Tyrosine	C9H11NO3	[NH3+][C@@H](CC1=CC=C(O)C=C1)C([O-])=O	HILIC	neg	msi 1	-2.658	3	9	13	24	0	0	2	3	26.68072	88	181.1889	1	1	0	286.0539
UMP	C9H13N2O9P	OC1C(O)C(OC1COP([O-])([O-])=O)N1C=CC(=O)[N-]C1=O	HILIC	neg	msi 1	-3.601	4	12	22	31	0	0	2	11	35.55593	169.3	321.1578	0	3	-3	605.8178
Urate	C5H4N4O3	O=C1NC2=C([N-]1)C(=O)[N-]C(=O)N2	HILIC	neg	msi 1	-1.272	0	7	13	14	0	0	2	7	16.93959	75.27	166.0946	0	2	-2	277.53
Xanthine	C5H4N4O2	O=C1NC(=O)C2=C([N-]1)N=CN2	HILIC	neg	msi 1	-0.48	0	6	12	14	0	0	2	6	16.80438	70.56	151.1031	0	1	-1	142.7949
2-Phosphoglycerate	C3H7O7P	OCC(OP(O)(O)=O)C(O)=O	CE	neg	msi 1	-2.952	4	6	10	17	0	0	3	7	18.52476	136.93	185.0494	0	3	-1	-987.026
Allo-threonine	C4H9NO3	C[C@H](O)[C@H]([NH3+])C(O)=O	CE	neg	msi 1	-0.796	2	5	7	18	0	0	3	3	17.21393	85.17	120.1273	1	1	1	558.1103
beta-NAD	C21H27N7O14P2	NC(=O)C1=C[N+](=CC=C1)C1OC(COP([O-])(=O)OP(O)(=O)OCC2OC(C(O)C2O)N2C=NC3=C(N)[NH+]=CN=C23)C(O)C1O	CE	neg	msi 2	-5.82	11	23	48	72	0	0	8	19	81.8182	338.17	664.4339	2	2	1	-31.5151
cyclic AMP	C10H12N5O6P	Nc1[nH+]cnc2n(cnc12)[C@@H]1O[C@@H]2COP([O-])(=O)O[C@H]2[C@H]1O	CE	neg	msi 1	-1.448	1	11	25	34	10	9	3	9	39.54352	168.73	329.2064	1	1	0	-15.968
Cystathionine	C7H14N2O4S	[NH3+][C@@H](CCSC[C@H]([NH3+])C([O-])=O)C([O-])=O	CE	neg	msi 2	-0.937	7	10	13	28	0	0	2	4	29.9631	160.84	222.2634	2	2	0	637.3009
Dihydroorotate	C5H6N2O4	OC(=O)C1CC(=O)NC(=O)N1	CE	neg	msi 1	-1.675	1	6	11	17	0	0	3	6	18.20876	95.5	158.1124	0	1	0	-356.578
Glucosamine 6-sulfate	C6H13NO8S	[NH3+][C@H]1[C@H](O)OC[C@H](COS([O-])(=O)=O)[C@@H](O)[C@@H]1O	CE	neg	msi 2	-3.385	3	9	16	29	0	0	4	8	29.64431	172.37	259.2357	1	1	0	-40.9978
Glutamate	C5H9NO4	[NH3+][C@@H](CCC(O)=O)C(O)=O	CE	neg	msi 1	-0.649	4	7	9	20	0	0	3	4	19.77593	102.24	148.1374	1	2	1	512.6116
Glutamine	C5H10N2O3	NC(=O)CC[C@H]([NH3+])C(O)=O	CE	neg	msi 2	-1.377	4	7	9	21	0	0	3	4	20.74072	108.03	147.1526	1	1	1	575.5214
Glutathione	C10H17N3O6S	[NH3+][C@@H](CCC(=O)N[C@@H](CS[C@@H](CC(=O)NCC(O)=O)C([O-])=O	CE	neg	msi 2	-2.011	11	13	19	37	0	0	4	8	39.94748	202.07	307.325	1	2	0	346.2862
Guanosine	C10H13N5O5	NC1=NC2=C(N=CN2[C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)N1	CE	neg	msi 1	-2.272	2	10	22	33	0	0	5	10	35.77831	153	283.2411	0	0	0	357.2154
Hypoxanthine	C5H4N4O	O=c1[nH]cnc2nc[nH]c12	CE	neg	msi 2	0.227	0	5	11	14	10	9	2	3	16.66917	74.43	136.1117	0	0	0	425.4546
Inosine	C10H12N4O5	OC[C@H]1O[C@H]([C@H](O)[C@@H]1O)N1C=[NH+]C2=C1NC=NC2=O	CE	neg	msi 1	0.237	2	10	21	32	0	0	5	8	34.67831	128.59	269.2344	1	0	0	54.94137
Lactose	C12H22O11	OC[C@H]1O[C@@H](O[C@H]2[C@H](CO)[C@@H](O[C@@H](O)[C@@H]2CO)[C@H](O)[C@@H](O)[C@H]1O	CE	neg	msi 1	-3.663	4	11	24	45	0	0	8	11	44.61145	189.53	342.297	0	0	0	-22.5571
N-Formylglycine	C3H5NO3	OC(=O)CNC=O	CE	neg	msi 1	-1.144	3	6	6	12	0	0	2	4	12.11997	66.4	103.0768	0	1	0	-234.038
Paraxanthine	C7H8N4O2	CN1C=NC2=C1C(=O)N(C)C(=O)N2	CE	neg	msi 2	-0.637	0	7	14	21	0	0	1	6	23.65834	65.01	180.1643	0	0	0	22.33727
Quinate	C7H12O6	O[C@@H]1C[C@@](O)(C[C@@H](O)[C@H]1O)C(O)=O	CE	neg	msi 2	-1.979	1	7	13	25	0	0	5	6	25.13352	118.22	192.1669	0	1	0	-33.8124
Succinate	C4H6O4	OC(=O)CCC(O)=O	CE	neg	msi 1	-0.71	3	6	7	14	0	0	2	4	14.24876	74.6	118.0882	0	2	0	-15.968
Trans-aconitate	C6H6O6	OC(=O)C\C(=C/C(O)=O)C(O)=O	CE	neg	msi 1	-1.034	4	7	11	18	0	0	3	6	19.37276	111.9	174.1085	0	3	0	-329.223
Tyrosine	C9H11NO3	[NH3+][C@@H](CC1=CC=C(O)C=C1)C([O-])=O	CE	neg	msi 2	-0.492	3	9	13	24	0	0	2	3	26.68072	88	181.1889	1	1	0	555.8521
Urate	C5H4N4O3	O=C1NC2=C(NC(=O)N1)C(=O)N2	CE	neg	msi 1	-1.191	0	7	13	16	0	0	4	7	18.27317	99.33	168.1105	0	0	0	-22.2508
1-Methyltryptophan	C12H14N2O2	Cn1cc(CC([NH3+])C(O)=O)c2ccccc12	HILIC	pos	msi 2	-3.637	7	10	13	27	0	0	2	5	29.29631	159.22	221.2555	2	2	1	153.2794
4-Acetamidobutanoate	C6H11NO3	CC(=O)NCCCC(O)=O	HILIC	pos	msi 1	-4.544	5	8	13	22	0	0	4	8	22.31034	161.18	208.1232	0	2	0	289.9254
Acetyl-carnitine	C9H17NO4	CC(=O)O[C@H](CC(O)=O)C[N+](C)(C)C	HILIC	pos	msi 1	-4.22	5	8	12	24	0	0	5	7	23.50872	141.28	195.1476	0	1	1	575.8082
Adenine	C5H5N5	NC1=C2NC=NC2=NC=[NH+]1	HILIC	pos	msi 2	-3.355	3	9	16	27	0	0	4	9	28.74272	172.38	258.1202	0	2	1	230.7173
Adenosine	C10H13N5O4	NC1=C2N=CN[C@@H]3O[C@H](CO)[C@@H](O)[C@H]3O)C2=NC=[NH+]1	HILIC	pos	msi 1	-3.777	4	7	9	18	0	0	1	4	18.44234	107.9	146.1215	1	2	1	198.5469
Adipoyl-carnitine	C13H23NO6	C[N+](C)(C)CC(CC(O)=O)OC(=O)CCCCC(O)=O	HILIC	pos	msi 1	-3.109	4	7	9	17	0	0	2	6	18.38955	122.69	170.058	0	2	1	643.8932
Alanine	C3H7NO2	C[C@@H]([NH3+])C(O)=O	HILIC	pos	msi 1	-2.272	2	10	22	33	0	0	5	10	35.77831	153	283.2411	0	0	1	602.4718
Allo-threonine	C4H9NO3	C[C@H](O)[C@H]([NH3+])C(O)=O	HILIC	pos	msi 1	-1.178	2	10	21	30	0	0	3	9	33.34472	114.95	267.2185	0	1	1	605.1535
Aminoadipate	C6H11NO4	[NH3+]C(CCCC(O)=O)C(O)=O	HILIC	pos	msi 1	-3.507	5	7	12	18	0	0	1	7	19.50797	140.62	189.1	0	3	1	689.1081
AMP	C10H14N5O7P	Nc1[nH+]cnc2n(cnc12)[C@@H]1O[C@H](COP(O)([O-])=O)[C@@H](O)[C@H]1O	HILIC	pos	msi 1	-1.8	4	7	8	20	0	0	1	2	21.73872	93.07	149.2126	1	1	0	650.2498
Arginine	C6H14N4O2	NC(N)=[NH+]CCC[C@H]([NH3+])C(O)=O	HILIC	pos	msi 1	-4.161	4	6	9	18	0	0	2	5	18.44234	89.46	146.1215	0	1	2	766.0688
Betaine	C5H11NO2	C[N+](C)(C)CC(O)=O	HILIC	pos	msi 1	-1.824	7	10	14	31	0	0	3	6	31.61869	109.69	218.2274	0	1	1	536.9307
Cadaverine	C5H14N2	[NH3+]CCCCC[NH3+]	HILIC	pos	msi 1	-1.522	3	8	12	23	0	0	1	2	25.87872	67.77	165.1895	1	1	2	813.7643
Cartinine	C7H15NO3	C[N+](C)(C)C[C@H](O)CC(O)=O	HILIC	pos	msi 1	-4.437	2	5	6	13	0	0	2	4	12.78676	86.38	104.0848	1	1	1	625.4292
Citrulline	C6H13N3O3	NC(=O)NCCC[C@H]([NH3+])C(O)=O	HILIC	pos	msi 1	-3.06	9	14	25	41	10	9	3	7	46.80548	151.17	331.3239	1	2	1	668.3863
cyclic AMP	C10H12N5O6P	Nc1[nH+]cnc2n(cnc12)[C@@H]1O[C@@H]2COP([O-])(=O)O[C@H]2[C@H]1O	HILIC	pos	msi 1	-2.658	3	9	13	24	0	0	2	3	26.68072	88	181.1889	1	1	0	490.9989
Cystathionine	C7H14N2O4S	[NH3+][C@@H](CCSC[C@H]([NH3+])C([O-])=O)C([O-])=O	HILIC	pos	msi 1	-3.601	4	12	22	31	0	0	2	11	35.55593	169.3	321.1578	0	3	0	811.758
dCMP	C9H14N3O7P	NC1=[NH+]C(=O)N(C=C1)C1CC(O)C(COP(O)([O-])=O)O1	HILIC	pos	msi 1	-1.272	0	7	13	14	0	0	2	7	16.93959	75.27	166.0946	0	2	2	718.1268
Decanoyl-carnitine	C17H33NO4	CCCCCCCCCC(=O)OC(CC(O)=O)C[N+](C)(C)C	HILIC	pos	msi 1	-0.48	0	6	12	14	0	0	2	6	16.80438	70.56	151.1031	0	1	1	244.953
Deoxyadenosine	C10H13N5O3	NC1=C2N=CN[C@H]3C[C@H](O)[C@@H](CO)O3)C2=NC=[NH+]1	HILIC	pos	msi 1	-1.325	2	10	20	32	0	0	4	7	34.8411	117.64	252.2303	1	0	1	170.9835
Ethanolamine	C2H7NO	[NH3+]CCO	HILIC	pos	msi 1	-0.741	1	4	3	12	0	0	2	1	10.75634	47.87	62.09111	1	0	1	788.6246
Glutamate	C5H9NO4	[NH3+][C@@H](CCC(O)=O)C(O)=O	HILIC	pos	msi 1	-0.649	4	7	9	20	0	0	3	4	19.77593	102.24	148.1374	1	2	1	705.4748
Glycocholic acid	C26H43NO6	C[C@H](CCC(=O)NCC(O)=O)[C@H]1CC[C@H]2[C@@H]3[C@H](O)C[C@@H]4C[C@H](O)CC[C@]4(C)[C@H]3C[C@H](O)[C@]12C	HILIC	pos	msi 1	3.265	7	18	36	76	0	0	5	7	80.3441	127.09	465.6237	0	1	0	442.347
Guanidinosuccinate	C5H9N3O4	NC(=[NH2+])N[C@@H](CC(O)=O)C(O)=O	HILIC	pos	msi 1	-3.865	5	7	11	22	0	0	5	6	21.97593	138.24	176.1508	4	2	1	718.395
Guanine	C5H5N5O	NC1=[NH+]C(=O)C2=C(N1)N=CN2	HILIC	pos	msi 1	0.057	0	6	12	17	0	0	4	5	19.10276	93.48	152.1342	1	0	1	297.817
Guanosine	C10H13N5O5	NC1=NC2=C(N=CN2[C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)N1	HILIC	pos	msi 1	-2.272	2	10	22	33	0	0	5	10	35.77831	153	283.2411	0	0	0	407.0109
Histidine	C6H9N3O2	[NH3+][C@@H](CC1=CNC=[NH+]1)C([O-])=O	HILIC	pos	msi 2	-0.455	3	7	11	21	0	0	3	3	22.13193	93.77	156.1627	2	1	1	515.4691
Inosine	C10H12N4O5	OC[C@H]1O[C@H]([C@H](O)[C@@H]1O)N1C=[NH+]C2=C1NC=NC2=O	HILIC	pos	msi 1	0.237	2	10	21	32	0	0	5	8	34.67831	128.59	269.2344	1	0	1	327.3499
Isobutyryl-carnitine	C11H22NO4	[H][C@@](CC(O)=O)(C[N+](C)(C)C)OC(=O)C(C)C	HILIC	pos	msi 1	0.026	7	8	15	38	0	0	1	4	38.33745	63.6	232.2971	2	1	1	458.5116
Isoleucine	C6H13NO2	C[C@H](C)[C@H]([NH3+])C(O)=O	HILIC	pos	msi 1	1.1	3	6	8	23	0	0	2	2	22.5991	64.94	132.1811	1	1	1	482.866
Kynurenine	C10H12N2O3	[NH3+][C@@H](CC(=O)C1=CC=CC=C1[NH3+])C([O-])=O	HILIC	pos	msi 1	-0.749	4	9	15	28	0	0	2	3	30.87431	112.48	209.2222	2	1	1	436.9473

Table B.3 Metabolites detected in *C. elegans* using CE-MS and HILIC-MS in positive and negative mode and their physicochemical properties.

Name	Formula	SMILES	method	ion	msi	XLogP	nRotB	nAtomLC	nB	nAtom	nAromBond	naAromAtom	nHBDon	nHBAcc	apol	TopoPSA	MW	nBase	nAcid	Charge	mobility / RT
Leucine	C6H13NO2	CC(C)CC(N)C(=O)=O	HILIC	pos	msi 2	1.311	3	6	8	23	0	0	2	2	22.5991	64.94	132.1811	1	1	1	462.3339
Lysine	C6H14N2O2	[NH3+]CCCCC[NH3+])C(O)=O	HILIC	pos	msi 2	0.285	5	8	9	26	0	0	3	2	25.03269	92.58	148.2037	2	1	2	788.5925
Methionine	C5H11NO2S	CSCC[C@H]([NH3+])C(O)=O	HILIC	pos	msi 1	0.847	4	7	8	21	0	0	2	2	22.40552	90.24	150.2206	1	1	1	492.2148
Methylglutaryl-carnitine	C13H23NO6	CC(C(O)=O)CC(=O)OC(CC(O)=O)C[N+](C)(C)C	HILIC	pos	msi 1	-0.284	10	11	19	44	0	0	2	6	44.79503	100.9	290.3333	2	2	1	643.8932
N-Acetylasparagine	C6H10N2O4	CC(=O)NC(=O)CC([NH3+])C([O-])=O	HILIC	pos	msi 1	-1.248	5	8	11	22	0	0	2	5	22.63593	113.94	174.1548	1	1	0	625.7132
N-Acetylgalactosamine	C8H15NO6	CC(=O)NC1C(O)OC(CO)C(O)C1O	HILIC	pos	msi 1	-1.488	3	9	15	30	0	0	5	7	29.9939	119.25	221.2081	0	0	0	433.7705
N-Acetylspermine	C12H28N4O	CC(=O)NCCC[NH2+]CCCC[NH2+]CCC[NH3+]	HILIC	pos	msi 2	0.352	13	16	16	48	0	0	4	2	46.99258	89.96	247.4012	5	0	3	910.2052
Ornithine	C5H12N2O2	[NH3+]CCCC([NH3+])C([O-])=O	HILIC	pos	msi 1	-0.554	4	7	8	22	0	0	2	2	21.27231	95.41	133.1691	2	1	1	791.1631
Pantothenate	C9H17NO5	CC(C)(CO)C(O)C(=O)NCCC(O)=O	HILIC	pos	msi 1	-1.343	7	10	14	32	0	0	4	6	32.28548	106.86	219.2353	0	1	0	425.8365
Phenylalanine	C9H11NO2	[NH3+][C@@H](CC1=CC=CC=C1)C(O)=O	HILIC	pos	msi 1	1.125	3	8	12	24	0	0	2	2	26.54552	64.94	166.1974	1	1	1	438.0012
Proline	C5H9NO2	[O-]C(=O)[C@@H]1CCC[NH2+]1	HILIC	pos	msi 1	-0.554	1	5	8	17	0	0	1	2	17.50514	56.74	115.1307	2	1	0	791.0877
Propanoylcarnitine	C10H19NO4	CCC(=O)O[C@H](CC(O)=O)C[N+](C)(C)C	HILIC	pos	msi 1	-0.267	7	8	14	35	0	0	1	4	35.24386	63.6	218.2705	2	1	1	500.103
Pyridoxamine	C8H12N2O2	Cc1[nH+]cc(CO)c(C[NH3+])c1O	HILIC	pos	msi 2	1.42	2	7	12	26	6	6	4	1	27.2191	82.24	170.2093	2	0	2	622.117
Quinine	C20H24N2O2	COc1ccc2[nH+]ccc([C@@H](O)[C@@H]3C[C@@H]4CC[N@H+]3C[C@@H]4C=C)c2c1	HILIC	pos	msi 2	3.483	4	13	27	50	11	10	3	1	56.34062	48.04	326.4334	4	0	2	164.3223
S-Adenosyl-methionine	C15H22N6O5S	C[S+](CC[C@H]([NH3+])C([O-])=O)C[C@H]1O[C@H](n2cnc3c(N)ncnc23)[C@@H](O)[C@H]1O	HILIC	pos	msi 2	-3.448	7	15	29	51	0	0	5	9	55.91303	185.41	400.4549	3	1	2	791.7178
Spermidine	C7H19N3	[NH3+]CCCC[NH2+]CCC[NH3+]	HILIC	pos	msi 1	0.526	7	10	9	32	0	0	3	0	30.28945	71.89	148.27	4	0	3	972.7641
Taurumuricholate	C26H45NO7S	C1[C@@]2([C@]3(CC[C@]4(C[C@@H](O)CC[C@]4([C@H](C)CCC(=O)NCCS([O-])(=O)=O)[C@H]4[C@@H](O)C[C@H]5[C@]6(C)CC[C@H](O)C[C@H]6CC[C@H]5[C@@]4(C)CC3)[H])[H])CC1	HILIC	pos	msi 1	4.044	8	19	38	79	0	0	4	8	84.71289	155.37	514.6971	0	1	-1	321.0611
Tauroursodeoxycholic acid	C26H45NO6S	C1[C@@]2([C@]3(CC[C@]4(C[C@@H](O)CC[C@]4([C@H](C)CCC(=O)NCCS(=O)(=O)O)[C@H]4[C@@H](O)C[C@H]5[C@]6(C)CC[C@H](O)C[C@H]6CC[C@H]5[C@@]4(C)CC3)[H])[H])CC1	HILIC	pos	msi 1	4.856	8	19	37	79	0	0	4	7	84.57769	132.31	499.7057	0	1	0	230.043
Tetradecenoyl-carnitine	C21H39NO4	CCCCCCCCCCCC=CC(=O)OC(CC(=O)[O-])C[N+](C)(C)C	HILIC	pos	msi 1	5.881	17	19	25	66	0	0	1	4	67.93972	63.6	370.5474	2	1	1	254.097
Tyrosine	C9H11NO3	[NH3+][C@@H](CC1=CC=C(O)C=C1)C([O-])=O	HILIC	pos	msi 2	-0.492	3	9	13	24	0	0	2	3	26.68072	88	181.1889	1	1	0	509.9058
UMP	C9H13N2O9P	OC[C@H]1O[C@H]([C@H](O)[C@@H]1OP(O)([O-])=O)N1C=CC(=O)NC1=O	HILIC	pos	msi 1	-4.204	4	11	22	33	0	0	4	11	36.88952	178.5	323.1737	0	2	-1	625.324
Uracil	C4H4N2O2	O=C1NC=CC(=O)N1	HILIC	pos	msi 1	-0.875	0	5	8	12	0	0	2	4	13.51117	58.2	112.0869	0	0	0	132.4679
Valeryl-carnitine	C12H23NO4	CCCCC(=O)O[C@@H](CC(O)=O)C[N+](C)(C)C	HILIC	pos	msi 1	0.871	9	10	16	41	0	0	1	4	41.43103	63.6	246.3237	2	1	1	415.1629
Valine	C5H11NO2	CC(C)[C@H]([NH3+])C(O)=O	HILIC	pos	msi 2	0.531	2	5	7	20	0	0	2	2	19.50552	64.94	118.1545	1	1	1	875.1558
Xanthine	C5H4N4O2	O=C1NC2=C(NC=N2)C(=O)N1	HILIC	pos	msi 1	-0.654	0	6	12	15	0	0	3	3	17.47117	82.59	152.1111	0	0	0	187.35
2-aminopentanoic acid	C5H11NO2	CCC(C[NH3+])C(O)=O	CE	pos	msi 1	0.803	3	6	7	20	0	0	2	2	19.50552	64.94	118.1545	1	1	1	743.005
4-Hydroxyphenylglycine	C8H9NO3	[NH3+][C@H](C([O-])=O)c1ccc(O)cc1	CE	pos	msi 2	0.038	2	8	12	21	6	6	2	2	23.58714	88	167.1623	1	1	0	523.8051
Adenine	C5H5N5	NC1=C2NC=NC2=NC=[NH+]1	CE	pos	msi 2	-0.846	0	5	11	16	0	0	3	4	18.30076	76.74	136.1348	1	0	1	1081.622
Adenosine	C10H13N5O4	NC1=C2N=CN([C@@H]3O[C@H](CO)[C@@H](O)[C@H]3O)C2=NC=[NH+]1	CE	pos	msi 1	-1.926	2	10	21	33	0	0	5	8	35.6431	137.87	268.2497	1	0	1	888.4997
Alanine	C3H7NO2	C[C@@H]([NH3+])C(O)=O	CE	pos	msi 1	-0.124	1	4	5	14	0	0	2	2	13.31834	64.94	90.10125	1	1	1	838.3743
Allo-threonine	C4H9NO3	C[C@H](O)[C@H]([NH3+])C(O)=O	CE	pos	msi 2	-0.796	2	5	7	18	0	0	3	3	17.21393	85.17	120.1273	1	1	1	620.6585
Aminoadipate	C6H11NO4	[NH3+]C(CCCC(O)=O)C(O)=O	CE	pos	msi 2	-0.291	5	8	10	23	0	0	3	4	22.86952	102.24	162.164	1	2	1	634.7875
AMP	C10H14N5O7P	Nc1[nH+]cnc2n(cnc12)[C@@H]1O[C@H](COP(O)([O-])=O)[C@@H](O)[C@H]1O	CE	pos	msi 1	-2.037	4	12	25	37	10	9	5	10	41.6791	199.96	347.2216	1	2	0	-16.4056
Arginine	C6H14N4O2	NC(N)=[NH+]CCC[C@H]([NH3+])C(O)=O	CE	pos	msi 2	-3.17	5	9	11	28	0	0	5	4	27.23269	130.95	176.2171	5	1	2	1074.636
Aspartate	C4H7NO4	[NH3+][C@@H](CC(=O)=O)C([O-])=O	CE	pos	msi 1	-1.488	3	6	8	16	0	0	2	4	16.01555	105.07	133.1029	1	2	0	-18.3967
Betaine	C5H11NO2	C[N+](C)(C)CC(O)=O	CE	pos	msi 1	-0.746	2	5	7	20	0	0	1	2	19.50552	37.3	118.1545	2	1	1	422.5898
Citrulline	C6H13N3O3	NC(=O)NCCC[C@H]([NH3+])C(O)=O	CE	pos	msi 2	-1.209	6	9	11	26	0	0	4	5	25.6011	120.06	176.1999	1	1	1	606.9257
Creatine	C4H9N3O2	CN(CC(O)=O)C(N)=[NH2+]	CE	pos	msi 1	-2.945	3	6	8	19	0	0	3	4	18.61193	92.15	132.1413	4	1	1	916.2468
Cystathionine	C7H14N2O4S	[NH3+][C@@H](CCSC[C@H]([NH3+])C([O-])=O)C([O-])=O	CE	pos	msi 1	-0.937	7	10	13	28	0	0	2	4	29.9631	160.84	222.2634	2	2	0	546.7837
Glucosamine 6-sulfate	C6H13NO8S	[NH3+][C@H]1[C@H](O)O[C@H](COS([O-])(=O)=O)[C@@H](O)[C@@H]1O	CE	pos	msi 1	-3.385	3	9	16	29	0	0	4	8	29.64431	172.37	259.2357	1	1	0	6.349645
Glutamate	C5H9NO4	[NH3+][C@@H](CCC(O)=O)C(O)=O	CE	pos	msi 1	-0.649	4	7	9	20	0	0	3	4	19.77593	102.24	148.1374	1	2	1	575.5214
Glutamine	C5H10N2O3	NC(=O)CC[C@H]([NH3+])C(O)=O	CE	pos	msi 2	-1.377	4	7	9	21	0	0	3	4	20.74072	108.03	147.1526	1	1	1	974.4524
glycine betaine	C5H11NO2	C[N+](C)(C)CC(O)=O	CE	pos	msi 2	-0.746	2	5	7	20	0	0	1	2	19.50552	37.3	118.1545	2	1	1	422.5898
Glycyl-proline	C7H12N2O3	[NH3+]CC(=O)N1CCCC1C(O)=O	CE	pos	msi 2	-0.727	3	7	12	25	0	0	2	4	25.59431	85.25	173.19	1	1	1	952.2581
Guanidinosuccinate	C5H9N3O4	NC(=[NH2+])N[C@@H](CC(=O)=O)C(O)=O	CE	pos	msi 1	-3.865	5	7	11	22	0	0	5	6	21.97593	138.24	176.1508	4	2	1	724.3186
Guanine	C5H5N5O	NC1=[NH+]C(=O)C2=C(N1)N=CN2	CE	pos	msi 2	0.057	0	6	12	17	0	0	4	5	19.10276	93.48	152.1342	1	0	1	992.1556
Guanosine	C10H13N5O5	NC1=NC2=C(N=CN2[C@@H]2O[C@H](CO)[C@@H](O)[C@H]2O)C(=O)N1	CE	pos	msi 1	-2.272	2	10	22	33	0	0	5	10	35.77831	153	283.2411	0	0	0	421.3335
Hypoxanthine	C5H4N4O	O=c1[nH]cnc2nc[nH]c12	CE	pos	msi 1	0.227	0	5	11	14	10	9	2	3	16.66917	74.43	136.1117	0	0	0	64.54549
Inosine	C10H12N4O5	OC[C@H]1O[C@H]([C@H](O)[C@@H]1O)N1C=[NH+]C2=C1NC=NC2=O	CE	pos	msi 1	0.237	2	10	21	32	0	0	5	8	34.67831	128.59	269.2344	1	0	0	-22.5571
Isobutyryl-carnitine	C11H22NO4	[H][C@@](CC(O)=O)(C[N+](C)(C)C)OC(=O)C(C)C	CE	pos	msi 2	0.026	7	8	15	38	0	0	1	4	38.33745	63.6	232.2971	2	1	1	1091.883
Isoleucine	C6H13NO2	CC[C@H](C)[C@H]([NH3+])C(O)=O	CE	pos	msi 1	1.1	3	6	8	23	0	0	2	2	22.5991	64.94	132.1811	1	1	1	724.1674
Kynurenine	C10H12N2O3	[NH3+][C@@H](CC(=O)C1=CC=CC=C1[NH3+])C([O-])=O	CE	pos	msi 1	-0.749	4	9	15	28	0	0	2	3	30.87431	112.48	209.2222	2	1	1	680.1713
Methionine	C5H11NO2S	CSCC[C@H]([NH3+])C(O)=O	CE	pos	msi 1	0.847	4	7	8	21	0	0	2	2	22.40552	90.24	150.2206	1	1	1	643.6581
N2-Acetyl-lysine	C8H16N2O3	CC(=O)N[C@@H](CCCC[NH3+])C(O)=O	CE	pos	msi 2	0.027	7	9	12	30	0	0	3	4	30.02148	94.04	189.2325	1	1	1	906.8514
N-Acetylgalactosamine	C8H15NO6	CC(=O)NC1C(O)OC(CO)C(O)C1O	CE	pos	msi 1	-1.488	3	9	15	30	0	0	5	7	29.9939	119.25	221.2081	0	0	0	-15.2304
N-Acetylhistidine	C8H11N3O3	CC(=O)NC(Cc1c[NH+]=cN1)C(O)=O	CE	pos	msi 2	-0.546	5	8	14	26	0	0	4	5	27.78752	92.4	198.1995	1	1	1	861.9443
N-acetylmannosamine	C8H15NO6	[H][C@@]1(O)O[C@H](CO)[C@@H](O)[C@@H](O)[C@H]1NC(C)=O	CE	pos	msi 1	-1.488	3	9	15	30	0	0	5	7	29.9939	119.25	221.2081	0	0	0	-15.2304
N-acetylserine	C5H9NO4	CC(=O)N[C@@H](CO)C(O)=O	CE	pos	msi 1	-1.514	4	6	9	19	0	0	3	5	19.10914	86.63	147.1295	0	1	0	-26.97
Niacin	C6H5NO2	OC(=O)C1=CN=CC=C1	CE	pos	msi 2	-0.508	1	6	9	14	0	0	1	3	16.57997	49.66	123.1096	0	1	0	647.153
Nicotinamide	C6H6N2O	NC(=O)C1=C[NH+]=CC=C1	CE	pos	msi 2	-0.366	1	6	9	16	0	0	2	2	18.22955	57.06	123.1328	1	0	1	1971.189
N-Methylaspartate	C5H9NO4	C[NH2+][C@H](CC(O)=O)C([O-])=O	CE	pos	msi 1	-1.52	4	6	9	19	0	0	2	4	19.10914	94.04	147.1295	2	2	0	366.6015
N-methyl-glutamate	C6H11NO4	C[NH2+]C(CCC(O)=O)C(O)=O	CE	pos	msi 2	-0.681	5	7	10	23	0	0	3	4	22.86952	91.21	162.164	2	2	1	512.618
Noradrenaline	C8H11NO3	[NH3+]C(O)c1ccc(O)c(O)c1	CE	pos	msi 2	0.421	2	8	12	24	6	6	4	1	25.58752	88.33	170.1861	1	0	1	1094.079
Norleucine	C6H13NO2	CCCC[C@H]([NH3+])C(O)=O	CE	pos	msi 1	1.372	4	7	8	23	0	0	2	2	22.5991	64.94	132.1811	1	1	1	725.4029
O-Acetylserine	C5H9NO4	CC(=O)O[C@@H](CN)C([O-])=O	CE	pos	msi 1	-0.997	4	7	9	19	0	0	1	4	19.10914	94.07	147.1295	1	1	0	502.7501
Ophthalmic acid	C11H19N3O6	CC[C@H](NC(=O)CC[C@H]([NH3+])C([O-])=O)C(=O)NCC(O)=O	CE	pos	msi 2	-1.421	11	13	19	40	0	0	5	8	40.80786	160.44	290.2954	1	2	1	433.5451
Oxoproline	C5H7NO3	[O-]C(=O)C1CC(=O)C[NH2+]1	CE	pos	msi 1	-1.793	1	6	9	16	0	0	1	3	16.97355	73.81	129.1142	2	1	0	-27.2857
Paracetamol	C8H9NO2	CC(=O)Nc1ccc(O)cc1	CE	pos	msi 2	0.87	2	8	11	20	6	6	2	2	22.78514	49.33	151.1629	0	0	0	-0.20533
Phenylalanine	C9H11NO2	[NH3+][C@@H](CC1=CC=CC=C1)C(O)=O	CE	pos	msi 1	1.125	3	8	12	24	0	0	2	2	26.54552	64.94	166.1974	1	1	1	501.6971
Pipecolic acid	C6H11NO2	OC(=O)[C@@H]1CCCC[NH2+]1	CE	pos	msi 2	0.285	1	6	9	21	0	0	2	2	21.26552	53.91	130.1652	2	1	1	1094.205
Proline	C5H9NO2	[O-]C(=O)[C@@H]1CCC[NH2+]1	CE	pos	msi 1	-0.554	1	5	8	17	0	0	1	2	17.50514	56.74	115.1307	2	1	0	476.2338
Propanoylcarnitine	C10H19NO4	CCC(=O)O[C@H](CC(O)=O)C[N+](C)(C)C	CE	pos	msi 1	-0.267	7	8	14	35	0	0	1	4	35.24386	63.6	218.2705	2	1	1	947.0492
Purine	C5H4N4	N1C=NC2=NC=[NH+]C=C12	CE	pos	msi 1	-0.512	0	5	10	14	0	0	2	3	16.53397	50.72	121.1202	1	0	1	865.8998
Serine	C3H7NO3	[NH3+][C@@H](CO)C(O)=O	CE	pos	msi 1	-1.256	2	5	6	15	0	0	3	3	14.12034	85.17	106.1007	1	1	1	695.9802
Spermidine	C7H19N3	[NH3+]CCCC[NH2+]CCC[NH3+]	CE	pos	msi 1	0.526	7	10	9	32	0	0	3	0	30.28945	71.89	148.27	4	0	3	2841.987
Threitol	C4H10O4	OC[C@H](O)[C@@H](O)CO	CE	pos	msi 1	-2.552	3	6	7	18	0	0	4	4	16.91593	80.92	122.12	0	0	0	-40.7429
Trans-aconitate	C6H6O6	OC(=O)\C(=C/C(O)=O)C(O)=O	CE	pos	msi 1	-1.034	4	7	11	18	0	0	3	6	19.37276	111.9	174.1085	0	3	0	24.25124
Trigonelline	C7H7NO2	C[N+]1=CC(=CC=C1)C(O)=O	CE	pos	msi 1	0.616	1	6	10	18	0	0	1	2	20.35834	40.31	138.1442	1	1	1	577.0442
Tryptophan	C11H12N2O2	[NH3+][C@H](CC1=CNC2=CC=CC=C12)C(O)=O	CE	pos	msi 1	0.677	3	9	16	28	0	0	3	3	31.83231	76.97	205.2335	1	1	1	642.9119
Tyrosine	C9H11NO3	[NH3+][C@@H](CC1=CC=C(O)C=C1)C([O-])=O	CE	pos	msi 2	-0.492	3	9	13	24	0	0	2	3	26.68072	88	181.1889	1	1	0	593.902
Uracil	C4H4N2O2	O=C1NC=CC(=O)N1	CE	pos	msi 1	-0.875	0	5	8	12	0	0	2	4	13.51117	58.2	112.0869	0	0	0	-8.71451
Urate	C5H4N4O3	O=C1NC2=C(N1)C(=O)NC(=O)N2	CE	pos	msi 1	-1.191	0	7	13	16	0	0	4	7	18.27317	99.33	168.1105	0	0	0	-5.87923
Uridine	C9H12N2O6	OC[C@H]1O[C@H]([C@H](O)[C@H]1O)N1C=CC(=O)NC1=O	CE	pos	msi 1	-2.117	2	10	18	29	0	0	4	8	30.85352	119.33	244.2017	0	0	0	21.7267
Urocanic acid	C6H6N2O2	OC(=O)\C=C\C1=CN=[NH+]1	CE	pos	msi 2	0.614	2	7	10	17	0	0	3	3	19.03155	63.3	139.1322	1	1	1	1580.788
Valine	C5H11NO2	CC(C)[C@H]([NH3+])C(O)=O	CE	pos	msi 2	0.531	2	5	7	20	0	0	2	2	19.50552	64.94	118.1545	1	1	1	970.2733

Table B.3 (Continued) Metabolites detected in *C. elegans* using CE-MS and HILIC-MS in positive and negative mode and their physicochemical properties.

	Substance	[M-H]-	[M+H]+	mobility
negative EOF marker	Ethylsulfate	124.99		-1937
neutral EOF marker	Paracetamol	150.05	152.07	0
positive EOF marker	Procaine		237.16	1239.2

Table B.4 Electroosmotic flow (EOF) markers added to each sample and used to perform effective mobility scale transformation of CE-MS data.

method	ion	mean	n	min	max	sd	RSD %
CE	neg	7051	8	51	40468	13678	194
CE	pos	5305	48	23	78679	12948	244
HILIC	neg	49878	6	5239	211436	81561	164
HILIC	pos	117800	38	82	2548220	417474	354

Table B.5 Summary of intensity values of commonly detected features in HILIC and CE over all *C. elegans* QC samples

C Appendix Chapter 4

C.1 Methods

C.1.1 Overview of the APEX workflow

The automated **A**nnotation **P**ropagation through multiple **EX**perimental networks (APEX) workflow was defined as follows:

```
for each manually annotated seed node S_i (GPNAE / ascarosides):
  get the set of neighbors N(S_i)

    for each neighbor node N_j in N(S_i):
        for each mass difference edge connecting S_i to N_j:
            if m(N_j) < m(S_i) and the corresponding chemical formula (formula(S_i)
            - formula(N_j)) is not in the chemical formula of S_i:
                  remove the edge

            if N_j is not annotated yet:
                  annotate N_j with the chemical formula corresponding to
                  (formula(N_j) +/- formula(S_i)), if available

                  add new attributes to N_j containing the networks connecting N_j
                  and S_i, the seed node S_i, and other available attributes (i.e.
                  values of spectral similarity, mass-difference, and homologous
                  series)
            else:
                  if the number of networks connecting N_j and S_i is greater than
                  the current annotation or if the number is the same but the mass
                  difference |m(N_j) - m(S_i)| is smaller than the existing
                  annotation:
                        replace the existing annotation with the new one by:

                        annotate N_j with the chemical formula corresponding to
                        (formula(N_j) +/- formula(S_i)), if available

                        add new attributes to N_j containing the networks
                        connecting N_j and S_i, the seed node S_i, and other
                        available attributes (i.e. values of spectral similarity,
                        mass-difference, and homologous series)
```

C.2 Tables

Name	Formula difference	mass
Acetylation (+C2H2O)	C2H2O	42.01056469
Amination (+NH)	NH	15.01089904
Amination (+NH2)	NH2	16.01872407
Amination (+NH3)	NH3	17.02654911
C	C	12
C2	C2	24
C2H2	C2H2	26.01565006
C2H4O	C2H4O	44.02621475
C3H6	C3H6	42.04695019
C4H8	C4H8	56.06260026
C5H2	C5H2	62.01565006
Carboxylation (+CO2)	CO2	43.98982924
CH4	CH4	16.03130013
Condensation (+H2O)	H2O	18.01056469
Deamination (+CH2O)	CH2O	30.01056468
Disaccharide (+C12H20O11)	C12H20O11	340.100561
Ethylation (+C2H4)	C2H4	28.03130013
Formylation (+CO)	CO	27.99491462
Hexose (C6H10O5)	C6H10O5	162.0528234
Hydrogenation (+H2)	H2	2.015650064
Hydroxylation (+O)	O	15.99491462
Isoprenylation (+C5H8)	C5H8	68.06260026
Methylation (+CH2)	CH2	14.01565006
O2	O2	31.98982924

Table C.1 (Bio)Chemical transformation list used for generating the mass difference networks for the annotation of GPNAEs

Name	Formula difference	mass
ascr ->hbas (+C7H4O2)	C7H4O2	120.02113
ascr ->icas (+C9H5NO)	C9H5NO	143.037114
ascr ->mbas (+C5H6O)	C5H6O	82.041865
ascr#1	C13H22O5	258.146725
C	C	12
C2H2	C2H2	26.01565006
C2H4O	C2H4O	44.02621475
C3H6	C3H6	42.04695019
C3H6O	C3H6O	58.04186481
C4H8	C4H8	56.06260026
C4H8O	C4H8O	72.05751488
Carboxylation (+CO2)	CO2	43.98982924
Condensation (+H2O)	H2O	18.01056469
Ethylation (+C2H4)	C2H4	28.03130013
Formylation (+CO)	CO	27.99491462
Hexose (C6H10O5)	C6H10O5	162.0528234
Hydrogenation (+H2)	H2	2.015650064
Hydroxylation (+O)	O	15.99491462
Methylation (+CH2)	CH2	14.01565006
Phosphorylation (+HPO3)	HPO3	79.96633041

Table C.2 (Bio)Chemical transformation list used for generating the mass difference networks for the annotation of ascarosides and MOGLs

Number of manual annotations in SNS	Overlap between all SNS	Overlap between SNS 1 + SNS 2	Overlap between SNS 1 + SNS 3	Overlap between SNS 2 + SNS 3
1	1 (20%)	0	0	0
2	1 (33%)	0	0	0
5	2 (18%)	3 (27%)	4 (36%)	0

Table C.3 Overlap of molecular formulas predicted only keeping annotations having a match in 3 layers by APEX by using different seed nodes in three sets

Number of seed nodes	Matches in SNS 1	Matches in SNS 2	Matches in SNS 2
1	5/5	3/3	15/15
2	18/19	10/10	20/20
5	36/36	38/39	34/35
10	36/38	43/44	46/48
19	48/50	48/50	48/50

Table C.4 Overlap of molecular formula predicted by APEX and all manually annotated compounds (93 in total)

Number of seed nodes	No of APEX annotations in SNS 1	No of APEX annotations in SNS 2	No of APEX annotations in SNS 3
1	53	41	48
2	51	58	67
5	127	112	114
10	142	179	173
19	204	204	204

Table C.5 Total number of annotations made with different seed nodes. The more Seed nodes the more annotations

Bibliography

[1] Timothy D. Veenstra. *Omics in Systems Biology: Current Progress and Future Outlook*. 2021. DOI: 10.1002/pmic.202000235.

[2] Hans Winkler. *Verbreitung und Ursache der Parthenogenesis im Pflanzen- und Tierreiche / von Dr. Hans Winkler.* 2011. DOI: 10.5962/bhl.title.1460.

[3] Bob Kuska. "Beer, Bethesda, and Biology: How "Genomics" Came Into Being". In: *JNCI: Journal of the National Cancer Institute* 90.2 (Jan. 1998), pp. 93–93. ISSN: 0027-8874. DOI: 10.1093/jnci/90.2.93. eprint: https://academic.oup.com/jnci/article-pdf/90/2/93/18080264/90-2-93.pdf. URL: https://doi.org/10.1093/jnci/90.2.93.

[4] Joachim Weischenfeldt et al. "Phenotypic impact of genomic structural variation: insights from and for human disease". In: *Nature Reviews Genetics* 14.2 (2013), pp. 125–138. ISSN: 1471-0064. DOI: 10.1038/nrg3373. URL: https://doi.org/10.1038/nrg3373.

[5] Victor E Velculescu et al. "Characterization of the Yeast Transcriptome". In: *Cell* 88.2 (1997), pp. 243–251. ISSN: 0092-8674. DOI: https://doi.org/10.1016/S0092-8674(00)81845-0. URL: https://www.sciencedirect.com/science/article/pii/S0092867400818450.

[6] Stanislaw Supplitt et al. *Current achievements and applications of transcriptomics in personalized cancer medicine*. 2021. DOI: 10.3390/ijms22031422.

[7] Marc R. Wilkins et al. "Progress with Proteome Projects: Why all Proteins Expressed by a Genome Should be Identified and How To Do It". In: *Biotechnology and Genetic Engineering Reviews* 13.1 (1996). PMID: 8948108, pp. 19–50. DOI: 10.1080/02648725.1996.10647923. eprint: https://doi.org/10.1080/02648725.1996.10647923. URL: https://doi.org/10.1080/02648725.1996.10647923.

[8] Bilal Aslam et al. "Proteomics: Technologies and Their Applications". In: *Journal of Chromatographic Science* 55.2 (Jan. 2017), pp. 182–196. ISSN: 0021-9665. DOI: 10.1093/chromsci/bmw167. eprint: https://academic.oup.com/chromsci/article-pdf/55/2/182/9487766/bmw167.pdf. URL: https://doi.org/10.1093/chromsci/bmw167.

[9] Safa Al-Amrani et al. "Proteomics: Concepts and applications in human medicine". In: *World Journal of Biological Chemistry* 12 (5 2021). ISSN: 1949-8454. DOI: 10.4331/wjbc.v12.i5.57.

[10] Jing Wu, Yi Ding Chen, and Wei Gu. *Urinary proteomics as a novel tool for biomarker discovery in kidney diseases*. 2010. DOI: 10.1631/jzus.B0900327.

[11] B. P. Bowen and T. R. Northen. "Dealing with the unknown: metabolomics and metabolite atlases". In: *J Am Soc Mass Spectrom* 21.9 (2010). 1879-1123 Bowen, Benjamin P Northen, Trent R Journal Article Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, Non-P.H.S. Review United States 2010/05/11 J Am Soc Mass Spectrom. 2010 Sep;21(9):1471-6. doi: 10.1016/j.jasms.2010.04.003. Epub 2010 Apr 12., pp. 1471–6. ISSN: 1044-0305. DOI: 10.1016/j.jasms.2010.04.003.

[12] Tanvir Sajed et al. "ECMDB 2.0: A richer resource for understanding the biochemistry of E. coli". In: *Nucleic Acids Research* 44.D1 (Oct. 2015), pp. D495–D501. ISSN: 0305-1048. DOI: 10.1093/nar/gkv1060. eprint: https://academic.oup.com/nar/article-pdf/44/D1/D495/9482189/gkv1060.pdf. URL: https://doi.org/10.1093/nar/gkv1060.

[13] David S. Wishart et al. “HMDB 4.0: the human metabolome database for 2018”. In: *Nucleic Acids Research* 46.D1 (2018), pp. D608–D617. ISSN: 0305-1048. DOI: 10.1093/nar/gkx1089. URL: https://doi.org/10.1093/nar/gkx1089.

[14] David S. Wishart et al. “HMDB 3.0—The Human Metabolome Database in 2013”. In: *Nucleic Acids Research* 41.D1 (2013), pp. D801–D807. ISSN: 0305-1048. DOI: 10.1093/nar/gks1065. URL: https://doi.org/10.1093/nar/gks1065.

[15] Richard A. Dixon and Dieter Strack. *Phytochemistry meets genome analysis, and beyond.....* 2003. DOI: 10.1016/S0031-9422(02)00712-4.

[16] Farit Mochamad Afendi et al. “KNApSAcK Family Databases: Integrated Metabolite–Plant Species Databases for Multifaceted Plant Research”. In: *Plant and Cell Physiology* 53.2 (Nov. 2011), e1–e1. ISSN: 0032-0781. DOI: 10.1093/pcp/pcr165. eprint: https://academic.oup.com/pcp/article-pdf/53/2/e1/17115220/pcr165.pdf. URL: https://doi.org/10.1093/pcp/pcr165.

[17] Oliver Fiehn. “Metabolomics - The link between genotypes and phenotypes”. In: *Plant Molecular Biology* 48 (1-2 2002). ISSN: 01674412. DOI: 10.1023/A:1013713905833.

[18] K. Dettmer and B. D. Hammock. “Metabolomics–a new exciting field within the "omics" sciences”. In: *Environ Health Perspect* 112.7 (2004), A396–7. ISSN: 0091-6765 (Print) 0091-6765. DOI: 10.1289/ehp.112-1241997.

[19] K. Dettmer, P. A. Aronov, and B. D. Hammock. “Mass spectrometry-based metabolomics”. In: *Mass Spectrom Rev* 26.1 (2007), pp. 51–78. ISSN: 0277-7037 (Print) 0277-7037. DOI: 10.1002/mas.20108.

[20] Balasubramanian Chellammal Muthubharathi, Thirumugam Gowripriya, and Krishnaswamy Balamurugan. *Metabolomics: small molecules that matter more.* 2021. DOI: 10.1039/d0mo00176g.

[21] Markus R. Wenk. *The emerging field of lipidomics.* 2005. DOI: 10.1038/nrd1776.

[22] Chang Wang et al. “Plasma phospholipid metabolic profiling and biomarkers of type 2 diabetes mellitus based on high-performance liquid chromatography/electrospray mass spectrometry and multivariate statistical analysis”. In: *Analytical Chemistry* 77 (13 2005). ISSN: 00032700. DOI: 10.1021/ac0481001.

[23] Min Yi et al. “Integrated Metabolomic and Lipidomic Analysis Reveals the Neuroprotective Mechanisms of Bushen Tiansui Formula in an A β 1-42-Induced Rat Model of Alzheimer’s Disease”. In: *Oxidative Medicine and Cellular Longevity* 2020 (2020). ISSN: 19420994. DOI: 10.1155/2020/5243453.

[24] Lisa M. Butler et al. *Lipids and cancer: Emerging roles in pathogenesis, diagnosis and therapeutic intervention.* 2020. DOI: 10.1016/j.addr.2020.07.013.

[25] Manish Sud et al. “LMSD: LIPID MAPS structure database”. In: *Nucleic Acids Research* 35 (SUPPL. 1 2007). ISSN: 03051048. DOI: 10.1093/nar/gkl838.

[26] Eoin Fahy et al. *Update of the LIPID MAPS comprehensive classification system for lipids.* 2009. DOI: 10.1194/jlr.R800095-JLR200.

[27] G. X. Yang, X. Li, and M. Snyder. “Investigating metabolite-protein interactions: an overview of available techniques”. In: *Methods* 57.4 (2012), pp. 459–66. ISSN: 1046-2023 (Print) 1046-2023. DOI: 10.1016/j.ymeth.2012.06.013.

[28] Gert Wohlgemuth et al. “The chemical translation service-a web-based tool to improve standardization of metabolomic reports”. In: *Bioinformatics* 26 (20 2010). ISSN: 13674803. DOI: 10.1093/bioinformatics/btq476.

[29] Yannick Djoumbou Feunang et al. “ClassyFire: automated chemical classification with a comprehensive, computable taxonomy”. In: *Journal of Cheminformatics* (2016). ISSN: 17582946. DOI: 10.1186/s13321-016-0174-y.

[30] Stephen G. Oliver et al. "Systematic functional analysis of the yeast genome". In: *Trends in Biotechnology* 16.9 (1998), pp. 373–378. ISSN: 0167-7799. DOI: https://doi.org/10.1016/S0167-7799(98)01214-1. URL: https://www.sciencedirect.com/science/article/pii/S0167779998012141.

[31] Warwick B. Dunn, Nigel J. C. Bailey, and Helen E. Johnson. "Measuring the metabolome: current analytical technologies". In: *Analyst* 130 (5 2005), pp. 606–625. DOI: 10.1039/B418288J. URL: http://dx.doi.org/10.1039/B418288J.

[32] Ewa Urbanczyk-Wochniak and Alisdair R. Fernie. "Metabolic profiling reveals altered nitrogen nutrient regimes have diverse effects on the metabolism of hydroponically-grown tomato (Solanum lycopersicum) plants". In: *Journal of Experimental Botany* 56.410 (Dec. 2004), pp. 309–321. ISSN: 0022-0957. DOI: 10.1093/jxb/eri059. eprint: https://academic.oup.com/jxb/article-pdf/56/410/309/16921450/eri059.pdf. URL: https://doi.org/10.1093/jxb/eri059.

[33] Oliver Fiehn et al. "Metabolite profiling for plant functional genomics". In: *Nature Biotechnology* 18 (11 2000). ISSN: 10870156. DOI: 10.1038/81137.

[34] Ute Roessner et al. "Metabolic Profiling Allows Comprehensive Phenotyping of Genetically or Environmentally Modified Plant Systems". In: *The Plant Cell* 13.1 (Jan. 2001), pp. 11–29. ISSN: 1040-4651. DOI: 10.1105/tpc.13.1.11. eprint: https://academic.oup.com/plcell/article-pdf/13/1/11/35422768/plcell_v13_1_11.pdf. URL: https://doi.org/10.1105/tpc.13.1.11.

[35] Fang Chen et al. "Profiling phenolic metabolites in transgenic alfalfa modified in lignin biosynthesis". In: *Phytochemistry* 64.5 (2003), pp. 1013–1021. ISSN: 0031-9422. DOI: https://doi.org/10.1016/S0031-9422(03)00463-1. URL: https://www.sciencedirect.com/science/article/pii/S0031942203004631.

[36] Edda von Roepenack-Lahaye et al. "Profiling of Arabidopsis Secondary Metabolites by Capillary Liquid Chromatography Coupled to Electrospray Ionization Quadrupole Time-of-Flight Mass Spectrometry". In: *Plant Physiology* 134.2 (Feb. 2004), pp. 548–559. ISSN: 0032-0889. DOI: 10.1104/pp.103.032714. eprint: https://academic.oup.com/plphys/article-pdf/134/2/548/38701306/plphys_v134_2_548.pdf. URL: https://doi.org/10.1104/pp.103.032714.

[37] Eng Soon Teoh. *Secondary Metabolites of Plants*. 2016. DOI: 10.1007/978-3-319-24274-3_5.

[38] Saurabh Pagare et al. "Secondary metabolites of plants and their role: Overview". In: *Current Trends in Biotechnology and Pharmacy* 9 (3 2015). ISSN: 22307303.

[39] David S. Wishart. *Applications of metabolomics in drug discovery and development*. 2008. DOI: 10.2165/00126839-200809050-00002.

[40] Charles L. Edelstein. "Biomarkers of Acute Kidney Injury". In: *Advances in Chronic Kidney Disease* 15 (3 2008). ISSN: 15485595. DOI: 10.1053/j.ackd.2008.04.003.

[41] Mikhail Bogdanov et al. "Metabolomic profiling to develop blood biomarkers for Parkinson's disease". In: *Brain* 131.2 (Feb. 2008), pp. 389–396. ISSN: 0006-8950. DOI: 10.1093/brain/awm304. eprint: https://academic.oup.com/brain/article-pdf/131/2/389/1132748/awm304.pdf. URL: https://doi.org/10.1093/brain/awm304.

[42] Xianlin Han et al. "Substantial sulfatide deficiency and ceramide elevation in very early Alzheimer's disease: Potential role in disease pathogenesis". In: *Journal of Neurochemistry* 82 (4 2002). ISSN: 00223042. DOI: 10.1046/j.1471-4159.2002.00997.x.

[43] Joanne T. Brindle et al. "Rapid and noninvasive diagnosis of the presence and severity of coronary heart disease using 1H-NMR-based metabonomics". In: *Nature Medicine* 8 (12 2002). ISSN: 10788956. DOI: 10.1038/nm802.

[44] Warwick B. Dunn et al. “Serum metabolomics reveals many novel metabolic markers of heart failure, including pseudouridine and 2-oxoglutarate”. In: *Metabolomics* 3 (4 2007). ISSN: 15733882. DOI: 10.1007/s11306-007-0063-5.

[45] Natalie J. Serkova et al. “The metabolites citrate, myo-inositol, and spermine are potential age-independent markers of prostate cancer in human expressed prostatic secretions”. In: *The Prostate* 68.6 (2008), pp. 620–628. DOI: https://doi.org/10.1002/pros.20727. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pros.20727. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/pros.20727.

[46] Alyssa K. Kosmides et al. “Metabolomic fingerprinting: Challenges and opportunities”. In: *Critical Reviews in Biomedical Engineering* 41 (3 2013). ISSN: 0278940X. DOI: 10.1615/CritRevBiomedEng.2013007736.

[47] Kunle Odunsi et al. “Detection of epithelial ovarian cancer using 1H-NMR-based metabonomics”. In: *International Journal of Cancer* 113 (5 2005). ISSN: 00207136. DOI: 10.1002/ijc.20651.

[48] Arthur S. Edison et al. “The Time Is Right to Focus on Model Organism Metabolomes”. In: *Metabolites* 6.1 (2016). ISSN: 2218-1989. DOI: 10.3390/metabo6010008. URL: https://www.mdpi.com/2218-1989/6/1/8.

[49] Sastia P. Putri et al. *Current metabolomics: Practical applications*. 2013. DOI: 10.1016/j.jbiosc.2012.12.007.

[50] Shunsuke Hayashi et al. “A novel application of metabolomics in vertebrate development”. In: *Biochemical and Biophysical Research Communications* 386.1 (2009), pp. 268–272. ISSN: 0006-291X. DOI: https://doi.org/10.1016/j.bbrc.2009.06.041. URL: https://www.sciencedirect.com/science/article/pii/S0006291X09011747.

[51] Kamilla Sofie Pedersen et al. “Metabolomic signatures of inbreeding at benign and stressful temperatures in Drosophila melanogaster”. In: *Genetics* 180 (2 2008). ISSN: 00166731. DOI: 10.1534/genetics.108.089144.

[52] Anders Malmendal et al. “Metabolomic profiling of heat stress: Hardening and recovery of homeostasis in Drosophila”. In: *American Journal of Physiology - Regulatory Integrative and Comparative Physiology* 291 (1 2006). ISSN: 03636119. DOI: 10.1152/ajpregu.00867.2005.

[53] Laurence Coquin et al. “Metabolomic and flux-balance analysis of age-related decline of hypoxia tolerance in Drosophila muscle tissue”. In: *Molecular Systems Biology* 4 (2008). ISSN: 17444292. DOI: 10.1038/msb.2008.71.

[54] Jeffrey R. Idle and Frank J. Gonzalez. “Metabolomics”. In: *Cell Metabolism* 6.5 (2007), pp. 348–351. ISSN: 1550-4131. DOI: https://doi.org/10.1016/j.cmet.2007.10.005. URL: https://www.sciencedirect.com/science/article/pii/S1550413107002987.

[55] Sydney Brenner. “Nature’s Gift to Science (Nobel Lecture)”. In: *ChemBioChem* 4.8 (2003), pp. 683–687. ISSN: 1439-4227. DOI: 10.1002/cbic.200300625. URL: https://chemistry-europe.onlinelibrary.wiley.com/doi/abs/10.1002/cbic.200300625.

[56] A. Barriere and M. A. Felix. “High local genetic diversity and low outcrossing rate in Caenorhabditis elegans natural populations”. In: *Current Biology* 15.13 (2005), pp. 1176–1184. ISSN: 0960-9822. DOI: 10.1016/j.cub.2005.06.022. URL: %3CGo%20to%20ISI%3E://WOS:000230662400021.

[57] Mark Lucanic et al. “A Simple Method for High Throughput Chemical Screening in Caenorhabditis Elegans”. In: *Journal of visualized experiments : JoVE* 133 (2018), p. 56892. ISSN: 1940-087X. DOI: 10.3791/56892. URL: https://pubmed.ncbi.nlm.nih.gov/29630057%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5933220/.

[58] P. J. Hu. "Dauer". In: *WormBook* (2007), pp. 1–19. ISSN: 1551-8507 (Electronic) 1551-8507 (Linking). DOI: 10.1895/wormbook.1.144.1. URL: https://www.ncbi.nlm.nih.gov/pubmed/17988074.

[59] Titus Kaletta and Michael O. Hengartner. "Finding function in novel targets: C. elegans as a model organism". In: *Nature Reviews Drug Discovery* 5.5 (2006), pp. 387–399. ISSN: 1474-1784. DOI: 10.1038/nrd2031. URL: https://doi.org/10.1038/nrd2031.

[60] Andrew Fire et al. "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans". In: *Nature* 391.6669 (1998), pp. 806–811. ISSN: 1476-4687. DOI: 10.1038/35888. URL: https://doi.org/10.1038/35888.

[61] Pierre Gönczy et al. "Functional genomic analysis of cell division in C. elegans using RNAi of genes on chromosome III". In: *Nature* 408.6810 (2000), pp. 331–336. ISSN: 1476-4687. DOI: 10.1038/35042526. URL: https://doi.org/10.1038/35042526.

[62] Craig C. Mello and Darryl Conte. "Revealing the world of RNA interference". In: *Nature* 431.7006 (2004), pp. 338–342. ISSN: 1476-4687. DOI: 10.1038/nature02872. URL: https://doi.org/10.1038/nature02872.

[63] Andrew G. Fraser et al. "Functional genomic analysis of C. elegans chromosome I by systematic RNA interference". In: *Nature* 408.6810 (2000), pp. 325–330. ISSN: 1476-4687. DOI: 10.1038/35042517. URL: https://doi.org/10.1038/35042517.

[64] Lisa Timmons and Andrew Fire. "Specific interference by ingested dsRNA". In: *Nature* 395.6705 (1998), pp. 854–854. ISSN: 1476-4687. DOI: 10.1038/27579. URL: https://doi.org/10.1038/27579.

[65] J. A. Arribere et al. "Efficient marker-free recovery of custom genetic modifications with CRISPR/Cas9 in Caenorhabditis elegans". In: *Genetics* 198.3 (2014), pp. 837–46. ISSN: 0016-6731 (Print) 0016-6731. DOI: 10.1534/genetics.114.169730.

[66] Vinci Au et al. "CRISPR/Cas9 Methodology for the Generation of Knockout Deletions in Caenorhabditis elegans". In: *G3: Genes|Genomes|Genetics* 9.1 (2019), pp. 135–144. DOI: 10.1534/g3.118.200778. URL: https://www.g3journal.org/content/ggg/9/1/135.full.pdf.

[67] D. J. Dickinson et al. "Streamlined Genome Engineering with a Self-Excising Drug Selection Cassette". In: *Genetics* 200.4 (2015), pp. 1035–49. ISSN: 0016-6731 (Print) 0016-6731. DOI: 10.1534/genetics.115.178335.

[68] Chintan J. Joshi et al. "StanDep: Capturing transcriptomic variability improves context-specific metabolic models". In: *PLOS Computational Biology* 16.5 (2020), e1007764. DOI: 10.1371/journal.pcbi.1007764. URL: https://doi.org/10.1371/journal.pcbi.1007764.

[69] Hyun-Min Kim and Monica P. Colaiácovo. "CRISPR-Cas9-Guided Genome Engineering in Caenorhabditis elegans". In: *Current Protocols in Molecular Biology* 129.1 (2019), e106. ISSN: 1934-3639. DOI: https://doi.org/10.1002/cpmb.106. URL: https://currentprotocols.onlinelibrary.wiley.com/doi/abs/10.1002/cpmb.106.

[70] A. Paix et al. "High Efficiency, Homology-Directed Genome Editing in Caenorhabditis elegans Using CRISPR-Cas9 Ribonucleoprotein Complexes". In: *Genetics* 201.1 (2015), pp. 47–54. ISSN: 0016-6731 (Print) 0016-6731. DOI: 10.1534/genetics.115.179382.

[71] J. D. Ward. "Rapid and precise engineering of the Caenorhabditis elegans genome with lethal mutation co-conversion and inactivation of NHEJ repair". In: *Genetics* 199.2 (2015), pp. 363–77. ISSN: 0016-6731 (Print) 0016-6731. DOI: 10.1534/genetics.114.172361.

[72] Daniel J. Dickinson and Bob Goldstein. "CRISPR-Based Methods for Caenorhabditis elegans Genome Engineering". In: *Genetics* 202.3 (2016), pp. 885–901. ISSN: 1943-2631 0016-6731. DOI: 10.1534/genetics.115.182162. URL: https://pubmed.ncbi.nlm.nih.gov/26953268%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4788126/.

[73] Alisdair R. Fernie et al. "Metabolite profiling: from diagnostics to systems biology". In: *Nature Reviews Molecular Cell Biology* 5.9 (2004), pp. 763–769. ISSN: 1471-0080. DOI: 10.1038/nrm1451. URL: https://doi.org/10.1038/nrm1451.

[74] Ralph J DeBerardinis and Craig B Thompson. "Cellular Metabolism and Disease: What Do Metabolic Outliers Teach Us?" In: *Cell* 148.6 (2012), pp. 1132–1144. ISSN: 0092-8674. DOI: https://doi.org/10.1016/j.cell.2012.02.032. URL: http://www.sciencedirect.com/science/article/pii/S0092867412002322.

[75] Stephen G. Oliver et al. "Systematic functional analysis of the yeast genome". In: *Trends in Biotechnology* 16.9 (1998), pp. 373–378. ISSN: 0167-7799. DOI: https://doi.org/10.1016/S0167-7799(98)01214-1. URL: http://www.sciencedirect.com/science/article/pii/S0167779998012141.

[76] Alexander B. Artyukhin et al. "Metabolomic "Dark Matter" Dependent on Peroxisomal β-Oxidation in Caenorhabditis elegans". In: *Journal of the American Chemical Society* 140.8 (2018), pp. 2841–2852. ISSN: 0002-7863. DOI: 10.1021/jacs.7b11811. URL: https://doi.org/10.1021/jacs.7b11811.

[77] Ricardo R. da Silva, Pieter C. Dorrestein, and Robert A. Quinn. "Illuminating the dark matter in metabolomics". In: *Proceedings of the National Academy of Sciences* 112.41 (2015), pp. 12549–12550. DOI: 10.1073/pnas.1516878112. URL: https://www.pnas.org/content/pnas/112/41/12549.full.pdf.

[78] Yevgeniy Izrayelit et al. "Targeted Metabolomics Reveals a Male Pheromone and Sex-Specific Ascaroside Biosynthesis in Caenorhabditis elegans". In: *ACS Chemical Biology* 7.8 (2012), pp. 1321–1325. ISSN: 1554-8929. DOI: 10.1021/cb300169c. URL: https://doi.org/10.1021/cb300169c.

[79] Jan M. Falcke et al. "Linking Genomic and Metabolomic Natural Variation Uncovers Nematode Pheromone Biosynthesis". In: *Cell Chemical Biology* 25.6 (2018), 787–796.e12. ISSN: 2451-9456. DOI: https://doi.org/10.1016/j.chembiol.2018.04.004. URL: http://www.sciencedirect.com/science/article/pii/S2451945618301181.

[80] Chirag Pungaliya et al. "A shortcut to identifying small molecule signals that regulate behavior and development in Caenorhabditis elegans". In: *Proceedings of the National Academy of Sciences* 106.19 (2009), pp. 7708–7713. DOI: 10.1073/pnas.0811918106. URL: https://www.pnas.org/content/pnas/106/19/7708.full.pdf.

[81] Stephan H. von Reuss et al. "Comparative Metabolomics Reveals Biogenesis of Ascarosides, a Modular Library of Small-Molecule Signals in C. elegans". In: *Journal of the American Chemical Society* 134.3 (2012), pp. 1817–1824. ISSN: 0002-7863. DOI: 10.1021/ja210202y. URL: https://doi.org/10.1021/ja210202y.

[82] Stephan H. von Reuss, Franziska Dolke, and Chuanfu Dong. "Ascaroside Profiling of Caenorhabditis elegans Using Gas Chromatography–Electron Ionization Mass Spectrometry". In: *Analytical Chemistry* 89.19 (2017), pp. 10570–10577. ISSN: 0003-2700. DOI: 10.1021/acs.analchem.7b02803. URL: https://doi.org/10.1021/acs.analchem.7b02803.

[83] Hugo Aguilaniu, Paola Fabrizio, and Michael Witting. "The Role of Dafachronic Acid Signaling in Development and Longevity in Caenorhabditis elegans: Digging Deeper Using Cutting-Edge Analytical Chemistry". In: *Frontiers in endocrinology* 7 (2016), pp. 12–12. ISSN: 1664-2392. DOI: 10.3389/fendo.2016.00012. URL: https://pubmed.ncbi.nlm.nih.gov/26903948%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4749721/.

[84] Sider Penkov et al. “Maradolipids: Diacyltrehalose Glycolipids Specific to Dauer Larva in Caenorhabditis elegans”. In: *Angewandte Chemie International Edition* 49.49 (2010), pp. 9430–9435. ISSN: 1433-7851. DOI: 10.1002/anie.201004466. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201004466.

[85] Florian M. Geier et al. “Cross-Platform Comparison of Caenorhabditis elegans Tissue Extraction Strategies for Comprehensive Metabolome Coverage”. In: *Analytical Chemistry* 83.10 (2011), pp. 3730–3736. ISSN: 0003-2700. DOI: 10.1021/ac2001109. URL: https://doi.org/10.1021/ac2001109.

[86] Emelyne Teo et al. “Metabolic stress is a primary pathogenic event in transgenic Caenorhabditis elegans expressing pan-neuronal human amyloid beta”. In: *eLife* 8 (2019), e50069. ISSN: 2050-084X. DOI: 10.7554/eLife.50069. URL: https://pubmed.ncbi.nlm.nih.gov/31610847%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6794093/.

[87] Stefan Zdraljevic et al. “Natural variation in C. elegans arsenic toxicity is explained by differences in branched chain amino acid metabolism”. In: *eLife* 8 (2019), e40260. ISSN: 2050-084X. DOI: 10.7554/eLife.40260. URL: https://doi.org/10.7554/eLife.40260.

[88] Stefanie Müthel et al. “The conserved histone chaperone LIN-53 is required for normal lifespan and maintenance of muscle integrity in Caenorhabditis elegans”. In: *Aging Cell* 18.6 (2019), e13012. ISSN: 1474-9718. DOI: 10.1111/acel.13012. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/acel.13012.

[89] Arwen W. Gao et al. “Identification of key pathways and metabolic fingerprints of longevity in C. elegans”. In: *Experimental Gerontology* 113 (2018), pp. 128–140. ISSN: 0531-5565. DOI: https://doi.org/10.1016/j.exger.2018.10.003. URL: http://www.sciencedirect.com/science/article/pii/S0531556518301141.

[90] J. Folch, M. Lees, and G. H. Sloane Stanley. “A simple method for the isolation and purification of total lipides from animal tissues”. In: *The Journal of biological chemistry* 226.1 (1957), pp. 497–509. URL: https://www.scopus.com/inward/record.uri?eid=2-s2.0-70449158340%5C&partnerID=40%5C&md5=4f05700fa89b7f7a1293ca5bef961778.

[91] E. G. Bligh and W. J. Dyer. “A RAPID METHOD OF TOTAL LIPID EXTRACTION AND PURIFICATION”. In: *Canadian Journal of Biochemistry and Physiology* 37.8 (1959), pp. 911–917. ISSN: 0576-5544. DOI: 10.1139/o59-099. URL: https://doi.org/10.1139/o59-099.

[92] Vitali Matyash et al. “Lipid extraction by methyl-tert-butyl ether for high-throughput lipidomics”. In: *Journal of lipid research* 49.5 (2008), pp. 1137–1146. ISSN: 0022-2275. DOI: 10.1194/jlr.D700041-JLR200. URL: https://pubmed.ncbi.nlm.nih.gov/18281723%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2311442/.

[93] Maja Klapper et al. “Methyl group donors abrogate adaptive responses to dietary restriction in C. elegans”. In: *Genes & Nutrition* 11.1 (2016), p. 4. ISSN: 1865-3499. DOI: 10.1186/s12263-016-0522-4. URL: https://doi.org/10.1186/s12263-016-0522-4.

[94] Samuel S. W. Szeto, Stacey N. Reinke, and Bernard D. Lemire. “1H NMR-based metabolic profiling reveals inherent biological variation in yeast and nematode model systems”. In: *Journal of Biomolecular NMR* 49.3 (2011), pp. 245–254. ISSN: 1573-5001. DOI: 10.1007/s10858-011-9492-6. URL: https://doi.org/10.1007/s10858-011-9492-6.

[95] John L. Markley et al. “The future of NMR-based metabolomics”. In: *Current Opinion in Biotechnology* 43 (2017), pp. 34–40. ISSN: 0958-1669. DOI: https://doi.org/10.1016/j.copbio.2016.08.001. URL: http://www.sciencedirect.com/science/article/pii/S0958166916301768.

[96] Cecilia Castro et al. "A study of Caenorhabditis elegans DAF-2 mutants by metabolomics and differential correlation networks". In: *Molecular bioSystems* 9.7 (2013), pp. 1632–1642. ISSN: 1742-2051 1742-206X. DOI: 10.1039/c3mb25539e. URL: https://pubmed.ncbi.nlm.nih.gov/23475189%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7002165/.

[97] Qin-Li Wan et al. "Metabolomic signature associated with reproduction-regulated aging in Caenorhabditis elegans". In: *Aging* 9.2 (2017), pp. 447–474. ISSN: 1945-4589. DOI: 10.18632/aging.101170. URL: https://doi.org/10.18632/aging.101170.

[98] Artur B. Lourenço et al. "Analysis of the effect of the mitochondrial prohibitin complex, a context-dependent modulator of longevity, on the C. elegans metabolome". In: *Biochimica et Biophysica Acta (BBA) - Bioenergetics* 1847.11 (2015), pp. 1457–1468. ISSN: 0005-2728. DOI: https://doi.org/10.1016/j.bbabio.2015.06.003. URL: http://www.sciencedirect.com/science/article/pii/S0005272815001127.

[99] Francois-Pierre J. Martin et al. "Metabotyping of Caenorhabditis elegans and their Culture Media Revealed Unique Metabolic Phenotypes Associated to Amino Acid Deficiency and Insulin-Like Signaling". In: *Journal of Proteome Research* 10.3 (2011), pp. 990–1003. ISSN: 1535-3893. DOI: 10.1021/pr100703a. URL: https://doi.org/10.1021/pr100703a.

[100] Silke Fuchs et al. "A metabolic signature of long life in Caenorhabditis elegans". In: *BMC Biology* 8.1 (2010), p. 14. ISSN: 1741-7007. DOI: 10.1186/1741-7007-8-14. URL: https://doi.org/10.1186/1741-7007-8-14.

[101] David I. Schlipalius et al. "A Core Metabolic Enzyme Mediates Resistance to Phosphine Gas". In: *Science* 338.6108 (2012), pp. 807–810. DOI: 10.1126/science.1224951. URL: https://science.sciencemag.org/content/sci/338/6108/807.full.pdf.

[102] Jagan Srinivasan et al. "A blend of small molecules regulates both mating and development in Caenorhabditis elegans". In: *Nature* 454.7208 (2008), pp. 1115–1118. ISSN: 1476-4687. DOI: 10.1038/nature07168. URL: https://doi.org/10.1038/nature07168.

[103] Helen J. Atherton et al. "A comparative metabolomic study of NHR-49 in Caenorhabditis elegans and PPAR-α in the mouse". In: *FEBS Letters* 582.12 (2008), pp. 1661–1666. ISSN: 0014-5793. DOI: 10.1016/j.febslet.2008.04.020. URL: https://febs.onlinelibrary.wiley.com/doi/abs/10.1016/j.febslet.2008.04.020.

[104] Yong Jin An et al. "Metabotyping of the C. elegans sir-2.1 Mutant Using in Vivo Labeling and 13C-Heteronuclear Multidimensional NMR Metabolomics". In: *ACS Chemical Biology* 7.12 (2012), pp. 2012–2018. ISSN: 1554-8929. DOI: 10.1021/cb3004226. URL: https://doi.org/10.1021/cb3004226.

[105] M. Osman Sheikh et al. "Correlations Between LC-MS/MS-Detected Glycomics and NMR-Detected Metabolomics in Caenorhabditis elegans Development". In: *Frontiers in Molecular Biosciences* 6.49 (2019). ISSN: 2296-889X. DOI: 10.3389/fmolb.2019.00049. URL: https://www.frontiersin.org/article/10.3389/fmolb.2019.00049.

[106] Florian M. Geier, Armand M. Leroi, and Jacob G. Bundy. "13C Labeling of Nematode Worms to Improve Metabolome Coverage by Heteronuclear Nuclear Magnetic Resonance Experiments". In: *Frontiers in Molecular Biosciences* 6.27 (2019). ISSN: 2296-889X. DOI: 10.3389/fmolb.2019.00027. URL: https://www.frontiersin.org/article/10.3389/fmolb.2019.00027.

[107] David S. Wishart et al. "HMDB: the Human Metabolome Database". In: *Nucleic Acids Research* 35.suppl1 (2007), pp. D521–D526. ISSN: 0305-1048. DOI: 10.1093/nar/gkl923. URL: https://doi.org/10.1093/nar/gkl923.

[108] David S. Wishart et al. "HMDB: a knowledgebase for the human metabolome". In: *Nucleic Acids Research* 37.suppl1 (2009), pp. D603–D610. ISSN: 0305-1048. DOI: 10.1093/nar/gkn810. URL: https://doi.org/10.1093/nar/gkn810.

[109] David S. Wishart et al. "PathBank: a comprehensive pathway database for model organisms". In: *Nucleic Acids Research* 48.D1 (2020), pp. D470–D478. ISSN: 0305-1048. DOI: 10.1093/nar/gkz861. URL: https://doi.org/10.1093/nar/gkz861.

[110] Parag Mahanti et al. "Comparative Metabolomics Reveals Endogenous Ligands of DAF-12, a Nuclear Hormone Receptor, Regulating C. elegans Development and Lifespan". In: *Cell Metabolism* 19.1 (2014), pp. 73–83. ISSN: 1550-4131. DOI: https://doi.org/10.1016/j.cmet.2013.11.024. URL: http://www.sciencedirect.com/science/article/pii/S1550413113004993.

[111] Benjamin J. Blaise et al. "Metabotyping of Caenorhabditis elegans reveals latent phenotypes". In: *Proceedings of the National Academy of Sciences* 104.50 (2007), pp. 19808–19812. DOI: 10.1073/pnas.0707393104. URL: https://www.pnas.org/content/pnas/104/50/19808.full.pdf.

[112] Clément Pontoizeau et al. "Metabolomics Analysis Uncovers That Dietary Restriction Buffers Metabolic Changes Associated with Aging in Caenorhabditis elegans". In: *Journal of Proteome Research* 13.6 (2014), pp. 2910–2919. ISSN: 1535-3893. DOI: 10.1021/pr5000686. URL: https://doi.org/10.1021/pr5000686.

[113] Alan Wong et al. "μHigh Resolution-Magic-Angle Spinning NMR Spectroscopy for Metabolic Phenotyping of Caenorhabditis elegans". In: *Analytical Chemistry* 86.12 (2014), pp. 6064–6070. ISSN: 0003-2700. DOI: 10.1021/ac501208z. URL: https://doi.org/10.1021/ac501208z.

[114] D. Sakellariou, G. Le Goff, and J. F. Jacquinot. "High-resolution, high-sensitivity NMR of nanolitre anisotropic samples by coil spinning". In: *Nature* 447.7145 (2007), pp. 694–697. ISSN: 1476-4687. DOI: 10.1038/nature05897. URL: https://doi.org/10.1038/nature05897.

[115] Roy Teranishi et al. "Gas chromatography of volatiles from breath and urine". In: *Analytical Chemistry* 44.1 (1972), pp. 18–20. ISSN: 0003-2700. DOI: 10.1021/ac60309a012. URL: https://doi.org/10.1021/ac60309a012.

[116] Carsten Jaeger et al. "Metabolomic changes in Caenorhabditis elegans lifespan mutants as evident from GC–EI–MS and GC–APCI–TOF–MS profiling". In: *Metabolomics* 10.5 (2014), pp. 859–876. ISSN: 1573-3890. DOI: 10.1007/s11306-014-0637-y. URL: https://doi.org/10.1007/s11306-014-0637-y.

[117] Ian D. Wilson et al. "High Resolution "Ultra Performance" Liquid Chromatography Coupled to oa-TOF Mass Spectrometry as a Tool for Differential Metabolic Pathway Profiling in Functional Genomic Studies". In: *Journal of Proteome Research* 4.2 (2005), pp. 591–598. ISSN: 1535-3893. DOI: 10.1021/pr049769r. URL: https://doi.org/10.1021/pr049769r.

[118] Janna Hastings et al. "Multi-Omics and Genome-Scale Modeling Reveal a Metabolic Shift During C. elegans Aging". In: *Frontiers in Molecular Biosciences* 6.2 (2019). ISSN: 2296-889X. DOI: 10.3389/fmolb.2019.00002. URL: https://www.frontiersin.org/article/10.3389/fmolb.2019.00002.

[119] Rawi Ramautar, Govert W. Somsen, and Gerhardus J. de Jong. "CE-MS for metabolomics: Developments and applications in the period 2016–2018". In: *ELECTROPHORESIS* 40.1 (2019), pp. 165–179. ISSN: 0173-0835. DOI: 10.1002/elps.201800323. URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/elps.201800323.

[120] Wei Zhang and Rawi Ramautar. "CE-MS for metabolomics: Developments and applications in the period 2018-2020." eng. In: *Electrophoresis* 42.4 (Feb. 2021), pp. 381–401. ISSN: 1522-2683 (Electronic); 0173-0835 (Print); 0173-0835 (Linking). DOI: 10.1002/elps.202000203.

[121] Giuseppe Paglia and Giuseppe Astarita. "Metabolomics and lipidomics using traveling-wave ion mobility mass spectrometry". In: *Nature Protocols* 12.4 (2017), pp. 797–813. ISSN: 1750-2799. DOI: 10.1038/nprot.2017.013. URL: https://doi.org/10.1038/nprot.2017.013.

[122] Dominik Schwudke et al. "Top-Down Lipidomic Screens by Multivariate Analysis of High-Resolution Survey Mass Spectra". In: *Analytical Chemistry* 79.11 (2007), pp. 4083–4093. ISSN: 0003-2700. DOI: 10.1021/ac062455y. URL: https://doi.org/10.1021/ac062455y.

[123] Parise Henry et al. "Fatty acids composition of Caenorhabditis elegans using accurate mass GCMS-QTOF". In: *Journal of environmental science and health. Part. B, Pesticides, food contaminants, and agricultural wastes* 51.8 (2016), pp. 546–552. ISSN: 1532-4109 0360-1234. DOI: 10.1080/03601234.2016.1170555. URL: https://pubmed.ncbi.nlm.nih.gov/27166662%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5052121/.

[124] Mark Lucanic et al. "N-acylethanolamine signalling mediates the effect of diet on lifespan in Caenorhabditis elegans". In: *Nature* 473.7346 (2011), pp. 226–229. ISSN: 1476-4687. DOI: 10.1038/nature10007. URL: https://doi.org/10.1038/nature10007.

[125] Andrew Folick et al. "Lysosomal signaling molecules regulate longevity in Caenorhabditis elegans". In: *Science* 347.6217 (2015), pp. 83–86. DOI: 10.1126/science.1258857. URL: https://science.sciencemag.org/content/sci/347/6217/83.full.pdf.

[126] Wenqing Wang et al. "Comparative Metabolomic Profiling Reveals That Dysregulated Glycolysis Stemming from Lack of Salvage NAD+ Biosynthesis Impairs Reproductive Development in Caenorhabditis elegans". In: *Journal of Biological Chemistry* 290.43 (2015), pp. 26163–26179. DOI: 10.1074/jbc.M115.662916. URL: http://www.jbc.org/content/290/43/26163.abstract.

[127] Tzu-Ling Chen et al. "Impaired embryonic development in glucose-6-phosphate dehydrogenase-deficient Caenorhabditis elegans due to abnormal redox homeostasis induced activation of calcium-independent phospholipase and alteration of glycerophospholipid metabolism". In: *Cell Death & Disease* 8.1 (2018), e2545–e2545. ISSN: 2041-4889. DOI: 10.1038/cddis.2016.463. URL: https://doi.org/10.1038/cddis.2016.463.

[128] Mai-Britt Mosbech et al. "Functional Loss of Two Ceramide Synthases Elicits Autophagy-Dependent Lifespan Extension in C. elegans". In: *PLOS ONE* 8.7 (2013), e70087. DOI: 10.1371/journal.pone.0070087. URL: https://doi.org/10.1371/journal.pone.0070087.

[129] Verena Schmökel et al. "Genetics of Lipid-Storage Management in Caenorhabditis elegans Embryos". In: *Genetics* 202.3 (2016), pp. 1071–1083. DOI: 10.1534/genetics.115.179127. URL: https://www.genetics.org/content/genetics/202/3/1071.full.pdf.

[130] Roel Van Assche et al. "Metabolic profiling of a transgenic Caenorhabditis elegans Alzheimer model". In: *Metabolomics : Official journal of the Metabolomic Society* 11.2 (2015), pp. 477–486. ISSN: 1573-3882 1573-3890. DOI: 10.1007/s11306-014-0711-5. URL: https://pubmed.ncbi.nlm.nih.gov/25750603%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4342517/.

[131] Li Ma et al. "Systems biology analysis using a genome-scale metabolic model shows that phosphine triggers global metabolic suppression in a resistant strain of C. elegans". In: *bioRxiv* (2017), p. 144386. DOI: 10.1101/144386. URL: https://www.biorxiv.org/content/biorxiv/early/2017/05/31/144386.full.pdf.

[132] Michael Witting et al. "DI-ICR-FT-MS-based high-throughput deep metabotyping: a case study of the Caenorhabditis elegans–Pseudomonas aeruginosa infection model". In: *Analytical and Bioanalytical Chemistry* 407.4 (2015), pp. 1059–1073. ISSN: 1618-2650. DOI: 10.1007/s00216-014-8331-5. URL: https://doi.org/10.1007/s00216-014-8331-5.

[133] Geert Depuydt et al. "LC–MS Proteomics Analysis of the Insulin/IGF-1-Deficient Caenorhabditis elegans daf-2(e1370) Mutant Reveals Extensive Restructuring of Intermediary Metabolism". In: *Journal of Proteome Research* 13.4 (2014), pp. 1938–1956. ISSN: 1535-3893. DOI: 10.1021/pr401081b. URL: https://doi.org/10.1021/pr401081b.

[134] Arwen W. Gao et al. "A sensitive mass spectrometry platform identifies metabolic changes of life history traits in C. elegans". In: *Scientific Reports* 7.1 (2017), p. 2408. ISSN: 2045-2322. DOI: 10.1038/s41598-017-02539-w. URL: https://doi.org/10.1038/s41598-017-02539-w.

[135] Hyeon-Cheol Lee et al. "Caenorhabditis elegans mboa-7, a member of the MBOAT family, is required for selective incorporation of polyunsaturated fatty acids into phosphatidylinositol". In: *Molecular biology of the cell* 19.3 (2008), pp. 1174–1184. ISSN: 1939-4586 1059-1524. DOI: 10.1091/mbc.e07-09-0893. URL: https://pubmed.ncbi.nlm.nih.gov/18094042%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2262980/.

[136] Vincent Menuz et al. "Protection of C. elegans from Anoxia by HYL-2 Ceramide Synthase". In: *Science* 324.5925 (2009), pp. 381–384. DOI: 10.1126/science.1168532. URL: https://science.sciencemag.org/content/sci/324/5925/381.full.pdf.

[137] Cyrus Papan et al. "Systematic Screening for Novel Lipids by Shotgun Lipidomics". In: *Analytical Chemistry* 86.5 (2014), pp. 2703–2710. ISSN: 0003-2700. DOI: 10.1021/ac404083u. URL: https://doi.org/10.1021/ac404083u.

[138] Sarah K. Davies et al. "The mutational structure of metabolism in Caenorhabditis elegans". In: *Evolution; international journal of organic evolution* 70.10 (2016), pp. 2239–2246. ISSN: 1558-5646 0014-3820. DOI: 10.1111/evo.13020. URL: https://pubmed.ncbi.nlm.nih.gov/27465022%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5050113/.

[139] Qin-Li Wan et al. "The Effects of Age and Reproduction on the Lipidome of Caenorhabditis elegans". In: *Oxidative Medicine and Cellular Longevity* 2019 (2019), p. 5768953. ISSN: 1942-0900. DOI: 10.1155/2019/5768953. URL: https://doi.org/10.1155/2019/5768953.

[140] C. L. Perez and M. R. Van Gilst. "A 13C isotope labeling strategy reveals the influence of insulin signaling on lipogenesis in C. elegans". In: *Cell Metab* 8.3 (2008), pp. 266–74. ISSN: 1932-7420 (Electronic) 1550-4131 (Linking). DOI: 10.1016/j.cmet.2008.08.007. URL: https://www.ncbi.nlm.nih.gov/pubmed/18762027.

[141] Xun Shi et al. "Regulation of lipid droplet size and phospholipid composition by stearoyl-CoA desaturase". In: *Journal of lipid research* 54.9 (2013), pp. 2504–2514. ISSN: 1539-7262 0022-2275. DOI: 10.1194/jlr.M039669. URL: https://pubmed.ncbi.nlm.nih.gov/23787165%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3735947/.

[142] Aude D. Bouagnon et al. "Intestinal peroxisomal fatty acid β-oxidation regulates neural serotonin signaling through a feedback mechanism". In: *PLOS Biology* 17.12 (2019), e3000242. DOI: 10.1371/journal.pbio.3000242. URL: https://doi.org/10.1371/journal.pbio.3000242.

[143] Tracy L. Vrablik and Jennifer L. Watts. "Polyunsaturated fatty acid derived signaling in reproduction and development: insights from Caenorhabditis elegans and Drosophila melanogaster". In: *Molecular reproduction and development* 80.4 (2013), pp. 244–259. ISSN: 1098-2795 1040-452X. DOI: 10.1002/mrd.22167. URL: https://pubmed.ncbi.nlm.nih.gov/23440886%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4350910/.

[144] Felipe Macedo et al. "Lipase-like 5 enzyme controls mitochondrial activity in response to starvation in Caenorhabditis elegans". In: *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* 1865.2 (2020), p. 158539. ISSN: 1388-1981. DOI: https://doi.org/10.1016/j.bbalip.2019.158539. URL: http://www.sciencedirect.com/science/article/pii/S1388198119301908.

[145] Simon Haeussler et al. “Autophagy compensates for defects in mitochondrial dynamics”. In: *PLOS Genetics* 16.3 (2020), e1008638. DOI: 10.1371/journal.pgen.1008638. URL: https://doi.org/10.1371/journal.pgen.1008638.

[146] Kathrine B. Dall et al. “HLH-30 dependent rewiring of metabolism during starvation in C. elegans”. In: *bioRxiv* (2020), p. 2020.06.26.170555. DOI: 10.1101/2020.06.26.170555. URL: https://www.biorxiv.org/content/biorxiv/early/2020/06/26/2020.06.26.170555.full.pdf.

[147] Tesfahun Dessale Admasu et al. “Drug Synergy Slows Aging and Improves Healthspan through IGF and SREBP Lipid Signaling”. In: *Developmental Cell* 47.1 (2018), 67–79.e5. ISSN: 1534-5807. DOI: https://doi.org/10.1016/j.devcel.2018.09.001. URL: http://www.sciencedirect.com/science/article/pii/S1534580718306993.

[148] Tesfahun Dessale Admasu et al. “Lipid profiling of C. elegans strains administered pro-longevity drugs and drug combinations”. In: *Scientific Data* 5.1 (2018), p. 180231. ISSN: 2052-4463. DOI: 10.1038/sdata.2018.231. URL: https://doi.org/10.1038/sdata.2018.231.

[149] S. M. Lam et al. “Sequestration of polyunsaturated fatty acids in membrane phospholipids of Caenorhabditis elegans dauer larva attenuates eicosanoid biosynthesis for prolonged survival”. In: *Redox Biology* 12 (2017), pp. 967–977. ISSN: 2213-2317. DOI: 10.1016/j.redox.2017.05.002. URL: %3CGo%20to%20ISI%3E://WOS:000403328700096.

[150] J. T. Hannich et al. “Structure and conserved function of iso-branched sphingoid bases from the nematode Caenorhabditis elegans”. In: *Chem Sci* 8.5 (2017), pp. 3676–3686. ISSN: 2041-6520 (Print) 2041-6520 (Linking). DOI: 10.1039/c6sc04831e. URL: https://www.ncbi.nlm.nih.gov/pubmed/30155209.

[151] Victoria Hänel, Christian Pendleton, and Michael Witting. “The sphingolipidome of the model organism Caenorhabditis elegans”. In: *Chemistry and Physics of Lipids* 222 (2019), pp. 15–22. ISSN: 0009-3084. DOI: https://doi.org/10.1016/j.chemphyslip.2019.04.009. URL: http://www.sciencedirect.com/science/article/pii/S0009308419300428.

[152] Hung-Chi Yang et al. “IDH-1 deficiency induces growth defects and metabolic alterations in GSPD-1-deficient Caenorhabditis elegans”. In: *Journal of Molecular Medicine* 97.3 (2019), pp. 385–396. ISSN: 1432-1440. DOI: 10.1007/s00109-018-01740-2. URL: https://doi.org/10.1007/s00109-018-01740-2.

[153] Hunbeen Kim et al. “JAK/STAT and TGF-ß activation as potential adverse outcome pathway of TiO2NPs phototoxicity in Caenorhabditis elegans”. In: *Scientific Reports* 7.1 (2017), p. 17833. ISSN: 2045-2322. DOI: 10.1038/s41598-017-17495-8. URL: https://doi.org/10.1038/s41598-017-17495-8.

[154] Lloyd W. Sumner et al. “Proposed minimum reporting standards for chemical analysis”. In: *Metabolomics* 3.3 (2007), pp. 211–221. ISSN: 1573-3890. DOI: 10.1007/s11306-007-0082-2. URL: https://doi.org/10.1007/s11306-007-0082-2.

[155] Carlos Guijas et al. “METLIN: A Technology Platform for Identifying Knowns and Unknowns”. In: *Analytical Chemistry* 90.5 (2018), pp. 3156–3164. ISSN: 0003-2700. DOI: 10.1021/acs.analchem.7b04424. URL: https://doi.org/10.1021/acs.analchem.7b04424.

[156] Eldon L. Ulrich et al. “BioMagResBank”. In: *Nucleic Acids Research* 36.suppl1 (2007), pp. D402–D408. ISSN: 0305-1048. DOI: 10.1093/nar/gkm957. URL: https://doi.org/10.1093/nar/gkm957.

[157] M. Sud et al. “LMSD: LIPID MAPS structure database”. In: *Nucleic Acids Res* 35.Database issue (2007), pp. D527–32. ISSN: 0305-1048 (Print) 0305-1048. DOI: 10.1093/nar/gkl838.

[158] F. Buchel et al. "Path2Models: large-scale generation of computational models from biochemical pathway maps". In: *BMC Syst Biol* 7 (2013), p. 116. ISSN: 1752-0509 (Electronic) 1752-0509 (Linking). DOI: 10.1186/1752-0509-7-116. URL: https://www.ncbi.nlm.nih.gov/pubmed/24180668.

[159] J. Gebauer et al. "A Genome-Scale Database and Reconstruction of Caenorhabditis elegans Metabolism". In: *Cell Syst* 2.5 (2016), pp. 312–22. ISSN: 2405-4712 (Print) 2405-4712 (Linking). DOI: 10.1016/j.cels.2016.04.017. URL: https://www.ncbi.nlm.nih.gov/pubmed/27211858.

[160] L. S. Yilmaz and A. J. Walhout. "A Caenorhabditis elegans Genome-Scale Metabolic Network Model". In: *Cell Syst* 2.5 (2016), pp. 297–311. ISSN: 2405-4712 (Print) 2405-4712 (Linking). DOI: 10.1016/j.cels.2016.04.012. URL: https://www.ncbi.nlm.nih.gov/pubmed/27211857.

[161] Janna Hastings et al. "WormJam: A consensus C. elegans Metabolic Reconstruction and Metabolomics Community and Workshop Series". In: *Worm* 6.2 (2017), e1373939. ISSN: 2162-4046 2162-4054. DOI: 10.1080/21624054.2017.1373939. URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5612476/.

[162] Michael Witting. "Suggestions for Standardized Identifiers for Fatty Acyl Compounds in Genome Scale Metabolic Models and Their Application to the WormJam Caenorhabditis elegans Model". In: *Metabolites* 10.4 (2020), p. 130. ISSN: 2218-1989. URL: https://www.mdpi.com/2218-1989/10/4/130.

[163] Michael Witting et al. "Modeling Meets Metabolomics—The WormJam Consensus Model as Basis for Metabolic Studies in the Model Organism Caenorhabditis elegans". In: *Frontiers in Molecular Biosciences* 5.96 (2018). ISSN: 2296-889X. DOI: 10.3389/fmolb.2018.00096. URL: https://www.frontiersin.org/article/10.3389/fmolb.2018.00096.

[164] Arnald Alonso, Sara Marsal, and Antonio Julià. "Analytical Methods in Untargeted Metabolomics: State of the Art in 2015". In: *Frontiers in Bioengineering and Biotechnology* 3.23 (2015). ISSN: 2296-4185. DOI: 10.3389/fbioe.2015.00023. URL: https://www.frontiersin.org/article/10.3389/fbioe.2015.00023.

[165] Jan Krumsiek et al. "Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data". In: *BMC Systems Biology* 5.1 (2011), p. 21. ISSN: 1752-0509. DOI: 10.1186/1752-0509-5-21. URL: https://doi.org/10.1186/1752-0509-5-21.

[166] J. P. N. Hattwell et al. "Using Genome-Scale Metabolic Networks for Analysis, Visualization, and Integration of Targeted Metabolomics Data". In: *Methods Mol Biol* 2104 (2020), pp. 361–386. ISSN: 1064-3745. DOI: 10.1007/978-1-0716-0239-3$_$18.

[167] Lutfu Safak Yilmaz et al. "Modeling tissue-relevant Caenorhabditis elegans metabolism at network, pathway, reaction, and metabolite levels". In: *Molecular Systems Biology* 16.10 (2020), e9649. ISSN: 1744-4292. DOI: https://doi.org/10.15252/msb.20209649. URL: https://www.embopress.org/doi/abs/10.15252/msb.20209649.

[168] Masaharu Uno and Eisuke Nishida. "Lifespan-regulating genes in C. elegans". In: *npj Aging and Mechanisms of Disease* 2.1 (2016), p. 16010. ISSN: 2056-3973. DOI: 10.1038/npjamd.2016.10. URL: https://doi.org/10.1038/npjamd.2016.10.

[169] Jeevan K. Prasain et al. "Comparative Lipidomics of Caenorhabditis elegans Metabolic Disease Models by SWATH Non-Targeted Tandem Mass Spectrometry". In: *Metabolites* 5.4 (2015), pp. 677–696. ISSN: 2218-1989. URL: https://www.mdpi.com/2218-1989/5/4/677.

[170] Britta Spanier et al. "Comparison of lipidome profiles of Caenorhabditis elegans—results from an interlaboratory ring trial". In: *Metabolomics* 17.3 (2021), p. 25. ISSN: 1573-3890. DOI: 10.1007/s11306-021-01775-6. URL: https://doi.org/10.1007/s11306-021-01775-6.

[171] Michael Klass and David Hirsh. "Non-ageing developmental variant of Caenorhabditis elegans". In: *Nature* 260.5551 (1976), pp. 523–525. ISSN: 1476-4687. DOI: 10.1038/260523a0. URL: https://doi.org/10.1038/260523a0.

[172] Jagan Srinivasan et al. "A Modular Library of Small Molecule Signals Regulates Social Behaviors in Caenorhabditis elegans". In: *PLOS Biology* 10.1 (2012), e1001237. DOI: 10.1371/journal.pbio.1001237. URL: https://doi.org/10.1371/journal.pbio.1001237.

[173] Michael Witting and Philippe Schmitt-Kopplin. "The Caenorhabditis elegans lipidome: A primer for lipid analysis in Caenorhabditis elegans". In: *Archives of Biochemistry and Biophysics* 589 (2016), pp. 27–37. ISSN: 0003-9861. DOI: https://doi.org/10.1016/j.abb.2015.06.003. URL: http://www.sciencedirect.com/science/article/pii/S0003986115002672.

[174] Jennifer L. Watts and John Browse. "Genetic dissection of polyunsaturated fatty acid synthesis in Caenorhabditis elegans". In: *Proceedings of the National Academy of Sciences* 99.9 (2002), pp. 5854–5859. DOI: 10.1073/pnas.092064799. URL: https://www.pnas.org/content/pnas/99/9/5854.full.pdf.

[175] Paula Ann Hutzell, HUTZELL PA, and KRUSBERG LR. "Fatty acid compositions of Caenorhabditis elegans and C. briggsae". In: (1982).

[176] Tamotsu Tanaka et al. "Effects of growth temperature on the fatty acid composition of the free-living nematodeCaenorhabditis elegans". In: *Lipids* 31.11 (1996), pp. 1173–1178. ISSN: 0024-4201.

[177] Kyleann K Brooks, Bin Liang, and Jennifer L Watts. "The influence of bacterial diet on fat storage in C. elegans". In: *PloS one* 4.10 (2009), e7545. ISSN: 1932-6203.

[178] Xun Shi et al. "A Caenorhabditis elegans model for ether lipid biosynthesis and function". In: *Journal of lipid research* 57.2 (2016), pp. 265–275. ISSN: 1539-7262 0022-2275. DOI: 10.1194/jlr.M064808. URL: https://pubmed.ncbi.nlm.nih.gov/26685325%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4727422/.

[179] Jacques Bodennec et al. "Purification of Sphingolipid Classes by Solid-Phase Extraction with Aminopropyl and Weak Cation Exchanger Cartridges". In: *Methods in Enzymology*. Ed. by Alfred H. Merrill and Yusuf A. Hannun. Vol. 312. Academic Press, 2000, pp. 101–114. ISBN: 0076-6879. DOI: https://doi.org/10.1016/S0076-6879(00)12902-7. URL: http://www.sciencedirect.com/science/article/pii/S0076687900129027.

[180] J. T. Hannich, K. Umebayashi, and H. Riezman. "Distribution and functions of sterols and sphingolipids". In: *Cold Spring Harb Perspect Biol* 3.5 (2011). ISSN: 1943-0264 (Electronic) 1943-0264 (Linking). DOI: 10.1101/cshperspect.a004762. URL: https://www.ncbi.nlm.nih.gov/pubmed/21454248.

[181] Sebastian Boland et al. "Phosphorylated glycosphingolipids essential for cholesterol mobilization in Caenorhabditis elegans". In: *Nature Chemical Biology* 13.6 (2017), pp. 647–654. ISSN: 1552-4469. DOI: 10.1038/nchembio.2347. URL: https://doi.org/10.1038/nchembio.2347.

[182] Xue Zhao et al. "Resolving modifications on sphingoid base and N-acyl chain of sphingomyelin lipids in complex lipid extracts". In: *Analytical Chemistry* 92.21 (2020), pp. 14775–14782.

[183] Stefan Gerdt et al. "Isolation, characterization and immunolocalization of phosphorylcholine-substituted glycolipids in developmental stages of Caenorhabditis elegans". In: *European Journal of Biochemistry* 266.3 (1999), pp. 952–963. ISSN: 0014-2956. DOI: 10.1046/j.1432-1327.1999.00937.x. URL: https://febs.onlinelibrary.wiley.com/doi/abs/10.1046/j.1432-1327.1999.00937.x.

[184] David J. Chitwood et al. "Sterol metabolism in the nematodeCaenorhabditis elegans". In: *Lipids* 19.7 (1984), pp. 500–506. ISSN: 0024-4201. DOI: 10.1007/bf02534482. URL: https://aocs.onlinelibrary.wiley.com/doi/abs/10.1007/BF02534482.

[185] M. Witting, U. Schmidt, and H. J. Knolker. "UHPLC-IM-Q-ToFMS analysis of maradolipids, found exclusively in Caenorhabditis elegans dauer larvae". In: *Anal Bioanal Chem* (2021). ISSN: 1618-2650 (Electronic) 1618-2642 (Linking). DOI: 10.1007/s00216-021-03172-3. URL: https://www.ncbi.nlm.nih.gov/pubmed/33575816.

[186] Robin Drechsler et al. "HPLC-Based Mass Spectrometry Characterizes the Phospholipid Alterations in Ether-Linked Lipid Deficiency Models Following Oxidative Stress". In: *PLOS ONE* 11.11 (2016), e0167229. DOI: 10.1371/journal.pone.0167229. URL: https://doi.org/10.1371/journal.pone.0167229.

[187] Lorissa J. Smulan et al. "Cholesterol-Independent SREBP-1 Maturation Is Linked to ARF1 Inactivation". In: *Cell reports* 16.1 (2016), pp. 9–18. ISSN: 2211-1247. DOI: 10.1016/j.celrep.2016.05.086. URL: https://pubmed.ncbi.nlm.nih.gov/27320911%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4934023/.

[188] Yasmine J. Liu et al. "Mitochondrial translation and dynamics synergistically extend lifespan in C. elegans through HLH-30". In: *bioRxiv* (2019), p. 871079. DOI: 10.1101/871079. URL: https://www.biorxiv.org/content/biorxiv/early/2019/12/10/871079.full.pdf.

[189] Blair C. R. Dancy et al. "13C- and 15N-Labeling Strategies Combined with Mass Spectrometry Comprehensively Quantify Phospholipid Dynamics in C. elegans". In: *PLOS ONE* 10.11 (2015), e0141850. DOI: 10.1371/journal.pone.0141850. URL: https://doi.org/10.1371/journal.pone.0141850.

[190] A. Antebi. "Nuclear receptor signal transduction in C. elegans". In: *WormBook* (2015), pp. 1–49. ISSN: 1551-8507 (Electronic) 1551-8507 (Linking). DOI: 10.1895/wormbook.1.64.2. URL: https://www.ncbi.nlm.nih.gov/pubmed/26069085.

[191] R. R. Zwaal et al. "Two neuronal G proteins are involved in chemosensation of the Caenorhabditis elegans Dauer-inducing pheromone". In: *Genetics* 145.3 (1997), pp. 715–727. ISSN: 0016-6731. URL: %3CGo%20to%20ISI%3E://WOS:A1997WK68900017.

[192] M. M. Perez-Jimenez et al. "Steroid hormones sulfatase inactivation extends lifespan and ameliorates age-related diseases". In: *Nat Commun* 12.1 (2021), p. 49. ISSN: 2041-1723 (Electronic) 2041-1723 (Linking). DOI: 10.1038/s41467-020-20269-y. URL: https://www.ncbi.nlm.nih.gov/pubmed/33397961.

[193] Samantha L. Hughes et al. "The Metabolomic Responses of Caenorhabditis elegans to Cadmium Are Largely Independent of Metallothionein Status, but Dominated by Changes in Cystathionine and Phytochelatins". In: *Journal of Proteome Research* 8.7 (2009), pp. 3512–3519. ISSN: 1535-3893. DOI: 10.1021/pr9001806. URL: https://doi.org/10.1021/pr9001806.

[194] Oliver A. H. Jones et al. "Potential New Method of Mixture Effects Testing Using Metabolomics and Caenorhabditis elegans". In: *Journal of Proteome Research* 11.2 (2012), pp. 1446–1453. ISSN: 1535-3893. DOI: 10.1021/pr201142c. URL: https://doi.org/10.1021/pr201142c.

[195] Gita Sudama et al. "Metabolic profiling in Caenorhabditis elegans provides an unbiased approach to investigations of dosage dependent lead toxicity". In: *Metabolomics : Official journal of the Metabolomic Society* 9.1 (2013), pp. 189–201. ISSN: 1573-3882 1573-3890. DOI: 10.1007/s11306-012-0438-0. URL: https://pubmed.ncbi.nlm.nih.gov/23335868%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3548106/.

[196] A. Ray et al. "Mitochondrial dysfunction, oxidative stress, and neurodegeneration elicited by a bacterial metabolite in a C. elegans Parkinson's model". In: *Cell Death Dis* 5 (2014), e984. ISSN: 2041-4889 (Electronic). DOI: 10.1038/cddis.2013.513. URL: https://www.ncbi.nlm.nih.gov/pubmed/24407237.

[197] F. Cioffi, R. H. I. Adam, and K. Broersen. "Molecular Mechanisms and Genetics of Oxidative Stress in Alzheimer's Disease". In: *J Alzheimers Dis* 72.4 (2019), pp. 981–1017. ISSN: 1875-8908 (Electronic) 1387-2877 (Linking). DOI: 10.3233/JAD-190863. URL: https://www.ncbi.nlm.nih.gov/pubmed/31744008.

[198] B. I. Giasson et al. "Oxidative damage linked to neurodegeneration by selective alpha-synuclein nitration in synucleinopathy lesions". In: *Science* 290.5493 (2000), pp. 985–9. ISSN: 0036-8075 (Print) 0036-8075 (Linking). DOI: 10.1126/science.290.5493.985. URL: https://www.ncbi.nlm.nih.gov/pubmed/11062131.

[199] P. O. Helmer et al. "Investigation of cardiolipin oxidation products as a new endpoint for oxidative stress in C. elegans by means of online two-dimensional liquid chromatography and high-resolution mass spectrometry". In: *Free Radic Biol Med* 162 (2021), pp. 216–224. ISSN: 1873-4596 (Electronic) 0891-5849 (Linking). DOI: 10.1016/j.freeradbiomed.2020.10.019. URL: https://www.ncbi.nlm.nih.gov/pubmed/33127566.

[200] Daniel L. Halligan and Peter D. Keightley. "Spontaneous Mutation Accumulation Studies in Evolutionary Genetics". In: *Annual Review of Ecology, Evolution, and Systematics* 40.1 (2009), pp. 151–172. DOI: 10.1146/annurev.ecolsys.39.110707.173437. URL: https://www.annualreviews.org/doi/abs/10.1146/annurev.ecolsys.39.110707.173437.

[201] Cassandra Coburn et al. "Anthranilate Fluorescence Marks a Calcium-Propagated Necrotic Wave That Promotes Organismal Death in C. elegans". In: *PLOS Biology* 11.7 (2013), e1001613. DOI: 10.1371/journal.pbio.1001613. URL: https://doi.org/10.1371/journal.pbio.1001613.

[202] Martijn P. van Iersel et al. "The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services". In: *BMC Bioinformatics* 11.1 (2010), p. 5. ISSN: 1471-2105. DOI: 10.1186/1471-2105-11-5. URL: https://doi.org/10.1186/1471-2105-11-5.

[203] Rajarshi Guha. "Chemical Informatics Functionality in R". In: *Journal of Statistical Software; Vol 1, Issue 5 (2007)* (2007). URL: https://www.jstatsoft.org/v018/i05%20http://dx.doi.org/10.18637/jss.v018.i05.

[204] Gerhard Liebisch et al. "Shorthand notation for lipid structures derived from mass spectrometry". In: *Journal of lipid research* 54.6 (2013), pp. 1523–1530. ISSN: 1539-7262 0022-2275. DOI: 10.1194/jlr.M033506. URL: https://pubmed.ncbi.nlm.nih.gov/23549332%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3646453/.

[205] D. Bensaddek et al. "Micro-proteomics with iterative data analysis: Proteome analysis in C. elegans at the single worm level". In: *Proteomics* 16.3 (2016), pp. 381–92. ISSN: 1615-9861 (Electronic) 1615-9853 (Linking). DOI: 10.1002/pmic.201500264. URL: https://www.ncbi.nlm.nih.gov/pubmed/26552604.

[206] T. W. Harris et al. "WormBase: a modern Model Organism Information Resource". In: *Nucleic Acids Res* 48.D1 (2020), pp. D762–D767. ISSN: 1362-4962 (Electronic) 0305-1048 (Linking). DOI: 10.1093/nar/gkz920. URL: https://www.ncbi.nlm.nih.gov/pubmed/31642470.

[207] J. J. van Deemter, F. J. Zuiderweg, and A. Klinkenberg. "Longitudinal diffusion and resistance to mass transfer as causes of nonideality in chromatography". In: *Chemical Engineering Science* 50 (24 Dec. 1995), pp. 3869–3882. ISSN: 0009-2509. DOI: 10.1016/0009-2509(96)81813-6.

[208] Eva-Maria Harrieder et al. "Current state-of-the-art of separation methods used in LC-MS based metabolomics and lipidomics". In: *Journal of Chromatography B* 1188 (2022), p. 123069. ISSN: 1570-0232. DOI: https://doi.org/10.1016/j.jchromb.2021.123069. URL: https://www.sciencedirect.com/science/article/pii/S157002322100550X.

[209] Andrew J. Alpert. "Hydrophilic-interaction chromatography for the separation of peptides, nucleic acids and other polar compounds". In: *Journal of Chromatography A* 499 (C Jan. 1990), pp. 177–196. ISSN: 0021-9673. DOI: 10.1016/S0021-9673(00)96972-3.

[210] Bogusław Buszewski and Sylwia Noga. *Hydrophilic interaction liquid chromatography (HILIC)-a powerful separation technique.* Jan. 2012. DOI: 10.1007/s00216-011-5308-5.

[211] Andrew J. Alpert. "Effect of salts on retention in hydrophilic interaction chromatography". In: *Journal of Chromatography A* 1538 (Feb. 2018), pp. 45–53. ISSN: 18733778. DOI: 10.1016/j.chroma.2018.01.038.

[212] Haibin Li et al. "A systematic investigation of the effect of sample solvent on peak shape in nano- and microflow hydrophilic interaction liquid chromatography columns". In: *Journal of Chromatography A* 1655 (Oct. 2021), p. 462498. ISSN: 0021-9673. DOI: 10.1016/J.CHROMA.2021.462498.

[213] D.N. Heiger. *High Performance Capillary Electrophoresis: An Introduction.* Hewlett-Packard, 1992. URL: https://books.google.de/books?id=GUUdywAACAAJ.

[214] D. Corradini and L. Sprecacenere. "Dependence of the Electroosmotic Flow in Bare Fused-Silica Capillaries from pH, Ionic Strength and Composition of Electrolyte Solutions Tailored for Protein Capillary Zone Electrophoresis". In: *Chromatographia* 58.9 (2003), pp. 587–596. ISSN: 1612-1112. DOI: 10.1365/s10337-003-0098-3. URL: https://doi.org/10.1365/s10337-003-0098-3.

[215] N. Drouin et al. "Effective mobility as a robust criterion for compound annotation and identification in metabolomics: Toward a mobility-based library". In: *Anal Chim Acta* 1032 (2018), pp. 178–187. ISSN: 0003-2670. DOI: 10.1016/j.aca.2018.05.063.

[216] M. Ariel Geer Wallace and James P. McCord. "Chapter 16 - High-resolution mass spectrometry". In: *Breathborne Biomarkers and the Human Volatilome (Second Edition).* Ed. by Jonathan Beauchamp, Cristina Davis, and Joachim Pleil. Second Edition. Boston: Elsevier, 2020, pp. 253–270. ISBN: 978-0-12-819967-1. DOI: https://doi.org/10.1016/B978-0-12-819967-1.00016-5. URL: https://www.sciencedirect.com/science/article/pii/B9780128199671000165.

[217] Petrus W. Lindenburg et al. *Developments in Interfacing Designs for CE–MS: Towards Enabling Tools for Proteomics and Metabolomics.* Mar. 2015. DOI: 10.1007/s10337-014-2795-5.

[218] Richard D. Smith, Charles J. Barinaga, and Harold R. Udseth. "Improved Electrospray Ionization Interface for Capillary Zone Electrophoresis-Mass Spectrometry". In: *Analytical Chemistry* 60 (18 1988). ISSN: 15206882. DOI: 10.1021/ac00169a022.

[219] Johannes Schlecht et al. "nanoCEasy: An Easy, Flexible, and Robust Nanoflow Sheath Liquid Capillary Electrophoresis-Mass Spectrometry Interface Based on 3D Printed Parts". In: *Analytical Chemistry* 93 (44 2021). ISSN: 15206882. DOI: 10.1021/acs.analchem.1c03213.

[220] Stephanie L. Collins et al. "Current Challenges and Recent Developments in Mass Spectrometry–Based Metabolomics". In: *Annual Review of Analytical Chemistry* 14.1 (2021). PMID: 34314226, pp. 467–487. DOI: 10.1146/annurev-anchem-091620-015205. eprint: https://doi.org/10.1146/annurev-anchem-091620-015205. URL: https://doi.org/10.1146/annurev-anchem-091620-015205.

[221] Hiroshi Tsugawa et al. "A cheminformatics approach to characterize metabolomes in stable-isotope-labeled organisms". In: *Nature Methods* 16 (4 2019). ISSN: 15487105. DOI: 10.1038/s41592-019-0358-2.

[222] Vladimir Shulaev. "Metabolomics technology and bioinformatics". In: *Briefings in Bioinformatics* 7.2 (May 2006), pp. 128–139. ISSN: 1467-5463. DOI: 10.1093/bib/bbl012. eprint: https://academic.oup.com/bib/article-pdf/7/2/128/868824/bbl012.pdf. URL: https://doi.org/10.1093/bib/bbl012.

[223] Ibrahim Karaman, Rui Climaco Pinto, and Gonçalo Graça. "Chapter Eight - Metabolomics Data Preprocessing: From Raw Data to Features for Statistical Analysis". In: *Data Analysis for Omic Sciences: Methods and Applications*. Ed. by Joaquim Jaumot, Carmen Bedia, and Romà Tauler. Vol. 82. Comprehensive Analytical Chemistry. Elsevier, 2018, pp. 197–225. DOI: https://doi.org/10.1016/bs.coac.2018.08.003. URL: https://www.sciencedirect.com/science/article/pii/S0166526X18300746.

[224] Colin A. Smith et al. "XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification". In: *Analytical Chemistry* 78.3 (2006), pp. 779–787. ISSN: 0003-2700. DOI: 10.1021/ac051437y. URL: https://doi.org/10.1021/ac051437y.

[225] Tomáš Pluskal et al. "MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data". In: *BMC Bioinformatics* 11 (2010). ISSN: 14712105. DOI: 10.1186/1471-2105-11-395.

[226] Kendra J. Adams et al. "Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics". In: *Journal of Proteome Research* 19 (4 2020). ISSN: 15353907. DOI: 10.1021/acs.jproteome.9b00640.

[227] Hannes L. Röst et al. *OpenMS: A flexible open-source software platform for mass spectrometry data analysis*. 2016. DOI: 10.1038/nmeth.3959.

[228] Matthew C. Chambers et al. "A cross-platform toolkit for mass spectrometry and proteomics". In: *Nature Biotechnology* 30.10 (2012), pp. 918–920. ISSN: 1546-1696. DOI: 10.1038/nbt.2377. URL: https://doi.org/10.1038/nbt.2377.

[229] Yang Chen, En-Min Li, and Li-Yan Xu. "Guide to Metabolomics Analysis: A Bioinformatics Workflow". In: *Metabolites* 12.4 (2022). ISSN: 2218-1989. DOI: 10.3390/metabo12040357. URL: https://www.mdpi.com/2218-1989/12/4/357.

[230] Carlos Guijas et al. "METLIN: A Technology Platform for Identifying Knowns and Unknowns". In: *Analytical Chemistry* 90.5 (2018). PMID: 29381867, pp. 3156–3164. DOI: 10.1021/acs.analchem.7b04424. eprint: https://doi.org/10.1021/acs.analchem.7b04424. URL: https://doi.org/10.1021/acs.analchem.7b04424.

[231] Mingxun Wang et al. *Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking*. 2016. DOI: 10.1038/nbt.3597.

[232] Royston Goodacre et al. "Proposed minimum reporting standards for data analysis in metabolomics". In: *Metabolomics* 3 (3 2007). ISSN: 15733882. DOI: 10.1007/s11306-007-0081-3.

[233] Gerhard Liebisch et al. "Update on LIPID MAPS classification, nomenclature, and shorthand notation for MS-derived lipid structures". In: *Journal of Lipid Research* 61.12 (2020), pp. 1539–1555. ISSN: 0022-2275. DOI: https://doi.org/10.1194/jlr.S120001025. URL: https://www.sciencedirect.com/science/article/pii/S0022227520600177.

[234] Sebastian Böcker et al. "SIRIUS: decomposing isotope patterns for metabolite identification†". In: *Bioinformatics* 25.2 (Nov. 2008), pp. 218–224. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btn603. eprint: https://academic.oup.com/bioinformatics/article-pdf/25/2/218/48983335/bioinformatics_25_2_218.pdf. URL: https://doi.org/10.1093/bioinformatics/btn603.

[235] Sebastian Böcker and Florian Rasche. "Towards de novo identification of metabolites by analyzing tandem mass spectra". In: *Bioinformatics* 24.16 (Aug. 2008), pp. i49–i55. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btn270. eprint: https://academic.oup.com/bioinformatics/article-pdf/24/16/i49/49053673/bioinformatics_24_16_i49.pdf. URL: https://doi.org/10.1093/bioinformatics/btn270.

[236] K. Dührkop et al. “Searching molecular structure databases with tandem mass spectra using CSI:FingerID”. In: *Proc Natl Acad Sci U S A* 112.41 (2015), pp. 12580–5. ISSN: 0027-8424 (Print) 0027-8424. DOI: 10.1073/pnas.1509788112.

[237] Sebastian Wolf et al. “In silico fragmentation for computer assisted identification of metabolite mass spectra”. In: *BMC Bioinformatics* 11 (2010). ISSN: 14712105. DOI: 10.1186/1471-2105-11-148.

[238] Christoph Ruttkies et al. “MetFrag relaunched: Incorporating strategies beyond in silico fragmentation”. In: *Journal of Cheminformatics* 8 (1 2016). ISSN: 17582946. DOI: 10.1186/s13321-016-0115-9.

[239] Clément Frainay and Fabien Jourdan. “Computational methods to identify metabolic sub-networks based on metabolomic profiles”. In: *Briefings in Bioinformatics* 18.1 (Jan. 2016), pp. 43–56. ISSN: 1467-5463. DOI: 10.1093/bib/bbv115. eprint: https://academic.oup.com/bib/article-pdf/18/1/43/25408409/bbv115.pdf. URL: https://doi.org/10.1093/bib/bbv115.

[240] Douglas B. Kell and Royston Goodacre. *Metabolomics and systems pharmacology: Why and how to model the human metabolic network for drug discovery*. 2014. DOI: 10.1016/j.drudis.2013.07.014.

[241] Martin Loos and Heinz Singer. “Nontargeted homologue series extraction from hyphenated high resolution mass spectrometry data”. In: *Journal of Cheminformatics* 9 (1 2017). ISSN: 17582946. DOI: 10.1186/s13321-017-0197-z.

[242] Robin Schmid et al. “Ion identity molecular networking for mass spectrometry-based metabolomics in the GNPS environment”. In: *Nature Communications* 12 (1 2021). ISSN: 20411723. DOI: 10.1038/s41467-021-23953-9.

[243] Ludovic Cottret and Fabien Jourdan. *Graph methods for the investigation of metabolic networks in parasitology*. 2010. DOI: 10.1017/S0031182010000363.

[244] Clément Frainay et al. “MetaboRank: Network-based recommendation system to interpret and enrich metabolomics results”. In: *Bioinformatics* 35 (2 2019), pp. 274–283. ISSN: 14602059. DOI: 10.1093/bioinformatics/bty577.

[245] Karoline Faust et al. “Pathway discovery in metabolic networks by subgraph extraction”. In: *Bioinformatics* 26 (9 2010). ISSN: 13674803. DOI: 10.1093/bioinformatics/btq105.

[246] Jan Schellenberger et al. “Quantitative prediction of cellular metabolism with constraint-based models: The COBRA Toolbox v2.0”. In: *Nature Protocols* 6 (9 2011). ISSN: 17542189. DOI: 10.1038/nprot.2011.308.

[247] Zoran Nikoloski, Richard Perez-Storey, and Lee J. Sweetlove. “Inference and Prediction of Metabolic Network Fluxes”. In: *Plant Physiology* 169.3 (Sept. 2015), pp. 1443–1455. ISSN: 0032-0889. DOI: 10.1104/pp.15.01082. eprint: https://academic.oup.com/plphys/article-pdf/169/3/1443/38117089/plphys_v169_3_1443.pdf. URL: https://doi.org/10.1104/pp.15.01082.

[248] Karthik Raman and Nagasuma Chandra. “Flux balance analysis of biological systems: applications and challenges”. In: *Briefings in Bioinformatics* 10.4 (Mar. 2009), pp. 435–449. ISSN: 1467-5463. DOI: 10.1093/bib/bbp011. eprint: https://academic.oup.com/bib/article-pdf/10/4/435/841894/bbp011.pdf. URL: https://doi.org/10.1093/bib/bbp011.

[249] Adam Amara et al. *Networks and Graphs Discovery in Metabolomics Data Analysis and Interpretation*. 2022. DOI: 10.3389/fmolb.2022.841373.

[250] Franco Moritz et al. “Characterization of poplar metabotypes via mass difference enrichment analysis”. In: *Plant, Cell & Environment* 40.7 (2017), pp. 1057–1073. DOI: https://doi.org/10.1111/pce.12878. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/pce.12878. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/pce.12878.

[251] Fabien Jourdan et al. "MetaNetter: inference and visualization of high-resolution metabolomic networks". In: *Bioinformatics* 24.1 (Nov. 2007), pp. 143–145. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btm536. eprint: https://academic.oup.com/bioinformatics/article-pdf/24/1/143/49044008/bioinformatics_24_1_143.pdf. URL: https://doi.org/10.1093/bioinformatics/btm536.

[252] Minoru Kanehisa et al. "KEGG: new perspectives on genomes, pathways, diseases and drugs". In: *Nucleic Acids Research* 45.D1 (Nov. 2016), pp. D353–D361. ISSN: 0305-1048. DOI: 10.1093/nar/gkw1092. eprint: https://academic.oup.com/nar/article-pdf/45/D1/D353/8846820/gkw1092.pdf. URL: https://doi.org/10.1093/nar/gkw1092.

[253] James G. Jeffryes et al. "MINEs: Open access databases of computationally predicted enzyme promiscuity products for untargeted metabolomics". In: *Journal of Cheminformatics* 7 (1 2015). ISSN: 17582946. DOI: 10.1186/s13321-015-0087-1.

[254] Noushin Hadadi et al. "ATLAS of Biochemistry: A Repository of All Possible Biochemical Reactions for Synthetic Biology and Metabolic Engineering Studies". In: *ACS Synthetic Biology* 5 (10 2016). ISSN: 21615063. DOI: 10.1021/acssynbio.6b00054.

[255] Sébastien Moretti et al. "MetaNetX/MNXref: unified namespace for metabolites and biochemical reactions in the context of metabolic models". In: *Nucleic Acids Research* 49.D1 (Nov. 2020), pp. D570–D574. ISSN: 0305-1048. DOI: 10.1093/nar/gkaa992. eprint: https://academic.oup.com/nar/article-pdf/49/D1/D570/35364710/gkaa992.pdf. URL: https://doi.org/10.1093/nar/gkaa992.

[256] Larry A. Lerno, J. Bruce German, and Carlito B. Lebrilla. "Method for the identification of lipid classes based on referenced kendrick mass analysis". In: *Analytical Chemistry* 82 (10 2010). ISSN: 00032700. DOI: 10.1021/ac100556g.

[257] Kris Morreel et al. "Systematic structural characterization of metabolites in Arabidopsis via candidate substrate-product pair networks". In: *Plant Cell* (2014). ISSN: 1532298X. DOI: 10.1105/tpc.113.122242.

[258] Hannes Doerfler et al. "mzGroupAnalyzer-predicting pathways and novel chemical structures from untargeted high-throughput metabolomics data". In: *PLoS ONE* 9 (5 2014). ISSN: 19326203. DOI: 10.1371/journal.pone.0096188.

[259] Thomas Naake and Alisdair R. Fernie. "MetNet: Metabolite Network Prediction from High-Resolution Mass Spectrometry Data in R Aiding Metabolite Annotation". In: *Analytical Chemistry* 91 (3 2019). ISSN: 15206882. DOI: 10.1021/acs.analchem.8b04096.

[260] Jeramie Watrous et al. "Mass spectral molecular networking of living microbial colonies". In: *Proceedings of the National Academy of Sciences* 109.26 (2012), E1743–E1752. DOI: 10.1073/pnas.1203689109. eprint: https://www.pnas.org/doi/pdf/10.1073/pnas.1203689109. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1203689109.

[261] Louis Félix Nothias et al. "Feature-based molecular networking in the GNPS analysis environment". In: *Nature Methods* 17 (9 2020). ISSN: 15487105. DOI: 10.1038/s41592-020-0933-6.

[262] Florian Huber et al. "Spec2Vec: Improved mass spectral similarity scoring through learning of structural relationships". In: *PLoS Computational Biology* 17 (2 2021). ISSN: 15537358. DOI: 10.1371/JOURNAL.PCBI.1008724.

[263] Shipei Xing et al. "Retrieving and Utilizing Hypothetical Neutral Losses from Tandem Mass Spectra for Spectral Similarity Analysis and Unknown Metabolite Annotation". In: *Analytical Chemistry* 92 (21 2020). ISSN: 15206882. DOI: 10.1021/acs.analchem.0c02521.

[264] Chaojie Wang, Jin Du, and Xiaodan Fan. "High-dimensional correlation matrix estimation for general continuous data with Bagging technique". In: *Machine Learning* 111 (8 2022). ISSN: 15730565. DOI: 10.1007/s10994-022-06138-3.

[265] Leonardo Perez De Souza et al. "Network based strategies in metabolomics data analysis and interpretation: from molecular networking to biological interpretation". In: *Expert Review of Proteomics* (2020), pp. 1744–8387. ISSN: 1478-9450. DOI: 10.1080/14789450.2020.1766975. URL: https://doi.org/10.1080/14789450.2020.1766975.

[266] Peter Langfelder and Steve Horvath. "WGCNA: An R package for weighted correlation network analysis". In: *BMC Bioinformatics* 9 (2008). ISSN: 14712105. DOI: 10.1186/1471-2105-9-559.

[267] Atsushi Fukushima. "DiffCorr: An R package to analyze and visualize differential correlations in biological networks". In: *Gene* 518 (1 2013). ISSN: 03781119. DOI: 10.1016/j.gene.2012.11.028.

[268] Ph Schmitt-Kopplin et al. "Capillary electrophoresis for the simultaneous separation of selected carboxylated carbohydrates and their related 1, 4-lactones". In: *Journal of Chromatography A* 807.1 (1998), pp. 89–100. ISSN: 0021-9673.

[269] Philippe Schmitt-Kopplin et al. "Mobility distribution of synthetic and natural polyelectrolytes with capillary zone electrophoresis". In: *Journal of AOAC International* 82.6 (1999), pp. 1594–1603. ISSN: 1060-3271.

[270] Agnes Fekete et al. "Development of a capillary electrophoretic method for the analysis of low-molecular-weight amines from metal working fluid aerosols and ambient air". In: *Electrophoresis* 27.5-6 (2006), pp. 1237–1247. ISSN: 0173-0835.

[271] L. Martens et al. "mzML–a community standard for mass spectrometry data". In: *Mol Cell Proteomics* 10.1 (2011), R110.000133. ISSN: 1535-9476 (Print) 1535-9476. DOI: 10.1074/mcp.R110.000133.

[272] Johannes Rainer et al. "A Modular and Expandable Ecosystem for Metabolomics Data Annotation in R". In: 12.2 (2022), p. 173. ISSN: 2218-1989. URL: https://www.mdpi.com/2218-1989/12/2/173.

[273] Jan Stanstrup et al. "The metaRbolomics Toolbox in Bioconductor and beyond". In: 9.10 (2019), p. 200. ISSN: 2218-1989. URL: https://www.mdpi.com/2218-1989/9/10/200.

[274] S. Codesido et al. "New insights into the conversion of electropherograms to the effective electrophoretic mobility scale". In: *Electrophoresis* 42.19 (2021), pp. 1875–1884. ISSN: 1522-2683 (Electronic) 0173-0835 (Linking). DOI: 10.1002/elps.202000333. URL: https://www.ncbi.nlm.nih.gov/pubmed/34216494.

[275] Laurent Gatto, Sebastian Gibb, and Johannes Rainer. "MSnbase, Efficient and Elegant R-Based Processing and Visualization of Raw Mass Spectrometry Data". In: *Journal of Proteome Research* 20.1 (2021), pp. 1063–1069. ISSN: 1535-3893. DOI: 10.1021/acs.jproteome.0c00313. URL: https://doi.org/10.1021/acs.jproteome.0c00313.

[276] Laurent Gatto and Kathryn S. Lilley. "MSnbase-an R/Bioconductor package for isobaric tagged mass spectrometry data visualization, processing and quantitation". In: *Bioinformatics* 28.2 (2011), pp. 288–289. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btr645%JBioinformatics. URL: https://doi.org/10.1093/bioinformatics/btr645.

[277] Colin A. Smith et al. "XCMS: Processing Mass Spectrometry Data for Metabolite Profiling Using Nonlinear Peak Alignment, Matching, and Identification". In: *Analytical Chemistry* 78.3 (2006), pp. 779–787. ISSN: 0003-2700. DOI: 10.1021/ac051437y. URL: https://doi.org/10.1021/ac051437y.

[278] S Codesido et al. "A Mature ROMANCE: A Matter of Quantity and How Two Can Be Better than One". In: *ChemRxiv* (2020). DOI: 10.26434/chemrxiv.12287048.v1.

[279] Jennifer Wild et al. "Metabolomics for improved treatment monitoring of phenylketonuria: urinary biomarkers for non-invasive assessment of dietary adherence and nutritional deficiencies". In: *Analyst* 144.22 (2019), pp. 6595–6608. ISSN: 0003-2654. DOI: 10.1039/C9AN01642B. URL: http://dx.doi.org/10.1039/C9AN01642B.

[280] Wei Zhang et al. "Utility of sheathless capillary electrophoresis-mass spectrometry for metabolic profiling of limited sample amounts". In: *Journal of Chromatography B* 1105 (2019), pp. 10–14. ISSN: 1570-0232. DOI: https://doi.org/10.1016/j.jchromb.2018.12.004. URL: https://www.sciencedirect.com/science/article/pii/S1570023218314983.

[281] Miguel Fernández-García et al. "Comprehensive Examination of the Mouse Lung Metabolome Following Mycobacterium tuberculosis Infection Using a Multiplatform Mass Spectrometry Approach". In: *Journal of Proteome Research* 19.5 (2020). doi: 10.1021/acs.jproteome.9b00868, pp. 2053–2070. ISSN: 1535-3893. DOI: 10.1021/acs.jproteome.9b00868. URL: https://doi.org/10.1021/acs.jproteome.9b00868.

[282] L. Salzer and M. Witting. "Quo Vadis Caenorhabditis elegans Metabolomics-A Review of Current Methods and Applications to Explore Metabolism in the Nematode". In: *Metabolites* 11.5 (2021). ISSN: 2218-1989 (Print) 2218-1989 (Linking). DOI: 10.3390/metabo11050284. URL: https://www.ncbi.nlm.nih.gov/pubmed/33947148.

[283] Safa Beydoun et al. "An alternative food source for metabolism and longevity studies in Caenorhabditis elegans". In: *Communications Biology* 4.1 (2021), p. 258. ISSN: 2399-3642. DOI: 10.1038/s42003-021-01764-4. URL: https://doi.org/10.1038/s42003-021-01764-4.

[284] A. W. Gao et al. "Natural genetic variation in C. elegans identified genomic loci controlling metabolite levels". In: *Genome Res* 28.9 (2018), pp. 1296–1308. ISSN: 1088-9051 (Print) 1088-9051. DOI: 10.1101/gr.232322.117.

[285] Marte Molenaars et al. "Metabolomics and lipidomics in Caenorhabditis elegans using a single-sample preparation". In: *Disease Models & Mechanisms* 14.4 (2021). ISSN: 1754-8403. DOI: 10.1242/dmm.047746. URL: https://doi.org/10.1242/dmm.047746.

[286] Nicolas Drouin et al. "Evaluation of ion mobility in capillary electrophoresis coupled to mass spectrometry for the identification in metabolomics". In: *ELECTROPHORESIS* 42.4 (2021), pp. 342–349. ISSN: 0173-0835. DOI: https://doi.org/10.1002/elps.202000120. URL: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/elps.202000120.

[287] Jordy J. Hsiao et al. "Improved LC/MS Methods for the Analysis of Metal-Sensitive Analytes Using Medronic Acid as a Mobile Phase Additive". In: *Analytical Chemistry* 90.15 (2018), pp. 9457–9464. ISSN: 0003-2700. DOI: 10.1021/acs.analchem.8b02100. URL: https://doi.org/10.1021/acs.analchem.8b02100.

[288] L. Salzer, M. Witting, and P. Schmitt-Kopplin. "MobilityTransformR: an R package for effective mobility transformation of CE-MS data". In: *Bioinformatics* 38.16 (2022), pp. 4044–4045. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/btac441.

[289] Johannes Rainer et al. "A Modular and Expandable Ecosystem for Metabolomics Data Annotation in R". In: *Metabolites* 12.2 (2022), p. 173. ISSN: 2218-1989. URL: https://www.mdpi.com/2218-1989/12/2/173.

[290] Kai Dührkop et al. "Searching molecular structure databases with tandem mass spectra using CSI:FingerID". In: *Proceedings of the National Academy of Sciences* 112.41 (2015), pp. 12580–12585. DOI: doi:10.1073/pnas.1509788112. URL: https://www.pnas.org/doi/abs/10.1073/pnas.1509788112.

[291] Martin A. Hoffmann et al. "High-confidence structural annotation of metabolites absent from spectral libraries". In: *Nature Biotechnology* 40.3 (2022), pp. 411–421. ISSN: 1546-1696. DOI: 10.1038/s41587-021-01045-9. URL: https://doi.org/10.1038/s41587-021-01045-9.

[292] Santiago Codesido et al. "New insights into the conversion of electropherograms to the effective electrophoretic mobility scale". In: *ELECTROPHORESIS* 42.19 (2021), pp. 1875–1884. ISSN: 0173-0835. DOI: https://doi.org/10.1002/elps.202000333. URL: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/elps.202000333.

[293] Miranda G. M. Kok, Govert W. Somsen, and Gerhardus J. de Jong. "Comparison of capillary electrophoresis–mass spectrometry and hydrophilic interaction chromatography–mass spectrometry for anionic metabolic profiling of urine". In: *Talanta* 132 (2015), pp. 1–7. ISSN: 0039-9140. DOI: https://doi.org/10.1016/j.talanta.2014.08.047. URL: https://www.sciencedirect.com/science/article/pii/S0039914014007279.

[294] Jeanneth Vásconez et al. "Comparison of capillary electrophoresis and zwitterionic-hydrophilic interaction capillary liquid chromatography with ultraviolet and mass spectrometry detection for the analysis of microRNA biomarkers". In: *Talanta* 219 (2020), p. 121263. ISSN: 0039-9140. DOI: https://doi.org/10.1016/j.talanta.2020.121263. URL: https://www.sciencedirect.com/science/article/pii/S0039914020305543.

[295] Philippe Schmitt-Kopplin et al. "Mobility distribution of synthetic and natural polyelectrolytes with capillary zone electrophoresis". In: *Journal of AOAC International* 82.6 (1999), pp. 1594–1603. ISSN: 1060-3271.

[296] Morteza Gholi Khaledi. *High-performance capillary electrophoresis : theory, techniques, and applications.* Chemical analysis. New York: Wiley, 1998, xxxii, 1047 p. ISBN: 0471148512 (cloth alk. paper). URL: Contributor%20biographical%20information%20http://www.loc.gov/catdir/bios/wiley042/97035765.html%20Publisher%20description%20http://www.loc.gov/catdir/description/wiley033/97035765.html%20Table%20of%20Contents%20http://www.loc.gov/catdir/toc/onix04/97035765.html.

[297] Isabelle Kohler et al. "Hydrophilic interaction chromatography – mass spectrometry for metabolomics and proteomics: state-of-the-art and current trends". In: *Microchemical Journal* 175 (2022), p. 106986. ISSN: 0026-265X. DOI: https://doi.org/10.1016/j.microc.2021.106986. URL: https://www.sciencedirect.com/science/article/pii/S0026265X21010729.

[298] Jörg Martin Büscher et al. "Cross-Platform Comparison of Methods for Quantitative Metabolomics of Primary Metabolism". In: *Analytical Chemistry* 81.6 (2009). doi: 10.1021/ac8022857, pp. 2135–2143. ISSN: 0003-2700. DOI: 10.1021/ac8022857. URL: https://doi.org/10.1021/ac8022857.

[299] Warwick B. Dunn et al. "Mass appeal: metabolite identification in mass spectrometry-focused untargeted metabolomics". In: *Metabolomics* 9.1 (2013), pp. 44–66. ISSN: 1573-3890. DOI: 10.1007/s11306-012-0434-4. URL: https://doi.org/10.1007/s11306-012-0434-4.

[300] Graham Mullard et al. "A new strategy for MS/MS data acquisition applying multiple data dependent experiments on Orbitrap mass spectrometers in non-targeted metabolomic applications". In: *Metabolomics* 11.5 (2015), pp. 1068–1080. ISSN: 1573-3890. DOI: 10.1007/s11306-014-0763-6. URL: https://doi.org/10.1007/s11306-014-0763-6.

[301] E. C. Rickard, M. M. Strohl, and R. G. Nielsen. "Correlation of electrophoretic mobilities from capillary electrophoresis with physicochemical properties of proteins and peptides". In: *Analytical Biochemistry* 197.1 (1991), pp. 197–207. ISSN: 0003-2697. DOI: https://doi.org/10.1016/0003-2697(91)90379-8. URL: https://www.sciencedirect.com/science/article/pii/0003269791903798.

[302] S. Johnson and S. I. Imai. "NAD (+) biosynthesis, aging, and disease". In: *F1000Res* 7 (2018). 2046-1402 Johnson, Sean Orcid: 0000-0003-4051-2804 Imai, Shin-Ichiro R01 AG037457/AG/NIA NIH HHS/United States R01 AG047902/AG/NIA NIH HHS/United States Journal Article Review England 2018/05/11 F1000Res. 2018 Feb 1;7:132. doi: 10.12688/f1000research.12120.1. eCollection 2018., p. 132. ISSN: 2046-1402 (Print) 2046-1402. DOI: 10.12688/f1000research.12120.1.

[303] J. R. Revollo, A. A. Grimm, and S. Imai. "The NAD biosynthesis pathway mediated by nicotinamide phosphoribosyltransferase regulates Sir2 activity in mammalian cells". In: *J Biol Chem* 279.49 (2004). Revollo, Javier R Grimm, Andrew A Imai, Shin-ichiro Journal Article Research Support, Non-U.S. Gov't United States 2004/09/24 J Biol Chem. 2004 Dec 3;279(49):50754-63. doi: 10.1074/jbc.M408388200. Epub 2004 Sep 20., pp. 50754–63. ISSN: 0021-9258 (Print) 0021-9258. DOI: 10.1074/jbc.M408388200.

[304] Sarah K. Davies, Jacob G. Bundy, and Armand M. Leroi. "Metabolic Youth in Middle Age: Predicting Aging in Caenorhabditis elegans Using Metabolomics". In: *Journal of Proteome Research* 14.11 (2015), pp. 4603–4609. ISSN: 1535-3893. DOI: 10.1021/acs.jproteome.5b00442. URL: https://doi.org/10.1021/acs.jproteome.5b00442.

[305] Yoko Honda, Masashi Tanaka, and Shuji Honda. "Trehalose extends longevity in the nematode Caenorhabditis elegans". In: *Aging Cell* 9.4 (2010), pp. 558–569. ISSN: 1474-9718. DOI: https://doi.org/10.1111/j.1474-9726.2010.00582.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1474-9726.2010.00582.x.

[306] Geert Depuydt et al. "Increased Protein Stability and Decreased Protein Turnover in the Caenorhabditis elegans Ins/IGF-1 daf-2 Mutant". In: *The Journals of Gerontology: Series A* 71.12 (2016), pp. 1553–1559. ISSN: 1079-5006. DOI: 10.1093/gerona/glv221. URL: https://doi.org/10.1093/gerona/glv221.

[307] Nae-Cherng Yang, Yu-Hung Cho, and Inn Lee. "The Lifespan Extension Ability of Nicotinic Acid Depends on Whether the Intracellular NAD+ Level Is Lower than the Sirtuin-Saturating Concentrations". In: *International Journal of Molecular Sciences* 21.1 (2020), p. 142. ISSN: 1422-0067. URL: https://www.mdpi.com/1422-0067/21/1/142.

[308] O. Höcker, C. Montealegre, and C. Neusüß. "Characterization of a nanoflow sheath liquid interface and comparison to a sheath liquid and a sheathless porous-tip interface for CE-ESI-MS in positive and negative ionization". In: *Anal Bioanal Chem* 410.21 (2018). 1618-2650 Höcker, Oliver Montealegre, Cristina Neusüß, Christian Comparative Study Journal Article Germany 2018/06/27 Anal Bioanal Chem. 2018 Aug;410(21):5265-5275. doi: 10.1007/s00216-018-1179-3. Epub 2018 Jun 26., pp. 5265–5275. ISSN: 1618-2642. DOI: 10.1007/s00216-018-1179-3.

[309] Mehdi Moini. "Simplifying CEMS Operation. 2. Interfacing Low-Flow Separation Techniques to Mass Spectrometry Using a Porous Tip". In: *Analytical Chemistry* 79.11 (2007). doi: 10.1021/ac0704560, pp. 4241–4246. ISSN: 0003-2700. DOI: 10.1021/ac0704560. URL: https://doi.org/10.1021/ac0704560.

[310] N. Swainston et al. "Recon 2.2: from reconstruction to model of human metabolism". In: *Metabolomics* 12 (2016), p. 109. ISSN: 1573-3882 (Print) 1573-3882. DOI: 10.1007/s11306-016-1051-4.

[311] J. L. Robinson et al. "An atlas of human metabolism". In: *Sci Signal* 13.624 (2020). ISSN: 1945-0877 (Print) 1945-0877. DOI: 10.1126/scisignal.aaz1482.

[312] R. Breitling et al. "Ab initio prediction of metabolic networks using Fourier transform mass spectrometry data". In: *Metabolomics* 2.3 (2006), pp. 155–164. ISSN: 1573-3882 (Print) 1573-3882. DOI: 10.1007/s11306-006-0029-z.

[313] D. Tziotis, N. Hertkorn, and P. Schmitt-Kopplin. "Kendrick-analogous network visualisation of ion cyclotron resonance Fourier transform mass spectra: improved options for the assignment of elemental compositions and the classification of organic molecular complexity". In: *Eur J Mass Spectrom (Chichester)* 17.4 (2011), pp. 415–21. ISSN: 1469-0667 (Print) 1469-0667. DOI: 10.1255/ejms.1135.

[314] K. E. V. Burgess et al. "MetaNetter 2: A Cytoscape plugin for ab initio network analysis and metabolite feature classification". In: *J Chromatogr B Analyt Technol Biomed Life Sci* 1071 (2017), pp. 68–74. ISSN: 1570-0232 (Print) 1570-0232. DOI: 10.1016/j.jchromb.2017.08.015.

[315] A. B. Artyukhin et al. "Starvation-induced collective behavior in C. elegans". In: *Sci Rep* 5 (2015), p. 10647. ISSN: 2045-2322. DOI: 10.1038/srep10647.

[316] M. J. Helf et al. "Comparative metabolomics with Metaboseek reveals functions of a conserved fat metabolism pathway in C. elegans". In: *Nat Commun* 13.1 (2022), p. 782. ISSN: 2041-1723. DOI: 10.1038/s41467-022-28391-9.

[317] T. Stiernagle. "Maintenance of C. elegans". In: *WormBook* (2006), pp. 1–11. ISSN: 1551-8507 (Electronic) 1551-8507 (Linking). DOI: 10.1895/wormbook.1.101.1. URL: https://www.ncbi.nlm.nih.gov/pubmed/18050451.

[318] Henry H. Le et al. "Modular metabolite assembly in Caenorhabditis elegans depends on carboxylesterases and formation of lysosome-related organelles". In: *eLife* 9 (2020), e61886. ISSN: 2050-084X. DOI: 10.7554/eLife.61886. URL: https://doi.org/10.7554/eLife.61886.

[319] Gabriel M. Simon and Benjamin F. Cravatt. "Anandamide Biosynthesis Catalyzed by the Phosphodiesterase GDE1 and Detection of Glycerophospho-N-acyl Ethanolamine Precursors in Mouse Brain*". In: *Journal of Biological Chemistry* 283.14 (2008), pp. 9341–9349. ISSN: 0021-9258. DOI: https://doi.org/10.1074/jbc.M707807200. URL: https://www.sciencedirect.com/science/article/pii/S0021925820529001.

[320] Cosima Damiana Calvano et al. "Structural Characterization of Neutral Saccharides by Negative Ion MALDI Mass Spectrometry Using a Superbasic Proton Sponge as Deprotonating Matrix". In: *Journal of The American Society for Mass Spectrometry* 28.8 (2017), pp. 1666–1675. ISSN: 1879-1123. DOI: 10.1007/s13361-017-1679-y. URL: https://doi.org/10.1007/s13361-017-1679-y.

[321] R. A. Butcher. "Small-molecule pheromones and hormones controlling nematode development". In: *Nat Chem Biol* 13.6 (2017), pp. 577–586. ISSN: 1552-4450 (Print) 1552-4450. DOI: 10.1038/nchembio.2356.

[322] C. J. J. Wrobel et al. "Combinatorial Assembly of Modular Glucosides via Carboxylesterases Regulates C. elegans Starvation Survival". In: *J Am Chem Soc* 143.36 (2021), pp. 14676–14683. ISSN: 0002-7863 (Print) 0002-7863. DOI: 10.1021/jacs.1c05908.

[323] Jan Krumsiek et al. "Mining the Unknown: A Systems Approach to Metabolite Identification Combining Genetic and Metabolic Information". In: *PLOS Genetics* 8.10 (2012), e1003005. DOI: 10.1371/journal.pgen.1003005. URL: https://doi.org/10.1371/journal.pgen.1003005.

[324] V. González-Ruiz et al. "ROMANCE: A new software tool to improve data robustness and feature identification in CE-MS metabolomics". In: *Electrophoresis* 39.9-10 (2018), pp. 1222–1232. ISSN: 0173-0835. DOI: 10.1002/elps.201700427.

[325] S Codesido et al. "A Mature ROMANCE: A Matter of Quantity and How Two Can Be Better than One". In: *ChemRxiv* (2020). DOI: 10.26434/chemrxiv.12287048.v1.

[326] Víctor González-Ruiz et al. "ROMANCE: A new software tool to improve data robustness and feature identification in CE-MS metabolomics". In: 39.9-10 (2018), pp. 1222–1232. ISSN: 0173-0835. DOI: https://doi.org/10.1002/elps.201700427. URL: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/elps.201700427.

[327] Wei Zhang and Rawi Ramautar. "CE-MS for metabolomics: Developments and applications in the period 2018–2020". In: *ELECTROPHORESIS* 42.4 (2021), pp. 381–401. ISSN: 0173-0835. DOI: https://doi.org/10.1002/elps.202000203. URL: https://doi.org/10.1002/elps.202000203.

[328] M. G. Kok, G. J. de Jong, and G. W. Somsen. "Sensitivity enhancement in capillary electrophoresis-mass spectrometry of anionic metabolites using a triethylamine-containing background electrolyte and sheath liquid". In: *Electrophoresis* 32.21 (2011), pp. 3016–24. ISSN: 0173-0835. DOI: 10.1002/elps.201100271.

[329] Sandi Azab, Ritchie Ly, and Philip Britz-McKibbin. "Robust Method for High-Throughput Screening of Fatty Acids by Multisegment Injection-Nonaqueous Capillary Electrophoresis–Mass Spectrometry with Stringent Quality Control". In: *Analytical Chemistry* 91.3 (2019), pp. 2329–2336. ISSN: 0003-2700. DOI: 10.1021/acs.analchem.8b05054. URL: https://doi.org/10.1021/acs.analchem.8b05054.

[330] J. Burghoorn et al. "Dauer pheromone and G-protein signaling modulate the coordination of intraflagellar transport kinesin motor proteins in C. elegans". In: *Journal of Cell Science* 123.12 (2010), pp. 2076–2083. ISSN: 0021-9533. DOI: 10.1242/jcs.062885. URL: %3CGo%20to%20ISI%3E://WOS:000278277700011.

[331] Tanize Acunha et al. "Anionic metabolite profiling by capillary electrophoresis–mass spectrometry using a noncovalent polymeric coating. Orange juice and wine as case studies". In: *Journal of Chromatography A* 1428 (2016), pp. 326–335. ISSN: 0021-9673. DOI: https://doi.org/10.1016/j.chroma.2015.08.001. URL: https://www.sciencedirect.com/science/article/pii/S0021967315011243.

[332] R. A. Butcher et al. "Small-molecule pheromones that control dauer development in Caenorhabditis elegans". In: *Nature Chemical Biology* 3.7 (2007), pp. 420–422. ISSN: 1552-4450. DOI: 10.1038/nchembio.2007.3. URL: %3CGo%20to%20ISI%3E://WOS:000247462800017.

[333] T. Jendoubi. "Approaches to Integrating Metabolomics and Multi-Omics Data: A Primer". In: *Metabolites* 11.3 (2021). ISSN: 2218-1989 (Print) 2218-1989. DOI: 10.3390/metabo11030184.

List of Figures

List of Tables