Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht-Karls-University

Heidelberg

Presented by

Veronika Saharuka, M.Chem.

born in: Jurmala, Latvia

Oral examination: 12 April 2022

Application of MALDI mass spectrometry for spatially resolved, untargeted metabolomics

Referees: Prof. Dr. Rudiger Hell
Dr. Kiran Patil

Summary

The field of metabolism research evaluates the molecular basis of many important phenomena in biology and medicine, ranging from cellular function to systemic-level metabolic diseases. Spatial metabolomics investigates these phenomena *in situ*, by mapping metabolites in their native spatial context. Recent technical advances led to the development of matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS) imaging, now established as a key technology for spatial metabolomics. MALDI-MS imaging allows for simultaneous and label-free detection of a wide range of analytes, such as metabolites, lipids, peptides and drugs at single-cell resolution. In this dissertation, I have looked into the capabilities this technology offers for untargeted metabolomics, and applied it to a scientific question that would be difficult to address with other technologies. The first aim of this dissertation was to create a systematic account of metabolite detectability by the common MALDI-MS imaging protocols. Despite the abundance of studies that focus on experimental or theoretical aspects of MALDI-MS, the scope of metabolite detectability by different protocols is not fully established, and choosing a suitable protocol for the detection of metabolites of interest remains a non-trivial task. To address this, I have developed experimental and computational tools for the preparation and analysis of a reference standard sample containing a wide selection of biologically-relevant metabolites. The comparison of 24 MALDI-MS protocols has shown the suitability of MS imaging for untargeted metabolomics and clarified which methods could be applicable for the analysis of individual chemical classes and biochemical pathways. The broad applicability of the obtained results was demonstrated through analyses of standard samples on other MS imaging technologies, as well as in comparison with biological tissue data sets. A community web resource was created to facilitate sharing of results and serve as an aid for the selection of the most suitable protocols for future spatial metabolomics experiments. The second aim of this dissertation was to study the associations between spatial differences in metabolism and composition of microbiota in the murine gut. This topic, important for the understanding of host-microbe interactions and their role in health and disease, remains underexplored in the field of microbiome research. By applying MALDI-MS imaging to this question, I discovered regions with distinct metabolic profiles in the whole faecal-matter filled intestine that previously were undetectable by bulk metabolomics. Next, I examined the contributions of individual metabolites, and found that both host physiology and bacterial metabolism contribute to the formation of the observed regions. Additionally, in a proof-of-principle experiment, I have shown that MALDI-MS imaging can serve as a method for determining bacterial localisation in high spatial resolution. Combining the information on the localisation of metabolites and bacteria is a way forward for obtaining direct functional insights into host-microbe interactions by a single technique, simultaneously, in high spatial resolution. Taken together, the computational developments and experimental results obtained in this dissertation help advance the field of spatially resolved untargeted metabolomics.

Zusammenfassung

Auf dem Gebiet der Stoffwechselforschung werden die molekularen Grundlagen vieler wichtiger Phänomene in Biologie und Medizin untersucht, die von Zellfunktionen bis zu systemischen Stoffwechselkrankheiten reichen. Die räumlich aufgelöste Metabolomik (spatial metabolomics) untersucht diese Phänomene in situ, indem sie Metabolite in ihrem nativen räumlichen Kontext kartiert. Die technischen Fortschritte der letzten zwei Jahrzehnte führten zur Entwicklung der matrixunterstützten Laserdesorptions-/Ionisations-Massenspektrometrie (MALDI-MS), die sich inzwischen als Schlüsseltechnologie für die räumliche Metabolomik etabliert hat. MALDI-MS-Bildgebung ermöglicht den gleichzeitigen und markierungsfreien Nachweis eines breiten Spektrums von Analyten wie Metabolite, Lipide, Peptide und Arzneimittel bei einer Auflösung, die die Analyse einzelner Zellen erlaubt. In dieser Dissertation habe ich die Möglichkeiten untersucht, die diese Technologie für die nicht zielgerichtete Metabolomik bietet, und sie auf eine wissenschaftliche Frage angewandt, die mit anderen Technologien nur schwer zu beantworten ist. Das erste Ziel dieser Dissertation bestand darin, eine offen zugängliche und systematische Übersicht über die Nachweisbarkeit von Metaboliten mit den gängigen Protokollen der MALDI-MS-Bildgebung zu erstellen. Trotz der Fülle von Studien, die sich mit experimentellen oder theoretischen Aspekten der MALDI-MS befassen, ist der Umfang der Nachweisbarkeit von Metaboliten mit verschiedenen Protokollen nicht vollständig geklärt, was die Wahl eines geeigneten Protokolls für den Nachweis von Metaboliten zu einer nicht trivialen Aufgabe macht. Um dieses Problem zu lösen, habe ich experimentelle und computergestützte Werkzeuge für die Vorbereitung und Analyse einer Referenzstandardprobe entwickelt, die eine breite Auswahl biologisch relevanter Metabolite enthält. Der Vergleich von 24 MALDI-MS-Protokollen hat gezeigt, dass sich die MS-Bildgebung für die ungezielte Metabolomik eignet, und es wurde geklärt, welche Methoden für die Analyse einzelner chemischer Klassen und biochemischer Stoffwechselwege anwendbar sind. Die breite Anwendbarkeit der erhaltenen Ergebnisse wurde durch die Analyse von Standardproben mit anderen MS-Technologien gezeigt, sowie im Vergleich zu biologischen Gewebedaten. Es wurde eine gemeinschaftliche Web-Ressource geschaffen, die den Austausch von Ergebnissen erleichtert und als Hilfsmittel für die Auswahl der am besten geeigneten Protokolle für künftige räumliche Metabolomik-Experimente dienen soll. Das zweite Ziel dieser Dissertation war die Untersuchung der Zusammenhänge zwischen räumlichen Unterschieden im Stoffwechsel und Zusammensetzung der Darmflora der Maus. Dieses Thema, das für das Verständnis der Wirt-Mikroben-Interaktionen und ihrer Rolle bei Gesundheit und Krankheit wichtig ist, ist in der Mikrobiom-Forschung noch nicht ausreichend erforscht. Durch die Anwendung von MALDI-MS-Bildgebungsverfahren auf diese Frage entdeckte ich Regionen mit unterschiedlichen Stoffwechsel-Profilen im gesamten Darminhalt, die zuvor durch metabolomische bulk-Analyse nicht nachweisbar waren. Anschließend untersuchte ich die räumliche Verteilung einzelner Metaboliten und stellte fest, dass sowohl die Physiologie des Wirts als auch der bakterielle Stoffwechsel zur räumlichen Verteilung der Metabolite beitragen. Darüber hinaus habe ich in

einem Grundsatznachweis-Experiment gezeigt, dass die MALDI-MS-Bildgebung als Methode zur Bestimmung der bakteriellen Lokalisierung in hoher räumlicher Auflösung dient, und zwar durch einen Biotypisierungsansatz. Die Kombination der Informationen über die Lokalisierung von Metaboliten und Bakterien ist ein Weg, um mit einer einzigen Technik und gleichzeitig mit hoher räumlicher Auflösung direkte funktionelle Erkenntnisse über die Interaktion zwischen Mikroben und deren Wirt zu gewinnen. Zusammenfassend tragen die in dieser Dissertation erzielten technologischen Entwicklungen und experimentellen Ergebnisse dazu bei, das Feld der räumlich aufgelösten, nicht zielgerichteten Metabolomik voranzubringen.

Acknowledgements

My PhD journey would not be the same without great colleagues, dear friends and beloved family who have supported, encouraged and inspired me throughout the last four years.

First, I would like to thank my supervisors, Dr. Theodore Alexandrov and Dr. Michael Zimmermann, for welcoming me into your groups, for your scientific guidance and for your support through all the highs and lows that we faced.

I am grateful to all of the Alexandrov and Zimmerman group current and past members, to my other EMBL colleagues and collaborators from other institutes, for your help, advice and time we spent together. It was a pleasure to work with you. I would particularly like to thank Andrew Palmer and Diana Ordoñez-Rueda who went out of their way to answer my countless questions and teach me the techniques I needed as my PhD began. I owe my deepest gratitude to Måns Ekelöf, Vitaly Kovalev, Sergio Triana and Lachlan Stuart, who were always there for me. Thank you for your generosity. I am also thankful for the opportunity I had to supervise and work together with my students, Mengze Zhang and Simon Rapp, you made it a great experience.

I wish to acknowledge the help of collaborators that were not yet mentioned, Boyao Zhang, Alessio Milanese, Lucas Vieira, Prasad Phappale, George Matfei, Bernhard Drotleff, without your contributions this work would not be the same. I thank the members of my thesis advisory committee, Prof. Dr. Rainer Pepperkok, Dr. Kiran Patil and Prof. Dr.

Rüdiger Hell for your insightful comments and suggestions that were important to shaping my PhD. I would like to give special thanks to the people who helped to improve this dissertation with suggestions for its scientific content and the language choice, in particular Martijn Molenaar, Kevin Titeca, Natalie Horvat, Mohammed Shahraz and Sharath Menon, as well as Sara Goetz for German translation. Your suggestions were invaluable.

I would like to extend my thanks to the people who encouraged me to pursue a PhD. Dr. Neil Burton, I would have not considered postgraduate studies and a change from chemistry to biology without your sincere advice, and I thank you for it. Laurent Gireaud, Yann Waye Keuong, I am thankful to you for helping me navigate my first job in analytical chemistry. The chance you gave me to learn about mass spectrometry is what brought me where I am now. I would also like to thank Delia Bucher and Dr. Steeve Boulant. Your patient guidance was fundamental for my first steps in molecular biology.

I thank my dear friends for being a source of inspiration, happiness and comfort when I needed it most. Meeting you was one of the most important things that happened in these years. Finally, I would like to express my deepest appreciation to my family, who supported me throughout this journey.

Abbreviations

9AA, 9-aminoacridine
CE, Capillary electrophoresis
CHCA, Alpha-Cyano-4-hydroxycinnamic acid
CL, Cardiolipin
ClCCA, 4-Chloro-alpha-cyanocinnamic acid
CMBT, 5-chloro-2-mercaptobenzothiazole
CoA, Coenzyme A
Da, Dalton
DAN, 1,5-diaminonaphtalene
DESI, Desorption electrospray ionisation
DG, Diglyceride
DHB, 2,5-dihydroxybenzoic acid
DHAP, 2,5-dihydroxybenzophenone
DMAN, Bis(dimethyl-amino)naphthalene
e, Electron
ESI, Electrospray ionisation
FISH, Fluorescence *in situ* hybridisation
FN, False negative
FP, False positive
FT, Fourier transform
FTICR, Fourier-transform ion cyclotron resonance
GI, Gastrointestinal
HILIC, Hydrophilic interaction liquid chromatography
HPLC, High-performance liquid chromatography
ITO, Indium tin oxide
KEGG, Kyoto Encyclopedia of Genes and Genomes
LAESI, Laser ablation electrospray ionisation
LA-ICP, Laser Ablation – Inductively Coupled Plasma
LC, Liquid chromatography
MALDESI, Matrix-assisted laser desorption electrospray ionisation
MALDI, Matrix-assisted laser desorption/ionisation
MALDI2 (or MALDI-2), Matrix-assisted laser desorption/ionisation with laser-induced post-ionisation
MAPS, 4-maleicanhydridoproton sponge
MG, Monoglyceride
MS, Mass spectrometry
MS/MS, Tandem mass spectrometry

m/z, Mass-to-charge ratio
NEDC, N-(1-naphthyl)ethylenediaminedihydrochloride
NMR, Nuclear magnetic resonance
NOR, Norharmane
PA, Phosphatidic acid
PC, Phosphatidylcholine
PCA, Principal component analysis
PDMS, Polydimethylsiloxane
PE, Phosphatidylethanolamine
PG, Phosphatidylglycerol
PI, Phosphatidylinositol
pNA, Para-nitroaniline
PNDI (or PNDI-T2), Copolymer of naphthalene diimide and thiophene
PS, Phosphatidylserine
SIMS, Secondary-ion mass spectrometry
TG, Triglyceride
TN, True negative
TOF, Time of flight
TP, True positive
RNA, Ribonucleic acid
UMAP, Uniform Manifold Approximation and Projection

Contents

Chapter 1

Introduction

1.1 Spatial metabolomics

Spatial metabolomics is a young field in metabolism research that looks at the localisation of metabolites in their native spatial context and encompasses an extensive range of topics, including metabolic diseases, immunometabolism, cancer, stem cell metabolism, microbiome research, and more.[1] The spatial nature of metabolism manifests at a range of different scales. On the subcellular level, different biochemical pathways are known to operate in different compartments of the cell, such as the nucleus, mitochondria, and cytoplasm.[2] In addition, inter-organelle metabolite transport and signalling are known to take place.[3, 4] Metabolite-driven communication between single cells is also essential, as was shown, for instance, in intestinal crypts, where direct interaction between paneth cells and neighbouring stem cells sustain the self-renewal of the intestinal epithelium.[5] On a higher level, metabolic heterogeneity within tissues can affect local biological function. For example, metabolic differences in the tumour microenvironment impair the functions of immune cells, thereby allowing tumour progression.[6] Furthermore, localised metabolism can influence inter-organ communication. A link between intestinal carbohydrate digestion and gamete production was shown in *Drosophila melanogaster* driven by the spatial proximity of the two organs.[7] Spatial aspects of metabolism can also extend to multi-organism systems, such as interactions between marine organisms

and their symbionts.[8]

On each of these levels, spatial metabolomics allows for the sampling of molecular processes underlying the observed phenotypes. Compared to the measurements of genomics, transcriptomics or proteomics technologies, metabolite levels are directly influenced through biochemical pathways and thus provide the most direct insight into the observed molecular phenotypes.[9] At the same time, the readout of spatial metabolomics is unique and complementary to those obtained by other "omics technologies". Spatial multi-omics was named among seven technologies to watch in 2022, as the strategies to perform simultaneous spatial transcriptomics and proteomics emerge.[10] Perhaps, spatial metabolomics could soon join their ranks. Integration of spatially-resolved transcript, enzyme and metabolite information could provide unprecedented opportunities for spatial systems biology research and would be even more potent than the integration of such readouts that have been done in the past at a bulk level.[11]

1.2 Mass spectrometry imaging

Mass spectrometry imaging is a key technology currently used in spatial metabolomics analysis. Matrix-assisted laser desorption/ionisation (MALDI) was one of the first "soft" ionisation techniques that, together with electrospray ionisation, caused a rapid increase in applications for mass spectrometry in biological sciences.[12, 13, 14] Laser sampling had been used previously as an ion source for mass spectrometry, but the addition of a laser-absorbing matrix, a layer of organic molecules with favourable properties for ionising other substances, led to unprecedented sensitivity and preservation of intact analytes. In particular, the capacity for measuring intact proteins led to the rise of an entirely new field of mass-spectrometry-based proteomics, and a shared Nobel prize in 2002, only 15 years after its inception.

The potential for imaging was realised early on, as the surface sampling nature of MALDI makes it naturally suited to spatial analysis.[15] In a typical MALDI imaging experiment, the sample is prepared for analysis on a flat surface. A suitable matrix compound is applied as a homogeneous layer covering the entire sample, typically through spraying or sublimation. The matrix-treated sample is analysed by scanning a pulsed UV laser beam across the surface. The mass spectrum of each position is recorded into a "data cube" of mass spectra (intensity versus mass-to-charge ratio) and Cartesian coordinates. Specific ion images are generated from this data structure by selecting specific mass windows and plotting the aggregate intensity in space, typically as a two-dimensional heatmap.

MALDI-MS imaging has seen a variety of applications in basic and applied research.[16] Many recent studies have focused on molecular-level histopathological diagnosis based on mass spectrometry profiles from patient biopsies.[17] Another important application of MALDI-MS imaging is to study drug metabolism and pharmacokinetics, for example through analysis of whole-body mouse tissue.[18] Furthermore, while not using the spatial aspect, MALDI-MS is a clinically approved technology for the identification of pathogenic bacteria.[19] In basic research MALDI-MS imaging was applied to, for example, multi-organ tissue analysis in mice which led to discovery of bacterial metabolites that take part in gut-brain axis signalling.[20] Probing complex multi-organism environments, such as the rhizosphere of the plants, or intestinal microbiome, is another prominent application of this technology.[21] More recently, applications to single-cell metabolomics analysis were also enabled by MALDI-MS.[22]

In comparison with bulk metabolomics, such as LC-MS, MALDI-MS imaging lacks separation, is overall less sensitive, and meets complications when it comes to structural analysis due to the low volumes of sampled material. Nevertheless, recent developments are improving each of these points. For example, coupling of ion mobility spectrometry

to an MS imaging instrument can provide separation of ions prior to detection based on their collisional cross section.[23] Also, an addition of a second laser for post-ionisation of the ablated sample plume, referred to as MALDI-2, can increase the instrument's sensitivity.[24] Ultimately, these disadvantages are an acceptable trade-off for the single-cell or near-single-cell resolution attainable by this technology, and its suitability for high-throughput analysis.

While this work is focused primarily on MALDI-MSI, it should be noted that other ionisation techniques can also be used to perform spatially resolved mass spectrometry. Electrospray-based ionisation can be adapted to spatial analysis either directly (desorption electrospray ionisation or DESI[25]) or as a secondary ionisation step after laser sampling (MALDESI[26], LAESI[27]). Harder ionisation sources such as SIMS[28] or LA-ICP[29] can be of interest for spatial elemental analysis, notably for multi-channel imaging of labelled antibodies (imaging mass cytometry[30]), although these technologies only provide indirect insight into the metabolic composition. Notably, besides non-MS based techniques, spatial metabolomics can also be performed by Raman spectroscopy, which analyses the vibrational frequencies of the light-excited molecules.[31] Beyond this, analysis of individual metabolites could be done by the use of metabolite sensors and fluorescent microscopy.[32]

1.3 This dissertation in global context

In this dissertation, I pursued two different research interests, combined under the umbrella of spatial metabolomics.

The first study, described in chapter 2, assessed the suitability of MALDI-MS imaging for untargeted metabolomics analysis. Despite a relatively long MALDI-MS history, metabolite analysis came into the spotlight only recently. The empirical accounts of the state of the art are scattered among individual publications that focus on the analysis of

low molecular weight compounds in different biological samples, with different protocols and using different instruments. In addition, a theory able to predict whether a compound of interest can be ionised with a particular MALDI matrix is still lacking. As a result, the understanding of the applicability of different methods in MALDI-MS to metabolite detection, and its variability between different instruments is lacking. This complicates the design of new metabolomics experiments and brings uncertainty into the interpretation of acquired data. The study in chapter 2 addresses this challenge, as it aims to create a comprehensive account of metabolite detectability by most common MALDI MS imaging methods.

The second study, described in chapter 3, focused on the spatial analysis of host-microbe interactions in the mouse intestine using MALDI-MS imaging. The variation in bacterial compositions of the gut is recognised in the field of microbiome research. It remains unclear, however, whether and how the function of microbes and host-microbiome interactions change in different parts of the gut. This question can be addressed by MALDI-MS imaging, which provides both high spatial resolution and direct functional insights through the measurement of metabolites. The study thus aims to characterise spatial variations in the metabolic composition of the gut and find associations between these variations and the localisation of gut bacteria. Furthermore, it explores the suitability of untargeted metabolite imaging for bacterial identification in intestinal tissues.

In the following chapters, a detailed introduction and motivation for both of these studies are provided. There is a separate discussion and conclusions for each study putting the obtained results into a broader perspective.

Chapter 2

Investigation into the detection of metabolites with MALDI-MS using chemical standards

2.1 Introduction

The context of the problem

Matrix-assisted laser desorption/ionisation mass spectrometry (MALDI-MS) was invented over thirty years ago, and its main early application was ionisation and analysis of proteins deposited on a metal plate.[12, 13] Subsequent advances in sampling technology have made it applicable to analyte imaging in biological tissues.[15] The development of commercially available instruments with high-resolution mass analysers (such as QTOF, Orbitrap and FTICR) has enabled the collection of informative mass spectra in the lower mass range (under 1500 Da), previously crowded by overlapping, unresolved matrix peaks.[33] This has made MALDI-MS a popular technology for the imaging of peptides, lipids, and, more recently, metabolites.[34]

In biology, the analysis of metabolites is especially interesting because metabolite levels, directly influenced by biochemical reactions, often

correlate better with the observed phenotype compared to the readouts from genomics, transcriptomics or proteomics technologies.[9] Untargeted analysis of metabolites in non-imaging mass spectrometry has proven its importance, finding applications, for example, in biomarker discovery for diseases, development of new pesticides in agriculture, or studying host-microbiome response to a drug treatment.[35] Now untargeted metabolomics is also becoming important in mass spectrometry (MS) imaging, in areas such as clinical tissue sample classification, e.g. into healthy or tumour-bearing (”molecular histopathology”), or in single-cell analysis.[16]

Untargeted metabolite analysis by MALDI-MS imaging, however, is a challenging task. One challenge stems from the high chemical diversity of metabolites. In MALDI-MS, analyte ionisation is facilitated by a matrix substance, applied over the sample. While dozens of MALDI matrices are described in the literature, there is no universally applicable matrix that enables the detection of all metabolites. An incomplete understanding of ion formation in the MALDI plume means it is not yet possible to theoretically predict the best protocol for a molecule of interest. What makes matters worse is the lack of comprehensive empirical data on metabolite-compatible protocols.

Another challenge arises from the sheer complexity of mass spectra in the lower mass range. A study that set out to interpret 25 thousand features in metabolite LC/MS spectrum of E.Coli extract found that the signal could be attributable to at most 1000 unique metabolites.[36] MALDI-MS spectra can be assumed to be no less complex, especially considering additional signals generated by the MALDI matrix itself, both on its own and through reactions between matrix and analyte. Researchers who perform untargeted metabolite MS imaging often resort to some sort of database search for peak annotation. In combination with structural validation through MS/MS or similar approaches, this has led to important discoveries. Nevertheless, in the context of untargeted imaging some

uncertainty remains. For example, it is impossible to distinguish the case when a metabolite was not detected because its abundance was below the limit of detection, from that where its detection was incompatible with the chosen MS imaging protocol, unless a chemical standard is available for a control measurement. However, when all metabolites are of potential interest, especially in preliminary, high-throughput studies, validating every single metabolite becomes impractical.

A systematic account of metabolite detectability by various MS imaging protocols would help to address both of these challenges and make untargeted metabolite imaging more effective, interpretable and impactful.

Previous research

A number of thorough reviews have gathered and organised information about the detectability of different compounds by different MALDI matrices based on published studies. For example, applicability of 25 different matrices was discussed for mass spectrometric characterisation of lipids, discussing each lipid subclass individually.[37] Another review discussed the use of at least 76 matrices or matrix combinations for analysis of low molecular weight compounds, boldly posing the question of whether the search for new matrices is ”an endless story?”.[33] Among these, the matrices that had previously been applied for metabolite analysis had some overlap in covered molecular classes. Besides regular MALDI matrices, there are also numerous protocols for targeted analysis through chemical derivatisation.[38] These modify compounds with functional groups of interest (e.g. primary amine, phenolic carboxyl, thiol, carboxylate, etc.) directly on the sample, improving ionisation and detection of the target analytes. In recent years, an increasing number of ”rationally designed” matrices have also emerged, summarised in a comprehensive review.[39] For example, MALDI matrices can be derivatised with heavy substituents or attached to polymers to reduce the number of matrix-related peaks in the lower mass range, which improves sensitivity in small molecule analysis. Many of these have not

been tried with metabolites yet since they are not commercially available.

Early studies looked for an explanation of the observed ion yields in terms of different physicochemical properties.[40] One study compared the ionisation of three peptides by six MALDI matrices with acid/base properties of analytes and matrices in the gas phase, concluding that there is no simple relationship. Other studies took an empirical approach. In one, the sensitivity of five MALDI matrices was assessed for lipid detection in murine liver tissue.[41] This study found differences in ionisation among 12 lipid subclasses. Another study compared the detectability of lipids with 4 matrices between two MALDI-MS technologies, using a mix of lipid standards spiked into tissue homogenates or applied on top of tissue sections.[42] Such experimental approaches provide a good insight into the detectability of compounds by MALDI-MS protocols and could be extended to metabolite analysis.

Aims of the study

This work aims to produce an openly accessible and systematic account of metabolite detectability by the most common mass spectrometry imaging methods. The objectives of the study are as follows.

1. To develop a standard sample and a procedure (experimental and computational) that allows broad assessment of metabolite detectability for MS imaging protocols.

2. To compare the detection coverage of metabolites from the library by multiple MS imaging protocols in the host laboratory on a MALDI-Orbitrap instrument.

3. To distribute identical samples to prominent laboratories in the MS imaging community and compare metabolome detection coverage among different instruments (ionisation method and detection).

4. To make the results into an interactive online resource, as well as provide open access to all experimental data.

Significance of obtained results

Results obtained in the present study will assist the protocol choice for metabolite analysis by MALDI-MS imaging in several ways. Firstly, the users of the developed online resource will be able to visualise how different methods compare for detecting a molecule of interest (e.g. in terms of intensity). Secondly, the users will get information on additional ion species, such as metal cation adducts, matrix adducts, neutral losses, that are important for each metabolite. Thirdly, the readers and the users will get an overview of how each method performs for a selection of metabolites, such as those in a particular molecular class or metabolic pathway. Importantly, the limitations of the methods (for a subset of metabolites) will be determined, which will allow excluding unreasonable annotations when performing untargeted metabolomics on biological samples. A comparison of the data obtained by many imaging labs will shed light on variability between different techniques. Finally, a planned publication of all raw experimental data from the study will allow scientists to test further hypotheses.

Outlook

The experimental method and computational pipeline developed in the course of this study can be expanded to include more MS imaging methods in future, as new methods emerge and instruments improve. The results already obtained could serve as a benchmark for method transfer and validation. In addition, the detectability of other compounds of interest could be studied in the same way if the standard sample is prepared with those compounds instead of metabolites - this may, for example, be interesting for imaging of drugs or plant metabolites.

2.2 Study design

To reach the aim of this study and assess metabolite detectability by a mass spectrometry imaging method, careful choice of standard sample, development of the robust experimental protocol and computational analysis is needed. The methodology section explains decisions made in this regard during the project.

2.2.1 Choice of the standard sample

To characterise the performance of any analytical method, it is necessary to use a standard sample of known properties. When choosing a standard sample, three options had been considered.

1. Using a biological sample, such as animal tissue, could provide relevant results when a similar tissue is a sample of interest. Nevertheless, it is less applicable for broader metabolite profiling because the chemical composition of such a sample is not strictly defined, and different tissues vary in their metabolite content. For example, it would be difficult and laborious to determine whether a metabolite is not detected because of the limitations of a mass spectrometry method, or because it is absent in the sample.

2. Using a standard reference material with a more defined chemical composition is a second possibility. For example, 1950–Metabolites in Frozen Human Plasma from National Institute of Standards and Technology (NIST) [43] has 100 metabolites that are certified and several hundred detected by the metabolomics community. Such a standard performs best with methods that include a separation step, such as liquid chromatography. The limitation of analysing a mixture of metabolites by MS imaging, which does not have a separation step, is that detection of some compounds may be suppressed. This is because all analytes are ionised simultaneously, leading to a bias towards metabolites with energetically or kinetically favourable ionisation reactions.

3. Using individual pure chemical standards avoids the suppression concern and benefits from analysis simplicity. This approach was chosen for the study. It is important to note that using individual chemical standards provides the best performance measure of a method under ideal conditions, which may differ from the performance using a complex biological sample. As described in section 2.3.2 this concern was addressed by comparing results obtained with standard samples to the MS imaging data sets of animal tissue deposited in a public repository.

The choice of metabolites comprising the standard sample was driven by the desire to represent the most important primary metabolites common for mammals, bacteria and yeast, and to capture diversity present among naturally occurring metabolites. Compounds were chosen manually (see appendix A), and the resulting selection was assessed by the following three metrics.

Metric 1: Chemical diversity

Metabolites differ in molecular structure and chemical properties. To assess the diversity of chemical properties, chosen compounds were summarised by chemical classification (see figure 2.1 and appendix B).

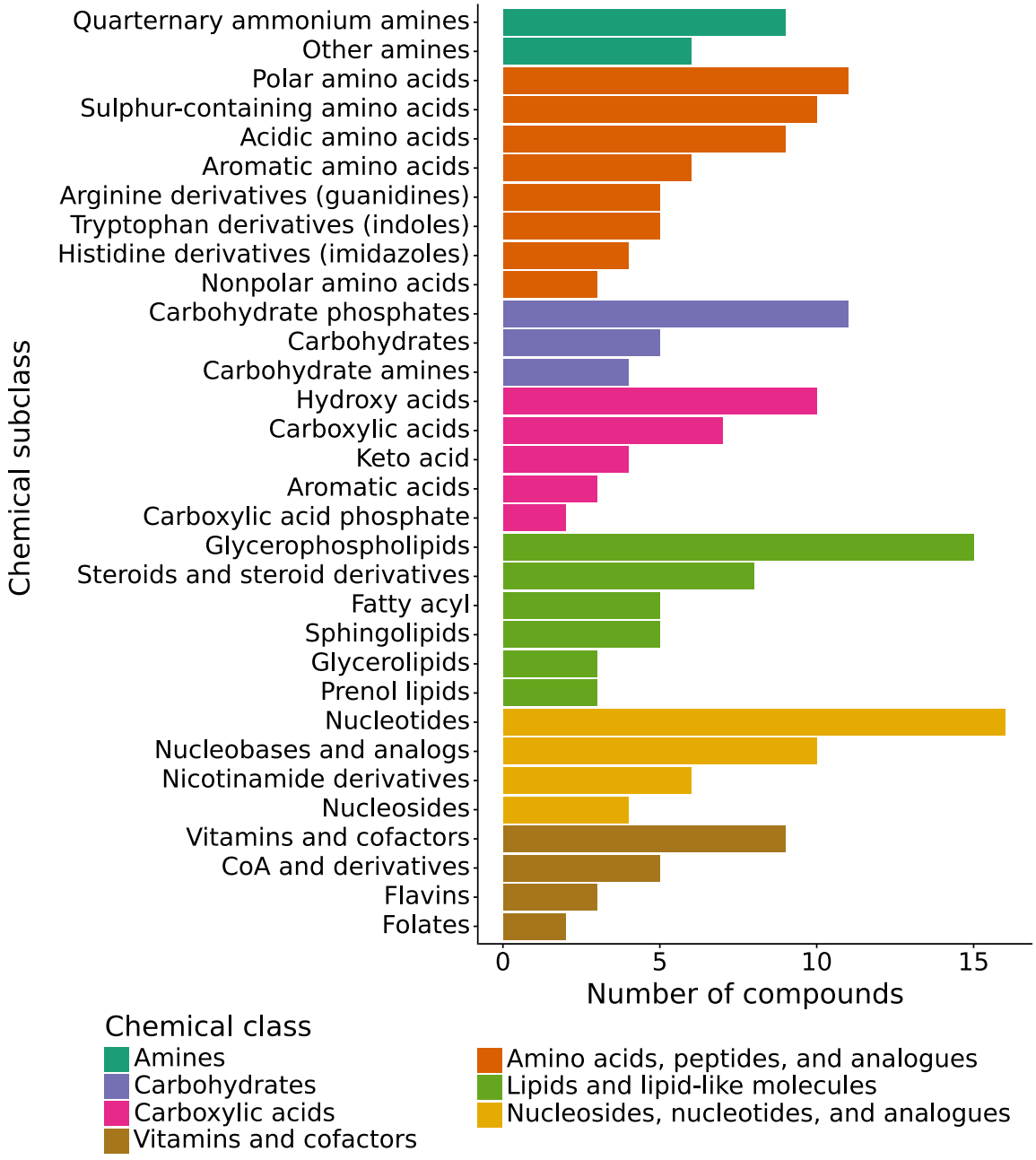


Figure 2.1 Breakdown of metabolites chosen for the standard sample by chemical class and subclass. Two-step classification was constructed with broader chemical classes and their subclasses. The classification used here borrows from automatic chemical classification ClassyFire [44] but was adapted in a way that classification terms are the most relevant and understandable in the metabolomics community. This figure shows that a wide range of chemical classes was represented in the standard sample.

To assess the structural diversity of the chosen molecules, chemical space was constructed using structures of all metabolites from a mouse genome-scale metabolic model [45]. To do this, metabolite SMILES (chemoinformatic representation of molecular structure) were downloaded from the BiGG open-source repository [46]. SMILES were used to calculate a vector where every bit represents a fragment of the molecule, using the ”Morgan/Circular fingerprint” method in RDkit Python library [47] as shown in example figure 2.2.

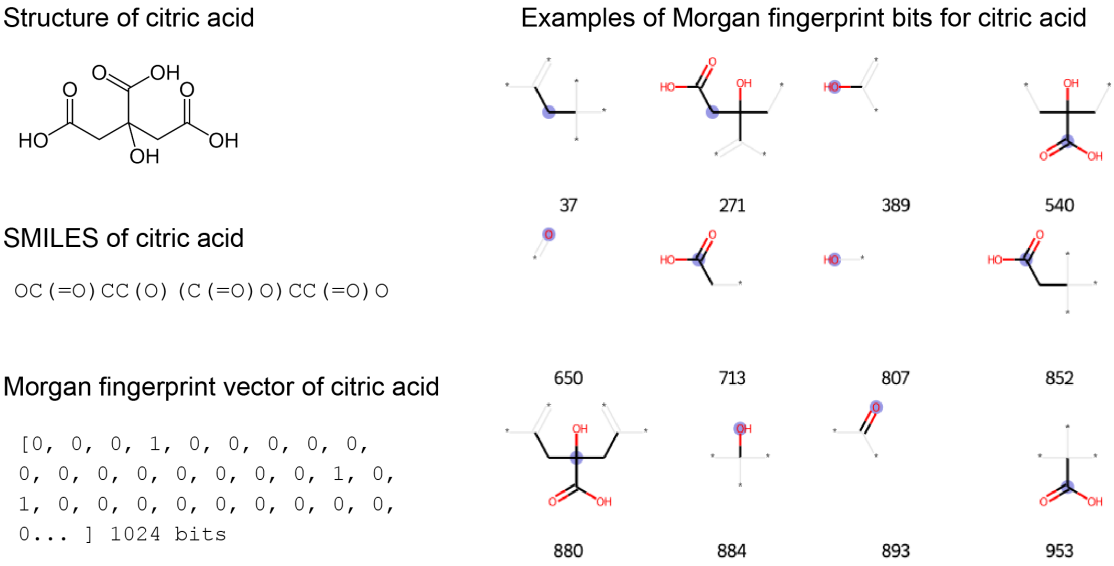


Figure 2.2 Example of Morgan fingerprint for citric acid.

Then, principal component analysis (PCA) was used to reduce the dimensionality of chemical space and visualise it as 2D and 3D plots, see figure 2.3. The fact that chosen metabolites span the entire space of all murine metabolites provides qualitative evidence that chosen compounds are structurally as diverse as what is observed among natural metabolites. The human or yeast genome-scale metabolic model could have been used to obtain a similar result, as the majority of the primary metabolites are conserved between these organisms.

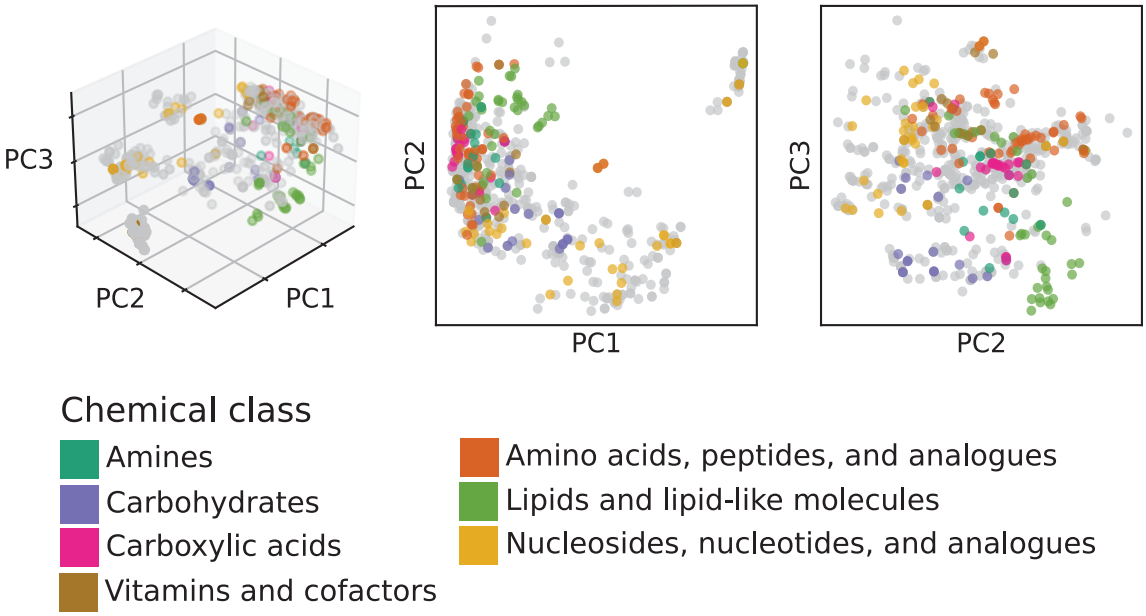


Figure 2.3 Three different views of the chemical space (Morgan fingerprint PCA plot). Grey dots represent all mouse metabolites from the genome-scale metabolic model (iMM1865), while coloured dots represent metabolites chosen for the standard sample. Distance between two molecules decreases with increasing structural similarity. This figure shows that metabolites selected for the standard sample span the entire chemical space occupied by all mouse metabolites.

Metric 2: Biological function

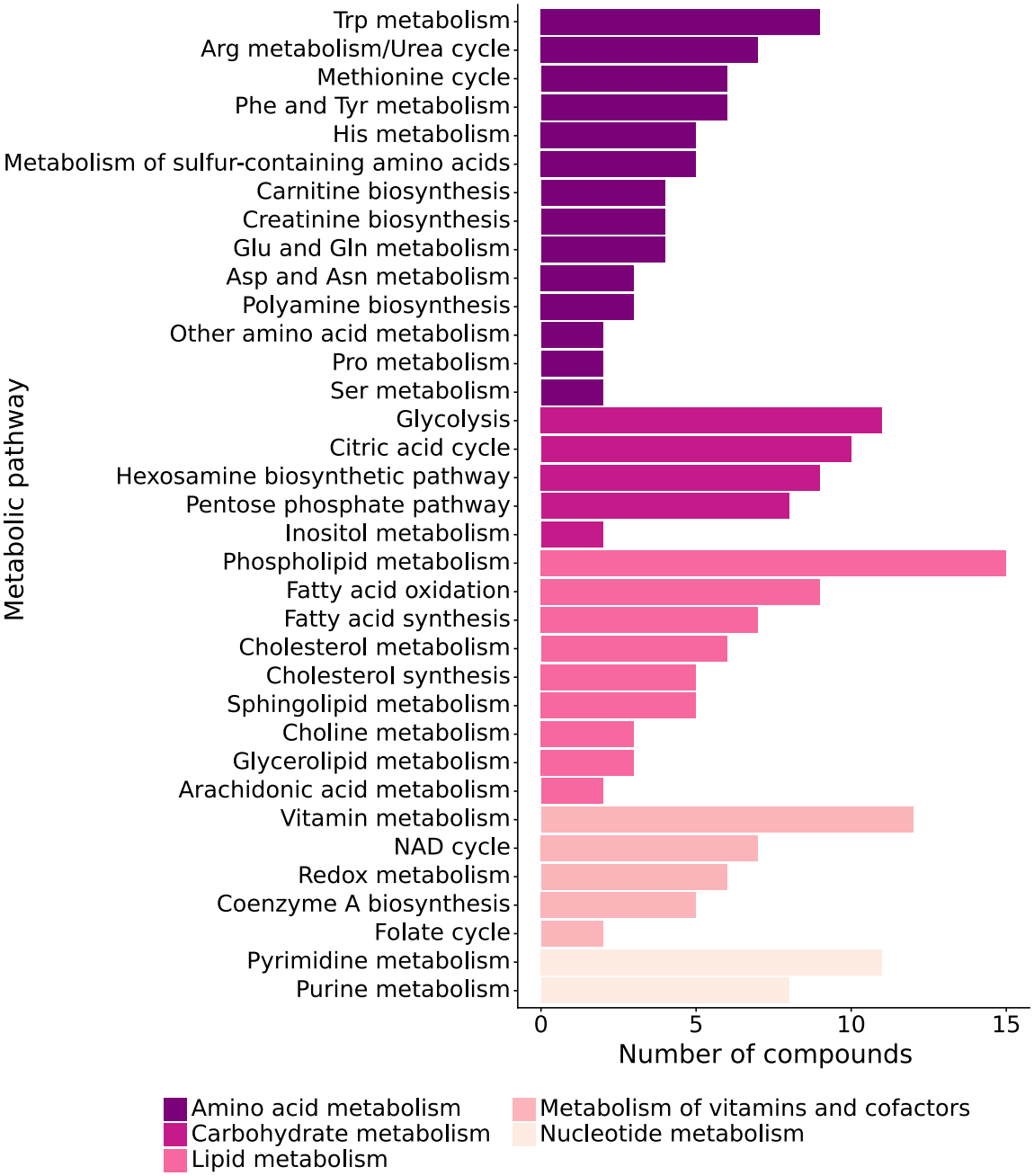


Figure 2.4 Breakdown of metabolites chosen for the standard sample by biochemical pathway. Pathway names were obtained from the Roche biochemical pathway map [48] and grouped in broader categories to obtain a two-level classification. This figure shows that metabolites from a wide range of biochemical pathways were included in the standard sample.

Metabolites participate in different biochemical pathways. To assess diversity in this category, the chosen compounds were summarised by the biochemical pathways they take part in (see figure 2.4 and appendix C). Representation in figure 2.5 shows chosen metabolites highlighted on a map of biochemical pathways in a web-based tool IPath3 [49].

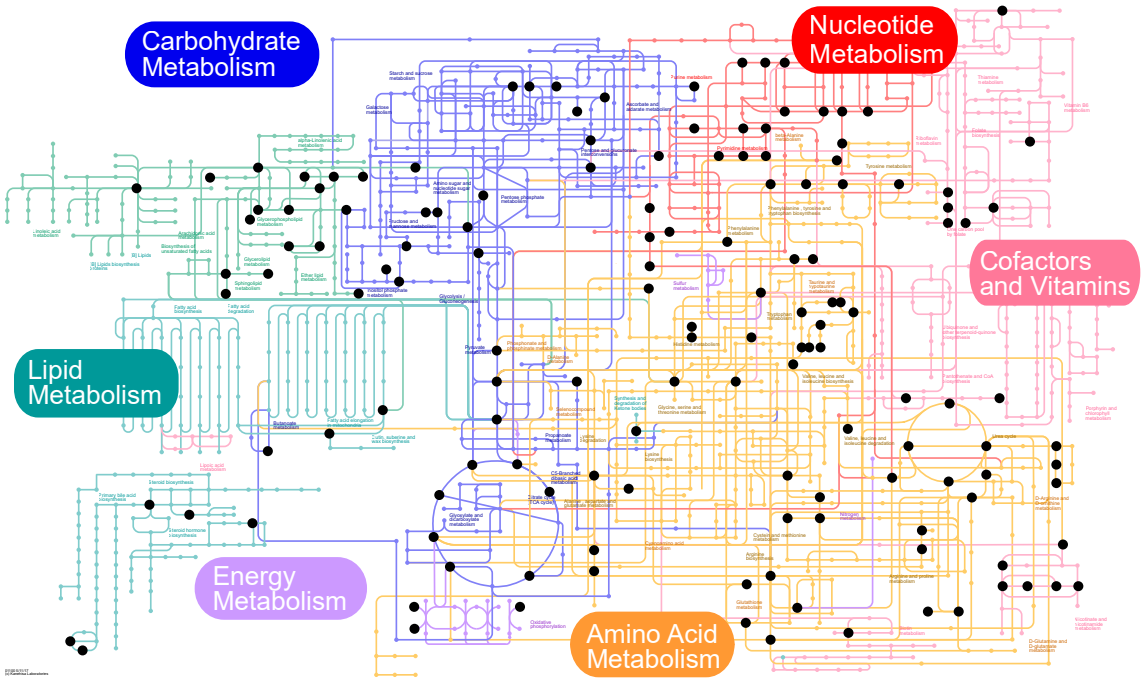


Figure 2.5 Map of common biochemical pathways. Each line is a reaction and each node is a metabolite. Highlighted nodes represent metabolites chosen for the standard sample. This visualisation excludes most of the chosen lipids, because these lipids lack non-generic KEGG identifiers required for the visualisation.

Metric 3: Natural abundance

Concentrations of metabolites in living cells span several orders of magnitude. To make sure that most of the abundant metabolites are captured in the standard sample, the chosen compounds were compared to a study where absolute abundance was determined for over 100 metabolites in *Escherichia coli*.[50] Figure 2.6 shows that a good coverage of highly abundant metabolites was achieved among the compounds chosen for the standard sample. It would be beneficial to make a similar comparison with absolute abundances of metabolites in mammalian and yeast cells once such data becomes available in the literature.

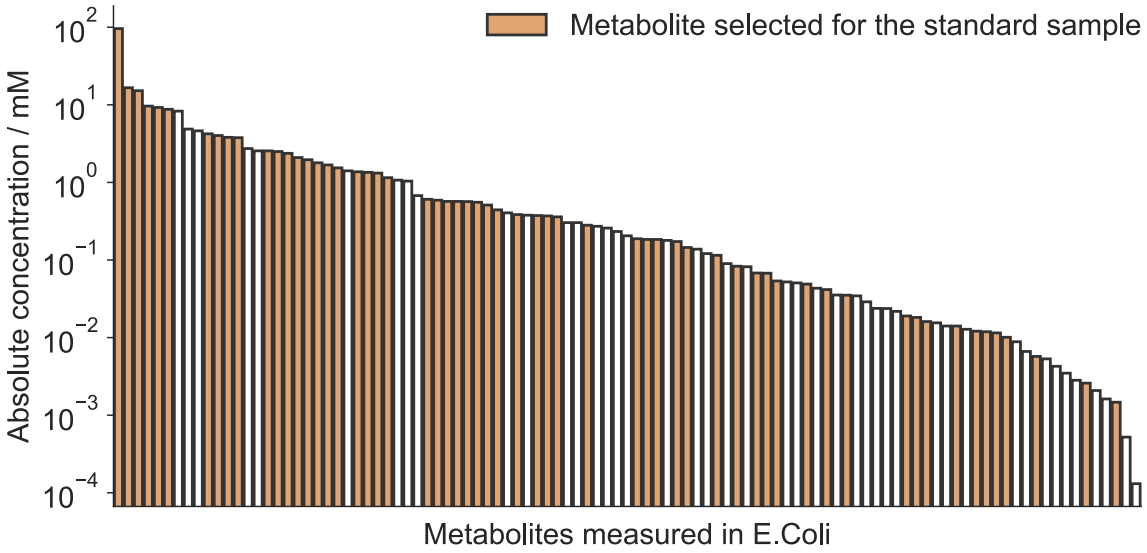


Figure 2.6 Absolute concentrations of metabolites measured in E.Coli [50]. All bars show metabolites from the cited study, ordered by concentration from high to low, left to right. Orange bars show metabolites that were also included in the standard sample. This figure shows that most of the highly abundant metabolites are represented among the compounds chosen for the standard sample preparation.

2.2.2 Development of a protocol to prepare the standard sample

Considerations for standard sample preparation

Besides the chemical composition, there are several requirements the standard sample had to meet.

- First, it should be possible to make tens to hundreds of copies of the standard sample in a reproducible way.
- Second, the sample preparation procedure should be as fast, easy and affordable as possible.
- Third, the standard sample should be compatible with MS imaging instruments in terms of shape (flat, rectangular) and material (conductivity).
- Fourth, it should be possible to image the standard sample with sufficient spatial resolution in a reasonable time (for example, overnight).

It was decided that an array of metabolite spots on a glass slide is a good compromise for these criteria for most of the imaging set-ups.

Several ways to make a metabolite array on a glass slide were considered. The scale of the study meant that compound deposition needed to be automated. In industry, high-throughput sample preparation for MALDI-MS screenings is currently done using automatic non-contact printers that deposit nanolitre-volume droplets onto steel MALDI target plates into wells with hydrophobic boundaries.[51]

1. A bench-top non-contact droplet dispenser (I.DOT, Dispendix, Stuttgart, Germany) was evaluated both with regular glass slides and patterned glass slides with hydrophilic micro-wells (Aquarray, Eggenstein, Germany). The limitation of this system was that it produced localised droplets only on specialised glass slides and only with water solutions, while most metabolites require some organic solvent to be dissolved. In particular for lipids, solvent compatibility was a limiting factor.

2. Next, an automatic contact printer (ChipWriter Compact robot, Bio-Rad Laboratories, Hercules, U.S.A.) formerly used for RNAi array preparation [52] was tested. This system could print spots of metabolite-gelatine mix in a regular pattern directly from a 384-well plate. While regular slides could be used, the limitations were that only aqueous solutions could be used to minimise evaporation and that only viscous, gelatine-containing solutions could be used for printing, which may have caused the ion signal suppression. Additionally, substantial spot-to-spot metabolite carryover was observed.

3. Finally, an automatic compound dispenser for thin layer chromatography (Automatic TLC sampler 4, CAMAG, Muttenz, Switzerland) was tested with the most success. Compatible with both regular glass slides and organic solvents, it dispensed the metabolite solutions using syringe nanospray and achieved uniform compound deposition onto the surface.

Standard sample preparation

The standard sample preparation procedure is shown in figure 2.7.

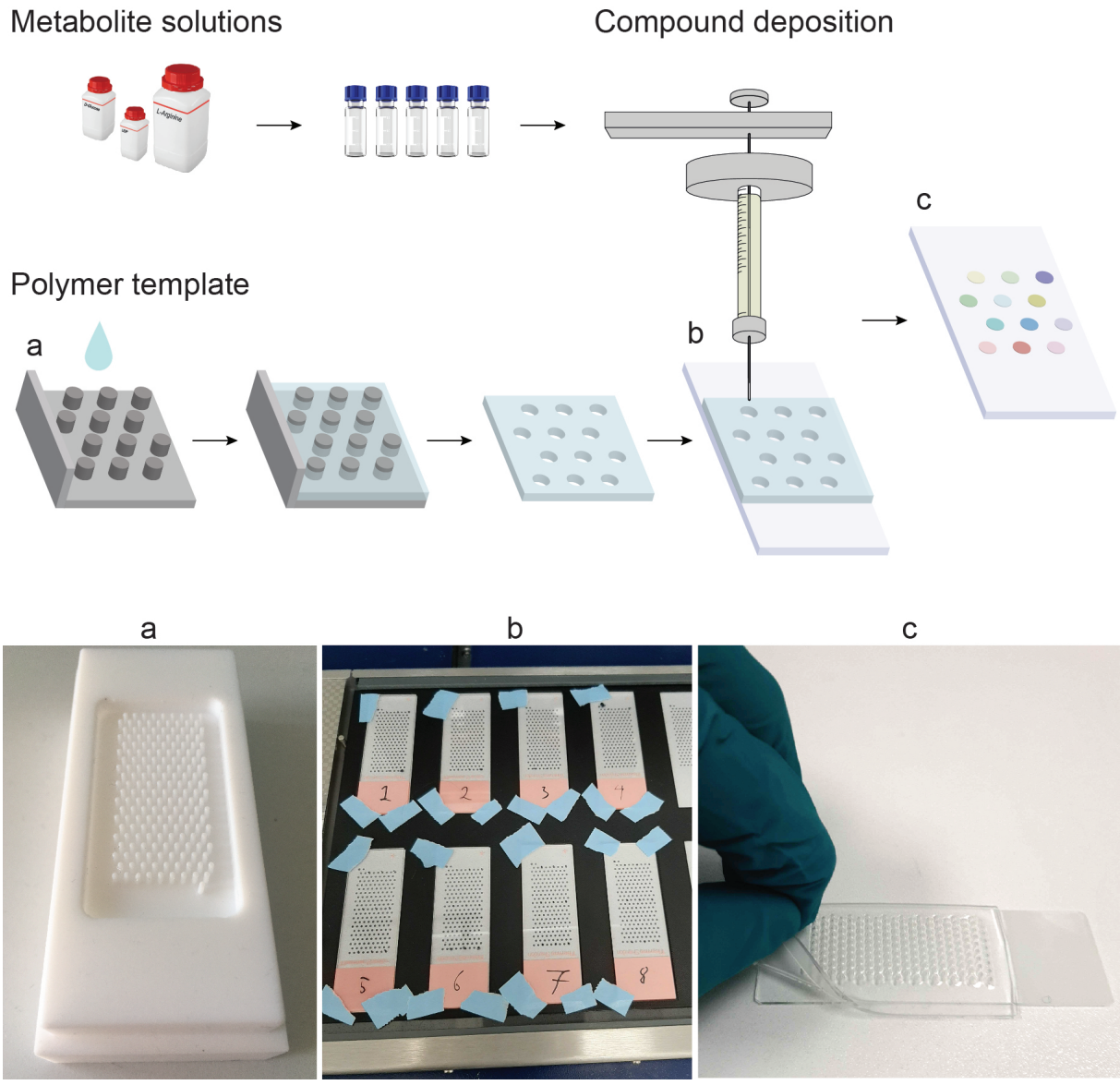


Figure 2.7 To produce a standard sample, first, individual equimolar solutions of every metabolite were prepared. The automatic compound dispenser was programmed to dispense the compound in a regular pattern onto the glass slides. To prevent metabolite cross-contamination between neighbouring spots, a polymer template was used, which formed leak-tight wells when placed on the glass slide. After compounds were deposited, the template was removed. Standard slides were stored in an argon atmosphere at -80 Celsius until analysis.

Metabolite solutions A two-step procedure was used to prepare metabolite solutions. First, a stock solution was prepared in a solvent system suitable for a given metabolite. Information on the preparation of all 172 metabolite solutions can be found in appendix D. Solubility information was collected from literature and vendor documents and adapted as necessary. Next, the stock solution was diluted to 200 micromolar in a volatile solvent to produce a quickly drying nanospray and ensure uniform deposition. Here two solvent systems were used: chloroform-methanol for lipophilic compounds and methanol-water for hydrophilic compounds. All solutions were stored in HPLC vials at -80 Celsius after preparation. Every day a fresh set of vials was thawed and used, to minimise differences between batches of standard samples prepared on different days. The solutions were analysed using HILIC and reversed-phase LC-MS/MS before slide preparation, and again after the end of slide preparation to verify the presence of the standard compound and to monitor potential degradation related to storage, which was found to be minimal.

Polymer template To prepare a polymer template, a custom mould with 180 regularly spaced columns was designed, and precision-cut from a solid polyoxymethylene block. A curable PDMS silicone, Sylgard 184, was set in the mould. When released from the mould, the PDMS template looked like a thin silicone sheet with 180 holes. When placed onto a glass slide, the template produced 20 rows and 9 columns of leak-tight wells, each well 1 mm in diameter. The mould design was such that the resulting polymer template could be placed on the glass slide with the well grid reproducibly aligned.

Compound deposition The automatic compound dispenser had a stage for thin layer chromatography silica plates. Instead of a silica plate, a compatible sample holder that could hold 10 glass slides was designed. The sample holder was prepared from a plastic sheet on the bottom glued to a 3D-printed layer on top. Since the glass slide position

in the sample holder was fixed, and the PDMS template was aligned the same relative to each slide, the positions of the wells on the glass slides were the same every time that the glass slide was put in the machine. The compound dispenser was then programmed to deposit 2 microlitres of one compound into each well. This way, up to 10 copies of the standard sample could be prepared simultaneously. Over 150 copies of the standard sample were produced this way.

2.2.3 Data acquisition

Standard samples were used to acquire data using different instruments and protocols. A common procedure applied to all samples, where a rectangular area covering all metabolite spots was analysed by an imaging mass spectrometer in raster mode. Within this rectangular area, sampling was spaced by 100-150 micrometres, to collect tens of spectra per metabolite spot.

Comparison of MALDI-MS imaging protocols

In the first part of the project, standard samples were used to measure metabolite detectability in the host laboratory on the atmospheric pressure MALDI-Orbitrap MS system. For this, twelve established sample preparation protocols using MALDI matrices previously described in the literature were chosen. The matrices are briefly described below, and the chemical structures are shown in Figure 2.8.

Positive ionisation mode matrices

- DHB (2,5-dihydroxybenzoic acid) was chosen because it is one of the most commonly used positive ionisation mode matrices in MALDI-MS imaging.[53]
- DHAP (2,5-dihydroxybenzophenone) is a derivative of DHB. It has been shown to ionise phospholipids in both positive and negative ionisation modes.[54] DHAP has a much smaller crystal size than

DHB, which enables its use for high spatial resolution imaging. DHAP is not vacuum-stable and thus can be used only with atmospheric pressure ionisation sources. A vacuum-stable derivative of DHAP was described in the literature but was not commercially available.[55]

- CHCA (alpha-Cyano-4-hydroxycinnamic acid) is a classic positive ionisation mode matrix used in the early days of MALDI-MS for protein analysis.[53] The advantage of CHCA is, like with DHAP, its small crystal size (on the order of nanometres), which is required for high spatial resolution imaging applications.

- ClCCA (4-Chloro-alpha-cyanocinnamic acid) was found through a screen of CHCA derivatives, as a matrix that enhances peptide signal in proteomic analyses.[56] The superior performance of ClCCA to CHCA was observed only at lower ionising laser wavelengths such as 337 nm. While the instrument used for this project has a higher ionising laser wavelength (343 nm), it was hypothesised that ClCCA may still be superior to CHCA in that its chlorine-containing fragments would have less chance to be isobaric with metabolites of interest. ClCCA, like CHCA, forms small crystals that suit the needs of high spatial resolution imaging.

- CMBT (5-chloro-2-mercaptobenzothiazole) was discovered two decades ago to be a useful MALDI matrix for the analysis of peptides, oligosaccharides and glycolipids.[57] Recently it was used for lipid imaging in the brain tissue in positive ionisation mode.[41]

Negative ionisation mode matrices

- 9AA (9-aminoacridine) was originally presented as a negative mode matrix for solutions of peptides, oligonucleotides and low molecular weight compounds with acidic protons, such as, for example, carboxylic acids, sulphonates and amines.[58] Later it was also used for MALDI-MS imaging, for example of lipids in the brain tissue.[59]

- NEDC (N-(1-naphthyl) ethylenediamine dihydrochloride) was shown to detect glycerophospholipids and low molecular weight compounds, for example, glucose, in animal tissue in negative ionisation mode.[60]

- MAPS (4-maleicanhydridoproton sponge) is a derivative of a matrix bis(dimethyl-amino)naphthalene (proton sponge, DMAN). DMAN readily generates a wide range of deprotonated metabolite ions, when mixed with a sample of interest in a solution.[61] However, DMAN is too volatile and produces too large crystals to be used for imaging (the finding from the own work of the author of this dissertation). MAPS, on the other hand, was synthesised to be vacuum-stable and was reported to detect some low-mass range analytes during tissue imaging in negative ionisation mode.[62]

Dual mode matrices

- DAN (1,5-diaminonaphthalene) was used as a dual-mode matrix for lipid analysis in intestinal tissue,[63] and for small metabolite localisation in plant tissue.[64]

- pNA (4-nitroaniline) was introduced as a matrix for the analysis of proteins, nucleotides, and, more recently, lipids. pNA was shown to outperform DHB in both positive and negative ionisation modes for imaging applications.[65] pNA is suitable for atmospheric and intermediate pressure systems due to its moderate volatility.

- NOR (norharmane) was used as a MALDI matrix useful for dual-polarity imaging of lipids in biological tissues.[66]

- PNDI (or PNDI-T2, a copolymer of naphthalene diimide and bithiophene) is a polymer studied in the context of photovoltaics that was recently tested as a MALDI matrix. It was reported to generate none or few matrix peaks in the spectra, enhance analyte detection, and work in both ionisation modes.[67]

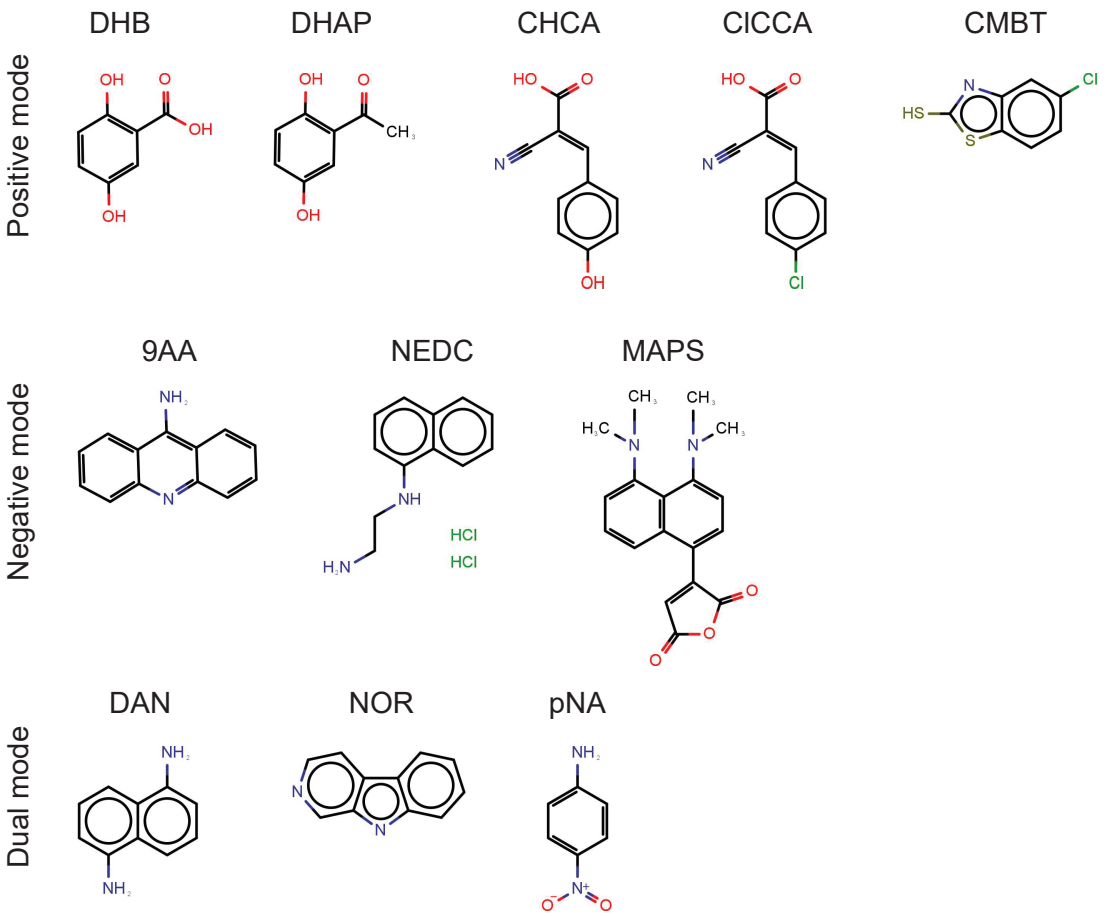


Figure 2.8 Chemical structures of MALDI matrices used in the project. Abbreviations are explained in the paragraph above the figure.

MALDI matrices were applied to the standard samples by automatic spray coating, and a decision was made to apply the same amount of every matrix (except MAPS and PNDI-T2 due to their high molecular mass). This way the same matrix to analyte ratio could be maintained for most protocols, in the attempt to make the results as comparable as possible.

Since MALDI-MS imaging usually consumes very little sample, the standard sample slide could be analysed four times with a minimum offset determined by the radius of the ablation spot. This allowed acquiring data in positive and negative ionisation modes, and over two mass ranges: m/z 75-350 and m/z 300-1200. Splitting of the mass range was necessary due to limitations of the mass spectrometer. These pairs of

mass ranges were later merged computationally.

Survey of MS imaging instruments

In the second part of the project, the standard samples were used for inter-laboratory comparison of metabolite detectability using different MS imaging technologies. The participating laboratories were recruited among the best experts in the field, including several that actively develop their own instrumentation. The instruments were chosen to represent the variety and current state of the art of imaging technologies. The limiting factor for including a certain technology was its capability to image a wide range of masses with high mass accuracy and a low degree of fragmentation. This restricted the range of techniques to the ”soft” ionisation methods, coupled to high-resolution time of flight (TOF) or Fourier transform (FT) mass analysers. The letter with description of requirements sent out to the collaborators can be found in appendix E. The instruments used are summarised in Table 2.1.

The participants were asked to produce two replicate images in both ionisation modes. To allow this, sets of four standard sample slides were provided to collaborators, one set for each instrument evaluated. Standard samples prepared on conductive indium tin oxide (ITO) coated slides were provided for some MALDI instruments operated at vacuum. The collaborators were asked to follow the protocols that they commonly use for biological samples. For this reason, methods that did require matrix application did not use all the same compound, but rather followed the procedure most established in the collaborator laboratory, and the most suitable for the instrument and the sample provided.

Sample (pos / neg)	Institution	Ion source	Analyser	Substrate	Matrix (pos / neg)
7A / 7J	EMBL	AP-MALDI	Orbitrap	Glass	DHB / DAN
12C / 12E	HS Mannheim	MALDI	FTICR	ITO	DHB / DAN
12G / 12I	HS Mannheim	MALDI	qTOF	ITO	DHB / DAN
9A / 9E	NCSU	MALDESI	Orbitrap	Glass	Ice
1A / 1C	AstraZeneca	DESI	Orbitrap	Glass	-
8A / 8D	JLU-Gießen	AP-MALDI	Orbitrap	Glass	DHB / 9AA
13G / 13J	Bruker	MALDI	qTOF	ITO	DHB / NEDC
10D / 10C	WWU Münster	MALDI-2	qTOF	ITO	DHAP
4F / 3A	WWU Münster	MALDI-2	Orbitrap	Glass	DHAP
1J / 1H	UT Austin	DESI	Orbitrap	Glass	-

Table 2.1 Table of MS imaging instruments involved in the data acquisition. ”pos” means positive ionisation mode, ”neg” means negative ionisation mode. ITO refers to indium-tin-oxide-coated conductive glass slides.

The most differentiating property of considered technologies is the method of sampling and ionisation.

- In matrix-assisted laser desorption/ionisation (MALDI), ions are produced directly from a matrix-treated sample surface by laser irradiation, typically in the near-ultraviolet range. MALDI ionisation is characterised by the generation of exclusively singly-charged molecular ions and the need for a suitable matrix. The inter-laboratory comparison included several common variations of this method.

- Atmospheric pressure (AP)-MALDI indicates that the ablation and ionisation are done under atmospheric conditions rather than in the vacuum of the mass spectrometer as is more common.[68] AP-MALDI has the benefit of compatibility with matrices that would sublimate in vacuum conditions, but is somewhat more complicated to interface with mass spectrometers as the produced ions must be transferred from atmospheric pressure into the vacuum system of the mass spectrometer.

- MALDI-2 refers to instruments where the plume of ablated material

is further post-ionised with a second laser. Irradiation with the second laser could increase sensitivity at the expense of additional complexity of instrument construction and protocol development (for example, not all matrices are equally compatible with MALDI-2).[24]

- Desorption electrospray ionisation (DESI), similar to MALDI, combines sampling and ionisation into one step. This technique bombards the surface of the sample with a stream of charged solvent droplets and directs generated secondary ions to the mass analyser.[25] The resulting spectra are similar to direct infusion electrospray ionisation, notably with the possibility of generating multiply charged ions for large analytes, such as proteins.

- The hybrid MALDESI instrument represented in this study uses DESI type ionisation but utilises an infrared laser probe for sampling.[26] Common to such an electrospray-based source is that no organic matrix application is required for ionisation. This allows samples to be analysed fresh or frozen, as well as desiccated.

The selection of mass analysers reflects the available commercial options for instruments with a sufficiently high mass resolving power and mass accuracy. In particular, the data analysis pipeline using METASPACE for image annotation (discussed in the following section 2.2.4) requires consistently well-calibrated spectra across an entire image. For this project, an absolute limit of 10 ppm mass error was imposed. The instruments chosen for this study used either quadrupole-time of flight (qTOF) systems or FT mass analysers: Orbitrap or Ion Cyclotron Resonance (FTICR). qTOF performance depends almost entirely on the instrument construction, especially the length of the flight path. FT instruments measure oscillations of trapped ions, and so could increase resolution and mass accuracy for longer cycle times. All the instruments included here were capable of operating within the chosen mass error tolerance. Thus, the spectral data were processed identically.

2.2.4 Data analysis

Mass spectrometry data were used to generate ion images for each metabolite present in the standard sample. Over 35 theoretical ions with all combinations of adducts ($+H^+$, $+Na^+$, $+K^+$, $-e$ for positive ionisation mode; $-H^+$, $+Cl^-$, $+e$ for negative ionisation mode) and neutral losses (water, carbon dioxide, ammonia, hydrogen, phosphate), as well as neutral gain of dihydrogen (reduction), were considered (e.g. $[cholesterol + H - H_2O]^+$). One ion image per theoretical ion was generated using the METASPACE [69] custom database functionality, resulting in approximately 4000 images per data set. Since some theoretical ions were not detected, or not chemically plausible, or were isobaric with matrix peaks, part of the generated images was expected to be background noise. To distinguish between real signal and background noise, an image classification procedure was developed.

Rule-based ion image classification

First, the positions of all metabolite spots in the ion images had to be found. Since the number of rows and columns in the metabolite grid was known, it was sufficient to determine the location of only three spots in the ion images to calculate the coordinates of all remaining spots in the grid. For this purpose, along with other metabolite solutions, three ”guide spots” were deposited onto the standard sample containing highly-ionisable compounds: rhodamine B and caffeine. An interactive Python tool was developed, where a ”guide spot” ion image was shown, and where the user could provide coordinates of the spot centres by clicking on them (see figure 2.9). The output of the tool was a mask image of the same dimensions as the ion image, where each pixel was assigned a label: 0 for background pixels, or between 1 and 180 for pixels assigned to metabolite spots.

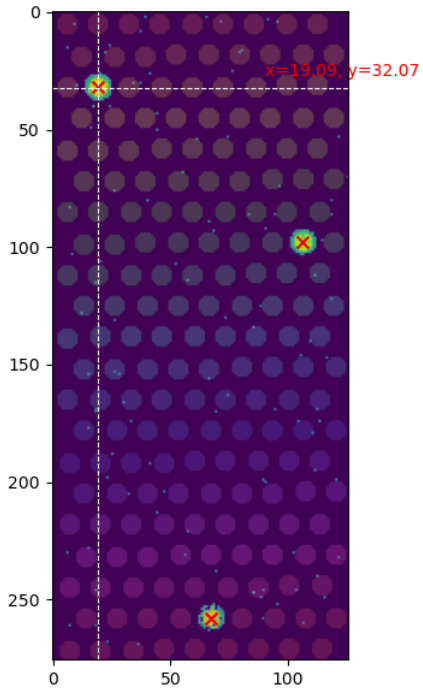


Figure 2.9 Screenshot of interactive grid-fitting Python tool that was used for finding positions of individual spots. Shown is an ion image of three ”guide spots” overlaid with computed spot locations.

Next, for every spot and metabolite ion expected in this spot, the following metrics were calculated.

- **Intensity ratio** : the ratio of average ion intensity inside the metabolite spot to the average intensity outside any spot (i.e. in background)
- **Occupancy ratio** : the ratio of the fraction of non-zero pixels inside the metabolite spot to the fraction of non-zero pixels in background

A rule-based classifier based on the two metrics mentioned above could help determine whether a metabolite ion of interest is detected in a spot where it has been deposited. An ion was considered to be detected if the signal was either localised (occupancy ratio above selected threshold) or intense (intensity ratio above selected threshold). This way it was possible to keep both faint but localised ions and bright ions occurring in noisy images, as shown in figure 2.10.

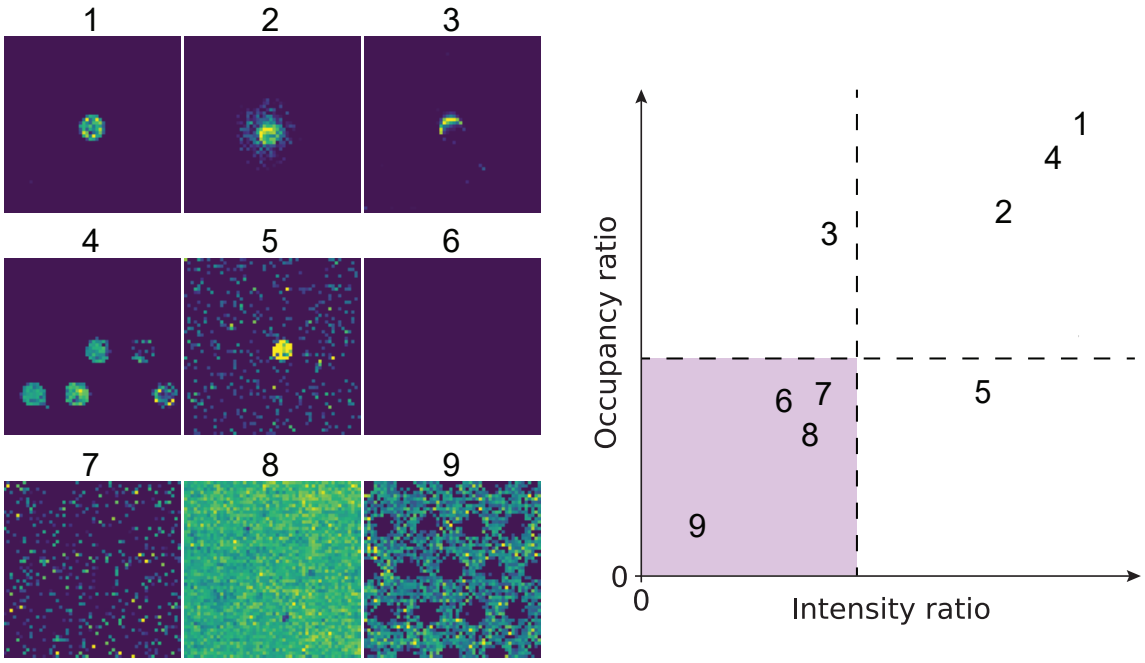


Figure 2.10 A classifier based on two metrics: intensity ratio and occupancy ratio. The panel on the left shows different types of ion images occurring in the data, with signal intensity increasing from dark blue to yellow. The schematic on the right shows how ion images can be separated into two groups by setting thresholds on the two metrics. When either of the metrics is above its threshold, the image is classified as a real signal (white area), otherwise, the image is classified as background noise (shaded area). Ion images 1-5 were classified as a real signal: localised in the spot (1), diffused out of the spot (2), present only in part of the spot (3), detected in multiple additional spots (4), and isobaric with weaker background signal (5). Ion images 6-9 were classified as background noise: images with less than two nonzero pixels (6), sparse random signal (7), uniform signal (8), images with stronger signal outside of the spots than inside.

To assess the accuracy of this classification method, a subset of ion images was shown to a collaborator, who decided whether to label each image as a real signal or as noise. These labels were treated as so-called ground truth. Then, the same images were classified automatically. Using ground truth labels, numbers of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) were counted. Accuracy was calculated as a fraction of correct predictions among all predictions: $Accuracy = \frac{TP+TN}{TP+FP+TN+FN}$. The advantages of this classification approach were the ease of result interpretation and 70-80 per cent classification accuracy. Nevertheless, the necessity to choose metric thresholds manually for every data set was a disadvantage, because it

was time-intensive and introduced bias in the analysis.

Ion image classification with machine learning

Several improvements on the original classifier allowed abandoning manual threshold selection and improving classification accuracy to over 95 per cent.

Firstly, instead of setting thresholds and then calculating accuracy using a ground truth image set, a machine learning approach was used. In this approach, ground truth image labels and metrics were fed into the algorithm, which trained a classification model. Then, this model was applied to the unseen data. I.e. for every new image, it could predict the probability of the ion being detected (rather than noise).

Secondly, when calculating classification accuracy, a five-fold cross-validation procedure was used. This means that the ground truth image set was split into a training set and a test set. Each time, a model was trained using the training set, then this model was used to predict labels of the test set, and accuracy was calculated using the predicted labels of the test set. This process was repeated five times and resulting accuracy values were averaged, giving a more reliable estimate of model performance.

Thirdly, the machine learning algorithms are capable of combining information from many metrics when training a model. 4 machine learning algorithms and 20 metrics were evaluated to maximise classification accuracy. The best performing algorithm was CatBoost[70] and the list of metrics was expanded to also include:

- **Spot occupancy**. This metric ensured that ion images with too few nonzero pixels in the metabolite spot were less likely to be classified as a real signal.
- **Average ion intensity in spot normalised by total ion count**. Including spot intensity in the model improved the classification accuracy. Instead of using the original intensity, the normalisation by total ion count was introduced to make the model applicable

to all data sets acquired on different instruments with different protocols.

- **Average ion intensity in all other spots**. This metric was added so that ion images with localised signals in multiple spots were less likely to be classified as a real signal.

Ion detectability and intensity metrics

Once all the ion images were classified with a machine learning model, two additional metrics, which were needed for comparing the detectability of metabolites, were calculated.

- **Background-corrected spot intensity**. This is a difference between the average ion intensity inside the spot and the average intensity of the metabolite outside of any spot. This metric was calculated to be able to compare the intensities of ions even when the ion image contains uniform background noise.

- **Prediction label**. Ion image was labelled as a real signal if its probability value from machine learning was 80 per cent or higher, otherwise, the image was labelled as noise. 80 per cent cutoff was chosen because for the downstream analysis a highly specific classifier (few false positives = fewer noise ion images classified as a real signal) is preferred over a highly sensitive classifier (few false negatives = fewer good ion images classified as noise), as shown in figure 2.11

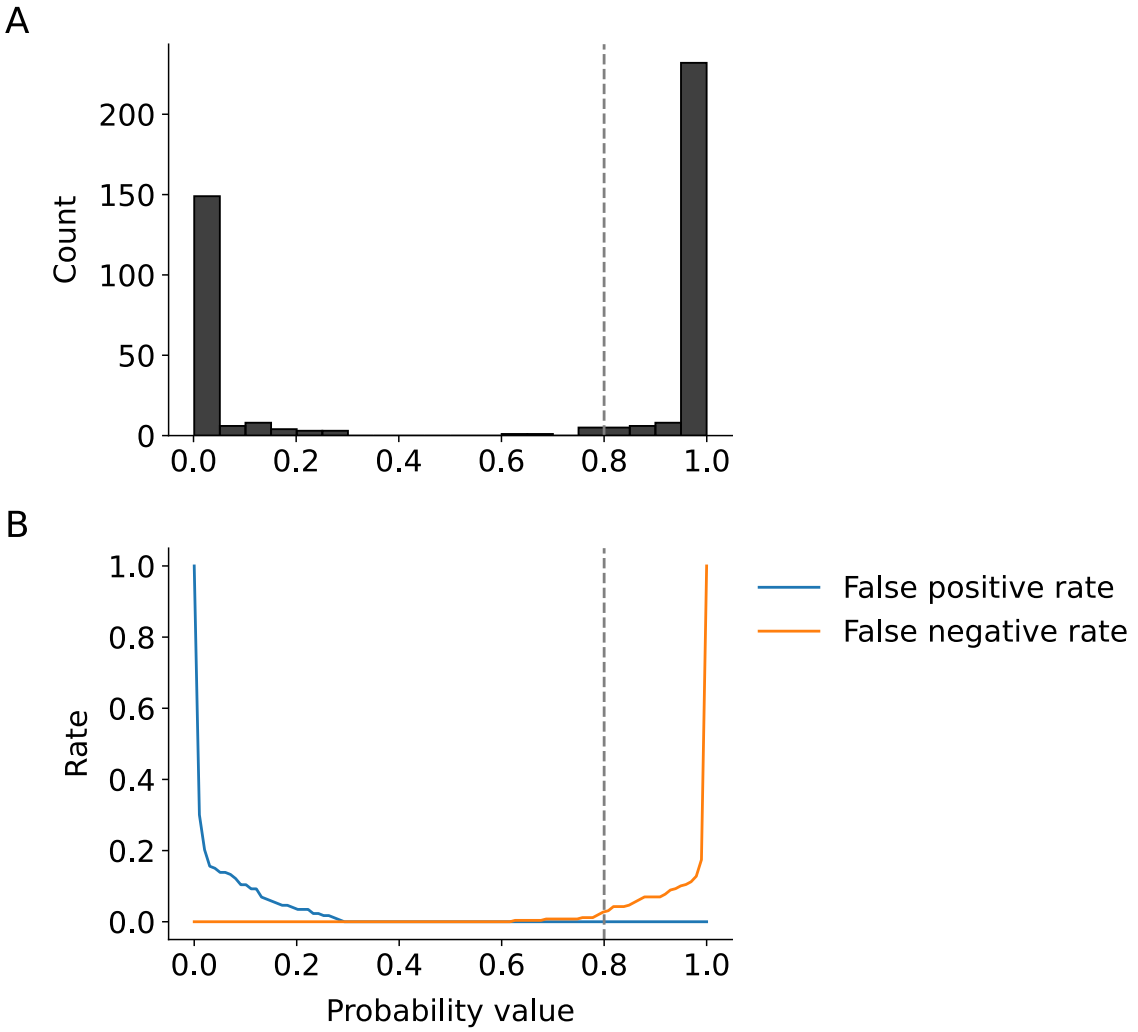


Figure 2.11 A) A histogram of probability values predicted by the machine learning model for a labelled set of ion images. The dotted line shows an 80 per cent cutoff for the probability value. When the prediction value is above the cutoff, ion images are labelled as a real signal. B) The higher the cutoff, the fewer false positives and the more false negatives result from classification. False positive rate = $\frac{FP}{(FP+TN)}$. False negative rate = $\frac{FN}{(FN+TP)}$.

2.3 Results and discussion

2.3.1 Comparison of MALDI-MS imaging protocols

Numerous sample preparation protocols are used for MALDI-MS imaging. Twelve established protocols from the literature were selected as described in section 2.2.3. The detection of metabolites comprising the standard sample was evaluated using these protocols, and the results are discussed in this section.

What is overall metabolite detectability?

When asking the question ”Are these metabolites detectable by a protocol/method of interest?”, there is no resource to go to for getting a quick answer. The researcher must look for relevant publications or go through the data sets deposited in public databases. Analysing the standard sample developed in the course of this work provides information on the detectability of 172 metabolites at once.

A metabolite was considered to be detected, if at least one of its ions was given a prediction label of a real signal (as discussed in section 2.2.4, subsection ”Ion detectability and intensity metrics”). The following ions were considered: $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$, and $[M]^{\cdot+}$ for positive ionisation mode; $[M-H]^+$, $[M+Cl]^-$, and $[M]^{\cdot-}$ for negative ionisation mode. Among all methods, the detectability of metabolites in the standard sample was ample: 91 per cent in positive ionisation mode and 82 per cent in negative ionisation mode. In both polarities, half of the metabolites were detected by eight or more methods (see figure 2.12). Among 11 metabolites that were not detected at all, some were obscured by background signal from isobaric MALDI matrix peaks (for example, lactic, butyric and acetoacetic acids), some were not found in the compound solutions by LC-MS, possibly due to hydrolysis or preparation error (phosphoserine, dihydroxyacetophenone), others were detected in the compound solutions by LC-MS, but were not detected

by MALDI-MS in any ion form that was considered (arachidonic acid, carbamoyl phosphate, cholesteryl acetate).

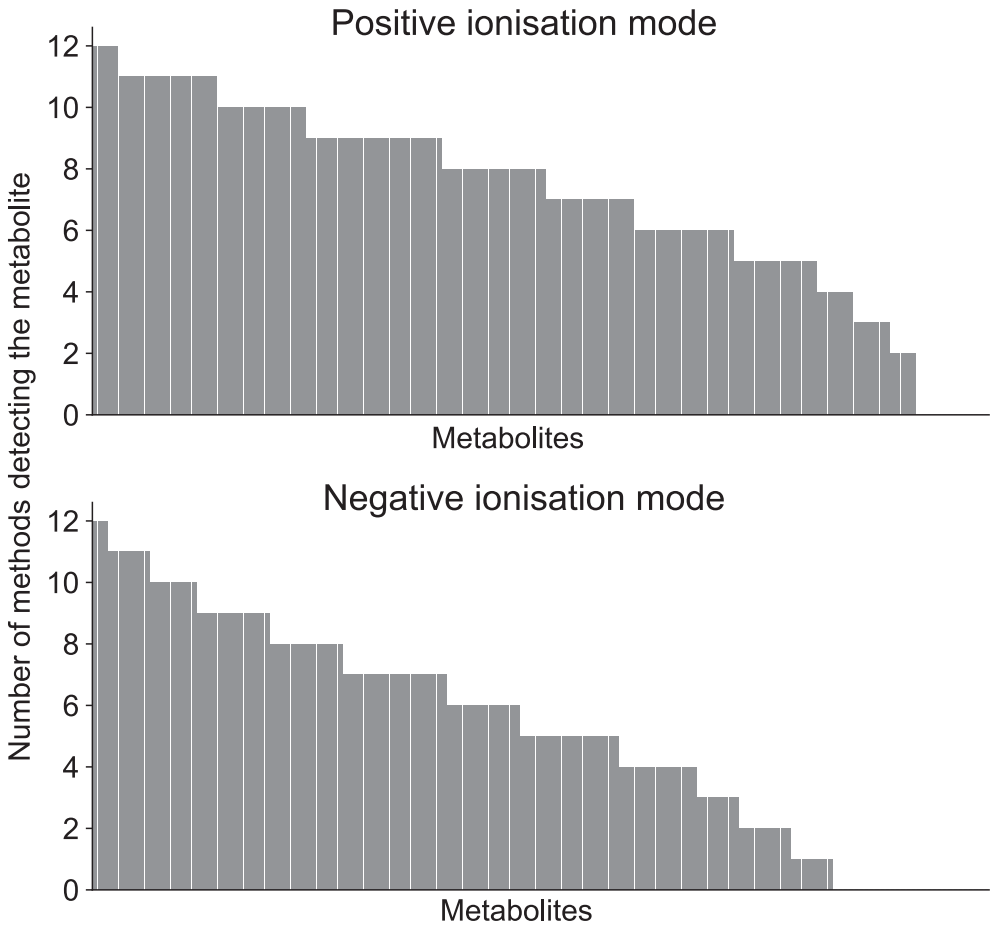


Figure 2.12 The number of methods detecting each metabolite from the standard sample in positive and negative ionisation modes. One bar represents one metabolite.

As shown in figures 2.13 and 2.14, in positive ionisation mode, most chemical classes of metabolites were well-detected. In negative ionisation mode, as expected, molecules that could not take on a negative charge were lacking detections: amines, triglycerides and positively charged metabolites (such as cholines, carnitines, vitamin B1 and nicotinamide riboside). Monoglycerides and diglycerides that rarely ionise as deprotonated species in LC-MS, were also not detected here. These results showed that generally, MALDI-MS imaging is suitable for detecting a wide range of metabolites.

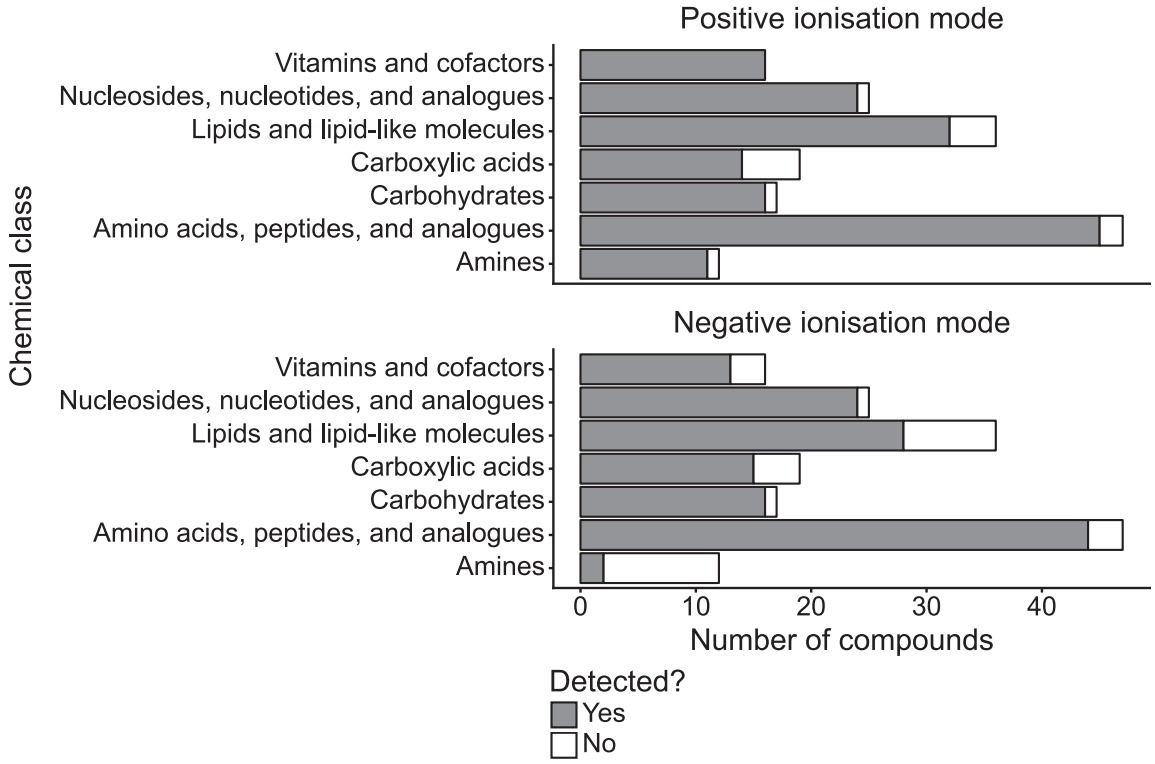


Figure 2.13 Numbers of total and detected metabolites per chemical class in positive and negative ionisation mode.

What are the differences between the methods?

When comparing individual method results, it became apparent that the sample preparation protocol choice is important for a successful experiment. Figure 2.15 shows that the number of detected metabolites by one protocol varied between 10 and 150. In positive ionisation mode, classic matrices for positive ionisation mode, DHB and CHCA, were among the best performers. ClCCA, a chlorinated CHCA analogue, detected a similarly high number of metabolites. The other two positive mode matrices, DHAP and CMBT, lagged by 20-30 metabolites. In negative ionisation mode, 9AA detected the most compounds. The results for the dual-mode matrices DAN, NOR and pNA were comparable to single-mode matrices, with pNA performing slightly better in positive mode, and DAN and NOR performing better in negative mode. Choosing dual-mode matrices thus might be beneficial for samples that can be analysed in both polarities. NEDC, MAPS and PNDI protocols led to the detection of fewer than 50 metabolites in either ionisation mode, which makes them less promising candidates for untargeted metabolite imaging, at least with the matrix application protocols used in this work.

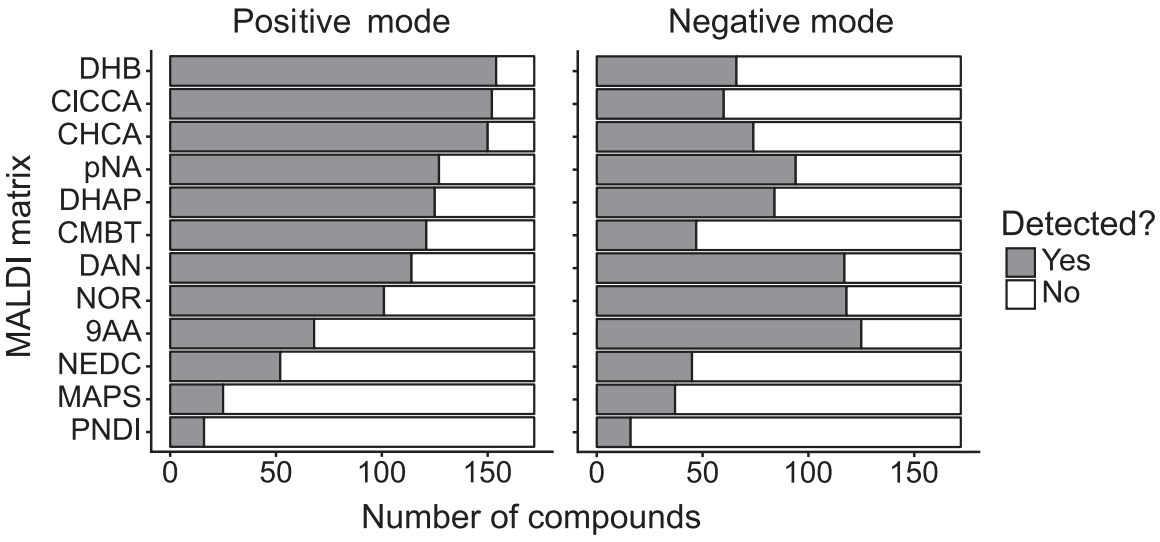


Figure 2.15 Numbers of detected metabolites per MALDI matrix sample preparation protocol in positive and negative ionisation modes.

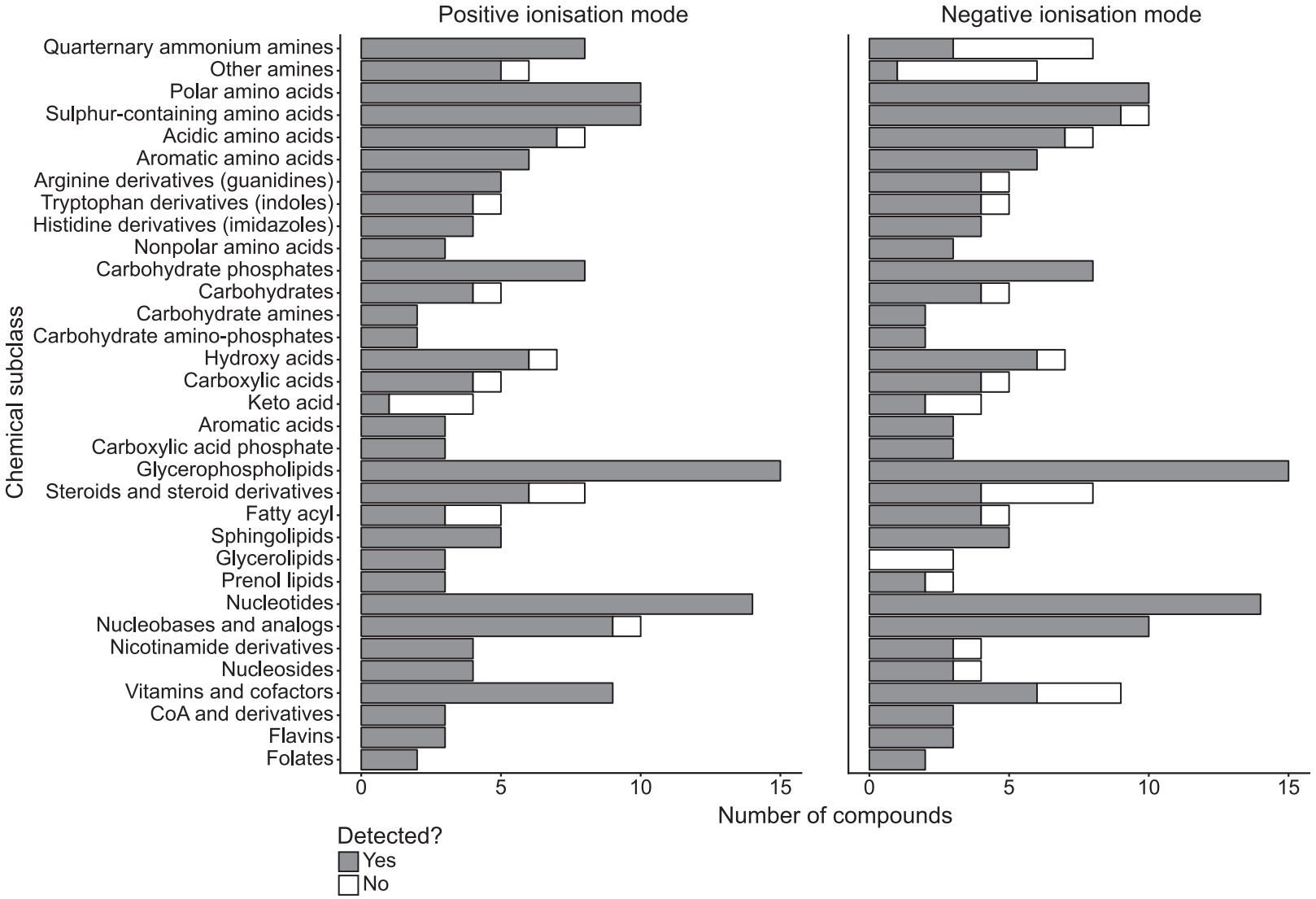


Figure 2.14 Numbers of total and detected metabolites per chemical subclass in positive and negative ionisation modes.

Which molecules are detected by which method?

To choose the matrices allowing to achieve the best metabolite coverage, matrix detection specificity of molecules from individual chemical classes needs to be considered. Figures 2.16 , 2.17 and 2.18 show the fraction of detected metabolites in each chemical class and subclass, as well as the average intensity of detected metabolites (the intensity metric was calculated as discussed in section 2.2.4, subsection ”Ion detectability and intensity metrics”). A summary of findings per chemical class follows below.

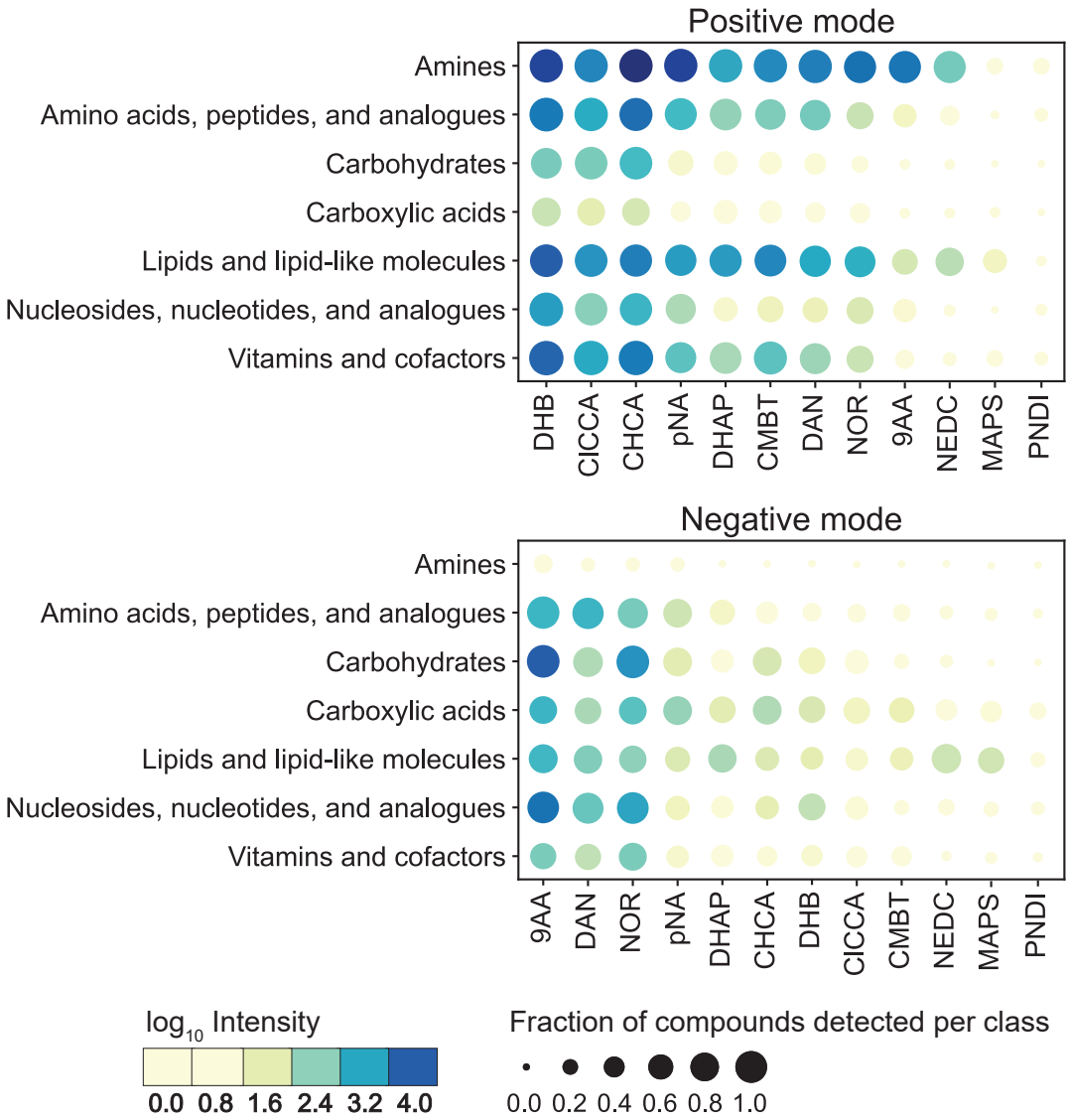


Figure 2.16 Detection of metabolites in each chemical class by different MALDI-MS methods. Radius of the dot shows what fraction of compounds in the given class was detected as any ion form. The colour of the dot represents average intensity among the detected compounds.

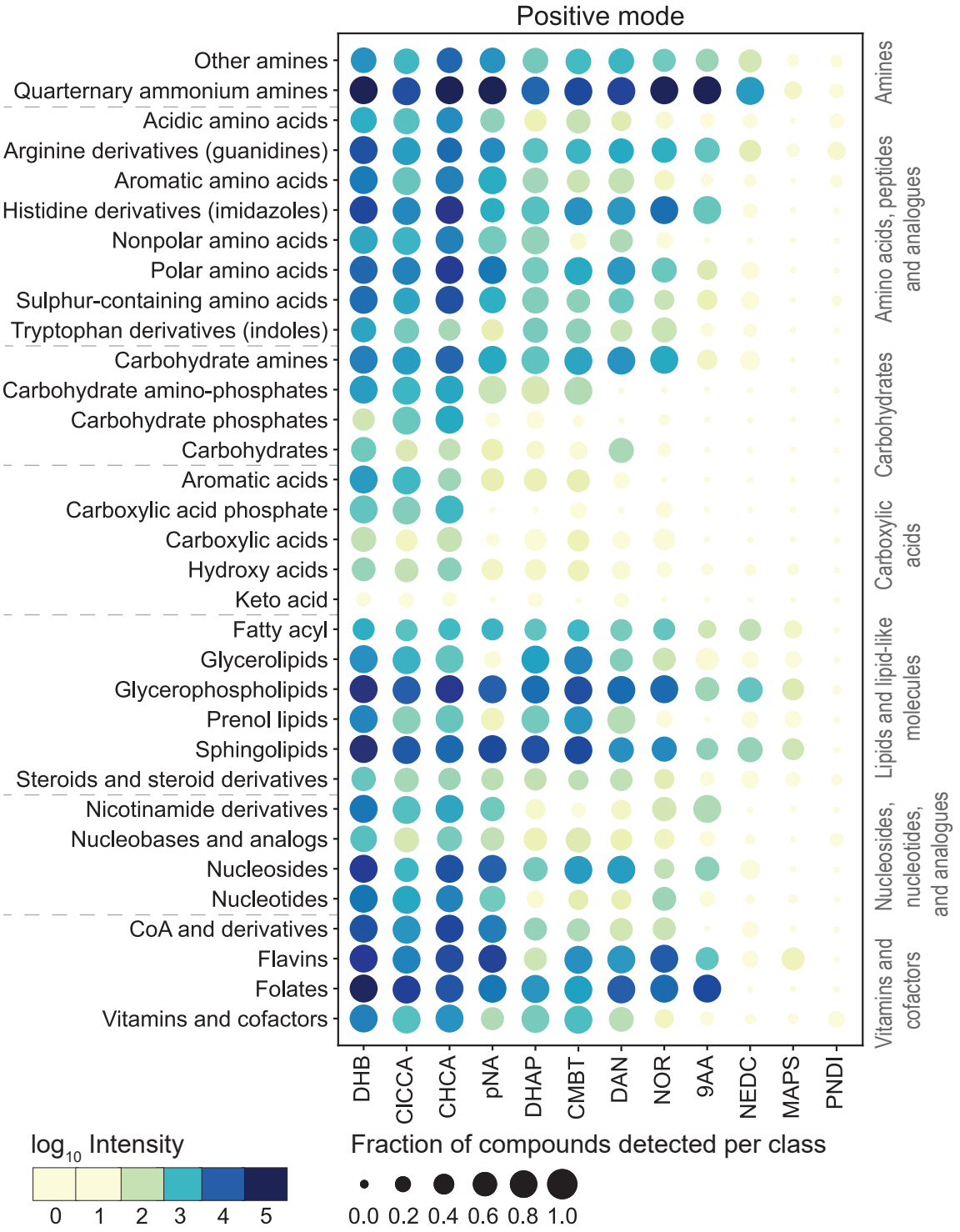


Figure 2.17 Detection of metabolites in each chemical subclass by different MALDI-MS methods in positive ionisation mode. The radius of the dot shows what fraction of compounds in the given class was detected as any ion form. The colour of the dot represents the average intensity among the detected compounds.

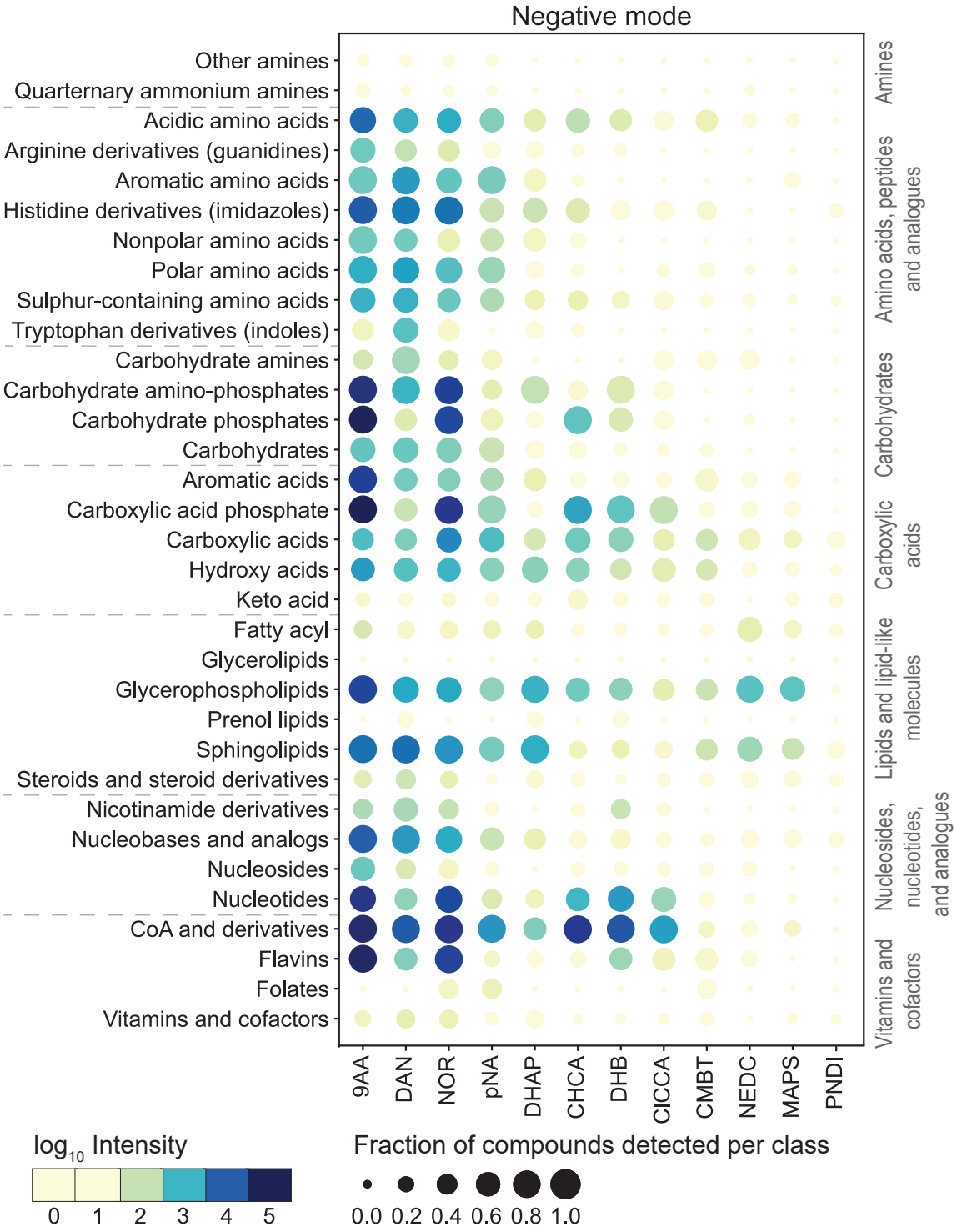


Figure 2.18 Detection of metabolites in each chemical subclass by different MALDI-MS methods in negative ionisation mode. The radius of the dot shows what fraction of compounds in the given class was detected as any ion form. The colour of the dot represents the average intensity among the detected compounds.

Amines The class ”Amines” includes polyamines, such as putrescine, amino acid-derived amines, such as histamine, and quaternary ammonium compounds, such as choline. Unsurprisingly, the detection of quaternary ammonium metabolites, which bear a positive charge, was very strong in positive ionisation mode as $[M]^+$ ion. The same applied to acylcarnitine and phosphatidylcholine (PC) in the ”Lipids...” class, and carnitine and trimethyllysine in the ”Amino acids...” class. Uncharged amines are readily protonated at the amine group. The amine group itself is unlikely to be deprotonated, and thus in negative mode, amines with no other major charge-carrying moieties are almost not detected. In very few cases, some amines can form a negatively charged ion with chloride adduct, for example, $[histamine + Cl]^-$ but such signals are low in intensity. Positive ionisation mode with any matrix among the ones considered in this work, except for MAPS and PNDI, is a method of choice if amines are of interest.

Amino acids, peptides, and analogues Amino acids carry both amine groups, prone to protonation and carboxylic acid groups, prone to deprotonation. Consequently, these metabolites were expected to be readily detectable in both ionisation modes. The expectations were fulfilled with mainly protonated and sodiated ions detected in positive mode, and predominantly deprotonated ions detected in negative mode. Metabolites of this class were further subdivided into subclasses, mainly by side-chain type. Only moderate differences were observed when comparing differences in both ionisation modes. Most notably, arginine and its derivatives, such as creatine, which carry a guanidine group that easily acquires a positive charge, were better detected in positive mode. Acidic amino acids, which have an additional carboxylic acid group, generated more ion intensity in the negative mode. ClCCA, an analogue of CHCA with chlorine substituent, meant to improve ionisation of peptides, yielded lower ion intensities than CHCA in this and other metabolite classes.

Carbohydrates Intact carbohydrates were better detected in negative ionisation mode, especially with 9AA and NOR matrices, mainly as $[M-H]^-$ and $[M+Cl]^-$ ion species. In positive mode, some sodiated ions were observed, while most of the signal was lost to fragmentation. Protonation of a hydroxyl group makes it into a good leaving group, which leads to the generation of fragment ions $[M+H-H_2O]^+$ions.

Carboxylic acids Carboxylic acids, such as succinic acid, were better detected in negative mode as deprotonated ions, although some signal was lost to decarboxylation, i.e. some $[M-H-CO_2]^-$ ions were observed. Lower in intensity, carboxylic acid ions were also observed in positive mode as sodiated species, as the sodium cation has a suitable radius to associate with the carbonyl group dipole. Only one metabolite in the "Keto acids" subclass was detected, likely because other metabolites were isobaric with matrix peaks, and their respective ion images did not pass machine learning classification.

Lipids and lipid-like molecules This class contains very diverse metabolites, thus it is best to consider individual subclasses if not individual metabolites, to gain a good insight into their detection. Overall, the variety of lipid subclasses in which metabolites were detected was broader in the positive mode. For example, glycerolipids, prenol lipids and fatty acyls were almost not detected in negative mode. Within the subclass "Steroids and steroid derivatives", more variety was represented in positive than in negative mode. However, steroids with acidic functional groups (taurocholic and cholic acids) or sugar attached (glycosyl cholesterol) could also be detected in negative mode. This represents the overall trend that also applies to the subclasses of glycerophospholipids and sphingolipids - detection of the ion depends on the functional groups present and not on the larger, hydrophobic part of the molecule. For example, ions of phosphatidylinositol (PI), a phospholipid carrying sugar phosphate as a lipid head group that readily forms a stable anion, are much more abundant in negative ionisation mode. By

the same logic, ions of sulphatide galactosylceramide (Sulpho Gal Cer), a sphingolipid carrying sulphonated sugar, was ionised predominantly in negative ionisation mode.

Nucleosides, nucleotides, and analogues Overall, metabolites of this class were detected in both ionisation modes, although some preferences for stronger ionisation were apparent among the chemical subclasses. Nicotinamide derivatives, such as NAD, were better detected in positive ionisation mode, likely due to a positive charge on the pyridine nitrogen. Nucleobases were detected in both ionisation modes, mainly as deprotonated, or protonated or sodiated ions. A negative charge could be stabilised through an aromatic ring or delocalisation into a carboxyl group, while the positive charge could be due to protonation of an amine group, or sodium cation association with carbonyl dipole. Nucleotides, especially those bearing multiple phosphate groups (such as UDP-glucose, CoA, FAD) were preferentially ionised in negative ionisation mode, as deprotonated species. A few nucleosides considered, on the contrary, were better detected in positive ionisation mode, in the protonated form. This may differ on nucleoside to nucleoside basis depending on which nucleobase hydrogen is replaced by a bond to the sugar moiety.

Vitamins and cofactors The class combined metabolites with more rare chemistries that are biologically relevant. While it is not possible to make any general statements about vitamins and cofactors, among the metabolites chosen for the standard sample, vitamins and folates were detected predominantly in positive ionisation mode.

Which method to choose?

In positive mode, the results for the top three matrices, DHB, CHCA and ClCCA, were very similar. These matrices demonstrate broad metabolite coverage and high ion yield for most of the chemical classes except carboxylic acids and carbohydrates, where small fraction metabolites are detected and with lower ion response. In negative mode, the range

of covered metabolite classes was narrower. 9AA and two dual-mode matrices DAN and NOR provided the best coverage and ion response.

The ultimate choice of the method for untargeted metabolite imaging depends on additional factors: spatial resolution, ionisation source, time and sample availability.

- For high spatial resolution imaging (i.e. pixel size of 5-10 micrometres), it is crucial to use matrices with crystal size not exceeding the desired spatial resolution. This is because the ionising laser supplies the energy to the entire crystal at once and the whole crystal then desorbs. If the crystal is 50 micrometres long, for example, the use of a 5-micrometre laser raster step is futile. Out of the matrices tested in this work, CHCA, ClCCA, DHAP, pNA and DAN have the crystal size required for high spatial resolution imaging.

- In this work atmospheric pressure ionisation source was used, however, some MS imaging instruments operate under intermediate pressure or vacuum. For such instruments, it is important to consider the vacuum stability of the matrix. In particular, pNA and DHAP may not be suitable for long acquisition runs under high-vacuum conditions as these matrices could evaporate and cause some artefacts in the data due to changing matrix to analyte ratio.

- If time allows, and enough sample is available, for example, as consecutive sections, it is best to use separate methods in positive and negative modes to get the best metabolite coverage. If time is a limiting factor, and the broadest coverage of metabolites is desired with no particular preference for detected chemical classes, it is best to use positive ionisation mode. If sample availability is the limiting factor, the best strategy could be to use a dual-mode matrix, such as DAN, NOR or pNA, and image the same sample twice in both polarities, offsetting the sampled locations by the diameter of a laser ablation mark.

To facilitate the choice of the method for targeted analysis of chemical classes, biological pathways or specific metabolites of interest, an interactive web application was designed that allows browsing and filtering of the collected data. This is discussed in more detail in section 2.3.4.

2.3.2 Biological relevance of the findings

MALDI-MS for individual metabolic pathways

MS imaging experiments are usually driven by a biological question, where the researcher might have more interest in certain biochemical pathways than in molecules of the same chemical class. As explained in section 2.2.1, metabolites comprising the standard sample were chosen keeping this in mind. To compare the suitability of MALDI-MS for studying individual metabolic pathways, a similar analysis as the one discussed in the previous section was completed, this time changing how metabolites were grouped on the Y-axis.

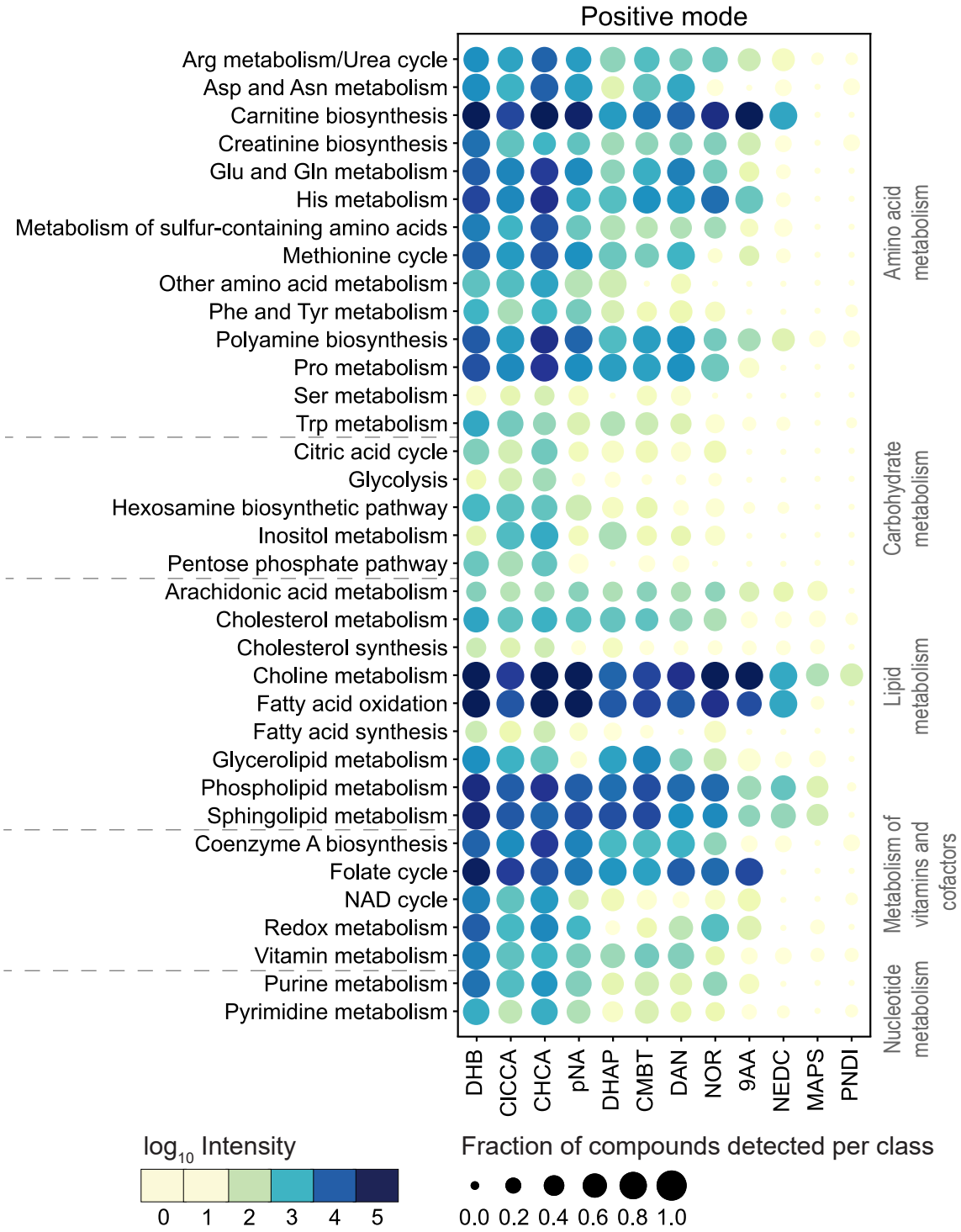


Figure 2.19 Detection of metabolites in each biochemical pathway by different MALDI-MS methods in positive ionisation mode. The radius of the dot shows what fraction of compounds in the given class was detected as any ion form. The colour of the dot represents the average intensity among the detected compounds.

In positive ionisation mode (figure 2.19), the overall coverage of amino acid metabolism was good except Ser metabolism, which was lacking in both polarities. The positive mode was also preferred over the negative mode for metabolic pathways related to arginine metabolism. These include:

- the urea cycle, that converts toxic ammonia to urea in the liver and kidneys, for further excretion (detected were Arg, ornithine, citrulline and argininosuccinate);
- the nitric oxide production that happens during the inflammatory response of macrophages;
- biosynthesis of polyamines (detected were putrescine, spermidine, spermine), that are important for development, growth and tissue repair;
- creatine metabolism (detected were creatine, creatinine), important for ATP turnover in muscle and brain.

The use of the positive ionisation mode was advantageous for looking at the one-carbon cycle, consisting of the folate cycle (detected were folate and tetrahydrofolate) and methionine cycle (detected were methionine, homocysteine, SAM and SAH). The one-carbon cycle is a source of one-carbon units (methyl groups). It is important for nucleotide biosynthesis, epigenetic regulation, and redox metabolism. Two vitamins associated with one-carbon metabolism: riboflavin and pyridoxine, were also better detected in positive mode.

The positive mode could be useful for studying fatty acid oxidation. Long-chain fatty acids are imported into mitochondria where they are broken down into acetyl-CoA units. Fatty acid transport is mediated by carnitine, through the formation of acylcarnitines. Among the detected metabolites representing this process were carnitine, its precursor trimethyllysine, palmitoylcarnitine, propionylcarnitine, and palmitoyl-CoA.

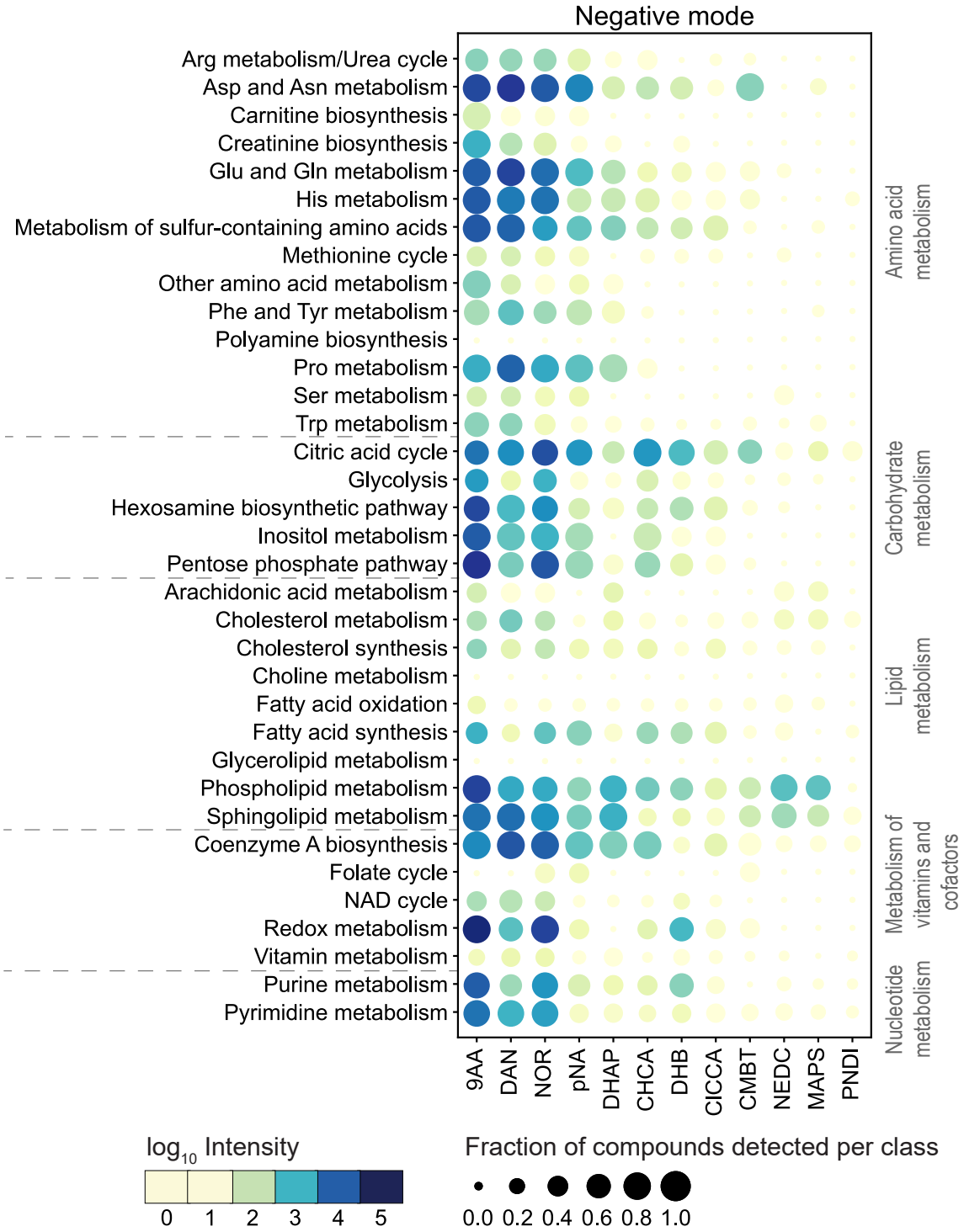


Figure 2.20 Detection of metabolites in each biochemical pathway by different MALDI-MS methods in negative ionisation mode. The radius of the dot shows what fraction of compounds in the given class was detected as any ion form. The colour of the dot represents the average intensity among the detected compounds.

The negative ionisation mode (figure 2.20) was found to provide better coverage of carbohydrate metabolism than the positive mode. In particular, for glycolysis (detected were glucose, glucose-6-phosphate, fructose-1,6,-biphosphate, 3-phosphoglycerate, phosphoenolpyruvate), citric acid cycle (detected were cis-aconitate, citrate, alpha-ketoglutarate, succinate, fumarate, malate, and acetyl-CoA), and for the pentose phosphate pathway (detected were phosphogluconic acid, ribulose-5-phosphate and ribose-5-phosphate) - which form the core of the biochemical pathways.

The hexosamine pathway was also well-covered in negative mode. This pathway generates a nucleotide sugar UDP-GlcNAc, which is involved in glycoprotein synthesis. UDP-GlcNAC and several other intermediates in this pathway were detected. Additionally, the purine and pyrimidine metabolism intermediates were better detected in positive than in negative mode. This might provide an additional insight into the same pathway when the metabolites leading up to UDP biosynthesis are detected.

Lastly, the negative mode was preferable to the positive mode for redox metabolism, for example, the detection of NAD, NADH, FMN and FAD was high.

Comparison between standard samples and biological samples

The analysis of pure chemical standards has advantages over the analysis of biological samples, such as simplicity of result interpretation and avoiding signal suppression by other analytes. It provides information about metabolite detectability under ideal conditions. While it is reasonable to expect that the metabolite detectability would be similar in biological tissues, it is not clear how much difference could be expected. To address this question, a comparison of metabolite detectability between standard sample data sets and biological tissue data sets available in the host laboratory was done.

For this, two replicates of standard sample data sets analysed with DHB matrix in the positive mode, and DAN matrix in the negative mode, were selected. These matrices are used most often in the host laboratory and thus have the most tissue data sets available for comparison. Only metabolites with mass above 100 Da were included in the analysis since lower masses were not frequently analysed with biological samples. 30 biological data sets were selected in positive mode including healthy tissues from wild type mice: intestine, brain, lungs, liver; tumour tissue: lungs; inflamed tissue: liver. 12 biological data sets were selected in negative mode: healthy tissues from wild type mice: intestine, brain, lungs; tumour tissue: lungs. The fraction of metabolites detected in each chemical subclass (shown previously in figure 2.1) was calculated separately for both standard sample replicates, and for all tissue data sets. To determine whether metabolites were detected in tissue data sets, first, ion images corresponding to the main adducts of every metabolite were generated: $+H^+$, $+Na^+$, $+K^+$ for positive ionisation more and $-H^+$, $+Cl^-$ for negative ionisation mode were considered. Then, the ion images were manually separated into the ions that correspond to the real signal, and the others that are background noise, by the researchers familiar with the data. Between 20 to 30 thousand metabolite images were classified this way. Metabolites that had at least one ion classified as detected were considered as detected in the corresponding ionisation mode.

The hypotheses held before this analysis were as follows.

1. Ions detected in tissue must be detectable in the standard sample.
2. Ions that are not detected in the standard sample cannot be detected in the standard sample.

Both hypotheses are based on the assumption that analysis of the pure chemical standards provides information about the detectability of the compounds under ideal conditions, and that ionisation of compounds in tissue is worse than in pure standards.

Figure 2.21 shows results of the analysis. In the positive mode with the DHB matrix, compounds from all chemical subclasses were more detectable in the standard samples. This was expected since some metabolites might not be present in the selected biological tissues, or they could fall below the limit of detection. In both modes, there were a few chemical classes, all metabolites of which were uniquely detected in the standard sample for this reason. In the negative mode, the results were the same for most of the chemical subclasses. Some subclasses were more detectable in biological tissues than in the standard sample. Others were only detected in biological samples and not in the standard sample. This could happen because of one limitation of the analysis. The ion images were generated using m/z values of metabolites of interest. In the standard sample, these correspond only to the compound deposited on the glass slide, however in tissue data sets, matches to isomers and isobars cannot be ruled out. Thus it is believed that the red and blue data points are misplaced because a few isomers or isobars to the target metabolite were abundant and localised in the tissues. Isobars could be excluded by considering the isotope pattern of the detected compounds, but the isomers could not be ruled out unless MS/MS analysis is carried out.

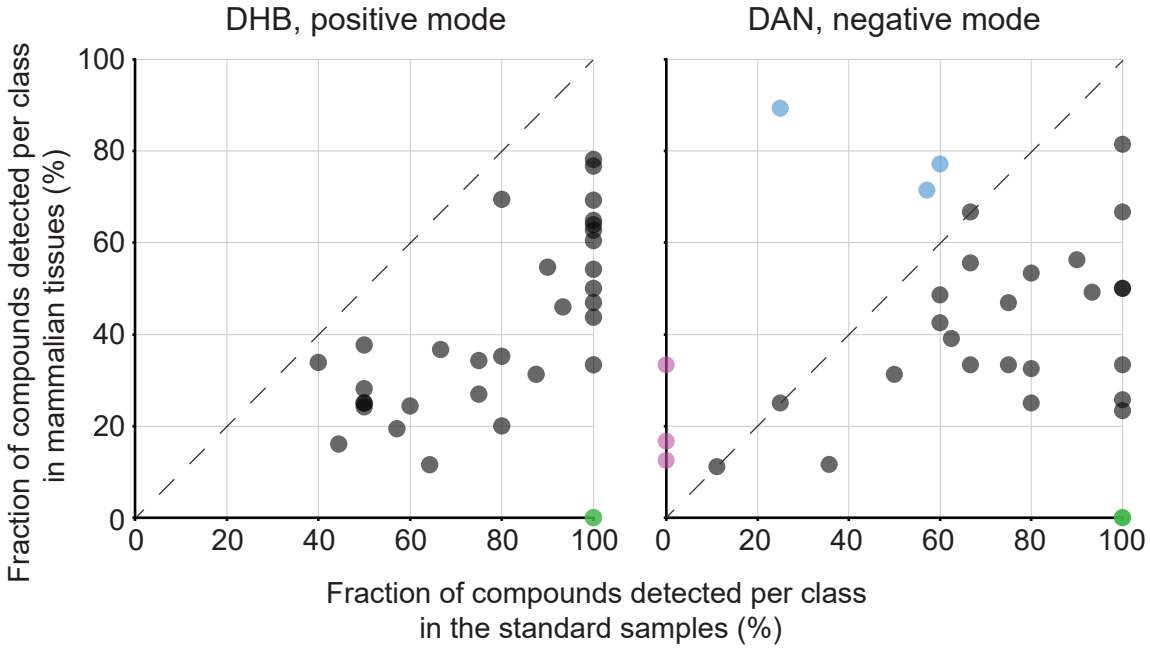


Figure 2.21 Comparison between metabolite detectability in the standard samples and biological samples. Each dot in the plot represents one chemical subclass from figure 2.1. The dashed diagonal line shows y=x. The chemical subclasses that fall on the diagonal are detected equally well in the standard samples and biological samples. The chemical subclasses that fall under the dashed line are more detectable in the standard sample, and the ones that fall in the upper triangle are more detectable in the biological data sets. Subclasses, shown in blue, were more detectable in biological tissues than in the standard sample. Others, shown in magenta, were only detected in biological samples and not in the standard sample. Subclasses shown in green were detected uniquely in the standard samples and not in biological data sets. Analysis was performed together with L. Vieira. M. Ekelöf contributed to ion image classification.

Overall, these results have demonstrated that the results on metabolite detectability obtained using standard samples could be broadly applicable to any biological sample of interest. Once more biological data sets are available, it would be beneficial to repeat this analysis including all MALDI matrices considered in this work for completeness.

2.3.3 Inter-laboratory survey

There are a variety of MS imaging technologies that could be used for untargeted metabolite imaging. To assess the metabolite detectability by these technologies, several leading laboratories in the field were invited to participate in the study, to analyse the developed standard samples. Nine instruments, in addition to the one available in the host lab, were selected, as described in section 2.2.3. The collaborators were asked to use sample preparation protocols and instrument settings that they commonly used for biological samples, to produce the data in positive and negative ionisation modes. In addition, two data sets from the host lab, matching this criterion, were selected for comparison.

The data received from collaborators were processed the same way as the data from the host lab. In contrast to the analysis presented in the previous sections, the intensities of the same metabolites could not be directly compared between different instruments. This is due to differences in the sample preparation protocols and sampling between technologies. An avenue for future work could be to find a suitable normalisation strategy that makes the intensities comparable, for example, total ion count normalisation. In the meantime, the detectability of metabolites per chemical class was compared, as shown in figure 2.22.

Overall, the data obtained among all participants showed broad metabolite detectability. The observations made in the previous sections from the comparison of different MALDI protocols were also apparent in this data. For example, amines were among the most detected metabolites in positive ionisation mode, and among the least detected in negative mode. For most technologies, carboxylic acids were better detected in negative ionisation mode, than in positive.

The results generated in the host laboratory using AP-MALDI-Orbitrap were as good as the results obtained by the majority of collaborators.

Between the two AP-MALDI-Orbitrap instruments in two different laboratories, the differences were minor. The results for atmospheric pressure AP-MALDI and vacuum MALDI (MALDI qTOF, MALDI FTICR) were very similar, with an exception of one data set in positive ionisation mode, where detectability was slightly lower in every chemical class. MALDI-2 performed best in positive ionisation mode with Orbitrap mass analysis. While it may be superior to regular MALDI for imaging biological samples at very low sampled volumes, it did not detect more metabolites in the standard sample than other techniques. This means that when the local concentration of the metabolite in the sample is not a limiting factor, its chemistry is. In negative mode, MALDI-2 detected the least metabolites in every chemical class. This might be because MALDI-2 provided too much ionising power and fragmented the metabolites, or because the protocol was not well-tuned for small molecule detection.

DESI detected more metabolites in negative mode than in positive mode. The two DESI instruments showed comparable results in positive ionisation mode, but different results in negative mode. As DESI could sometimes delocalise the analytes with its ionising droplet beam, the differences could be in part artefacts from automatic ion image analysis. A further investigation is needed to verify this. Detection of metabolites by MALDESI was most similar to MALDI in positive ionisation mode, and DESI in negative ionisation mode. In both polarities, MALDESI was among the best-performing methods.

Considering that each laboratory has used a different instrument and was not restricted in the method of choice, the similarity of obtained data is remarkable. These results show that overall, the results obtained in this study are comparable to those obtained by some of the leaders in the MS imaging field and relevant to its scientific community. In addition, the findings show that untargeted metabolite imaging could be easily accessible for all considered techniques. A detailed analysis of inter-laboratory data is in plans for future work. One aspect, however,

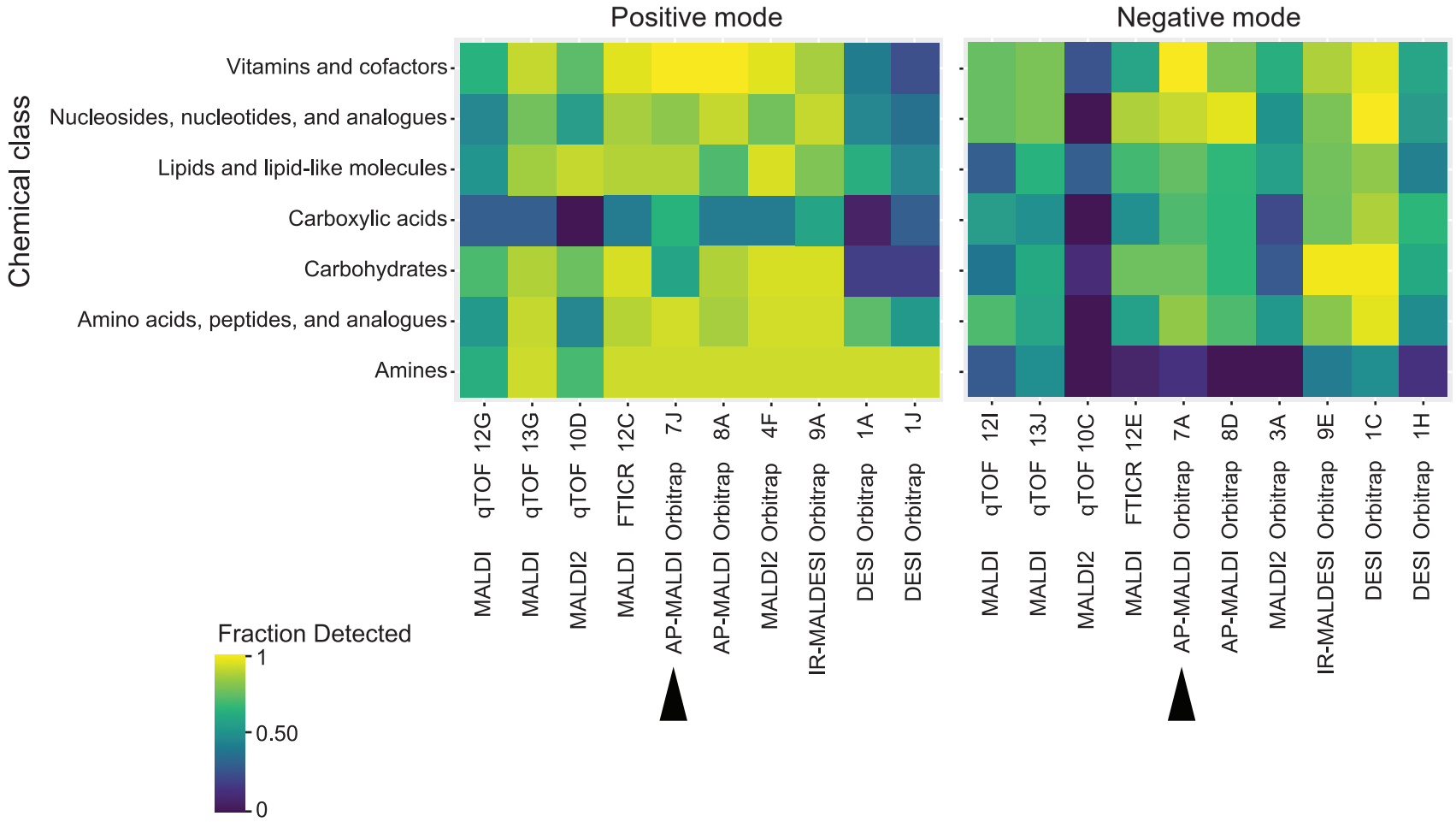


Figure 2.22 Comparison of metabolite detectability between different MS imaging technologies. The heatmap shows the fraction of detected metabolites per each chemical class by each instrument. Black triangles point at the data sets acquired in the host lab. This figure was prepared together with M.Ekelöf

has already been explored and is discussed in the section that follows.

Signal dilution

When analysing MS imaging data of metabolites, it is important to take into account signal dilution. By signal dilution here is meant a loss of signal due to the formation of multiple ions from one analyte (examples shown in figure 2.23. While for some molecules, $[M+H]^+$ or $[M-H]^-$ is the main and most intense ion, for some other molecules, like acetyl-CoA it is only $3/4$ of the more intense $[M+Na]^+$ ion. Signal dilution occurs not only through different adducts but also through neutral losses (fragmentation) or neutral gains (reduction or formation of matrix-analyte clusters). For example, a protonated ion of cholesterol was not detected and its sodiated ion appeared with relatively low intensity, as this metabolite predominantly ionises after a water loss. Other common losses include loss of phosphate, carboxylate (e.g. for orotic acid) and amine groups, as well as oxidation.

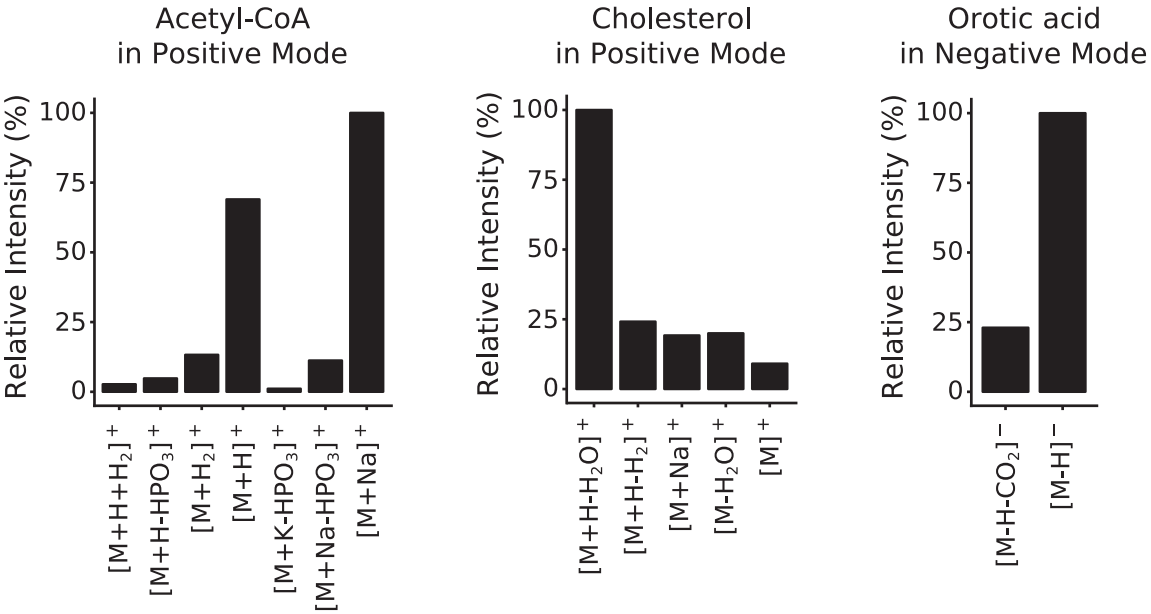


Figure 2.23 Examples of signal dilution. Relative abundances of detected cholesterol, acetyl-CoA and orotic acid ions. This figure was prepared by M. Ekelöf.

To gain a better understanding of the biggest contributors to signal dilution, and to answer the question of which ions are most abundant

for each metabolite class the following analysis was performed. First, the intensity of all detected ions per metabolite was summed. Next, the relative abundance of ions of one metabolite was calculated. Then, the results of each metabolite were averaged among the ten ”Interlab” data sets. Finally, the results were summarised per class as shown in figure 2.24.

- In the positive mode, the two main ions were $[M+H]^+$ and $[M+Na]^+$. The latter was the main ion form for lipids and lipid-like molecules, carbohydrates and carboxylic acids. For amino acids, nucleotides, vitamins, cofactors and related molecules, each ion form constituted only 30 per cent of the total detected signal. Potassiated ions were rarer. $[M]^+$ ions were observed for two groups of metabolites: metabolites that bear a positive charge in their database formula (such as vitamin B1 or choline), and metabolites that can eject an electron and form a stable radical, for example, vitamin E, which is a known radical scavenger and vitamin A, which has a long conjugated double bond system that could stabilise a radical by delocalisation. Loss of the amine group contributed to the signal dilution of amino acids and related metabolites. Similarly, loss of water was observed for certain lipids.

- In the negative mode, the dominant ion form was $[M-H]^-$. The exception was the class of amines, which overall was poorly detected in negative polarity, and predominantly yielded $[M+Cl]^-$ ions. Chlorinated ions were responsible for the signal dilution of carbohydrates, lipids and lipid-like molecules. For carboxylic acid, decarboxylation contributed substantially (ca. 10 per cent) to the signal dilution). Dephosphorylation was observed for nucleotides.

An additional observation that was made for MALDI-MS methods, was the formation of metabolite-matrix adducts, for example, $[M+matrix+H]^+$. This was previously reported for DHB, however in this work metabolite-adduct formation was observed for all matrices in both ionisation modes,

even underperforming NEDC, MAPS and PNDI matrices (for PNDI, adducts with monomer were observed).[71]

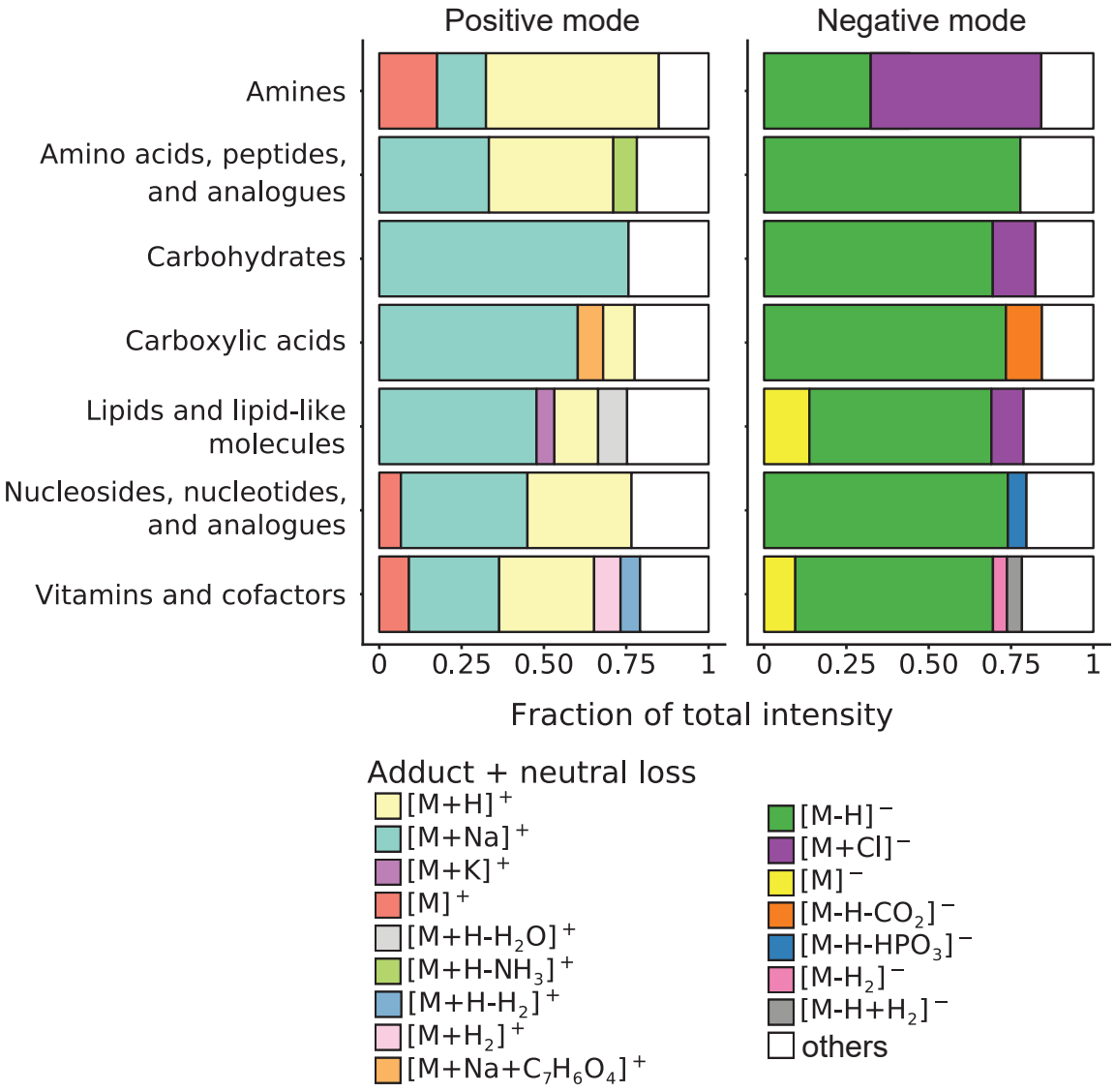


Figure 2.24 Composition of total detected signal per metabolite. Figure shows which ions contribute the most to the total detected metabolite intensity. This figure was prepared together with M. Ekelöf.

A previous study that compared the formation of protonated, sodiated and potassiated ions in MALDI-MS imaging among lipids, found that relative abundances of these ions for one compound could vary between the methods. [41] This is also expected for metabolites. Nevertheless, the trends observed by averaging results obtained on ten different instruments and from different laboratories are likely representative of

the chemistry of the analyte and are likely to hold for the majority of MS imaging methods.

To understand how much the signal dilution affects the results of the untargeted metabolite imaging, a comparison between different methods was done. Figure 2.25 shows what fraction of the total ion signal is explained by the most abundant ion. The results have shown that, on average, 65 to 85 per cent of total ion intensity was explained by the single most intense ion per metabolite, for all methods and in both ionisation modes. The obtained figures may be overestimated, because some signal dilution species might have gone unannotated (smaller fragments, unexpected neutral losses, etc.)

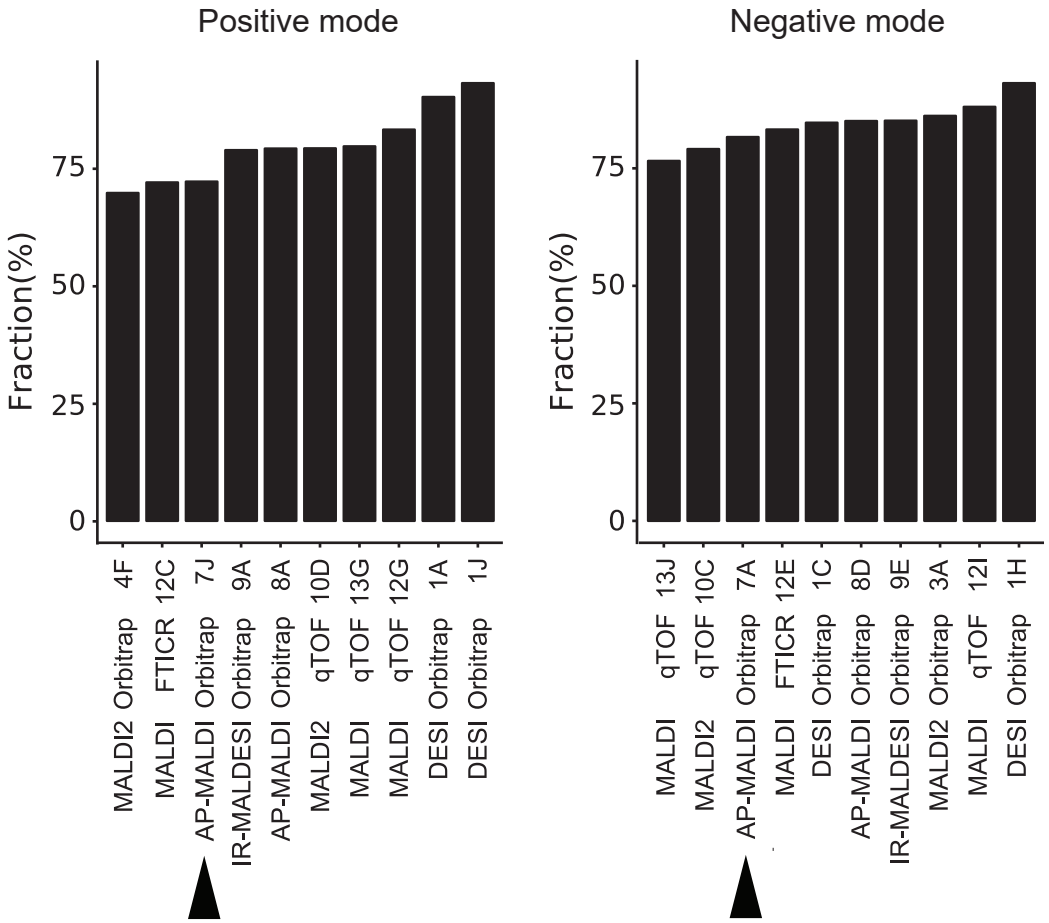


Figure 2.25 The effect of signal dilution on various MS imaging instruments. The chart shows the average abundance of the most intense ion compared to the abundance of all annotated ions detected per metabolite. Black triangles point at the data sets acquired in the host lab. This figure was prepared by M. Ekelöf.

To summarise, signal dilution was found to affect the detection of the majority of metabolites. For some metabolites, this effect was substantial and meant that more than one ion needs to be taken into account for correct result interpretation. When a broad detection of metabolites is of interest, however, it is sufficient to focus on the most abundant ions, if some information loss could be tolerated.

2.3.4 Result dissemination through an interactive application

In the course of this work, rich data on the detectability of 172 metabolites were collected considering: twelve MALDI-MS methods, two ionisation modes, over 35 ion forms per metabolite, and, in addition, ten MS imaging instruments in different laboratories. The previous sections have laid out the findings concerning the applicability of different methods to the untargeted imaging of metabolites. Additionally, the obtained data could serve as a valuable resource for the researchers interested in the targeted analysis, for example of a limited number of metabolites, chemical classes or biochemical pathways. For this reason, a prototype of an interactive web application was designed that allows browsing and filtering of the obtained results. This prototype was developed in collaboration (see the Contributions section for more detail).

The purpose of the application is to facilitate the dissemination of the results in the MS imaging community and to facilitate the choice of the method for targeted analysis of chemical classes, biological pathways or specific metabolites of interest. A screenshot of the application is shown in figure 2.26. In particular, the application can display two aspects of the pre-processed data per metabolite. The first is metabolite detectability, or prediction label (”real signal” or ”background noise”). The second is the average intensity of the metabolite in its spot, corrected by the background signal. Both metrics are described in detail in section 2.2.4. The user of the application can select any number of metabolites to view

on the Y-axis. Metabolites may be grouped by chemical (sub)class or biochemical pathway. On the X-axis, selected MALDI-MS methods, MS imaging instruments, or the ion forms contributing to the signal dilution can be shown. Additional filters allow further flexibility, for example, the ionisation mode of interest can be specified. Three main use cases are anticipated:

1. **Choosing the most suitable method for one metabolite of interest.** Here the application could show the total intensities of all detected ions per metabolite, separately for both ionisation modes, and all methods. The data could be presented in the heatmap form, similar to figure 2.22. Going forward, integration of the application with METASPACE[69] would be considered to show the user individual ion images for each ion on demand.

2. **Identifying the most abundant ion for a metabolite and a method of interest.**

3. **Choosing the most suitable method for a selection of metabolites, such as a molecular class or a metabolic pathway.** In this case, both detectability and intensity information could be shown, similar to how it was presented in figure 2.16.

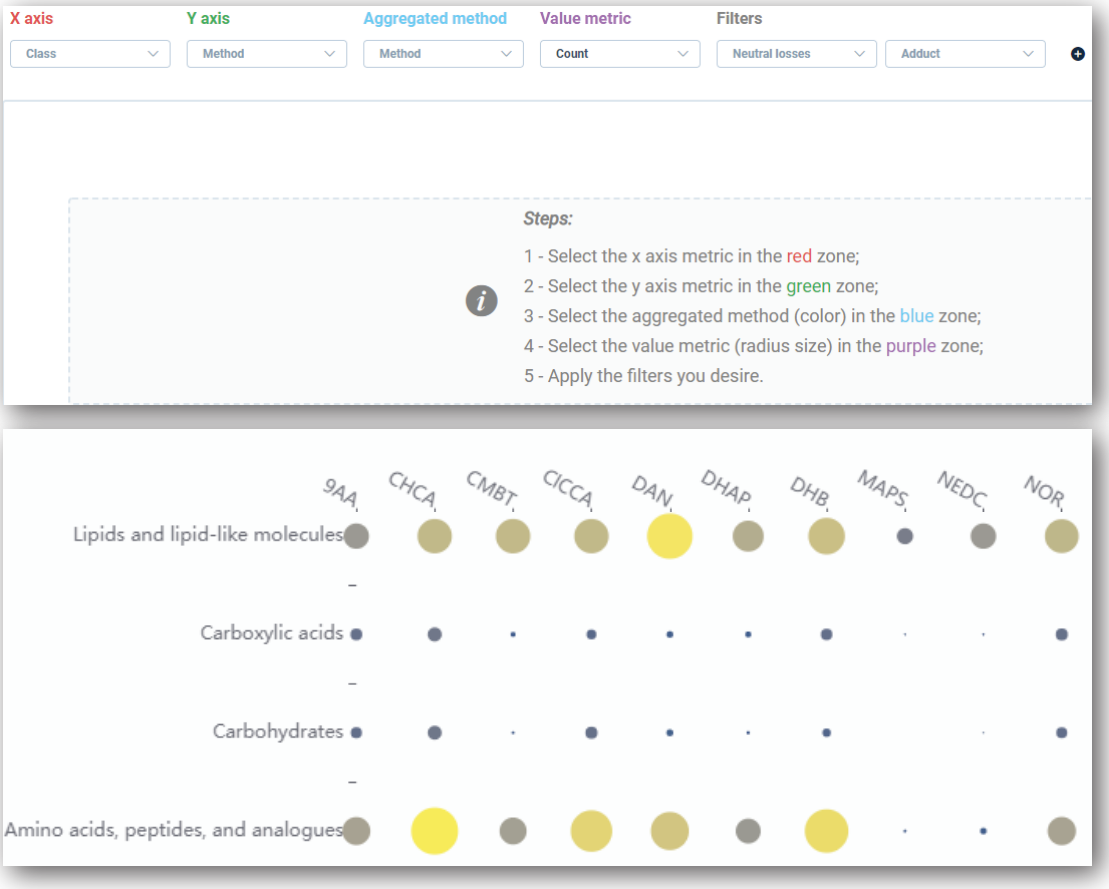


Figure 2.26 Screenshots of a prototype of interactive web application that allows visualising parts of the collected results. The prototype was created in collaboration with software developer L. Vieira.

The availability of this web application provides easy open access to all experimental data. It will allow other researchers to reproduce the visualisation of results discussed in this chapter. In addition, the researchers would be able to generate new hypotheses and draw as yet unexplored conclusions.

2.4 Conclusions

The aim of this work was to make a comprehensive assessment of MS imaging methods, in particular, MALDI-MS, for the untargeted analysis of metabolites relevant to biological samples. This section provides a summary of the main developments, discusses the significance of the obtained results and proposes directions for future studies.

2.4.1 Summary

First, a method was developed for the preparation and analysis of a new standard reference sample for MS imaging. To this end, 172 metabolites were selected to represent the diversity of chemical properties and biological functions that are expected in a typical mammalian metabolome. Metabolite solutions were sprayed onto glass slides in a regular array pattern, and in total 150 standard samples were produced. A part of these samples was analysed in the host laboratory on an AP-MALDI-Orbitrap instrument. The analysis compared 12 sample preparation protocols that included both established and recently developed MALDI matrices for which metabolite detection data is limited. The other standard samples were shared with collaborators for data acquisition on ten MS imaging instruments that represented the variety of ionisation sources and mass analysers commonly used in the MS imaging community. To analyse the data, a computational pipeline was developed, which would go from centroided MS data to the intensities of metabolite ions in each spot of the array. In the first step of the pipeline, the data was annotated using the custom database functionality of METASPACE. Then, the locations of the spots of the array were identified in the ion images using an interactive Python notebook. Ions specific to the metabolite spots were identified using machine-learning-based ion image classification. Finally, the intensities of the ions detected in metabolite spots were measured. Together, the acquired data represents a systematic account of metabolite detectability by the considered MS imaging methods.

Next, a comparison of metabolite detectability and ionisation efficiency was performed. The analysis determined which MALDI matrices in combination with which polarities are best suited for untargeted metabolite imaging by MALDI-MS. Furthermore, a detailed discussion explained which protocols are most applicable for the detection of metabolites from individual chemical classes and biochemical pathways. In addition to the assessment of metabolite detectability, the acquired data were used to study the effects of signal dilution. Signal dilution was found to lead to a measurable loss in sensitivity especially in positive ionisation mode. The common trends in signal dilution that should be taken into account when performing metabolite analysis were summarised for the main chemical classes. Remarkably, comparing metabolite detectability in the standard sample and in data sets of 30 mammalian tissues (acquired on the same instrument with the same MALDI matrices) showed that the results were in good agreement. I.e. the majority of metabolites found in tissue were also detectable in the standard sample, while the compounds that were not found in the standard sample, largely, were not detectable in tissue. A comparison of results in the context of an inter-laboratory survey showed that metabolite detectability is surprisingly similar across different MS imaging technologies. Both of these findings point to the broad applicability of this type of standard sample to the untargeted analysis of metabolites in biological samples. Finally, a prototype of an interactive application that helps browse and visualise the results was constructed.

2.4.2 Significance of the obtained results

The main achievement of this work is the collection of comprehensive data on the detectability of metabolites by a variety of protocols of MALDI-MS imaging. It expands the scope of previous studies that considered individual MALDI matrices or made comparisons for a smaller subset of analytes and protocols. Through the use of a defined reference standard sample, this study enabled direct comparison between MALDI-MS protocols, some of which have never been compared before. The

systematic approach undertaken here provides an unbiased empirical basis for common assumptions and puts all the findings in a broader context of untargeted metabolite analysis.

The data acquired from the metabolite arrays represents instrument performance under idealised conditions, that is, when the metabolite is sufficiently abundant in the sample and its detection is not impaired by major suppression effects. At the same time, the good correlation between standard sample measurements and tissue imaging shows that the metabolites used for standard sample preparation were well chosen and that the results obtained in this study are transferable and relevant to protocol choice for biological sample analysis.

Interestingly, the inter-laboratory survey of metabolite detectabilities found similar performance between MS imaging technologies using different ion sources (laser ablation, photoionisation or electrospray). This means that when any of these technologies are used with an established instrument method, they converge in their ability to generate molecular ions of abundant analytes. For sufficiently abundant analytes the detectability becomes limited mainly by the chemistry of the compound and selected matrix-polarity combination (or in the case of matrix-free MS imaging technologies, perhaps to a lesser degree, the choice solvent). The observed consistency of results between different laboratories and technologies means that the findings of this study can be used as a reference for metabolite detectability in the wider MS imaging community.

The results of this work would be the most beneficial for the MS imaging community when they are openly shared. For this, a prototype of an interactive result browser was created. In the future, this interactive application can be made available as a web resource, integrated with METASPACE data annotation platform. The value of this interactive application is that it will help the users to find the most suitable

MALDI-MS imaging protocol and polarity for a metabolite of interest, or a compromise solution that works well for several metabolites of interest (e.g. metabolites within one biochemical pathway). Additionally, the application can be used to identify the most abundant ion species for a given metabolite, thus informing the user of the potential loss of measurement sensitivity due to signal dilution. This web application will help researchers to plan targeted metabolite imaging experiments. The availability of the detailed results of this study could be particularly beneficial to scientists that are new to metabolite analysis by MS imaging. Likewise, the intention is to share the raw data along with a peer-reviewed publication. Availability of such data could be beneficial for the research groups that do not have access to equipment or resources for the collection of similar data sets but have a need for experimental data to inform future hypotheses.

2.4.3 Future outlook

Going forward, standard samples of the kind prepared in this study could be useful for benchmarking future protocols and instrumentation. For example, such a sample would be useful during the development of new MALDI matrices as it would allow a comparison of their performance to the protocols already analysed in this work. The standard sample would also help assess the changes in MS imaging instrument performance after, for example, switching to a newer version of an ionisation source, changing the laser and optics, or after major updates to the instrument software or firmware. Future studies could make use of the developed experimental and computational procedure to perform a similar survey including additional compounds of interest into the standard sample, such as plant metabolites, or particular classes of drugs. With additional validation, it might be possible to prepare a standard sample with metabolites deposited on substrates other than plain glass. For instance, sections of homogenised ("mimetic" [72]) tissue could be used to study suppression effects observed in biological tissue samples.

The results and raw data acquired in this work could be further analysed to examine aspects of metabolite ionisation that were not addressed here. For example, metabolite in-source fragmentation can be analysed in more detail than what was covered by the analysis of selected neutral losses in this study. It could be interesting to more accurately determine the extent to which such fragmentation occurs in a typical MS imaging experiment, and whether the fragmentation pattern resembles those found in MS/MS databases populated with data from LC-MS experiments. This information may be informative for increasing confidence in compound identification from MS1 spectra in the future.

The acquired data could find use beyond metabolite analysis. Examining the physicochemical properties in solution, gas phase, ground and excited states of the metabolites and MALDI matrices and the observed ionisation efficiencies could shed light on some aspects of the mechanism of ion formation in MALDI. This would complement already existing ionisation models, and perhaps, in the future, lead to the development of theoretical models for the prediction of ionisation efficiencies. A more detailed comparison of results between different instruments is another avenue to explore. For example, the comparison of atmospheric pressure and vacuum MALDI-MS could provide another perspective on the ionisation mechanisms, as the theory of MALDI largely originated from experiments with MALDI-TOF where ablation is performed under vacuum.

Overall, the outcomes of this work and the opportunities that they provide are believed to be of great use to the MS imaging community, making metabolite imaging more effective, interpretable and impactful.

2.5 Materials and methods

2.5.1 Preparation of standard samples

Chemicals

Analytical standards were purchased from the following vendors: MetaSci, Toronto, ON, Canada; Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany; Avanti Polar Lipids, Inc., Birmingham, AL, USA; GLSynthesis Inc, Worcester MA, USA. A complete list of analytical standards is shown in table A. From Fisher Scientific (Schwerte, Germany) the following chemicals were purchased: LC-MS grade acetonitrile, methanol and water; ACS grade chloroform; sodium hydroxide (aqueous, 2M); hydrochloric acid (aqueous, 37%), nitric acid (conc.). SYLGARD 184 silicone resin and curing agent were purchased from Biesterfeld, Hamburg, Germany. Nitrogen, argon and oxygen gas of 99.999% or higher purity were from EMBL internal supply. PNDI-T2 was purchased from Ossila BV, Leiden, The Netherlands. 4-maleicanhydridoproton sponge (MAPS) was kindly provided by Prof. Dr. Karsten Niehaus, Bielefeld University, Bielefeld, Germany. All other MALDI matrix substances were purchased from Sigma-Aldrich Chemie GmbH, Taufkirchen, Germany.

Solution preparation

Each standard compound was dissolved in a suitable combination of water, methanol or chloroform into concentrated stocks. A detailed list of protocols is provided in appendix D. Certain compounds that were not soluble to desired concentration in any of the solvent systems were solubilised through titration with hydrochloric acid or sodium hydroxide, added drop-wise until dissolved. Each stock solution was further diluted to final concentration in one of two standardised solvent systems: 9:1 methanol:water (v:v); and 2:1 chloroform:methanol (v:v).

LC-MS/MS

All standard solutions were analysed using LC-MS/MS before slide preparation started, and once again after the end of the four-month slide preparation process. The solutions were injected undiluted after internal standard addition, and run with two different rapid gradient methods determined by the solvent system.

1. For samples in methanol/water solution. Column: BEH Amide, 2.1x100 mm, 2.5 µm (Waters GmbH, Eschborn, Germany). Mobile phase A: water, 7.5 mM ammonium acetate, 0.05% ammonium hydroxide. Mobile phase B: acetonitrile. Flow rate 0.2 mL/min. Internal standard: 10 ppm ^{13}C labelled creatinine. Gradient described in the table below.

Time (min)	B (%)
0	85
2	85
12	10
14	10
14.1	85
16	85

Table 2.2 HILIC gradient

2. For samples in chloroform/methanol solution. Column: Kinetex C18 5 cm (Phenomenex, Aschaffenburg, Germany). Mobile phase A: 6:4 acetonitrile:water (v:v), 10 mM ammonium acetate. Mobile phase B: 9:1 isopropyl alcohol:acetonitrile (v:v), 5 mM ammonium acetate Flow rate 0.26 mL/min. Internal standard: 10 ppm ^{13}C labelled cholesterol + 10 ppm ^{13}C labelled oleic acid. Gradient described in table below.

Time (min)	B (%)
0	5
1	5
5	60
6	90
8	90
9	5
12	5

Table 2.3 RP gradient

All LC-MS/MS data were collected on a Q Exactive Plus mass spectrometer using an instrument method collecting full MS spectra in positive and negative modes, as well as data-dependent MS/MS. The chromatograms of target ions ($[M+H]^+$, $[M-H]^-$) were integrated with FreeStyle (Thermo Fisher Scientific, Bremen, Germany), using the Genesis algorithm for peak detection with manual corrections where necessary.

Preparation of PDMS polymer templates and microwell slides

PDMS polymer templates were cast using polyoxymethylene moulds produced by the EMBL Mechanical Workshop (technical drawing is shown in appendix G). Silicone casting and curing was done under cleanroom conditions. Sylgard 184 PDMS resin was mixed with curing agent according to manufacturer's instructions. For each casting, 4 g of the curing resin was poured into the mould immediately after mixing. After pouring, moulds were first degassed in a vacuum desiccator for 30 min, and then cured for 2 h at 65°C. After curing, the polymer templates were removed from the mould and attached to clean, dust-free glass slides. Prior to this, the glass slides were submerged in 2:1 nitric acid:hydrochloric acid (v:v) for 30 min to remove organic contamination, then rinsed thoroughly with water (MilliQ), dried under a stream of nitrogen and kept protected from dust until used.[73]

Deposition of metabolite solutions onto the glass slides

The metabolite solutions were dispensed into microwell slides using an ATS4 TLC spotter (CAMAG, Muttenz, Switzerland). A custom-made holder for standard size slides allowed production of batches of 10 slides in parallel. The "Free mode" functionality of the instrument software (WinCATS v. 1.4.4.6337) was used to create a spotting program dispensing 2 µL of each solution into the microwells. Each slide was spotted with an identical array of metabolites as shown in appendix F. After spotting, the PDMS template was removed, the finished slides were vacuum packed under argon gas and stored at -80°C until analysis.

2.5.2 Mass spectrometry imaging

Sample preparation for MALDI-MS imaging

Sample slides were removed from -80°C on the day of analysis and were warmed to room temperature for at least 1 h before the vacuum seal was broken. Spray-coating of matrix substances was performed using a TM Sprayer (HTX Technologies, Chapel Hill, NC, U.S.A.) according to the protocols in table 2.4.

Mass Spectrometry Imaging with AP-SMALDI5-Orbitrap MS

After matrix spray-coating, samples were imaged on an AP-SMALDI5 ion source (TransMIT, Giessen, Grenamy) coupled to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, Bremen, Gemrany). Instrument parameters were the same for all matrices, as listed in table 2.5.

Matrix	Solvent	Concentration (mg/mL)	Temp (°C)	Flowrate (mL/min)	Gas Pressure (psi)	Track Spacing (mm)	Velocity (mm/min)	Passes	Density (mg/mm^2)
1,5-DAN	7:3 ACN/H2O	10	85	0.07	10	3	1350	8	0.001383
2,5-DHB	7:3 ACN/H2O	10	85	0.068	10	3	1350	8	0.001343
NOR	1:1 CHCl3/MeOH	7	30	0.122	10	3	1350	7	0.001476
9AA	7:3 MeOH/H2O	5	75	0.07	10	3	1350	20	0.001728
CHCA	1:1 ACN/H2O	7	90	0.071	10	2	1350	9	0.001657
PNDI-2T	Toluene	2.5	30	0.1	10	3	600	10	0.001389
MAPS	Toluene	5	30	0.1	10	3	600	10	0.002778
NEDC	7:3 ACN/H2O	10	80	0.05	10	3	1350	13	0.001605
DHAP	7:3 ACN/H2O	10	85	0.067	10	3	1350	8	0.001323
ClCCA	1:1 ACN/H2O	7	90	0.07	10	2	1350	10	0.001815
CMBT	9:1 ACN/H2O	7	70	0.1	10	3	1350	10	0.001728
pNA	17:3 MeOH/H2O	11.45381	65	0.068	10	3	1350	6	0.001153

Table 2.4 TM Sprayer Protocols used for AP-SMALDI5 imaging. Solvent shorthand: ACN – acetonitrile; H2O – water; MeOH – methanol; CHCl - chloroform.

Parameter	Low Mass Range	High Mass Range
Attenuator Angle	29°	29°
Pitch (µm)	150	150
Mass Range (m/z)	70-350	300-1510
Voltage (pos / neg)	+3 kV / -3.1 kV	+3 kV / -3.1 kV
RF Level	50	100
Injection Time (ms)	500	500
Mass Resolving Power (nominal)	140000	140000

Table 2.5 Acquisition parameters

Each slide was imaged in both ion polarities at the indicated low and high mass ranges, for a total of four sequential acquisitions. Between each run, the imaging raster was offset manually by 50 µm to avoid sampling the same locations more than once. Typical analysis times were on the order of 10 h per image, or 40 h per slide.

2.5.3 Data analysis

Image data formatting

All raw data generated in the host laboratory was converted from the vendor proprietary data format to the open mzML format using the MSConvert software[74] using the vendor-supplied method for spectrum centroiding. From mzML format, the data was converted to imzML using imzMLConverter[75]. External collaborator data was converted separately by each participant lab and uploaded to METASPACE as centroided imzML files. Where the same sample had been imaged separately in a lower and a higher mass range for a given polarity, the images were merged pixel-by-pixel through a Python script utilising the Pyimzml library. For convenience, the output images were also rotated to a uniform orientation with the reference well at the top (as shown in appendix 6), and re-uploaded to METASPACE with the following standardised processing parameters. 10 ppm mass tolerance; Positive mode adducts: $[M]^+$, $[M+H]^+$, $[M+Na]^+$, $[M+K]^+$; Negative mode adducts: $[M]^-$, $[M-H]^-$, $[M+Cl]^-$; Neutral losses: $-H_2O$, $-2H$, $-CO_2$, $-CH_2O_3$, $-CH_2O_2$, $-HPO_3$, $-H_3PO_4$, $-NH_3$; Neutral gains: $+2H$, $+matrix$ (added individu-

ally for MALDI data sets). A map of well identity per pixel was generated for each data set by fitting reference points to image features with an interactive grid fitting tool developed for the purpose in Python.

Ion image classification

For ion image classification into ”real signal” and ”background noise” images, a machine-learning-based model was trained based on pre-computed set of metrics. As input features to the model, a variety of metrics were calculated for each pair of spot and ion image, where at least one peak in the ion image appeared inside the spot. The following metrics were included as model features:

- ”spot_intensity_tic_norm” - the average TIC-normalised intensity of the pixels inside the spot.
- ”spot_occupancy” - the percentage of pixels with a non-zero intensity in the spot.
- ”occupancy_vs_far_bg_ratio” - the ratio between the percentage of pixels with non-zero intensities inside the spot and in the “far background” (background area at least 4 spot-radii away from the spot).
- ”intensity_vs_far_bg_ratio” - the ratio between the average intensities inside the spot and in the far background.
- ”intensity_vs_other_spots_ratio” - the ratio between the average intensities inside the spot and in other spots.

A CatBoost[70] model was trained on a manually annotated data set of 284 ion images to identify ion images featuring localised intensity of a putative metabolite ion in its expected well location. The classifier’s raw probability output was used to classify images as detected or undetected. Only images with a probability score of 80% or higher were classified as ”real signal” and included in further analysis.

Assessment of chemical diversity of selected chemical standards

To construct the chemical space, SMILES of metabolites of mouse genome-scale metabolic model (iMM1865, [45]) were obtained through BiGG open-source repository[46]. SMILES for the compounds used in standard sample preparation were obtained from HMDB, where an HMDB identifier was available, or alternatively from the conversion of the chemical structure into SMILES in Marvin software. SMILES were used to calculate ”Morgan/Circular fingerprint” vectors using the RDkit Python library. The calculated vectors were then used to plot 2D and 3D PCA score plots representing chemical space.

Plots

Dot plots, bar plots, line plots, scatter plots and heatmaps were generated in Python. The map of common biochemical pathways with the chosen chemical standards highlighted was created using the interactive web-based tool IPath v.3[49], and KEGG identifiers obtained from HMDB for chemical standards included in the standard sample. Where the KEGG identifier was not available, the standard was excluded from the visualisation (this was the case for the majority of lipids).

Interactive application for result visualisation

In order to visualise the data generated in this study, a data browsing tool was built to plot charts on demand. The data browsing tool is a Single Page Application (SPA) built using the Javascript framework Vue and the chart library eCharts. The tool allows the user to visualise the data in two modes: as a Cartesian heatmap or as a scatter plot. The charts are dynamically built using three metrics: X-axis, Y-axis and aggregation metric. The axis-related metrics are used to get the chart Cartesian points and can accept as values any of the following metrics: polarity, adducts, neutral losses, matrix, molecule, technology, pathway, class and data set. The aggregation metric sets the colours of scatter chart dots and heatmap cell intensities by accepting any of

the following as metric values: prediction, intensity and simple count. Other than the three dynamic metrics, the scatter plot also reflects the count of X-axis metric found in the selected Y metric (i.e number of molecules/X-axis found using a specific technology/Y-axis) by changing the dot radius. It is possible to filter the data by using the additional metrics: polarity, adducts, neutral losses, matrix, value prediction, state prediction, technology, pathway, class and data set.

2.6 Contribution statement

All work described in this chapter was performed by the author, except for the following contributions from collaborators. This study was supervised by Dr. Theodore Alexandrov.

Study design

- Theodore Alexandrov invited participants for the inter-laboratory part of the study.
- Måns Ekelöf chose 26 lipids that were included in the standard sample.

Experimental

- Måns Ekelöf helped to prepare half of the metabolite solutions and standard sample slides.
- Måns Ekelöf and Prasad Phapale acquired and analysed LC/MS data on metabolite stability in -80 Celsius storage.
- Beate Neumann, Mengze Zhang, Måns Ekelöf took part in testing automatic compound printer and dispensers.
- The EMBL Mechanical workshop (Arthur Milberger, Tim Hettinger, Leo Burger) manufactured the mould for PDMS casting and 10-slide sample holder.

The following collaborators performed analysis of standard samples and shared their processed imaging data: Bruker Daltonik (Bremen, Germany), Hopf group (Hochschule Mannheim, Mannheim, Germany), AstraZeneca (Cambridge, U.K.), Spengler group (Justus-Liebig-Universität, Gießen, Germany), Dreisewerd group (University of Münster, Münster, Germany), Muddiman group (NCSU, Raleigh, NC, U.S.A.), Eberlin group (U.T. Austin, Austin, TX, U.S.A.).

Data analysis

- Lachlan Stuart developed an interactive grid fitting tool in Python.

- Lachlan Stuart performed machine learning model evaluation and wrote Python code for training and applying the model.
- Lucas Vieira developed the interface of the interactive web application.
- Lucas Vieira helped to make comparative analysis between metabolite detectability in standard samples and biological tissues.
- Måns Ekelöf contributed to the classification of ion images for the above-mentioned analysis.

Chapter 3

Spatial mapping of metabolic host-microbiome interactions in mouse intestine using MALDI-MS

3.1 Introduction

Context of the problem

Bacteria are known to facilitate essential metabolic transformations in the gut, such as digestion and the production of vitamins, immunomodulators and neurotransmitters. An increasing number of studies have found associations between disruptions in gut microbiome composition and the progression of various diseases (such as inflammatory bowel disease, colorectal cancer, or diabetes) which highlight the importance of understanding microbial-host interactions and the functional roles of gut bacteria.[76] The research on this topic, however, is hindered by two complications.

The first challenge is that the microbiome is not a homogeneous entity. For a long time, studies focused on probing an easily accessible microbial composition of stool samples in the search for biomarkers. Analysis of faecal pellets, however, only scratches the surface of true

bacterial complexity in the gut.[77] Indeed, metagenomic sequencing studies that collected samples at several locations along the gastrointestinal (GI) tract revealed differences in bacterial population and density along both longitudinal and cross-sectional axes of the gut.[78, 79] Moreover, a recent micro-scale analysis of three gut regions showed that associations of different bacterial taxa at each site were unique, suggesting that different microbial sub-communities exist in different parts of the gut.[80] This raises the questions, whether each microbial sub-community has a different function and where do the boundaries between bacterial sub-communities lie. Consequently, it becomes important to consider the spatial organisation of the microbiome, now also called ”gut biogeography”, when inquiring into this topic.[81]

The second challenge is that the function of microbiome is not a simple sum of the functions of individual bacteria that constitute it. A common approach to probing microbial activity and function is by metabolomics, which detects products of bacterial catabolism and anabolism. The metabolic activity of the microbe relies not only on microbe’s genetic makeup but also on the its environment. Analysis of an isolated bacterial strain in culture identifies some metabolites that a given strain would also produce *in vivo*, since it has the enzymatic machinery to do so. Nevertheless, a large part of metabolic activity is not captured, because the unique environment of the gut driving the metabolism (pH, low oxygen conditions, bacterial cell density and special substrate availability) cannot be faithfully recapitulated *ex vivo*. In addition, substrate sharing, or metabolic cross-feeding between bacteria, enables reactions that would not be possible for either of microbes, had they been grown in isolation.[82] Therefore, when studying the functions of the microbiome, it is important to consider bacterial sub-communities as a whole, and analyse them in the full context of the gut (for example, animal models).

Previous research

There is only a limited account of the spatial differences in metabolite distribution inside the gut, and how these differences can be attributed to bacterial sub-communities. Metabolomics by capillary electrophoresis mass spectrometry (CE-MS) was used early on to show differences in metabolites present in the colon of regular mice but absent in germ-free mice.[83] The same technique was used to compare metabolites from six sites along the GI tract, and to cluster metabolites into groups according to their abundance in each of the sites.[84] A study using nuclear magnetic resonance (NMR) spectroscopy characterised the metabolic differences between four physiological gut sections and suggested to combine this data with metagenomics analysis for a better interpretation of result.[85]

Spatial information from these bulk metabolomics approaches that sample a few locations along the gut is too coarse to offer a comprehensive view of most metabolite gradients in the gut. Mass spectrometry (MS) imaging, on the other hand, provides the high spatial resolution required for this. MS imaging was used for imaging of gut tissues previously. A combination of matrix-assisted laser desorption/ionisation (MALDI)- and desorption electrospray ionisation (DESI)-MS imaging was used to localise a bile acid metabolite produced by a bacterium in whole-body tissue sections of monocolonised mice. [86] Same technologies led to the discovery of two microbial metabolites found in both mouse gut and brain: the bacterial species producing these metabolites were identified by analysing individually cultured bacteria isolated from the same animals.[20]

While the above-mentioned studies did find evidence of bacterial metabolism in animal gut tissues, none of the studies related metabolomics findings to the spatial localisation of the microbes. This was achieved for the first time recently, by applying two techniques on the same tissue section of gills of a deep-sea mussel inhabited

by two symbiotic bacteria. Fluorescence in situ hybridisation (FISH) microscopy was used to locate bacteria in tissue using 16S ribosomal RNA, whereas MALDI-MS imaging was used to identify and locate microbial metabolites.[8]

Aims of the study

This work aimed to find associations between spatial differences in metabolism and composition of microbiota in the complex setting of the murine gut using best practices from previous studies described above. The objectives of the study were as follows.

1. To develop a sample preparation procedure, enabling multi-omics analysis of a whole, faecal-matter filled, mouse intestine.

2. To define the regions of distinct metabolism in the intestine, using MALDI-MS imaging.

3. To describe the anatomical and physiological significance of defined regions, in relation to the spatial differences in metabolite abundance.

4. To perform bulk sequencing on intestinal segments and relate spatial changes in the composition of the microbiome to the regions of distinct metabolism.

5. To isolate bacteria most differentially abundant in the gut, and construct a library of their metabolite profiles.

6. To check whether metabolites produced by individual bacteria are detected in the regions of distinct metabolism, and whether that is in agreement with where these bacteria are expected based on bulk sequencing data.

7. To investigate whether the distribution of bacterial species aligns with the distribution of the metabolites, using FISH imaging.

Significance of the obtained results

This work is the first to show the spatial differences in metabolite abundance in the whole, faecal-matter filled mouse intestine at high spatial resolution. The comprehensive MS imaging data of metabolite distributions in the gut acquired in this study were generally in agreement with previous studies. In addition, it led to new findings and allowed defining the regions of distinct metabolism in the gut. Together, these results could serve as a ”healthy” phenotype reference in future studies in the host laboratory, and lead to a better understanding of site-specific gastrointestinal metabolism. Moreover, future studies using disease models can benefit from these reported ”healthy” states and could shed light on some aspects of microbiota-associated diseases.[85]

While only metabolites (and lipids) annotated by public databases were considered, the data analysis pipeline developed in this study could be directly expanded to the non-annotated ions detected by imaging MS. This approach could be useful for mining the so-called ”dark metabolome” to find previously unknown metabolites that have unique distribution patterns in the gut. This study has shown the viability of the biotyping approach, which identifies the location of microbes in the intestinal tissue based on their metabolic fingerprint library. A potential direction of further research is to combine the analysis of unannotated metabolite ions along with the construction of a more extensive bacterial metabolite library. This would allow establishing new links between bacteria and the spatial context of intestinal metabolism.

This study has demonstrated how 16S microbiome analysis could be carried out to complement the results of spatial metabolomics. Coarse-grained bulk sequencing of intestinal segments was informative to identify bacteria of interest for the biotyping approach. In addition, it was shown that high spatial resolution FISH imaging could be carried out on the same tissue sections as used for spatial metabolomics. This finding establishes confidence in validation of results arising from combined

metabolomics and biotyping approaches.

3.2 Results and discussion

3.2.1 Development of the sample preparation procedure

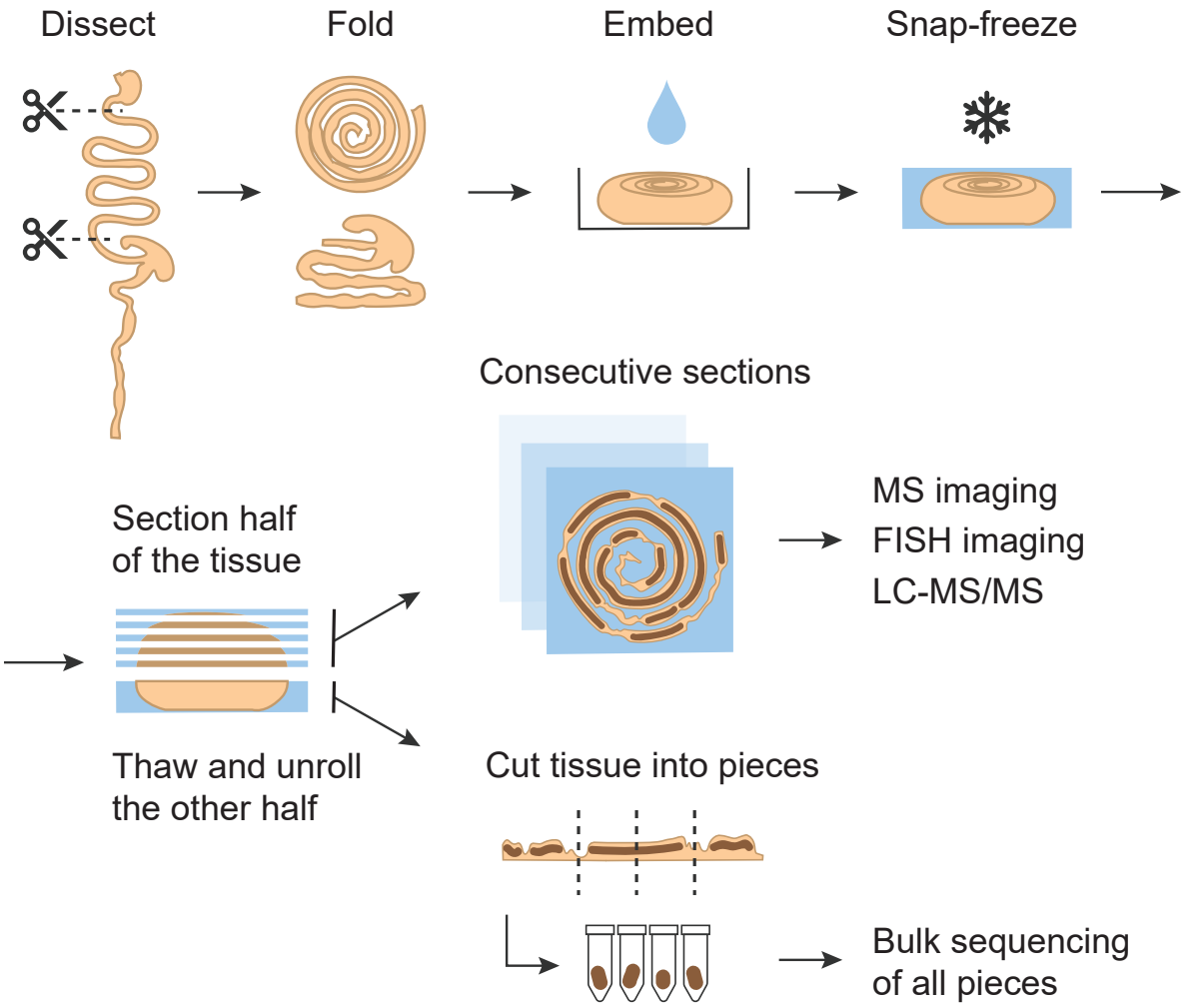


Figure 3.1 Sample preparation protocol. First, the intestine was dissected and split into two pieces: small intestine and large intestine. The small intestine was rolled into a spiral, the large intestine was folded in a "zig-zag" manner, and both were placed into plastic weighing boats. Then, the tissue was covered with a cryo-embedding medium and snap-frozen. Cryotome was used to collect consecutive sections from half of the frozen sample. Leftover tissue was then thawed, unrolled, segmented into 1-2 cm pieces and reserved for bulk tissue sequencing analysis.

The first step towards fulfilling the aim of this work was to develop a novel sample preparation procedure that enables multi-omics analysis of a whole, faecal-matter filled mouse intestine, as shown in figure 3.1. In particular, the samples from one animal were to be analysed by MS

imaging, FISH imaging, bulk sequencing and LC-MS/MS analysis.

Producing tissue sections of the faecal matter-filled intestine is challenging because the faecal matter is not attached to the rest of the tissue, and because it is fragile in a frozen state due to the high water content. Previous studies that prepared rodent intestinal tissue sections, either washed the faecal matter out during sample preparation ([20], [87]), or used filled intestine with sampling small transverse cross-sections ([88]).

To ensure tissue integrity, several details were important. First, it was critical to choose an embedding medium that, at the same time, did not crack while snap-freezing, produced sections that did not shatter or curl, and was sufficiently fluid to pour, while not suppressing the MS signal. Among several embedding media tested (Optimal Cutting Temperature compound, gelatin, carboxymethylcellulose (CMC), gelatin-CMC mixtures), 4% CMC solution performed best. Second, it was crucial for successful embedding to remove all bubbles from the cryo-embedding medium, and to freeze the sample quickly and uniformly. For this, the tissue in cryo-embedding medium was submerged into an isopentane bath that was pre-cooled to -78 Celsius on dry ice. Third, to avoid tissue disintegration when multiple washing steps are required (as in FISH staining), tissues were mounted on electrostatically charged adhesion glass slides. The only part of the intestine where faecal matter could not be completely retained was the distal colon, likely because the faecal matter gets rapidly dehydrated in the last part of the gut, and it no longer had enough water to bind to the slide during thaw-mounting.

For each frozen tissue block, at least 30 consecutive sections were collected, which allowed for the analysis of the same tissue by multiple techniques. For example, only two sections were required for MS imaging, one section - for FISH imaging, and three sections (pooled) - for LC-MS/MS analysis. Samples for at least five technical replicates were

collected per technique. Then, the leftover tissue was thawed, unrolled, segmented into 1-2 cm pieces and reserved for bulk tissue sequencing analysis.

To summarise, a simple method was developed that allowed collecting consecutive tissue sections that capture the whole mouse intestine, and that enabled MS imaging, FISH imaging, LC-MS/MS, and bulk tissue sequencing of the samples from one animal.

3.2.2 Imaging metabolites with MALDI-MS

Data acquisition and preprocessing

MALDI-MS imaging was performed on the sections of the small intestine and large intestine from four mice (two males and two females). To maximise the coverage of the compounds, data were acquired using positive and negative ionisation modes in the mass-to-charge ratio (m/z) range 80-1200. Data were recorded on a mass spectrometer equipped with an Orbitrap detector, ensuring high mass resolution. Tissues from three animals were acquired with a spatial resolution of 150 micrometres (pixel-to-pixel distance). From one animal, tissues were imaged with higher spatial resolution: pixel-to-pixel distance was 75 micrometres. In total, 16 data sets were recorded.

Figure 3.2 shows a mean spectrum of one of the small intestine data sets. MSIreader software detected approximately 3,800 peaks with a minimal intensity of 200 ions/second. At lower intensity thresholds, the number of peaks increased tenfold. These peaks correspond to tissue metabolite ions with various adducts, background ions (embedding material, MALDI matrix, embedding material), and fragments of both.

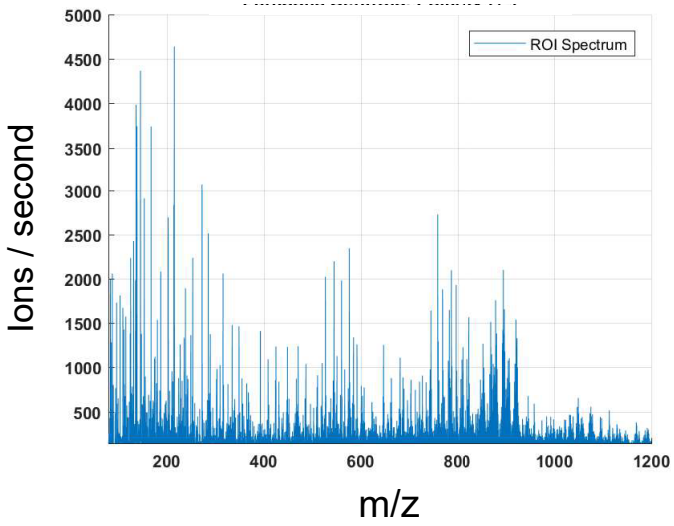


Figure 3.2 A mean mass spectrum of one of the small intestine tissue sections in positive ionisation mode.

To distinguish biologically relevant peaks from a background peaks, the machine learning image classification method described in detail in the previous chapter (section 2.2.4) was applied to classify ion images into on-sample images and background images. Briefly, for every tissue section data set, a binary mask image was drawn, where all pixels overlapping with tissue were labelled as 1 and all the other pixels as 0. For every ion image, the same metrics as in the previous chapter were calculated, namely, a ratio of average ion intensities between the two regions, and a ratio of occupancy (i.e. a fraction of non-zero pixels) between the two regions. A subset of user-labelled ion images was used to train the classification model. Then the performance of the model was evaluated, and finally, the model was used to classify all ion images. Figure 3.3 shows examples of a binary mask, an ion classified as on-sample, and an ion classified as background.

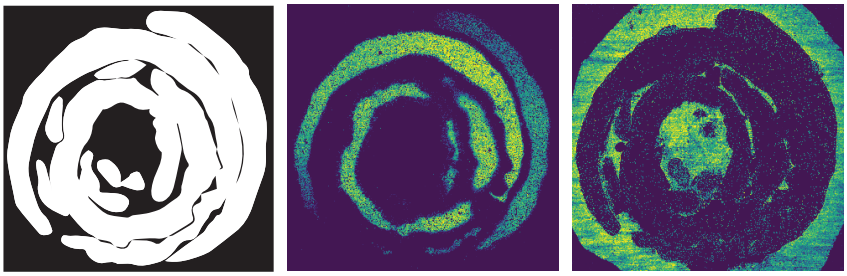


Figure 3.3 Left to right: example of a binary mask used for metric calculation for image classification model; ion classified as on-sample; ion classified as background.

Hypothetically, ion image classification could be applied to all ion images generated from tens of thousands of peaks from each of 16 data sets. In practice, the computational resources to achieve this in a reasonable time were lacking. Thus, it was decided to postpone such analysis, and focus on a subset of peaks from every data set instead.

To this end, the data sets were first annotated by the online annotation platform for MS imaging data (METASPACE).[69] METASPACE searches the data for signals corresponding to monoisotopic masses of metabolites present in a metabolite database. For every match, METASPACE calculates the annotation confidence score and false discovery rate. Finally, a list of ranked putative annotations, their m/z values and corresponding ion images are generated. The disadvantage of only looking at the annotated subset of the data is that metabolites absent from the databases are not captured. For annotation, several metabolite databases were combined:

- Human Metabolome Database, which is the most comprehensive collection of human metabolites, with over 114,000 compound entries.[89] While this is a database for the human metabolome, a substantial overlap with the mouse metabolome is expected due to the broad scope of included compounds.
- LIPID MAPS Structure Database, which collects more than 47,000 biologically relevant lipids and their structures.[90]

- Genome-scale metabolic model of Mus Musculus (iMM1415).[91]

The subset of annotated metabolites was subjected to manual ion image classification, yielding 200 labelled images that were then used to train the classification model. Figure 3.4 shows how many ions were classified as on-sample.

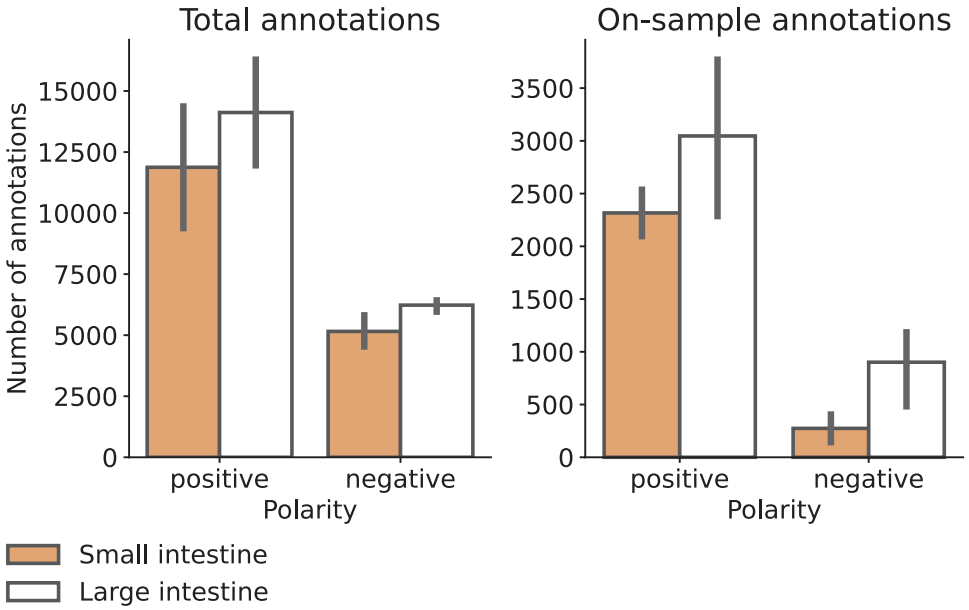


Figure 3.4 The number of ions annotated across all tissue data sets. Left: in total, right: classified as on-sample. In line with the data available in the literature, more features were observed in the large intestine than in the small intestine.[92] In addition, twice as many metabolites were detected in positive ionisation mode (shown in the plot as ”polarity”) than in negative ionisation mode.

Overview of detected metabolites

For the preliminary investigation of data, three common plots were created. The histogram of m/z values shows that a broad spectrum of metabolites was detected: small metabolites around m/z 200, fatty acids around m/z 500 and larger lipids around m/z 800, (see figure 3.5)

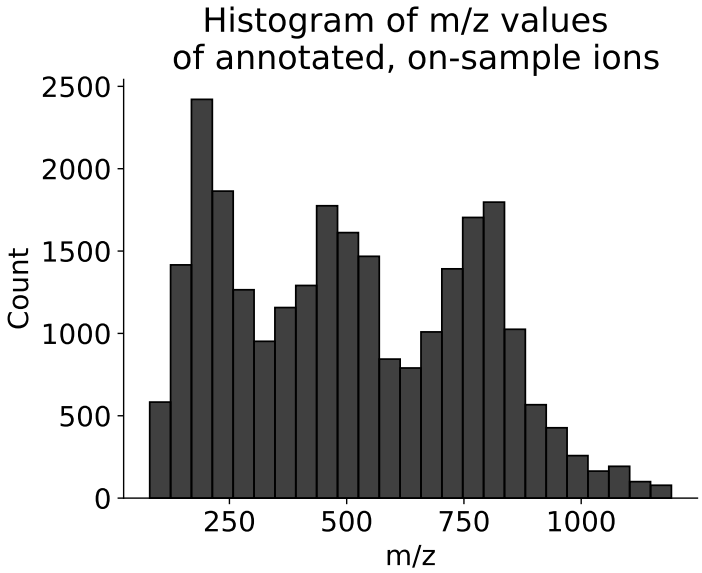


Figure 3.5 Histogram of m/z values of annotated ions classified as on-sample.

Plotting absolute mass defect over m/z in figure 3.6 shows differences in atomic composition among selected ions. Mass defect (in mass spectrometry) is obtained by subtracting the nearest integer mass from the mass of the ion. Since most of the ions in MALDI-MS are singly charged, m/z values can be used instead. Mass defect increases when hydrogen and nitrogen are added to the compound because mass defects of these atoms are: 0.008 for protium and 0.003 for nitrogen-14. The mass defect of carbon-12 is 0, therefore the addition of carbon does not change the mass defect. Mass defect decreases with the addition of, for example, phosphorus (-0.027), sulphur-32 (-0.028), oxygen-16 (-0.005).

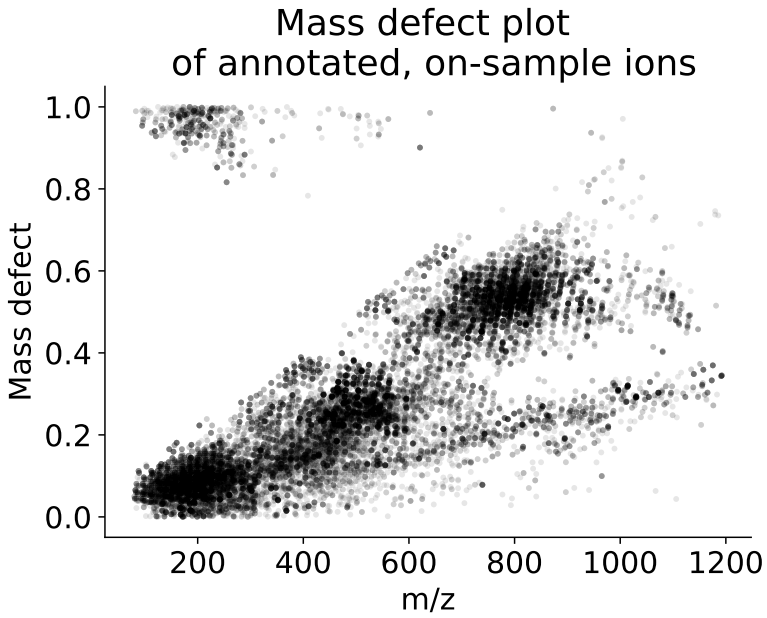


Figure 3.6 Mass defect plot of annotated ions classified as on-sample. The mass defect plot shows a set of data points in the top left corner for the small metabolites with increased content in these three atoms, for example, carbohydrate phosphates. In addition, the data points aligning along the main diagonal of the plot indicate the abundance of highly saturated lipids. A minor diagonal trend in the data on the bottom right side of the plot may indicate the abundance of polysaccharides, peptides or highly unsaturated lipids, such as steroids.

The Van Krevelen plot, shown in figure 3.7 is useful to see what classes of compounds have been annotated, using ratios of the number of hydrogen, carbons, oxygen, and, sometimes, nitrogen atoms in individual compounds. For comparison, a Van Krevelen diagram of compounds used in the metabolite standard sample in the previous chapter (section 2.2.1)() is shown and colour-coded by chemical class. Consistent with findings from the mass defect plot, the Van Krevelen plot shows the abundance of lipids and steroids among annotated metabolites, as well as compounds from all other classes, such as carbohydrates, and amino acids or peptides.

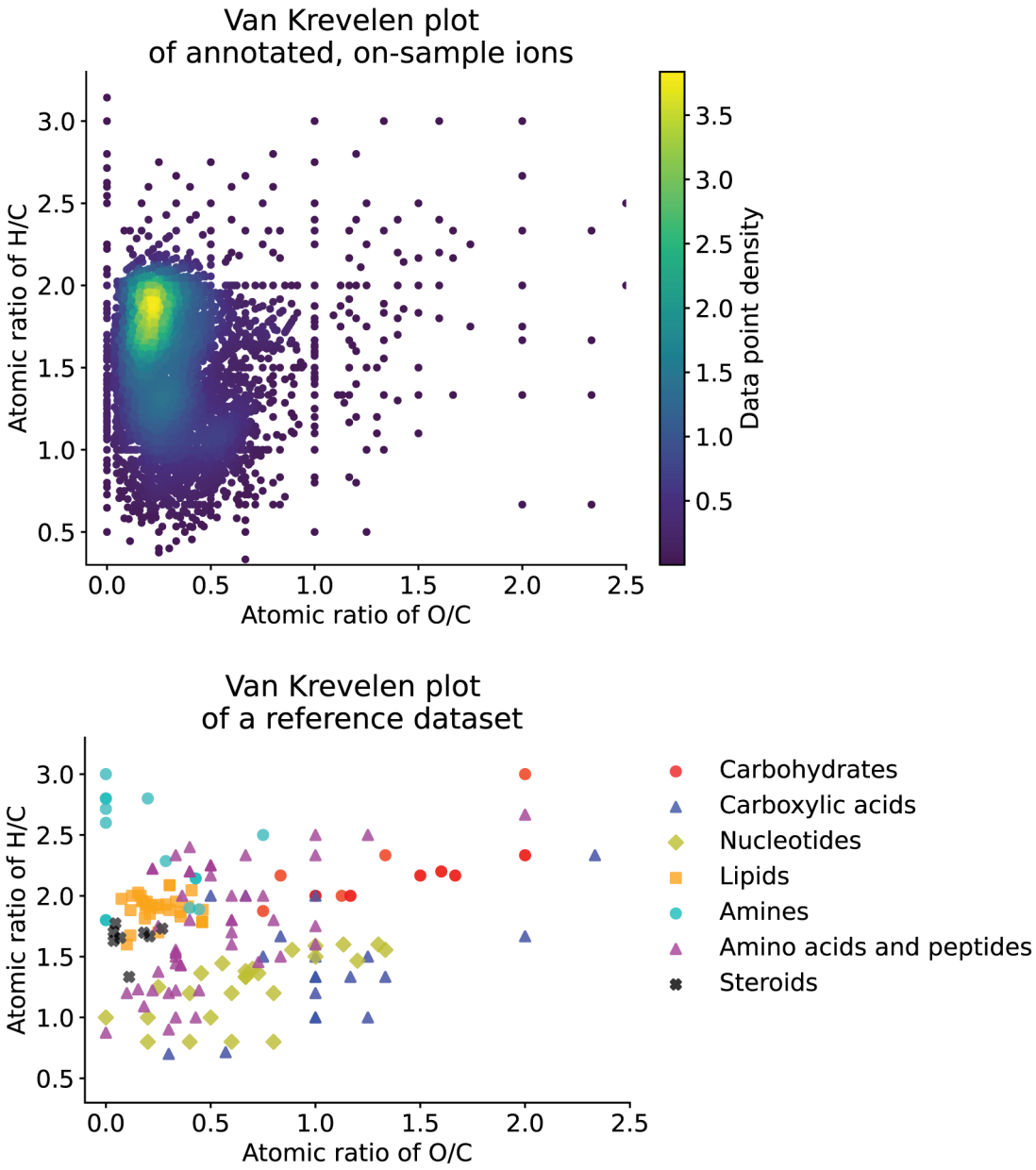


Figure 3.7 Van Krevelen plots. Top - a plot for annotated ions classified as on-sample. The plot is coloured by the local density of the data points. Bottom - a reference plot with a few representative metabolites of various chemical classes. The plot is coloured by chemical class.

In summary, these results show that a wide range of biologically relevant compounds in terms of mass, composition and chemical class were detected in whole intestine tissue sections by MALDI-MS imaging.

3.2.3 Spatial differences in metabolite composition of the gut tissue

Previous studies have observed differences in metabolite abundances in different parts of the mouse intestine by analysing tissue extracts using NMR spectroscopy and mass spectrometry.[85, 84] Such differences, however, have not been shown in MS imaging data yet. To investigate whether these differences are present in the obtained data, all spectra in the data set were first partitioned into groups (or ”clusters”) by similarity. Then, the spatial coordinates of each cluster of similar spectra were analysed.

Many approaches for clustering spectra exist.[93] Here, an established procedure, which is common for analyses of similar types of data (e.g. single-cell transcriptomics), was used. Initially, each spectrum was reduced to a vector of metabolite intensities. Intensities were log-transformed and scaled to zero mean and unit variance to ensure that ions with both high and low intensities contribute to the analysis equally. Here, only annotated metabolites that were classified as on-sample were used. The same procedure, however, would be applicable for any set of peaks. The spectra could be thought of as data points in high-dimensional space, with as many dimensions as the number of metabolites considered. Clustering data points in high-dimensional space is computationally intensive, and a typical solution is to use a dimensionality-reduction method. The combination of PCA (Principal component analysis) and UMAP (Uniform Manifold Approximation and Projection) was used in this work. UMAP creates a graph connecting the data points in high-dimensional space and then re-creates this graph in low-dimensional space. Finally, a clustering algorithm partitions the data points in low-dimensional space into groups. Here, Leiden clustering, one of the modern clustering algorithms, was applied.

Figure 3.8 shows clusters obtained from a section of the small intestine imaged in positive ionisation mode. The dots located close to one another

represent spectra with similar metabolite abundances. Clusters could be compared among each other to find differentially abundant metabolites. For example, the ion, putatively annotated as tetrasaccharide, was found to have 3-4-fold higher intensity in two clusters compared to the rest.

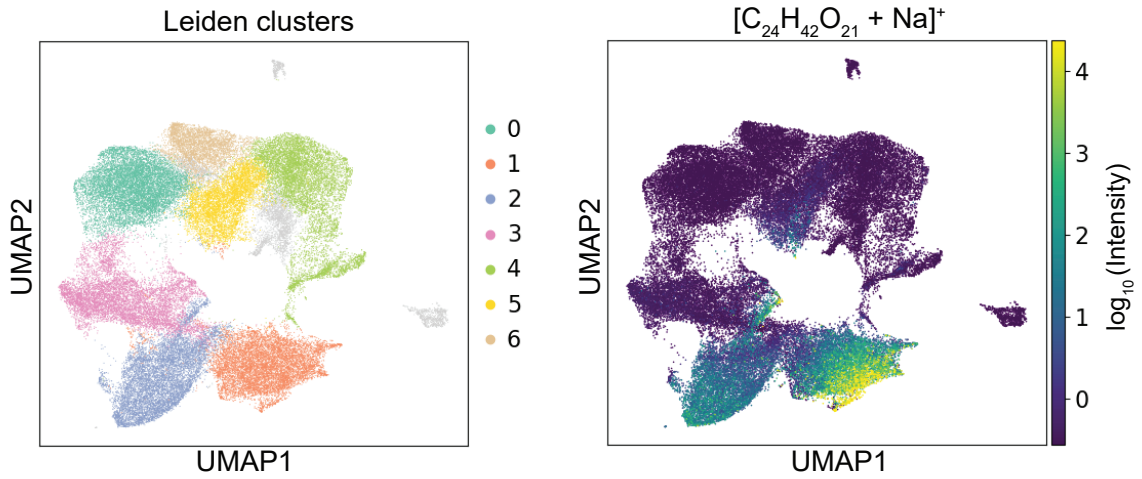


Figure 3.8 Left: UMAP plot for metabolites of one of the small intestine samples. One dot in the plot corresponds to one reduced mass spectrum in the data. Dots located close to one another represent spectra with similar metabolite abundances. The colour shows the partition of spectra into the Leiden clusters. Right: The same plot coloured by the intensity of a tetrasaccharide ion. The ion is more abundant in clusters 1 and 2. Similarly, intensities of other metabolites also differ between clusters. Clusters are defined only based on metabolite intensities.

Since each spectrum was collected from a known (X, Y) location of the sample, spatial coordinates of clusters can be retrieved and visualised. It could be seen in figure 3.9 that spectra that cluster together in the UMAP plot, originate from distinct regions in the tissue. The regions consist of connected pixels, and form structures of different sizes, which resemble anatomical features. Plotting the same example of a tetrasaccharide ion reveals an intensity gradient within the tissue.

The analysis of the MS imaging data thus reveals spatial differences in the metabolite composition of the gut tissue.

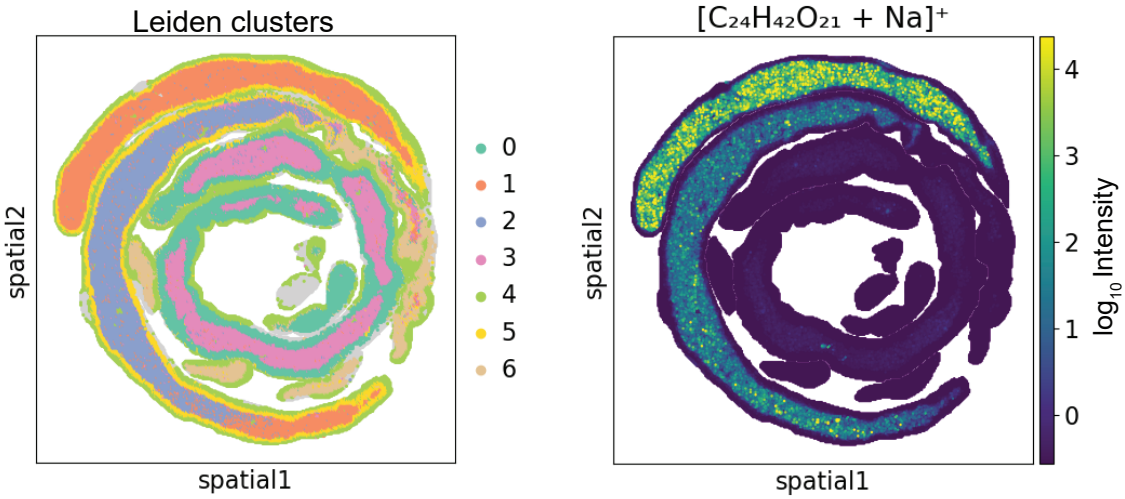


Figure 3.9 Left: UMAP clusters plotted in spatial coordinates. Right: intensity of a tetrasaccharide ion in spatial coordinates. Cluster colours used in this figure match cluster colours in previous figure 3.8.

3.2.4 Anatomical interpretation of spatial differences

To find which anatomical regions correspond to the clusters, the spatial plots were compared to the optical images of tissue sections recorded prior to MS imaging (the optical images are shown in figure 3.10).

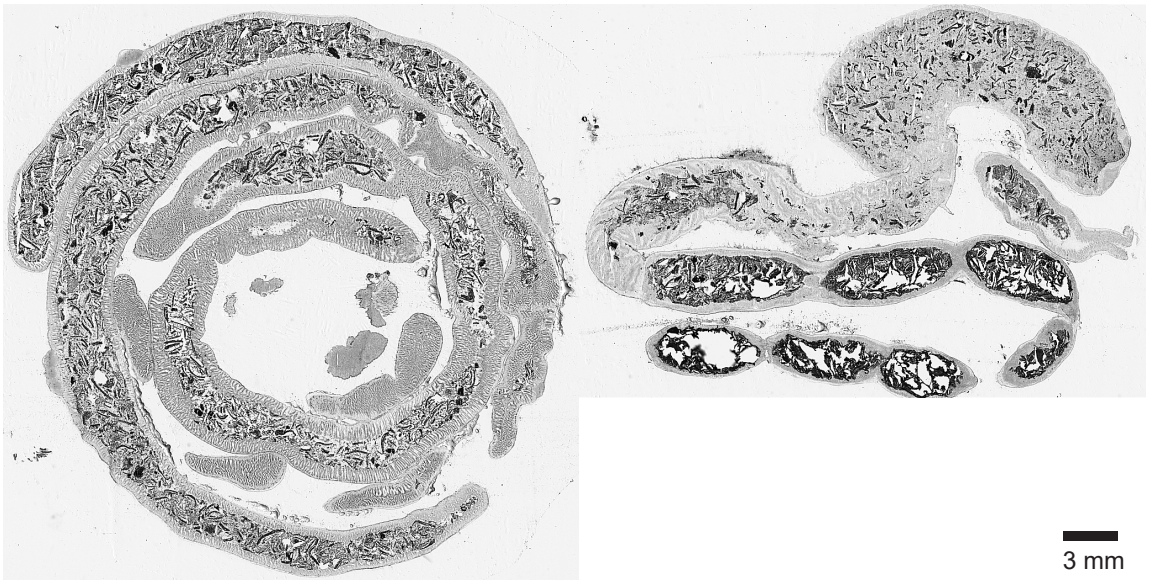


Figure 3.10 Transmission light microscopy images of tissue sections of small intestine (left) and large intestine (right).

To accurately perform image registration of the optical images and the MS imaging data sets, a set of points corresponding to the same morphological features was chosen in one optical image and one ion image from the data set, using Affinder - a plugin to the python image viewer Napari [94]. The points were used to compute the affine transformation matrix. Then, this matrix was used to scale, rotate and translate the optical image to match the spatial coordinates. Exploring interactive plots of clusters overlaid with optical images allowed matching clusters with anatomical features. As a reference, anatomical regions are schematically shown in figure 3.11. In general, the spatial arrangement of clusters was the same for the tissue sections coming from the same animal imaged in positive and negative ionisation mode. In addition, very similar clusters were identified in all males and females, except for a few low-quality data sets with high number of signal drop-outs.

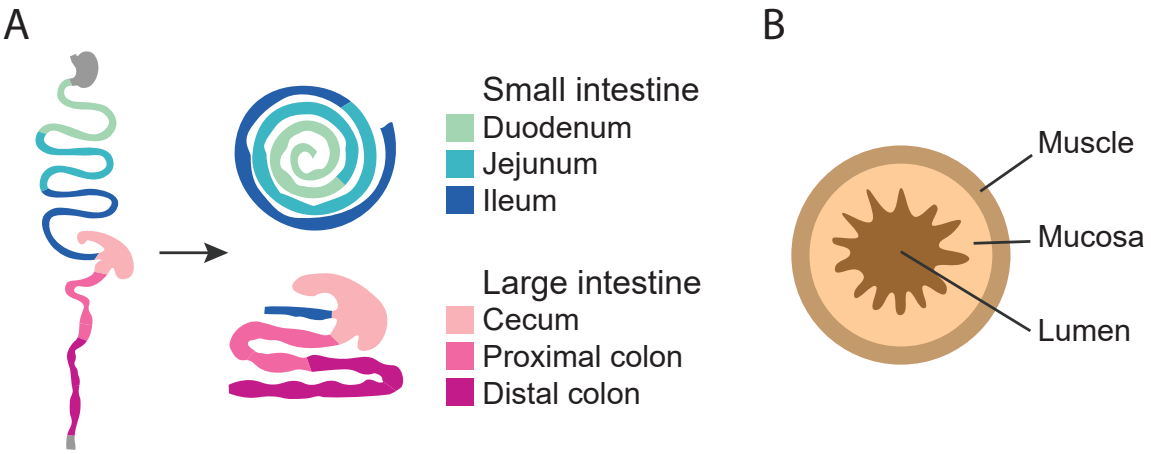


Figure 3.11 Schematic representation of anatomical features: (A) along the gut and in tissue sections, (B) in cross-section of the intestinal tube.

In the small intestine, as shown in figure 3.12, three major clusters aligned with the faecal-matter filled lumen of the gut: in duodenum, jejunum and ileum. The gut of the animal, for which the results are shown, was empty between duodenum and jejunum sections, but not between jejunum and ileum, where a transition of one cluster into the other was visible. Interestingly, the clustering approach also captured the differences in metabolic composition in the cross-sectional direction.

Namely, clusters of the mucosa and outer muscle layer were separated from one another and the luminal clusters. Mucosa was represented by two different clusters: one in the duodenum region, where the mucosal layer is thicker, and the other in the jejunum and ileum, where the mucosal layer is thinner. The metabolic differences between the two mucosal clusters might be explained by the fact that many nutrients are absorbed in the beginning of the small intestine.

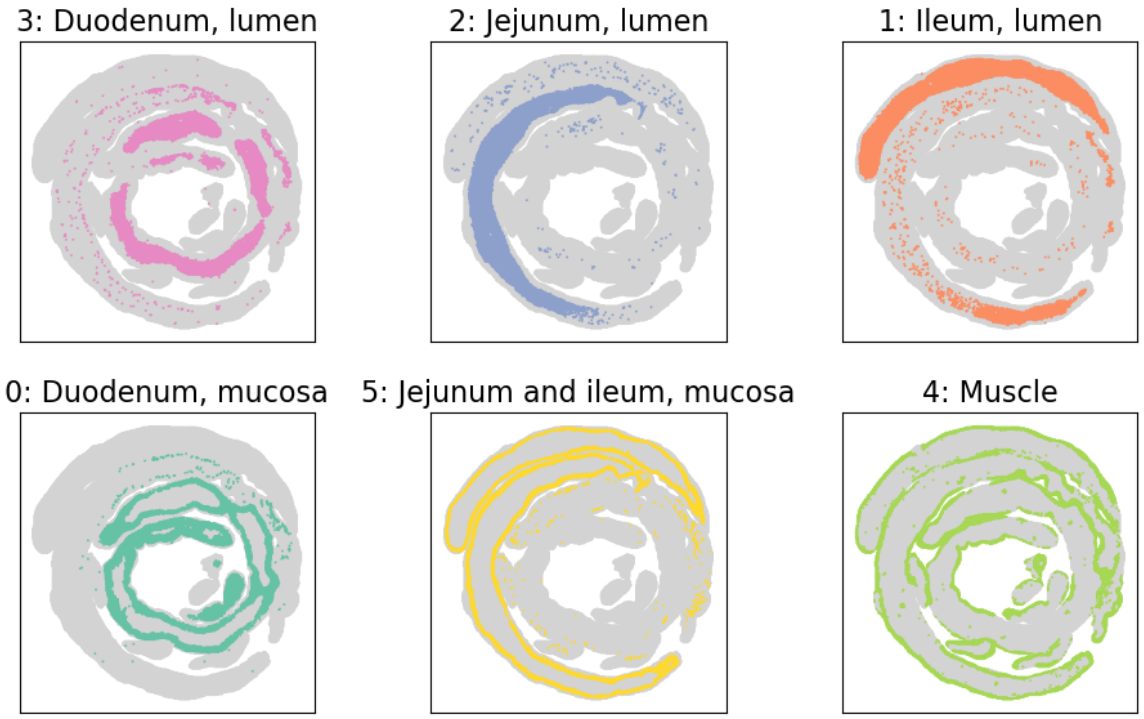


Figure 3.12 Main Leiden clusters of small intestine plotted in spatial coordinates and annotated.

As seen in the small intestine, the large intestine section shows (figure 3.13) three clusters aligned with the lumen of the gut: in the terminal ileum, cecum, and colon. One cluster aligned with the outer muscle layer. The mucosa of the ileum was represented by a different cluster than the mucosa of the rest of the large intestine. Unexpectedly, in the higher-resolution data sets, a cluster that was not visible in the lower resolution data sets was found to align just on top of the mucosal layer in the cecum. As shown in figure 3.14, this cluster was frequently only 1-2 pixels wide.

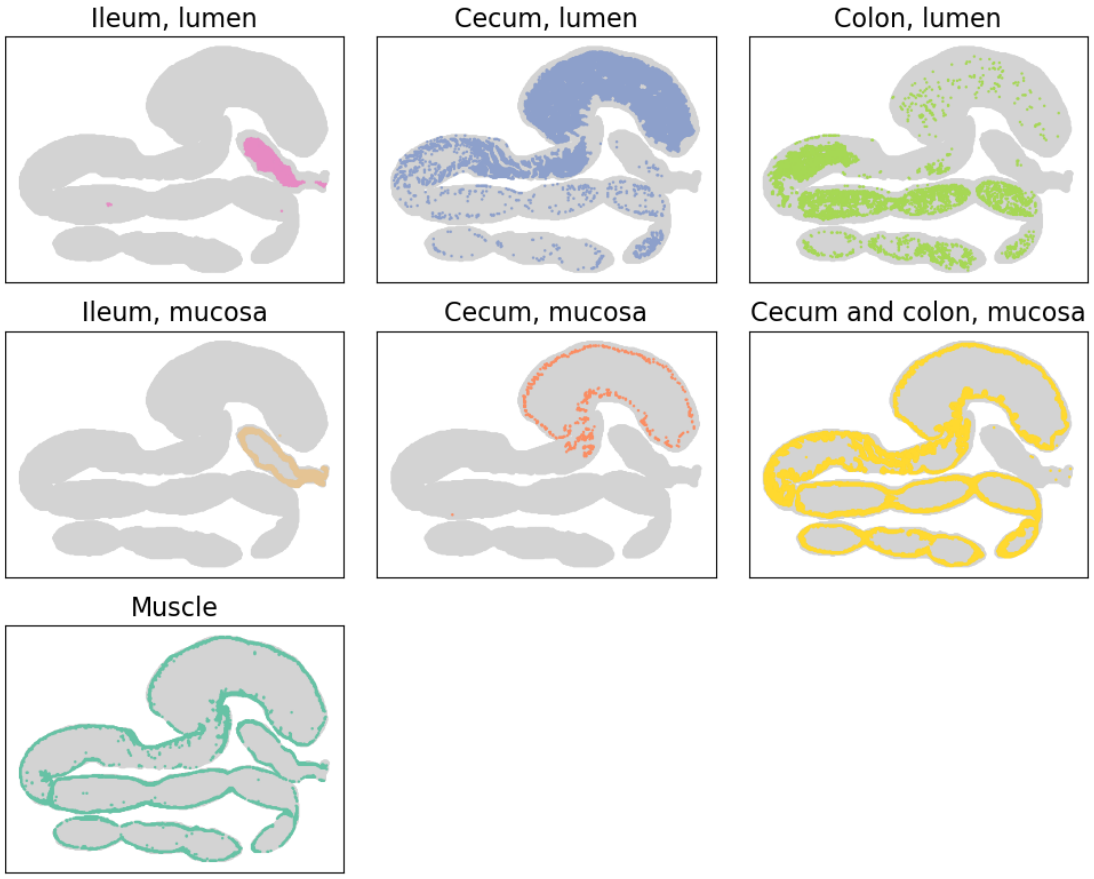


Figure 3.13 Main Leiden clusters of large intestine plotted in spatial coordinates and annotated.

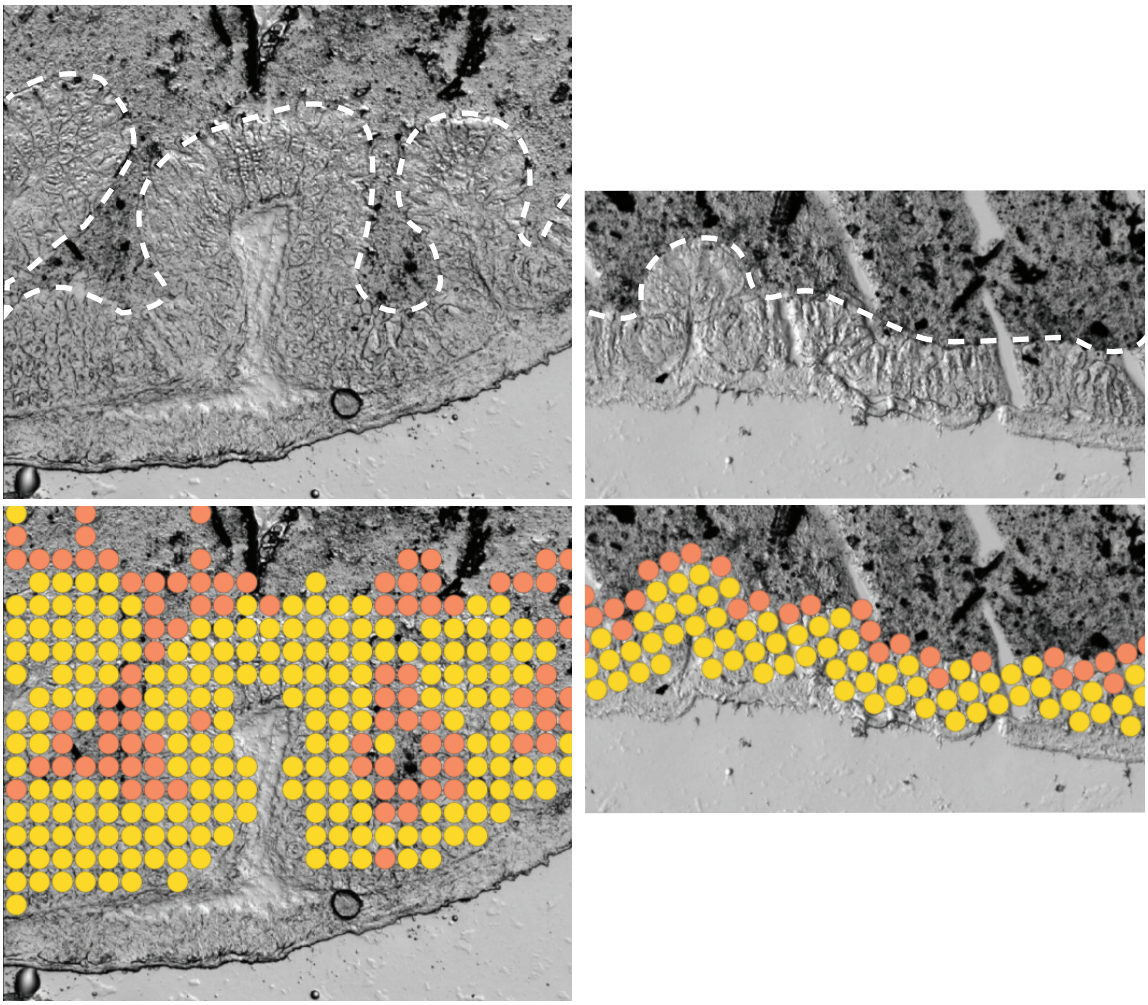


Figure 3.14 Two examples of the mucosa-associated cluster in the cecum. Top: optical image of the cecum, dashed line shows the boundary between mucosa and lumen. Bottom: optical image with overlaid clusters; yellow represents mucosa found in the entire large intestine, red found only in the cecum.

Collectively, these findings demonstrate that MALDI-MS imaging is capable of detecting differences in metabolite abundances in different parts of the mouse intestine, despite lacking ion separation and sampling significantly less material compared to bulk MS techniques. In addition, using high spatial resolution MS imaging is advantageous to resolve metabolically distinct regions at a small scale, such as mucosal or mucosa-associated regions. Going forward, it might be beneficial to perform imaging of a part of the tissue at higher spatial resolution (for example, 10 micrometres) to reliably visualise fine features.

3.2.5 Biological interpretation of spatial differences

Differences in the metabolite composition between tissue areas corresponding to the clusters in the lumen of the gut are likely driven by both host physiology and microbial metabolism, as was suggested by a study by Yamamoto et al.[84]. In their study, the authors showed that the abundances of certain metabolites vary along the gut, and that there are several common patterns that metabolite abundances can follow. Different patterns in metabolite abundance were observed for either germ-free or regular mice. Consequently, the authors suggested that both host processes (e.g. metabolite digestion by host enzymes and absorption) and bacterial activity can contribute to spatial variations in gut metabolite composition.

In order to accomplish biological interpretation of observed clusters, a two-step analysis was carried out. Only a part of the tissue corresponding to luminal clusters was considered. First, a pairwise comparison of each cluster against all other clusters was carried out to identify metabolites with at least a 5-fold change in intensity. This threshold was selected to capture the most differentially abundant metabolites and to exclude metabolites that show little change cluster-to cluster. 689 differentially abundant metabolites were identified. Next, the part of the tissue mask, corresponding to the luminal clusters was segmented into small areas using the image processing package Fiji, as shown in figure 3.15. To each segment, a unique number was assigned, starting from 1 (beginning of the intestine) and increasing with the distance from this starting point (thus, the 1st segment corresponds to the beginning of the duodenum, whereas the last segment corresponds to the end of the distal colon). As a result, the numbering represents the longitudinal axis of the gut, along which the metabolite intensities could be measured. For example, here shown is a variation in the intensity of a tetrasaccharide ion between intestinal segments.

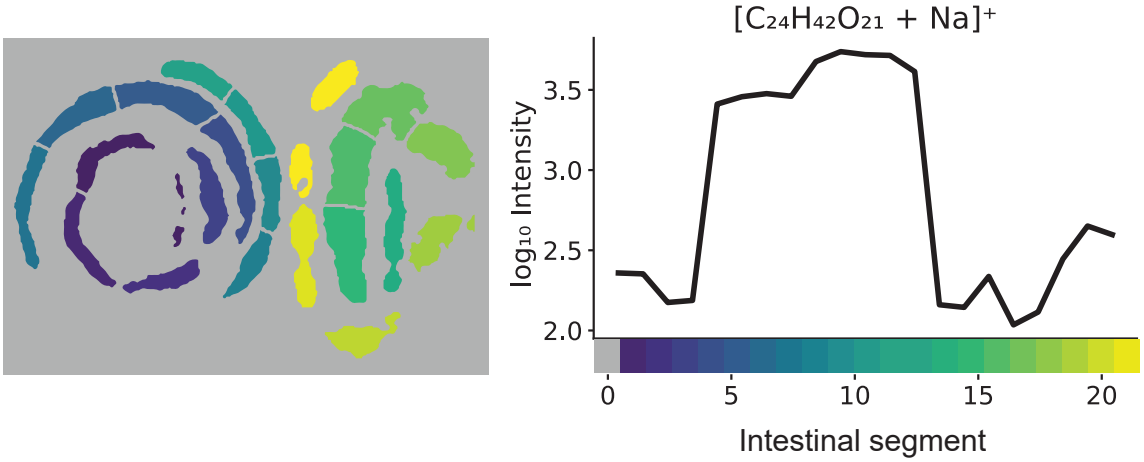


Figure 3.15 Left: segmentation of the tissue mask corresponding to the lumen of the gut into areas. Right: mean intensity of tetrasaccharide ion per intestinal segment; dashed line shows the transition between small and large intestine. The plot shows an accumulation of the ion in jejunum and ileum before a sharp drop at the entrance to cecum.

To be able to make a comparison between all metabolites, metabolite intensities were log-transformed, averaged per intestinal segment, then scaled to zero mean and unit variance (as in [84]), and subjected to hierarchical clustering. Hierarchical clustering generates a dendrogram of hierarchical relationships between all metabolites. Extracting the clusters using a chosen distance threshold, yields groups of metabolites whose intensity changes in a similar way along the gut. The two following figures show the analysis results. The overview figure 3.16 shows a heatmap of metabolite intensities along the gut, partitioned into nine groups. To aid the interpretation of the groups and the associated intensity trends, intensities of metabolites present in each group were averaged and the groups were reordered. The resulting condensed heatmap is visualised in figure 3.17. Furthermore, groups were given informative names.

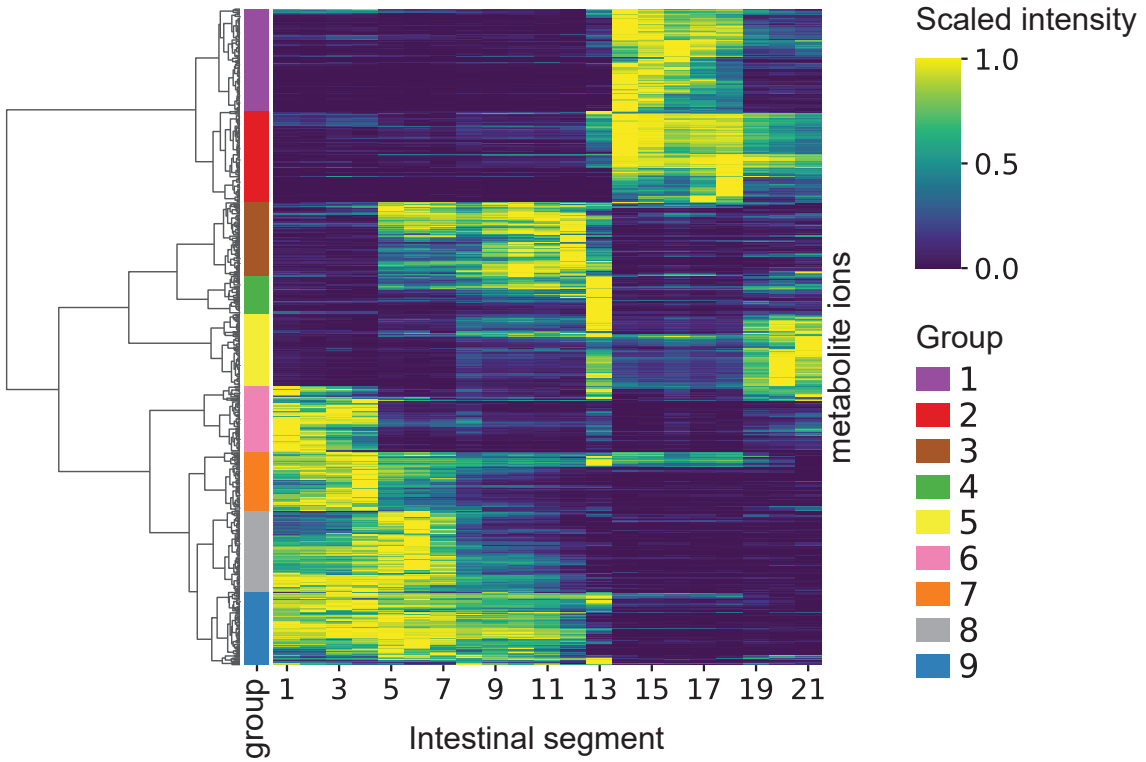


Figure 3.16 Heatmap with hierarchical clustering of scaled metabolite intensities along the gut. The X-axis represents the intestinal segments arranged in anatomical order along the GI tract. The Y-axis represents individual metabolite ions, partitioned into groups by similarity. Scaling of intensities includes 3 steps: log-transformation, scaling to zero mean and unit variance (to compute the dendrogram), and normalisation to fit [0,1] range (only for plotting).

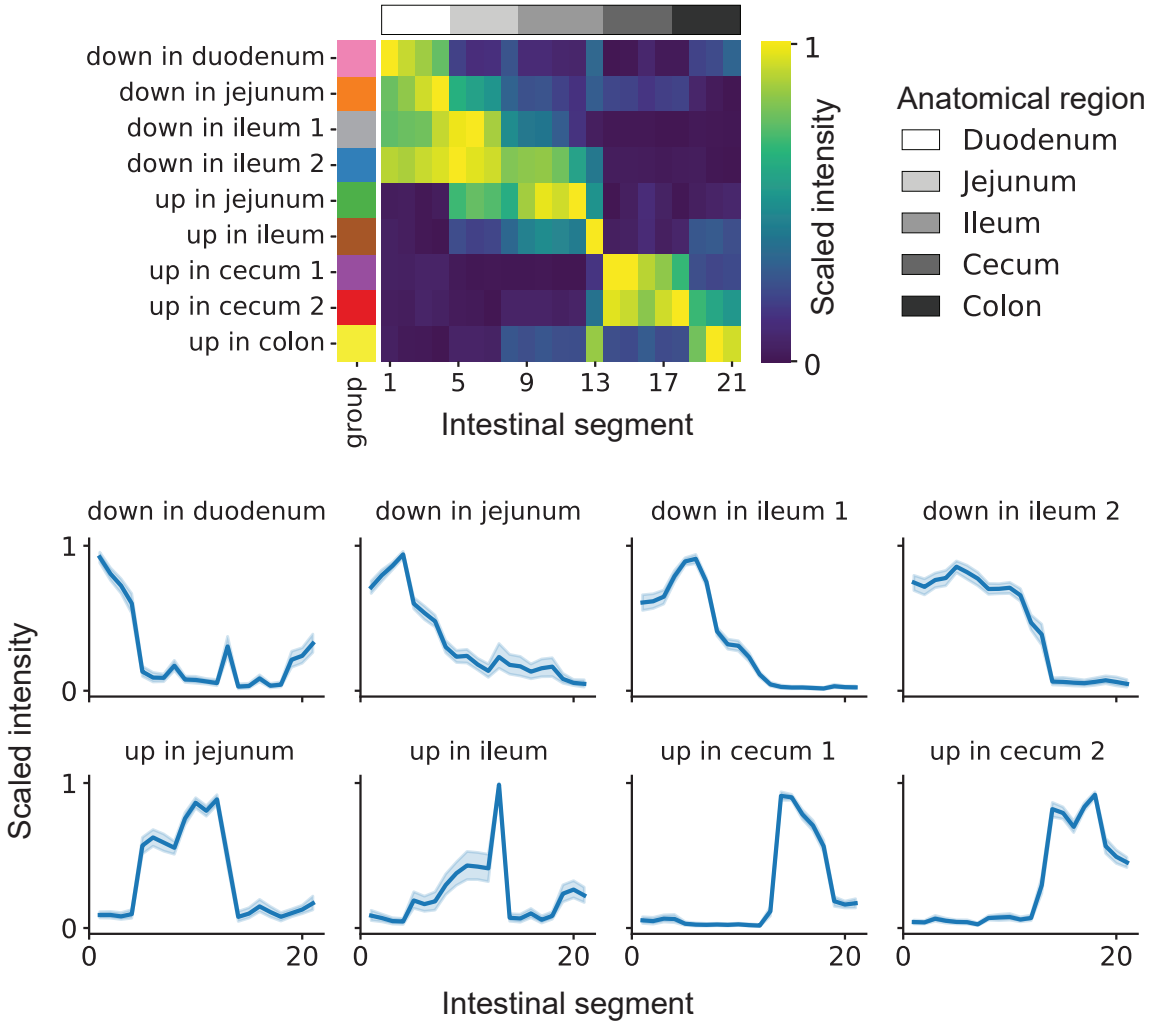


Figure 3.17 Condensed heatmap. Top: heatmap from figure 3.16 in condensed format with metabolite intensities averaged per group. Colours on the left-hand side match the group colour in figure 3.16. Bottom: plots of average intensities of all metabolites in a group, along the gut tube.

Three major trends were observed for metabolite ion intensities along the gut. The list of metabolites constituting each group can be found in Appendix H

- **Depletion in upper small intestine** Groups called ”down in duodenum” and ”down in jejunum” (figure 3.17) represent ions that decrease in abundance in the small intestine. Many lipid and lysolipid ions were detected in the ”down in duodenum” group. Indeed, the beginning of the small intestine is said to be the site of digestion and absorption of dietary lipids, where enzymes of bile and pancreas break down triglycerides and phospholipids into fatty acids.[95]

Digestion of proteins also takes place in this part of the gut, producing amino acids, some of which are absorbed. For example, ions of amino acids (e.g. Lys, Pro, Asn) were detected in group ”down in ileum”. Ions of groups ”down in ileum 1” and ”down in ileum 2” are depleted later towards the ileum. For example, terminal ileum is known to be a site of an active absorption of conjugated bile acids. A variety of conjugated bile acids ions were detected in these two groups.

- **Accumulation in the small intestine and depletion in the large intestine** Occurs in groups ”up in jejunum” and ”up in ileum” (figure 3.17). Ions of the group ”up in jejunum” increase in abundance between jejunum and ileum. An abundance of ions of the group ”up in ileum” rises sharply in the ileum. Both are depleted in the cecum. The accumulation of smaller metabolites could occur as digestion enzymes and bacteria break down proteins, carbohydrates and lipids. The cecum is known to have a rich and dense bacterial community, which could explain the sharp depletion of the accumulated metabolites as they reach this area of active microbial fermentation. For example, this behaviour was observed for a range of oligosaccharide ions detected in group ”up in ileum 1”. Interestingly, long-chain fatty acyl-coenzyme A (CoA) ions accumulated in group ”up in ileum 2”

- **Accumulation in the large intestine** Ions of groups ”up in cecum 1” and ”up in cecum 2” (figure 3.17) increase in abundance in the cecum. In the former, ions are depleted in the early proximal colon, while in the latter ions persist, and are excreted with faeces. The emergence of these ions could be a consequence of the metabolic activity of microbes. For example, nucleobase deconjugation from nucleotides by gut bacteria has been observed in cecum.[96] The ions of nucleobases were also detected in this study in both groups. The depletion of metabolites in group ”up in cecum 1” is likely driven by the absorption by the host or uptake by the bacteria.

In summary, this data suggests that spatial differences in metabolite composition in the gut tissue are driven by both physiological processes and microbial metabolism. The obtained results are in agreement with the study of Yamamoto et al.[84], that observed the same three trends as described above in regular mice. This study, however, did not have the spatial resolution to differentiate between finer patterns of metabolite accumulation and depletion. Unfortunately, the metabolite localisation data presented in this study is not publicly available, making further comparison impossible.

To gain a better understanding about the contributions of individual metabolites, the abundance patterns in each chemical class were examined and the findings were compared with what is known from the literature. Individual chemical classes are discussed as follows.

Lipids Lipid digestion in mice takes place in the stomach and small intestine. Dietary triglycerides and phospholipids are broken down into fatty acids and absorbed in the beginning of the small intestine.[95] Comparing the number of annotated lipids ions in each group showed that indeed the majority of the detected lipids come from the group ”down in duodenum”, meaning that these ions deplete early in the duodenum (figure 3.18). The majority of detected lipid ions were annotated as phospholipids. There were no triglyceride annotations among the considered groups of metabolites, although a few diglycerides were detected. A substantial number of lipid annotations were also found in groups ”up in cecum 1” and ”up in cecum 2” (metabolites that accumulate in the cecum). Among these ions, the majority were annotated as phosphatidylglycerols (PG) and phosphatidylethanolamines (PE). The reason for the accumulation of these lipids is unclear. One possibility is that these are bacterial lipids. PG and PE lipids, as well as bacterial cardiolipins (CL) that can be broken down into PG and PE lipids, are among the most abundant components of bacterial membranes.[97] The fact that the density of bacteria increases substantially going from

the ileum into the cecum, supports this interpretation. Host CL lipids, present in the gut as a result of epithelial cell turnover, could also give rise to PG ions, but not PE ions.[98] Interestingly, all ions that were highly differentially abundant in the mucosa-associated cluster in the cecum (discussed in figure 3.14), were annotated as PE lipids.

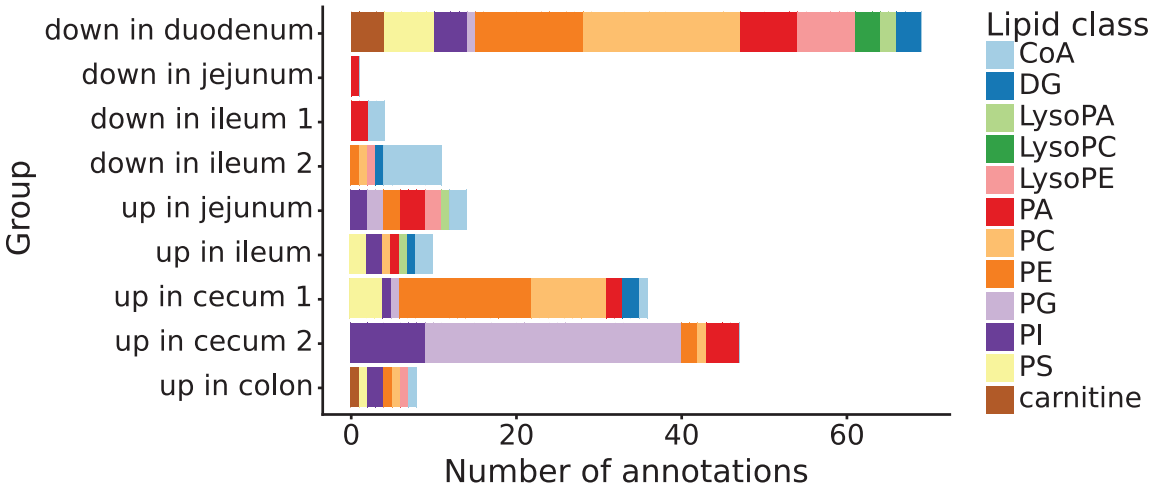


Figure 3.18 Lipid annotations among metabolites of each group. Most ions were annotated as lipids in groups ”down in duodenum” (metabolites that deplete through duodenum), ”up in cecum 1” and ”up in cecum 2” (accumulation in cecum).

Bile acids Bile acids play an important role in lipid solubilisation. Produced in the liver from cholesterol, primary bile acids are delivered into the duodenum in the conjugated form (e.g. taurocholate and glycocholate).[99] Conjugated bile acids are toxic to most microbes due to their surfactant properties. Bacteria deconjugate the amino acid using bile salt hydrolases. The resulting cholate, in contrast to the conjugated forms, is not reabsorbed and not recycled back into the liver from the small intestine. Instead, it arrives in the cecum where it is subjected to further microbial fermentation. Three fermentation reactions are known: dehydroxylation (to produce deoxycholate), oxidation (to produce oxocholate) and microbial reconjugation of bile acid to another amino acid (to produce, for example, tyrosocholate).[99] It is conceivable that a combination of fermentation reactions could occur yielding a variety of products. The modified (secondary) bile acids are reabsorbed in the colon and delivered back to the liver by enterohepatic circulation. The

fate of the secondary bile acids in the liver is to be conjugated (if not conjugated already) and, possibly, sulfonated, if the secondary bile acid is toxic to the host.[100] Secondary bile acids modified by the liver is secreted again into the intestine, which means that the bile acid pool may become quite diverse.

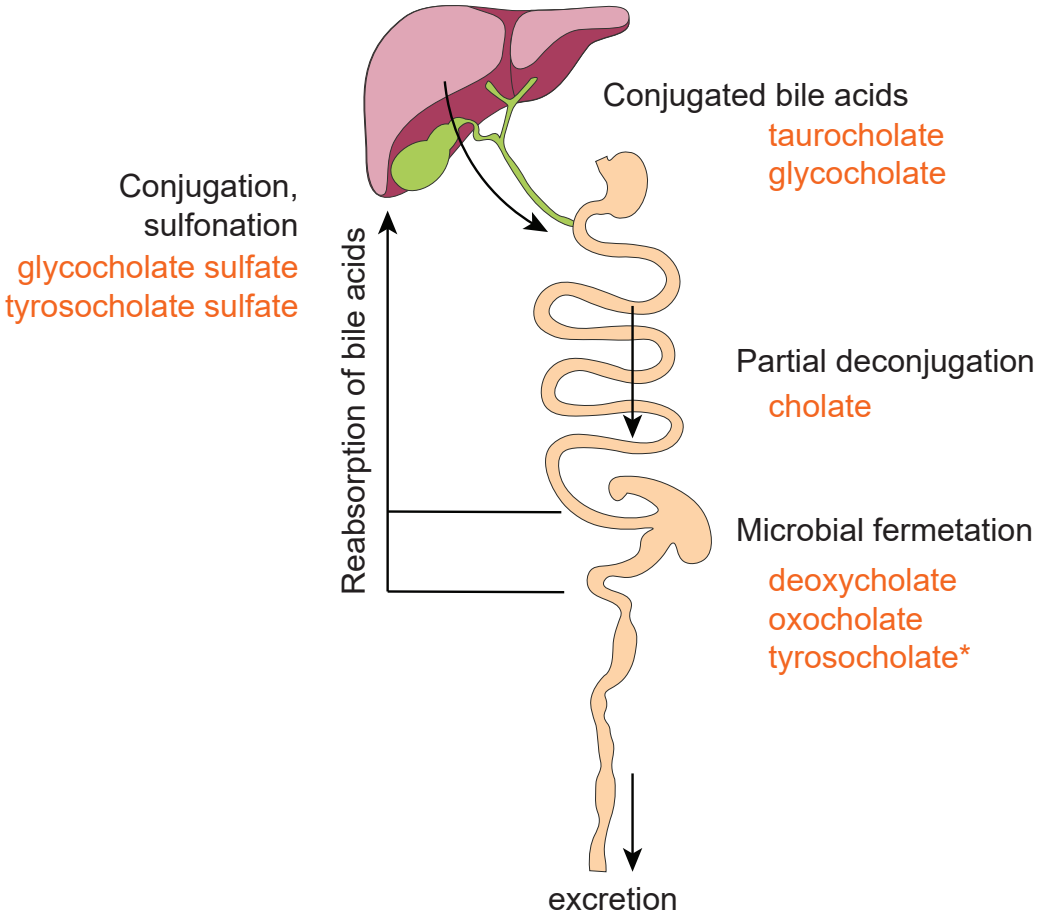


Figure 3.19 Schematic representation of bile acid metabolism. Examples of bile acid molecules are shown in red, ions corresponding to each example except for (*) were detected, as shown in the next figure.

Bile acids for every discussed step were detected, as shown in figure 3.20. Taurocholate and glycocholate were abundant through the whole small intestine and depleted in the ileum, as expected. Cholate was found in the entire gut. Other ion annotations should be interpreted with caution since many species have isomers. Secondary bile acid deoxycholate was detected in the large intestine. The abundance of sulphonated secondary bile acids in the small intestine (glycodeoxy-

cholate sulphate, tyrosocholate sulphate, glycolithocholate sulphate) might serve as indirect evidence of production of secondary bile acids by the microbiome. The sulphonated acids were not detected in the large intestine, likely because some bacteria are known to have desulphonases that could remove the sulphate functional group.

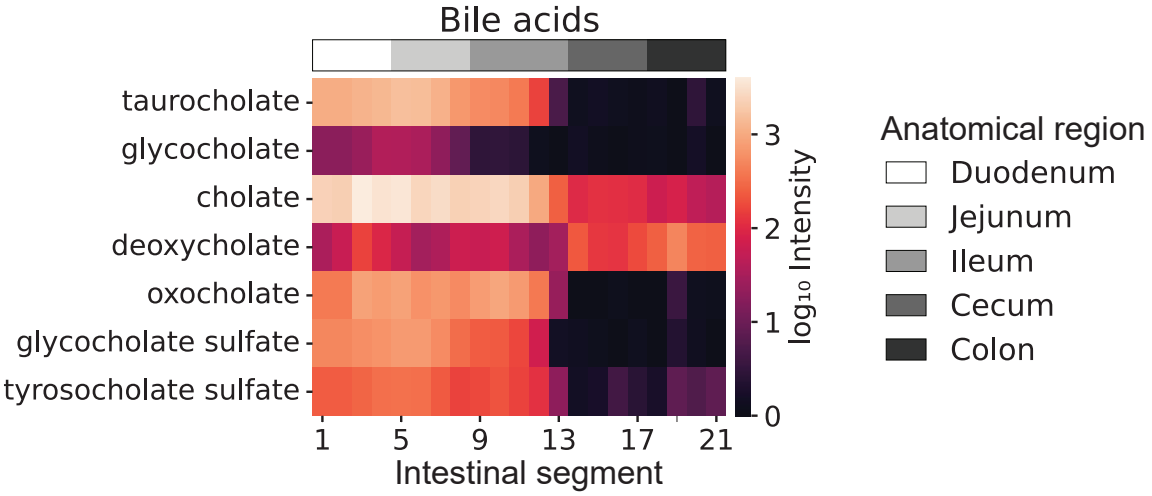


Figure 3.20 Abundance of bile acid ions along the gut.

Carbohydrates Complex carbohydrates are broken down into monosaccharides that are both absorbed by the host and used as a carbon source by the gut microbiome.[101] Indeed, a series of oligosaccharides ions that were observed in the mass defect plot (figure 3.6), were found to accumulate through jejunum and ileum, as shown in figure 3.21. Similarly, sugar alcohols and sugar acids increased in the same gut region. The level of monosaccharides that entered the gut was seen to drop through the duodenum and jejunum. Since MS1 acquisition does not allow distinguishing between isomers, the observed trends are the summed result of several ions that each have, potentially, unique dynamics. For example, glucose might have been depleted, while less digestible fructose might persist at low levels throughout the gut.

In the large intestine, where simple sugars are mostly depleted, certain bacterial species feed off metabolising digestion-resistant starch and fibres into short-chain fatty acids (SCFA): acetate, propionate and

butyrate.[101] These metabolites were not detectable by the MALDI-MS sample preparation protocol used in these experiments, however accumulation of hexose phosphate, an intermediate in glycolysis and SCFA production, was observed. Other microbes break down mucin - the glycosylated protein layer produced by the cells lining the intestine - to use as a carbon source.

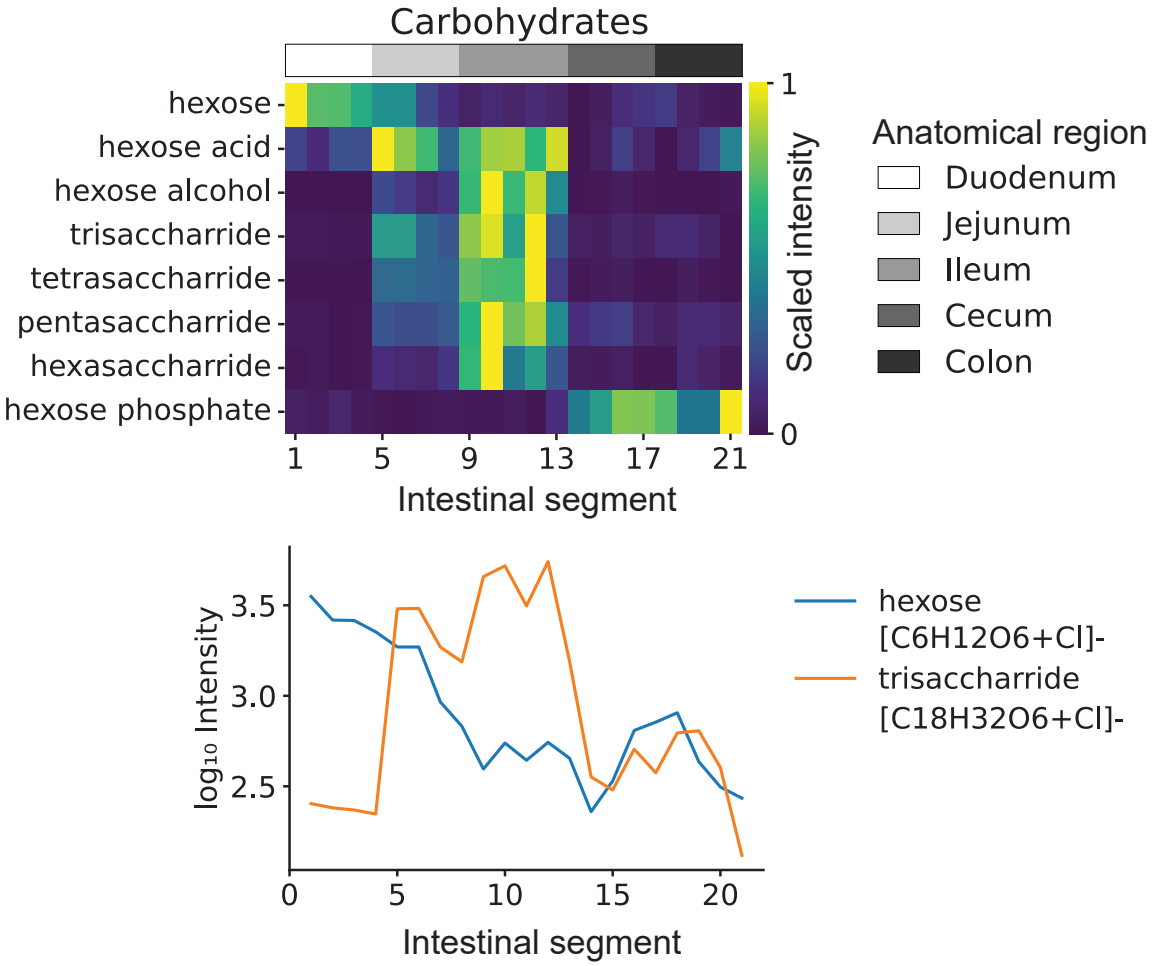


Figure 3.21 The abundance of carbohydrate ions along the gut. Left: heatmap of ion intensities scaled to zero mean and unit variance. Right: plot of log-transformed (unscaled) intensities for representative ions. Hexose ions (such as glucose) are depleted in the duodenum. Oligosaccharides, sugar alcohols and sugar acids accumulate through the duodenum and jejunum before being consumed in the large intestine. Sugar phosphates accumulate in the large intestine.

Amino acids Amino acids are produced in the intestine during protein digestion. While most are absorbed in the small intestine, some are metabolised by bacteria. For example, through decarboxylation, a variety

of amines is produced.[102] Detected in this study were polyamines (produced from arginine), tyramine (from tyrosine) and phenylethylamine (from phenylalanine). The source amino acids peaked in abundance at the end of the duodenum and were depleted throughout the rest of the small intestine. The abundance of corresponding amine products rose sharply upon entry to the cecum. For tryptophan, a series of known[103] indole derivatives was detected in the same region of the gut (figure 3.22).

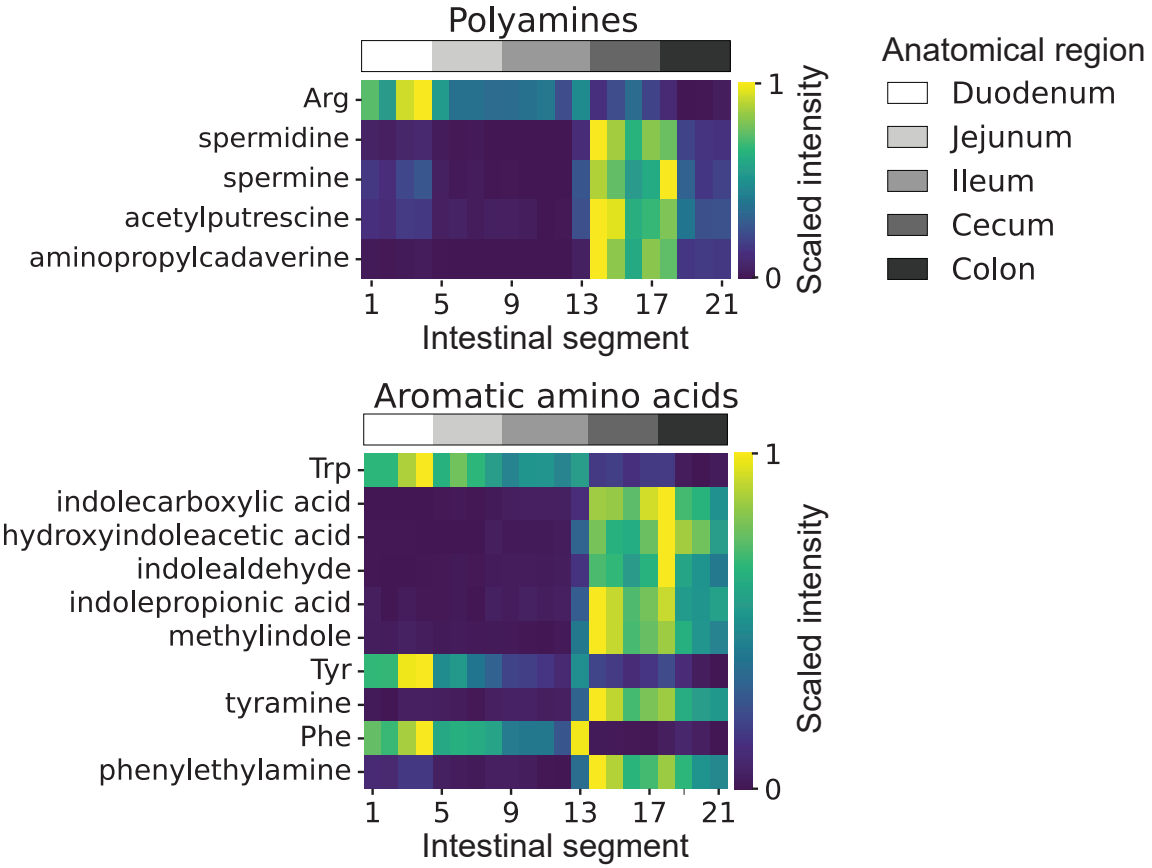


Figure 3.22 Abundance of ions related to amino acid digestion along the gut.

Nucleotides Among the differentially abundant ions with the high N/C ratio, several point toward microbial metabolism of nucleotides. Bacterial nucleotide phosphorylases cleave nucleotides into ribose phosphate and nucleobase in the cecum.[96] Ions of all five nucleobases (A,T, C, G, U) were found to accumulate in the cecum. While ribose phosphate was not detected, phosphoribosylamine was found to accumulate at the same

rate as nucleobases. It could be hypothesised that this compound is produced from ribose phosphate. Nucleobases could be used by bacteria for nucleotide synthesis, or absorbed by the host, and were found to be depleted in the proximal colon. Examples of these findings are shown in figure 3.23.

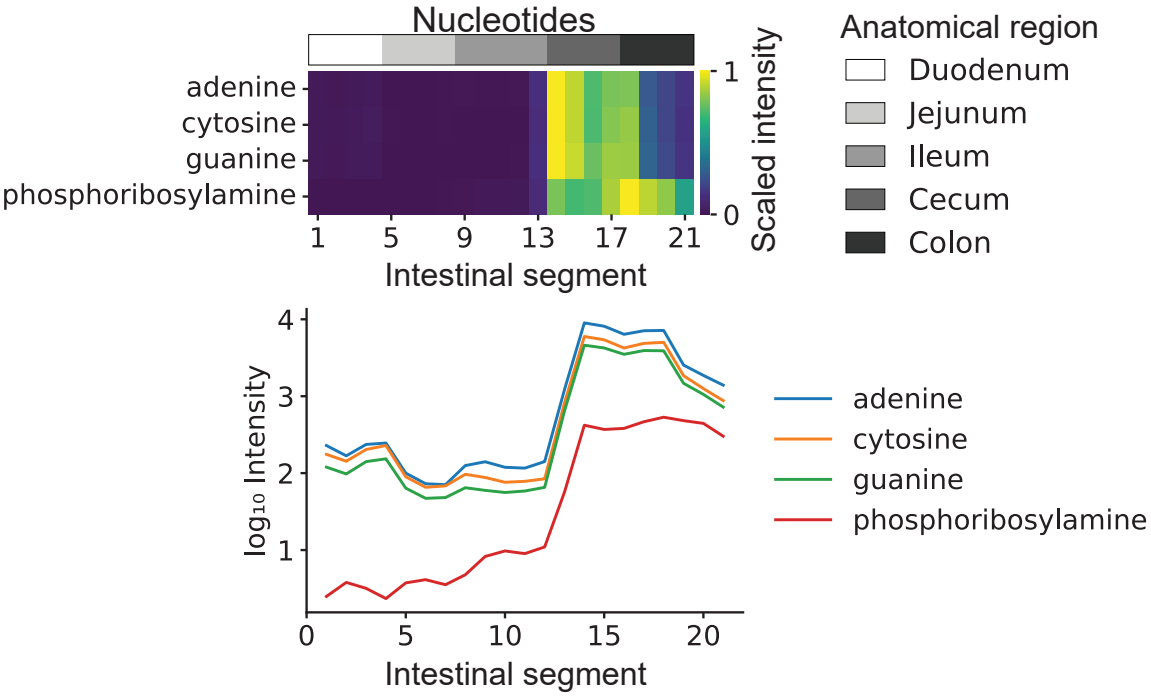


Figure 3.23 Abundance of some ions related to nucleotide digestion along the gut.

In conclusion, the biological significance of five clusters of MS spectra that align with the different parts of the lumen of the gut (duodenum, jejunum, ileum, cecum and colon) was investigated. The metabolites that were most differentially abundant between the clusters covered a variety of chemical classes and were found to follow three main patterns. The first pattern was faster or slower depletion upon entry into the small intestine (for example, for most lipids, and monosaccharides) as a result of absorption by host tissue or catabolic processes. The second pattern was the accumulation of the products of catabolism in the jejunum (for example, oligosaccharides, amino acids) and their reabsorption at the end of the ileum. The third pattern was another round of biochemical transformations leading to the accumulation of metabolites in the cecum

(for example, amino acid derivatives, secondary bile acids). In both small and large intestines evidence of known bacteria-facilitated metabolic transformations was observed. Two further unexplained findings were made: the accumulation of PG and PE lipids in the lumen of the cecum and the accumulation of a spatially localised group of PE lipids close to the cecal mucosa layer.

3.2.6 Longitudinal probing of the gut by bulk sequencing

Regions of distinct metabolism and gradients in the abundance of some metabolites along the gut, discussed in the previous section, arise from a combination of host physiology and bacterial metabolic activity. Studies in the past have observed variation in microbiome composition between different parts of the gut.[78, 79] Nevertheless, it is not well understood what is the relationship between spatial differences in bacterial abundance and spatial differences in the gut metabolome. To bridge this gap, an integrated multi-omics approach combining metabolomics and sequencing data was undertaken.

In published data, there is a lot of variation in the microbiomes detected in mouse intestines, likely due to a variety of factors such as animal strain, origin, age, sex, differences in housing and chow, sample preparation and sequencing analysis. For this reason, sequencing was performed on tissue from the same animal that was used for metabolite MS imaging. As explained in Section 3.2.2, the tissues leftover after tissue section preparation were thawed, and cut into 1-2 cm pieces. Together with collaborators (see the Contributions section), the luminal contents of each piece were subjected to 16S ribosomal RNA gene amplicon sequencing, obtaining a list of 101 unique bacterial genera and their relative abundances in samples collected from 4 animals.

Comparing the sequencing results from all 4 animals (in figure 3.24) showed that the main difference between samples was driven by their anatomical origin. In each animal, samples grouped into two main clusters by similarity: samples from the small intestine (duodenum, jejunum, ileum) and samples from the large intestine (cecum, colon). The second important driver of variance was the biological replicate. Females that were co-housed had much more similar microbiome composition than either of the males, raised in separate cages. Due to this inter-animal variability, correlation analysis was performed separately for each mouse.

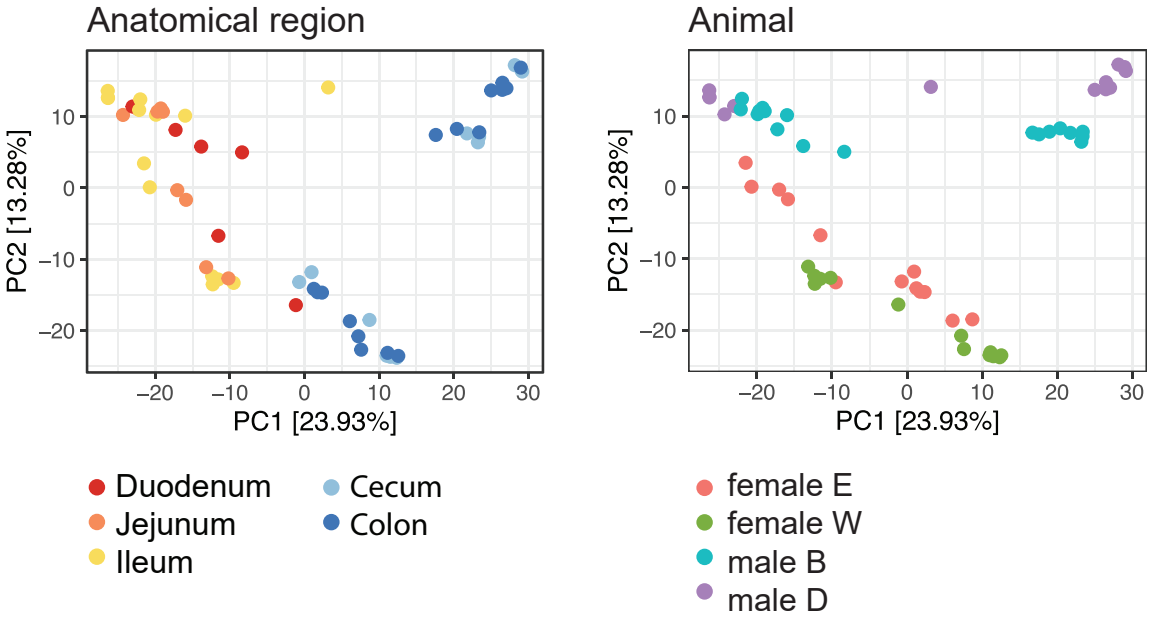


Figure 3.24 PCA score plots of sequenced samples. The experiments and analysis were performed together with A. Milanese and B. Zhang.

Figure 3.25 shows the relative abundance of bacterial genera in the gut of the same animal for which metabolomics results were presented in the previous sections. Some genera, including the most abundant ones, were observed throughout the whole intestine. In the small intestine, 6-7 dominant genera were detected, while the bacterial diversity in the large intensity was found to be richer (by measuring the Shannon index, data not shown). These findings are in line with published data.[80]

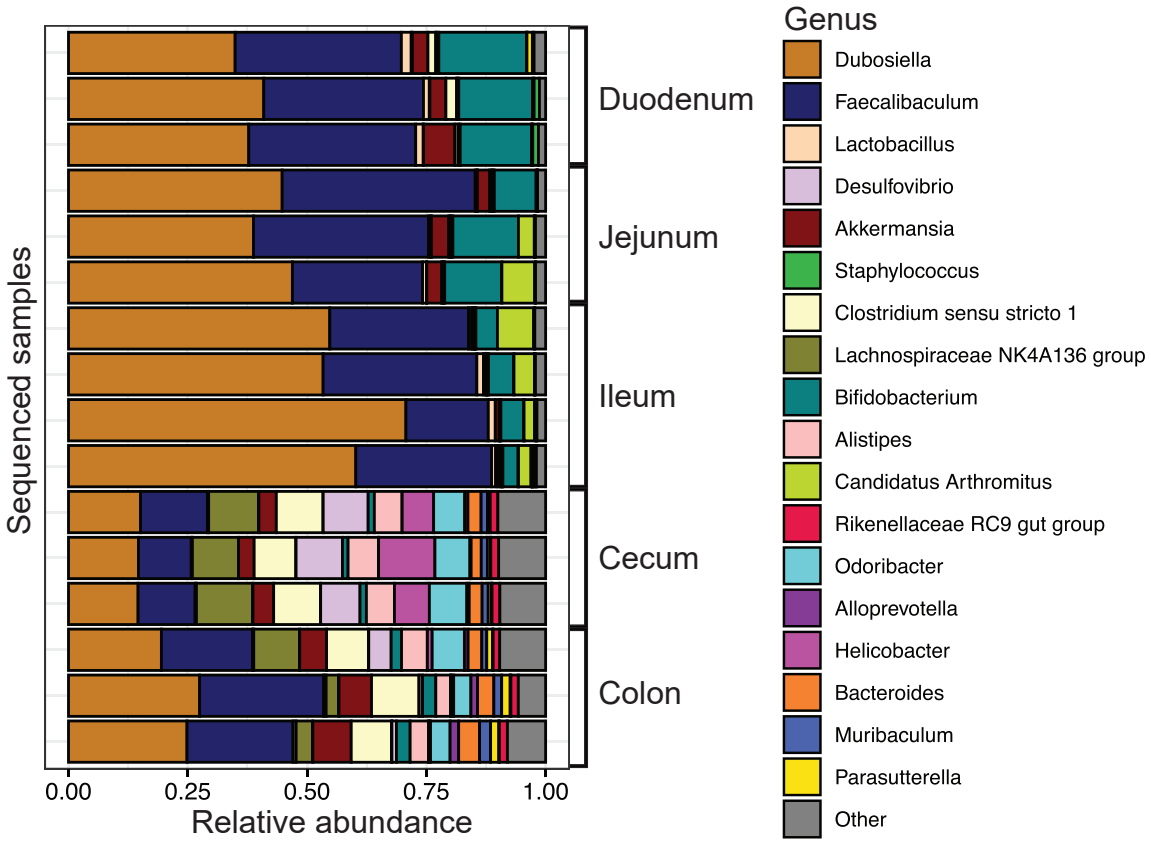


Figure 3.25 Relative abundance of 20 most abundant bacterial genera in the mouse intestine. These data correspond to the same animal for which metabolomics results were presented in the previous sections. The experiments and analysis were performed together with A. Milanese and B. Zhang.

Genera detected predominantly in the small intestine included: *Staphylococcus, Streptococcus, Lactococcus, Ligilactobacillus* (class *Bacilli*, phylum *Firmicutes*), *Candidatus Arthomitus* (phylum *Firmicutes*), *Afipia* and *Ralstonia* (phylum *Proteobacteria*) and *Bifidobacterium* (phylum *Actinobacteria*). The intestine is known to have a steep oxygen gradient between the stomach and rectum. Thus, the appearance of obligate or facultative aerobes, such the observed *Bacilli* strains, is expected in this region of the intestine. *Candidatus Arthomitus* (detected only in males) was previously observed in the terminal ileum by electron microscopy. It belongs to the group of segmented filamentous bacteria, which have the unique property to penetrate the mucus layer and bind to ileal epithelium.[104]

The majority of bacteria detected in the large intestine were from class *Clostridia* (phylum *Firmicutes*), e.g. *Lachnoclostridium*, 14 different *Lachnospiraceae, Ruminococcus*; phylum *Bacteroidetes*, e.g. *Bacteroides, Parabacteroides, Odoribacter, Alistipes, Alloprevotella*; and phylum *Proteobacteria*, e.g. *Desulfovibrio* and *Helicobacter*. These bacteria were either aerotolerant or obligate anaerobes, which is expected as the lumen of the large intestine is characterised by the absence of oxygen. Within the large intestine, some differences were observed between the cecum and colon. For example, *Helicobacter* had a higher relative abundance in the cecum than in the colon in all animals. Studied in the context of infection of the mucosal layer and disease, some *Helicobacter* species were previously observed in the cecal crypts with FISH microscopy.[105]

To find associations between bacteria and metabolites, a correlation score was computed between each metabolite-microbe pair. The relative abundance of bacterial genera was compared to the average intensity of metabolites in the same intestinal segments. Hierarchical clustering identified eight groups of bacteria that had similar correlation profiles (3.26). The largest differences were observed between groups 1-3 and 4-8. The differences between the groups of bacteria could be explained by the spatial origin of highly correlated metabolites (figure 3.27). Namely, bacteria of groups 1-3 predominantly correlated with metabolites localised in the small intestine, and bacteria of groups 4-8 predominantly correlated with metabolites of the large intestine. This was expected from the major changes in the composition between the small intestine and large intestine seen in the bacterial abundance data. Beyond this, the correlation between bacterial abundances and metabolite intensities did not explain the groups on metabolites observed earlier in this work (figure 3.17). This means that while metabolites showed spatial variation in accumulation and depletion along the GI tract, the bacterial abundances did not vary in a similar way.

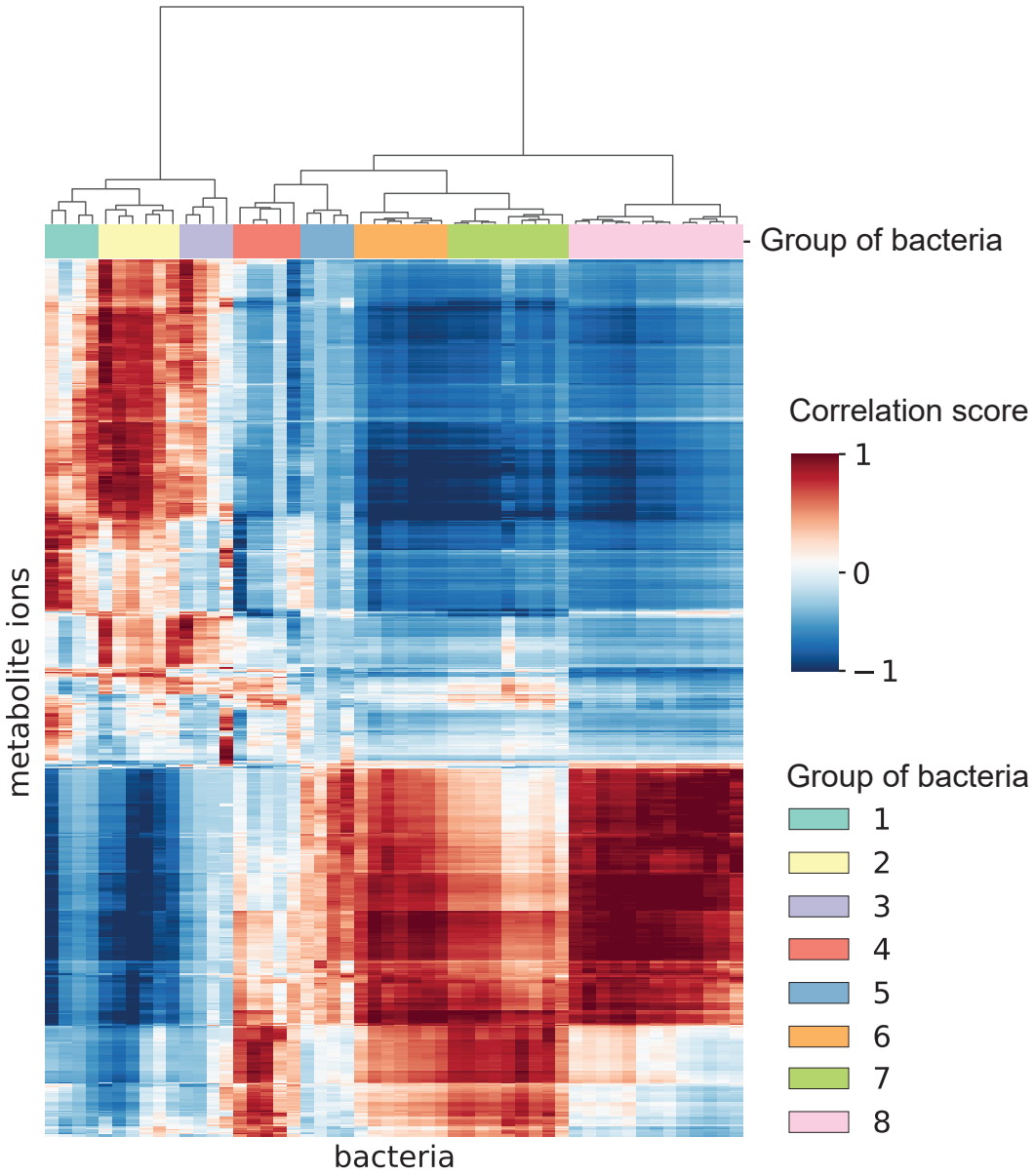


Figure 3.26 Pearson correlation matrix between 52 bacterial genera and 689 metabolite ions in one animal. A score of 1 refers to high correlation between the relative abundance of the microbe and mean metabolite ion intensity in the same area of the tissue, 0 to no correlation, -1 to anticorrelation. Bacteria were clustered into sets by similarity, using hierarchical clustering.

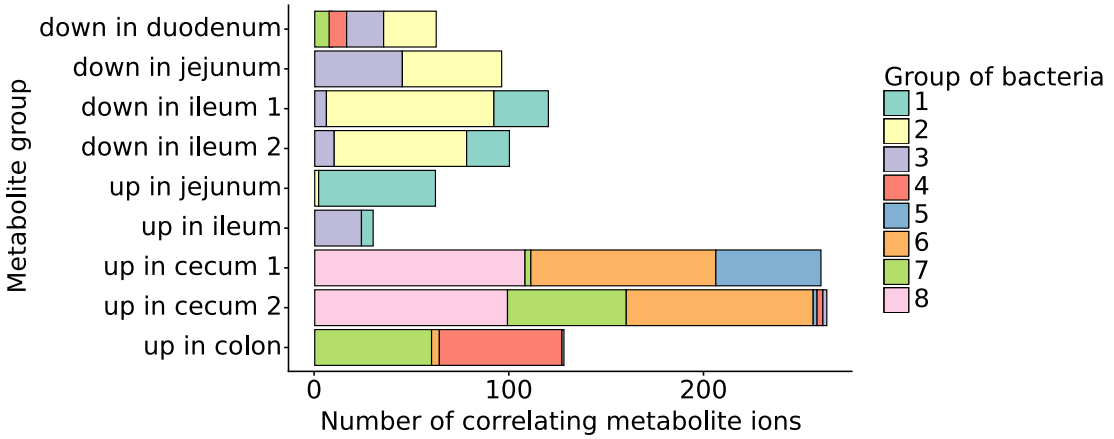


Figure 3.27 Metabolites that are highly correlated with bacterial sets. Metabolite groups that were defined by hierarchical clustering and shown in figure 3.17. A metabolite was considered highly correlated if its Pearson correlation score with any bacterium in the group was above 0.7. Groups of bacteria were defined by hierarchical clustering and shown in the previous figure 3.26.

The initial hypothesis was that each region of distinct metabolic composition (identified by plotting UMAP clusters in spatial coordinates), is inhabited by a distinct bacterial community. This appears not to be the case for the luminal samples that have been sequenced. The predominant difference was found between the bacterial compositions in the small and the large intestine. The possibility of finer spatial differences in the mucosa-associated microbiome, however, should not be discarded. The microbiome living close to mucosa might not be abundant enough for detection in the collected faecal matter samples for bulk sequencing, but it might have nonetheless an important local role in nutrient metabolism and absorption. A suggestion for future studies would be to repeat the bulk-sequencing experiment using a sample preparation protocol compatible with the analysis of the mucosal microbiome.

In case the spatial heterogeneity of both luminal and mucosal sites is low, there is another possible explanation why the spatial organisation of bacteria does not match the spatial variation in metabolite composition. Namely, it is important to consider the transcriptional activity of bacteria. That is, the same bacterium, confronted with a different microbial neighbourhood with its unique competition and cooperation

mechanisms, and different local nutrient concentrations, may have different transcriptional programmes activated. As a result, it cannot be ruled out that the metabolic activity of the same bacterium changes depending on its location. Spatial analysis technologies that provide simultaneous metatranscriptomic readout and distinguish between microbial taxa in high resolution are already emerging.[106, 107] In the near future, such technologies might shed more light on the observed regions of different metabolic composition.

3.2.7 Biotyping of bacteria based of the metabolite fingerprint

Another approach to find bacteria-metabolite associations could be undertaken by MS imaging. MALDI-MS has been used for years for the identification or ”biotyping” of individual microbes.[19, 108] In the biotyping approach, an extract or a smear of a microbial colony is analysed to acquire a mass spectrum of proteins, peptides and lipids, which is then compared to a spectral database containing many microorganisms. Here, the similarity between the spectra is the basis of bacterial identity assignments and is widely used in clinical applications for the detection of pathogens. ”Could a similar approach help identify the bacteria in tissues?” was asked in the course of this work. In comparison to biotyping approaches reported in the literature, the recorded mass spectra obtained for this study are in a lower mass range, thereby only capturing metabolites, oligopeptides (four to five amino acids are usually the largest ones that are detected), and smaller lipids. While such compounds are well-represented in mammalian metabolite and lipid databases, comprehensive databases with bacterial metabolites are still lacking data for many microbes. For example, only 20 bacterial genera commonly found in the mouse gut microbiome were found in the Natural Product Atlas, a database that collects information about microbial metabolites described in scientific literature.[109] For this reason, to help assess the feasibility of a biotyping approach in tissues, an in-house library was prepared.

Bacterial library preparation

To prepare the bacterial metabolite library, the microbes were isolated from mouse intestine. Individually grown bacteria were spotted on glass slides in a regular pattern and subjected to MS imaging. Analysis of the imaging data identified bacteria-specific metabolites. Metabolites that were unique to one or a few bacteria were added to the library.

For the isolation of the microbial strains, performed in collaboration, one male and one female mouse were used. The contents of the small intestine, cecum and large intestine were collected in a microaerobic chamber, to avoid compromised viability of some anaerobic bacteria. The suspension of the collected material was plated on agar, from which individual colonies were picked, sequenced, and expanded in liquid cultures, under suitable microaerobic conditions. In total, 96 colonies were isolated. Sequencing revealed many duplicate strains, and the final collection resulted in 11 different genera: four strains of *Lactobacillus*, three strains of *Bacteroides*, two strains of *Parabacteroides*, and one strain of each: *Lactococcus, Staphylococcus, Streptococcus*, Alistipes, Clostridium, Faecalibaculum*, Parasutterella** and *Pasteurella*. Certain isolates (marked with *) hardly expanded in liquid cultures. As a result, the collected bacteria represented only a partial portion of the most differentially abundant genera identified by the bulk sequencing. Follow-up studies could consider ordering additional bacteria externally, for example from the Mouse Intestinal Bacterial Collection (The Leibniz Institute DSMZ)[110] and include them in the library. Meanwhile, it was decided to proceed with available isolates, to establish the experimental and computational procedure of library preparation.

Two approaches were evaluated for bacterial isolate MS imaging: analysis of extracts and analysis of whole bacterial cells. In the first case, a recently reported single-phase butanol-methanol extraction protocol was used.[111] LC-MS comparison between this method and a more established two-phase Folch's extraction method (performed by EMBL

Metabolomics Core Facility), has shown that the extracted lipid content is very similar in both. Butanol-methanol extraction, in addition to lipids, extracted cellular metabolites. Since no phase separation was required for the collection of the extract, it could be possible to easily scale up the preparation of the bacterial extracts, and perform the extraction in deep-96-well plates using a pipetting robot. In future studies, this procedure could be developed into a high-throughput isolate extraction protocol.

In parallel, a simpler and faster procedure for analysis of whole bacterial cells was developed, as illustrated in figure 3.28. For this, pellets from individual liquid cultures were washed to remove media, which are high in salts and phosphates and interfere with the MS signal. Then, the bacterial suspensions were pipetted into microwells, that were created using a PDMS (polydimethylsiloxane) polymer template mounted onto a regular glass slide (preparation of these microwells is described in the previous chapter, section 2.2.2), and dried. The PDMS template was detached from the glass slide, leaving a regular pattern of bacterial spots on the glass slide. An empty well and a bacterium-free medium spot were included in the sample. The glass slide was then subjected to MS imaging.

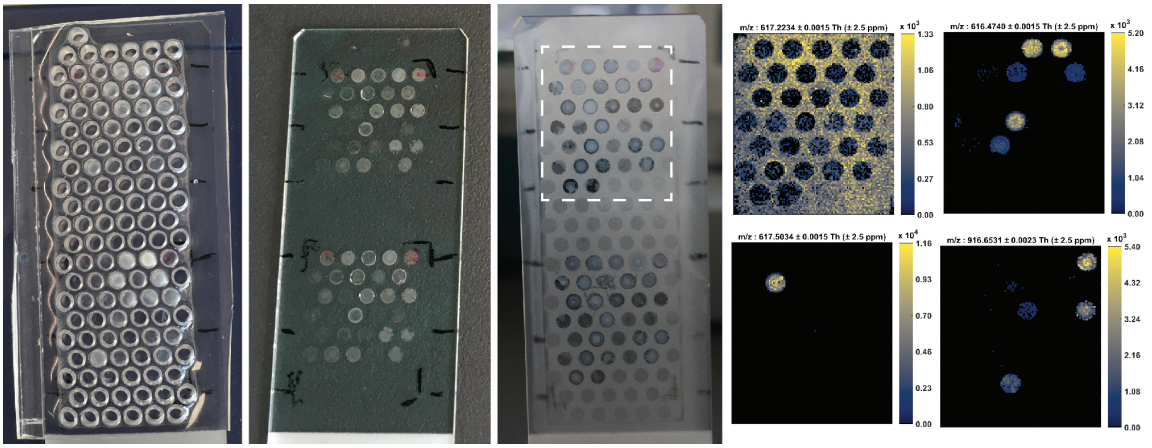


Figure 3.28 Analysis of whole bacterial cells. Left to right: suspension pipetted into microwells, dried bacterial spots after microwell removal, bacterial spots covered with MALDI matrix (dashed line shows one imaged area), four examples of ion images (obtained from the area shown by a dashed line). In the photos, not all the wells are filled with bacterial suspension.

To identify bacteria-specific ions, a similar approach to the data analysis procedure described in the previous chapter (section 2.2.4) was used. Briefly, all pixels belonging to the bacterial spots were identified using an interactive Python tool, as described before. Different in this procedure was the generation of a mean spectrum of pixels belonging to the bacterial spots, after which a peak centroid calculation was performed in MSiReader software. In total, approximately 12,000 peaks were selected in positive ionisation mode, and 6,600 peaks were selected in negative mode. For each peak, an ion image was generated. Following the machine-learning-based procedure, as before, ion images were classified into ”real signal” or ”background noise” images. Finally, calculating the average intensity of ions classified as ”real signal” in every bacterial spot resulted in metabolic fingerprints of individual bacterial strains. Ions that occurred in a blank well or a well containing medium were excluded from the analysis.

In total, the library of bacterial fingerprints included 978 bacteria-specific ions in positive ionisation mode and 657 ions in negative ionisation mode. These ions covered the entire mass range, with a substantial number of both small molecules (< m/z 500) and lipids (> m/z 500), as shown in the figure 3.29. In negative mode, more high-mass ions were detected. Most ions with lower m/z values were detected in the spots of multiple bacterial genera. These ions are likely to be primary metabolites common for all bacteria. A recent study on 833 commercially available, and mainly primary, metabolites has shown that the measurement of the levels of these compounds in bacteria by LC-MS cannot predict bacterial phylogenetic relationships.[112] Bacterial lipids or secondary metabolites of bacteria, however, were not considered by the authors. Among the ions included in bacterial library, higher mass ions, such as lipids (the main diagonal trend in the mass excess plots) and peptides, were mostly unique to one isolate or shared between a small number of isolates. These specific ions could be used as biomarkers and be helpful for the identification of bacteria in tissue.

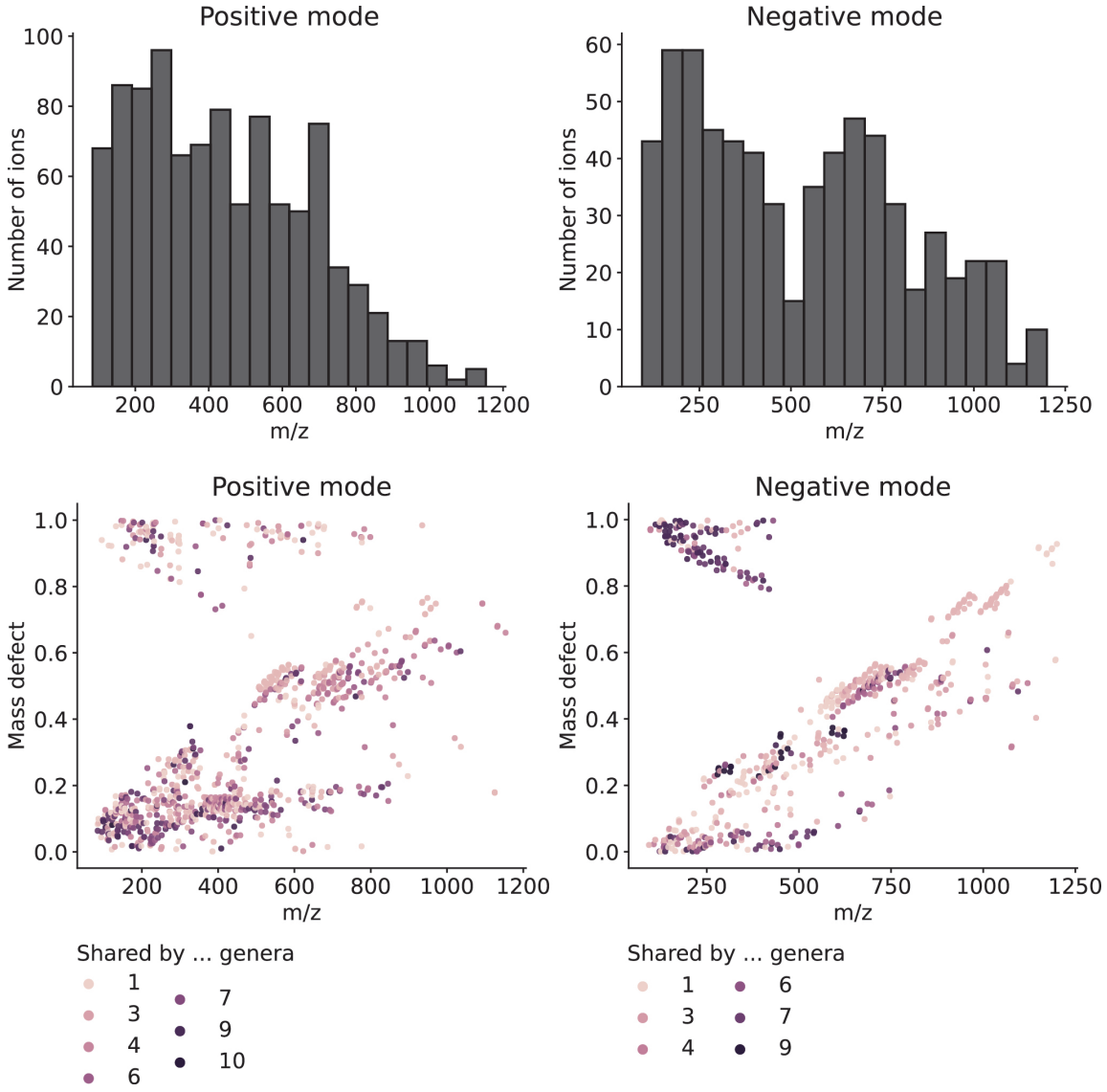
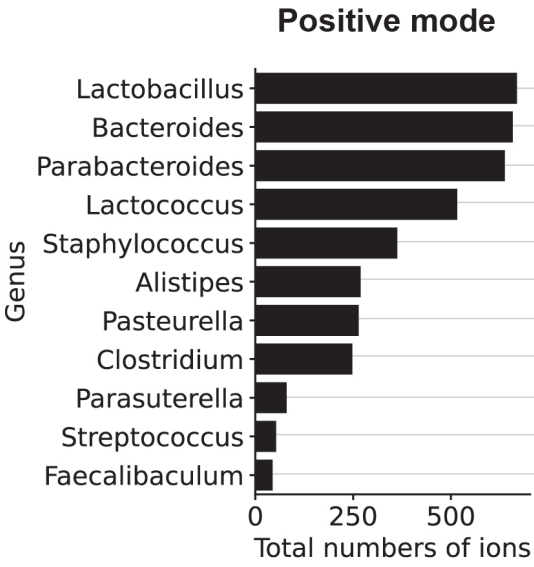


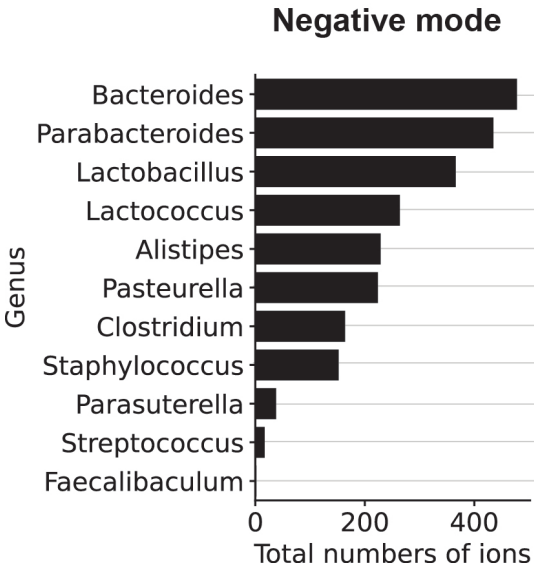
Figure 3.29 Description of identified bacteria-specific ions in positive and negative ionisation modes. Top: histogram of ion m/z values. Bottom: mass excess plot. Each dot in the mass excess plot represents one ion. The dots are coloured by the number of genera in which the ion was detected.

The total number of ions detected per bacterial genus varied with the density of bacterial cells in suspension used for sample preparation, as shown in figure 3.30. Bacterial isolates that expanded well in liquid cultures had the highest density and yielded between 400 and 700 bacteria-specific ions. Bacterial isolates with limited growth in liquid cultures led to the production of fewer than 100 ions. For future studies, it would be recommended to control bacterial density, e.g. by monitoring the optical density (OD600) of the liquid cultures by spectrophotometry

and collecting the same amount of bacterial cells. The samples with a good amount of bacterial matter showed a small number of unique ions per genus. This low number might be due to the fact that some isolates that are phylogenetically related may share specific biochemical pathways and produce the same compounds. This might be the case for example for *Bacteroides* and *Parabacteroides*, or *Lactococcus, Lactobacillus* and *Streptococcus*. This suggestion is supported by the observation that the number of unique ions shared between maximum two genera was considerably higher.



Total ions	Unique ions	Shared with one other genus	Density
669	75	144	Medium
658	19	169	High
638	20	157	High
516	13	52	Medium
363	23	40	Medium
269	18	27	Medium
264	1	3	Medium
248	1	1	Medium
80	5	7	Low
53	0	2	Low
44	0	3	Low



Total ions	Unique ions	Shared with one other genus	Density
478	24	190	High
435	3	155	High
366	29	72	Medium
264	1	21	Medium
229	53	59	Medium
224	5	15	Medium
164	2	3	Medium
152	5	8	Medium
38	1	1	Low
17	11	11	Low
2	0	1	Low

Figure 3.30 Number of bacterial ions detected per genus in positive and negative ionisation modes. Left: Total number of ions detected in each genus. Right: Table with mode information. ”Total ions”: same numbers as shown in the bar chart. ”Unique ions”: the number of ions only found in the genus of interest. ”Shared with one other genus”: the number of ions that are found in the genus of interest and maximum one more genus. ”Density”: the apparent optical density of suspension pipetted into the microwells: low density means that the suspension was transparent, high density means that suspension was opaque white.

Metabolic fingerprints of all isolates, except those with the lowest density, were subjected to hierarchical clustering, as shown in figure 3.31. For this analysis, only the information about the detection of an ion in bacterial isolate spots was used. Generally, bacteria from the same

genus clustered together. Within the class *Bacteroidia, Alistipes* genus had a distinct metabolic fingerprint from *Bacteroides* and *Parabacteroides* genera, which were sharing more features. Within the class *Bacilli*, genus *Lactobacillus* separated from genera *Lactococcus* and *Staphylococcus*. Genera *Clostridium* and *Pasteurella* had less than five unique ions detected per strain (see previous figure). As a result, these genera could not be distinguished from the others based on their metabolic fingerprint. Collectively, these findings suggest that metabolic fingerprints could be used to distinguish microbes at a certain taxonomic level. The exact nature of of this level (species, genus, family or class) remains to be determined in follow-up studies with more isolates.

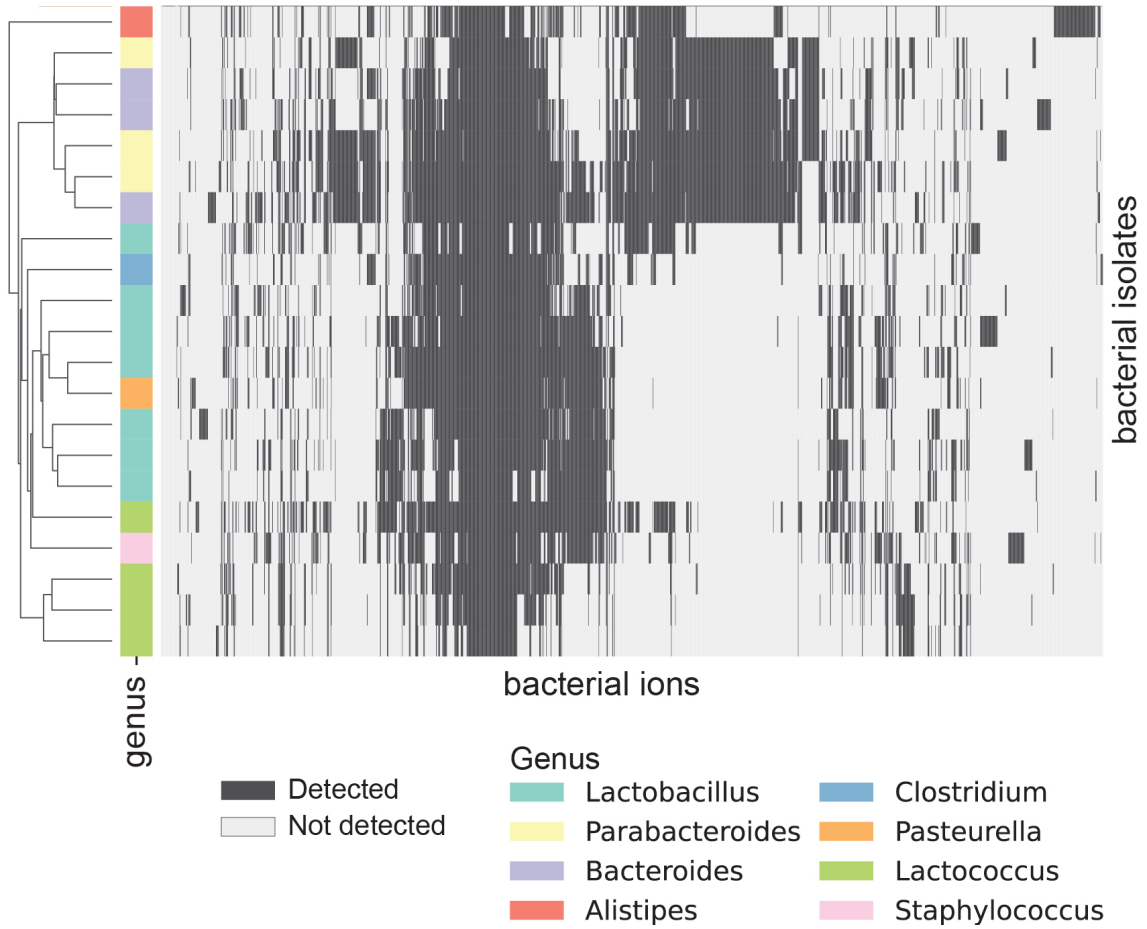


Figure 3.31 Hierarchical clustering of metabolic fingerprints of 21 bacterial isolates. Metabolic fingerprints include 1635 metabolites from both positive and negative ionisation modes. Clustering was performed on binary (detected/not detected) vectors using Jaccard distance. Side colour bar shows the genus of bacterial isolates.

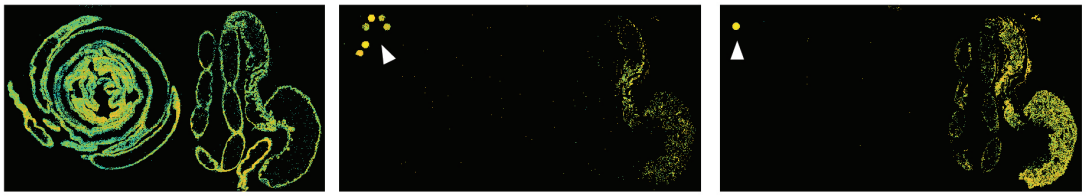
Identification of bacterial metabolites in tissue

Ions from the created bacterial library were matched to the intestinal tissue data sets, using a 2.5 ppm mass tolerance window. Among the ions that were unique for one genus or shared between two genera, 67 had localised distribution within the lumen of the gut (figure 3.32). 21 ions unique to *Alistipes* fell into the m/z range of 550-650 and localised in the cecum. Ions from the larger group that also localised in the cecum were unique to *Bacteroides, Parabacteroides* or (mostly) shared between both genera, consistent with the findings from the hierarchical clustering. These ions covered a broader m/z range of around 440-1040. For the bacterial class *Bacilli*, only one ion, unique to *Staphylococcus*, was found in the terminal ileum of tissue samples. The remaining classes contained few unique metabolites in the library and were not detected in tissues.

Detecting multiple ions in the same region of the tissue, with the m/z values matching to the unique ions of bacterial isolates, increases the probability that the ions in tissue are of bacterial origin. Of course, by only relying on MS1 measurements, there is always the possibility that the two ions with matching m/z values are isomers or isobars of database queries. The use of a mass analyser with high mass resolution as used in this study (i.e. an Orbitrap with resolution 140,000 at m/z 200) helps to maximally resolve the isobars. For a validation beyond the matching of ion m/z values, one could also compare isotope peak distributions of ions of interest between bacterial isolate and tissue data. Furthermore, MS/MS analysis could be performed either using a tissue section by MALDI-MS, or tissue extract — by LC-MS.

Genus	Bacterial ions detected in mouse tissue		Location
	Positive mode	Negative mode	
Alistipes	600.4391, 636.4056, 652.3800	376.2176, 560.4372, 572.4375, 573.4410, 574.4531, 575.4561, 576.4590, 588.4326, 588.4684, 589.4356, 589.4718, 590.4477, 590.4743, 602.4844, 603.4877, 604.4634, 616.4997, 617.4668	Cecum
Bacteroides and Parabacteroides	436.2061, 456.1727, 458.1880, 542.5166, 552.5375, 570.5481, 650.4780, 651.4816, 693.5571, 694.5604, 716.4648, 762.7354, 763.7394, 776.7517, 777.7550, 829.5734	434.1935, 677.4994, 677.5272, 691.5430, 707.5380, 778.5276, 779.5309, 782.5223, 796.5378, 797.5411, 798.5444, 799.5469, 812.5601, 824.5694, 858.6993, 992.7214, 993.7247, 1006.7368, 1007.7402, 1008.7432, 1020.7524, 1021.7561, 1022.7591, 1023.7618, 1034.7683, 1035.7718, 1036.7748	Cecum
Staphylococcus	222.0196	-	Terminal ileum

Examples



Reference | m/z 799.5469 (Bacteroides, Parabacteroides) | m/z 603.4877 (Alistipes)

Figure 3.32 Bacterial ions detected in intestinal tissue sections. Top: table with m/z values. Bottom: examples of ions detected in bacterial spots (shown by white triangle) and localised in intestinal tissue.

Most bacteria are not able to synthesise sphingolipids, lipids containing sphingoid backbones, making this lipid class suitable to serve as potential biomarker. Among the gut bacteria, only members of the phylum *Bacteroidetes*, such as *Alistipes, Bacteroides* and *Parabacteroides*, are known to synthesise these lipids.[113] Sulphonolipids, a subclass of sphingolipids with a sulphate functional group, were recently shown to be produced by only two bacterial genera in the mammalian gut: *Alistipes* and *Odoribacter*.[114] Of the ions that were unique for *Alistipes* isolate and found in the tissue sections, some indeed matched with the sulphonolipid ions described in this work. Figure 3.33 shows the structure of a sulphonolipid sulfobacin B, which was, together with ions containing

additional CH_2 units, detected in the data set. One of these ions, spectrum of which is shown in the same figure, comes from *Alistipes* isolate. The abundance of a characteristic pair of peaks at the M+2 position, one for ^{34}S isotope and the other for an isotope with two ^{13}C assures that the detected ion contains a sulphur atom. Similar spatial distributions of the isotope peaks were observed for both tissue and bacterial isolate data, strengthening confidence that sulphonolipids indeed were detected in tissue.

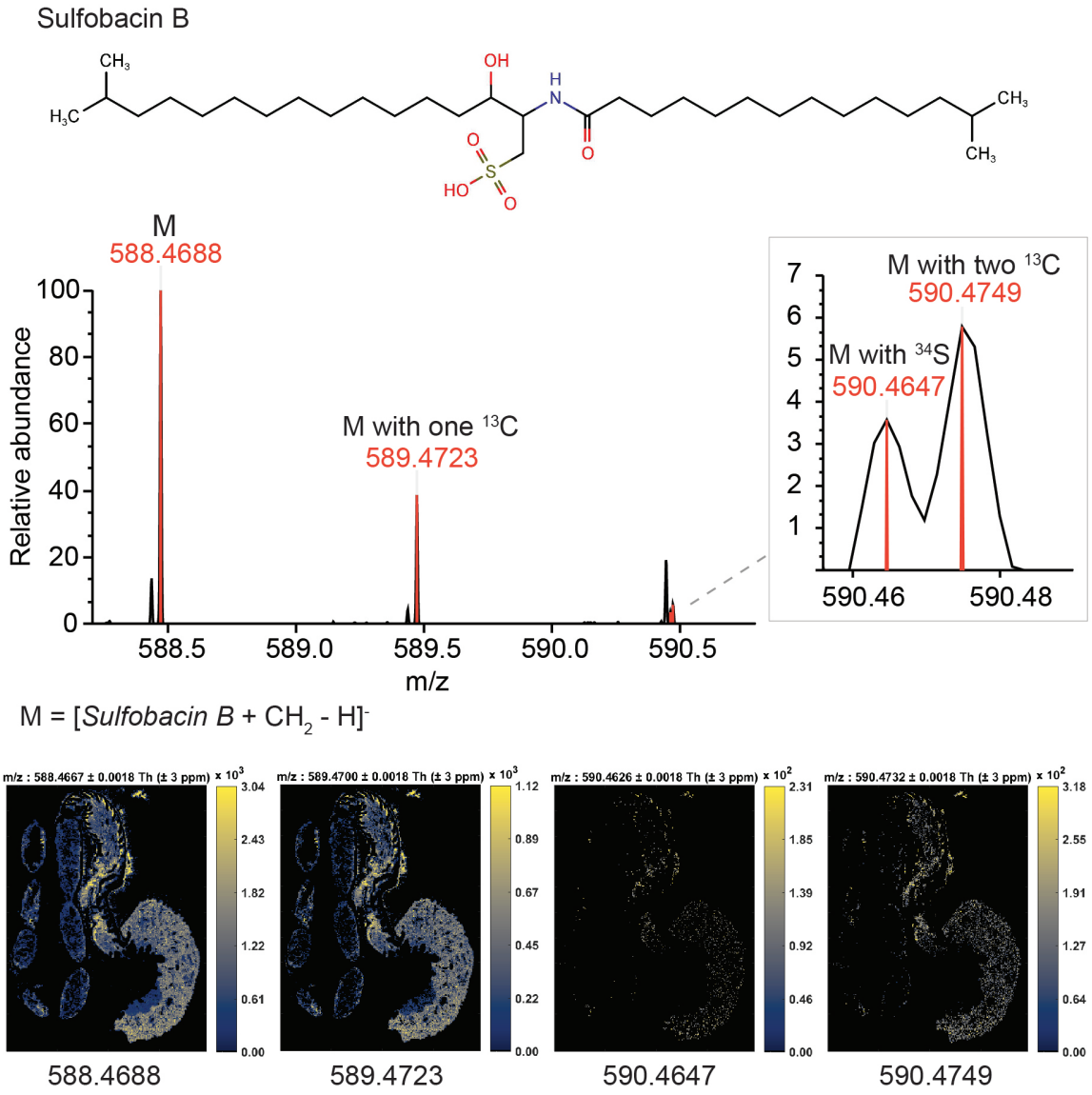
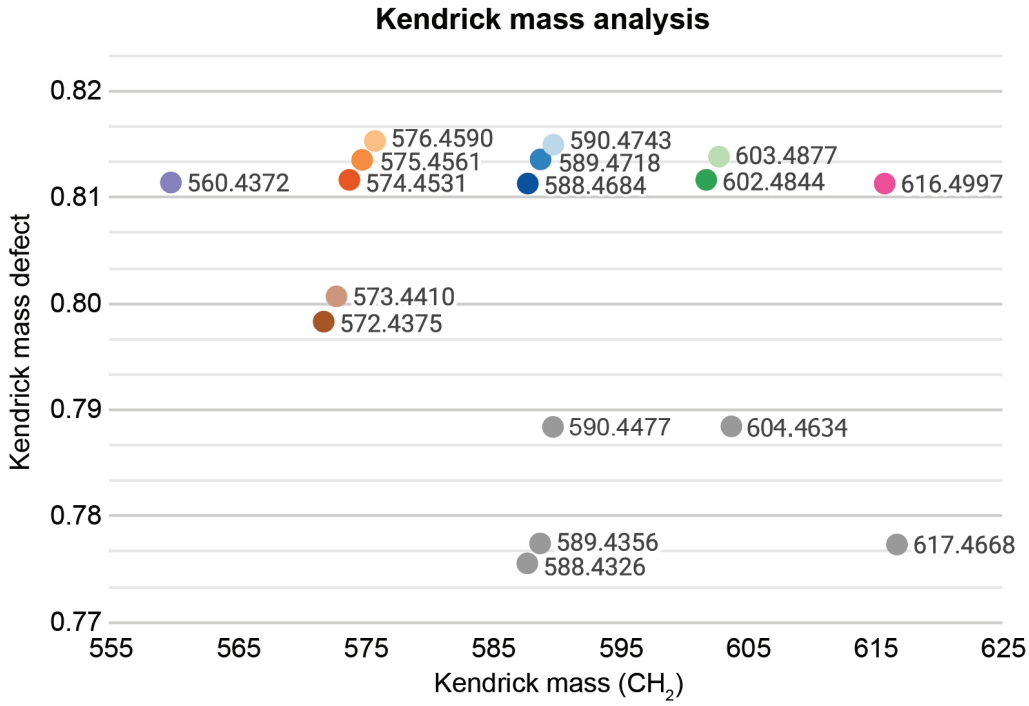


Figure 3.33 Detection of a sulfobacin B derivative (M) in *Alistipes* isolate and in tissue. Top: structure of sulfobacin B, middle: mass spectrum of M from the bacterial spot, bottom: ion images of M and its isotopic peaks in the large intestine.

Based on this finding, annotations could be proposed for other *Alistipes*' ions found in the tissue. To this end, a Kendrick mass defect plot was made (figure 3.34). The Kendrick mass defect plot is similar to the regular mass defect plot, discussed earlier, except that masses of the ions are scaled in such a way that two ions that differ by one CH_2 unit (or another unit if desired) fall on the horizontal line. Here, the Kendrick plot showed that five putative sulphonolipids — only differing from sulfobacin B by $(CH_2)_n$ — were detected and localised in tissue together with some isotope peaks. In addition, a reduced form of sulfobacin B and its first isotope were detected. Theoretical m/z values of proposed annotations differed from detected values on average by 3 ppm, which is within the acceptable range for the mass analyser used. No annotations were proposed for other ions.



m/z	Putative formula	Putative identity	Theoretical mass of the anion	Mass error / ppm
574.4531	$C_{32}H_{65}NO_5S$	Sulfobacin B (B)	574.4511	3.5
575.4561	$^{13}CC_{31}H_{65}NO_5S$	B isotope	575.4543	3.1
576.4590	$^{13}C_2C_{30}H_{65}NO_5S$	B isotope	576.4575	2.6
560.4372	$C_{31}H_{63}NO_5S$	B - CH_2	560.4354	3.2
572.4375	$C_{32}H_{63}NO_5S$	B - 2H	572.4354	3.6
573.4410	$^{13}CC_{31}H_{63}NO_5S$	B - 2H isotope	573.4387	4.0
588.4684	$C_{33}H_{67}NO_5S$	B + CH_2	588.4667	2.9
589.4718	$^{13}CC_{32}H_{67}NO_5S$	B + CH_2 isotope	589.4700	3.1
590.4743	$^{13}C_2C_{31}H_{67}NO_5S$	B + CH_2 isotope	590.4733	1.7
602.4844	$C_{34}H_{69}NO_5S$	B + $2CH_2$	602.4824	3.4
603.4877	$^{13}CC_{33}H_{69}NO_5S$	B + $2CH_2$ isotope	603.4856	3.4
616.4997	$C_{35}H_{71}NO_5S$	B + $3CH_2$	616.4980	2.7

Figure 3.34 Kendrick mass defect plot for ions unique to *Alistipes* isolate that were localised in intestinal tissue sections. Dots that align horizontally in the plot represent ions that differ by one or more CH2 units. Coloured dots have a proposed putative annotation in the table below the plot. Labels next to the dots show the experimental m/z value.

Taken together, these findings show that a biotyping approach is useful to generate hypotheses about bacteria present in the tissue, based on their metabolite and lipid fingerprint. Before bacteria could be reliably identified in the tissue, further validation is required. More extensive bac-

terial libraries are needed to draw conclusions about the typical numbers of unique ions that can be detected in tissue per isolate. A planned follow-up study would prepare more extensive bacterial libraries by obtaining bacteria from public repositories of bacterial strains, such as The Mouse Intestinal Bacterial Collection.[110] Besides adding more microbes to the library, it could be of interest to grow isolates in different media and under different oxygen conditions. This way a broader overview of what metabolites or lipids each bacterium can synthesise is obtained. The subsequent analysis of unique metabolites would determine which taxonomic resolution can be attained by the biotyping approach using the created library. Furthermore, the likelihood of confusion between two isolates would also be evaluated.

Validation approaches

The biotyping approach helped to identify the bacterial metabolites in the gut, which could be interpreted as biomarkers for the localisation of the corresponding microbes. These initial findings, however, need to be validated. This could be done with two approaches: validation of metabolite chemical structure and validation of bacterial localisation.

Validation of the chemical structure of detected metabolites can be easily done by LC-MS/MS. For this approach, the tissue extract has to be prepared. To facilitate performing LC-MS/MS on the tissue of the same animal that was used to prepare samples for MS imaging and bulk sequencing, first, a small evaluation was made. In this evaluation, 1, 3, 5 and 10 tissue sections, similar to the ones used for MS imaging, were homogenised, extracted and analysed (by the EMBL Metabolomics Core facility). This analysis showed that three (12 micrometre-thick) sections of either small or large intestine are sufficient to obtain spectra of good quality. Importantly, the residual embedding medium did not interfere with the extraction or analysis procedures. Thus, during the cryosectioning of intestinal tissue for MS-imaging, a small number of

consecutive tissue sections could be reserved and stored for later LC-MS/MS validation of the findings of interest.

To validate that bacteria are found in a specific location of the gut, FISH imaging can be used. In this approach, tissue sections are labelled with a fluorescent probe designed to bind to a specific region of bacterial ribosomal 16S RNA. The locations of the fluorescent probes are then visualised by a fluorescent microscope. It is not a common practice to combine FISH and MALDI-MS imaging on the same tissue section, however, a recent study achieved this by washing off the MALDI matrix from the samples of a deep-sea mussel and proceeding with a regular FISH protocol.[8] The challenge for repeating this procedure with the intestinal tissue section lies in the high number of permeabilisation and washing steps required for FISH, which may detach the faecal matter of the gut. Indeed, the attempts to fix freshly frozen, thawed and dried tissue sections, as well as tissue sections after MALDI imaging, with paraformaldehyde solution failed, as the contents of the gut detached and dissolved. A solution to this problem was to first dip the glass slide with a tissue section into pure ethanol. Ethanol denatures proteins and makes them sticky, resulting in better attachment to the glass. After the ethanol treatment, which at the same time also washed off the MALDI matrix, the tissue section was firmly attached to the glass slide throughout the whole procedure. Staining of fresh frozen tissue sections and tissue sections already analysed by MALDI-MS by a generic bacterial probe showed that both samples gave a good FISH signal for bacteria (figure 3.35). This observation indicates that MALDI-MS and FISH imaging can be combined on the same tissue section, rather than using two consecutive sections. Although some material is ablated by MALDI laser, the ablated region is as small as 5 micrometres in diameter, when low laser power is used. This means that both metabolomics, and fluorescent probe information can be obtained from virtually the same area of the tissue.

Currently, FISH remains an expensive and time-consuming technique limited to simultaneous analysis of only a few probes, limited by the number of fluorescent channels within a microscope. For this reason, it was not considered as a main technique used to localise the bacteria in the intestinal tissue, but only as a technique used to validate the biotyping approach. The technical limitations of FISH imaging are believed to change in the future, as more complex labelling strategies making use of barcodes and artificial intelligence emerge.[107]

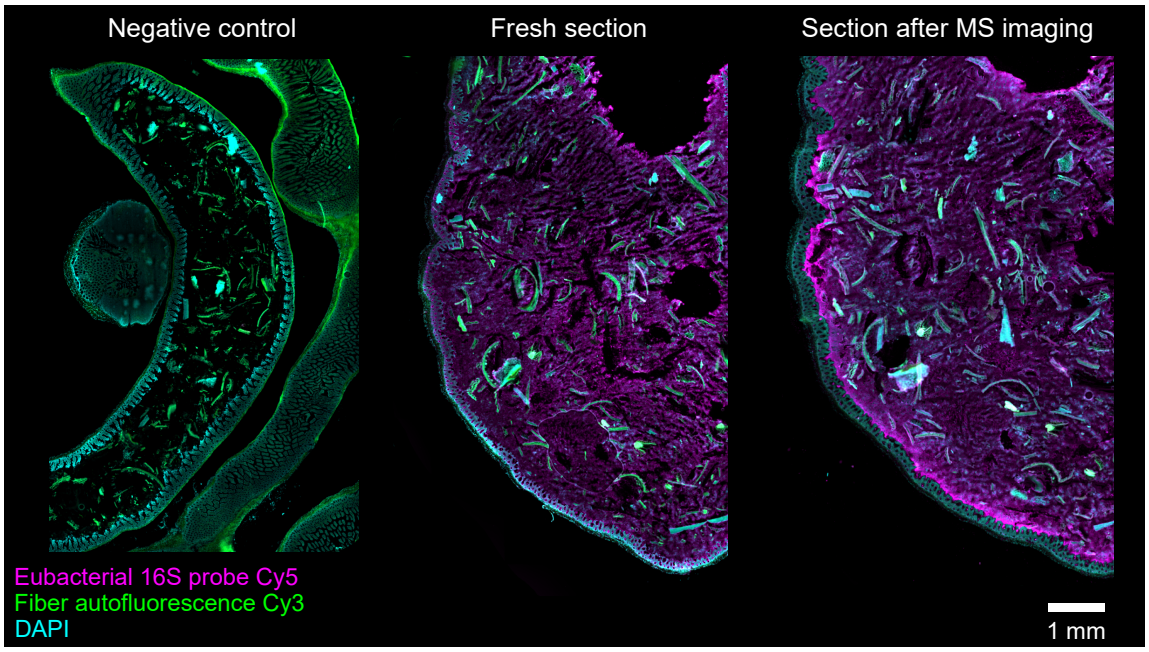


Figure 3.35 FISH imaging of intestinal sections. Left to right: negative control - probe for a gene absent in mouse; fresh frozen section stained with a generic bacterial probe; tissue section after MALDI-MS imaging stained with a generic bacterial probe. The bacterial probe signal was recorded in the Cy5 channel (pseudocoloured with magenta). Additionally, all sections were stained with DAPI to visualise the nuclei of the host cells (pseudocoloured blue). Autofluorescence of host cells and plant matter (fibres) was recorded in the Cy3 channel (pseudocoloured green). All probes used were obtained from a commercial RNAscope kit.

The planned follow-up studies include not only the expansion of the library of bacterial metabolite fingerprints but also the design of FISH probes that would allow the spatial localisation of isolates in tissue (at the genus level) to be validated. For this, the probes that match the ribosomal 16S RNA of target bacteria would be designed in collaboration with a bioinformatician. Specific regions of ribosomal 16S RNA conserved

between all the species of each genus of interest can be determined using a computational pipeline that was previously developed by the collaborators (unpublished work). The probes could then be manufactured by a commercial service (such as RNAscope).

If bacterial identity could be reliably identified by the biotyping approach, information about bacterial and metabolite localisation could be acquired simultaneously, using the same sample and same spatial resolution. This would allow for the identification of all metabolites that co-localise with a bacterium of interest. For example, correlation scores could be calculated for the ion image of sulfobacin B (as a representation of the abundance of *Alistipes* in bacterial tissue) and each of the other ions images generated from tissue section data. This would help identify highly-correlated metabolites associated with *Alistipes* genus better than the bulk sequencing approach, which only provides a very coarse spatial resolution as discussed earlier. Ultimately, the identification of the local bacteria-metabolite associations provides the foundation to understanding the local function of gut microbiome, a yet underexplored area in the field of microbiome research.

3.3 Conclusions

The aim of this work was to characterise spatial differences in metabolism in the mouse intestine and to study the contribution of gut microbiota to these differences. This section will provide a summary of the experimental work, discuss the significance of the obtained results and propose directions for future work.

Summary of the main findings

The first part of this study focused on the metabolism in the gut. With the aid of untargeted MS imaging and machine-learning-based ion image classification, the localisation of approximately 5,500 annotated metabolites in the gut was determined. The mass spectral information was used to identify regions with distinct metabolic compositions in the gut. These regions corresponded to anatomically relevant features, and distinguished between five luminal regions, four mucosal regions, and the muscle wall of the gut. The same regions were identified in four animals. Next, individual contributions of metabolites to the regions observed in the lumen of the gut were determined. The observations from 689 differentially abundant metabolites could be summarised into three main trends. First, some metabolites decreased in abundance along the small intestine. This decrease was driven by the digestion of dietary lipids, absorption of simple sugars, and production of primary bile acids by bacterial metabolism. Second, some metabolites accumulated in the small intestine and were depleted in the cecum. This group included amino acids, produced by protein digestion, and oligosaccharides, which are the breakdown products of larger carbohydrates. Third, some metabolites accumulated in the large intestine. Among these metabolites were lipids with possible bacterial origin, as well as metabolites produced as a result of bacterial metabolic activity (including amines, aromatic amino acid derivatives, secondary bile acids, and nucleobases). Thus, it was shown that both host physiology and bacterial metabolism contribute to the formation of the regions of distinct metabolism.

The second part of this study explored how the spatial differences in the bacterial composition of the gut are related to the identified regions with distinct metabolism. The developed sample preparation procedure allowed measuring relative bacterial abundances in the same tissue that was also used for MS imaging analysis. Bulk sequencing of intestinal segments revealed major differences in microbiome composition between the small and the large intestine. Furthermore, some bacteria localised in only one anatomical region (e.g. ileum, or cecum). Interestingly, the correlation analysis between bacterial abundances and metabolite intensities showed a clear separation between bacteria-metabolite associations in the small and the large intestine. Beyond this, however, patterns in relative bacterial abundance did not match the metabolite patterns. It was concluded that the spatial resolution of such a bulk sequencing approach is not sufficient to address the posed question. In an alternative ”biotyping” approach bacteria in tissue sections were localised based on their metabolite fingerprint. To this end, a library of metabolite fingerprints was constructed from 8 bacterial isolates. Remarkably, unique metabolites only present in the isolates of Alistipes, Bacteroides and Parabacteroides were detected in the cecum. Indeed, this location matched the locations of these bacteria as analysed by bulk sequencing. These proof-of-principle findings suggested that ”biotyping” could be a viable approach to determine bacterial localisations in tissues in high spatial resolution. Finally, a proof-of-principle procedure to validate the findings by LC-MS/MS and FISH imaging was established.

Significance of the obtained results

One of the main accomplishments of this work is the demonstration of how simultaneous analysis of spatial information of metabolites and bacterial composition in the mouse gut could be achieved using MALDI-MS imaging. This study presents the first description of regions with distinct metabolic profiles in the whole faecal-matter filled intestine. The high spatial resolution of MS imaging allow defining the regions that previously were undetectable by bulk metabolomics techniques. For example, a

less than 100-micrometre thick mucosa-associated region was identified in the cecum. The computational procedure used in this study not only defined the spatial boundaries of such regions but also identified the metabolites contributing to their formation. In addition, preliminary findings showed that it is possible to locate certain bacteria in tissue sections based on the metabolic fingerprints of cultured isolates. This procedure could enable the identification of more co-localised bacteria and metabolites in the future.

When more extensive libraries of metabolic fingerprints are available for bacterial localisation in tissues, the scientific approach presented in this work could be used to study host-microbiome interactions *in situ* in the complex setting of the murine gut. On the one hand, this approach would allow simultaneous, label-free localisation of many bacteria in the gut at the high spatial resolution, which would provide an orthogonal readout to what could be obtained by FISH imaging. On the other hand, this approach would also provide an insight into local bacterial function by identifying bacteria-associated metabolites.

Future outlook

The ability to identify co-localised metabolites and the bacteria could be of particular interest in mucosa-associated regions which represent the host-microbiome interface. Current MS imaging technologies allow imaging samples at a spatial resolution below 5 micrometres. Thus, one direction of future work could be the high spatial resolution analysis of mucosa and adjacent faecal matter region in different compartments of the GI tract. For example, if a particular metabolite involved in host-microbiome interactions is of interest, the analysis of biosynthetic genes of bacteria co-localised with this metabolite could lead to a better understanding of site-specific functional roles of individual microbes.

Furthermore, it could be informative to perform metabolite imaging and bacterial localisation through biotyping to compare the two conditions.

For example, wild-type versus germ-free or antibiotic-treated animals could be compared to better characterise microbiome contribution to the spatial heterogeneity observed in the gut metabolome. Despite lacking spatial resolution, such studies have been performed previously with success.[83, 84] Comparison of ”healthy” and ”diseased” conditions would benefit the research using mouse models of diseases for which microbiome changes are known to play an important role. For example, the localisation of bacteria and bacteria-associated metabolites in the tumours of the colorectal cancer model (e.g. APC gene knockout mice [115]) could help to identify the role of individual microbes in tumour development. In cases where disease-associated bacteria are known, but the molecular mechanism that causes the disease is unknown, the identification of metabolites co-localised with bacteria could serve as a lead to such molecular mechanism, or serve as a candidate biomarker for disease diagnosis.

To gain insights on commensals relevant to human health, gnotobiotic animals colonised with synthetic communities of human gut bacteria could be used. The analysis of the tissues of gnotobiotic models similar to that presented in this study could be of particular interest when studying the effects of perturbations (drugs, antibiotics, probiotics, etc.) on the local microbiome and metabolic function of different parts of the intestine. Where the perturbation of the microbiome is unwanted, the results of such a study could inform how the drug should be modified to avoid such side-effects. On the contrary, where the perturbation is desired, the results could help choose the most effective drug candidate in early-stage drug development.

In the broader picture, the scientific approach presented in this work could be extended to gain insights into the regions of distinct metabolite and microbial composition of other organs, such as lungs, and in other model organisms. In addition, such a study can be carried out on clinical samples where knowing both local bacterial composition and their

function can be therapeutically beneficial for certain cancer types, for example in biopsies of colorectal cancer. With continuous developments in imaging ”omics” technologies, the spatial bacteria-metabolite associations identified in this work may be complemented with information from other imaging modalities, such as spatial proteomics, in future.

3.4 Materials and Methods

3.4.1 Animal tissue collection

Animal handling

All experiments using mice were performed using protocols approved by the EMBL Institutional Animal Care and Use Committee. Conventional 9-12 week old C57BL/6 mice (derived from mice from Charles Rivers) were maintained with a 12-hour light/dark cycle in the animal house at EMBL. Three females came from the same litter and were housed in the same cage. Three males came from different litters and were kept in separate cages. All mice were provided with a standard, autoclaved mouse chow (Altromin, Germany) ad libitum. For sample collection, the mice were sacrificed by cervical dislocation.

Collection of intestinal tissue

Tissues were collected from two male and two female mice. The gastrointestinal tract was cut below the stomach, above the entrance to the cecum, and above the rectum to produce two pieces referred to throughout the text as the small intestine and the large intestine. The intestine was untangled, outer fat was removed and the intestine was folded into a compact shape in a 3x3 cm plastic weighing boat. The tissue was covered by warm carboxymethylcellulose (4 wt%, aqueous, stirred overnight, then degassed; Sigma-Aldrich), and any bubbles were removed with a pipette. The weighing boat was submerged for 1.5 min into isopentane (Sigma-Aldrich) pre-cooled on dry ice. Frozen samples were stored in -80°C.

Collection of intestinal content

For intestinal content collection, one male and one female mouse were used. Intestinal tissues from the small intestine, cecum and large intestine were separately collected in 50 mL tubes. After collection, tubes

were immediately cycled into an anaerobic chamber (Coy Laboratory Products, U.S.A.). Contents from each intestinal section were squeezed out from the tissues on sterile Petri dishes using 1 mL pipette tips, transferred into 2 mL tubes (Eppendorf, Germany) and vortexed in 3 mL of PBS buffer.

3.4.2 MALDI-MS metabolite imaging of intestinal tissue

Tissue sample preparation

Tissue samples were removed from the freezer and placed at -20°C for 1 h to equilibrate. Samples were sectioned using the Leica CM3050s cryotome (Leica Biosystems, Germany) with -20°C chamber temperature, -17°C object temperature, 12 µm thickness, and using an anti-roll plate. Part of the sections was thaw-mounted on Superfrost Plus wide microscopy slides (Thermo Scientific, Germany) and briefly air-dried, then dried in the desiccator for 1 hour. Tissue sections were put in sample holders, flushed with argon, vacuum sealed and stored at -80°C until analysis. Another part of the sections was transferred into 2 mL tubes (Eppendorf, Germany) and stored at -80°C until LC-MS/MS analysis. The tissue left over after sectioning was stored at -80°C until bulk sequencing analysis.

Optical microscopy of tissue sections

Samples were removed from -80°C and brought to room temperature for 1 h before opening the vacuum bag to avoid condensation. Tissue sections were imaged with the Nikon Ti-E microscope (Nikon Instruments; objective Plan Apo λ 4x, transmitted light illumination), acquiring multiple tiles in a large field of view. To avoid vignetting effect, flat-field correction was applied by dividing every tile by a pixel-wise median image from the glass area outside the sample scaled to [0, 1] range. Tiles were stitched into large images using the BigStitcher software package.

MALDI-MS imaging of tissue sections

After optical imaging, tissue sections were spray-coated using TM-Sprayer (HTX Technologies, U.S.A). For analysis in positive ionisation mode, samples were coated using 2,5-dihydroxybenzoic acid (DHB) solution (15 mg/mL, acetonitrile:water, 7:3, v:v; Sigma-Aldrich, Germany), for analysis in negative ionisation mode, samples were coated using 1,5-diaminonaphthalene (DAN) solution (10 mg/mL, acetonitrile:water, 7:3, v:v; Sigma-Aldrich, Germany). Water purified with the Milli-Q system (Merck, Germany) was used for solution preparation. The following spray-coating protocols were used. For DHB: temperature 75°C, flow rate 0.07 mL/min, gas pressure 10 psi, track spacing 2 mm, velocity 1350 mm/min, number of passes 8, drying time between passes 15 s. For DAN: same as DHB except track spacing was 2 mm. After matrix spray-coating, samples were imaged on an AP-SMALDI5 ion source (TransMIT, Giessen, Germany) coupled to a Q Exactive Plus mass spectrometer (Thermo Fisher Scientific, Bremen, Germany). Instrument parameters were the same for both matrices: attenuator angle 26°, pitch (75 µm for higher-resolution data sets and 150 µm for lower-resolution data sets), m/z ranges 80-320 and 300-1200, voltage (pos/neg) +3 kV / -3.1 kV, RF level (50 for low mass range, 100 for high mass range), injection time 500 ms, mass resolving power (nominal) 140,000. Each sample was imaged at both indicated low and high mass ranges. Between the two runs, the imaging raster was offset manually by 30 µm to avoid sampling the same locations more than once. After the acquisition run was finished, the samples were vacuum-sealed in an argon atmosphere and stored at -80°C until FISH imaging.

Analysis of MALDI-MS imaging data of tissue sections

The raw data was converted from the vendor proprietary data format to the open mzML format using the MSConvert[74] software, using the vendor-supplied method for spectrum centroiding. From mzML format, the data was converted to imzML using imzMLConverter[75]

software. Low mass range and high mass range data sets acquired from the same tissue section were merged into one data set with a Python script utilising the pyimzML Python library. The data sets were annotated using METASPACE online annotation platform, using databases HMDB (v.4), LIPID MAPS (v.2017-12-12) and a custom database created with metabolites the Genome-scale metabolic model of Mus Musculus (iMM1415, [91]). The annotated images were subjected to classification using a machine learning pipeline developed and described in detail in the previous chapter (section 2.2.4). Briefly, for every data set a binary mask was drawn in the Fiji imaging processing package, where all pixels overlapping with tissue were labelled as 1 and all the other pixels as 0. Two metrics were calculated for each ion image: a ratio of average ion intensities between the two regions, and a ratio of occupancy (i.e. a fraction of non-zero pixels) between the two regions. A subset of 262 ion images randomly drawn from all acquired mouse intestine data sets was manually labelled as ”real signal” or ”background noise”. CatBoost[70] machine learning model was trained on the labelled set of images and its performance was evaluated. Then, the model was applied to classify all remaining annotated ion images in all data sets. Analysis of ion images classified as ”real signal” was completed in Python, besides generic libraries using: PyKrev (for Van Krevelen plots), Scanpy (for UMAP analysis, Leiden clustering, differential abundance analysis), Affinder (a plugin to Napari Python image viewer [94], for overlaying clusters over optical images) and SciPy (for hierarchical clustering). Fiji imaging processing package was used to define intestinal segments over which metabolite intensities were averaged for analysis by hierarchical clustering.

3.4.3 Bulk sequencing of intestinal segments

DNA extraction of bacterial contents from mouse intestinal segments

Frozen intestinal tissues were thawed at room temperature. Intestines from each animal were cut into 1-2 cm segments. Contents from each intestinal segment were scraped out from tissues into 2 mL tubes (Eppendorf, Germany) used for bead-beating provided in the DNeasy PowerSoil Pro Kit (QIAGEN, Germany). DNA extraction was performed according to the manufacturer's protocol.

16S amplicon sequencing of bacterial communities from mouse intestinal segments

16S rRNA sequences were first amplified with the primer pair forward:

5'-CTCTTTCCCTACACGACGCTCTTCCGATCTGTGCCAGCMGCCGCGGTAA-3' and reverse:

5'-CTGGAGTTCAGACGTGTGCTCTTCCGATCTGGACTACHVGGGTWTCTAAT-3' at 0.25 µM per reaction in Phire Hot Start II PCR Master Mix (Thermo Scientific, Germany). 1 µL of DNA template was used in each 10 µL reaction. PCR conditions were: Initial denaturation 98°C 30 s, 15 cycles: denaturation 98°C 10 s, annealing 57°C 30 s, extension 72°C 20 s, final extension 72°C 10 min. 1µL of PCR reaction produces from the first PCR was transferred to the second 10 µL PCR reaction mix containing barcoded primer mix (IDT) and amplified with the PCR condition: initial denaturation 98°C 30 s, 15 cycles: denaturation 98°C 10 s, annealing 65°C 30 s, extension 72°C 20 s, final extension 65 °C 2 min. Reaction products from the second PCR reaction were run on a 2% agarose gel to see if a single clear band around 430 bp could be visualised. 2 µL of each sample were pooled together and submitted for 250 bp paired-end sequencing on MiSeq System (Illumina, U.S.A.) in the EMBL Genomics Core Facility.

Analysis of bulk sequencing data

After sequencing, the raw sequence reads were demultiplexed based on their unique primer barcode identifiers using MiSeq instrument software and prepared for further downstream processing and analysis by the EMBL Genomics Core Facility. The demultiplexed sequence reads were processed with DADA2 (Divisive Amplicon Denoising Algorithm), a tool able to identify the original sequences that generated the observed 16S sequences, known as Amplicon Sequence Variants (ASVs).[116] Using DADA2, raw sequence reads were quality filtered and denoised, after which paired-end reads were merged, and chimeric sequences discarded. A total of 7,957,964 reads were generated after Miseq paired-end sequencing, and 4,644,301 reads remained after filtering, denoising and chimeric sequences removal. The reads were denoised by DADA2 resulting in 5,558 ASVs, which have been taxonomically annotated using the DADA2 implementation of the RDP classifier[117] and the Silva database nr. 99 version 138 [118]. The ASVs that map to the same genus were summed together to create a taxonomic profile composed of 101 unique genera. To calculate the relative abundance, the count of each genus of each intestinal segment was divided by the total read count of that intestinal segment.

3.4.4 Construction of bacterial library for biotyping

Isolation and culturing of bacteria

After collection of intestinal contents, supernatants were diluted 10-fold serially before 50 µL of each dilution was plated on MGAM and BHI blood plates (made with a commercial recipe), respectively. Intestinal contents from the small intestine were incubated in a microaerobic chamber (Coy Laboratory Products, U.S.A.) containing 5% oxygen, while the contents from the distal intestine were cultured under anaerobic conditions, in order to mimic the oxygen conditions in vivo. After 24 h of incubation of plates containing different supernatant dilutions at 37°C, 48 colonies with as much diversity as possible from each animal's intestinal section

were picked into MGAM or BHI-S (made following ATCC's recipe) liquid medium for expansion. After reaching OD600 of 1, 20% glycerol stocks were made for all isolates and maintained at -80°C.

Bacterial isolate 16S identification

2 µL of isolate culture from 24-hour growth were lysed in 20 µL hot shot lysis buffer (25 mM NaOH aq., 0.2 mM EDTA aq.) by heating at 95°C for 30 min. 4 µL lysates were neutralised with 4 µL 40 mM Tris-HCl buffer in 36 µL EB buffer (QIAGEN, Germany). 16S rRNA sequences were amplified with primer pair 27f (5'-AGAGTTTGATCCTGGCTCAG-3') and 1492r (5'-ACGGCTACCTTGTTACGACTT-3') using DreamTaq Green PCR Master Mix (ThermoFisher, Germany) following the commercial protocol. PCR products were cleaned up using the QIAquick PCR Purification Kit (QIAGEN, Germany) and sent to Eurofins for Sanger sequencing. Sequences were BLASTed against "16S ribosomal RNA sequences" database using the NCBI nucleotide BLAST tool for isolate identification.

MALDI-MS imaging of bacterial isolates

Using the same procedure as described in the previous chapter, microwell slides were prepared using regular glass slides and PDMS polymer templates. Polymer templates were cast using polyoxymethylene moulds produced by the EMBL Mechanical Workshop. Silicone casting and curing was done under cleanroom conditions. Sylgard 184 PDMS resin (Biesterfeld, Hamburg, Germany) was mixed with a curing agent according to the manufacturer's instructions. For each casting, 4 g of the curing resin was poured into the mould immediately after mixing. After pouring, moulds were first degassed in a vacuum desiccator for 30 min and then cured for 2 h at 65°C. After curing, the polymer templates were removed from the mould and attached to glass slides to create microwells. 2 mL aliquots of bacterial isolate liquid cultures were washed with water thrice to remove media, at pelleted. 10 µL of water was added to each pellet to create a suspension. 1 µL of each bacterial cell suspension was pipetted

into microwells. The sample was air-dried for 30 min, after which the polymer template was removed. Then, the samples were spray-coated with MALDI matrix and MALDI-MS imaging was performed with the same instruments, protocols and parameters as used for tissue section imaging (described in section 3.4.2).

Analysis of MALDI-MS imaging data of bacterial isolates

The raw data were converted to imzML format (here without centroiding) and the mass ranges were merged as was described for tissue section MALDI-MS data (section 3.4.2). The locations of bacterial spots were determined using the interactive grid fitting tool in Python developed and described in detail in the previous chapter (section 2.2.4). The region of interest was created to include all bacterial spots and exclude spots with blank and media controls. The mean spectrum was generated for the region of interest in MSiReader software, and peak picking was performed using the ”Parabolic centroid” algorithm (mean intensity treshold=200). An ion image for every picked peak was generated and saved. The resulting ion images were subjected to machine-learning-based classification into the ”real signal” and ”background noise” images, in the same way, that was described for ion image classification of metabolite spots in the previous chapter (section 2.2.4), including the same metrics. The ion images that passed the classification contained only ions localised in bacterial spots (bacteria-specific ions). Bacterial identity was assigned to each spot, and the intensity of each bacteria-specific ion in each spot was measured. This information was further used for hierarchical clustering analysis, done using the SciPy Python library. For each m/z value of identified bacteria-specific ions, an ion image of each tissue data set was generated with 2.5 ppm tolerance, using MSiReader. For convenience, ion images generated for each m/z value were stitched into one multi-panel image showing the small and large intestines of four animals, and the array of bacterial isolate spots side-by-side, in Python. The panel images were manually inspected to identify bacterial ions consistently localised in the intestinal tissue of all animals.

3.4.5 LC-MS/MS and FISH imaging validation protocols

Tissue extract preparation for LC-MS analysis

Intestinal tissue sections were thawed on ice for 2 h and transferred to 5 mL tubes (Eppendorf, Germany). For lipidomics, analyte extraction was initiated by adding 1.5 mL 2-propanol to each sample, whereas for metabolomics 1.0 mL methanol (both LC-MS grade; Fisher Scientific, Germany) was used. Subsequently, tissues were homogenised for 5 min with a battery-operated pestle mixer. After incubation for 10 min at -80°C, the homogenates were vortexed for 30 s and placed in an iced ultrasonication bath for 5 min. Samples were then centrifuged for 15 min at 15000g and 4°C (5415R microcentrifuge; Eppendorf, Germany) and supernatants were transferred to fresh 5 mL tubes. Prior to analysis, samples were evaporated to dryness under a nitrogen stream and reconstituted with 100 µL 2-propanol:methanol (50:50, v:v) for lipidomics and 100 µL 85% acetonitrile (LC-MS grade; Fisher Scientific, Germany) for metabolomics.

FISH imaging

Samples were removed from -80°C and brought to room temperature for 1 h before opening the vacuum bag to avoid condensation. Both new tissue sections and tissue sections that were already analysed by MALDI-MS imaging were used. Each sample slide was dipped into ethanol for 10 s to fix the tissue to the glass slide and wash away the matrix, if present. Then, samples were fixed in PFA (4% v:v in PBS) for 1h at room temperature. The following steps were performed according to the manufacturer's protocol of the HiPlex RNAscope FISH staining kit (RNAscope, Abingdon, U.K.). Briefly, samples were dehydrated for 5 min with ethanol, then permeabilised with protease treatment for 30 min. The negative control probe provided in the kit was hybridised to the control sample. Probe EB-16S-rRNA was hybridised to the remaining samples, and amplified using provided hybridisation system. This probe

targets conservative regions in Eubacteria 16SrRNA. After hybridisation, samples were stained with DAPI and mounted with ProLong Gold Antifade. Fluorescent images were obtained using the Nikon Ti-E microscope (Nikon Instruments; objective Plan Flour 10x, fluorescence channels DAPI, Cy3, and Cy5).

3.5 Contribution statement

All work described in this chapter was performed by the author, except for the following contributions from collaborators. This work was supervised by Dr. Michael Zimmermann.

- George Maftei was responsible for the breeding and maintaining of the mice.
- George Maftei and Michael Zimmermann dissected the animals and prepared the tissue for embedding.
- Eleonora Mastrorilli prepared a custom database created with metabolites of the Genome-scale metabolic model of Mus Musculus (iMM1415).
- Boyao Zhang isolated bacteria from the mouse intestine, performed isolate sequencing, and carried out bulk tissue sequencing.
- The EMBL Mechanical workshop (Arthur Milberger, Tim Hettinger, Leo Burger) manufactured the mould for PDMS casting.
- Alessio Milanese and Boyao Zhang analysed the sequencing data.
- The EMBL Metabolomics Core Facility (Prasad Phapale, Bernhard Drotleff) extracted and analysed tissue samples by LC-MS.
- Mohammed Shahraz took part in sample preparation for FISH imaging.
- Alessio Milanese established a computational pipeline for the design of 16s rRNA probes for FISH imaging.

Bibliography

[1] Fredrik Bäckhed et al. "The next decade of metabolism". en. In: *Nat Metab* 1.1 (Jan. 2019), pp. 2–4.

[2] Kathryn E Wellen and Nathaniel W Snyder. "Should we consider subcellular compartmentalization of metabolites, and if so, how do we measure them?" en. In: *Curr. Opin. Clin. Nutr. Metab. Care* 22.5 (Sept. 2019), pp. 347–354.

[3] Antonella Chiapparino et al. "The orchestra of lipid-transfer proteins at the crossroads between metabolism and signaling". en. In: *Prog. Lipid Res.* 61 (Jan. 2016), pp. 30–39.

[4] Ruben Boon. "Metabolic Fuel for Epigenetic: Nuclear Production Meets Local Consumption". en. In: *Front. Genet.* 12 (Nov. 2021), p. 768996.

[5] Maria J Rodrıǵuez-Colman et al. "Interplay between metabolic identities in the intestinal crypt supports stem cell function". en. In: *Nature* 543.7645 (Mar. 2017), pp. 424–427.

[6] Chih-Hao Chang et al. "Metabolic Competition in the Tumor Microenvironment Is a Driver of Cancer Progression". en. In: *Cell* 162.6 (Sept. 2015), pp. 1229–1241.

[7] Bruno Hudry et al. "Sex Differences in Intestinal Carbohydrate Metabolism Promote Food Intake and Sperm Maturation". en. In: *Cell* 178.4 (Aug. 2019), 901–918.e16.

[8] Benedikt Geier et al. "Spatial metabolomics of in situ host-microbe interactions at the micrometre scale". en. In: *Nat Microbiol* 5.3 (Mar. 2020), pp. 498–510.

[9] Gary J Patti, Oscar Yanes, and Gary Siuzdak. "Innovation: Metabolomics: the apogee of the omics trilogy". en. In: *Nat. Rev. Mol. Cell Biol.* 13.4 (Mar. 2012), pp. 263–269.

[10] Michael Eisenstein. "Seven technologies to watch in 2022". en. In: *Nature* 601.7894 (Jan. 2022), pp. 658–661.

[11] Abhishek K Jha et al. "Network integration of parallel metabolic and transcriptional data reveals metabolic modules that regulate macrophage polarization". en. In: *Immunity* 42.3 (Mar. 2015), pp. 419–430.

[12] Michael Karas et al. "Matrix-assisted ultraviolet laser desorption of non-volatile compounds". In: *Int. J. Mass Spectrom. Ion Process.* 78 (Sept. 1987), pp. 53–68.

[13] Koichi Tanaka et al. "Protein and polymer analyses up tom/z 100 000 by laser ionization time-of-flight mass spectrometry". en. In: *Rapid Commun. Mass Spectrom.* 2.8 (Aug. 1988), pp. 151–153.

[14] Masamichi Yamashita and John B Fenn. "Electrospray ion source. Another variation on the free-jet theme". In: *J. Phys. Chem.* 88.20 (Sept. 1984), pp. 4451–4459.

[15] Richard M Caprioli, T B Farmer, and J Gile. "Molecular imaging of biological samples: localization of peptides and proteins using MALDI-TOF MS". en. In: *Anal. Chem.* 69.23 (Dec. 1997), pp. 4751–4760.

[16] Theodore Alexandrov. "Spatial Metabolomics and Imaging Mass Spectrometry in the Age of Artificial Intelligence". en. In: *Annu Rev Biomed Data Sci* 3 (July 2020), pp. 61–87.

[17] Rémi Longuespée et al. "MALDI mass spectrometry imaging: A cutting-edge tool for fundamental and clinical histopathology". en. In: *Proteomics Clin. Appl.* 10.7 (July 2016), pp. 701–719.

[18] John G Swales et al. "Mass spectrometry imaging and its application in pharmaceutical research and development: A concise review". In: *Int. J. Mass Spectrom.* 437 (Mar. 2019), pp. 99–112.

[19] Piseth Seng et al. “Ongoing revolution in bacteriology: routine identification of bacteria by matrix-assisted laser desorption ionization time-of-flight mass spectrometry”. en. In: *Clin. Infect. Dis.* 49.4 (Aug. 2009), pp. 543–551.

[20] Heather Hulme et al. “Microbiome-derived carnitine mimics as previously unknown mediators of gut-brain axis communication”. en. In: *Sci Adv* 6.11 (Mar. 2020), eaax6328.

[21] Dušan Veličković and Christopher R Anderton. “Mass spectrometry imaging: Towards mapping the elemental and molecular composition of the rhizosphere”. In: *Rhizosphere* 3 (June 2017), pp. 254–258.

[22] Luca Rappez et al. “SpaceM reveals metabolic states of single cells”. en. In: *Nat. Methods* 18.7 (July 2021), pp. 799–805.

[23] John A McLean, Whitney B Ridenour, and Richard M Caprioli. “Profiling and imaging of tissues by imaging ion mobility-mass spectrometry”. en. In: *J. Mass Spectrom.* 42.8 (Aug. 2007), pp. 1099–1105.

[24] Jens Soltwisch et al. “Mass spectrometry imaging with laser-induced postionization”. en. In: *Science* 348.6231 (Apr. 2015), pp. 211–215.

[25] Zoltán Takáts et al. “Mass spectrometry sampling under ambient conditions with desorption electrospray ionization”. en. In: *Science* 306.5695 (Oct. 2004), pp. 471–473.

[26] Jason S Sampson, Adam M Hawkridge, and David C Muddiman. “Generation and detection of multiply-charged peptides and proteins by matrix-assisted laser desorption electrospray ionization (MALDESI) Fourier transform ion cyclotron resonance mass spectrometry”. en. In: *J. Am. Soc. Mass Spectrom.* 17.12 (Dec. 2006), pp. 1712–1716.

[27] Peter Nemes and Akos Vertes. “Laser ablation electrospray ionization for atmospheric pressure, in vivo, and imaging mass spectrometry”. en. In: *Anal. Chem.* 79.21 (Nov. 2007), pp. 8098–8106.

[28] T L Colliver et al. “Atomic and molecular imaging at the single-cell level with TOF-SIMS”. en. In: *Anal. Chem.* 69.13 (July 1997), pp. 2225–2231.

[29] J Sabine Becker, Andreas Matusch, and Bei Wu. “Bioimaging mass spectrometry of trace elements - recent advance and applications of LA-ICP-MS: A review”. en. In: *Anal. Chim. Acta* 835 (July 2014), pp. 1–18.

[30] Charlotte Giesen et al. “Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry”. en. In: *Nat. Methods* 11.4 (Apr. 2014), pp. 417–422.

[31] Wei-Wen Chen et al. “Spectroscopic coherent Raman imaging of Caenorhabditis elegans reveals lipid particle diversity”. en. In: *Nat. Chem. Biol.* 16.10 (Oct. 2020), pp. 1087–1095.

[32] Qiuliyang Yu et al. “A biosensor for measuring NAD+ levels at the point of care”. en. In: *Nat Metab* 1.12 (Dec. 2019), pp. 1219–1225.

[33] Cosima Damiana Calvano et al. “MALDI matrices for low molecular weight compounds: an endless story?” en. In: *Anal. Bioanal. Chem.* 410.17 (July 2018), pp. 4015–4038.

[34] David A Pirman et al. “Changes in cancer cell metabolism revealed by direct sample analysis with MALDI mass spectrometry”. en. In: *PLoS One* 8.4 (Apr. 2013), e61379.

[35] Haitao Lv. “Mass spectrometry-based metabolomics towards understanding of gene functions with a diversity of biological contexts”. en. In: *Mass Spectrom. Rev.* 32.2 (Mar. 2013), pp. 118–128.

[36] Nathaniel G Mahieu and Gary J Patti. “Systems-Level Annotation of a Metabolomics Data Set Reduces 25 000 Features to Fewer than 1000 Unique Metabolites”. en. In: *Anal. Chem.* 89.19 (Oct. 2017), pp. 10397–10406.

[37] Jenny Leopold et al. “Recent Developments of Useful MALDI Matrices for the Mass Spectrometric Characterization of Lipids”. en. In: *Biomolecules* 8.4 (Dec. 2018).

[38] Qiuqin Zhou, Annabelle Fülöp, and Carsten Hopf. "Recent developments of novel matrices and on-tissue chemical derivatization reagents for MALDI-MSI". en. In: *Anal. Bioanal. Chem.* 413.10 (Apr. 2021), pp. 2599–2617.

[39] Zhi Qiao and Franziska Lissel. "MALDI Matrices for the Analysis of Low Molecular Weight Compounds: Rational Design, Challenges and Perspectives". en. In: *Chem. Asian J.* 16.8 (Apr. 2021), pp. 868–878.

[40] Maxim Dashtiev et al. "Positive and negative analyte ion yield in matrix-assisted laser desorption/ionization". In: *Int. J. Mass Spectrom.* 268.2 (Dec. 2007), pp. 122–130.

[41] William J Perry et al. "Uncovering matrix effects on lipid analyses in MALDI imaging mass spectrometry experiments". en. In: *J. Mass Spectrom.* 55.4 (Apr. 2020), e4491.

[42] Josiah C McMillen et al. "Effect of MALDI matrices on lipid analyses of biological tissues using MALDI-2 postionization mass spectrometry". en. In: *J. Mass Spectrom.* 55.12 (Dec. 2020), e4663.

[43] Yamil Simón-Manso et al. "Metabolite profiling of a NIST Standard Reference Material for human plasma (SRM 1950): GC-MS, LC-MS, NMR, and clinical laboratory analyses, libraries, and web-based resources". en. In: *Anal. Chem.* 85.24 (Dec. 2013), pp. 11725–11731.

[44] Yannick Djoumbou Feunang et al. "ClassyFire: automated chemical classification with a comprehensive, computable taxonomy". en. In: *J. Cheminform.* 8 (Nov. 2016), p. 61.

[45] Saeideh Khodaee et al. "iMM1865: A New Reconstruction of Mouse Genome-Scale Metabolic Model". en. In: *Sci. Rep.* 10.1 (Apr. 2020), p. 6177.

[46] Zachary A King et al. "BiGG Models: A platform for integrating, standardizing and sharing genome-scale models". en. In: *Nucleic Acids Res.* 44.D1 (Jan. 2016), pp. D515–22.

[47] David Rogers and Mathew Hahn. "Extended-connectivity fingerprints". en. In: *J. Chem. Inf. Model.* 50.5 (May 2010), pp. 742–754.

[48] Gerhard Michal. *Roche Biochemical Pathways, 4th Edition*. Accessed: 2020-12-01. URL: http://biochemical-pathways.com.

[49] Youssef Darzi et al. "iPath3.0: interactive pathways explorer v3". en. In: *Nucleic Acids Res.* 46.W1 (July 2018), W510–W513.

[50] Bryson D Bennett et al. "Absolute metabolite concentrations and implied enzyme active site occupancy in Escherichia coli". en. In: *Nat. Chem. Biol.* 5.8 (Aug. 2009), pp. 593–599.

[51] Virginia De Cesare et al. "High-throughput matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry-based deubiquitylating enzyme assay for drug discovery". en. In: *Nat. Protoc.* 15.12 (Dec. 2020), pp. 4034–4057.

[52] Beate Neumann et al. "High-throughput RNAi screening by time-lapse imaging of live human cells". en. In: *Nat. Methods* 3.5 (May 2006), pp. 385–390.

[53] Cosima Damiana Calvano et al. "MALDI matrices for low molecular weight compounds: an endless story?" In: *Anal. Bioanal. Chem.* 410.17 (July 2018), pp. 4015–4038.

[54] Takahiro Hayasaka et al. "Application of 2,5dihydroxyacetophenone with sublimation provides efficient ionization of lipid species by atmospheric pressure matrixassisted laser desorption/ionization imaging mass spectrometry". In: *Surf. Interface Anal.* 46.12-13 (2014), pp. 1219–1222.

[55] Junhai Yang, Jeremy L Norris, and Richard Caprioli. "Novel vacuum stable ketone-based matrices for high spatial resolution MALDI imaging mass spectrometry". en. In: *J. Mass Spectrom.* 53.10 (Oct. 2018), pp. 1005–1012.

[56] Thorsten W Jaskolla, Wolf-Dieter Lehmann, and Michael Karas. "4-Chloro-alpha-cyanocinnamic acid is an advanced, rationally designed MALDI matrix". en. In: *Proc. Natl. Acad. Sci. U. S. A.* 105.34 (Aug. 2008), pp. 12200–12205.

[57] Naxing Xu et al. “Mercaptobenzothiazoles: A new class of matrices for laser desorption ionization mass spectrometry”. In: *J. Am. Soc. Mass Spectrom.* 8.2 (Feb. 1997), pp. 116–124.

[58] Rachal L Vermillion-Salsbury and David M Hercules. “9-Aminoacridine as a matrix for negative mode matrix-assisted laser desorption/ionization”. en. In: *Rapid Commun. Mass Spectrom.* 16.16 (Aug. 2002), pp. 1575–1581.

[59] Christopher D Cerruti et al. “MALDI imaging and structural analysis of rat brain lipid negative ions with 9-aminoacridine matrix”. en. In: *Anal. Chem.* 84.5 (Mar. 2012), pp. 2164–2171.

[60] Jianing Wang et al. “MALDI-TOF MS imaging of metabolites with a N-(1-naphthyl) ethylenediamine dihydrochloride matrix and its application to colorectal cancer liver metastasis”. en. In: *Anal. Chem.* 87.1 (Jan. 2015), pp. 422–430.

[61] Rohit Shroff et al. “Acid–base-driven matrix-assisted mass spectrometry for targeted metabolomics”. In: *Proceedings of the National Academy of Sciences* 106.25 (June 2009), pp. 10092–10096.

[62] M Giampà et al. “Maleic anhydride proton sponge as a novel MALDI matrix for the visualization of small molecules (<250 m/z) in brain tumors by routine MALDI ToF imaging mass spectrometry”. en. In: *Chem. Commun.* 52.63 (Aug. 2016), pp. 9801–9804.

[63] Aurélien Thomas et al. “Sublimation of new matrix candidates for high spatial resolution imaging mass spectrometry of lipids: enhanced information in both positive and negative polarities after 1,5-diaminonapthalene deposition”. en. In: *Anal. Chem.* 84.4 (Feb. 2012), pp. 2048–2054.

[64] Andrew R Korte and Young Jin Lee. “MALDI-MS analysis and imaging of small molecule metabolites with 1,5-diaminonaphthalene (DAN)”. en. In: *J. Mass Spectrom.* 49.8 (Aug. 2014), pp. 737–741.

[65] Rory T Steven, Alan M Race, and Josephine Bunch. “para-Nitroaniline is a promising matrix for MALDI-MS imaging on intermediate pressure MS systems”. en. In: *J. Am. Soc. Mass Spectrom.* 24.5 (May 2013), pp. 801–804.

[66] Shane R Ellis et al. “More from less: high-throughput dual polarity lipid imaging of biological tissues”. en. In: *Analyst* 141.12 (June 2016), pp. 3832–3841.

[67] Kilian Horatz et al. “Conjugated Polymers as a New Class of Dual-Mode Matrices for MALDI Mass Spectrometry and Imaging”. en. In: *J. Am. Chem. Soc.* (Aug. 2018).

[68] V V Laiko, M A Baldwin, and A L Burlingame. “Atmospheric pressure matrix-assisted laser desorption/ionization mass spectrometry”. en. In: *Anal. Chem.* 72.4 (Feb. 2000), pp. 652–657.

[69] Andrew Palmer et al. “FDR-controlled metabolite annotation for high-resolution imaging mass spectrometry”. en. In: *Nat. Methods* 14.1 (Jan. 2017), pp. 57–60.

[70] Liudmila Prokhorenkova et al. “CatBoost: unbiased boosting with categorical features”. In: (June 2017). arXiv: 1706.09516 [cs.LG].

[71] Moritz Janda et al. “Determination of Abundant Metabolite Matrix Adducts Illuminates the Dark Metabolome of MALDI-Mass Spectrometry Imaging Datasets”. en. In: *Anal. Chem.* 93.24 (June 2021), pp. 8399–8407.

[72] M Reid Groseclose and Stephen Castellino. “A mimetic tissue model for the quantification of drug distributions by MALDI imaging mass spectrometry”. en. In: *Anal. Chem.* 85.21 (Nov. 2013), pp. 10099–10106.

[73] Andrew H Fischer et al. “Preparation of slides and coverslips for microscopy”. en. In: *CSH Protoc.* 2008 (May 2008), db.prot4988.

[74] Matthew C Chambers et al. “A cross-platform toolkit for mass spectrometry and proteomics”. en. In: *Nat. Biotechnol.* 30.10 (Oct. 2012), pp. 918–920.

[75] Alan M Race, Iain B Styles, and Josephine Bunch. “Inclusive sharing of mass spectrometry imaging data requires a converter for all”. en. In: *J. Proteomics* 75.16 (Aug. 2012), pp. 5111–5112.

[76] Jen Nguyen, Deanna M Pepin, and Carolina Tropini. "Cause or effect? The spatial organization of pathogens and the gut microbiota in disease". en. In: *Microbes Infect.* 23.6-7 (July 2021), p. 104815.

[77] Walid Mottawea et al. "The mucosal-luminal interface: an ideal sample to study the mucosa-associated microbiota and the intestinal microbial biogeography". en. In: *Pediatr. Res.* 85.6 (May 2019), pp. 895–903.

[78] Riccardo Vasapolli et al. "Analysis of Transcriptionally Active Bacteria Throughout the Gastrointestinal Tract of Healthy Individuals". en. In: *Gastroenterology* 157.4 (Oct. 2019), 1081–1092.e3.

[79] Dongyao Li et al. "Microbial Biogeography and Core Microbiota of the Rat Digestive Tract". en. In: *Sci. Rep.* 8 (Apr. 2017), p. 45840.

[80] Ravi U Sheth et al. "Spatial metagenomic characterization of microbial biogeography in the gut". en. In: *Nat. Biotechnol.* 37.8 (Aug. 2019), pp. 877–883.

[81] Gregory P Donaldson, S Melanie Lee, and Sarkis K Mazmanian. "Gut biogeography of the bacterial microbiota". en. In: *Nat. Rev. Microbiol.* 14.1 (Jan. 2016), pp. 20–32.

[82] Daniel Machado et al. "Polarization of microbial communities between competitive and cooperative metabolism". en. In: *Nat Ecol Evol* 5.2 (Feb. 2021), pp. 195–203.

[83] Mitsuharu Matsumoto et al. "Impact of intestinal microbiota on intestinal luminal metabolome". en. In: *Sci. Rep.* 2 (Jan. 2012), p. 233.

[84] Yuri Yamamoto et al. "A Metabolomic-Based Evaluation of the Role of Commensal Microbiota throughout the Gastrointestinal Tract in Mice". en. In: *Microorganisms* 6.4 (Sept. 2018).

[85] Silke S Heinzmann and Philippe Schmitt-Kopplin. "Deep metabotyping of the murine gastrointestinal tract for the visualization of digestion and microbial metabolism". en. In: *J. Proteome Res.* 14.5 (May 2015), pp. 2267–2277.

[86] Christopher M Rath et al. "Molecular analysis of model gut microbiotas by imaging mass spectrometry and nanodesorption electrospray ionization reveals dietary metabolite transformations". en. In: *Anal. Chem.* 84.21 (Nov. 2012), pp. 9259–9267.

[87] Anna Nilsson et al. "Mass Spectrometry Imaging proves differential absorption profiles of well-characterised permeability markers along the crypt-villus axis". en. In: *Sci. Rep.* 7.1 (July 2017), p. 6352.

[88] Ibrahim Kaya et al. "Dual polarity MALDI imaging mass spectrometry on the same pixel points reveals spatial lipid localizations at high-spatial resolutions in rat small intestine". en. In: *Anal. Methods* 10.21 (June 2018), pp. 2428–2435.

[89] David S Wishart et al. "HMDB 4.0: the human metabolome database for 2018". en. In: *Nucleic Acids Res.* 46.D1 (Jan. 2018), pp. D608–D617.

[90] Manish Sud et al. "LMSD: LIPID MAPS structure database". en. In: *Nucleic Acids Res.* 35.Database issue (Jan. 2007), pp. D527–32.

[91] Martin I Sigurdsson et al. "A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1". en. In: *BMC Syst. Biol.* 4 (Oct. 2010), p. 140.

[92] Rafael Montenegro-Burke et al. "Metabolomics activity screening of T cell-induced colitis reveals anti-inflammatory metabolites". en. In: *Sci. Signal.* 14.702 (Sept. 2021), eabf6584.

[93] Alex Dexter et al. "Two-Phase and Graph-Based Clustering Methods for Accurate and Efficient Segmentation of Large Mass Spectrometry Images". en. In: *Anal. Chem.* 89.21 (Nov. 2017), pp. 11293–11300.

[94] Nicholas Sofroniew et al. *Napari: a multi-dimensional image viewer for python*. Oct. 2021.

[95] Ingrid Lema et al. "Jejunum: The understudied meeting place of dietary lipids and the microbiota". en. In: *Biochimie* 178 (Nov. 2020), pp. 124–136.

[96] Robert W P Glowacki et al. "A Ribose-Scavenging System Confers Colonization Fitness on the Human Gut Symbiont Bacteroides thetaiotaomicron in a Diet-Specific Manner". en. In: *Cell Host Microbe* 27.1 (Jan. 2020), 79–92.e9.

[97] Christian Sohlenkamp and Otto Geiger. "Bacterial membrane lipids: diversity in structures and pathways". en. In: *FEMS Microbiol. Rev.* 40.1 (Jan. 2016), pp. 133–159.

[98] Stephen R Coats et al. "Cardiolipins Act as a Selective Barrier to Toll-Like Receptor 4 Activation in the Intestine". en. In: *Appl. Environ. Microbiol.* 82.14 (July 2016), pp. 4264–4278.

[99] Douglas V Guzior and Robert A Quinn. "Review: microbial transformations of human bile acids". en. In: *Microbiome* 9.1 (June 2021), p. 140.

[100] Paul A Dawson and Kenneth D R Setchell. "Will the real bile acid sulfotransferase please stand up? Identification of Sult2a8 as a major hepatic bile acid sulfonating enzyme in mice". en. In: *J. Lipid Res.* 58.6 (June 2017), pp. 1033–1035.

[101] Christophe Chassard and Christophe Lacroix. "Carbohydrates and the human gut microbiota". en. In: *Curr. Opin. Clin. Nutr. Metab. Care* 16.4 (July 2013), pp. 453–460.

[102] Kaitlyn Oliphant and Emma Allen-Vercoe. "Macronutrient metabolism by the human gut microbiome: major fermentation by-products and their impact on host health". en. In: *Microbiome* 7.1 (June 2019), p. 91.

[103] Yali Liu et al. "Gut Microbial Metabolites of Aromatic Amino Acids as Signals in Host-Microbe Interplay". en. In: *Trends Endocrinol. Metab.* 31.11 (Nov. 2020), pp. 818–834.

[104] Grant A Hedblom et al. "Segmented Filamentous Bacteria - Metabolism Meets Immunity". en. In: *Front. Microbiol.* 9 (Aug. 2018), p. 1991.

[105] Vivian Chan et al. "Visualization of Helicobacter species within the murine cecal mucosa using specific fluorescence in situ hybridization". en. In: *Helicobacter* 10.2 (Apr. 2005), pp. 114–124.

[106] Daniel Dar et al. "Spatial transcriptomics of planktonic and sessile bacterial populations at single-cell resolution". en. In: *Science* 373.6556 (Aug. 2021).

[107] Hao Shi et al. "Highly multiplexed spatial mapping of microbial communities". en. In: *Nature* 588.7839 (Dec. 2020), pp. 676–681.

[108] Praveen Rahi, Om Prakash, and Yogesh S Shouche. "Matrix-Assisted Laser Desorption/Ionization Time-of-Flight Mass-Spectrometry (MALDI-TOF MS) Based Microbial Identifications: Challenges and Scopes for Microbial Ecologists". en. In: *Front. Microbiol.* 7 (Aug. 2016), p. 1359.

[109] Jeffrey A van Santen et al. "The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery". en. In: *ACS Cent Sci* 5.11 (Nov. 2019), pp. 1824–1833.

[110] Ilias Lagkouvardos et al. "The Mouse Intestinal Bacterial Collection (miBC) provides host-specific insight into cultured diversity and functional potential of the gut microbiota". en. In: *Nat Microbiol* 1.10 (Aug. 2016), p. 16131.

[111] Zahir H Alshehry et al. "An Efficient Single Phase Method for the Extraction of Plasma Lipids". en. In: *Metabolites* 5.2 (June 2015), pp. 389–403.

[112] Shuo Han et al. "A metabolomics pipeline for the mechanistic interrogation of the gut microbiome". en. In: *Nature* 595.7867 (July 2021), pp. 415–420.

[113] Stacey L Heaver, Elizabeth L Johnson, and Ruth E Ley. "Sphingolipids in host-microbial interactions". en. In: *Curr. Opin. Microbiol.* 43 (June 2018), pp. 92–99.

[114] Alesia Walker et al. "Sulfonolipids as novel metabolite markers of Alistipes and Odoribacter affected by high-fat diets". en. In: *Sci. Rep.* 7.1 (Sept. 2017), p. 11047.

[115] Baktiar O Karim and David L Huso. "Mouse models for colorectal cancer". en. In: *Am. J. Cancer Res.* 3.3 (June 2013), pp. 240–250.

[116] Benjamin J Callahan et al. “DADA2: High-resolution sample inference from Illumina amplicon data”. en. In: *Nat. Methods* 13.7 (July 2016), pp. 581–583.

[117] Qiong Wang et al. “Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy”. en. In: *Appl. Environ. Microbiol.* 73.16 (Aug. 2007), pp. 5261–5267.

[118] Christian Quast et al. “The SILVA ribosomal RNA gene database project: improved data processing and web-based tools”. en. In: *Nucleic Acids Res.* 41.Database issue (Jan. 2013), pp. D590–6.

Appendices

Appendix A

Chemical standards with vendor information. Vendors represented:

1. MetaSci, Toronto, ON, Canada
2. Sigma-Aldrich Chemie, Taufkirchen, Germany
3. Avanti Polar Lipids, Birmingham, AL, U.S.A.
4. GLSynthesis, Worcester, MA, U.S.A.

Name	Formula	Monoisotopic mass	Vendor
2-Oxoglutaric acid	C5H6O5	146.0215	1
4,5-Dihydroorotic acid	C5H6N2O4	158.0328	1
5-Hydroxyindole-3-acetic acid	C10H9NO3	191.0582	1
Acetylcholine chloride	C7H16NO2	146.1181	1
Adenine hydrochloride	C5H5N5	135.0545	1
Adenosine 5'-diphosphate potassium salt	C10H15N5O10P2	427.0294	1
Adenosine 5'-monophosphate	C10H14N5O7P	347.0631	1
Adenosine 5'-triphosphate disodium salt	C10H16N5O13P3	506.9958	1
Agmatine sulfate	C5H14N4	130.1218	1
all-trans-Retinoic acid	C20H28O2	300.2089	1
alpha-tocopherol	C29H50O2	430.3811	1
Ascorbic acid	C6H8O6	176.0321	1
Biotin	C10H16N2O3S	244.0882	1
Butyric acid sodium salt	C4H8O2	88.0524	1
Carbamoyl phosphate dilithium salt hydrate	CH4NO5P	140.9827	1
Cholecalciferol	C27H44O	384.3392	1
Cholesterol	C27H46O	386.3549	1
Cholesteryl heptadecanoate	C44H78O2	638.6002	1
Cholic acid	C24H40O5	408.2876	1
Choline chloride	C5H14NO	104.1075	1
cis-Aconitic acid	C6H6O6	174.0164	1
Citric acid	C6H8O7	192.0270	1
Creatine anhydrous	C4H9N3O2	131.0695	1
Creatine phosphate disodium salt hydrate	C4H10N3O5P	211.0358	1
Creatinine	C4H7N3O	113.0589	1
Cyclic AMP	C10H12N5O6P	329.0525	1
Cytidine	C9H13N3O5	243.0855	1
Cytidine 5'-monophosphate	C9H14N3O8P	323.0519	1
Cytosine	C4H5N3O	111.0433	1
D-3-Phosphoglyceric acid disodium salt	C3H7O7P	185.9929	1
D-Fructose 1,6-bisphosphate trisodium salt octahydrate	C6H14O12P2	339.9961	1
D-Glucosamine 6-phosphate	C6H14NO8P	259.0457	1
D-Glucosamine hydrochloride	C6H13NO5	179.0794	1
D-Glucose	C6H12O6	180.0634	1
D-Glucose 6-phosphate sodium salt	C6H13O9P	260.0297	1
Dihydroxyacetone	C3H6O3	90.0317	1
Dopamine hydrochloride	C8H11NO2	153.0790	1

D-Ribose 5-phosphate disodium salt hydrate	C5H11O8P	230.0192	1
Epinephrine	C9H13NO3	183.0895	1
Estradiol	C18H24O2	272.1776	1
Flavin adenine dinucleotide disodium salt hydrate	C27H33N9O15P2	785.1571	1
Flavin mononucleotide sodium salt	C17H21N4O9P	456.1046	1
Folic acid	C19H19N7O6	441.1397	1
Fumaric acid	C4H4O4	116.0110	1
gamma-Aminobutyric acid	C4H9NO2	103.0633	1
Gluconic acid sodium salt	C6H12O7	196.0583	1
Glycerophosphocholine	C8H20NO6P	257.1028	1
Glycine	C2H5NO2	75.0320	1
Guanine	C5H5N5O	151.0494	1
Histamine dihydrochloride	C5H9N3	111.0796	1
Homocysteine	C4H9NO2S	135.0354	1
Hypoxanthine	C5H4N4O	136.0385	1
Indole	C8H7N	117.0578	1
Kynurenic acid	C10H7NO3	189.0426	1
Kynurenine	C10H12N2O3	208.0848	1
L-4-Hydroxyproline	C5H9NO3	131.0582	1
L-Acetylcarnitine hydrochloride	C9H17NO4	203.1158	1
L-Arginine	C6H14N4O2	174.1117	1
L-Asparagine	C4H8N2O3	132.0535	1
L-Aspartic acid	C4H7NO4	133.0375	1
L-Carnitine	C7H15NO3	161.1052	1
L-Carnosine	C9H14N4O3	226.1066	1
L-Citrulline	C6H13N3O3	175.0957	1
L-Cysteine	C3H7NO2S	121.0198	1
L-Cystine	C6H12N2O4S2	240.0239	1
L-Dihydroxyphenylalanine	C9H11NO4	197.0688	1
L-Glutamic acid	C5H9NO4	147.0532	1
L-Glutamine	C5H10N2O3	146.0691	1
L-glutathione oxidised	C20H32N6O12S2	612.1520	1
L-Glutathione reduced	C10H17N3O6S	307.0838	1
L-Histidine	C6H9N3O2	155.0695	1
L-Lactic acid sodium salt	C3H6O3	90.0317	1
L-Malic acid	C4H6O5	134.0215	1
L-Methionine	C5H11NO2S	149.0511	1
L-Phenylalanine	C9H11NO2	165.0790	1
L-Proline	C5H9NO2	115.0633	1
L-Serine	C3H7NO3	105.0426	1
L-Tryptophan	C11H12N2O2	204.0899	1
L-Tyrosine	C9H11NO3	181.0739	1
L-Valine	C5H11NO2	117.0790	1

Malonic acid disodium salt	C3H4O4	104.0110	1
Melatonin	C13H16N2O2	232.1212	1
myo-Inositol	C6H12O6	180.0634	1
N-Acetyl-D-glucosamine	C8H15NO6	221.0899	1
N-Acetyl-L-aspartic acid	C6H9NO5	175.0481	1
Niacinamide	C6H6N2O	122.0480	1
Nicotinamide adenine dinucleotide	C21H28N7O14P2	664.1170	1
Nicotinamide adenine dinucleotide (reduced) disodium salt hydrate	C21H29N7O14P2	665.1248	1
Nicotinic acid	C6H5NO2	123.0320	1
Ornithine hydrochloride	C5H12N2O2	132.0899	1
Orotic acid anhydrous	C5H4N2O4	156.0171	1
Oxalacetic acid	C4H4O5	132.0059	1
Palmitic acid	C16H32O2	256.2402	1
Pantothenic acid sodium salt	C9H17NO5	219.1107	1
Phosphoenolpyruvic acid potassium salt	C3H5O6P	167.9824	1
Phosphoserine	C3H8NO6P	185.0089	1
Phylloquinone	C31H46O2	450.3498	1
Propionylcarnitine hydrochloride	C10H19NO4	217.1314	1
Putrescine dihydrochloride	C4H12N2	88.1000	1
Pyridoxine hydrochloride	C8H11NO3	169.0739	1
Pyruvic acid	C3H4O3	88.0160	1
Quinolinic acid	C7H5NO4	167.0219	1
Riboflavin	C17H20N4O6	376.1383	1
S-(5'-Adenosyl)-L-methionine p-toluenesulfonate salt	C15H23N6O5S	399.1451	1
Serotonin hydrochloride	C10H12N2O	176.0950	1
Spermidine	C7H19N3	145.1579	1
Spermine tetrahydrochloride	C10H26N4	202.2157	1
Succinic acid	C4H6O4	118.0266	1
Taurine	C2H7NO3S	125.0147	1
Taurocholic acid sodium salt	C26H45NO7S	515.2917	1
Thiamine hydrochloride	C12H17N4OS	265.1123	1
Thymine	C5H6N2O2	126.0429	1
trans-Urocanic acid	C6H6N2O2	138.0429	1
Uracil	C4H4N2O2	112.0273	1
Ureidosuccinic acid	C5H8N2O5	176.0433	1
Uric acid	C5H4N4O3	168.0283	1
Uridine	C9H12N2O6	244.0695	1
Uridine 5'-diphosphate disodium salt hydrate	C9H14N2O12P2	404.0022	1
Uridine diphosphate alpha-D-glucose disodium salt	C15H24N2O17P2	566.0550	1
Xanthine	C5H4N4O2	152.0334	1
(R)-Mevalonic acid lithium salt	C6H12O4	148.0736	2

(R)-Pantetheine	C11H22N2O4S	278.1300	2
3-Hydroxy-3-methylglutaric acid	C6H10O5	162.0528	2
3-Hydroxyanthranilic acid	C7H7NO3	153.0426	2
4-nitrobenzyl pyridinium chloride	C12H11N2O2	215.0821	2
5-Methyltetrahydrofolic acid disodium salt	C20H25N7O6	459.1866	2
6-Phosphogluconic acid trisodium salt	C6H13O10P	276.0246	2
Acetoacetic acid lithium salt	C4H6O3	102.0317	2
Acetyl-CoA trisodium salt	C23H38N7O17P3S	809.1258	2
Arachidonic acid	C20H32O2	304.2402	2
Argininosuccinic acid lithium salt	C10H18N4O6	290.1226	2
Cholesteryl acetate	C29H48O2	428.3654	2
Dihydroxyacetone phosphate dilithium salt	C3H7O6P	169.9980	2
D-Ribulose 5-phosphate disodium salt	C5H11O8P	230.0192	2
L-Ribulose	C5H10O5	150.0528	2
Myo-inositol 1-phosphate dipotassium salt	C6H13O9P	260.0297	2
N6,N6,N6-Trimethyl-L-lysine hydrochloride	C9H20N2O2	188.1525	2
N-Acetylaspartylglutamic acid	C11H16N2O8	304.0907	2
N-Acetyl-D-galactosamine 6-phosphate	C8H16NO9P	301.0563	2
Nicotinamide riboside chloride	C11H15N2O5	255.0981	2
Nicotinamide ribotide	C11H15N2O8P	334.0566	2
S-(5'-Adenosyl)-L-homocysteine	C14H20N6O5S	384.1216	2
sn-Glycerol 3-phosphate lithium salt	C3H9O6P	172.0137	2
Uridine 5'-diphosphoglucuronic acid trisodium salt	C15H22N2O18P2	580.0343	2
Uridine 5'-diphospho-N-acetylglucosamine disodium salt	C17H27N3O17P2	607.0816	2
1-(10Z-heptadecenoyl)-2-hydroxy-sn-glycero-3-[phospho-L-serine] (sodium salt)	C23H44NO9P	509.2754	3
1-(10Z-heptadecenoyl)-2-hydroxy-sn-glycero-3-phospho-(1'-myo-inositol) (ammonium salt)	C26H49O12P	584.2962	3
1,2-dioleoyl-sn-glycero-3-phospho-(1'-myo-inositol-5'-phosphate) (diammonium salt)	C45H84O16P2	942.5235	3
1',3'-bis[1,2-dioleoyl-sn-glycero-3-phospho]-glycerol (disodium salt)	C81H150O17P2	1457.0350	3
1,3-dipentadecanoyl-2-oleoyl-glycerol	C51H96O6	804.7207	3
1-O-hexadecyl-2-oleoyl-sn-glycero-3-phosphocholine	C42H84NO7P	745.5985	3

1-oleoyl-2-hydroxy-sn-glycero-3-phosphate (sodium salt)	C21H41O7P	436.2590	3
1-oleoyl-rac-glycerol	C21H40O4	356.2927	3
1-palmitoyl-2-hydroxy-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt)	C22H45O9P	484.2801	3
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate (sodium salt)	C37H71O8P	674.4887	3
1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt)	C40H77O10P	748.5254	3
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoinositol (ammonium salt)	C43H81O13P	836.5415	3
1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (sodium salt)	C40H76NO10P	761.5207	3
1-pentadecanoyl-2-hydroxy-sn-glycero-3-phosphocholine	C23H48NO7P	481.3168	3
1-pentadecanoyl-2-oleoyl-sn-glycero-3-phosphocholine	C41H80NO8P	745.5622	3
1-stearoyl-2-arachidonoyl-sn-glycero-3-phosphoethanolamine	C43H78NO8P	767.5465	3
1-stearoyl-2-docosahexaenoyl-sn-glycerol	C43H72O5	668.5380	3
1-Stearoyl-2-Hydroxy-sn-Glycero-3-Phosphoethanolamine	C23H48NO7P	481.3168	3
3-O-sulfo-D-galactosyl-beta1-1'-N-[2''(S)-hydroxystearoyl]-D-erythro-sphingosine (ammonium salt), (synthetic)	C42H81NO12S	823.5480	3
beta-D-glucosyl cholesterol	C33H56O6	548.4077	3
C18:0 GM3 Ganglioside (synthetic, ammonium salt)	C59H108N2O21	1180.7450	3
Coenzyme A	C21H36N7O16P3S	767.1152	3
D-glucosyl-beta-1,1'-N-stearoyl-D-erythro-sphingosine	C42H81NO8	727.5962	3
N-behenoyl-D-erythro-sphingosine	C40H79NO3	621.6060	3
N-palmitoyl-D-erythro-sphingosylphosphorylcholine	C39H79N2O6P	702.5676	3
palmitoyl L-carnitine	C23H45NO4	399.3349	3
Palmitoyl-CoA triammonium salt	C37H66N7O17P3S	1005.3450	3
Prostaglandin E1	C20H34O5	354.2406	3
4-chlorobenzyl pyridinium chloride	C12H11ClN	204.0580	4
4-cyanobenzyl pyridinium chloride	C20H17N2	285.1392	4
4-methoxybenzyl pyridinium chloride	C13H14NO	200.1075	4
4-methylbenzyl pyridinium chloride	C13H14N	184.1126	4

Appendix B

Chemical classification of the compounds chosen for the standard sample preparation.

Name	Class	Subclass
Agmatine	Amines	Other amines
Carbamoyl phosphate	Amines	Other amines
Histamine	Amines	Other amines
Putrescine	Amines	Other amines
Spermidine	Amines	Other amines
Spermine	Amines	Other amines
Acetylcarnitine	Amines	Quarternary ammonium amines
Acetylcholine	Amines	Quarternary ammonium amines
CAR 16:0	Amines	Quarternary ammonium amines
Carnitine	Amines	Quarternary ammonium amines
Choline	Amines	Quarternary ammonium amines
Glycerophosphocholine	Amines	Quarternary ammonium amines
N6,N6,N6-Trimethyllysine	Amines	Quarternary ammonium amines
Propionylcarnitine	Amines	Quarternary ammonium amines
3-Hydroxyanthranilic acid	Amino acids, peptides, and analogues	Acidic amino acids
Argininosuccinic acid	Amino acids, peptides, and analogues	Acidic amino acids
Aspartic acid	Amino acids, peptides, and analogues	Acidic amino acids
Glutamic acid	Amino acids, peptides, and analogues	Acidic amino acids
N-Acetylaspartic acid	Amino acids, peptides, and analogues	Acidic amino acids
N-Acetylaspartylglutamic acid	Amino acids, peptides, and analogues	Acidic amino acids
Phosphoserine	Amino acids, peptides, and analogues	Acidic amino acids
Ureidosuccinic acid	Amino acids, peptides, and analogues	Acidic amino acids
Agmatine	Amino acids, peptides, and analogues	Arginine derivatives (guanidines)
Arginine	Amino acids, peptides, and analogues	Arginine derivatives (guanidines)
Creatine	Amino acids, peptides, and analogues	Arginine derivatives (guanidines)
Creatinine	Amino acids, peptides, and analogues	Arginine derivatives (guanidines)

Phosphocreatine	Amino acids, peptides, and analogues	Arginine derivatives (guanidines)
Dopamine	Amino acids, peptides, and analogues	Aromatic amino acids
Epinephrine	Amino acids, peptides, and analogues	Aromatic amino acids
Kynurenine	Amino acids, peptides, and analogues	Aromatic amino acids
L-Dopa	Amino acids, peptides, and analogues	Aromatic amino acids
Phenylalanine	Amino acids, peptides, and analogues	Aromatic amino acids
Tyrosine	Amino acids, peptides, and analogues	Aromatic amino acids
Carnosine	Amino acids, peptides, and analogues	Histidine derivatives (imidazoles)
Histamine	Amino acids, peptides, and analogues	Histidine derivatives (imidazoles)
Histidine	Amino acids, peptides, and analogues	Histidine derivatives (imidazoles)
Urocanic acid	Amino acids, peptides, and analogues	Histidine derivatives (imidazoles)
Glycine	Amino acids, peptides, and analogues	Nonpolar amino acids
Proline	Amino acids, peptides, and analogues	Nonpolar amino acids
Valine	Amino acids, peptides, and analogues	Nonpolar amino acids
4-Hydroxyproline	Amino acids, peptides, and analogues	Polar amino acids
Asparagine	Amino acids, peptides, and analogues	Polar amino acids
Carnitine	Amino acids, peptides, and analogues	Polar amino acids
Citrulline	Amino acids, peptides, and analogues	Polar amino acids
gamma-Aminobutyric acid	Amino acids, peptides, and analogues	Polar amino acids
Glutamine	Amino acids, peptides, and analogues	Polar amino acids
N6,N6,N6-Trimethyllysine	Amino acids, peptides, and analogues	Polar amino acids
Ornithine	Amino acids, peptides, and analogues	Polar amino acids
Pantothenic acid	Amino acids, peptides, and analogues	Polar amino acids
Serine	Amino acids, peptides, and analogues	Polar amino acids
Cysteine	Amino acids, peptides, and analogues	Sulphur-containing amino acids
Cystine	Amino acids, peptides, and analogues	Sulphur-containing amino acids
Glutathione	Amino acids, peptides, and analogues	Sulphur-containing amino acids

Homocysteine	Amino acids, peptides, and analogues	Sulphur-containing amino acids
Methionine	Amino acids, peptides, and analogues	Sulphur-containing amino acids
Oxidized glutathione	Amino acids, peptides, and analogues	Sulphur-containing amino acids
Pantetheine	Amino acids, peptides, and analogues	Sulphur-containing amino acids
S-Adenosylhomocysteine	Amino acids, peptides, and analogues	Sulphur-containing amino acids
S-Adenosylmethionine	Amino acids, peptides, and analogues	Sulphur-containing amino acids
Taurine	Amino acids, peptides, and analogues	Sulphur-containing amino acids
5-Hydroxyindoleacetic acid	Amino acids, peptides, and analogues	Tryptophan derivatives (indoles)
Indole	Amino acids, peptides, and analogues	Tryptophan derivatives (indoles)
Melatonin	Amino acids, peptides, and analogues	Tryptophan derivatives (indoles)
Serotonin	Amino acids, peptides, and analogues	Tryptophan derivatives (indoles)
Tryptophan	Amino acids, peptides, and analogues	Tryptophan derivatives (indoles)
Glucosamine	Carbohydrates	Carbohydrate amines
N-Acetylglucosamine	Carbohydrates	Carbohydrate amines
Glucosamine 6-phosphate	Carbohydrates	Carbohydrate amino-phosphates
N-Acetylgalactosamine 6-phosphate	Carbohydrates	Carbohydrate amino-phosphates
6-Phosphogluconic acid	Carbohydrates	Carbohydrate phosphates
Dihydroxyacetone phosphate	Carbohydrates	Carbohydrate phosphates
Fructose 1,6-bisphosphate	Carbohydrates	Carbohydrate phosphates
Glucose 6-phosphate	Carbohydrates	Carbohydrate phosphates
Glycerol 3-phosphate	Carbohydrates	Carbohydrate phosphates
Myo-inositol 1-phosphate	Carbohydrates	Carbohydrate phosphates
Ribose 5-phosphate	Carbohydrates	Carbohydrate phosphates
Ribulose 5-phosphate	Carbohydrates	Carbohydrate phosphates
Dihydroxyacetone	Carbohydrates	Carbohydrates
Gluconic acid	Carbohydrates	Carbohydrates
Glucose	Carbohydrates	Carbohydrates
myo-Inositol	Carbohydrates	Carbohydrates
Ribulose	Carbohydrates	Carbohydrates
Kynurenic acid	Carboxylic acids	Aromatic acids

Nicotinic acid	Carboxylic acids	Aromatic acids
Quinolinic acid	Carboxylic acids	Aromatic acids
3-Phosphoglyceric acid	Carboxylic acids	Carboxylic acid phosphate
6-Phosphogluconic acid	Carboxylic acids	Carboxylic acid phosphate
Phosphoenolpyruvic acid	Carboxylic acids	Carboxylic acid phosphate
Butyric acid	Carboxylic acids	Carboxylic acids
cis-Aconitic acid	Carboxylic acids	Carboxylic acids
Fumaric acid	Carboxylic acids	Carboxylic acids
Malonic acid	Carboxylic acids	Carboxylic acids
Succinic acid	Carboxylic acids	Carboxylic acids
3-Hydroxymethylglutaric acid	Carboxylic acids	Hydroxy acids
Ascorbic acid	Carboxylic acids	Hydroxy acids
Citric acid	Carboxylic acids	Hydroxy acids
Gluconic acid	Carboxylic acids	Hydroxy acids
Lactic acid	Carboxylic acids	Hydroxy acids
Malic acid	Carboxylic acids	Hydroxy acids
Mevalonic acid	Carboxylic acids	Hydroxy acids
2-Oxoglutaric acid	Carboxylic acids	Keto acid
Acetoacetic acid	Carboxylic acids	Keto acid
Oxalacetic acid	Carboxylic acids	Keto acid
Pyruvic acid	Carboxylic acids	Keto acid
Arachidonic acid	Lipids and lipid-like molecules	Fatty acyl
CAR 16:0	Lipids and lipid-like molecules	Fatty acyl
Palmitic acid	Lipids and lipid-like molecules	Fatty acyl
Palmitoyl-CoA	Lipids and lipid-like molecules	Fatty acyl
Prostaglandin E1	Lipids and lipid-like molecules	Fatty acyl
DG 18:0-22:6	Lipids and lipid-like molecules	Glycerolipids
MG 18:1	Lipids and lipid-like molecules	Glycerolipids
TG 15:0-18:1-15:0	Lipids and lipid-like molecules	Glycerolipids
Cardiolipin 18:1	Lipids and lipid-like molecules	Glycerophospholipids
Lyso PA 18:1	Lipids and lipid-like molecules	Glycerophospholipids
Lyso PC 15:0	Lipids and lipid-like molecules	Glycerophospholipids
Lyso PE 18:0	Lipids and lipid-like molecules	Glycerophospholipids
Lyso PG 16:0	Lipids and lipid-like molecules	Glycerophospholipids

Lyso PI 17:1	Lipids and lipid-like molecules	Glycerophospholipids
Lyso PS 17:1	Lipids and lipid-like molecules	Glycerophospholipids
PA 16:0-18:1	Lipids and lipid-like molecules	Glycerophospholipids
PC (O) C16-18:1	Lipids and lipid-like molecules	Glycerophospholipids
PC 15:0-18:1	Lipids and lipid-like molecules	Glycerophospholipids
PE 18:0-20:4	Lipids and lipid-like molecules	Glycerophospholipids
PG 16:0-18:1	Lipids and lipid-like molecules	Glycerophospholipids
PI 16:0-18:1	Lipids and lipid-like molecules	Glycerophospholipids
PI(5)P 18:1	Lipids and lipid-like molecules	Glycerophospholipids
PS (POPS) 16:0-18:1	Lipids and lipid-like molecules	Glycerophospholipids
alpha-tocopherol	Lipids and lipid-like molecules	Prenol lipids
Phylloquinone	Lipids and lipid-like molecules	Prenol lipids
Retinoic acid	Lipids and lipid-like molecules	Prenol lipids
Cer d8:1/22:0 Cer	Lipids and lipid-like molecules	Sphingolipids
GlcCer d18:1/18:0	Lipids and lipid-like molecules	Sphingolipids
GM3 Ganglioside 18:0	Lipids and lipid-like molecules	Sphingolipids
SM d18:1-16:0	Lipids and lipid-like molecules	Sphingolipids
Sulfo GalCer 18:0(2S-OH)	Lipids and lipid-like molecules	Sphingolipids
Cholecalciferol	Lipids and lipid-like molecules	Steroids and steroid derivatives
Cholesterol	Lipids and lipid-like molecules	Steroids and steroid derivatives
Cholesteryl acetate	Lipids and lipid-like molecules	Steroids and steroid derivatives
Cholesteryl ester 17:0	Lipids and lipid-like molecules	Steroids and steroid derivatives
Cholic acid	Lipids and lipid-like molecules	Steroids and steroid derivatives
Estradiol	Lipids and lipid-like molecules	Steroids and steroid derivatives
Glucosyl cholesterol	Lipids and lipid-like molecules	Steroids and steroid derivatives
Taurocholic acid	Lipids and lipid-like molecules	Steroids and steroid derivatives
Nicotinamide adenine dinucleotide (NAD)	Nucleosides, nucleotides, and analogues	Nicotinamide derivatives

Nicotinamide adenine dinucleotide, reduced (NADH)	Nucleosides, nucleotides, and analogues	Nicotinamide derivatives
Nicotinamide riboside	Nucleosides, nucleotides, and analogues	Nicotinamide derivatives
Nicotinamide ribotide	Nucleosides, nucleotides, and analogues	Nicotinamide derivatives
4,5-Dihydroorotic acid	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Adenine	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Cytosine	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Guanine	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Hypoxanthine	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Orotic acid	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Thymine	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Uracil	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Uric acid	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Xanthine	Nucleosides, nucleotides, and analogues	Nucleobases and analogs
Cytidine	Nucleosides, nucleotides, and analogues	Nucleosides
S-Adenosylhomocysteine	Nucleosides, nucleotides, and analogues	Nucleosides
S-Adenosylmethionine	Nucleosides, nucleotides, and analogues	Nucleosides
Uridine	Nucleosides, nucleotides, and analogues	Nucleosides
Acetyl-CoA	Nucleosides, nucleotides, and analogues	Nucleotides
Adenosine diphosphate	Nucleosides, nucleotides, and analogues	Nucleotides
Adenosine monophosphate	Nucleosides, nucleotides, and analogues	Nucleotides
Adenosine triphosphate	Nucleosides, nucleotides, and analogues	Nucleotides
Coenzyme A	Nucleosides, nucleotides, and analogues	Nucleotides
Cyclic adenosine monophosphate	Nucleosides, nucleotides, and analogues	Nucleotides
Cytidine monophosphate	Nucleosides, nucleotides, and analogues	Nucleotides
Flavin adenine dinucleotide	Nucleosides, nucleotides, and analogues	Nucleotides
Flavin mononucleotide	Nucleosides, nucleotides, and analogues	Nucleotides
Palmitoyl-CoA	Nucleosides, nucleotides, and analogues	Nucleotides

Uridine diphosphate	Nucleosides, nucleotides, and analogues	Nucleotides
Uridine diphosphate glucose	Nucleosides, nucleotides, and analogues	Nucleotides
Uridine diphosphate glucuronic acid	Nucleosides, nucleotides, and analogues	Nucleotides
Uridine diphosphate-N-acetylglucosamine	Nucleosides, nucleotides, and analogues	Nucleotides
Acetyl-CoA	Vitamins and cofactors	CoA and derivatives
Coenzyme A	Vitamins and cofactors	CoA and derivatives
Palmitoyl-CoA	Vitamins and cofactors	CoA and derivatives
Flavin adenine dinucleotide	Vitamins and cofactors	Flavins
Flavin mononucleotide	Vitamins and cofactors	Flavins
Riboflavin	Vitamins and cofactors	Flavins
Folic acid	Vitamins and cofactors	Folates
Tetrahydrofolic acid	Vitamins and cofactors	Folates
alpha-tocopherol	Vitamins and cofactors	Vitamins and cofactors
Biotin	Vitamins and cofactors	Vitamins and cofactors
Niacinamide	Vitamins and cofactors	Vitamins and cofactors
Nicotinic acid	Vitamins and cofactors	Vitamins and cofactors
Pantothenic acid	Vitamins and cofactors	Vitamins and cofactors
Phylloquinone	Vitamins and cofactors	Vitamins and cofactors
Pyridoxine	Vitamins and cofactors	Vitamins and cofactors
Retinoic acid	Vitamins and cofactors	Vitamins and cofactors
Thiamine	Vitamins and cofactors	Vitamins and cofactors

Appendix C

Biochemical pathways to which the compounds chosen for the standard sample preparation belong

Name	Pathway Group	Pathway
Agmatine	Amino acid metabolism	Arg metabolism/Urea cycle
Arginine	Amino acid metabolism	Arg metabolism/Urea cycle
Argininosuccinic acid	Amino acid metabolism	Arg metabolism/Urea cycle
Carbamoyl phosphate	Amino acid metabolism	Arg metabolism/Urea cycle
Citrulline	Amino acid metabolism	Arg metabolism/Urea cycle
Ornithine	Amino acid metabolism	Arg metabolism/Urea cycle
Asparagine	Amino acid metabolism	Asp and Asn metabolism
Aspartic acid	Amino acid metabolism	Asp and Asn metabolism
N-Acetylaspartic acid	Amino acid metabolism	Asp and Asn metabolism
Carnitine	Amino acid metabolism	Carnitine biosynthesis
N6,N6,N6-Trimethyllysine	Amino acid metabolism	Carnitine biosynthesis
Creatine	Amino acid metabolism	Creatinine biosynthesis
Creatinine	Amino acid metabolism	Creatinine biosynthesis
Phosphocreatine	Amino acid metabolism	Creatinine biosynthesis
gamma-Aminobutyric acid	Amino acid metabolism	Glu and Gln metabolism
Glutamic acid	Amino acid metabolism	Glu and Gln metabolism
Glutamine	Amino acid metabolism	Glu and Gln metabolism
N-Acetylaspartylglutamic acid	Amino acid metabolism	Glu and Gln metabolism
Carnosine	Amino acid metabolism	His metabolism
Histamine	Amino acid metabolism	His metabolism
Histidine	Amino acid metabolism	His metabolism
Urocanic acid	Amino acid metabolism	His metabolism
Cysteine	Amino acid metabolism	Metabolism of sulfur-containing amino acids
Cystine	Amino acid metabolism	Metabolism of sulfur-containing amino acids
Glutathione	Amino acid metabolism	Metabolism of sulfur-containing amino acids
Oxidized glutathione	Amino acid metabolism	Metabolism of sulfur-containing amino acids
Taurine	Amino acid metabolism	Metabolism of sulfur-containing amino acids
Homocysteine	Amino acid metabolism	Methionine cycle
Methionine	Amino acid metabolism	Methionine cycle
S-Adenosylhomocysteine	Amino acid metabolism	Methionine cycle
S-Adenosylmethionine	Amino acid metabolism	Methionine cycle
Glycine	Amino acid metabolism	Other amino acid metabolism
Valine	Amino acid metabolism	Other amino acid metabolism
Acetoacetic acid	Amino acid metabolism	Phe and Tyr metabolism

Dopamine	Amino acid metabolism	Phe and Tyr metabolism
Epinephrine	Amino acid metabolism	Phe and Tyr metabolism
L-Dopa	Amino acid metabolism	Phe and Tyr metabolism
Phenylalanine	Amino acid metabolism	Phe and Tyr metabolism
Tyrosine	Amino acid metabolism	Phe and Tyr metabolism
Putrescine	Amino acid metabolism	Polyamine biosynthesis
Spermidine	Amino acid metabolism	Polyamine biosynthesis
Spermine	Amino acid metabolism	Polyamine biosynthesis
4-Hydroxyproline	Amino acid metabolism	Pro metabolism
Proline	Amino acid metabolism	Pro metabolism
Phosphoserine	Amino acid metabolism	Ser metabolism
Serine	Amino acid metabolism	Ser metabolism
3-Hydroxyanthranilic acid	Amino acid metabolism	Trp metabolism
5-Hydroxyindoleacetic acid	Amino acid metabolism	Trp metabolism
Indole	Amino acid metabolism	Trp metabolism
Kynurenic acid	Amino acid metabolism	Trp metabolism
Kynurenine	Amino acid metabolism	Trp metabolism
Melatonin	Amino acid metabolism	Trp metabolism
Quinolinic acid	Amino acid metabolism	Trp metabolism
Serotonin	Amino acid metabolism	Trp metabolism
Tryptophan	Amino acid metabolism	Trp metabolism
2-Oxoglutaric acid	Carbohydrate metabolism	Citric acid cycle
Acetyl-CoA	Carbohydrate metabolism	Citric acid cycle
cis-Aconitic acid	Carbohydrate metabolism	Citric acid cycle
Citric acid	Carbohydrate metabolism	Citric acid cycle
Fumaric acid	Carbohydrate metabolism	Citric acid cycle
Malic acid	Carbohydrate metabolism	Citric acid cycle
Oxalacetic acid	Carbohydrate metabolism	Citric acid cycle
Succinic acid	Carbohydrate metabolism	Citric acid cycle
3-Phosphoglyceric acid	Carbohydrate metabolism	Glycolysis
Dihydroxyacetone	Carbohydrate metabolism	Glycolysis
Dihydroxyacetone phosphate	Carbohydrate metabolism	Glycolysis
Fructose 1,6-bisphosphate	Carbohydrate metabolism	Glycolysis
Glucose	Carbohydrate metabolism	Glycolysis
Glucose 6-phosphate	Carbohydrate metabolism	Glycolysis
Glycerol 3-phosphate	Carbohydrate metabolism	Glycolysis
Lactic acid	Carbohydrate metabolism	Glycolysis
Phosphoenolpyruvic acid	Carbohydrate metabolism	Glycolysis
Pyruvic acid	Carbohydrate metabolism	Glycolysis
Glucosamine	Carbohydrate metabolism	Hexosamine biosynthetic pathway
Glucosamine 6-phosphate	Carbohydrate metabolism	Hexosamine biosynthetic pathway
N-Acetylgalactosamine 6-phosphate	Carbohydrate metabolism	Hexosamine biosynthetic pathway

N-Acetylglucosamine	Carbohydrate metabolism	Hexosamine biosynthetic pathway
Uridine diphosphate glucose	Carbohydrate metabolism	Hexosamine biosynthetic pathway
Uridine diphosphate glucuronic acid	Carbohydrate metabolism	Hexosamine biosynthetic pathway
Uridine diphosphate-N-acetylglucosamine	Carbohydrate metabolism	Hexosamine biosynthetic pathway
myo-Inositol	Carbohydrate metabolism	Inositol metabolism
Myo-inositol 1-phosphate	Carbohydrate metabolism	Inositol metabolism
6-Phosphogluconic acid	Carbohydrate metabolism	Pentose phosphate pathway
Gluconic acid	Carbohydrate metabolism	Pentose phosphate pathway
Glucose 6-phosphate	Carbohydrate metabolism	Pentose phosphate pathway
Ribose 5-phosphate	Carbohydrate metabolism	Pentose phosphate pathway
Ribulose	Carbohydrate metabolism	Pentose phosphate pathway
Ribulose 5-phosphate	Carbohydrate metabolism	Pentose phosphate pathway
Arachidonic acid	Lipid metabolism	Arachidonic acid metabolism
Prostaglandin E1	Lipid metabolism	Arachidonic acid metabolism
Cholesteryl acetate	Lipid metabolism	Cholesterol metabolism
Cholesteryl ester 17:0	Lipid metabolism	Cholesterol metabolism
Cholic acid	Lipid metabolism	Cholesterol metabolism
Estradiol	Lipid metabolism	Cholesterol metabolism
Glucosyl cholesterol	Lipid metabolism	Cholesterol metabolism
Taurocholic acid	Lipid metabolism	Cholesterol metabolism
3-Hydroxymethylglutaric acid	Lipid metabolism	Cholesterol synthesis
Acetyl-CoA	Lipid metabolism	Cholesterol synthesis
Cholesterol	Lipid metabolism	Cholesterol synthesis
Mevalonic acid	Lipid metabolism	Cholesterol synthesis
Acetylcholine	Lipid metabolism	Choline metabolism
Choline	Lipid metabolism	Choline metabolism
Glycerophosphocholine	Lipid metabolism	Choline metabolism
Acetylcarnitine	Lipid metabolism	Fatty acid oxidation
CAR 16:0	Lipid metabolism	Fatty acid oxidation
Carnitine	Lipid metabolism	Fatty acid oxidation
Palmitoyl-CoA	Lipid metabolism	Fatty acid oxidation
Propionylcarnitine	Lipid metabolism	Fatty acid oxidation
Acetyl-CoA	Lipid metabolism	Fatty acid synthesis
Butyric acid	Lipid metabolism	Fatty acid synthesis
Citric acid	Lipid metabolism	Fatty acid synthesis
Malonic acid	Lipid metabolism	Fatty acid synthesis
Palmitic acid	Lipid metabolism	Fatty acid synthesis

DG 18:0-22:6	Lipid metabolism	Glycerolipid metabolism
MG 18:1	Lipid metabolism	Glycerolipid metabolism
TG 15:0-18:1-15:0	Lipid metabolism	Glycerolipid metabolism
Cardiolipin 18:1	Lipid metabolism	Phospholipid metabolism
Lyso PA 18:1	Lipid metabolism	Phospholipid metabolism
Lyso PC 15:0	Lipid metabolism	Phospholipid metabolism
Lyso PE 18:0	Lipid metabolism	Phospholipid metabolism
Lyso PG 16:0	Lipid metabolism	Phospholipid metabolism
Lyso PI 17:1	Lipid metabolism	Phospholipid metabolism
Lyso PS 17:1	Lipid metabolism	Phospholipid metabolism
PA 16:0-18:1	Lipid metabolism	Phospholipid metabolism
PC (O) C16-18:1	Lipid metabolism	Phospholipid metabolism
PC 15:0-18:1	Lipid metabolism	Phospholipid metabolism
PE 18:0-20:4	Lipid metabolism	Phospholipid metabolism
PG 16:0-18:1	Lipid metabolism	Phospholipid metabolism
PI 16:0-18:1	Lipid metabolism	Phospholipid metabolism
PI(5)P 18:1	Lipid metabolism	Phospholipid metabolism
PS (POPS) 16:0-18:1	Lipid metabolism	Phospholipid metabolism
Cer d8:1/22:0 Cer	Lipid metabolism	Sphingolipid metabolism
GlcCer d18:1/18:0	Lipid metabolism	Sphingolipid metabolism
GM3 Ganglioside 18:0	Lipid metabolism	Sphingolipid metabolism
SM d18:1-16:0	Lipid metabolism	Sphingolipid metabolism
Sulfo GalCer 18:0(2S-OH)	Lipid metabolism	Sphingolipid metabolism
Coenzyme A	Metabolism of vitamins and cofactors	Coenzyme A biosynthesis
Pantetheine	Metabolism of vitamins and cofactors	Coenzyme A biosynthesis
Pantothenic acid	Metabolism of vitamins and cofactors	Coenzyme A biosynthesis
Folic acid	Metabolism of vitamins and cofactors	Folate cycle
Tetrahydrofolic acid	Metabolism of vitamins and cofactors	Folate cycle
Niacinamide	Metabolism of vitamins and cofactors	NAD cycle
Nicotinamide adenine dinucleotide (NAD)	Metabolism of vitamins and cofactors	NAD cycle
Nicotinamide adenine dinucleotide, reduced (NADH)	Metabolism of vitamins and cofactors	NAD cycle
Nicotinamide riboside	Metabolism of vitamins and cofactors	NAD cycle
Nicotinamide ribotide	Metabolism of vitamins and cofactors	NAD cycle
Nicotinic acid	Metabolism of vitamins and cofactors	NAD cycle
Flavin adenine dinucleotide	Metabolism of vitamins and cofactors	Redox metabolism
Flavin mononucleotide	Metabolism of vitamins and cofactors	Redox metabolism

Nicotinamide adenine dinucleotide (NAD)	Metabolism of vitamins and cofactors	Redox metabolism
Nicotinamide adenine dinucleotide, reduced (NADH)	Metabolism of vitamins and cofactors	Redox metabolism
alpha-tocopherol	Metabolism of vitamins and cofactors	Vitamin metabolism
Ascorbic acid	Metabolism of vitamins and cofactors	Vitamin metabolism
Biotin	Metabolism of vitamins and cofactors	Vitamin metabolism
Cholecalciferol	Metabolism of vitamins and cofactors	Vitamin metabolism
Phylloquinone	Metabolism of vitamins and cofactors	Vitamin metabolism
Pyridoxine	Metabolism of vitamins and cofactors	Vitamin metabolism
Retinoic acid	Metabolism of vitamins and cofactors	Vitamin metabolism
Riboflavin	Metabolism of vitamins and cofactors	Vitamin metabolism
Thiamine	Metabolism of vitamins and cofactors	Vitamin metabolism
Adenine	Nucleotide metabolism	Purine metabolism
Adenosine diphosphate	Nucleotide metabolism	Purine metabolism
Adenosine monophosphate	Nucleotide metabolism	Purine metabolism
Adenosine triphosphate	Nucleotide metabolism	Purine metabolism
Cyclic adenosine monophosphate	Nucleotide metabolism	Purine metabolism
Guanine	Nucleotide metabolism	Purine metabolism
Hypoxanthine	Nucleotide metabolism	Purine metabolism
Xanthine	Nucleotide metabolism	Purine metabolism
4,5-Dihydroorotic acid	Nucleotide metabolism	Pyrimidine metabolism
Cytidine	Nucleotide metabolism	Pyrimidine metabolism
Cytidine monophosphate	Nucleotide metabolism	Pyrimidine metabolism
Cytosine	Nucleotide metabolism	Pyrimidine metabolism
Orotic acid	Nucleotide metabolism	Pyrimidine metabolism
Thymine	Nucleotide metabolism	Pyrimidine metabolism
Uracil	Nucleotide metabolism	Pyrimidine metabolism
Ureidosuccinic acid	Nucleotide metabolism	Pyrimidine metabolism
Uric acid	Nucleotide metabolism	Pyrimidine metabolism
Uridine	Nucleotide metabolism	Pyrimidine metabolism
Uridine diphosphate	Nucleotide metabolism	Pyrimidine metabolism

Appendix D

Preparation protocols used for metabolite standards. Diluent solvent systems:

1) 9:1 methanol/water (v/v)
2) 1:2 methanol/chloroform (v/v).

Titrant shorthand:

NaOH – sodium hydroxide (aqueous)

HCl – hydrochloric acid (aqueous)

Name	Stock concentration / mM	Stock solvent	Dilution / µM	Diluent	Titrant
D-Glucose	98.3	water	200	1	
D-Glucose 6-phosphate sodium salt	39.0	water	200	1	
D-Fructose 1,6-bisphosphate trisodium salt octahydrate	79.8	water	200	1	
Dihydroxyacetone phosphate dilithium salt	5.5	water	200	1	
D-3-Phosphoglyceric acid disodium salt	21.4	water	200	1	
Phosphoenolpyruvic acid potassium salt	63.7	water	200	1	
Pyruvic acid	211.4	water	200	1	
Acetyl-CoA trisodium salt	1.1	water	200	1	
L-Lactic acid sodium salt	93.6	water	200	1	
6-Phosphogluconic acid trisodium salt	2.9	water	200	1	
D-Ribulose 5-phosphate disodium salt	3.6	water	200	1	
D-Ribose 5-phosphate disodium salt hydrate	40.8	water	200	1	
Dihydroxyacetone	169.8	water	200	1	
Gluconic acid sodium salt	87.6	water	200	1	
L-Ribulose	199.8	water	200	1	
Citric acid	79.1	water	200	1	
cis-Aconitic acid	76.5	water	200	1	
2-Oxoglutaric acid	158.8	water	200	1	
Succinic acid	124.7	water	200	1	
Fumaric acid	137.0	methanol	200	2	
L-Malic acid	79.1	water	200	1	
Oxalacetic acid	77.2	water	200	1	
Uridine diphosphate alpha-D-glucose disodium salt	8.2	water	200	1	
Uridine 5'-diphosphoglucuronic acid trisodium salt	1.5	water	200	1	
D-Glucosamine 6-phosphate	20.4	water	200	1	

N-Acetyl-D-galactosamine 6-phosphate	3.3	water	200	1	
Uridine 5'-diphospho-N-acetylglucosamine disodium salt	5.4	water	200	1	
N-Acetyl-D-glucosamine	51.0	water	200	1	
myo-Inositol	174.8	water	200	1	
Ascorbic acid	98.2	water	200	1	
Myo-inositol 1-phosphate dipotassium salt	28.5	water	200	1	
Palmitic acid	67.5	methanol	200	2	
(R)-Mevalonic acid lithium salt	6.5	water	200	1	
Cholesterol	26.9	1:5 chloroform/methanol	200	2	
Malonic acid disodium salt	33.3	water	200	1	
3-Hydroxy-3-methylglutaric acid	6.2	water	200	1	
Palmitoyl-CoA triammonium salt	4.7	2:1:1 methanol/water/chloroform	200	2	
L-Carnitine	93.6	water	200	1	
Cholic acid	46.7	methanol	200	2	
Glycine	251.8	water	200	1	
Taurocholic acid sodium salt	26.8	40% methanol	200	1	
Estradiol	40.8	methanol	200	2	
L-Asparagine	160.3	20% methanol	200	1	HCl
L-Aspartic acid	119.6	20% methanol	200	1	HCl
N-Acetyl-L-aspartic acid	100.7	water	200	1	
L-Glutamine	109.5	water	200	1	
L-Glutamic acid	201.8	20% methanol	200	1	HCl
N-Acetylaspartylglutamic acid	3.3	water	200	1	
gamma-Aminobutyric acid	164.9	water	200	1	
L-Phenylalanine	110.4	20% methanol	200	1	HCl
L-Tyrosine	112.0	20% methanol	200	1	HCl
L-Dihydroxyphenylalanine	99.9	20% methanol	200	1	HCl

Dopamine hydrochloride	69.6	water	200	1	
Epinephrine	47.3	90% methanol	200	1	NaOH
Acetoacetic acid lithium salt	9.3	water	200	1	
L-Histidine	92.2	water	200	1	
L-Carnosine	74.3	water	200	1	
trans-Urocanic acid	97.4	40% methanol	200	1	NaOH
Histamine dihydrochloride	97.8	water	200	1	
L-Arginine	109.1	water	200	1	
L-Citrulline	134.1	water	200	1	
Argininosuccinic acid lithium salt	3.4	water	200	1	
Ornithine hydrochloride	102.0	water	200	1	
Carbamoyl phosphate dilithium salt hydrate	63.4	water	200	1	
L-Proline	271.0	water	200	1	
L-4-Hydroxyproline	176.2	water	200	1	
Putrescine dihydrochloride	87.5	water	200	1	
Spermidine	477.8	water	200	1	
Spermine tetrahydrochloride	33.0	water	200	1	
Creatine anhydrous	117.4	water	200	1	
Creatinine	282.0	40% methanol	200	1	
Agmatine sulfate	55.6	water	200	1	
Folic acid	45.8	40% methanol	200	1	NaOH
5-Methyltetrahydrofolic acid disodium salt	2.0	40% methanol	200	1	
L-Methionine	179.6	20% methanol	200	1	HCl
L-Cystine	142.3	20% methanol	200	1	HCl
L-glutathione oxidised	24.3	water	200	1	
L-Glutathione reduced	36.7	water	200	1	
Taurine	112.3	water	200	1	
S-(5'-Adenosyl)-L-methionine p-toluenesulfonate salt	27.0	water	200	1	

S-(5'-Adenosyl)-L-homocysteine	2.6	water	200	1	
Homocysteine	102.8	water	200	1	
L-Cysteine	203.0	water	200	1	HCl
N6,N6,N6-Trimethyl-L-lysine hydrochloride	9.4	water	200	1	
L-Valine	398.6	water	200	1	
Pantothenic acid sodium salt	55.6	water	200	1	
(R)-Pantetheine	3.6	methanol	200	1	
Coenzyme A	3.0	water	200	1	
L-Serine	211.2	water	200	1	
L-Tryptophan	96.0	water	200	1	HCl
Kynurenine	24.0	water	200	1	HCl
Kynurenic acid	43.0	50% methanol	200	1	NaOH
3-Hydroxyanthranilic acid	6.5	methanol	200	1	
Quinolinic acid	118.5	20% methanol	200	1	NaOH
Nicotinamide adenine dinucleotide	15.5	water	200	1	
Nicotinamide adenine dinucleotide (reduced) disodium salt hydrate	15.1	water	200	1	
Niacinamide	147.4	water	200	1	
Nicotinic acid	139.2	95% methanol	200	1	NaOH
Serotonin hydrochloride	111.4	water	200	1	
5-Hydroxyindole-3-acetic acid	68.0	ethanol	200	1	
Melatonin	111.5	methanol	200	2	
Nicotinamide riboside chloride	3.4	water	200	1	
Nicotinamide ribotide	3.0	water	200	1	
Hypoxanthine	62.4	water	200	1	NaOH
Adenosine 5'-monophosphate	60.3	water	200	1	NaOH
Adenine hydrochloride	57.9	water	200	1	
Xanthine	84.8	10% methanol	200	1	NaOH
Guanine	91.5	water	200	1	NaOH

Adenosine 5'-diphosphate potassium salt	30.9	water	200	1	
Adenosine 5'-triphosphate disodium salt	55.3	water	200	1	
Cyclic AMP	24.1	water	200	1	
Ureidosuccinic acid	72.3	water	200	1	
4,5-Dihydroorotic acid	105.0	water	200	1	
Orotic acid anhydrous	67.0	40% methanol	200	1	NaOH
Uridine	100.3	water	200	1	
Uracil	161.5	water	200	1	NaOH
Thymine	119.7	water	200	1	NaOH
Cytidine 5'-monophosphate	60.7	water	200	1	
Cytidine	42.3	water	200	1	
Cytosine	106.3	water	200	1	HCl
Uridine 5'-diphosphate disodium salt hydrate	72.5	water	200	1	
Choline chloride	208.4	water	200	1	
Acetylcholine chloride	239.5	water	200	1	
Thiamine hydrochloride	47.1	water	200	1	
Riboflavin	27.4	water	200	1	NaOH
Pyridoxine hydrochloride	56.4	water	200	1	
Biotin	59.3	40% methanol	200	1	NaOH
Cholecalciferol	76.7	methanol	200	2	
alpha-tocopherol	165.4	methanol	200	2	
Phylloquinone	163.7	chloroform	200	2	
Flavin mononucleotide sodium salt	22.4	water	200	1	
Flavin adenine dinucleotide disodium salt hydrate	13.8	water	200	1	
Creatine phosphate disodium salt hydrate	57.2	water	200	1	
Glycerophosphocholine	47.4	water	200	1	
Propionylcarnitine hydrochloride	75.7	40% methanol	200	1	
L-Acetylcarnitine hydrochloride	51.1	water	200	1	
Phosphoserine	55.4	water	200	1	

Uric acid	81.3	water	200	1	NaOH
Cholesteryl acetate	5.3	chloroform	200	2	
Indole	126.3	methanol	200	2	
D-Glucosamine hydrochloride	80.7	water	200	1	
Butyric acid sodium salt	193.5	water	200	1	
sn-Glycerol 3-phosphate lithium salt	5.8	water	200	1	
all-trans-Retinoic acid	45.9	chloroform	200	2	
Arachidonic acid	3.3	chloroform	200	2	
1-pentadecanoyl-2-oleoyl-sn-glycero-3-phosphocholine	1.3	chloroform	200	2	
1,3-dipentadecanoyl-2-oleoyl-glycerol	1.2	chloroform	200	2	
1-oleoyl-rac-glycerol	2.8	chloroform	200	2	
beta-D-glucosyl cholesterol	3.0	2:1 chloroform/methanol	200	2	
1',3'-bis[1,2-dioleoyl-sn-glycero-3-phospho]-glycerol (disodium salt)	16.6	chloroform	200	2	
1-stearoyl-2-docosahexaenoyl-sn-glycerol	3.0	chloroform	200	2	
1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (sodium salt)	12.8	chloroform	200	2	
1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt)	13.0	chloroform	200	2	
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate (sodium salt)	35.9	chloroform	200	2	
1-(10Z-heptadecenoyl)-2-hydroxy-sn-glycero-3-phospho-(1'-myo-inositol) (ammonium salt)	0.2	1:1 chloroform/methanol	50	2	
1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoinositol (ammonium salt)	0.1	chloroform	50	2	
1,2-dioleoyl-sn-glycero-3-phospho-(1'-myo-inositol-5'-phosphate) (diammonium salt)	0.2	chloroform	50	2	
1-stearoyl-2-arachidonoyl-sn-glycero-3-phosphoethanolamine	13.0	chloroform	200	2	

1-pentadecanoyl-2-hydroxy-sn-glycero-3-phosphocholine	8.5	1:1 chloroform/methanol	200	2	
1-Stearoyl-2-Hydroxy-sn-Glycero-3-Phosphoethanolamine	1.1	1:1 chloroform/methanol	200	2	
1-oleoyl-2-hydroxy-sn-glycero-3-phosphate (sodium salt)	10.9	chloroform	200	2	
1-palmitoyl-2-hydroxy-sn-glycero-3-phospho-(1'-rac-glycerol) (sodium salt)	8.7	methanol	200	2	
1-(10Z-heptadecenoyl)-2-hydroxy-sn-glycero-3-[phospho-L-serine] (sodium salt)	4.7	1:1 chloroform/methanol	200	2	
C18:0 GM3 Ganglioside (synthetic, ammonium salt)	0.1	methanol	10	2	
N-behenoyl-D-erythro-sphingosine	2.7	2:1 chloroform/methanol	200	2	
D-glucosyl-beta-1,1'-N-stearoyl-D-erythro-sphingosine	6.9	1:1 chloroform/methanol	200	2	
N-palmitoyl-D-erythro-sphingosylphosphorylcholine	7.1	chloroform	200	2	
3-O-sulfo-D-galactosyl-beta1-1'-N-[2"(S)-hydroxystearoyl]-D-erythro-sphingosine (ammonium salt), (synthetic)	1.2	chloroform	200	2	
palmitoyl L-carnitine	6.3	methanol	200	2	
1-O-hexadecyl-2-oleoyl-sn-glycero-3-phosphocholine	3.4	chloroform	200	2	
Prostaglandin E1	1.4	1:1 chloroform/methanol	200	2	
Cholesteryl heptadecanoate	11.9	1:1 methanol/chloroform	200	2	
4-nitrobenzyl pyridinium chloride	1.4	methanol	200	2	
4-methoxybenzyl pyridinium chloride	0.9	methanol	200	2	
4-methylbenzyl pyridinium chloride	4.3	methanol	200	2	
4-cyanobenzyl pyridinium chloride	0.6	methanol	200	2	
4-chlorobenzyl pyridinium chloride	4.1	methanol	200	2	

Appendix E

Letter to collaborators with instructions for METASPACE Interlab Imaging Mass Spectrometry Study

"First of all, thank you very much for your participation!

We are sending you 5 identical slides (75x25x1 mm) with authentic standards spotted as indicated at the end of this file. Please let the sample warm up to the room temperature (~1h) before opening the vacuum-sealed bag.

We ask you to pick two protocols that are commonly used in your lab for biological samples: one for positive mode, and one for negative mode. For each protocol, please collect 2 datasets that will be used as replicates. For example, in our team we will analyze 2 slides with the DAN matrix in negative mode and 2 slides with the DHB matrix in positive mode as these two protocols are the most commonly used in our team. One slide is provided as a backup.

For the imaging mass spectrometry analysis:

- The neutral monoisotopic masses of the standards range from 75 to 1500 Da. If you consider splitting up the m/z range, please submit to METASPACE the individual datasets as separate submissions, indicating for each submission the m/z range in the "Additional information" metadata field (see example below). The number of required submissions to METASPACE will increase depending on the number of m/z ranges used. For example, when using 1 m/z range, 4 datasets are to be submitted (2 replicates * 2 protocols * 1 m/z range); however, when using 2 m/z ranges, 8 datasets are to be submitted (2 replicates * 2 protocols * 2 m/z ranges).

- Please acquire data within a rectangular area comfortably covering all the spots on the glass slide. Silver pen marks on the back side of the slide show where the spots are located. If needed, the same slide can be sampled more than once if undersampled.

- Set the mass resolving power at m/z 400 to be as close as possible to 120 000 and not lower than 50 000

- Try to ensure that the data has less than 5 ppm mass error throughout the whole m/z range (in case you see stronger deviations, please contact us)

- Use the pixel size / pitch of 100 µm in both x and y (approximately 80 000 pixels for 4 cm x 2 cm). On instruments that cannot do this, please use the closest setting.

- Make sure the data is centroided prior to submission to METASPACE

- Prepare and store your raw data (uncentroided) for sharing it with us later - it will be deposited into a public repository

- If data is acquired using polarity switching, please split into separate files as
METASPACE only accepts single polarity data.

For each dataset in METASPACE, please also provide the metadata as detailed as possible, in particular, please include:

- details on the sample preparation (DESI settings, MALDI matrix application parameters)

- sample number as printed on the slide label, in the Additional Information metadata field.

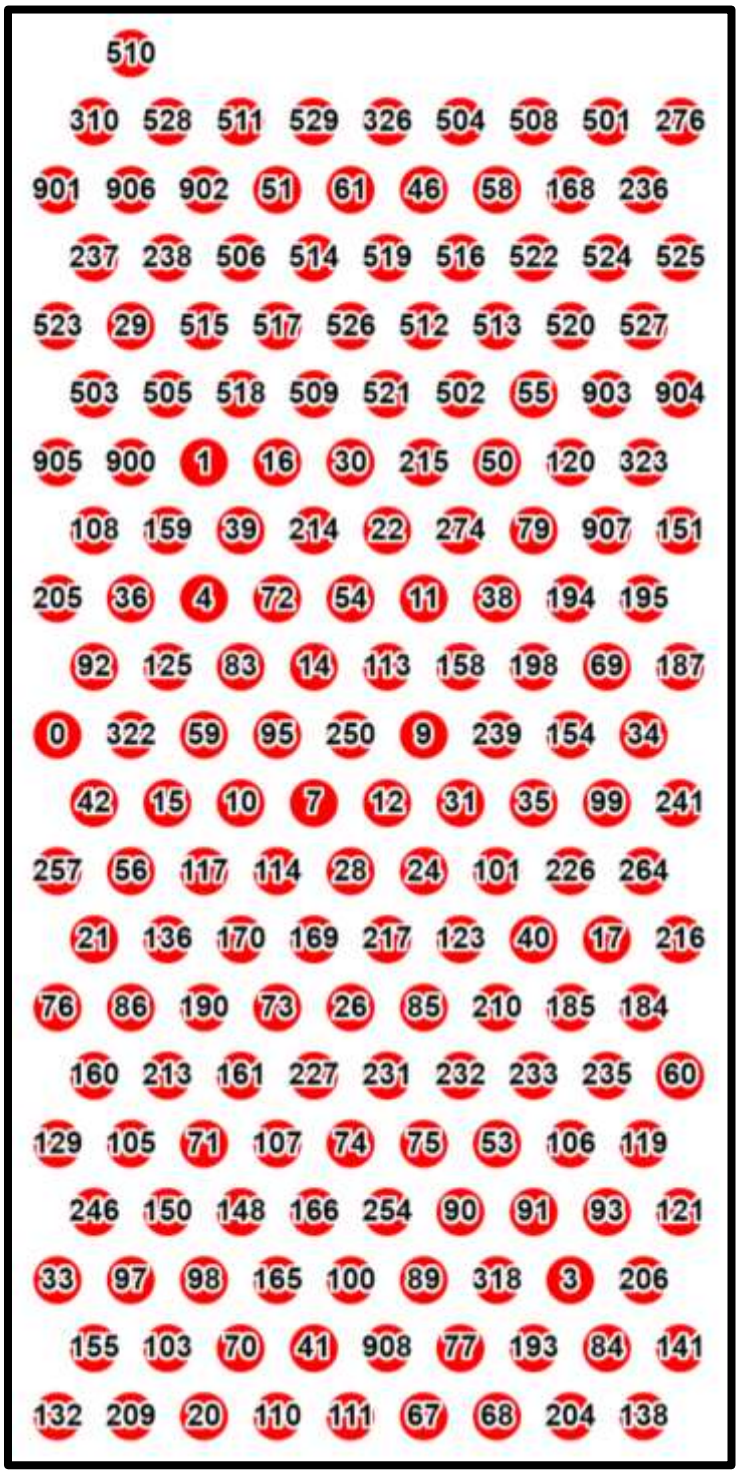
Please submit the datasets to METASPACE within 1 month. After submission, please email us. If you submit the datasets as private, for this particular study you give us (the Alexandrov team at EMBL) a special permission to look at them, work with them, and use them for this study. We will evaluate the datasets, review their quality and get back to you with a summary of our assessment. At that point, you can still withdraw the datasets from the study.

If you have any questions, please email.

Thank you!
Veronika, Mans, Theodore
Alexandrov team, EMBL”

Appendix F

Slide Layout with key



Key	Name
0	Glucose
1	Glucose 6-phosphate
1	Glucose 6-phosphate
3	Fructose 1,6-bisphosphate
4	Dihydroxyacetone phosphate
7	3-Phosphoglyceric acid
9	Phosphoenolpyruvic acid
10	Pyruvic acid
11	Acetyl-CoA
11	Acetyl-CoA
11	Acetyl-CoA
12	Lactic acid
14	6-Phosphogluconic acid
15	Ribulose 5-phosphate
16	Ribose 5-phosphate
17	Dihydroxyacetone
20	Gluconic acid
21	Ribulose
22	Citric acid
22	Citric acid
24	cis-Aconitic acid
26	2-Oxoglutaric acid
28	Succinic acid
29	Fumaric acid
30	Malic acid
31	Oxalacetic acid
33	Uridine diphosphate glucose
34	Uridine diphosphate glucuronic acid
35	Glucosamine 6-phosphate
36	N-Acetylgalactosamine 6-phosphate
38	Uridine diphosphate-N-acetylglucosamine
39	N-Acetylglucosamine
40	myo-Inositol
41	Ascorbic acid
42	Myo-inositol 1-phosphate
46	Palmitic acid
50	Mevalonic acid
51	Cholesterol
53	Malonic acid
54	3-Hydroxymethylglutaric acid
55	Palmitoyl-CoA
56	Carnitine

56	Carnitine
58	Cholic acid
59	Glycine
60	Taurocholic acid
61	Estradiol
67	Asparagine
68	Aspartic acid
69	N-Acetylaspartic acid
70	Glutamine
71	Glutamic acid
72	N-Acetylaspartylglutamic acid
73	gamma-Aminobutyric acid
74	Phenylalanine
75	Tyrosine
76	L-Dopa
77	Dopamine
79	Epinephrine
83	Acetoacetic acid
84	Histidine
85	Carnosine
86	Urocanic acid
89	Histamine
90	Arginine
91	Citrulline
92	Argininosuccinic acid
93	Ornithine
95	Carbamoyl phosphate
97	Proline
98	4-Hydroxyproline
99	Putrescine
100	Spermidine
101	Spermine
103	Creatine
105	Creatinine
106	Agmatine
107	Folic acid
108	Tetrahydrofolic acid
110	Methionine
111	Cystine
113	Oxidized glutathione
114	Glutathione
117	Taurine
119	S-Adenosylmethionine
120	S-Adenosylhomocysteine

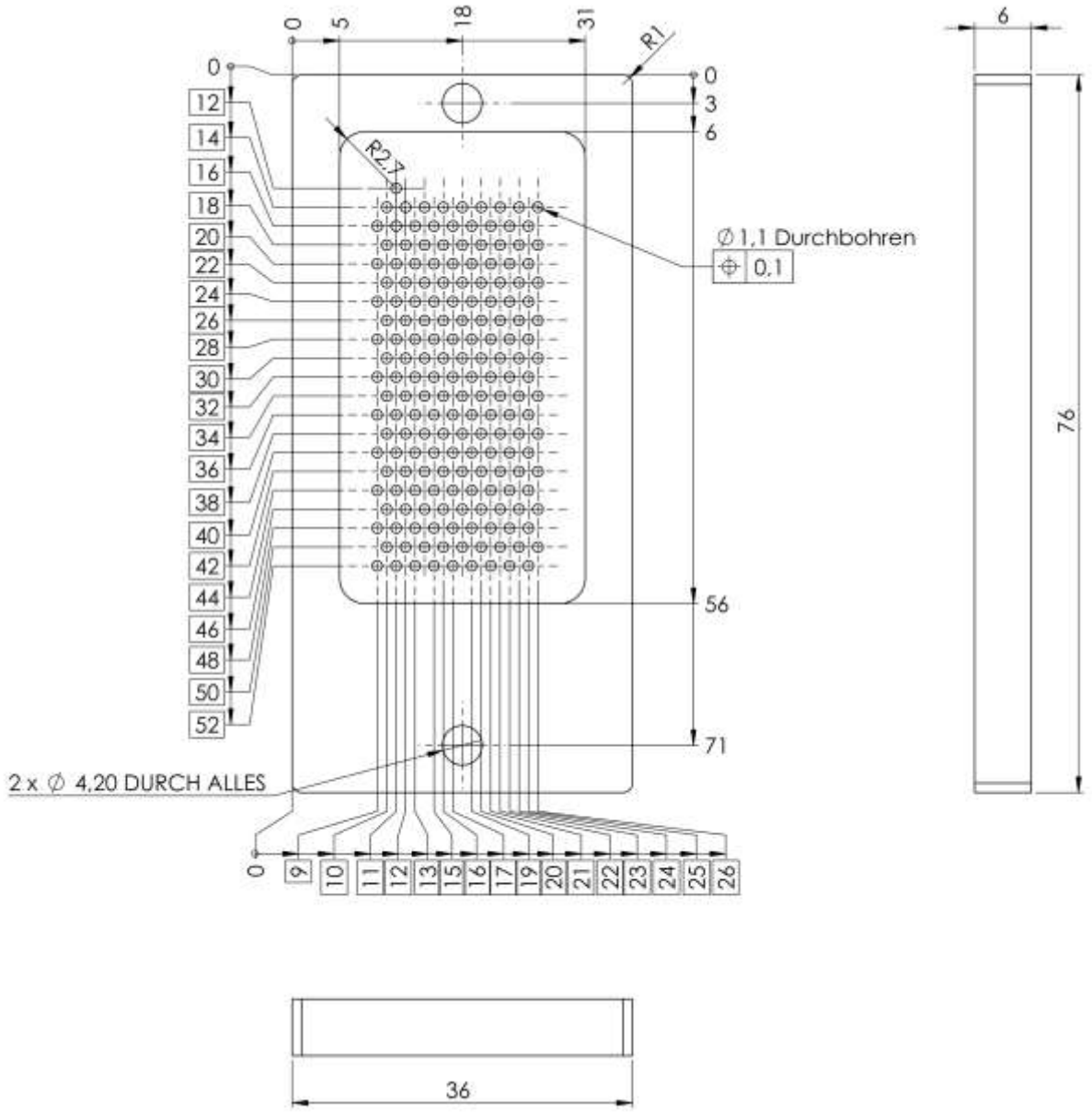
121	Homocysteine
123	Cysteine
125	N6,N6,N6-Trimethyllysine
129	Valine
132	Pantothenic acid
136	Pantetheine
138	Coenzyme A
141	Serine
148	Tryptophan
150	Kynurenine
151	Kynurenic acid
154	3-Hydroxyanthranilic acid
155	Quinolinic acid
158	Nicotinamide adenine dinucleotide (NAD)
158	Nicotinamide adenine dinucleotide (NAD)
159	Nicotinamide adenine dinucleotide, reduced (NADH)
159	Nicotinamide adenine dinucleotide, reduced (NADH)
160	Niacinamide
161	Nicotinic acid
165	Serotonin
166	5-Hydroxyindoleacetic acid
168	Melatonin
169	Nicotinamide riboside
170	Nicotinamide ribotide
184	Hypoxanthine
185	Adenosine monophosphate
187	Adenine
190	Xanthine
193	Guanine
194	Adenosine diphosphate
195	Adenosine triphosphate
198	Cyclic adenosine monophosphate
204	Ureidosuccinic acid
205	4,5-Dihydroorotic acid
206	Orotic acid
209	Uridine
210	Uracil
213	Thymine
214	Cytidine monophosphate
215	Cytidine
216	Cytosine
217	Uridine diphosphate
226	Choline
227	Acetylcholine

231	Thiamine
232	Riboflavin
233	Pyridoxine
235	Biotin
236	Cholecalciferol
237	alpha-tocopherol
238	Phylloquinone
239	Flavin mononucleotide
241	Flavin adenine dinucleotide
246	Phosphocreatine
250	Glycerophosphocholine
254	Propionylcarnitine
257	Acetylcarnitine
264	Phosphoserine
274	Uric acid
276	Cholesteryl acetate
310	Indole
318	Glucosamine
322	Butyric acid
323	Glycerol 3-phosphate
326	Retinoic acid
501	Arachidonic acid
502	PC 15:0-18:1
503	TG 15:0-18:1-15:0
504	MG 18:1
505	Glucosyl cholesterol
506	Cardiolipin 18:1
508	DG 18:0-22:6
509	PS (POPS) 16:0-18:1
510	PG 16:0-18:1
511	PA 16:0-18:1
512	Lyso PI 17:1
513	PI 16:0-18:1
514	PI(5)P 18:1
515	PE 18:0-20:4
516	Lyso PC 15:0
517	Lyso PE 18:0
518	Lyso PA 18:1
519	Lyso PG 16:0
520	Lyso PS 17:1
521	GM3 Ganglioside 18:0
522	Cer d8:1/22:0 Cer
523	GlcCer d18:1/18:0
524	SM d18:1-16:0

525	Sulfo GalCer 18:0(2S-OH)
526	CAR 16:0
527	PC (O) C16-18:1
528	Prostaglandin E1
529	Cholesteryl ester 17:0
901	4-nitrobenzyl pyridinium chloride
902	4-methoxybenzyl pyridinium chloride
903	4-methylbenzyl pyridinium chloride
904	4-cyanobenzyl pyridinium chloride
905	4-chlorobenzyl pyridinium chloride
906-908	Rhodamine B + Caffeine (guide spots)
900	Blank

Appendix G

Technical drawing of the mould for casting polymer templates for microwell preparation (all units in mm)



Appendix H

Differentially abundant metabolites in each metabolite groups. Some putative annotations of drugs were removed as irrelevant.

group	m/z	putative ion	putative annotations
up in cecum 1	107.036	C4H6N4O-H2O-H-	5-Aminoimidazole-4-carboxamide
up in cecum 1	110.036	C4H5N3O-H-	Cytosine
up in cecum 1	119.036	C5H4N4-H-	Purine
up in cecum 1	129.102	C6H14N2O2-H2O+H+	Lysine
up in cecum 1	131.118	C6H14N2O+H+	N-Acetylputrescine
up in cecum 1	133.016	C5H4N4O2-H2O-H-	Xanthine
up in cecum 1	143.118	C7H16N2O2-H2O+H+	Putreanine, N(6)-Methyllysine
up in cecum 1	146.165	C7H19N3+H+	Spermidine
up in cecum 1	158.046	C6H9NO4-H-	N-methyl-2-oxoglutaramic acid
up in cecum 1	159.077	C6H12N2O3-H-	Alanyl-Alanine
up in cecum 1	160.181	C8H21N3+H+	Aminopropylcadaverine, 4,4-Diaminodibutylamine
up in cecum 1	166.072	C6H7N5O+H+	N2-Methylguanine, 7-Methylguanine
up in cecum 1	168.067	C8H11NO3-H-	Pyridoxine, 5-Hydroxydopamine, Norepinephrine,
up in cecum 1	168.077	C7H9N3O2+H+	(S)-Spinacine
up in cecum 1	171.006	C3H9O6P-H-	Glycerol 3-phosphate, Beta-Glycerophosphoric ac
up in cecum 1	174.077	C7H13NO4-H-	N-Carboxyethyl-g-aminobutyric acid, hydroxyvaler
up in cecum 1	184.121	C7H19N3+K+	Spermidine
up in cecum 1	185.069	C6H14N2O2+K+	L-Lysine
up in cecum 1	185.128	C9H16N2O2+H+	N-(3-acetamidopropyl)pyrrolidin-2-one
up in cecum 1	188.176	C9H21N3O+H+	N1-Acetylspermidine, N8-Acetylspermidine
up in cecum 1	196.073	C8H11N3O3-H-	N-Acetylhistidine
up in cecum 1	197.092	C9H12N2O3+H+	5-Nitro-2-propoxyaniline, 2-(Ethylamino)-4,5-dihyd
up in cecum 1	200.068	C7H13N3O5-H2O-H-	Serylasparagine, Asparaginyl-Serine
up in cecum 1	201.088	C8H14N2O4-H-	Prolyl-Serine, Serylproline
up in cecum 1	203.150	C8H18N4O2+H+	Asymmetric dimethylarginine, Symmetric dimethyl
up in cecum 1	203.223	C10H26N4+H+	Spermine
up in cecum 1	205.155	C9H20N2O3+H+	3-Hydroxy-N6,N6,N6-trimethyl-L-lysine
up in cecum 1	210.088	C9H15N3O4-H2O-H-	Prolyl-Asparagine, Asparaginyl-Proline
up in cecum 1	214.119	C9H17N3O4-H2O+H+	Asparaginyl-Valine
up in cecum 1	223.972	C3H8NO6P+K+	Phosphoserine
up in cecum 1	227.033	C6H13O7P-H-	Mevalonate-P
up in cecum 1	232.082	C10H19NO2S2-H2O+H+	S-Acetyldihydrolipoamide, S-Acetyldihydrolipoami
up in cecum 1	240.135	C11H21N3O4-H2O-H-	Glutaminylleucine, Leucyl-Glutamine, Glutaminylis
up in cecum 1	241.108	C12H20O6-H2O-H-	Glycerol tripropanoate
up in cecum 1	248.067	C11H11N3O5-H2O+H+	Isoniazid alpha-ketoglutaric acid
up in cecum 1	250.093	C10H19NO2S2+H+	S-Acetyldihydrolipoamide, S-Acetyldihydrolipoami
up in cecum 1	250.093	C10H13N5O4-H2O+H+	Deoxyguanosine, Adenosine
up in cecum 1	255.110	C10H18N4O5-H2O-H-	Glutaminylglutamine
up in cecum 1	264.087	C13H15NO6-H2O+H+	4-Hydroxyphenylacetylglutamine

up in cecum 1	264.087	C14H17NS2+H+	Dimethylthiambutene
up in cecum 1	341.269	C20H36O4+H+	PGE1 alcohol, 11-deoxy-PGF1b, 11-deoxy-PGF1a
up in cecum 1	341.269	C21H36NO+Na+	Tridihexethyl
up in cecum 1	341.269	C18H40NO2+K+	sphinganine(1+)
up in cecum 1	357.279	C24H38O3-H2O+H+	7-Oxo-5alpha-cholan-24-oic Acid, 3alpha-Hydroxy
up in cecum 1	434.191	C17H34NO8P+Na+	PE(6:0/6:0)
up in cecum 1	439.234	C23H38O9-H2O-H-	6-Gingerdiol 5-O-beta-D-glucopyranoside, 6-Ginge
up in cecum 1	469.257	C21H43O9P-H-	PG(15:0/0:0)
up in cecum 1	469.260	C28H40O7-H2O-H-	Hydrocortisone butyrate propionate
up in cecum 1	483.273	C22H45O9P-H-	PG(16:0/0:0)
up in cecum 1	497.252	C22H43O10P-H-	PG(8:0/8:0)
up in cecum 1	509.288	C24H47O9P-H-	PG(18:1(9Z)/0:0)
up in cecum 1	537.283	C25H49O11P-H2O-H-	PI(P-16:0/0:0)
up in cecum 1	539.299	C25H51O11P-H2O-H-	PI(O-16:0/0:0)
up in cecum 1	553.315	C26H51O10P-H-	PG(10:0/10:0)
up in cecum 1	561.202	C28H39ClO7+K+	4-Deoxyphysalolactone
up in cecum 1	562.241	C30H37NO8+Na+	5,9,11-trihydroxyprosta-6E,14Z-dien-1-oate
up in cecum 1	567.294	C26H51O12P-H2O-H-	PI(17:0/0:0)
up in cecum 1	567.296	C33H44O8-H-	Helvolic acid
up in cecum 1	573.221	C30H36N2O9-H2O+Na+	Indacaterol-8-O-glucuronide
up in cecum 1	581.310	C27H53O12P-H2O-H-	6-O-(1-O-stearoyl-sn-glycero-3-phosphono)-1D-m
up in cecum 1	595.325	C28H55O12P-H2O-H-	PI(19:0/0:0)
up in cecum 1	605.308	C29H51O12P-H2O+H+	PI(20:3(8Z,11Z,14Z)/0:0)
up in cecum 1	643.264	C29H51O12P-H2O+K+	PI(20:3(8Z,11Z,14Z)/0:0)
up in cecum 1	691.492	C37H73O9P-H-	PG(P-18:0/13:0), PG(O-16:0/15:1(9Z)), PG(P-16:0
up in cecum 1	691.548	C40H78O8-H2O+Na+	1-O-alpha-D-glucopyranosyl-(2-tetradecanoyloxy)-
up in cecum 1	705.508	C38H75O9P-H-	PG(P-20:0/12:0), PG(P-16:0/16:0), PG(O-16:0/16:
up in cecum 1	707.487	C37H73O10P-H-	PG(16:0/15:0), PG(18:0/13:0), PG(17:0/14:0), PG(
up in cecum 1	710.546	C39H80NO7P-H2O+Na+	PC(O-18:0/13:0), PE(O-16:0/18:0), PE(O-18:0/16:
up in cecum 1	719.523	C39H77O9P-H-	PG(P-16:0/17:0), PG(O-16:0/17:1(9Z)), PG(O-18:0
up in cecum 1	721.502	C38H75O10P-H-	PG(13:0/19:0), PG(15:0/17:0), PG(20:0/12:0), PG(
up in cecum 1	729.508	C40H77O10P-H2O-H-	LBPA(16:0/18:1(9Z)), PG(18:1(9Z)/16:0), PG(20:0
up in cecum 1	731.523	C40H79O10P-H2O-H-	PG(12:0/22:0), PG(13:0/21:0), PG(14:0/20:0), PG(
up in cecum 1	733.502	C39H75O10P-H-	PG(13:0/20:1(11Z)), PG(20:1(11Z)/13:0), PG(18:0
up in cecum 1	735.518	C39H77O10P-H-	PG(12:0/21:0), PG(15:0/18:0), PG(19:0/14:0), PG(
up in cecum 1	749.497	C39H77O12P-H2O-H-	PI(O-18:0/12:0), PI(O-16:0/14:0)
up in cecum 1	757.539	C42H81O10P-H2O-H-	PG(16:0/20:1(11Z)), PG(21:0/15:1(9Z)), PG(17:1(9
up in cecum 1	761.534	C41H79O10P-H-	PG(16:0/18:0(11Cp)), PG(18:0/17:1(9Z)), PG(13:0
up in cecum 1	773.534	C42H79O10P-H-	LBPA(18:1(9Z)/18:1(9Z)), PG(18:1(9Z)/18:1(9Z)),
up in cecum 1	787.549	C43H81O10P-H-	PG(20:2(11Z,14Z)/17:0), PG(20:1(11Z)/17:1(9Z)),
up in cecum 1	793.478	C45H71O8P+Na+	PA(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/20:3(8Z,11Z,14Z
up in cecum 1	793.478	C42H75O9P+K+	PG(P-16:0/20:4(5Z,8Z,11Z,14Z)), PG(O-16:0/20:5

up in cecum 1	801.565	C44H83O10P-H-	PG(21:0/17:2(9Z,12Z)), PG(20:2(11Z,14Z)/18:0), F
up in cecum 1	809.452	C45H71O8P+K+	PA(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/20:3(8Z,11Z,14Z
up in cecum 1	809.473	C45H73O10P-H2O+Na+	PG(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/17:2(9Z,12Z)), F
up in cecum 1	809.473	C40H76O13P2-H2O+H+	PGP(18:2(9Z,12Z)/16:0), PGP(18:1(9Z)/16:1(9Z)),
up in cecum 1	809.473	C42H75O10P+K+	PG(18:2(9Z,12Z)/18:2(9Z,12Z)), PG(18:0/18:4(6Z,
up in cecum 1	817.560	C44H85O12P-H2O-H-	PI(P-18:0/17:0), PI(P-16:0/19:0), PI(O-20:0/15:1(9
up in cecum 1	825.447	C45H73O10P-H2O+K+	PG(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/17:2(9Z,12Z)), F
up in cecum 1	835.488	C42H78O13P2-H2O+H+	PGP(18:3(9Z,12Z,15Z)/18:0), PGP(18:3(6Z,9Z,12Z
up in cecum 1	835.489	C44H77O10P+K+	PG(18:1(9Z)/20:4(5Z,8Z,11Z,14Z)), PG(20:5(5Z,8Z
up in cecum 1	863.517	C42H84O13P2-H2O+Na+	PGP(i-12:0/i-24:0), PGP(18:0/18:0)
up in cecum 1	863.520	C44H82O13P2-H2O+H+	PGP(18:0/20:3(8Z,11Z,14Z)), PGP(18:0/20:3(5Z,8
up in cecum 1	863.520	C46H81O10P+K+	PG(20:5(5Z,8Z,11Z,14Z,17Z)/20:0), PG(20:2(11Z,
up in cecum 1	971.481	C50H78O13P2+Na+	PGP(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/22:5(7Z,10Z,1
up in cecum 1	971.485	C48H76O21-H2O+H+	1alpha,3beta,22R-Trihydroxyergosta-5,24E-dien-2
up in cecum 2	100.039	C4H7NO3-H2O+H+	2S-amino-3-oxo-butanoic acid
up in cecum 2	112.050	C4H5N3O+H+	Cytosine
up in cecum 2	122.096	C8H11N+H+	Phenylethylamine
up in cecum 2	124.039	C6H7NO3-H2O+H+	2-aminomuconic acid semialdehyde
up in cecum 2	124.087	C6H11N3O-H2O+H+	L-Histidinol
up in cecum 2	125.071	C6H8N2O+H+	2-Methoxy-3-methylpyrazine, Methylimidazole ace
up in cecum 2	127.087	C6H12N2O2-H2O+H+	Adipamide
up in cecum 2	128.071	C6H11NO3-H2O+H+	3-oxo-5S-amino-hexanoic acid
up in cecum 2	132.032	C5H5N5O-H2O-H-	2-Hydroxyadenine, Guanine
up in cecum 2	132.081	C9H9N+H+	skatole
up in cecum 2	134.032	C4H5N3O+Na+	Cytosine
up in cecum 2	136.062	C5H5N5+H+	Adenine
up in cecum 2	138.091	C8H11NO+H+	Tyramine
up in cecum 2	144.048	C6H11NO2S-H2O+H+	trans-S-(1-Propenyl)-L-cysteine
up in cecum 2	146.060	C9H7NO+H+	indolealdehyde
up in cecum 2	150.006	C4H5N3O+K+	Cytosine
up in cecum 2	150.042	C5H5N5O-H-	2-Hydroxyadenine, Guanine
up in cecum 2	152.057	C5H5N5O+H+	2-Hydroxyadenine, Guanine
up in cecum 2	154.086	C8H13NO3-H2O+H+	N-butanoyl-lhomoserine lactone
up in cecum 2	155.081	C7H12N2O3-H2O+H+	Prolylglycine, Glycylproline
up in cecum 2	157.108	C6H14N4O2-H2O+H+	D-Arginine, L-Arginine
up in cecum 2	158.044	C5H7N5O-H2O+Na+	FAPy-adenine
up in cecum 2	158.081	C7H13NO4-H2O+H+	N-Carboxyethyl-g-aminobutyric acid, hydroxyvaler
up in cecum 2	162.055	C9H7NO2+H+	2-Indolecarboxylic acid, Quinoline-4,8-diol, 4,6-Dih
up in cecum 2	164.071	C9H11NO3-H2O+H+	o-Tyrosine, L-Tyrosine
up in cecum 2	166.050	C8H9NO4-H2O+H+	4-Pyridoxic acid
up in cecum 2	168.065	C8H9NO3+H+	Pyridoxal
up in cecum 2	168.066	C6H15N2O2-H2O+K+	L-Lysine

up in cecum 2	168.147	C7H19N3+Na+	Spermidine
up in cecum 2	169.097	C8H12N2O2+H+	Pyridoxamine
up in cecum 2	171.113	C8H16N2O3-H2O+H+	Glycyl-Isoleucine, Glycylleucine, N-Alpha-acetyllys
up in cecum 2	174.018	C5H5N5+K+	Adenine
up in cecum 2	174.039	C5H5N5O+Na+	2-Hydroxyadenine, Guanine
up in cecum 2	174.112	C8H15NO3+H+	Hexanoylglycine, N-Acetylisoleucine, N-Acetylleuc
up in cecum 2	174.124	C7H15N3O2+H+	apo-3-methylcrotonoyl-CoA:carbon-dioxide ligase
up in cecum 2	184.060	C8H9NO4+H+	4-Pyridoxic acid
up in cecum 2	186.076	C6H15N2O2+K+	L-Lysine
up in cecum 2	188.092	C8H13NO4+H+	N-(3S-hydroxy-butanoyl)-homoserine lactone
up in cecum 2	189.160	C9H20N2O2+H+	N6,N6,N6-Trimethyl-L-lysine
up in cecum 2	190.013	C5H5N5O+K+	2-Hydroxyadenine, Guanine
up in cecum 2	190.086	C11H11NO2+H+	Indole-3-methyl acetate, Indole-3-propionic acid
up in cecum 2	192.065	C10H11NO4-H2O+H+	Hydroxyphenylacetylglycine
up in cecum 2	198.085	C6H13N3O3+Na+	Argininic acid, Citrulline
up in cecum 2	206.081	C11H11NO3+H+	5-Methoxyindoleacetate, Indolelactic acid, 3-Indole
up in cecum 2	208.060	C10H9NO4+H+	4-(2-Aminophenyl)-2,4-dioxobutanoic acid
up in cecum 2	208.061	C8H13N2O2+K+	Pyridoxamine
up in cecum 2	209.128	C11H16N2O2+H+	Descarbonyl-lacosamide
up in cecum 2	210.076	C10H11NO4+H+	Hydroxyphenylacetylglycine
up in cecum 2	214.047	C10H9NO3+Na+	5-Hydroxyindoleacetic acid, 2-Oxindole-3-acetate
up in cecum 2	223.144	C12H18N2O2+H+	3-Hydroxymonoethylglycinexylidide
up in cecum 2	226.047	C11H9NO3+Na+	Indolepyruvate
up in cecum 2	226.047	C6H14NO7P-H2O+H+	1-alkyl-2-acylglycerophosphoethanolamine
up in cecum 2	226.048	C8H13NO4+K+	N-(3S-hydroxy-butanoyl)-homoserine lactone
up in cecum 2	226.082	C9H13N3O5-H2O+H+	Cytidine
up in cecum 2	227.139	C11H20N2O4-H2O+H+	Leucyl-Hydroxyproline, Isoleucyl-Hydroxyproline, H
up in cecum 2	228.145	C9H19N5O3-H2O+H+	Alanyl-Arginine, Arginyl-Alanine
up in cecum 2	230.042	C10H9NO4+Na+	4-(2-Aminophenyl)-2,4-dioxobutanoic acid
up in cecum 2	230.042	C5H12NO7P+H+	5-Phosphoribosylamine
up in cecum 2	230.042	C7H13NO5+K+	2-amino-2,3,7-trideoxy-D-lyxo-hept-6-ulosonic acid
up in cecum 2	237.160	C13H20N2O2+H+	N-(4-aminobutyl)-3-(4-hydroxyphenyl)propanimidic
up in cecum 2	259.022	C6H13O9P-H-	Beta-D-Glucose 6-phosphate, D-Mannose 1-phosp
up in cecum 2	283.972	C8H8NO6P+K+	Pyridoxal 5-phosphate
up in cecum 2	292.082	C14H15NO7-H2O+H+	Indoxyl glucuronide
up in cecum 2	391.285	C24H40O4-H-	3alpha,15alpha-dihydroxy-5beta-cholan-24-oic aci
up in cecum 2	438.265	C27H39NO5-H2O-H-	N-docosahexaenoyl glutamic acid
up in cecum 2	455.137	C21H26O9S+H+	{10-butanoyl-5-hydroxy-2,2-dimethyl-8-oxo-6-prop
up in cecum 2	517.098	C24H22O14-H2O+H+	5,4-Dihydroxy-3,3-dimethoxy-6,7-methylenedioxyf
up in cecum 2	570.546	C35H71NO4+H+	Cer(d18:0/h17:0)
up in cecum 2	603.451	C32H65N2O6P-H-	PE-Cer(d14:1(4E)/16:0)
up in cecum 2	611.537	C39H72O3+Na+	1-hexadecanyl-2-(8-3-ladderane-octanyl)-sn-glyce

up in cecum 2	625.553	C40H74O3+Na+	1-(10-methyl-hexadecanyl-2-(8-3-ladderane-octan
up in cecum 2	648.461	C34H68NO8P-H-	PE(13:0/16:0), PE(12:0/17:0), PE(14:0/15:0), PE(1
up in cecum 2	662.477	C35H70NO8P-H-	PE-NMe2(14:0/14:0), PC(15:0/12:0), PC(14:0/13:0
up in cecum 2	672.460	C36H66NO8P+H+	PE(14:1(9Z)/17:2(9Z,12Z)), PE(13:0/18:3(9Z,12Z,
up in cecum 2	676.492	C36H72NO8P-H-	PE(16:0/15:0), PC(9:0/19:0), PC(8:0/20:0), PC(7:0
up in cecum 2	677.512	C39H74O6+K+	Trilauroyl-glycerol
up in cecum 2	678.507	C36H72NO8P+H+	PE(16:0/15:0), PC(9:0/19:0), PC(8:0/20:0), PC(7:0
up in cecum 2	679.512	C38H74O9-H2O+Na+	1-(O-alpha-D-glucopyranosyl)-29-keto-(1,3R,31R)-
up in cecum 2	686.473	C35H70NO8P+Na+	PE-NMe2(14:0/14:0), PC(15:0/12:0), PC(14:0/13:0
up in cecum 2	686.475	C37H70NO9P-H2O+H+	PS(P-16:0/15:1(9Z))
up in cecum 2	688.431	C37H66NO8P-H2O+Na+	PE(14:0/18:4(6Z,9Z,12Z,15Z)), PE(18:3(9Z,12Z,15
up in cecum 2	688.431	C34H68NO8P+K+	PE(13:0/16:0), PE(12:0/17:0), PE(14:0/15:0), PE(1
up in cecum 2	693.483	C38H73O8P-H2O+Na+	PA(17:1(9Z)/18:0), PA(16:1(9Z)/19:0), PA(21:0/14
up in cecum 2	693.485	C45H68O5-H2O+Na+	DG(20:4(5Z,8Z,11Z,14Z)/22:6(4Z,7Z,10Z,13Z,16Z
up in cecum 2	693.485	C40H71O8P-H2O+H+	PA(17:2(9Z,12Z)/20:2(11Z,14Z)), PA(22:4(7Z,10Z,
up in cecum 2	693.485	C42H70O5+K+	DG(19:1(9Z)/20:5(5Z,8Z,11Z,14Z,17Z)/0:0)iso2, D
up in cecum 2	700.492	C38H74NO9P-H2O-H-	PS(P-20:0/12:0), PS(O-18:0/14:1(9Z)), PS(O-16:0/
up in cecum 2	702.507	C38H72NO8P+H+	PE(13:0/20:2(11Z,14Z)), PC(13:0/17:2(9Z,12Z)), P
up in cecum 2	703.452	C35H69O10P+Na+	PG(13:0/16:0), PG(12:0/17:0), PG(17:0/12:0), PG(
up in cecum 2	716.463	C39H70NO8P-H2O+Na+	PE(14:0/20:4(5Z,8Z,11Z,14Z)), PE(22:4(7Z,10Z,13
up in cecum 2	716.463	C36H72NO8P+K+	PE(16:0/15:0), PC(9:0/19:0), PC(8:0/20:0), PC(7:0
up in cecum 2	716.465	C41H68NO8P-H2O+H+	PE(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/14:1(9Z)), PE(18
up in cecum 2	724.431	C37H70NO9P-H2O+K+	PS(P-16:0/15:1(9Z))
up in cecum 2	738.447	C41H68NO8P-H2O+Na+	PE(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/14:1(9Z)), PE(18
up in cecum 2	738.447	C38H72NO9P-H2O+K+	PS(P-18:0/14:1(9Z)), PS(P-16:0/16:1(9Z))
up in cecum 2	740.465	C43H68NO8P-H2O+H+	PE(18:4(6Z,9Z,12Z,15Z)/20:5(5Z,8Z,11Z,14Z,17Z)
up in cecum 2	754.421	C41H68NO8P-H2O+K+	PE(22:6(4Z,7Z,10Z,13Z,16Z,19Z)/14:1(9Z)), PE(18
up in cecum 2	778.524	C40H80NO12P-H2O-H-	PI-Cer(t18:0/16:0), PI-Cer(t20:0/14:0), PI-Cer(d18:
up in jejunum	157.037	C4H6N4O3-H-	Allantoin
up in jejunum	181.072	C6H14O6-H-	Galactitol
up in jejunum	205.068	C6H14O6+Na+	Galactitol
up in jejunum	217.048	C6H14O6+Cl-	Galactitol
up in jejunum	218.103	C9H17NO5-H-	Pantothenic acid
up in jejunum	220.118	C9H17NO5+H+	Pantothenic acid
up in jejunum	307.151	C12H24N2O7-H-	1-(5-Amino-5-carboxypentyl)amino-1-deoxyfructos
up in jejunum	309.166	C12H24N2O7+H+	1-(5-Amino-5-carboxypentyl)amino-1-deoxyfructos
up in jejunum	331.148	C12H24N2O7+Na+	1-(5-Amino-5-carboxypentyl)amino-1-deoxyfructos
up in jejunum	341.143	C12H24N4O7-H2O+Na+	N2-Fructopyranosylarginine
up in jejunum	438.130	C20H18F7NO2+H+	(2S,3R)-2-(1R)-1-3,5-bis(trifluoromethyl)phenyleth
up in jejunum	438.239	C25H31N3O4+H+	N1,N10-Dicoumaroylspermidine
up in jejunum	460.221	C25H31N3O4+Na+	N1,N10-Dicoumaroylspermidine
up in jejunum	460.222	C20H42NO7P-H2O+K+	PE(15:0/0:0), PC(12:0/0:0), LysoPE(0:0/15:0)

up in jejunum	476.195	C25H31N3O4+K+	N1,N10-Dicoumaroylspermidine
up in jejunum	497.148	C17H30O15+Na+	a-L-Arabinofuranosyl-(1->2)-a-D-mannopyranosyl-
up in jejunum	521.128	C18H32O16-H2O+Cl-	trisaccharide
up in jejunum	527.158	C24H32O12S-H2O+H+	Estriol 3-sulfate 16-glucuronide
up in jejunum	527.158	C18H32O16+Na+	Maltotriose, Raffinose
up in jejunum	539.138	C18H32O16+Cl-	Maltotriose, Raffinose
up in jejunum	544.136	C23H29N3O10S-H2O+Na+	2-amino-4-({1-(carboxymethyl)-C-hydroxycarbonin
up in jejunum	665.178	C36H36O11-H2O+K+	4-O-Methylneobavaisoflavone 7-O-(2-p-coumaroy
up in jejunum	665.215	C24H42O21-H-	Glycogen, Maltotetraose, Stachyose
up in jejunum	667.205	C38H34N2O9-H2O+Na+	9-(6,11-dihydroxy-2,2,5-trimethyl-10-oxo-3,4,5,10-
up in jejunum	689.211	C24H42O21+Na+	Glycogen, Maltotetraose, Stachyose
up in jejunum	691.216	C36H36N4O8+K+	Dehydroisocoproporphyrinogen
up in jejunum	701.191	C24H42O21+Cl-	Glycogen, Maltotetraose, Stachyose
up in jejunum	703.188	C33H38O18-H2O-H-	6-(3-Hydroxy-3-methylglutaroyl)isoviolanthin
up in jejunum	759.403	C46H60O7+Cl-	Bis(4-methoxybenzoyl)-3a,29-dihydroxy-8-multiflo
up in jejunum	841.201	C34H42O23+Na+	Laricitrin 3,7,5-triglucoside
up in jejunum	849.227	C30H52O26-H2O+K+	Maltopentaose
up in jejunum	851.264	C30H52O26+Na+	Maltopentaose
up in jejunum	863.244	C30H52O26+Cl-	Maltopentaose
up in jejunum	867.238	C30H52O26+K+	Maltopentaose
up in jejunum	948.476	C44H77N3O15P2-H-	CDP-DG(i-14:0/18:2(9Z,11Z)), CDP-DG(18:2(9Z,1
up in jejunum	960.237	C33H52N7O18P3S+H+	(+)-7-Isojasmonic acid CoA, 3(S)-3-hydroxydodec
up in jejunum	965.513	C45H86O17P2-H2O+Na+	CL(8:0/8:0/10:0/10:0), CL(8:0/8:0/8:0/i-12:0), CL(8
up in jejunum	966.509	C44H69N15O9S-H2O+H+	Adrenorphin
up in jejunum	977.217	C41H48O27-H2O+Na+	Malvidin 3-O-(6-O-(4-O-malonyl-alpha-rhamnopyr
up in jejunum	989.320	C36H62O31-H-	Maltohexaose
up in jejunum	995.306	C36H62O31-H2O+Na+	Maltohexaose
up in jejunum	996.308	C36H64N7O17P3S-H2O+Na+	isopentadecanoyl-CoA
up in jejunum	1012.282	C36H64N7O17P3S-H2O+K+	isopentadecanoyl-CoA
up in jejunum	1013.317	C36H62O31+Na+	Maltohexaose
up in jejunum	1014.318	C36H64N7O17P3S+Na+	isopentadecanoyl-CoA
up in jejunum	1014.320	C40H53N11O18+K+	5-Methyltetrahydropteroylpentaglutamate
up in jejunum	1014.321	C38H62N7O17P3S+H+	10Z-heptadecenoyl-CoA
up in jejunum	1029.291	C36H62O31+K+	Maltohexaose
up in jejunum	1030.292	C39H62N7O17P3S-H2O+Na+	Stearidonoyl-CoA
up in jejunum	1030.292	C36H64N7O17P3S+K+	isopentadecanoyl-CoA
up in jejunum	1175.370	C42H72O36+Na+	Maltoheptaose
up in ileum	290.146	C10H19N5O5+H+	Aspartyl-Arginine, Arginyl-Aspartic acid
up in ileum	352.042	C10H12N5O6P+Na+	Adenosine 2,3-cyclic phosphate, Cyclic AMP
up in ileum	377.144	C16H26O11-H2O+H+	1-(3-Methyl-2-butenoyl)-6-apiosylglucose
up in ileum	433.235	C26H36O5-H2O+Na+	2-Angeloyl-9-(3-methyl-2E-pentenoyl)-2b,9a-dihyd
up in ileum	433.235	C21H37O7P+H+	PA(18:3(6Z,9Z,12Z)/0:0), PA(18:3(9Z,12Z,15Z)/0:(

up in ileum	433.235	C23H38O5+K+	Nor-omega-muricholic acid, Nor-beta-muricholic a
up in ileum	463.184	C24H30N2O7-H2O+Na+	2-Hydroxy-desipramine glucuronide
up in ileum	463.186	C19H37O8P+K+	PA(8:0/8:0)
up in ileum	471.191	C21H37O7P+K+	1-(9Z,12Z,15Z-octadecatrienoyl)-glycero-3-phosph
up in ileum	479.158	C24H30N2O7-H2O+K+	2-Hydroxy-desipramine glucuronide
up in ileum	501.124	C21H26O15-H2O+H+	3,4,5-trihydroxy-6-{2-hydroxy-5-(1E)-3-oxo-3-(3,4,5
up in ileum	545.129	C26H26O14-H2O+H+	3,4,5-trihydroxy-6-{5-hydroxy-4-oxo-2-phenyl-8-(3,
up in ileum	653.115	C26H30O16S+Na+	{2-(2-{4-3-(2,4-dihydroxyphenyl)-3-oxoprop-1-en-1
up in ileum	663.177	C27H36O20-H2O+H+	6-{5-(1E)-3-(6-{3,4-dihydroxy-2,5-bis(hydroxymeth
up in ileum	753.254	C28H48N2O21-H2O+Na+	O-b-D-Gal-(1->3)-O-O-b-D-Gal-(1->4)-2-(acetylam
up in ileum	853.270	C46H44O16+H+	6-2-(5-{3-(2E)-3-(2,4-dihydroxyphenyl)prop-2-enoy
up in ileum	961.247	C56H42O14+Na+	2-3-(3,5-dihydroxyphenyl)-6-hydroxy-2-(4-hydroxy
up in ileum	977.221	C56H42O14+K+	2-3-(3,5-dihydroxyphenyl)-6-hydroxy-2-(4-hydroxy
up in ileum	996.310	C40H53N11O18-H2O+K+	5-Methyltetrahydropteroylpentaglutamate
up in ileum	996.310	C38H62N7O17P3S-H2O+H+	10Z-heptadecenoyl-CoA
up in ileum	1030.295	C41H60N7O17P3S-H2O+H+	timnodonyl coenzyme A, timnodonyl coenzyme A,
up in ileum	1157.359	C42H72O36-H2O+Na+	Maltoheptaose
up in ileum	1191.343	C42H72O36+K+	Maltoheptaose
up in colon	149.046	C5H10O5-H-	2-deoxy-ribonic acid, 2-Deoxyribonic acid
up in colon	181.051	C9H10O4-H-	3-(3-Hydroxyphenyl)-3-hydroxypropanoic acid, Ho
up in colon	181.051	C5H13NO4P-H-	choline phosphate(1-), choline phosphate(1-), chol
up in colon	193.051	C10H10O4-H-	2E,4Z,6Z,8Z-Decatetraenedioic acid, 2E,4E,6Z,8Z
up in colon	195.066	C10H12O4-H-	3-(3-hydroxy-2-methoxyphenyl)propanoic acid, 3-h
up in colon	207.066	C11H14O5-H2O-H-	3-(2-hydroxy-3,4-dimethoxyphenyl)propanoic acid,
up in colon	209.046	C10H10O5-H-	Vanilpyruvic acid
up in colon	211.061	C10H12O5-H-	Vanillactic acid
up in colon	269.066	C12H12O7+H+	(2Z)-3-(6,7-dimethoxy-2H-1,3-benzodioxol-5-yl)-2-
up in colon	271.061	C15H12O5-H-	2-Hydroxydihydrodaidzein, Licodione, (-)-Glycinol,
up in colon	287.104	C15H16N2O4-H-	3-Carboxy-2,3,4,9-tetrahydro-1H-pyrido3,4-bindole
up in colon	309.083	C11H18O10-H-	3,4,5-trihydroxy-6-{3-hydroxy-2-(hydroxymethyl)-2-
up in colon	391.066	C18H16O11-H2O+H+	3,4,5-trihydroxy-6-({8-methoxy-2-oxo-2H-furo2,3-h
up in colon	393.082	C18H16O10+H+	5,7,3,4,5-Pentahydroxy-3,6,8-trimethoxyflavone
up in colon	395.389	C26H52O2-H-	26:0(18Me,24Me), 26:0(18Me), Phthioic acid, 4-M
up in colon	417.082	C20H18O11-H2O+H+	Quercetin 3-xyloside, 6-Hydroxyluteolin 7-xyloside
up in colon	433.076	C20H18O12-H2O+H+	Myricetin 3-xyloside, Myricetin 3-alpha-L-arabinofu
up in colon	451.087	C20H18O12+H+	Myricetin 3-xyloside, Myricetin 3-alpha-L-arabinofu
up in colon	495.381	C31H52O3+Na+	26,27-diethyl-1alpha,25-dihydroxyvitamin D3 / 26,2
up in colon	495.381	C28H56O4+K+	MG(a-25:0/0:0/0:0)rac, MG(0:0/a-25:0/0:0)rac
up in colon	495.383	C33H52O4-H2O+H+	Methyl 3b-hydroxy-13(18)-oleanen-28-oate
up in colon	519.113	C24H22O13+H+	5,6,7-Trihydroxyflavone 7-(6-malonylglucoside), A
up in colon	569.361	C31H57O8P-H2O-H-	PA(14:1(9Z)/14:1(9Z))
up in colon	571.108	C27H22O14+H+	Kaempferol 3-(2-galloyl-alpha-L-arabinopyranosid

up in colon	591.134	C27H28O16-H2O+H+	Kaempferol 3-rhamnoside-7-galacturonide, Luteoli
up in colon	595.129	C26H28O17-H2O+H+	Myricetin 3-sambubioside
up in colon	621.145	C28H30O17-H2O+H+	Herbacetin 7-(6-quinoylglucoside), Isoetin 7-glucos
up in colon	623.161	C28H30O16+H+	Kaempferol 3-(3-acetyl-alpha-L-arabinopyranosyl)-
up in colon	637.140	C28H28O17+H+	Acacetin 7-glucuronosyl-(1->2)-glucuronide, Luteo
up in colon	647.204	C24H42O21-H2O-H-	Glycogen, Maltotetraose, Stachyose
up in colon	653.171	C29H34O18-H2O+H+	Limocitrol 3-alpha-L-arabinopyranosyl-(1->3)-galac
up in colon	663.156	C30H32O18-H2O+H+	Kaempferol 3-(6G-malonylneohesperidoside), Lute
up in colon	665.171	C30H32O17+H+	Apigenin 7-(6-malonylneohesperidoside), Pelargon
up in colon	679.150	C30H32O19-H2O+H+	Quercetin 3-(6-malonylglucoside)-7-rhamnoside, C
up in colon	681.166	C30H34O19-H2O+H+	Malvidin 3-glucoside-5-(6-acetylglucoside)
up in colon	681.187	C27H36O20+H+	6-{5-(1E)-3-(6-{3,4-dihydroxy-2,5-bis(hydroxymeth
up in colon	699.177	C30H34O19+H+	Malvidin 3-glucoside-5-(6-acetylglucoside)
up in colon	701.215	C40H40O10-H2O+K+	2-{6-2,4-dihydroxy-5-methoxy-3-(3-methylbut-2-en
up in colon	711.177	C31H34O19+H+	8-Hydroxyluteolin 7-6-acethylallosyl-(1->2)-3-acety
up in colon	713.192	C31H36O19+H+	Syringetin 3-(6-acetylglucosyl)(1->3)-galactoside,
up in colon	717.210	C40H38O10+K+	3-(2,4-dihydroxyphenyl)-1-{3-5-(2,4-dihydroxypher
up in colon	725.192	C32H36O19+H+	8-Hydroxyluteolin 4-methyl ether 7-(6-acetylallosyl
up in colon	741.187	C32H38O21-H2O+H+	Quercetin 3-sambubioside-7-glucoside, Isoorientin
up in colon	743.203	C32H38O20+H+	Quercetin 3-xylosyl-(1->2)-rhamnosyl-(1->6)-gluco
up in colon	771.198	C33H38O21+H+	Kaempferol 3-rutinoside-7-glucuronide, 6-Hydroxy
up in colon	799.252	C30H45N6O16P+Na+	5,6,7,8-Tetrahydromethanopterin
up in colon	801.208	C34H40O22+H+	Delphinidin 3-(2-xylosylgalactoside)-5-(6-acetylglu
up in colon	801.208	C34H42O23-H2O+H+	Laricitrin 3,7,5-triglucoside
up in colon	809.257	C30H52O26-H2O-H-	Maltopentaose
up in colon	815.224	C35H44O23-H2O+H+	Limocitrol 3-alpha-L-arabinopyranosyl-(1->3)galac
up in colon	815.226	C30H45N6O16P+K+	5,6,7,8-Tetrahydromethanopterin
up in colon	819.219	C34H42O23+H+	Laricitrin 3,7,5-triglucoside
up in colon	833.235	C35H44O23+H+	Limocitrol 3-alpha-L-arabinopyranosyl-(1->3)galac
up in colon	835.260	C46H44O16-H2O+H+	6-2-(5-{3-(2E)-3-(2,4-dihydroxyphenyl)prop-2-enoy
up in colon	873.229	C37H44O24+H+	Kaempferol 3-neohesperidoside-7-(6-malonylgluco
up in colon	933.251	C39H48O26+H+	6-{2-(3,4-dihydroxyphenyl)-4-oxo-6,8-bis3,4,5-trihy
up in colon	943.236	C56H42O14-H2O+Na+	2-3-(3,5-dihydroxyphenyl)-6-hydroxy-2-(4-hydroxy
up in colon	945.252	C56H42O13+Na+	5-4-(1,2-dihydroxy-2-{3-6-hydroxy-3-(3-hydroxyph
up in colon	1011.280	C36H62O31-H2O+K+	Maltohexaose
down in duodenum	193.035	C6H10O7-H-	D-tagaturonic acid
down in duodenum	275.208	C12H26N4O3+H+	Lysyl-Lysine
down in duodenum	339.069	C17H19ClN2OS-H2O+Na+	hydroxychlorpromazine, Chlorpromazine-N-oxide
down in duodenum	339.070	C9H15N4O8P+H+	AICAR
down in duodenum	400.342	C23H45NO4+H+	O-palmitoylcarnitine, L-palmitoylcarnitine, Palmitoy
down in duodenum	412.210	C19H35NO6+K+	O-(11-carboxyundecanoyl)carnitine
down in duodenum	459.248	C21H41O7P+Na+	lysophosphatidic acid, PA(0:0/18:1(9Z)), PA(18:1(9

down in duodenum	461.264	C21H43O7P+Na+	PA(18:0/0:0), LPA(0:0/18:0)
down in duodenum	475.222	C21H41O7P+K+	lysophosphatidic acid, PA(0:0/18:1(9Z)), PA(18:1(
down in duodenum	476.275	C21H44NO7P+Na+	PE(16:0/0:0), PC(13:0/0:0), LysoPE(0:0/16:0)
down in duodenum	479.334	C26H48O6+Na+	Byrsonic acid
down in duodenum	483.248	C23H41O7P+Na+	PA(20:3(8Z,11Z,14Z)/0:0)
down in duodenum	485.264	C23H45O8P-H2O+Na+	PA(10:0/10:0)
down in duodenum	485.266	C25H41O7P+H+	LysoPA(22:5(7Z,10Z,13Z,16Z,19Z)/0:0), LysoPA(2
down in duodenum	487.279	C23H45O7P+Na+	PA(20:1(11Z)/0:0)
down in duodenum	494.324	C29H45NO4+Na+	Cervonyl carnitine
down in duodenum	494.324	C24H50NO8P-H2O+H+	PS(O-18:0/0:0)
down in duodenum	496.340	C29H47NO4+Na+	Clupanodonyl carnitine, Docosa-4,7,10,13,16-pent
down in duodenum	496.340	C24H50NO7P+H+	PE(19:0/0:0), PC(16:0/0:0), PC(O-14:0/2:0), PC(16
down in duodenum	502.290	C23H46NO7P+Na+	LysoPE(0:0/18:1(11Z)), LysoPE(18:1(11Z)/0:0), PI
down in duodenum	504.306	C23H48NO7P+Na+	PE(18:0/0:0), PC(15:0/0:0), LysoPE(0:0/18:0)
down in duodenum	516.306	C24H48NO7P+Na+	PE(19:1(9Z)/0:0), PC(16:1(9Z)/0:0), PC(16:1(9E)/0
down in duodenum	516.308	C26H46NO7P+H+	PC(18:4(6Z,9Z,12Z,15Z)/0:0), PC(18:4(9E,11E,13
down in duodenum	518.322	C24H50NO7P+Na+	PE(19:0/0:0), PC(16:0/0:0), PC(O-14:0/2:0), PC(16
down in duodenum	518.324	C26H50NO8P-H2O+H+	PC(16:1(9Z)/2:0)
down in duodenum	520.280	C26H46NO7P-H2O+Na+	PC(18:4(6Z,9Z,12Z,15Z)/0:0), PC(18:4(9E,11E,13
down in duodenum	520.280	C23H48NO7P+K+	PE(18:0/0:0), PC(15:0/0:0), LysoPE(0:0/18:0)
down in duodenum	520.340	C26H52NO8P-H2O+H+	PC(2:0/16:0), PS(P-20:0/0:0), PC(9:0/9:0), PC(16:
down in duodenum	521.345	C28H52O8-H2O+Na+	Menthol propylene glycol carbonate
down in duodenum	522.355	C26H52NO7P+H+	PC(P-16:0/2:0), PC(O-16:1(11Z)/2:0), PC(18:1(6Z
down in duodenum	524.371	C31H51NO4+Na+	tetracosapentaenoyl carnitine, tetracosapentaenoy
down in duodenum	524.371	C26H54NO7P+H+	PC(O-16:0/2:0), PC(18:0/0:0), PE(21:0/0:0), PC(0:
down in duodenum	526.290	C25H46NO7P+Na+	LysoPE(0:0/20:3(8Z,11Z,14Z)), LysoPE(0:0/20:3(5
down in duodenum	532.280	C27H46NO7P-H2O+Na+	LysoPE(22:5(7Z,10Z,13Z,16Z,19Z)/0:0), LysoPE(2
down in duodenum	532.280	C24H50NO8P-H2O+K+	PS(O-18:0/0:0)
down in duodenum	534.295	C27H48NO7P-H2O+Na+	LysoPE(0:0/22:4(7Z,10Z,13Z,16Z)), PE(22:4(7Z,1(
down in duodenum	534.296	C24H50NO7P+K+	PE(19:0/0:0), PC(16:0/0:0), PC(O-14:0/2:0), PC(16
down in duodenum	535.301	C24H49O9P+Na+	PG(18:0/0:0)
down in duodenum	539.319	C27H48O9+Na+	27-nor-campestan-3beta,4beta,5alpha,6alpha,7be
down in duodenum	542.322	C26H52NO8P-H2O+Na+	PC(2:0/16:0), PS(P-20:0/0:0), PC(9:0/9:0), PC(16:
down in duodenum	542.324	C28H48NO7P+H+	PC(20:5(5Z,8Z,11Z,14Z,17Z)/0:0)
down in duodenum	544.337	C26H54NO8P-H2O+Na+	PS(O-20:0/0:0)
down in duodenum	544.340	C28H50NO7P+H+	LysoPC(20:4(8Z,11Z,14Z,17Z)), PC(0:0/20:4(5Z,8
down in duodenum	546.353	C26H54NO7P+Na+	PC(O-16:0/2:0), PC(18:0/0:0), PE(21:0/0:0), PC(0:
down in duodenum	560.311	C26H52NO7P+K+	PC(P-16:0/2:0), PC(O-16:1(11Z)/2:0), PC(18:1(6Z
down in duodenum	595.264	C25H49O11P+K+	PI(P-16:0/0:0)
down in duodenum	595.285	C25H49O12P+Na+	PI(16:0/0:0), 1-hexadecanoyl-sn--glycero-3-phospl
down in duodenum	595.288	C27H47O12P+H+	PI(18:3(9Z,12Z,15Z)/0:0), PI(18:3(6Z,9Z,12Z)/0:0)
down in duodenum	657.361	C35H54O10+Na+	Pandaroside B

down in duodenum	681.538	C45H74O2+Cl-	CE(18:3(6Z,9Z,12Z)), CE(18:3(9Z,12Z,15Z))
down in duodenum	683.554	C45H76O2+Cl-	CE(18:2(9Z,12Z))
down in duodenum	715.579	C47H80O2+K+	18:2 Sitosteryl ester, 20:2 Cholesteryl ester, 18:1 S
down in duodenum	737.192	C33H38O20-H2O+H+	Kaempferol 3-6-(3-hydroxy-3-methylglutaryl)glucos
down in duodenum	755.182	C36H36O19-H2O+H+	Kaempferol 3-(6-caffeylglucosyl)-(1->2)-galactosid
down in duodenum	796.546	C42H80NO9P+Na+	PS(P-20:0/16:1(9Z)), PS(O-18:0/18:2(9Z,12Z)), PS
down in duodenum	904.389	C39H71N3O15P2-H2O+K+	CDP-DG(i-15:0/i-12:0), CDP-DG(i-14:0/i-13:0), CD
down in duodenum	979.575	C49H89O18P-H2O+H+	PIM1(18:2(9Z,12Z)/16:0), PIM1(18:0/16:2(9Z,12Z)
down in duodenum	1056.602	C51H95N3O15P2-H2O+Na+	CDP-DG(i-22:0/i-17:0), CDP-DG(i-22:0/a-17:0), CD
down in duodenum	1072.576	C51H95N3O15P2-H2O+K+	CDP-DG(i-22:0/i-17:0), CDP-DG(i-22:0/a-17:0), CD
down in jejunum	116.071	C5H9NO2+H+	L-Proline, D-Proline
down in jejunum	132.102	C6H13NO2+H+	2-Amino-4-methylpentanoic acid, 2S-Amino-3S-me
down in jejunum	133.097	C5H12N2O2+H+	2,5-Diaminopentanoic acid
down in jejunum	134.045	C4H7NO4+H+	1-deoxy-1-(N6-lysino)-D-fructose, D-Aspartic acid,
down in jejunum	156.077	C6H9N3O2+H+	L-Histidine
down in jejunum	169.095	C6H14N2O2+Na+	L-Lysine, D-Lysine
down in jejunum	171.017	C4H8N2O3+K+	L-Asparagine, D-Asparagine, Glycyl-glycine, Ureid
down in jejunum	175.119	C6H14N4O2+H+	D-Arginine, L-Arginine
down in jejunum	182.081	C9H11NO3+H+	o-Tyrosine, L-Tyrosine
down in jejunum	187.108	C8H16N2O4-H2O+H+	N-(3-carboxypropanoyl)-N-hydroxyputrescine
down in jejunum	189.123	C8H16N2O3+H+	Glycyl-Isoleucine, Glycylleucine, N-Alpha-acetyllys
down in jejunum	197.101	C6H14N4O2+Na+	D-Arginine, L-Arginine
down in jejunum	203.139	C9H18N2O3+H+	Alanyl-Leucine, Alanyl-Isoleucine, Leucyl-Alanine,
down in jejunum	204.134	C8H17N3O3+H+	Lysyl-Glycine, Glycyl-Lysine
down in jejunum	209.081	C6H14N4O2+Cl-	D-Arginine, L-Arginine
down in jejunum	213.075	C6H14N4O2+K+	D-Arginine, L-Arginine
down in jejunum	213.123	C10H16N2O3+H+	L-prolyl-L-proline
down in jejunum	215.033	C6H12O6+Cl-	scyllo-Inositol, L-Sorbose, Alpha-D-Glucose, D-chi
down in jejunum	260.197	C12H25N3O3+H+	Lysyl-Leucine, Lysyl-Isoleucine
down in jejunum	276.155	C11H21N3O5+H+	gamma-Glutamyllysine, Epsilon-(gamma-Glutamy
down in jejunum	283.129	C13H18N2O5+H+	Hydroxy-Iacosamide, Threoninyl-Tyrosine, Tyrosyl
down in jejunum	454.293	C26H43NO5-H2O+Na+	Glycoursodeoxycholic acid, Glycochenodeoxychol
down in jejunum	454.293	C21H44NO7P+H+	PE(16:0/0:0), PC(13:0/0:0), LysoPE(0:0/16:0)
down in jejunum	454.293	C23H45NO5+K+	2-Hydroxyhexadecanoylcarnitine
down in jejunum	472.282	C29H41NO4-H2O+Na+	N-arachidonoyl tyrosine
down in jejunum	472.282	C26H43NO4+K+	Lithocholic acid glycine conjugate
down in jejunum	536.244	C26H45NO7S-H2O+K+	taurocholic acid
down in jejunum	538.260	C26H45NO6S+K+	taurodeoxycholic acid
down in jejunum	561.210	C27H40O11-H2O+K+	Tetrahydroaldosterone-3-glucuronide
down in jejunum	576.180	C30H29N3O7S+H+	7,8-Dihydro-7-hydroxy-8-S-glutathionyl-benzoapyr
down in jejunum	578.215	C30H37NO8+K+	5,9,11-trihydroxyprosta-6E,14Z-dien-1-oate
down in jejunum	590.215	C31H37NO8+K+	Endoxifen O-glucuronide

down in jejunum	592.232	C28H37N3O9S+H+	4-Hydroxyestrone-2-S-glutathione, 2-Hydroxyestro
down in jejunum	597.355	C37H50O5+Na+	3-Benzoyloxy-11-oxo-12-ursen-28-oic acid, 3-Ben
down in jejunum	647.394	C40H56O8-H2O+H+	3-O-trans-Feruloyleuscaphic acid
down in jejunum	677.254	C44H40O8-H2O-H-	2,4-Dihydroxy-6-methoxy-3-(2-hydroxybenzyl)-5-(3
down in jejunum	712.395	C35H66NO10P-H2O+K+	PS(15:1(9Z)/14:0), PS(15:0/14:1(9Z)), PS(14:0/15
down in jejunum	717.346	C31H59O17P-H2O+H+	LPIM1(16:0/0:0)
down in jejunum	739.208	C33H38O19+H+	Malvidin 3-glucoside-5-(6-malonylglucoside), Kaer
down in jejunum	827.291	C45H46O15+H+	6-{4-6-(2,4-dihydroxyphenyl)-2-{4-(E)-2-(2,4-dihydı
down in jejunum	865.247	C45H46O15+K+	6-{4-6-(2,4-dihydroxyphenyl)-2-{4-(E)-2-(2,4-dihydı
down in jejunum	920.206	C30H48N7O18P3S+H+	7-methyl-3-oxooct-6-enoyl-CoA
down in jejunum	939.602	C49H88O15+Na+	Glycerol 2-(9Z,12Z-octadecadienoate) 1-hexadeca
down in jejunum	945.231	C56H42O12+K+	5-4-(2-{3-3-(3,5-dihydroxyphenyl)-6-hydroxy-2-(4-ł
down in jejunum	1049.514	C47H89O19P3-H-	PIP2(22:2(13Z,16Z)/16:0), PIP2(20:2(11Z,14Z)/18
down in jejunum	1112.504	C52H85NO23-H2O+K+	(23R)-Acetoxytomatine
down in jejunum	1114.519	C50H73N15O12+K+	Bradykinin hydroxyproline
down in jejunum	1138.316	C54H59O28-H2O+H+	Cyanidin 3-(6-feruloyl-2-sinapoylsophoroside) 5-gl
down in ileum 1	319.243	C18H38O2Si-H2O+Na+	Trimethylsilyl n-pentadecanoic acid
down in ileum 1	337.253	C24H34O2-H2O+H+	5beta-Chola-3,8(14),11-trien-24-oic Acid, delta2-T
down in ileum 1	337.253	C18H38O2Si+Na+	Trimethylsilyl n-pentadecanoic acid
down in ileum 1	371.258	C24H34O3+H+	(22E)-3beta-Hydroxychola-5,16,22-trien-24-oic Ac
down in ileum 1	373.274	C24H38O4-H2O+H+	3beta,7beta-Dihydroxychol-5-en-24-oic Acid, 7alpł
down in ileum 1	411.251	C24H36O4+Na+	6beta-Hydroxy-3-oxochol-4-en-24-oic Acid, 3alpha
down in ileum 1	411.251	C19H39O7P+H+	LPA(0:0/16:0), PA(16:0/0:0), 1-palmitoylglycerol 3-
down in ileum 1	411.251	C21H40O5+K+	DG(8:0/0:0/10:0), DG(8:0/10:0/0:0), DG(10:0/0:0/8
down in ileum 1	445.235	C24H40O6-H2O+K+	3alpha,7alpha,12alpha,16alpha-tetrahydroxy-5bet
down in ileum 1	456.308	C26H43NO4+Na+	Lithocholic acid glycine conjugate
down in ileum 1	463.272	C27H42N2O3-H2O+K+	N-palmitoyl tryptophan
down in ileum 1	480.279	C26H45NO6S-H2O-H-	taurodeoxycholic acid
down in ileum 1	481.283	C27H42N2O3+K+	N-palmitoyl tryptophan
down in ileum 1	488.298	C26H43NO6+Na+	glycocholic acid
down in ileum 1	496.273	C26H43NO7S-H2O+H+	Sulfoglycolithocholate, Sulfoglycolithocholic acid
down in ileum 1	498.288	C26H45NO7S-H2O+H+	Tauro-b-muricholic acid, Taurohyocholic acid, Tau
down in ileum 1	518.255	C26H43NO7S-H2O+Na+	Sulfoglycolithocholate, Sulfoglycolithocholic acid
down in ileum 1	520.270	C26H45NO7S-H2O+Na+	Tauro-b-muricholic acid, Taurohyocholic acid, Tau
down in ileum 1	522.286	C26H45NO6S+Na+	Tauroursodeoxycholic acid, Taurodeoxycholic acic
down in ileum 1	526.256	C26H42NO9P-H2O+H+	PS(20:5(5Z,8Z,11Z,14Z,17Z)/0:0)
down in ileum 1	536.265	C26H43NO7S+Na+	Sulfoglycolithocholate, Sulfoglycolithocholic acid
down in ileum 1	544.231	C30H37NO8-H2O+Na+	5,9,11-trihydroxyprosta-6E,14Z-dien-1-oate
down in ileum 1	544.267	C26H42NO9P+H+	PS(20:5(5Z,8Z,11Z,14Z,17Z)/0:0)
down in ileum 1	545.236	C27H40O11-H2O+Na+	Tetrahydroaldosterone-3-glucuronide
down in ileum 1	545.236	C24H42O11+K+	(3b,9R)-5-Megastigmene-3,9-diol 9-apiosyl-(1->6)-
down in ileum 1	545.270	C25H40N2O11+H+	Perindopril Acyl-beta-D-glucuronide

down in ileum 1	545.272	C28H42O9+Na+	14-O-(alpha-D-mannopyranosyl)-7S,14S-dihydrox
down in ileum 1	545.272	C23H45O12P+H+	PI(14:0/0:0)
down in ileum 1	552.239	C26H43NO7S+K+	Sulfoglycolithocholate, Sulfoglycolithocholic acid
down in ileum 1	554.255	C26H45NO7S+K+	taurocholic acid
down in ileum 1	557.217	C33H32O8+H+	2-{2,6-dihydroxy-4-6-hydroxy-7-(3-methylbut-2-en-
down in ileum 1	560.204	C30H37NO8-H2O+K+	5,9,11-trihydroxyprosta-6E,14Z-dien-1-oate
down in ileum 1	561.246	C28H42O9+K+	14-O-(alpha-D-mannopyranosyl)-7S,14S-dihydrox
down in ileum 1	562.269	C28H48NO7P-H2O+K+	PC(20:5(5Z,8Z,11Z,14Z,17Z)/0:0)
down in ileum 1	567.293	C26H49O12P-H2O+H+	PI(17:1(9Z)/0:0), PI(17:1(10Z)/0:0)
down in ileum 1	567.293	C28H48O9+K+	campest-22E-en-3beta,4beta,5alpha,6alpha,8beta
down in ileum 1	572.204	C31H37NO8-H2O+K+	Endoxifen O-glucuronide
down in ileum 1	578.245	C26H45NO10S2-H2O+H+	Taurocholic acid 3-sulfate
down in ileum 1	592.209	C29H37N3O8S-H2O+Na+	2-amino-4-({1-(carboxymethyl)-C-hydroxycarbonin
down in ileum 1	661.280	C33H48N2O11S-H2O-H-	6-{(16R)-5,11-dihydroxy-8,8,10,16-tetramethyl-3-1
down in ileum 1	680.286	C32H53NO11S-H2O+K+	(3a,5b)-24-oxo-24-(2-sulfoethyl)aminocholan-3-yl-
down in ileum 1	849.273	C45H46O15+Na+	6-{4-6-(2,4-dihydroxyphenyl)-2-{4-(E)-2-(2,4-dihydr
down in ileum 1	889.264	C56H42O12-H2O+H+	5-4-(2-{3-3-(3,5-dihydroxyphenyl)-6-hydroxy-2-(4-h
down in ileum 1	1006.292	C37H62N7O17P3S-H2O+Na+	Palmitoyl-CoA (n-C16:0CoA), Palmitoyl-CoA (n-C1
down in ileum 1	1080.404	C44H74N7O17P3S-H2O+H+	(2E)-Tricosenoyl-CoA
down in ileum 1	1098.545	C50H73N15O12+Na+	Bradykinin hydroxyproline
down in ileum 1	1099.545	C53H90O22-H2O+K+	Methyl protogracillin, Methylprotodioscin
down in ileum 1	1111.493	C52H80O24+Na+	Medicagenic acid 3-O-b-D-glucuronide 28-O-b-D-g
down in ileum 1	1160.298	C54H59O28-H2O+Na+	Cyanidin 3-(6-feruloyl-2-sinapoylsophoroside) 5-gl
down in ileum 2	90.091	C4H11NO+H+	Dimethylethanolamine
down in ileum 2	106.050	C3H7NO3+H+	2-Amino-3-hydroxypropanoic acid
down in ileum 2	133.061	C4H8N2O3+H+	L-Asparagine, D-Asparagine, Glycyl-glycine, Ureid
down in ileum 2	155.043	C4H8N2O3+Na+	L-Asparagine, D-Asparagine, Glycyl-glycine, Ureid
down in ileum 2	166.053	C5H11NO3S+H+	Methionine sulfoxide
down in ileum 2	173.104	C6H14N4O2-H-	D-Arginine, L-Arginine
down in ileum 2	188.068	C9H11NO2+Na+	L-Phenylalanine
down in ileum 2	188.068	C9H13NO3-H2O+Na+	Normetanephrine, Epinephrine
down in ileum 2	188.068	C6H15NO3+K+	Triethanolamine
down in ileum 2	195.051	C6H12O7-H-	Gulonic acid, Galactonic acid, Gluconic acid
down in ileum 2	205.097	C11H12N2O2+H+	L-Tryptophan, D-Tryptophan
down in ileum 2	229.155	C11H20N2O3+H+	Leucylproline, Isoleucylproline
down in ileum 2	242.066	C10H11NO6+H+	2-{(2,4-dihydroxy-5-methoxyphenyl)(hydroxy)meth
down in ileum 2	248.904	C7H5Cl3O+K+	1,3,5-Trichloro-2-methoxybenzene
down in ileum 2	299.076	C13H14O8+H+	6-(benzoyloxy)-3,4,5-trihydroxyoxane-2-carboxylic
down in ileum 2	353.247	C24H34O3-H2O+H+	3-Oxo-4,6-choladienoic acid
down in ileum 2	355.263	C24H36O3-H2O+H+	(20S)-3beta-Hydroxychola-5,16-dien-24-oic Acid, 3
down in ileum 2	365.105	C12H22O11+Na+	D-Maltose, Turanose, Melibiose, Cellobiose, Sucro
down in ileum 2	405.265	C24H38O5-H-	3alpha,6beta-Dihydroxy-12-oxo-5beta-cholan-24-o

down in ileum 2	407.280	C24H40O5-H-	3beta,7alpha,12beta-Trihydroxy-5alpha-cholan-24
down in ileum 2	429.261	C24H38O5+Na+	3alpha,6beta-Dihydroxy-12-oxo-5beta-cholan-24-c
down in ileum 2	429.264	C26H36O5+H+	2-Angeloyl-9-(3-methyl-2E-pentenoyl)-2b,9a-dihyd
down in ileum 2	431.277	C24H40O5+Na+	3beta,7alpha,12beta-Trihydroxy-5alpha-cholan-24
down in ileum 2	431.279	C26H38O5+H+	17-phenyl-trinor-PGF2alpha isopropyl ester
down in ileum 2	447.251	C27H38O5-H2O+Na+	2,9-Bis(3-methyl-2E-pentenoyl)-2b,9a-dihydroxy-4
down in ileum 2	447.251	C22H39O7P+H+	PA(19:3(10Z,13Z,16Z)/0:0)
down in ileum 2	447.251	C24H40O5+K+	3beta,7alpha,12beta-Trihydroxy-5alpha-cholan-24
down in ileum 2	465.241	C25H41O7P-H2O-H-	LysoPA(22:5(7Z,10Z,13Z,16Z,19Z)/0:0), LysoPA(2
down in ileum 2	465.241	C26H38O5+Cl-	2-(2-Methylbutanoyl)-9-(3-methyl-2E-pentenoyl)-2l
down in ileum 2	494.258	C26H43NO7S-H2O-H-	Sulfoglycolithocholate, Sulfoglycolithocholic acid
down in ileum 2	496.274	C26H45NO7S-H2O-H-	taurocholic acid
down in ileum 2	498.289	C26H45NO6S-H-	Tauroursodeoxycholic acid, Taurodeoxycholic acic
down in ileum 2	510.253	C26H43NO8S-H2O-H-	glycochenodeoxycholic acid 7-sulfate, N-(3a,5b,7b
down in ileum 2	512.269	C26H43NO7S-H-	Sulfoglycolithocholate, Sulfoglycolithocholic acid
down in ileum 2	514.284	C26H45NO7S-H-	Tauro-b-muricholic acid, Taurohyocholic acid, Tau
down in ileum 2	517.294	C26H49O9P-H2O-H-	PG(20:2(11Z,14Z)/0:0)
down in ileum 2	526.248	C26H43NO9S-H2O-H-	N-(3a,5b,7a,12a)-3,7-dihydroxy-24-oxo-12-(sulfoxy
down in ileum 2	528.264	C26H43NO8S-H-	glycochenodeoxycholic acid 7-sulfate, N-(3a,5b,7b
down in ileum 2	530.279	C26H45NO8S-H-	Taurocholate
down in ileum 2	534.239	C25H42NO7P+Cl-	LysoPE(0:0/20:5(5Z,8Z,11Z,14Z,17Z)), LysoPE(20
down in ileum 2	536.255	C25H44NO7P+Cl-	LysoPE(0:0/20:4(5Z,8Z,11Z,14Z)), LysoPE(0:0/20
down in ileum 2	538.281	C26H45NO7S+Na+	Tauro-b-muricholic acid, Taurohyocholic acid, Tau
down in ileum 2	543.256	C23H43O12P+H+	PI(14:1(9Z)/0:0)
down in ileum 2	544.259	C26H43NO9S-H-	N-(3a,5b,7a,12a)-3,7-dihydroxy-24-oxo-12-(sulfoxy
down in ileum 2	573.247	C27H45O12P-H2O-H-	PI(18:4(6Z,9Z,12Z,15Z)/0:0)
down in ileum 2	574.222	C28H37N3O9S-H2O+H+	4-Hydroxyestrone-2-S-glutathione, 2-Hydroxyestro
down in ileum 2	576.237	C28H39N3O9S-H2O+H+	4-Hydroxy-17beta-estradiol-2-S-glutathione
down in ileum 2	582.272	C33H37N5O5-H-	Dihydroergotamine
down in ileum 2	743.304	C37H52O13+K+	25-Acetyl-6,7-didehydrofevicordin F 3-glucoside
down in ileum 2	749.395	C41H63O8P+Cl-	PA(20:5(5Z,8Z,11Z,14Z,17Z)/18:4(6Z,9Z,12Z,15Z)
down in ileum 2	814.491	C40H77NO11S+Cl-	3-O-Sulfogalactosylceramide (d18:1/16:0)
down in ileum 2	913.283	C50H52O15-H2O+K+	6-{2-(4-{6-2,4-dihydroxy-3-(3-methylbut-2-en-1-yl)b
down in ileum 2	929.257	C56H42O12+Na+	5-4-(2-{3-3-(3,5-dihydroxyphenyl)-6-hydroxy-2-(4-h
down in ileum 2	1008.308	C37H64N7O17P3S-H2O+Na+	Palmitelaidoyl-CoA, palmitoleoyl-CoA, (E)-hexade
down in ileum 2	1024.282	C37H64N7O17P3S-H2O+K+	Palmitelaidoyl-CoA, palmitoleoyl-CoA, (E)-hexade
down in ileum 2	1049.301	C62H48O16+H+	6-(3-{4-(E)-2-{3-3-(3,5-dihydroxyphenyl)-6-hydroxy